READ ME for Harry's work on the CRU TS2.1/3.0 datasets, 2006-2009!

1. Two main filesystems relevant to the work:

/cru/dpe1a/f014
/cru/tyn1/f014

Both systems copied in their entirety to /cru/cruts/

Nearly 11,000 files! And about a dozen assorted 'read me' files addressing
individual issues, the most useful being:

fromdpe1a/data/stnmon/doc/oldmethod/f90_READ_ME.txt
fromdpe1a/code/linux/cruts/_READ_ME.txt
fromdpe1a/code/idl/pro/README_GRIDDING.txt

(yes, they all have different name formats, and yes, one does begin '_'!)

2. After considerable searching, identified the latest database files for
tmean:

fromdpe1a/data/cruts/database/+norm/tmp.0311051552.dtb
fromdpe1a/data/cruts/database/+norm/tmp.0311051552.dts

(yes.. that is a directory beginning with '+'!)

3. Successfully ran anomdtb.f90 to produce anomaly files (as per item 7
in the '_READ_ME.txt' file). Had to make some changes to allow for the
move back to alphas (different field length from the 'wc -l' command).

4. Successfully ran the IDL regridding routine quick_interp_tdm.pro
(why IDL?! Why not F90?!) to produce '.glo' files.

5. Currently trying to convert .glo files to .grim files so that we can
compare with previous output. However the progam suite headed by
globulk.f90 is not playing nicely - problems with it expecting a defunct
file system (all path widths were 80ch, have been globally changed to 160ch)
and also no guidance on which reference files to choose. It also doesn't
seem to like files being in any directory other than the current one!!

6. Temporarily abandoned 5., getting closer but there's always another
problem to be evaded. Instead, will try using rawtogrim.f90 to convert
straight to GRIM. This will include non-land cells but for comparison
purposes that shouldn't be a big problem... [edit] noo, that's not gonna
work either, it asks for a 'template grim filepath', no idea what it wants
(as usual) and a serach for files with 'grim' or 'template' in them does
not bear useful fruit. As per usual. Giving up on this approach altogether.

7. Removed 4-line header from a couple of .glo files and loaded them into
Matlab. Reshaped to 360r x 720c and plotted; looks OK for global temp
(anomalies) data. Deduce that .glo files, after the header, contain data
taken row-by-row starting with the Northernmost, and presented as '8E12.4'.
The grid is from -180 to +180 rather than 0 to 360.
This should allow us to deduce the meaning of the co-ordinate pairs used to
describe each cell in a .grim file (we know the first number is the lon or
column, the second the lat or row - but which way up are the latitudes? And
where do the longitudes break?
There is another problem: the values are anomalies, wheras the 'public'
.grim files are actual values. So Tim's explanations (in _READ_ME.txt) are
incorrect..

8. Had a hunt and found an identically-named temperature database file which
did include normals lines at the start of every station. How handy - naming
two different files with exactly the same name and relying on their location
to differentiate! Aaarrgghh!! Re-ran anomdtb:

```
crua6[/cru/cruts/rerun1/data/cruts/rerun1work] ./anomdtb

   > ***** AnomDTB: converts .dtb to anom .txt for gridding *****

   > Enter the suffix of the variable required:
.tmp
   > Select the .cts or .dtb file to load:
tmp.0311051552.dtb
   > Specify the start,end of the normals period:
1961,1990
   > Specify the missing percentage permitted:
25
   > Data required for a normal:           23
   > Specify the no. of stdevs at which to reject data:
3
   > Select outputs (1=.cts,2=.ann,3=.txt,4=.stn):
3
   > Check for duplicate stns after anomalising? (0=no,>0=km range)
8
   > Select the generic .txt file to save (yy.mm=auto):
rr2.txt
   > Select the first,last years AD to save:
1901,2002
   > Operating...
   > Failed to find file.
   > Enter the file, with suffix: .dts
tmp.0311051552.dts
Values loaded: 1255171542;  No. Stations:      12155
   > NORMALS            MEAN percent      STDEV percent
   >         .dtb    5910325    86.6
   >         .cts     575661     8.4    6485986    95.0
   > PROCESS        DECISION percent %of-chk
   > no lat/lon        12043     0.2     0.2
   > no normal        335741     4.9     4.9
   > out-of-range      31951     0.5     0.5
   > duplicated       341323     5.0     5.3
   > accepted        6107721    89.4
   > Dumping years 1901-2002 to .txt files...

crua6[/cru/cruts/rerun1/data/cruts/rerun1work]
```

9. Ran the IDL function:

```
IDL> quick_interp_tdm2,1901,2002,'rr2glofiles/rr2grid.',1200,gs=0.5,dumpglo='dumpglo',pts_prefix='rr2txtfiles/rr2.'
% Compiled module: QUICK_INTERP_TDM2.
% Compiled module: GLIMIT.
Defaults set
    1901
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: STRIP.
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
```

```
    1902
 (etc)
    2002
IDL>
```

This produces anomoly files even when given a normals-added database.. doesn't create the CLIMATOLOGY. However we do have it, both in the 'normals' directory of the user data directory, and in the dpela 'cru_cl_1.0' folder! The relevant file is 'clim.6190.lan.tmp'. Obviously this is for land only.

10. Trying to compare .glo and .grim
Wrote several programs to assist with this process. Tried creating anomalies from the .grim files, using the published climatology. Then tried to compare with the glo files I'd produced (this is all for 1961-1970). Couldn't get a sensible grid layout for the glo files! Eventually resorted to visualisation - looks like the .glo files are 'regular' grid format after all (longitudes change fastest). Don't understand why the comparison program had so much trouble getting matched cells!

11. Decided to concentrate on Norwich. Tim M uses Norwich as the example on the website, so we know it's at (363,286). Wrote a prog to extract the relevant 1961-1970 series from the published output, the generated .glo files, and the published climatology. Prog is norwichtest.for. Prog also creates anomalies from the published data, and raw data from the generated .glo data. Then Matlab prog plotnorwich.m plots the data to allow comparisons.
First result: works perfectly, except that the .glo data is all zeros. This means I still don't understand the structure of the .glo files. Argh!

12. Trying something *else*. Will write a prog to convert the 1961-1970 .glo files to a single file with 120 columns and a row for each non-zero cell. It will be slow. It is a nuisance because the site power os off this weekend (and it's Friday afternoon) so I will get it running at home. Program is glo2vec.for, and yup it is slow. Started a second copy on uealogin1 and it's showing signs of overtaking the crua6 version that started on Friday (it's Tuesday now). I'm about halfway through and the best correlation so far (as tested by norwichcorr.for) is 0.39 at (170,135) (lon,lat).

13. Success! I would crack open a bottle of bubbly but it's only 11.25am. The program norwichcorr.for found a correlation for the norwich series at (363, 286) of 1.00! So we have found the published Norwich series in the grids I produced. A palpable sense of relief pervades the office :-) It's also the grid reference given by Tim for Norwich. So how did I miss it earlier??

14. Wrote a program ('glo2grim.for') to do what I cannot get Tim's 'raw2grim.f90', ie, convert .glo files to GRIM format. It's slow but sure. In parallel, a quick prog called grimcmp.for which compares two GRIM-format files. It produces brief stats. At time of writing, just over 4000 cells have been converted, and the output of grimcmp is:

```
uealogin1[/cru/cruts/rerun1/data/cruts/rerun1work] ./grimcmp

Welcome to the GRIM Comparer

Please enter the first grim file (must be complete!):  cru_ts_2_10.1961-1970.tmp
Please enter the second grim file (may be incomplete): glo2grim1.out

File glo2grim1.out terminated prematurely after     4037 records.

SUMMARY FROM GRIMCMP

Files compared:
1. cru_ts_2_10.1961-1970.tmp
2. glo2grim1.out

Total Cells Compared              4037
Total 100% Matches                   0
Cells with Corr. == 1.00             0  ( 0.0%)
Cells with 0.90<=Corr<=0.99       3858  (95.6%)
Cells with 0.80<=Corr<=0.89        119  ( 2.9%)
Cells with 0.70<=Corr<=0.79         25  ( 0.6%)
```

..which is good news! Not brilliant because the data should be identical.. but good because the correlations are so high! This could be a result of my mis-setting of the parameters on Tim's programs (although I have followed his recommendations wherever possible), or it could be a result of Tim using the Beowulf 1 cluster for the f90 work. Beowulf 1 is now integrated in to the latest Beowulf cluster so it may not be practical to test that theory.

15. All change! My 'glo2grim1' program was presciently named as it's now up to v3! My attempt to speed up early iterations by only reading as much of each glo file as was needed was really stupidly coded and hence the poor results. Actually they're worryingly good as the data was effectively random :-0
We are now on-beam and initial results are very very promising:

```
uealogin1[/cru/cruts/rerun1/data/cruts/rerun1work] ./grimcmp3x

File glo2grim3.out terminated prematurely after      143 records.

SUMMARY FROM GRIMCMP

Files compared:
1. cru_ts_2_10.1961-1970.tmp
2. glo2grim3.out

Total Cells Compared               143
Total 100% Matches                  12
Cells with Corr. == 1.00            12  ( 8.4%)
Cells with 0.96<=Corr<=0.99        130  (90.9%)
Cells with 0.90<=Corr<=0.95          1  ( 0.7%)
Cells with 0.80<=Corr<=0.89          0  ( 0.0%)
Cells with 0.70<=Corr<=0.79          0  ( 0.0%)
```

..so all correlations are >= 0.9 and all but one are >=0.96! with 12 complete (100% identical) matches I think we can safely say we are producing the data Tim produced. The variations can

```
be accounted for as rounding errors due to different hardware
and compilers, I reckon..

16. So, it seemed like a good time to start a Precip run. With
a bit of luck this would go as smoothly as the Temperature run,
ho, ho, ho. The first problem was that anomdtb kept crashing:

crua6[/cru/cruts/rerun1/data/cruts/rerun2work] ./anomdtb

   > ***** AnomDTB: converts .dtb to anom .txt for gridding *****

   > Enter the suffix of the variable required:
.pre
   > Will calculate percentage anomalies.
   > Select the .cts or .dtb file to load:
pre.0312031600.dtb
   > Specify the start,end of the normals period:
1961,1990
   > Specify the missing percentage permitted:
25
   > Data required for a normal:           23
   > Specify the no. of stdevs at which to reject data:
3
   > Select outputs (1=.cts,2=.ann,3=.txt,4=.stn):
3
   > Check for duplicate stns after anomalising? (0=no,>0=km range)
8
   > Select the generic .txt file to save (yy.mm=auto):
rr2pre.txt
   > Select the first,last years AD to save:
1901,2002
   > Operating...
Values loaded: 1258818288;  No. Stations:      12732
   > NORMALS            MEAN percent      STDEV percent
   >         .dtb    2635549    29.6
forrtl: error (75): floating point exception
IOT trap (core dumped)
crua6[/cru/cruts/rerun1/data/cruts/rerun2work]

..not good! Tried recompiling for uealogin1.. AARGGHHH!!! Tim's
code is not 'good' enough for bloody Sun!! Pages of warnings and
27 errors! (full results in 'anomdtb.uealogin1.compile.results').

17. Inserted debug statements into anomdtb.f90, discovered that
a sum-of-squared variable is becoming very, very negative! Key
output from the debug statements:

OpEn=   16.00, OpTotSq=    4142182.00, OpTot= 7126.00
DataA val =       93, OpTotSq=       8649.00
DataA val =      172, OpTotSq=      38233.00
DataA val =      950, OpTotSq=     940733.00
DataA val =      797, OpTotSq=    1575942.00
DataA val =      293, OpTotSq=    1661791.00
DataA val =       83, OpTotSq=    1668680.00
DataA val =      860, OpTotSq=    2408280.00
DataA val =      222, OpTotSq=    2457564.00
DataA val =      452, OpTotSq=    2661868.00
DataA val =      561, OpTotSq=    2976589.00
DataA val =    49920, OpTotSq=-1799984256.00
DataA val =      547, OpTotSq=-1799684992.00
DataA val =      672, OpTotSq=-1799233408.00
DataA val =      710, OpTotSq=-1798729344.00
DataA val =      211, OpTotSq=-1798684800.00
DataA val =      403, OpTotSq=-1798522368.00
OpEn=   16.00, OpTotSq=-1798522368.00, OpTot=56946.00
forrtl: error (75): floating point exception
IOT trap (core dumped)

..so the data value is unbfeasibly large, but why does the
sum-of-squares parameter OpTotSq go negative?!!

Probable answer: the high value is pushing beyond the single-
precision default for Fortran reals?

Value located in pre.0312031600.dtb:

-400002  3513   3672  309 HAMA                SYRIA         1985 2002   -999     -999
6190  842  479 3485  339  170  135  106    0    9  243  387  737
1985  887  582   93   16   17    0    0    0    0  352  221  627
1986  899  252  172  527  173   30    0    0    0   84  496  570
1987  578  349  950  191    4    0    0    0    0  343  462  929
1988 1044  769  797  399   11  903  218    0    0  163  517 1181
1989  269   62  293    3   13    0    0    0    0  101  292  342
1990  328  276   83  135  224    0    0    0    0   87  343  230
1991 1297  292  860  320   70    0    0    0    0  206  298  835
1992  712 1130  222   39  339  301    0    0    0    0  909  351
1993  726  609  452   82  672    3    0    0    0   34  183  351
1994  625  661  561   41  155    0    0    0   22  345  953 1072
1995  488-9999-9999  182-9999    0-9999    0    0    0  754-9999
1996-9999  40949920-9999   82    0-9999    0   36  414  112  312
1997-9999  339  547-9999  561-9999    0    0   54  155  265  962
1998 1148  289  672  496-9999    0    0-9999    9   21-9999 1206
1999  343  379  710  111    0    0    0-9999-9999-9999  132  285
2000 1518  399  211  354   27    0-9999    0   27  269  316 1057
2001  370-9999-9999  273  452    0-9999-9999-9999  290  356-9999
2002  871  329  403  111  233-9999    0    0-9999-9999  377 1287

(value is for March 1996)

Action: value replaced with -9999 and file renamed:

pre.0312031600H.dtb   (to indicate I've fixed it)

.dts file also renamed for consistency.

anomdtb then runs fine!! Producing the usual txt files.

18. Ran the IDL gridding routine for the precip files:

quick_interp_tdm2,1901,2002,'rr2preglofiles/rr2pregrid.',450,gs=0.5,dumpglo='dumpglo',pts_prefix='rr2pretxtfiles/rr2pre.'

..and this is where it gets CRAZY. Instead of running normally,
this time I get:

IDL> quick_interp_tdm2,1901,1910,'rr2glofiles2/rr2grid.',1200,gs=0.5,dumpglo='dumpglo',pts_prefix='rr2txtfiles/rr2.'

limit=glimit(/all) ; sets limit to global field
               ^
% Syntax error.
  At: /cru/cruts/fromdpe1a/code/idl/pro/quick_interp_tdm2.pro, Line 38
```

```
    lim=glimit(/all)
               ^
% Syntax error.
  At: /cru/cruts/fromdpe1a/code/idl/pro/quick_interp_tdm2.pro, Line 122

    r=area_grid(pts2(n,1),pts2(n,0),pts2(n,2),gs*2.0,bounds,dist,angular=angular)
                                                               ^
% Syntax error.
  At: /cru/cruts/fromdpe1a/code/idl/pro/quick_interp_tdm2.pro, Line 183
% Compiled module: QUICK_INTERP_TDM2.
% Attempt to call undefined procedure/function: 'QUICK_INTERP_TDM2'.
% Execution halted at:  $MAIN$
IDL>

.. WHAT?! Now it's not precompiling its functions for some reason!
What's more - I cannot find the 'glimit' function anywhere!!

Eventually (the following day) I found glimit and area_grid, they are
in Mark New's folder: /cru/u2/f080/Idl. Since this is in $IDL_PATH I
have no idea why they're not compiling! I manually compiled them with
.compile, and the errors vanished! Though not for long:

IDL> .compile /cru/u2/f080/Idl/glimit.pro
% Compiled module: GLIMIT.
IDL> .compile /cru/u2/f080/Idl/area_grid.pro
% Compiled module: AREA_GRID.
IDL> quick_interp_tdm2,1901,1910,'rr2glofiles2/rr2grid.',1200,gs=0.5,dumpglo='dumpglo',pts_prefix='rr2txtfiles/rr2.'
% Compiled module: QUICK_INTERP_TDM2.
Defaults set
    1901
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Variable is undefined: STRIP.
% Execution halted at:  QUICK_INTERP_TDM2  215 /cru/cruts/fromdpe1a/code/idl/pro/quick_interp_tdm2.pro
%                       $MAIN$
IDL>

Was this a similar problem? Unfortunately not:

IDL> .compile /cru/u2/f080/Idl/strip.pro
% Compiled module: STRIP.
IDL> quick_interp_tdm2,1901,1910,'rr2glofiles2/rr2grid.',1200,gs=0.5,dumpglo='dumpglo',pts_prefix='rr2txtfiles/rr2.'
Defaults set
    1901
% Variable is undefined: STRIP.
% Execution halted at:  QUICK_INTERP_TDM2  215 /cru/cruts/fromdpe1a/code/idl/pro/quick_interp_tdm2.pro
%                       QUICK_INTERP_TDM2  215 /cru/cruts/fromdpe1a/code/idl/pro/quick_interp_tdm2.pro
%                       $MAIN$
IDL>

..so it looks like a path problem. I wondered if the NFS errors that have
been plagueing crua6 work for some time now might have prevented IDL from
adding the correct directories to the path? After all the help file does
mention that IDL discards any path entries that are inaccessible.. so if
the timeout is a few seconds that would explain it. So I restarted IDL,
and PRESTO! It worked. I then tried the precip veriosn - and it worked
too!

IDL> quick_interp_tdm2,1901,2002,'rr2preglofiles/rr2pregrid.',450,gs=0.5,dumpglo='dumpglo',pts_prefix='rr2pretxtfiles/rr2pre.'
% Compiled module: QUICK_INTERP_TDM2.
% Compiled module: GLIMIT.
Defaults set
    1901
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: STRIP.
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
    1902
(etc)
    2001
    2002
IDL>

I then ran glo2grim4.for to convert from percentage anomalies to real
(10ths of a mm) values. Initial results are not as good as temperature,
but mainly above 0.96 so obviously on the right track.

However..

19. Here is a little puzzle. If the latest precipitation database file
contained a fatal data error (see 17. above), then surely it has been
altered since Tim last used it to produce the precipitation grids? But
if that's the case, why is it dated so early? Here are the dates:

/cru/dpe1a/f014/data/cruts/database/+norm/pre.0312031600.dtb
- directory date is 23 Dec 2003

/cru/tyn1/f014/ftpfudge/data/cru_ts_2.10/data_dec/cru_ts_2_10.1961-1970.pre.Z
- directory date is 22 Jan 2004 (original date not preserved in zipped file)
- internal (header) date is also '22.01.2004 at 17:57'

So what's going on? I don't see how the 'final' precip file can have been
produced from the 'final' precipitation database, even though the dates
imply that. The obvious conclusion is that the precip file must have been
produced before 23 Dec 2003, and then redated (to match others?) in Jan 04.

20. Secondary Variables - Eeeeeek!! Yes the time has come to attack what even
Tim seems to have been unhappy about (reading between the lines). To assist
me I have 12 lines in the gridding ReadMe file.. so par for the course.
Almost immediately I hit that familiar feeling of ambiguity: the text
suggests using the following three IDL programs:
     frs_gts_tdm.pro
     rd0_gts_tdm.pro
     vap_gts_anom.pro
So.. when I look in the code/idl/pro/ folder, what do I find? Well:

 3447 Jan 22  2004 fromdpe1a/code/idl/pro/frs_gts_anom.pro
 2774 Jun 12  2002 fromdpe1a/code/idl/pro/frs_gts_tdm.pro

 2917 Jan  8  2004 fromdpe1a/code/idl/pro/rd0_gts_anom.pro
 2355 Jun 12  2002 fromdpe1a/code/idl/pro/rd0_gts_tdm.pro

 5880 Jan  8  2004 fromdpe1a/code/idl/pro/vap_gts_anom.pro

In other words, the *anom.pro scripts are much more recent than the *tdm
scripts. There is no way of knowing which Tim used to produce the current
public files. The scripts differ internally but - you guessed it! - the
descriptions at the start are identical. WHAT IS GOING ON? Given that the
```

```
'README_GRIDDING.txt' file is dated 'Mar 30  2004' we will have to assume
that the originally-stated scripts must be used.

To begin with, we need binary output from quick_interp_tdm2, so it's run
again for tmp and pre, and (for the first time) for dtr. This time, the
command line looks like this for tmp:
IDL> quick_interp_tdm2,1901,2002,'idlbinout/idlbin',1200,gs=2.5,dumpbin='dumpbin',pts_prefix='tmp_txt_4idl/tmp.'
This gives screen output for each year, typically:
    1991
grid 1991 non-zero    0.9605    2.0878    2.1849 cells=    27048
And produces output files (in, in this case, 'idlbinout/'), like this:
-rw-------   1 f098     cru       248832 Sep 21 12:20 idlbin_tmp/idlbin_tmp1991

At this point, did some logical renaming. So..
.txt files (pre-IDL) are typically 'tmp.1901.01.txt' in 'tmp_txt_4idl/'
binary files (post-IDL) are typically 'idlbin_tmp1991' in 'idlbin_tmp/'.
These changes rolled back to the quoted command lines, to avoid confusion.

Next, precip command line:
IDL> quick_interp_tdm2,1901,2002,'idlbin_pre/idlbin_pre',450,gs=2.5,dumpbin='dumpbin',pts_prefix='pre_txt_4idl/pre.'
(note new filenaming schema)
This gives example screen output:
    1991
grid 1991 non-zero   -4.8533   36.2155   51.0738 cells=    51060
And produces output files like:
-rw-------   1 f098     cru       248832 Sep 21 12:50 idlbin_pre/idlbin_pre1991

Finally for the primaries, the first stab at dtr. Ran anomdtb with the
database file dtr.0312221128.dtb, and the standard/recommended responses.
Screen output:
   > NORMALS            MEAN percent      STDEV percent
   >         .dtb          0      0.0
   >         .cts    3375441    84.1    3375441    84.1
   > PROCESS        DECISION percent %of-chk
   > no lat/lon         3088     0.1     0.1
   > no normal        638538    15.9    15.9
   > out-of-range      70225     1.7     2.1
   > duplicated       135457     3.4     4.1
   > accepted        3167636    78.9
   > Dumping years 1901-2002 to .txt files...

Then for the gridding:
IDL> quick_interp_tdm2,1901,2002,'idlbin_dtr/idlbin_dtr',750,gs=2.5,dumpbin='dumpbin',pts_prefix='dtr_txt_4idl/dtr.'
Giving screen output:
    1991
grid 1991 non-zero   -0.3378    1.6587    1.7496 cells=     3546
And files such as:
-rw-------   1 f098     cru       248832 Sep 21 13:39 idlbin_dtr/idlbin_dtr1991

And.. at this point, I read the ReadMe file properly. I should be gridding at
2.5 degrees not 0.5 degrees! For some reason, secondary variables are not
derived from the 0.5 degree grids. Re-did all three generations (the sample
command lines and outputs above have been altered to reflect this, to avoid
confusion).

So, to the generation of the synthetic grids.

Tried running frs_gts_tdm but it complained it couldn't find the normals file:

IDL> frs_gts_tdm,dtr_prefix='idlbin_dtr/idlbin_dtr',tmp_prefix='idlbin_tmp/idlbin_tmp',1901,2002,outprefix='syngrid_frs/syngrid_frs'
% Compiled module: FRS_GTS_TDM.
% Attempt to call undefined procedure/function: 'FRS_GTS_TDM'.
% Execution halted at:  $MAIN$
IDL> frs_gts,dtr_prefix='idlbin_dtr/idlbin_dtr',tmp_prefix='idlbin_tmp/idlbin_tmp',1901,2002,outprefix='syngrid_frs/syngrid_frs'
% Compiled module: RDBIN.
% Compiled module: STRIP.
ls: /home/cru/f098/ml/gts/frs/glo/glo.frs.norm not found
ls: /home/cru/f098/ml/gts/frs/glo/glo.frs.norm.Z not found
ls: /home/cru/f098/ml/gts/frs/glo/glo.frs.norm.gz not found
% READF: End of file encountered. Unit: 99, File: foo
% Execution halted at:  RDBIN              25 /cru/u2/f080/Idl/rdbin.pro
%                       FRS_GTS            18 /cru/cruts/fromdpela/code/idl/pro/frs_gts_tdm.pro
%                       $MAIN$
IDL>

However when I eventually found what I hope is the normals file:

/cru/cruts/fromdpela/data/grid/twohalf/glo25.frs.6190

..and altered the IDL prog to read it.. same error! Turns out it's preferring
to pick up Mark N's version so tried explicitly compiling,
('.compile xxxxxx.pro') that worked, in that the error changed:

IDL> frs_gts,dtr_prefix='idlbin_dtr/idlbin_dtr',tmp_prefix='idlbin_tmp/idlbin_tmp',1901,2002,outprefix='syngrid_frs/syngrid_frs'
% Compiled module: RDBIN.
% Compiled module: STRIP.
yes
% Variable is undefined: NF.
% Execution halted at:  RDBIN              68 /cru/u2/f080/Idl/rdbin.pro
%                       FRS_GTS            21 /cru/cruts/fromdpela/code/idl/pro/frs_gts_tdm.pro
%                       $MAIN$
IDL>

So what is this mysterious variable 'nf' that isn't being set? Well strangely,
it's in Mark N's 'rdbin.pro'. I say strangely because this is a generic prog
that's used all over the place! Nonetheless it does have what certainly looks
like a bug:

    38   if keyword_set(gridsize) eq 0 then begin
    39    info=fstat(lun)
    40    if keyword_set(seas) then info.size=info.size*2.0
    41    if keyword_set(ann) then info.size=info.size*12.0
    42    nlat=sqrt(info.size/48.0)
    43    gridsize=180.0/nlat
    44    if keyword_set(quiet) eq 0 then print,'filesize=',info.size
    45    if keyword_set(quiet) eq 0 then print,'gridsize=',gridsize
    46   endif
    47   if keyword_set(had) then had=1 else had=0
    48   if keyword_set(echam) then echam=1 else echam=0
    49   if keyword_set(gfdl) then gfdl=1 else gfdl=0
    50   if keyword_set(ccm) then ccm=1 else ccm=0
    51   if keyword_set(csiro) then csiro=1 else csiro=0
    52  ;create array to read data into
    53   if keyword_set(seas) then nf=6 else nf=12
    54   if keyword_set(ann) then nf=1
    55   defxyz,lon,lat,gridsize,grid=grid,nf=nf,had=had,echam=echam,gfdl=gfdl,ccm=ccm,csiro=csiro
    56   if keyword_set(quiet) eq 0 then help,grid
    57   grid=fix(grid)
    58  ;read data
    59   readu,lun,grid
```

```
    60   close,lun
    61   spawn,string('rm -f ',fff)
    62  endif else begin
    63   openr,lun,fname
    64  ; check file size and work out grid spacing if gridsize isn't set
    65   if keyword_set(gridsize) eq 0 then begin
    66    info=fstat(lun)
    67    if keyword_set(quiet) eq 0 then print,'yes'
    68    nlat=sqrt((info.size/nf)/4.0)
    69    gridsize=180.0/nlat
    70    if keyword_set(quiet) eq 0 then print,'filesize=',info.size
    71    if keyword_set(quiet) eq 0 then print,'gridsize=',gridsize
    72   endif
    73   if keyword_set(seas) then nf=6.0 else nf=12.0
    74   if keyword_set(ann) then nf=1
```

In other words, 'nf' is set in the first conditional set of statements, but in
the alternative (starting on #62) it is only set AFTER it's used
(set #73,#74; used #68). So I shifted #73 and #74 to between #64 and #65, and..
with precompiling to pick up the local version of rdbin, too.. it worked!
Er, perhaps.

Lots of screen output, and lots of files. A set of synthetic grids in 'syngrid_frs/' as requested, typically:

```
-rw-------   1 f098     cru        20816 Sep 17 22:10 syngrid_frs/syngrid_frs1991.Z
```

..but also a set of some binariy files in the working directory! They look like this:

```
-rw-------   1 f098     cru        51542 Sep 17 22:10 glo.frs.1991.Z
```

Having read the program it looks as though the latter files are absolutes,
whereas the former are anomalies. With this in mind, they are renamed:

```
glo.frs.1991 -> glo.frs.abs.1991
```

..and put into folder syngrid_frs_abs/

Then - a real setback. Looked for a database file for frost.. nothing. Is
this a real secondary parameter? Answer: yes. Further digging revealed that
quick_interp_tdm2.pro has a 'nostn' command line option. It's undocumented,
as usual, but it does seem to avoid the use of the 'pts_prefix' option.. so
I set it, and it at least *ran* for the full term (though very slow compared
to primary variables)!

```
IDL> quick_interp_tdm2,1901,2002,'glo_frs_grids/frs.grid.',750,gs=0.5,dumpglo='dumpglo',nostn=1,synth_prefix='syngrid_frs/syngrid_frs'
```

It does produce output grids. Without converting to absolutes with the normals file,
it's hard to know if they're realistic.

Then, I moved on to rd0 (wet-day frequency). This time, when I searched for the
normals files required ('glo.pre.norm' and 'glo.rd0.norm'), I could not (as before)
find exact matches. The difference this time is that the program checks that the
normals file supplied is a 0.5-degree grid, so glo25.pre.6190 failed. This implies
to me that my approach to frs (above) was wrong as well. Where is the documenatation
to explain all this?!

Finally - a breakthrough. A search of Mark New's old directory hierarchy revealed
what look like the required files:

```
crua6[/cru/mark1/f080] find . -name 'glo.*.norm*'
./gts/cld/glo/glo.cld.norm.Z
./gts/dtr/glo_old/glo.dtr.norm.Z
./gts/frs/glo.frs.norm.Z
./gts/frs/glo/glo.frs.norm.Z
find: cannot open < ./gts/frs/glo_txt >
./gts/pre/glo_quick_abs/glo.pre.norm.Z
./gts/pre/glo_quick_log/glo.pre.norm.Z
./gts/pre/glo_spl/glo.pre.norm.Z
find: cannot open < ./gts/pre_perc/station_list >
./gts/rad/glo/glo.rad.norm.Z
./gts/rd0/glo/glo.rd0.norm.Z
./gts/rd0/glo_old/glo.rd0.norm.Z
./gts/sunp/glo/glo.sunp.norm
./gts/sunp/means/glo.sunp.norm.Z
./gts/tmp/glo/glo.tmp.norm.Z
./gts/tmp/glo_old/glo.tmp.norm.Z
find: cannot open < ./gts/tmp/station_list >
./gts/vap/glo/glo.vap.norm.Z
./gts/wnd/glo/glo.wnd.norm.Z
```

A listing of /cru/mark1/f080/gts gives:

```
drwxr-x---   2 f080     cru         1024 Sep 12  2005 cdrom
drwxr-x---  10 f080     cru        57344 Nov  1  2001 cld
drwxr-xr-x  19 f080     cru        24576 Feb 27  2001 dtr
drwxr-x---   2 f080     cru         8192 Feb 25  1998 elev
drwxr-x---   2 f080     cru         8192 Jun  8  1998 euroclivar
-rw-r-----   1 f080     cru            0 Aug  3  1999 foo
drwxr-x---   6 f080     cru         8192 Aug  6  2002 frs
-rw-r-x---   1 f080     cru          438 May 12  1998 gts.errors
-rw-r-----   1 f080     cru           10 Jul 21  1999 in
drwxr-x---   5 f080     cru         8192 Jan  6  1999 jiang
drwxr-x---   2 f080     cru         8192 Apr  7  1998 landsea
-rw-r-----   1 f080     cru          240 May 12  1998 normal.errors
drwxr-x---   5 f080     cru         8192 Aug  6  2002 plots
drwxr-xr-x  12 f080     cru       106496 May 22  2000 pre
drwxr-x---   9 f080     cru       114688 Aug  6  2002 pre_perc
drwxr-x---   4 f080     cru         1024 Jan  6  1999 rad
drwxr-x--x   6 f080     cru         8192 Nov  1  2001 rd0
-rwxr-xr--   1 f080     cru         1779 Dec  5  1997 readme.txt
drwxr-x---   8 f080     cru         1024 Apr  5  2000 reg_series
drwxr-x---   3 f080     cru         1024 Oct 18  1999 reh
drwxr-x---   2 f080     cru         8192 Jan 19  2000 scengen
drwxr-x---   5 f080     cru        24576 Nov  5  1998 sunp
drwxr-x---   2 f080     cru         1024 Aug  6  2002 test
drwxr-x---   4 f080     cru         1024 Aug  3  1999 tmn
drwxr-xr-x  20 f080     cru       122880 Mar 19  2002 tmp
drwxr-x---   4 f080     cru         1024 Aug  3  1999 tmx
drwxr-x---   6 f080     cru         1024 Jul  8  1998 ukcip
drwxr-x---   5 f080     cru         8192 Nov  5  2001 vap
drwxr-x---   4 f080     cru         1024 Jul  2  1998 wnd
```

And a listing of, for example, the 'frs' directory:

```
drwxr-x---   2 f080     cru        16384 Jul 18  2002 glo
-rw-r-x---   1 f080     cru       433393 Aug 12  1998 glo.frs.1961.Z
-rw-r-x---   1 f080     cru       321185 Aug 12  1998 glo.frs.ano.1961.Z
-rw-r-x---   1 f080     cru       740431 Aug 12  1998 glo.frs.norm.Z
drwxr-xr-x   2 f080     cru        16384 Jul 27  1999 glo25
drwx------   2 f080     cru         8192 Jul 18  2002 glo_txt
```

```
drwxr-xr-x   2 f080     cru         8192 Aug 28  1998 means

So, the following were copied to the working area:

cp /cru/mark1/f080/gts/frs/glo.frs.norm.Z /cru/cruts/rerun1/data/cruts/rerun_synth/
cp /cru/mark1/f080/gts/cld/glo/glo.cld.norm.Z /cru/cruts/rerun1/data/cruts/rerun_synth/
cp /cru/mark1/f080/gts/dtr/glo_old/glo.dtr.norm.Z /cru/cruts/rerun1/data/cruts/rerun_synth/

precip looked like it might be a problem (3 matching files, see above),
but on investigation they were found to be identical! Wonderful.

cp /cru/mark1/f080/gts/pre/glo_quick_log/glo.pre.norm.Z /cru/cruts/rerun1/data/cruts/rerun_synth/
cp /cru/mark1/f080/gts/rad/glo/glo.rad.norm.Z /cru/cruts/rerun1/data/cruts/rerun_synth/
cp /cru/mark1/f080/gts/rd0/glo/glo.rd0.norm.Z /cru/cruts/rerun1/data/cruts/rerun_synth/

There were two 'sunp' norm files, but one was 0 bytes in length.

cp /cru/mark1/f080/gts/sunp/means/glo.sunp.norm.Z /cru/cruts/rerun1/data/cruts/rerun_synth/
cp /cru/mark1/f080/gts/tmp/glo/glo.tmp.norm.Z /cru/cruts/rerun1/data/cruts/rerun_synth/
cp /cru/mark1/f080/gts/vap/glo/glo.vap.norm.Z /cru/cruts/rerun1/data/cruts/rerun_synth/
cp /cru/mark1/f080/gts/wnd/glo/glo.wnd.norm.Z /cru/cruts/rerun1/data/cruts/rerun_synth/

The synthetics generation was then re-run for frs (records above have
been modified to reflect this).

Next, rd0. Synthetics generated OK..

IDL> rd0_gts,1901,2002,1961,1990,outprefix="syngrid_rd0/syngrid_rd0",pre_prefix="idlbin_pre/idlbin_pre"

..until the end:

    2001
yes
filesize=      248832
gridsize=      2.50000
    2002
yes
filesize=      248832
gridsize=      2.50000
% Program caused arithmetic error: Floating divide by 0
% Program caused arithmetic error: Floating illegal operand
IDL>

However, all synthetic grids appear to have been written OK, including 2002.

Grid generation proceeded without error:

IDL> quick_interp_tdm2,1901,2002,'glo_rd0_grids/rd0.grid.',450,gs=0.5,dumpglo='dumpglo',nostn=1,synth_prefix='syngrid_rd0/syngrid_rd0'


Onto vapour pressure, and the crunch. For here, the recommended program for
synthetic grid production is 'vap_gts_anom.pro'. In fact, there is no sign
of a 'vap_gts_tdm.pro'. And, in the program notes, it reads:

; required inputs are:
; ** vapour pressure and temperature normals on 2.5deg grid
;       (these come ready-supplied for a 1961-90 normal period)
; ** temp and dtr monthly anomalies on 2.5deg grid, including normal period

So, we face a situation where some synthetics are built with 0.5-degree
normals, and others are built with 2.5-degree normals. I can find no
documentation of this. There are '*_anom.pro' versions of the frs and rd0
programs, both of which use 2.5-degree normals, however they are dated
Jan 2004, and Tim's Read_Me (which refers to the '*_tdm.pro' 0.5-degree
versions) is dated end March 2004, so we have to assume these are his
best suggestions.

The 2.5 normals are found here:

> ls -l /cru/cruts/fromdpe1a/data/grid/twohalf/
total 1248
-rwxr-xr-x   1 f098     cru       248832 Jan  9  2004 glo25.frs.6190
-rwxr-xr-x   1 f098     cru       248832 Jan  8  2004 glo25.pre.6190
-rwxr-xr-x   1 f098     cru       248832 Jan  8  2004 glo25.rd0.6190
-rwxr-xr-x   1 f098     cru       248832 Jan  7  2004 glo25.tmp.6190
-rwxr-xr-x   1 f098     cru       248832 Jan  6  2004 glo25.vap.6190
-rwxr-xr-x   1 f098     cru           86 Feb 25  2004 readme.txt

readme.txt:
2.5deg climatology files
Tim Mitchell, 25.2.04

These are in Mark New's binary format
(end)

Set up the required inputs, and ran it:

IDL> vap_gts_anom,dtr_prefix='idlbin_dtr/idlbin_dtr',tmp_prefix='idlbin_tmp/idlbin_tmp',1901,2002,outprefix='syngrid_vap/syngrid_vap',dumpbin=1

Producing screen output like this:
1991 vap (x,s2,<<,>>):  0.000493031  0.000742087   -0.0595093      1.86497

And output files like this:
-rw-------   1 f098     cru       248832 Sep 22 10:56 syngrid_vap/syngrid_vap1991

On, without further ado, to the gridding. For this secondary, there *are* database
files, so the 'nostn' option is not used, and anomdtb.f is wheeled out again
to construct .txt files for the run:

crua6[/cru/cruts/rerun1/data/cruts/rerun_vap] ./anomdtb

   > ***** AnomDTB: converts .dtb to anom .txt for gridding *****

   > Enter the suffix of the variable required:
.vap
   > Select the .cts or .dtb file to load:
vap.0311181410.dtb
   > Specify the start,end of the normals period:
1961,1990
   > Specify the missing percentage permitted:
25
   > Data required for a normal:           23
   > Specify the no. of stdevs at which to reject data:
3
   > Select outputs (1=.cts,2=.ann,3=.txt,4=.stn):
3
   > Check for duplicate stns after anomalising? (0=no,>0=km range)
8
   > Select the generic .txt file to save (yy.mm=auto):
vap.txt
```

```
   > Select the first,last years AD to save:
1901,2002
   > Operating...
Values loaded: 1239868112;  No. Stations:       7691
   > NORMALS            MEAN percent      STDEV percent
   >         .dtb     887754    46.9
   >         .cts      34175     1.8     921929    48.7
   > PROCESS        DECISION percent %of-chk
   > no lat/lon          105     0.0     0.0
   > no normal        969384    51.3    51.3
   > out-of-range       2661     0.1     0.3
   > duplicated        25557     1.4     2.8
   > accepted         893711    47.3
   > Dumping years 1901-2002 to .txt files...

crua6[/cru/cruts/rerun1/data/cruts/rerun_vap]

Moved straight onto the gridding, which, of course, failed:

IDL> quick_interp_tdm2,1901,2002,'glo_vap_grids/vap.grid.',1000,gs=0.5,dumpglo='dumpglo',synth_prefix='syngrid_vap/syngrid_vap',pts_prefix='../rerun_vap/vap_txt_4idl/vap.'
Defaults set
    1901
    1902
% Array dimensions must be greater than 0.
% Execution halted at:  QUICK_INTERP_TDM2   88 /cru/cruts/fromdpe1a/code/idl/pro/quick_interp_tdm2.pro
%                       QUICK_INTERP_TDM2   88 /cru/cruts/fromdpe1a/code/idl/pro/quick_interp_tdm2.pro
%                       $MAIN$
IDL>

This turns out to be because of the sparcity of VAP station measurements in the
early years. The program cannot handle anom files of 0 length, even though it
checks the length! Bizarre. The culprit is 'vap.1902.03.txt', the only month to
have no station reading at all (45 months have only 1 however). I decided to mod
the program to use the 'nostn' option if the length is 0. Hope that's right - the
synthetics are read in first and the station data is added to that grid so this
should be OK.. and it looks OK:

IDL> quick_interp_tdm2,1901,2002,'vap.grid.',1000,gs=0.5,dumpglo='dumpglo',synth_prefix='syngrid_vap/syngrid_vap',pts_prefix='../rerun_vap/vap_txt_4idl/vap.'
% Compiled module: GLIMIT.
Defaults set
    1901
    1902
no stations found in: ../rerun_vap/vap_txt_4idl/vap.1902.03.txt
    1903

(..etc..)

Pause for reflection: the list of CRU_TS_2.1 parameters is as follows:
pre  primary, done
tmp  primary, done
tmx  derived, not done
tmn  derived, not done
dtr  primary, done
vap  secondary, done
cld/spc  secondary, not done
wet  secondary, done
frs  secondary, done

Now the interesting thing is that the 'Read Me' file for gridding only
mentions frs, rd0 (which I'm assuming == wet) and vap. How, then, do I
produce cld/spc and the two derived vars??

Well, there's a /cru/cruts/fromdpe1a/code/idl/pro/cal_cld_gts_tdm.pro,
also:
/cru/cruts/fromdpe1a/code/idl/pro/cloudcorrspc.pro
/cru/cruts/fromdpe1a/code/idl/pro/cloudcorrspcann.pro
/cru/cruts/fromdpe1a/code/idl/pro/cloudcorrspcann9196.pro

Loading just the first program opens up another huge can o' worms. The
program description reads:

pro cal_cld_gts_tdm,dtr_prefix,outprefix,year1,year2,info=info
; calculates cld anomalies using relationship with dtr anomalies
; reads coefficients from predefined files (*1000)
; reads DTR data from binary output files from quick_interp_tdm2.pro (binfac=1000)
; creates cld anomaly grids at dtr grid resolution
; output can then be used as dummy input to splining program that also
;  includes real cloud anomaly data

So, to me this identifies it as the program we cannot use any more because
the coefficients were lost. As it says in the gridding read_me:

Bear in mind that there is no working synthetic method for cloud, because Mark New
   lost the coefficients file and never found it again (despite searching on tape
   archives at UEA) and never recreated it. This hasn't mattered too much, because
   the synthetic cloud grids had not been discarded for 1901-95, and after 1995
   sunshine data is used instead of cloud data anyway.

But, (Lord how many times have I used 'however' or 'but' in this file?!!), when
you look in the program you find that the coefficient files are called:

rdbin,a,'/cru/tyn1/f709762/cru_ts_2.0/_constants/_7190/a.25.7190',gridsize=2.5
rdbin,b,'/cru/tyn1/f709762/cru_ts_2.0/_constants/_7190/b.25.7190',gridsize=2.5

And, if you do a search over the filesystems, you get:

crua6[/cru/cruts] ls fromdpe1a/data/grid/cru_ts_2.0/_makecld/_constants/_7190/spc2cld/_ann/
a.25.01.7190.glo.Z  a.25.05.7190.glo.Z  a.25.09.7190.glo.Z  a.25.7190.eps.Z     b.25.04.7190.glo.Z  b.25.08.7190.glo.Z  b.25.12.7190.glo.Z
a.25.02.7190.glo.Z  a.25.06.7190.glo.Z  a.25.10.7190.glo.Z  b.25.01.7190.glo.Z  b.25.05.7190.glo.Z  b.25.09.7190.glo.Z  b.25.7190.eps.Z
a.25.03.7190.glo.Z  a.25.07.7190.glo.Z  a.25.11.7190.glo.Z  b.25.02.7190.glo.Z  b.25.06.7190.glo.Z  b.25.10.7190.glo.Z
a.25.04.7190.glo.Z  a.25.08.7190.glo.Z  a.25.12.7190.glo.Z  b.25.03.7190.glo.Z  b.25.07.7190.glo.Z  b.25.11.7190.glo.Z
crua6[/cru/cruts] ls fromdpe1a/data/grid/cru_ts_2.0/_makecld/_constants/_7190/spc2cld/_mon/
a.25.01.7190.glo.Z  a.25.05.7190.glo.Z  a.25.09.7190.glo.Z  a.25.7190.eps.Z     b.25.04.7190.glo.Z  b.25.08.7190.glo.Z  b.25.12.7190.glo.Z
a.25.02.7190.glo.Z  a.25.06.7190.glo.Z  a.25.10.7190.glo.Z  b.25.01.7190.glo.Z  b.25.05.7190.glo.Z  b.25.09.7190.glo.Z  b.25.7190.eps.Z
a.25.03.7190.glo.Z  a.25.07.7190.glo.Z  a.25.11.7190.glo.Z  b.25.02.7190.glo.Z  b.25.06.7190.glo.Z  b.25.10.7190.glo.Z
a.25.04.7190.glo.Z  a.25.08.7190.glo.Z  a.25.12.7190.glo.Z  b.25.03.7190.glo.Z  b.25.07.7190.glo.Z  b.25.11.7190.glo.Z

So.. we don't have the coefficients files (just .eps plots of something). But
what are all those monthly files? DON'T KNOW, UNDOCUMENTED. Wherever I look,
there are data files, no info about what they are other than their names. And
that's useless.. take the above example, the filenames in the _mon and _ann
directories are identical, but the contents are not. And the only difference
is that one directory is apparently 'monthly' and the other 'annual' - yet
both contain monthly files.

Lots of further investigation.. probably the most useful program found is
cal_cld_gts_tdm.pro, the description of which reads as follows:

pro cal_cld_gts_tdm,dtr_prefix,outprefix,year1,year2,info=info
; calculates cld anomalies using relationship with dtr anomalies
```

```
; reads coefficients from predefined files (*1000)
; reads DTR data from binary output files from quick_interp_tdm2.pro (binfac=1000)
; creates cld anomaly grids at dtr grid resolution
; output can then be used as dummy input to splining program that also
;  includes real cloud anomaly data

It also tellingly contains:
; unnecessary because 61-90 normals have already been created
; print, "@@@@@ looking for 2.5 deg DTR 1961-90 @@@@@"
; mean_gts,'~/m1/gts/dtr/glo25/glo25.dtr.',nor1,nor2
; mean_gts_tdm,'/cru/mark1/f080/gts/dtr/glo25/glo25.dtr.',nor1,nor2
;print, "@@@@@ looking for 2.5 deg DTR normal @@@@@"
;; rdbin,dtrnor,'~/m1/gts/dtr/glo25/glo25.dtr.'+string(nor1-1900,nor2-1900,form='(2i2.2)')
;dtrnorstr='/cru/mark1/f080/gts/dtr/glo25/glo25.dtr.'+string(nor1-1900,nor2-1900,form='(2i2.2)')
;rdbin,dtrnor,dtrnorstr

The above has seemingly been replaced with:
rdbin,a,'/cru/tyn1/f709762/cru_ts_2.0/_constants/_7190/a.25.7190',gridsize=2.5
rdbin,b,'/cru/tyn1/f709762/cru_ts_2.0/_constants/_7190/b.25.7190',gridsize=2.5

These are the files that have been lost according to the gridding read_me
(see above).

The conclusion of a lot of investigation is that the synthetic cloud grids
for 1901-1995 have now been discarded. This means that the cloud data prior
to 1996 are static.

Edit: have just located a 'cld' directory in Mark New's disk, containing
over 2000 files. Most however are binary and undocumented..

Eventually find fortran (f77) programs to convert sun to cloud:

sh2cld_tdm.for     converts sun hours monthly time series to cloud percent
sp2cld_m.for       converts sun percent monthly time series to cloud oktas

There are also programs to convert sun parameters:

sh2sp_m.for        sun hours to sun percent
sh2sp_normal.for   sun hours monthly .nrm to sunshine percent
sh2sp_tdm.for      sun hours monthly time series to sunshine percent

AGREED APPROACH for cloud (5 Oct 06).

For 1901 to 1995 - stay with published data. No clear way to replicate
process as undocumented.

For 1996 to 2002:
 1. convert sun database to pseudo-cloud using the f77 programs;
 2. anomalise wrt 96-00 with anomdtb.f;
 3. grid using quick_interp_tdm.pro (which will use 6190 norms);
 4. calculate (mean9600 - mean6190) for monthly grids, using the
    published cru_ts_2.0 cloud data;
 5. add to gridded data from step 3.

This should approximate the correction needed.

On we go.. firstly, examined the spc database.. seems to be in % x10.
Looked at published data.. cloud is in % x10, too.
First problem: there is no program to convert sun percentage to
cloud percentage. I can do sun percentage to cloud oktas or sun hours
to cloud percentage! So what the hell did Tim do?!! As I keep asking.

Examined the program that converts sun % to cloud oktas. It is
complicated! Have inserted a line to multiple the result by 12.5 (the
result is in oktas*10 and ranges from 0 to 80, so the new result will
range from 0 to 1000).

Next problem - which database to use? The one with the normals included
is not appropriate (the conversion progs do not look for that line so
obviously are not intended to be used on +norm databases). The non
normals databases are either Jan 03 (in the '_ateam' directory) or
Dec 03 (in the regular database directory). The newer database is
smaller! So more weeding than planting in 2003. Unfortunately both
databases contain the 6190 normals line, just unpopulated. So I will
go with the 'spc.0312221624.dtb' database, and modify the already-
modified conversion program to process the 6190 line.

Then - comparing the two candidate spc databases:

spc.0312221624.dtb
spc.94-00.0312221624.dtb

I find that they are broadly similar, except the normals lines (which
both start with '6190') are very different. I was expecting that maybe
the latter contained 94-00 normals, what I wasn't expecting was that
thet are in % x10 not %! Unbelievable - even here the conventions have
not been followed. It's botch after botch after botch. Modified the
conversion program to process either kind of normals line.

Decided to go with the 'spc.94-00.0312221624.dtb' database, as it
hopefully has some of the 94-00 normals in. I just wish I knew more.

Conversion was hampered by the discovery that some stations have a mix
of % and % x10 values! So more mods to Hsp2cldp_m.for. Then conversion,
producing cldfromspc.94000312221624.dtb. Copied the .dts file across
as is, not sure what it does unfortunately (or can't remember!).

After conversion, ran anomdtb:

crua6[/cru/cruts/rerun1/data/cruts/rerun_cld] ./anomdtb

   > ***** AnomDTB: converts .dtb to anom .txt for gridding *****

   > Enter the suffix of the variable required:
.cld
   > Select the .cts or .dtb file to load:
cldfromspc.94000312221624.dtb

   > Specify the start,end of the normals period:
1994,2000
   > Specify the missing percentage permitted:
25
   > Data required for a normal:            6
   > Specify the no. of stdevs at which to reject data:
3
   > Select outputs (1=.cts,2=.ann,3=.txt,4=.stn):
3
   > Check for duplicate stns after anomalising? (0=no,>0=km range)
8
   > Select the generic .txt file to save (yy.mm=auto):
```

```
cldfromspc.txt
   > Select the first,last years AD to save:
1994,2002
   > Operating...

   >         .cts      96309    19.6     280712    57.2
   > PROCESS        DECISION percent %of-chk
   > no lat/lon            0     0.0     0.0
   > no normal        209619    42.8    42.8
   > out-of-range     177298    36.2    63.2
   > duplicated          154     0.0     0.1
   > accepted         103260    21.1
   > Dumping years 1994-2002 to .txt files...

crua6[/cru/cruts/rerun1/data/cruts/rerun_cld]

Then ran quick_interp_tdm2:

IDL> .compile /cru/cruts/fromdpela/code/idl/pro/quick_interp_tdm2.pro
% Compiled module: QUICK_INTERP_TDM2.
IDL> .compile /cru/cruts/fromdpela/code/idl/pro/rdbin.pro
% Compiled module: RDBIN.
IDL> quick_interp_tdm2,1994,2002,'glo_from_idl/cld.',600,gs=0.5,pts_prefix='txt_4_idl/cldfromspc.',dumpglo='dumpglo'
Defaults set
    1994
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: STRIP.
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
    1995
    1996
    1997
    1998
    1999
    2000
    2001
    2002
IDL>

Tadaa: .glo files produced for 1994 to 2002.

Then retracked to produce regular 0.5-degree grids for dtr (having only
produced 2.5-degree binaries for synthetics earlier):

IDL> quick_interp_tdm2,1901,2002,'glo_dtr_grids/dtr.',750,gs=0.5,pts_prefix='dtr_txt_4idl/dtr.',dumpglo='dumpglo'

That went off without any apparent hitches, so I wrote a fortran prog,
'maxminmaker.for', to produce tmn and tmx grids from tmp and dtr. It ran.

However - yup, more problems - when I checked the inputs and outputs I found
that in numerous instances there was a value for mean temperature in the grid,
with no corresponding dtr value. This led to tmn = tmx = tmp for thos cells.
NOT GOOD.

Actually, what was NOT GOOD was my grasp of context. Oh curse this poor
memory! For the IDL gridding program produces ANOMALIES not ACTUALS.

Wrote a program, 'glo2abs.for' does a file-for-file conversion of .glo
files (as produced by quick_interp_tdm2.pro) to absolute-value files (also
gridded and with headers). After some experiments realised that the .glo
anomalies are in degrees, but the normals are in 10ths of a degree :-)

Produced absolutes for TMP. Then wrote a program, 'cmpcruts.for', to
compare the absolute grids with the published cru_ts_2.10 data. The
comparison simply measures absolute differences between old and new, and
categorises as either (1) identical, (2) within 0.5 degs, (3) within 1 deg,
(4) over 1 deg apart. Results for temperature (TMP):

 Identical   <0.5deg  0.5-1deg     >1deg
  30096176  48594200   2755281   1076423

And for temperature range (DTR):

  45361058  31267870   3893754   1999398

These are very promising. The vast majority in both cases are within 0.5
degrees of the published data. However, there are still plenty of values
more than a degree out.

The total number of comparisons is 67420*102*12 = 82,522,080

It seems prudent to add percentage calculations..

TMP:
Final Diff Totals:   30096176  48594200   2755281   1076423
Percentages:            36.47     58.89      3.34      1.30

TMP has a comforting 95%+ within half a degree, though one still wonders
why it isn't 100% spot on..

DTR:
Final Diff Totals:   45361058  31267870   3893754   1999398
Percentages:            54.97     37.89      4.72      2.42

DTR fares perhaps even better, over half are spot-on, though about
7.5% are outside a half.

However, it's not such good news for precip (PRE):
Final Diff Totals:   11492331  21163924   9264554  40601271
Percentages:            13.93     25.65     11.23     49.20

21. A little experimentation goes a short way..

I tried using the 'stn' option of anomdtb.for. Not completely sure what
it's supposed to do, but no matter as it didn't work:

crua6[/cru/cruts/rerun1/data/cruts/rerun_pre] ./anomdtb

   > ***** AnomDTB: converts .dtb to anom .txt for gridding *****

   > Enter the suffix of the variable required:
.pre
   > Will calculate percentage anomalies.
   > Select the .cts or .dtb file to load:
pre.0312031600H.dtb

   > Specify the start,end of the normals period:
1961,1990
   > Specify the missing percentage permitted:
```

```
25
   > Data required for a normal:           23
   > Specify the no. of stdevs at which to reject data:
5
   > Select outputs (1=.cts,2=.ann,3=.txt,4=.stn):
4
   > Check for duplicate stns after anomalising? (0=no,>0=km range)
8
   > Select the .stn file to save:
pre.fromanomdtb.stn
   > Enter the correlation decay distance:
450
   > Submit a grim that contains the appropriate grid.
   > Enter the grim filepath:
cru_ts_2_10.1961-1970.pre

   > Grid dimensions and domain size:      720     360   67420
   > Select the first,last years AD to save:
1901,2002
   > Operating...

   > NORMALS            MEAN percent      STDEV percent
   >         .dtb    2635548    29.6
   >         .cts    4711327    52.8    7325296    82.2
   > PROCESS        DECISION percent %of-chk
   > no lat/lon       20761     0.2     0.2
   > no normal       1585342    17.8    17.8
   > out-of-range      20249     0.2     0.3
   > duplicated       317035     3.6     4.3
   > accepted        6972308    78.2
   > Calculating station coverages...
   > ##### WithinRange: Alloc: DataB #####
forrtl: severe (174): SIGSEGV, segmentation fault occurred
crua6[/cru/cruts/rerun1/data/cruts/rerun_pre]

..knowing how long it takes to debug this suite - the experiment
endeth here. The option (like all the anomdtb options) is totally
undocumented so we'll never know what we lost.

22. Right, time to stop pussyfooting around the niceties of Tim's labyrinthine software
suites - let's have a go at producing CRU TS 3.0! since failing to do that will be the
definitive failure of the entire project..

Firstly, we need to identify the updated data files. I acquired the following:

iran_asean_GHCN_WWR-CD_save50_CLIMAT_MCDW_updat_merged renamed to pre.0611301502.dat
newbigfile0606.dat renamed to tmp.0611301507.dat
glseries_tmn_final_merged renamed to tmn.0611301516.dat
glseries_tmx_final_merged renamed to tmx.0611301516.dat
anders9106m.dat renamed to tmp9106.0612011708.dat

..and established a directory hierarchy under /cru/cruts/version_3_0

Next step, convert the various db formats to the CRU TS one. Made a visual
comparison which indicated that it would work. Unfortunately it will mean
losing the 'extra' fields that have been tacked onto the headers willy-nilly
as they are undocumented. Furthermore the two extra fields in the CRU TS
format are undocumented, as far as I can see! So I wrote headergetter.for
to produce stats on the CRU TS headers. It looks for violations of the
mandatory blank spaces, and for variations in the two extra fields. Sample
output for temperature and precip:

Header report for tmp.0311051552.dtb
Produced by headgetter.for
Total Records Read:    12155

BLANKS (expected at 8,14,21,26,47,61,66,71,78)
  position   missed
       8        0
      14        0
      21        0
      26        0
      47        0
      61        0
      66        0
      71        0
      78        2

EXTRA FIELD 1 (72:77)
       type detected   counted
Missing Value Code       12155
Possible F.P. Value          0
Possible Exp. Value          0
Integer Value Found          0
Real Value Found             0
Unidentifiable               0

EXTRA FIELD 2 (79:86)
       type detected   counted
Missing Value Code         709
Possible F.P. Value        697
Possible Exp. Value          0
Integer Value Found      10749
Real Value Found             0
Unidentifiable               0

ENDS

Header report for pre.0312031600.dtb
Produced by headgetter.for
Total Records Read:    12732

BLANKS (expected at 8,14,21,26,47,61,66,71,78)
  position   missed
       8        0
      14        0
      21        0
      26        0
      47        0
      61        0
      66        0
      71        0
      78      154

EXTRA FIELD 1 (72:77)
       type detected   counted
Missing Value Code       12732
Possible F.P. Value          0
Possible Exp. Value          0
Integer Value Found          0
```

```
Real Value Found             0
Unidentifiable               0

EXTRA FIELD 2 (79:86)
      type detected   counted
Missing Value Code        3635
Possible F.P. Value        437
Possible Exp. Value          0
Integer Value Found       8660
Real Value Found             0
Unidentifiable               0

ENDS
```

As can be seen, there are no unidentifiable headers - hurrah! - but quite a few violations of the boundary between the two extra fields, particularly in the precip database. On examination, the culprits are all African stations. The two tmp exceptions:

```
 641080  -330   1735  324 BANDUNDU             DEM REP CONGO 1961 1990   -99908
 642200  -436   1525  445 KINSHASA/BINZA       DEM REP CONGO 1960 1990   -99920
```

And samples of the pre exceptions:

```
-656002   698   -958  150 SUAKOKO              LIBERIA       1951 1970   -999123008050
-655327   727   -723  350 KOUIBLY              IVORY COAST   1977 1990   -999109001290
-655001  1320   -235  332 GOURCY               BURKINA FASO  1956 1980   -999120001240
-618504   788  -1118 -999 KENEMA/FARM          SIERRA LEONE  1951 1972   -999139003500
-612067  1407   -307  253 KORO                 MALI          1958 1989   -999127002650
```

So the first extra field is apparently unused! It would be a handy place for the 6-character data-code and valid-start-year from the temperature db.

On to a more detailed look at the cru precip format; not sure whether there are two extra fields or one, and what the sizes are. A quick hack through the headers is not pleasing. There appears to be only one field, but it can have up to nine (9) digits in it, and at least three missing value codes:

```
6785300-1863  2700 1080HWANGE/N.P.A.       ZIMBABWE     19621996              40
8100100  680 -5820    2GEORGETOWN          GUYANA       18462006             -99
6274000 1420  2460 1160KUTUM                SUDAN        19291990             194
6109200-9999-99999 -999UNKNOWN              NIGER        19891989            -999
6542000  945    -2  197YENDI                 GHANA        19071997            8010
6544200  672  -160  293KUMASI                GHANA        19062006           17009
6122306 1670  -299  267KABARA                MALI         19231989          270022
6193128   32   672 -999SAO TOME             SAO TOME     19391973         8888888
6266000 1850  3180  249KARIMA                SUDAN        19172006        18315801
6109905 1208  -367  315OUARKOYE              BURKINA FASO 19601980       120002470
```

unimpressed

This is irritating as it means precip has only 9 fields and I can't do a generic mapping from any cru format to cru ts.

As a glutton for punishment I then looked at the tmin/tmax db format. Looks like two extra fields (i6,i7) with mvcs of 999999 and 8888888 respectively. However *sigh* inspection reveals the following two possibilities:

```
 851300 3775 -2568   17PONTA DELGADA        PORTUGAL     18652004 9999998888888
 851500 3697 -2517  100SANTA MARIA A        ACORES       19542006 -77777  8888888
```

Isn't that marvellous? These can't even be read with a consistent header format!

So, the approach will be to read exactly ONE extra field. For cru tmp that will be the i2+i4 anders/best-start codes as one. For cru pre it will be the amazing multipurpose, multilength field. For cru tmnx it will be the first field, which is at least stable at i6.

Conversions/corrections performed:

Temperature

```
Converted tmp.0611301507.dat to tmp.0612081033.dat

Found one corrupted station name:
BEFORE
 911900   209   1564   20  HI*KAHULUI WSO (PUU NENE)         1954 1990 101954  -999.00
AFTER
 911900   209   1564   20 KAHULUI ARPT/MAUI    HAWAII        1954 1990 101954  -999.00
```

Precipitation

```
Converted pre.0611301502.dat to pre.0612081045.dat

Found one corrupted station name:
BEFORE
4125600 2358  5828   15SEEB AP./=MUSCAT*0.9OMAN         18932006          301965
AFTER
4125600  2358   5828   15 SEEB INTL/MUSCAT     OMAN          1893 2006   -999  -999.00
```

(DL later reported that the name wasintended to signify that the data had been corrected by a factor of 0.9 when data from another station was incorporated to extend the series - this was Mike Hulme's work)

Write db2dtb.for, which converts any of the CRU db formats to the CRU TS format.

Started work on mergedb.for, which should merge a primary database with and incoming database of the same (CRU TS) format. Quite complicated. No operator interventions, just a log file of failed attempts - but hooks left in for op sections in case this turns out to be the main programmatic deliverable to BADC!

23. Interrupted work on mergedb.for in order to trial a precip gridding for 3.0. This required another new proglet, addnormline.for, which adds a normals line below each header. It fills in the normals values if the condisions are met (75% of values, or 23 for the 30 year period).

Initial results promising.. ran it for precip, it added normals lines OK, a total of 15942 with 6003 missing value lines. No errors, and no ops interventions because the file didn't have normals lines before!

'Final' precip file: pre.0612151458.dtb

Tried running anomdtb.f90.. failed because it couldn't find the .dts file! No matter that it doesn't need it - argh!

Examined existing .dts files.. not sure what they're for. Headers are identical to the .dtb file, all missing values are retained, all other values are replaced with one of several code numbers, no idea what they mean.

```
Wrote 'falsedts.for' to produce dummy .dts files with all zeros in place of real
data values. Produced pre.0612151458.dts.

Added normals line, producing: pre.0612181221.dtb
Re-produced matching pre.0612181221.dts file.

Tried running anomdtb.f90 again. This time it crashed at record #1096. Wrote a proglet
'findstn.for' to find the n-th station in a dtb file, pulled out 1096:

      0   486  10080 1036 BUKIT LARUT          MALAYSIA      1951 1988   -999  -999.00
6190 2094 2015 2874 3800 4619 3032 5604 3718 4626 5820 5035 3049
1951 3330 2530 2790 5660 4420 4030 1700 2640 8000 5950 6250 2020

(snipped normal years)

1979  110 1920 1150 5490 3140 308067100 2500 4860 4280 4960 1600

Uh-oh! That's 6.7m of rain in July 1979? Looks like a factor-of-10 problem. Confirmed
with DL and changed to 6710.

Next run, crashed at #4391, CHERRAPUNJI, the wettest place in the world. So here, the
high values are realistic. However I did notice that the missing value code was -10
instead of -9999! So modified db2dtb.for to fix that and re-produced the precip database
as pre.0612181214.dat. This then had to have normals recalculated for it (after fixing
#1096).

Finally got it through anomdtb.for AND quick_interp_tdm2 - without crashing! IDL was even
on the ball with the missing months at the end of 2006:


IDL> quick_interp_tdm2,1901,2006,'preglo/pregrid.',450,gs=0.5,dumpglo='dumpglo',pts_prefix='preanoms/pre.'
% Compiled module: QUICK_INTERP_TDM2.
% Compiled module: GLIMIT.
Defaults set
    1901
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: STRIP.
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
    1902
    1903
(etc)
    2005
    2006
no stations found in: preanoms/pre.2006.09.txt
no stations found in: preanoms/pre.2006.10.txt
no stations found in: preanoms/pre.2006.11.txt
no stations found in: preanoms/pre.2006.12.txt

All good. Wrote mergegrids.for to create the more-familiar decadal and full-series
files from the monthly *.glo.abs ones.

Then.. like an idiot.. I had to test the data! Duh.

Firstly, wrote mmeangrid.for and cmpmgrids.m to get a visual comparison of old and
new precip grids (old being CRU TS 2.10). This showed variations in 'expected' areas
where changes had been made, it the Southern tip of Greenland.

Next, Phil requested some statistical plots of percentage change in annual totals,
and long-term trends. Wrote 'anntots.for' to convert monthly gridded files into
yearly totals files. Then tried to write precipchecker.m to do the rest in Matlab..
it wasn't having it, OUT OF MEMORY! Bah. So wrote 'prestats.for' to calculate the
final stats, for printing with an emasculated precipchecker.m. BUT.. it wouldn't
work, and on investigating I found 200-odd stations with zero precipitation for
the entire 1901-2006 period! Modified anntots.for to dump a single grid with those
cells that remained at zero marked, then plotted.

Zero cells in North Africa and the Western coast of South America. None in the
CRU TS 2.10 precip grids :-(

Next step, produce a list of cell centres of the offending cells. wrote a quick
proglet, 'idzerocells.for'. Then 'getcellstations.for', which, given a CRUTS DB
file and a list of lat/lon values, extracts all stations lying inside the cells
listed.

Uh-oh. Looked in the new pre db and found 15 stations for 257 zero cells! They are:

6061170  2810    670  381 FT FLATTERS          ALGERIA       1925 1965   -999  -999.00
6064000  2650    840  559 FORT POLIGNAC        ALGERIA       1925 2006   -999  -999.00
6262000  2080   3260  470 STATION NO. 6        SUDAN         1950 1988   -999  -999.00
8450100  -810  -7900   26 TRUJILLO             PERU          1961 2006   -999  -999.00
8453100  -920  -7850   10 CHIMBOTE             PERU          1961 2006   -999  -999.00
8462800 -1200  -7710   13 LIMA-CALLAO/INTL.AP. PERU          1961 2006   -999  -999.00
8463100 -1210  -7700  137 LIMATAMBO/C.DE MARTE PERU          1927 1980   -999  -999.00
8469100 -1380  -7630    6 PISCO                PERU          1942 2006   -999  -999.00
8540600 -1850  -7030   29 ARICA/CHACALLUTA     CHILE         1903 2006   -999  -999.00
8541700 -2020  -7020    6 IQUIQUE/CAVANCHA     CHILE         1886 1986   -999  -999.00
8541800 -2053  -7018   52 IQUIQUE DIEGO ARACEN CHILE         1989 2006   -999  -999.00
8700494  -707  -7957  150 CAYALTI              PERU          1934 1959   -999  -999.00
8700562 -1203  -7703  137 LIMA                 PERU          1929 1963   -999  -999.00
8700581 -1207  -7717   13 LA PUNTA (NA         PERU          1939 1963   -999  -999.00
9932040  2810    670  381 FT FLATTER           ALGERIA       1925 1965   -999  -999.00

Looked for the same zero cell stations in the old pre db (pre.0312031600.dtb) and only
found 10:

-854031 -2021  -7015    5 IQUIQUE/CAVANCHA     CHILE         1899 1986   -999     0.00
-843002 -1210  -7700  135 LIMATAMBO            PERU          1927 1980   -999
-603550  2810    670  381 FT FLATTER           ALGERIA       1925 1965   -999  -999.00
 606400  2650    841  558 ILLIZI/ILLIRANE      ALGERIA       1925 2002   -999     -999
 626200  2075   3255  468 STATION NO. 6        SUDAN         1950 1988   -999  -999.00
 845010  -810  -7903   30 TRUJILLO/MARTINEZ    PERU          1961 2002   -999     -999
 845310  -916  -7851   11 CHIMBOTE/TENIENTE    PERU          1961 2001   -999
 846280 -1200  -7711   13 LIMA/JORGE CHAVEZ    PERU          1961 2002   -999     -999
 846910 -1375  -7628    7 PISCO (CIV/MIL)      PERU          1942 2002   -999     -999
 854180 -2053  -7018   52 IQUIQUE/DIEGO ARAC   CHILE         1989 2002   -999  -999.00

So why does the old db result in no 'zero' cells, and the new db give us over 250? I
wondered if normals might be the answer, but none of the 10 stations from the old db
have in-db normals, wheras three of the new db have:

8453100  -920  -7850   10 CHIMBOTE             PERU          1961 2006   -999  -999.00
6190   19   59   36   18    5    0    3    0    0    1   10    5
8469100 -1380  -7630    6 PISCO                PERU          1942 2006   -999  -999.00
6190    3    0    3    0    0    1    1    3    1    4    0    0
8540600 -1850  -7030   29 ARICA/CHACALLUTA     CHILE         1903 2006   -999  -999.00
6190    1    3    0    0    0    2    2    2    2    0    0    0
```

So these alone ought to guarantee three of the cells being nonzero - they should have
the bloody normals in! So the next check has to be the climatology, that which provides
the cell-by-cell normals..

A check of the gridded climatology revealed that all 257 'zero cells' have their
climatologies set to zero, too. This was partially checked in the GRIM-format climatology
just in case!

Next, a focus: on CHIMBOTE (see header line above). This has real data (not just zeros).
It is in cell (162,203), or (-9.25,-78.75) [lat, lon in both cases]. So we extract the
full timeseries for that cell from the published 2.10 (1901-2002) GRIM file:

```
Grid-ref= 203, 162
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    0    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    7    0    3
    2    0    0    0    2    0    0    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    0    0    0    5    0    3
    2    0    0    0    2    0    2    0    0    5    0    3
    2    0    0    0    2    0    0    0    0    5    0    3
    0    0    0    0    2    0    0    0    0    0    0    0
    0    0    0    0    2    0    0    0    0    0    0    2
    0    0    0    0    2    5    6    0    0    0    0    3
    2    3    0    0    0    0   17    0    0    4    0    3
    2    0    0    0    3    0    2    0    0    2    0    3
    0    0    0    0    0    0   14    0    0    9    0    0
    0    0    0    0    0    0    0    0    0    2    0    2
    0    0    0    0    0    0   12    0    0    0    0    5
    0    0    0    0    0    0    0    0    0    3    0    2
    0    0    0    0    0    0   10    0    0    0    0    2
    0    0    0    0    3    0   11    0    0    2    0    3
    0    0    0    0    2    0    0    0    0    0    0    2
    0    0    0    0    0    0    0    0    0    4    0    0
    3    0    0    0    0    0   15    0    0    0    0    2
    0    0    0    0    0    0    0    0    0    4    3    2
    5    0    0    0    0    0    0    0    0   12    0    3
    0    0    2    2    4    2    0    0    2    3    0    3
    0    0    0    0    3    0    0    2    0    2    2    3
    0    0    0    0    0    0    0    3    0    0    0    0
    0    0    0    0    0    0    0    0    0    0    0    0
    0    0    0    7    0    3    0    0    0    0    0    0
    0    0    2    3    0    0    0    4    0    0   12    0
    0    0    9    0    0    0    0    0    0    0    0    0
    0    0    0    0    0    0    0    0    0    0    0    0
    0    6    2    0    0    0    6    0    0    0    0    0
    0    0    0    0    0    0    0    2    2    0    0    0
    0    0    0    0    0    3    0    0    0    0    0    0
    0    0    0    0    2    2    0    0    0    0    0    0
    0    2    0    0    0    0    0    2    0    0    0    0
    0    0    0    7    0    0    0    2    0    0    0    3
    2    0    7    0    0    2    0    0    2    0    0    0
    0    0    0    7    0    0    2    2    2    0    0    0
    8    0    2    0    0    0    0    2    0    0    0    0
    0    7    0    0    0    0    0    0    0    0    0    0
    0    0    2    0    0    0    0    0    0    0    0    0
    0    0    0    0    0    0    0    0    0    2    0    0
    0    0    0    0    2    0    0    0    0    0    0    0
    2    0    0    0    0    2    0    0    0    0   10    0
    3    0    0    0    0    0    9    0    0    0    0    3
    0    0    0    0    0    0    0    0    0    3    0    5
    4    0    0    2   10    2    0    0    0    0    0    4
    0    0    0    0    0    0    0    0    2    5    0    0
    0    0    0    0    0    0    9    0    0    0    0    0
    0    0    0    0    0    0    0    3    0    0    0    0
    0    0    0    0    0    0    0    0    0    5    0    0
    3    0    0    0    0    0    0    0    0    0    0    0
    2    0    0    0    0    0    0    0    3    0    0    0
    0    0    0    0    8    2    0    0    0    0    0    3
    0    0    2    0    2    0    0    0    0    0    2    3
    0    0    0    0    2    0    2    0    0    5    0    2
    0    0    0    0    2    0    2    0    0    0    0    0
    0    0    0    0    0    0    0    0    0    5    0    0
    0    0    0    0    2    0    0    2    0    0    0    0
    2    0    2    0    0    0    0    0    0    5    0    0
    0    0    0    0   11    0    2    0    0    4    0    3
    2    3    2    0   13    0    0    0    0    0    0    0
    2    6    0    3    0    0    0    0    2    3    0    7
    2    0    0    0    2    0    0    0    0    0    0    3
    0    0    0    0    0    0    2    0    0    0    0    2
    0    0    0    0    0    0    0    0    0    0    0    3
```

..yet in the 3.00 version, it's all zeros!

Only one thing for it.. examine the attempt at regenerating 2.10.
Unfortunately - well, interestingly then - this gave the same

```
zero cells as the 3.00 generation! So it's something to do with
the process, not the database (or the climatology, assuming that
has remained constant, which I gather it has).

Update: aha! Phil pointed out that for precip the climatology
is used as a MULTIPLIER. So if the clim hasn't changed, the
cells should always have been zero regardless of actual data.

As I should have remembered:

crua6[/cru/cruts/version_3_0/primaries/precip] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.pre
Enter a name for the gridded climatology file: clim.6190.lan.pre.grid
Enter the path and stem of the .glo files: preglo/pregrid.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: pregrid/
Now, CONCENTRATE. Addition or Percentage (A/P)? P
Right, erm.. off I jolly well go!
pregrid.01.1901.glo
pregrid.02.1901.glo
(etc)

Decided to read Mitchell & Jones 2005 again. Noticed that the
limit for SD when anomalising should be 4 for precip, not 3! So
re-ran with that:

crua6[/cru/cruts/version_3_0/primaries/precip] ./anomdtb

   > ***** AnomDTB: converts .dtb to anom .txt for gridding *****

   > Enter the suffix of the variable required:
.pre
   > Will calculate percentage anomalies.
   > Select the .cts or .dtb file to load:
pre.0612181221.dtb
 pre.0612181221.dtb


 /tmp_mnt/cru-auto/cruts/version_3_0/primaries/precip/pre.0612181221.dtb


   > Specify the start,end of the normals period:
1961,1990
   > Specify the missing percentage permitted:
25
   > Data required for a normal:           23
   > Specify the no. of stdevs at which to reject data:
4
   > Select outputs (1=.cts,2=.ann,3=.txt,4=.stn):
3
   > Check for duplicate stns after anomalising? (0=no,>0=km range)
8
   > Select the generic .txt file to save (yy.mm=auto):
pre4sd.txt
   > Select the first,last years AD to save:
1901,2006
   > Operating...
 /tmp_mnt/cru-auto/cruts/version_3_0/primaries/precip/pre.0612181221.dtb


 /tmp_mnt/cru-auto/cruts/version_3_0/primaries/precip/pre.0612181221.dtb


 /tmp_mnt/cru-auto/cruts/version_3_0/primaries/precip/pre.0612181221.dts


 /tmp_mnt/cru-auto/cruts/version_3_0/primaries/precip/pre.0612181221.dts


   > NORMALS            MEAN percent      STDEV percent
   >         .dtb    7315040    73.8
 made it to here
   >         .cts     299359     3.0    7613600    76.8
   > PROCESS        DECISION percent %of-chk
   > no lat/lon        17527     0.2     0.2
   > no normal       2355659    23.8    23.8
   > out-of-range      13253     0.1     0.2
   > duplicated       586206     5.9     7.8
   > accepted        6934807    70.0
   > Dumping years 1901-2006 to .txt files...


This is not as good a percentage as for 2.10:

   > NORMALS            MEAN percent      STDEV percent
   >         .dtb          0     0.0
   >         .cts    3375441    84.1    3375441    84.1
   > PROCESS        DECISION percent %of-chk
   > no lat/lon         3088     0.1     0.1
   > no normal        638538    15.9    15.9
   > out-of-range      70225     1.7     2.1
   > duplicated       135457     3.4     4.1
   > accepted        3167636    78.9
   > Dumping years 1901-2002 to .txt files...

But the actual number of accepted values is more than TWICE 2.10!

Of course, the same 257 gridcells are zeros, because the multiplicative
normals are still zero.

For reference, these are the results for the 3 SD limit of 3.00:

   > NORMALS            MEAN percent      STDEV percent
   >         .dtb    7315040    73.8
 made it to here
   >         .cts     284160     2.9    7598401    76.7
   > PROCESS        DECISION percent %of-chk
   > no lat/lon        17527     0.2     0.2
   > no normal       2370858    23.9    24.0
   > out-of-range      32379     0.3     0.4
   > duplicated       583193     5.9     7.8
   > accepted        6903495    69.7
   > Dumping years 1901-2006 to .txt files...

So we've only gained 0.3% of values, a real figure of 31312 values.
```

Conclusion: stick with a 3 Standard Deviation limit, like the
Read_Me says.

24. (cont of 22 really)

Restarted work on mergedb.for. Decided I was taking the wrong approach,
so the interruption was probably a GOOD THING.

The process now is to read in the header lines AND line numbers from
the main database, and to then process the incoming database one record
at a time. It's more logical and haivng the line numbers will speed
things up enormously (well it has done on previous occasions).

The biggest immediate problem was the loss of an hour's edits to the
program, when the network died.. no explanations from anyone, I hope
it's not a return to last year's troubles.

(some weeks later)

well, it compiles OK, and even runs enthusiastically. However there are
loads of bugs that I now have to fix. Eeeeek. Timesrunningouttimesrunningout.

(even later)

Getting there.. still ironing out glitches and poor programming.

25. Wahey! It's halfway through April and I'm still working on it. This
surely is the worst project I've ever attempted. Eeeek. I think the main
problem is the rather nebulous concept of the automatic updater. If I
hadn't had to write it to add the 1991-2006 temperature file to the 'main'
one, it would probably have been a lot simpler. But that one operation has
proved so costly in terms of time, etc that the program has had to bend
over backwards to accommodate it. So yes, in retrospect it was not a
brilliant idea to try and kill two birds with one stone - I should have
realised that one of the birds was actually a pterodactyl with a temper
problem.

Success!

```
crua6[/cru/cruts/version_3_0/db/testmergedb] ./mergedb
**************************************************
*                  MERGEDB                       *
*                                                *
*  Merging of two database files                 *
*             Ops ID: f098xxxx                   *
*               Date:   12:17  25/04/07          *
*  The Session ID is: 0704251217.f098xxxx        *
*  (log file 'mergedb.0704251217.f098xxxx.log')  *
*                                                *
*  Please choose the mode of working.            *
*  This program can either run..                 *
*  [1] Interactively, (in which case an operator *
*  must be present throughout to make decision), *
*  or [2] in Batch mode, (in which case it may   *
*  be left unattended). If Batch mode is used, a *
*  file of outstanding issues will be saved for  *
*  later [3] resolution by an operator.          *
*                                                *
*  [1] Interactive (operator) processing         *
*  [2] Batch (no operator) processing            *
*  [3] Operator processing of saved batch        *
*  [4] Run a previously-saved action file        *
*                                                *
*  Please enter 1,2,3 or 4: 4
*  RUN ACTION FILE MODE                          *
*                                                *
*  Enter the ACTion filename, or 'x' for a list: x
*  The  1 most recent ACT files:
*   1. mergedb.0704201343.f098xxxx.act           *
*  Enter a number or 0 for none of the above: 1
*  Enter 'Y' to run this file or 'N' to abort: Y
*                                                *
*  Creation date/time:  13:43 20/04/07           *
*  Batch initiator was: f098                     *

*  Number of actions/requests:      2586
*  This ACT file derived from original OPS file: *
*  mergedb.0704201210.f098xxxx.ops               *
*  Main (existing) Database:      tmp.0702091122.dtb
*  Secondary (incoming) Database: tmp.0612081519.dat
*  Parameter is 'tmp' - confirm (Y/N): Y

*  Actions Completed!                            *
*          Thank You for using MERGEDB!          *
**************************************************
```

..well, 'success' in the sense that it ran and apparently all the data's
in the right place, in tmp.0704251819.dtb.

26. OK, now to merge in the US stations. First, wrote 'us2cru' to convert
the marksusanonwmocru.dat file to the 'standard' format we're using. That
worked OK. Then used 'addnormline' to, well - add a normals line. Only 17
out of 1035 stations ended up with missing normals, which is pretty good!

The with-normals US database file is tmp.0704251654.dat.

Now, I knew that using mergedb as it stands would not work. It expects to
be updating the existing records, and actions like 'addnew' require OPS
to confirm each one. So I thought it best to add an OPS clause to auto-
confirm additions where there's no WMO match and the data density is OK,
say 50% or higher. Unfortunately, that didn't work either, and rather than
spend even more time debugging mergedb.for, I knocked off simpleaddnew.for,
which adds two non-overlapping databases. The resultant file, with all
three partial databases, is tmp.0704271015.dtb.

27. Well, enough excuses - time to remember how to do the anomalising and
gridding things! Fisrtly, ran 'addnormline' just to ensure all normals are
up to date. The result was 8 new sets of normals, so well worth doing. The
database is now:

tmp.0704292158.dtb

Ran 'anomdtb' - got caught out by the requirement for a companion '.dts'
file again, ran 'falsedts.for' and carried on.. would still be nice to be
sure that it's not something meaningful **sigh**.

Output:

```
<BEGIN QUOTE>
crua6[/cru/cruts/version_3_0/primaries/temp] ./anomdtb
```

```
   > ***** AnomDTB: converts .dtb to anom .txt for gridding *****

   > Enter the suffix of the variable required:
.tmp
   > Select the .cts or .dtb file to load:
tmp.0704292158.dtb
 tmp.0704292158.dtb


 /tmp_mnt/cru-auto/cruts/version_3_0/primaries/temp/tmp.0704292158.dtb


   > Specify the start,end of the normals period:
1961,1990
   > Specify the missing percentage permitted:
25
   > Data required for a normal:           23
   > Specify the no. of stdevs at which to reject data:
3
   > Select outputs (1=.cts,2=.ann,3=.txt,4=.stn):
3
   > Check for duplicate stns after anomalising? (0=no,>0=km range)
8
   > Select the generic .txt file to save (yy.mm=auto):
tmp.txt
   > Select the first,last years AD to save:
1901,2006
   > Operating...
 /tmp_mnt/cru-auto/cruts/version_3_0/primaries/temp/tmp.0704292158.dtb


 /tmp_mnt/cru-auto/cruts/version_3_0/primaries/temp/tmp.0704292158.dtb


 /tmp_mnt/cru-auto/cruts/version_3_0/primaries/temp/tmp.0704292158.dts


 /tmp_mnt/cru-auto/cruts/version_3_0/primaries/temp/tmp.0704292158.dts


   > Failed to find file.
   > Enter the file, with suffix: .dts
tmp.0704292158.dts
 tmp.0704292158.dts


 /tmp_mnt/cru-auto/cruts/version_3_0/primaries/temp/tmp.0704292158.dts


   > NORMALS            MEAN percent      STDEV percent
   >         .dtb    3330007    81.3
 made it to here
   >         .cts      92803     2.3    3422810    83.6
   > PROCESS        DECISION percent %of-chk
   > no lat/lon            0     0.0     0.0
   > no normal        671592    16.4    16.4
   > out-of-range        744     0.0     0.0
   > duplicated      4102723   100.2   119.9
   > accepted        -680657   -16.6
   > Dumping years 1901-2006 to .txt files...

crua6[/cru/cruts/version_3_0/primaries/temp]
<END QUOTE>

.. which is a trifle worrying! And looking at the .txt files, they look
rather odd as well - for instance, tmp.1953.03.txt starts like this:

    7.09    0.87    10.0      0.10000  10010
    7.83   -1.55    28.0     -4.80000  10080
    6.97   -1.89    10.0      0.90000   -999
    6.97   -1.89   100.0      0.50000  10260
    7.45   -1.90    16.0     -3.10000  10280
    6.95   -2.55   129.0      3.70000  10650
    7.04   -3.11    14.0      0.00000  10980
    6.60   -0.20     0.0      1.20000  11000
    6.73   -1.44    13.0      1.60000   -999
    6.68   -1.40    39.0      2.20000  11530

Now, do those first two columns look like lat & lon to you? Me neither,
here's what the old version of the same file looks like:

   60.00  -20.00  -999.0      0.40000-990007
   62.00  -33.00  -999.0     -0.40000-990002
   56.50  -51.00     0.0     -0.50000-990000
    6.90  122.06     6.0     -0.60000   -999
   13.13  123.73    17.0      0.20000   -999
   14.52  121.00    15.0      0.60000   -999
   18.37  121.63     4.0      1.10000   -999
    6.90  122.00     6.0     -0.60000   -999
   10.70  122.50    14.0     -0.10000   -999
   13.13  123.73    19.0      0.10000   -999

In fact, the first two columns never get outside of +/- 30. Oh bugger.
What the HELL is going on?!

Decided to pursue that worrying (and impossible) 'duplicates' figure.

The function 'sort' was used to sort the database so that any duplicate
lines would be together - then 'uniq' was used to pull out duplicates.
There were quite a few dupes, and one or two triples too, like these:

crua6[/cru/cruts/version_3_0/primaries/temp] grep -n '1984  \-83  \-46   22   55  126  154  222  215  159   63   32  \-62' tmp.0704292158.dtb
195789:1984  -83  -46   22   55  126  154  222  215  159   63   32  -62
254265:1984  -83  -46   22   55  126  154  222  215  159   63   32  -62
254380:1984  -83  -46   22   55  126  154  222  215  159   63   32  -62

These are from the following stations:
 720344   408   1158 1539 ELKO-FAA-AP---------USA---------   1870 1996 301870  -999.00
 725837   408   1158 1549  NV ELKO FAA AP                    1930 1990 101930  -999.00
 725910   401   1223  103 RED BLUFF            USA           1878 2006 101878  -999.00

The past two are consecutive stations.

Looking at the last two.. it seems that 725910 has 725837's data!

1977   71  124  118  184  167  275  283  280  230  190  126   99
1978  107  114  149  144  208  248  289  282  232  220  118   72
1979   85   99  139  150  218  256  282  258  253  189  117   94
1980   99  121  119  156  192  216  275  262  241  196  128  102
```

```
1981   14   19   49   90  123  196  233  227  164   71   47   11
1982  -49  -14   32   57  114  164  206  214  148   74   11  -23
1983   -9   -1   54   59  114  167  204  223  170  104   25  -19
1984  -83  -46   22   55  126  154  222  215  159   63   32  -62

Ascan be seen, 1981 sees a complete chance in range, especially for
Autumn/Winter. In fact, from 1981 to 1990, 725910 is a copy of
725837! It then reverts to the original range for the rest of the run.
So.. did the merging program do this? Unfortunately, yes. Check dates:

crua6[/cru/cruts/version_3_0/db/testmergedb] grep -n 'RED BLUFF' tmp.0*.*
tmp.0612081519.dat:28595: 725910   401   1223  103 RED BLUFF            USA           1991 2006 101991  -999.00
tmp.0702091122.dtb:171674: 725910   401   1223  103 RED BLUFF            USA           1878 1980 101878  -999.00
tmp.0704251819.dtb:200331: 725910   401   1223  103 RED BLUFF            USA           1878 2006 101878  -999.00
tmp.0704271015.dtb:254272: 725910   401   1223  103 RED BLUFF            USA           1878 2006 101878  -999.00
tmp.0704292158.dtb:254272: 725910   401   1223  103 RED BLUFF            USA           1878 2006 101878  -999.00
crua6[/cru/cruts/version_3_0/db/testmergedb]

The first file is the 1991-2006 update file. The second is the original
temperature database - note that the station ends in 1980.

It has *inherited* data from the previous station, where it had -9999
before! I thought I'd fixed that?!!!

/goes off muttering to fix mergedb.for for the five hundredth time

Miraculously, despite being dog-tired at nearly midnight on a Sunday, I
did find the problem. I was clearing the data array but not close enough
to the action - when stations were being passed through (ie no data to
add to them) they were not being cleaned off the array afterwards. Meh.

Wrote a specific routine to clear halves of the data array, and back to
square one. Re-ran the ACT file to merge the x-1990 and 1991-2006 files.
Created an output file exactly the same size as the last time (phew!)
but with..

crua6[/cru/cruts/version_3_0/db/testmergedb] comm -12 tmp.0704292355.dtb tmp.0704251819.dtb |wc -l
    285516
crua6[/cru/cruts/version_3_0/db/testmergedb] wc -l tmp.0704292355.dtb
    285829 tmp.0704292355.dtb

.. 313 lines different. Typically:

14881,14886c14881,14886
< 1965-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
< 1966-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
< 1967-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
< 1968-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
< 1969-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
< 1970-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
---
> 1965 -221 -177 -234 -182   -5    6   24   36  -15  -91 -100 -221
> 1966 -272 -194 -248 -192  -66   10   27   45  -12  -75 -139 -228
> 1967 -201 -243 -196 -158  -26    1   40   30  -18  -89 -183 -172
> 1968 -253 -256 -253 -107  -42   10   46   33  -21  -64 -134 -195
> 1969 -177 -202 -248 -165  -33    8   42   50   -1  -89 -157 -204
> 1970 -237 -192 -217 -160  -87    6   30   25   -5  -55 -143 -222

ie, what should have been missing data is now missing data again:

200436,200445c200436,200445
< 1981-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
< 1982-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
< 1983-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
< 1984-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
< 1985-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
< 1986-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
< 1987-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
< 1988-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
< 1989-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
< 1990-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
---
> 1981   14   19   49   90  123  196  233  227  164   71   47   11
> 1982  -49  -14   32   57  114  164  206  214  148   74   11  -23
> 1983   -9   -1   54   59  114  167  204  223  170  104   25  -19
> 1984  -83  -46   22   55  126  154  222  215  159   63   32  -62
> 1985  -57  -29   17   89  122  181  244  188  121   79  -11  -50
> 1986    2   31   66   72  113  187  194  214  116   78   11  -39
> 1987  -59   -5   30   97  131  177  193  192  153  101   21  -35
> 1988  -65  -15   29   80  108  184  222  198  138  116    8  -57
> 1989 -113  -54   53   94  113  164  215  186  143   78    8  -24
> 1990  -24  -30   49  100  100  166  214  194  177   77    9  -97

Hurrah!

So the interim database file is tmp.0704292355.dtb. Now to re-add
the US station dataset with simpleaddnew.for.

crua6[/cru/cruts/version_3_0/db/testmergedb] ./simpleaddnew

SIMPLYADDNEW - add stations to a database
This program assumes the two databases have
NO COMMON STATIONS and will fail (stop) if
any are found.

Please enter the main database: tmp.0704292355.dtb

Please enter the new database: tmp.0704251654.dat
Please enter a 3-character parameter code: tmp
Output database is: tmp.0704300053.dtb
crua6[/cru/cruts/version_3_0/db/testmergedb]

So now we have the combined database again, a bit quicker than
last time: tmp.0704300053.dtb. Pity we slid into May: I was hoping
to only be FIVE MONTHS late.

What's worse - there are STILL duplicate non-missing lines, 210 of
them. The first example is this:

1835   92   73  141  187  260  279  281  288  241  195  183  106

Which belongs to this in the original database (tmp.0702091122.dtb):

 722080   329    800   15 CHARLESTON, S. CAROL UNITED STATES 1823 1990 101823  -999.00
6190   84  100  142  180  224  257  274  270  245  191  145  104

..and to this in the US database (tmp.0704251654.dat):

 720467   328    799    3 CHARLESTON-CITY-----USA---------   1835 1996 301835  -999.00
6190   91  106  144  186  227  260  277  272  249  199  154  112
```

These two stations obviously have a lot in common - though not
everything, as their normals (shown) differ. In fact, on examination
the US database record is a poor copy of the main database one, it
has more missing data and so forth. By 1870 they have diverged, so
in this case it's probably OK.. but what about the others? I just do
not have the time to follow up everything. We'll have to take 210
year repetitions as 'one of those things'.

..actually, I decided in the end to follow up all 210 of them. The
likelihood is that the number is far greater, since the filtering
that gave the 210 figure excluded any lines with two or more
consecutive missing values (to avoid hundreds of just-missing-value
lines). Also I spotted some instances where data lines would be
identical but for one or more missing values in one of the stations.

After checking, I found that the majority of the duplications were
between the original database and the US database, with just a couple
of 'linked' stations within the original database, and half a dozen
in the 1991-2006 update file. One surprise was that stations I'm sure
I rejected ended up marked as 'addnew' in the .act file - quite
unsettling!

Rather foolishly, perhaps, I decided to have a go at interactively
incorporating the US data rather than using 'simplyaddnew'. However,
progress was so slow (because of the high number of 'near matches')
that this approach was abandoned.

Tried 'anomdtb' with the fixed final file (tmp.0704300053.dtb)...
no better! The crucial bits:

<BEGIN QUOTE>
```
   > NORMALS             MEAN percent      STDEV percent
   >         .dtb    3323823    81.3
 made it to here
   >         .cts      91963     2.2    3415786    83.5
   > PROCESS        DECISION percent %of-chk
   > no lat/lon            0     0.0     0.0
   > no normal        675037    16.5    16.5
   > out-of-range        744     0.0     0.0
   > duplicated      4100117   100.2   120.1
   > accepted        -685075   -16.7
   > Dumping years 1901-2006 to .txt files...
   > Failed to create file. Try again.
   > Enter the file, with suffix: .ann
tmp.ann
   > Failed to create file. Try again.
   > Enter the file, with suffix: .ann
h.ann

crua6[/cru/cruts/version_3_0/primaries/temp]
```
<END QUOTE>

So the 'duplicated' figure is slightly lower.. but what's this
error with the '.ann' file?! Never seen before. Oh GOD if I
could start this project again and actually argue the case for
junking the inherited program suite!!

OK.. the .ann file was simply that it refuses to overwrite any
existing one. Meh. It's happy to overwrite the log file of
course - nice bit of logic there.

and the duplicates? Well I inserted a debug line where the
decision is made. Here's an example:

```
 712600 vs.  727340:      4.7     8.4     4.7     8.4 ->      0.0km
```

Here the two WMO codes look OK (though others are -999 which
seems unlikely) but the two lat/lon pairs? Ooops. Here are the
actual headers:

```
 712600   465    845  187 Sault Ste Marie A    CANADA        1945 2006 361945  -999.00
 727340   465    844  220 SAULT-STE-MARIE----- USA---------  1888 2006 101888  -999.00
```

So, uhhhh.. what in tarnation is going on? Just how off-beam
are these datasets?!!

Not sure why the lats & lons are a factor of 10 too low - may
be intentional though it wasn't happening before.

Ran with the original database:

<BEGIN QUOTE>
```
   > NORMALS             MEAN percent      STDEV percent
   >         .dtb    2113609    81.7
 made it to here
   >         .cts          0     0.0    2113608    81.7
   > PROCESS        DECISION percent %of-chk
   > no lat/lon            0     0.0     0.0
   > no normal        474422    18.3    18.3
   > out-of-range      68179     2.6     3.2
   > duplicated       923258    35.7    45.1
   > accepted        1122172    43.4
   > Dumping years 1901-1990 to .txt files...
```
<END QUOTE>

The lats & lons look the same.. but a lot less duplicates!

WHY? Well, it could just be those pesky US stations.. so
why not compare the two bespoke log files (as excerpted above)?

Immediately, another baffler: the log file from the run of
the 'final' database has lots of 'DEBUG DETAIL' information,
but the log file from the run of the original database does not!
So cropping those away with a judicious 'tail'.. I ran comm:

```
crua6[/cru/cruts/version_3_0/primaries/temp] comm -23 log_anomdtb_H.0702091122.dat barelog_anomdtb_H.0704300053.dat |wc -l
       200
crua6[/cru/cruts/version_3_0/primaries/temp] comm -13 log_anomdtb_H.0702091122.dat barelog_anomdtb_H.0704300053.dat | wc -l
      2572
crua6[/cru/cruts/version_3_0/primaries/temp] comm -12 log_anomdtb_H.0702091122.dat barelog_anomdtb_H.0704300053.dat | wc -l
      1809
```

So 200 duplication events are unique to the older database,
and 2572 are unique to the new database - with 1809 common
to both. A quick look at the 2572 'new' ones showed a majority
of those with the first WMO as -999: this is the key. The
databases do not have any records with WMO=-999 as far as I know,
so something is going on..

28. With huge reluctance, I have dived into 'anomdtb' - and already I have
that familiar Twilight Zone sensation.

I have found that the WMO Code gets set to -999 if *both* lon and lat are
missing. However, the following points are relevant:

* LoadCTS multiplies non-missing lons by 0.1, so they range from -18 to +18
  with missing value codes passing through AS LONG AS THEY ARE -9999. If they
  are -999 they will be processed and become -99.9. It is not clear why lats
  are not treated in the same way!

* The subroutine 'Anomalise' in anomdtb checks lon and lat against a simple
  'MissVal', which is defined as -999. This will catch lats of -999 but not
  lons of -9999.

* This does still not explain how we get so many -999 codes.. unless we don't
  and it's just one or two?

And the real baffler:

* If the code is -999 because lat and lon are both missing - how the bloody
  hell does it know there's a duplication within 8km?!!!

.. ah, OK. well for a start, the last point above does not apply - not one
case of the code being set to -999 because of lat/lon missing. In fact, I
hate to admit it, bit it is *sort of* clever - the code is set to -999 to
prevent it being used again, because the distance/duplication checker will
not make a distance comparison if either code is -999. So HOW COME loads of
the duplicates have a code of -999?!!!

The plot thickens.. I changed the exclusion tests in the duplication loops
from:

```
    if (AStn(XAStn).NE.MissVal) then
```

to:

```
    if (int(AStn(XAStn)).NE.-999) then
```

This made NO DIFFERENCE. So having tested to ensure that the first of the
pair hasn't already been used - we then use it! What's more I've noticed
that it's usually the one 'incorporated' in the previous iteration!

Consider:

```
  67700 vs.  160660:      4.6    -0.9     4.6    -0.9 ->      5.4km
   -999 vs.  160707:      4.6    -0.9     4.6    -0.9 ->      2.2km
   -999 vs.  160800:      4.6    -0.9     4.5    -0.9 ->      7.3km
   -999 vs.  160811:      4.6    -0.9     4.6    -0.9 ->      5.8km
```

Here we can see (check the first set of lat/lons) that, after being
incorporated into 160660, 67700 goes on to also be incorporated into
160707, 160800 and 160811! So the same data could end up in three
other stations. It gets worse!! Because later on, we find:

```
 160660 vs.  160707:      4.6    -0.9     4.6    -0.9 ->      7.9km
   -999 vs.  160800:      4.6    -0.9     4.5    -0.9 ->      7.0km
   -999 vs.  160811:      4.6    -0.9     4.6    -0.9 ->      5.8km
 160707 vs.  160800:      4.6    -0.9     4.5    -0.9 ->      7.9km
   -999 vs.  160811:      4.6    -0.9     4.6    -0.9 ->      6.6km
 160800 vs.  160811:      4.5    -0.9     4.6    -0.9 ->      2.2km
```

So three of those recipients have gone on to be incorporated into one
of them (160811). But although in this case 67700 is within 8km of
160811, there is no guarantee! Indeed, with this system, the 'chosen'
station may hop all over the place in <8km steps, collecting data as
it goes. In a densely-packed area this could drastically reduce the
number of stations. Then there's these:

```
  85997 vs.  390000:    -10.0   -20.0   -10.0   -20.0 ->      0.0km
   -999 vs.  685807:    -10.0   -20.0   -10.0   -20.0 ->      0.0km
   -999 vs.  688607:    -10.0   -20.0   -10.0   -20.0 ->      0.0km
   -999 vs.  967811:    -10.0   -20.0   -10.0   -20.0 ->      0.0km
   -999 vs.  968531:    -10.0   -20.0   -10.0   -20.0 ->      0.0km
```

as might be guessed, they all end up incorporated into 968531 - but
no surprise seeing as their lats & lons are rubbish!!! Oh Tim what
have you done, man? [actually - what he's done is to let missing
lats & lons through. Missing lon code is -1999 not -9999 so these
figures are the roundings]

All that said, the biggest worry is still the lats & lons themselves.
They just don't look realistic. Lats appear to have been reduced by
a factor of 10 too, even though I can't find the code for that. And
(from the top example) is 67700 really 5.4km from 160660?

```
  67700   460    -90  273 LUGANO               SWITZERLAND   1864 2006 101864  -999.00
 160660   456    -87 -999 MILANO MALPENSA      ITALY         1961 1970 101961  -999.00
```

Of course not! It's just over 50km. I do not understand why the lats
& lons have been scaled, when the stated distance threshold has not.

At least I've found *where* they are scaled, in LoadCTS (crutsfiles.f90):

```
    if (StnInfo(XStn,2).NE.LatMissVal) Lat (XStn) = real(StnInfo(XStn,2)) / real(LatFactor)
    if (StnInfo(XStn,3).NE.LonMissVal) Lon (XStn) = real(StnInfo(XStn,3)) / real(LonFactor)
```

Looking at how LoadCTS is called from anomdtb..

```
subroutine LoadCTS (StnInfo,StnLocal,StnName,StnCty,Code,Lat,Lon,Elv,OldCode,Data,YearAD,&
        NmlData,DtbNormals,CallFile,Hulme,Legacy,HeadOnly,HeadForm,LongType,Silent,Extra,PhilJ, &
        YearADMin,YearADMax,Source,SrcCode,SrcSuffix,SrcDate, &
        LatMV,LonMV,ElvMV,DataMV,LatF,LonF,ElvF,NmlYr0,NmlYr1,NmlSrc,NmlInc)

  call LoadCTS (StnInfoA,StnLocalA,StnNameA,StnCtyA,Code=AStn,OldCode=AStnOld, &
                        Lat=ALat,Lon=ALon,Elv=AElv,DtbNormals=DtbNormalsA, &
                        Data=DataA,YearAD=AYearAD,CallFile=LoadFileA,silent=1)  ! get .dtb file
```

.. we see that Legacy is not passed. This means that.. (from LoadCTS):

```
LatFactor=100 ; LonFactor=100 ; ElvFactor=1                    ! usual/hulme hdr factors
if (present(Legacy)) then
  LatFactor=10 ; LonFactor=10 ; ElvFactor=1                    ! legacy hdr factors
end if
if (present(LatF)) LatFactor = LatF                            ! custom hdr factors
if (present(LonF)) LonFactor = LonF
if (present(ElvF)) ElvFactor = ElvF
```

..LatFactor and LonFactor are set to 100.

So I added a specific pair of arguments, LatF=10,LonF=10, and got:

```
   > NORMALS            MEAN percent      STDEV percent
```

```
   >         .dtb    3323823    81.3
 made it to here
   >         .cts      91963     2.2    3415786    83.5
   > PROCESS        DECISION percent %of-chk
   > no lat/lon            0     0.0     0.0
   > no normal        675037    16.5    16.5
   > out-of-range        744     0.0     0.0
   > duplicated        53553     1.3     1.6
   > accepted        3361489    82.2
   > Dumping years 1901-2006 to .txt files...

Hurrah! Looking at the log it is still ignoring the -999 Code and re-intgrating stations..
but not to any extent worth worrying about. Not when duplications are down to 1.3% :-)))

Then got a mail from PJ to say we shouldn't be excluding stations inside 8km anyway - yet
that's in IJC - Mitchell & Jones 2005! So there you go. Ran again with 0km as the distance:

   > NORMALS            MEAN percent      STDEV percent
   >         .dtb    3323823    81.3
 made it to here
   >         .cts      91963     2.2    3415786    83.5
   > PROCESS        DECISION percent %of-chk
   > no lat/lon            0     0.0     0.0
   > no normal        675037    16.5    16.5
   > out-of-range        744     0.0     0.0
   > accepted        3415042    83.5
   > Dumping years 1901-2006 to .txt files...

Which hasn't saved much as it turns out. In fact, I must conclude that an inquiring mind is
a very dangerous thing - I decided to see what difference it made, turning off the proximity
duplicate detection and elimination:

crua6[/cru/cruts/version_3_0/primaries/temp] wc -l */*1962.12.txt
      2773 oldtxt/old.1962.12.txt
      3269 tmptxt0km/tmp.1962.12.txt
      3308 tmptxt8km/tmp.1962.12.txt

So.. 'oldtxt' is before I fixed the lat/lon scaling problem. But look at the last two - I
got MORE results when I used an elimination radius! Whaaaaaaaaat?!!!

/goes home in a huff

/gets out of huff and goes into house, checks things and thinks hard

Okay, I guess if we don't do the roll-duplicates-together thing, then we could lose data
because the 'rolled' station (ie the one subsumed into its neighbour) might have useful
years but no normals, so that data would be lost?

29. I suddenly thought - what about the Australian data? But luckily that's just tmax/tmin
so I can roll that into the next database work.

30. Being an idiot much experience I decided to go back to the 'perfectly-good' precip
generation for v3.0 and re-do the anomalies with the new anomdtb. At 8km, we got the
duplicates down from 5.9% to 2.1%:

<OLD ANOMDTB WITH LATLON PROBS>
   > NORMALS            MEAN percent      STDEV percent
   >         .dtb    7315040    73.8
 made it to here
   >         .cts     299359     3.0    7613600    76.8
   > PROCESS        DECISION percent %of-chk
   > no lat/lon         17527     0.2     0.2
   > no normal        2355659    23.8    23.8
   > out-of-range       13253     0.1     0.2
   > duplicated        586206     5.9     7.8
   > accepted         6934807    70.0
   > Dumping years 1901-2006 to .txt files...

<NEW ANOMDTB WITH LATLON 'FIXED'>
   > NORMALS            MEAN percent      STDEV percent
   >         .dtb    7315040    73.8
 made it to here
   >         .cts     299359     3.0    7613600    76.8
   > PROCESS        DECISION percent %of-chk
   > no lat/lon         17527     0.2     0.2
   > no normal        2355659    23.8    23.8
   > out-of-range       13253     0.1     0.2
   > duplicated        207391     2.1     2.8
   > accepted         7313622    73.8
   > Dumping years 1901-2006 to .txt files...

And, of course, all in with 0km range:

   > NORMALS            MEAN percent      STDEV percent
   >         .dtb    7315040    73.8
 made it to here
   >         .cts     299359     3.0    7613600    76.8
   > PROCESS        DECISION percent %of-chk
   > no lat/lon         17527     0.2     0.2
   > no normal        2355659    23.8    23.8
   > out-of-range       13253     0.1     0.2
   > accepted         7521013    75.9
   > Dumping years 1901-2006 to .txt files...

Happy? well.. no. Because something is happening for precip that does not happen for
temp! But of course. Here are the first few lines from various 1962.12 text files..

tmptxt8km/tmp.1962.12.txt
   70.90    8.70    10.0      2.10000  10010
   78.30  -15.50    28.0     -3.30000  10080
   69.70  -18.90    10.0     -1.40000   -999
   69.70  -18.90   100.0     -1.50000  10260
   74.50  -19.00    16.0     -1.20000  10280
   69.50  -25.50   129.0     -3.10000  10650
   70.40  -31.10    14.0     -0.20000  10980
   66.00   -2.00     0.0      0.50000  11000
   67.30  -14.40    13.0     -1.00000  11520
   66.80  -14.00    39.0     -0.70000  11530

tmptxt0km/tmp.1962.12.txt
   70.90    8.70    10.0      2.10000  10010
   78.30  -15.50    28.0     -3.30000  10080
   69.70  -18.90    10.0     -1.40000  10250
   69.70  -18.90   100.0     -1.50000  10260
   74.50  -19.00    16.0     -1.20000  10280
   69.50  -25.50   129.0     -3.10000  10650
   70.40  -31.10    14.0     -0.20000  10980
   66.00   -2.00     0.0      0.50000  11000
   67.30  -14.40    13.0     -1.00000  11520
   66.80  -14.00    39.0     -0.70000  11530
```

```
preanoms/pre.1962.12.txt (old anomdtb output)
   61.00   10.60   190.0     48.20000-511900
   54.45   -6.07   116.0     -3.70000   -999
   50.83   -4.55    15.0    -22.40000-389870
   50.22   -5.30    76.0     39.70000   -999
   50.63   -3.45     9.0    -28.10000-388730
   51.43   -2.67    51.0    -36.90000   -999
   51.05   -3.60   314.0    -27.80000-386030
   51.72   -2.77   245.0    -37.70000-385850
   51.62   -3.97    10.0    -46.10000-384130
   52.35   -3.82   301.0     -4.40000-380860

pretxt8km/pre.1962.12.txt
  610.00  106.00   190.0     48.20000-511900
  544.50  -60.70   116.0     -3.70000-392380
  508.30  -45.50    15.0    -22.40000-389870
  502.20  -53.00    76.0     39.70000-389280
  506.30  -34.50     9.0    -28.10000-388730
  514.30  -26.70    51.0    -36.90000-386780
  510.50  -36.00   314.0    -27.80000-386030
  517.20  -27.70   245.0    -37.70000-385850
  516.20  -39.70    10.0    -46.10000-384130
  523.50  -38.20   301.0     -4.40000-380860

pretxt0km/pre.1962.12.txt
  610.00  106.00   190.0     48.20000-511900
  544.50  -60.70   116.0     -3.70000-392380
  508.30  -45.50    15.0    -22.40000-389870
  502.20  -53.00    76.0     39.70000-389280
  506.30  -34.50     9.0    -28.10000-388730
  514.30  -26.70    51.0    -36.90000-386780
  510.50  -36.00   314.0    -27.80000-386030
  517.20  -27.70   245.0    -37.70000-385850
  516.20  -39.70    10.0    -46.10000-384130
  523.50  -38.20   301.0     -4.40000-380860

..As a result of fixing the lats and lons for temperature, and indeed
precip it seems, we have buggered up the outputs!!! Obviously the
correction factor is expecting 100 not 10, but why isn't this a problem
for temperature?! Went back and ran exactly the same version of anomdtb
on temperature - exactly the same as last time (2nd from top above). So
it is precip specific (or, erm, .not.temp specific?).

On the other hand, we've fixed the -999 WMO codes..

..and actually, those anomalies had better be percentage anomalies!

(checks a few) - yes, they are :-)

So oookay, LoadCTS reports the divisor is still 10 for lon/lat, so the
stored values for the first station (-511900, BIRI) should be 61 and 10.6,
sounds about right for Norway. The bit in anomdtb (actually the subroutine
'Dumping', LOL) that writes the .txt files just writes directly from the
arrays.. so they must have been modified somewhere in 'Anomalise' (there's
nothing else in 'Dumping'). Modified anomdtb to dump the first station's
lat & lon at key stages - they were too high throughout, so LoadCTS assumed
to be the troublemaker. Modified LoadCTS in the same way, and it was
holding them at x100 from their true values, ie 61.0 -> 6100. It was about
now that I spotted something I'd not thought to examine before: precip
headers use two decimal places for their coordinates!

Temperature header:
  10010   709     87   10 Jan Mayen            NORWAY        1921 2006 341921  -999.00

Precipitation header:
 100100  7093   -867   10 JAN MAYEN            NORWAY        1921 2006   -999  -999.00

So.. this begs the question, how does the software suite know which it's got?
By rights it should look at the most extreme values for each.. something tells
me that's not the case. Decided to look at the ranges of values for different
versions of the databases, starting with temperature:

crua6[/cru/cruts] head -1 fromdpe1a/data/cruts/database/+norm/tmp.0311051552.dtb
-990017 -9999 -99999 -999 UNKNOWN             MARINE        1948 1990   -999  -999.00
crua6[/cru/cruts] head -1 fromdpe1a/data/cruts/database/+norm/_old/tmp.0310311715.dtb
-176000  3520   3330  220 NICOSIA             CYPRUS        1932 1974   -999   nocode
crua6[/cru/cruts] head -1 rerun1/data/cruts/rerun_tmp/tmp.0311051552.dtb
-990017 -9999 -99999 -999 UNKNOWN             MARINE        1948 1990   -999  -999.00
crua6[/cru/cruts] head -1 rerun1/data/cruts/rerun_tmp/tmp.0311051552n.dtb
-990017 -9999 -99999 -999 UNKNOWN             MARINE        1948 1990   -999  -999.00
crua6[/cru/cruts] head -1 rerun1/data/cruts/rerun_tmp/database/+norm/_old/tmp.0310311715.dtb
-176000  3520   3330  220 NICOSIA             CYPRUS        1932 1974   -999   nocode
crua6[/cru/cruts] head -1 rerun1/data/cruts/rerun_tmp/database/+norm/tmp.0311051552.dtb
-990017 -9999 -99999 -999 UNKNOWN             MARINE        1948 1990   -999  -999.00
crua6[/cru/cruts] head -1 rerun1/data/cruts/rerun_tmp/database/tmp.0311051552.dtb
-990017 -9999 -99999 -999 UNKNOWN             MARINE        1948 1990   -999  -999.00
crua6[/cru/cruts] head -1 version_3_0/primaries/temp/tmp.0702091122.dtb
  10010   709     87   10 Jan Mayen            NORWAY        1921 1990 341921  -999.00
crua6[/cru/cruts] head -1 version_3_0/primaries/temp/tmp.0704300053.dtb
  10010   709     87   10 Jan Mayen            NORWAY        1921 2006 341921  -999.00
crua6[/cru/cruts] head -1 version_3_0/db/testmergedb/tmp.0702091122.dtb
  10010   709     87   10 Jan Mayen            NORWAY        1921 1990 341921  -999.00
crua6[/cru/cruts] head -1 version_3_0/db/testmergedb/tmp.0704292355.dtb
  10010   709     87   10 Jan Mayen            NORWAY        1921 2006 341921  -999.00
crua6[/cru/cruts] head -1 version_3_0/db/testmergedb/badtimeline/tmp.0704251819.dtb
  10010   709     87   10 Jan Mayen            NORWAY        1921 2006 341921  -999.00
crua6[/cru/cruts] head -1 version_3_0/db/testmergedb/badtimeline/tmp.0704271015.dtb
  10010   709     87   10 Jan Mayen            NORWAY        1921 2006 341921  -999.00
crua6[/cru/cruts] head -1 version_3_0/db/testmergedb/badtimeline/tmp.0704292158.dtb
  10010   709     87   10 Jan Mayen            NORWAY        1921 2006 341921  -999.00
crua6[/cru/cruts] head -1 version_3_0/db/testmergedb/tmp.0704300053.dtb
  10010   709     87   10 Jan Mayen            NORWAY        1921 2006 341921  -999.00
crua6[/cru/cruts] head -1 version_3_0/db/tmp.0702091122.dtb
  10010   709     87   10 Jan Mayen            NORWAY        1921 1990 341921  -999.00
crua6[/cru/cruts] head -1 version_3_0/db/tmp.0704300053.dtb
  10010   709     87   10 Jan Mayen            NORWAY        1921 2006 341921  -999.00

Without going any further, it's obvious that LoadCTS is going to have to auto-
sense the lat and lon ranges. Missing value codes can then be derived - if it
always returns actual (unscaled) degrees (to one or two decimal places) then
any value lower than -998 will suffice for both parameters. However, this does
make me wonder why it wasn't done like that. Is there a likelihood of the
programs being used on a spatial subset of stations? Say, English? Then lon
would never get into double figures, though lat would.. well let's just hope
not! *laughs hollowly*

Okay.. so I wrote extra code into LoadCTS to detect Lat & Lon ranges. It excludes any
values for which the modulus of 100 is -99, so hopefully missing value codes do not
conribute. The factors are set accordingly (to 10 or 100). I had to default to 1 which
```

```
is a pity. Once you've got the factors, detection of missing values can be a simple
out-of-range test.

However *sigh* this led me to examine the detection of 'non-standard longitudes' - a
small section of code that converts PJ-style reversed longitudes, or 0-360 ones, to
regular -180 (W) to +180 (E). This code is switched on by the presence of the
'LongType' flag in the LoadCTS call - the trouble is, THAT FLAG IS NEVER SET BY
ANOMDTB. There is a declaration 'integer :: QLongType' but that is never referred to
again. Just another thing I cannot understand, and another reason why this should all
have been rewritten from scratch a year ago!

So, I wrote 'revlons.for' - a proglet to reverse all longitude values in a database
file. Ran it on the temperature database (final):

<BEGIN QUOTE>
crua6[/cru/cruts/version_3_0/db/testmergedb] ./revlons
REVLONS - Reverse All Longitudes!

This nifty little proglet will fix all of your
longitudes so that they point the right way, ie,
positive = East of Greenwich, negative = West.

..of course, if they are already fixed, this will
UNfix them. I am not that smart! So be careful!!

Please enter the database to be fixed: tmp.0704300053.dtb

Output file will be: tmp.0705101334.dtb
Confirm this filename (Y/N): Y

Log file will be:    tmp.0705101334.log

  5065 stations written to tmp.0705101334.dtb

<END QUOTE>

Thus the 'final' temperature database is now tmp.0705101334.dtb.

Re-ran anomdtb - with working lat/lon detection and missing lat/lon value
detection - for both precip and temperature. This should ensure that all
WMO codes are present and all lats and lons are correct.

Temp:
<BEGIN QUOTE>
   > ***** AnomDTB: converts .dtb to anom .txt for gridding *****

   > Enter the suffix of the variable required:
.tmp
   > Select the .cts or .dtb file to load:
tmp.0705101334.dtb

   > Specify the start,end of the normals period:
1961,1990
   > Specify the missing percentage permitted:
25
   > Data required for a normal:           23
   > Specify the no. of stdevs at which to reject data:
3
   > Select outputs (1=.cts,2=.ann,3=.txt,4=.stn):
3
   > Check for duplicate stns after anomalising? (0=no,>0=km range)
0
   > Select the generic .txt file to save (yy.mm=auto):
tmp.txt
   > Select the first,last years AD to save:
1901,2006
   > Operating...

   > NORMALS            MEAN percent      STDEV percent
   >         .dtb    3323823    81.3
   >         .cts      91963     2.2    3415786    83.5
   > PROCESS        DECISION percent %of-chk
   > no lat/lon         1993     0.0     0.0
   > no normal        673044    16.5    16.5
   > out-of-range        744     0.0     0.0
   > accepted        3415042    83.5
   > Dumping years 1901-2006 to .txt files...
<END QUOTE>

Precip:
<BEGIN QUOTE>
crua6[/cru/cruts/version_3_0/primaries/precip] ./anomdtb

   > ***** AnomDTB: converts .dtb to anom .txt for gridding *****

   > Enter the suffix of the variable required:
.pre
   > Will calculate percentage anomalies.
   > Select the .cts or .dtb file to load:
pre.0612181221.dtb

   > Specify the start,end of the normals period:
1961,1990
   > Specify the missing percentage permitted:
25
   > Data required for a normal:           23
   > Specify the no. of stdevs at which to reject data:
4
   > Select outputs (1=.cts,2=.ann,3=.txt,4=.stn):
3
   > Check for duplicate stns after anomalising? (0=no,>0=km range)
0
   > Select the generic .txt file to save (yy.mm=auto):
pre.txt
   > Select the first,last years AD to save:
1901,2006
   > Operating...

   > NORMALS            MEAN percent      STDEV percent
   >         .dtb    7315040    73.8
   >         .cts     299359     3.0    7613600    76.8
   > PROCESS        DECISION percent %of-chk
   > no lat/lon        17911     0.2     0.2
   > no normal       2355275    23.8    23.8
   > out-of-range      13253     0.1     0.2
   > accepted        7521013    75.9
   > Dumping years 1901-2006 to .txt files...
<END QUOTE>

Note that precip accepted values is up to 75.9%, I honestly don't
```

think we'll get higher.

Decided to process temperature all the way. Ran IDL:

IDL> quick_interp_tdm2,1901,2006,'tmpglo/tmpgrid.',1200,gs=0.5,dumpglo='dumpglo',pts_prefix='tmp0km0705101334txt/tmp.'

then glo2abs, then mergegrids, to produce monthly output grids. It apparently worked:

```
-rw-------   1 f098     cru      138964083 May 13 20:42 cru_ts_3_00.1901.2006.tmp.dat.gz
-rw-------   1 f098     cru        7852589 May 13 20:42 cru_ts_3_00.2001.2006.tmp.dat.gz
-rw-------   1 f098     cru       13108065 May 13 20:39 cru_ts_3_00.1991.2000.tmp.dat.gz
-rw-------   1 f098     cru       13106515 May 13 20:36 cru_ts_3_00.1981.1990.tmp.dat.gz
-rw-------   1 f098     cru       13106963 May 13 20:33 cru_ts_3_00.1971.1980.tmp.dat.gz
-rw-------   1 f098     cru       13123939 May 13 20:30 cru_ts_3_00.1961.1970.tmp.dat.gz
-rw-------   1 f098     cru       13120586 May 13 20:26 cru_ts_3_00.1951.1960.tmp.dat.gz
-rw-------   1 f098     cru       13120691 May 13 20:23 cru_ts_3_00.1941.1950.tmp.dat.gz
-rw-------   1 f098     cru       13130077 May 13 20:20 cru_ts_3_00.1931.1940.tmp.dat.gz
-rw-------   1 f098     cru       13104881 May 13 20:16 cru_ts_3_00.1921.1930.tmp.dat.gz
-rw-------   1 f098     cru       13094948 May 13 20:13 cru_ts_3_00.1911.1920.tmp.dat.gz
-rw-------   1 f098     cru       13085509 May 13 17:08 cru_ts_3_00.1901.1910.tmp.dat.gz
```

As a reminder, these output grids are based on the tmp.0705101334.dtb database, with no
merging of neighbourly stations and a limit of 3 standard deviations on anomalies.

Decided to (re-) process precip all the way, in the hope that I was in the zone or
something. Started with IDL:

IDL> quick_interp_tdm2,1901,2006,'preglo/pregrid.',450,gs=0.5,dumpglo='dumpglo',pts_prefix='pre0km0612181221txt/pre.'

Then glo2abs, then mergegrids.. all went fine, apparently.

31. And so.. to DTR! First time for generation I think.

Wrote 'makedtr.for' to tackle the thorny problem of the tmin and tmax databases not
being kept in step. Sounds familiar, if worrying. am I the first person to attempt
to get the CRU databases in working order?!! The program pulls no punches. I had
already found that tmx.0702091313.dtb had seven more stations than tmn.0702091313.dtb,
but that hadn't prepared me for the grisly truth:

<BEGIN QUOTE>
```
crua6[/cru/cruts/version_3_0/db/dtr] ./makedtr

      MAKEDTR - Produce a DTR database

  This program takes as its input a database of
  of minimum temperatures and another of maximum
  temperatures, and produces a database of diurnal
  temperatures. If the input databases are found
  to be out of synchronisation, the option is also
  given to save synchronised versions.

So, may I please have the tmin database? tmn.0702091139.dtb

May I please now have the tmax database? tmx.0702091313.dtb

The output database will now be called:  dtr.0705152339.dtb

  IMPORTANT: PLEASE READ! (it's good for you)

  The databases you gave are NOT synchronised!

tmn.0702091139.dtb has   42 'extra' stations

tmx.0702091313.dtb has   49 'extra' stations
You have the choice of quitting, or of allowing
me to create two new synchronised databases,
which will be saved and used to create the dtr db

  Enter Q to Quit, S to Synchronise: S
  New tmin database is: tmn.0705152339.dtb
  Discarded tmin stations here: tmn.0702091139.dtb.del
  New tmax database is: tmx.0705152339.dtb
  Discarded tmax stations here: tmx.0702091313.dtb.del
Number of stations to process:    14267
```
<END QUOTE>

Yes, the difference is a lot more than seven! And the program helpfully dumps a listing
of the surplus stations to the log file. Not a pretty sight.

Unfortunately, it hadn't worked either. It turns out that there are 3518 stations in
each database with a WMO Code of '   0'. So, as the makedtr program indexes on the
WMO Code.. you get the picture. *cries*

Rewrote as makedtr2, which uses the first 20 characters of the header to match:

<BEGIN QUOTE>
```
      MAKEDTR2 - Produce a DTR database

  This program takes as its input a database of
  of minimum temperatures and another of maximum
  temperatures, and produces a database of diurnal
  temperatures. If the input databases are found
  to be out of synchronisation, the option is also
  given to save synchronised versions.

So, may I please have the tmin database? tmn.0702091139.dtb

May I please now have the tmax database? tmx.0702091313.dtb

The output database will now be called:  dtr.0705162028.dtb

  IMPORTANT: PLEASE READ! (it's good for you)

  The databases you gave are NOT synchronised!

tmn.0702091139.dtb has  203 'extra' stations

tmx.0702091313.dtb has  209 'extra' stations
You have the choice of quitting, or of allowing
me to create two new synchronised databases,
which will be saved and used to create the dtr db

  Enter Q to Quit, S to Synchronise: S
  New tmin database is: tmn.0705162028.dtb
  Discarded tmin stations here: tmn.0702091139.dtb.del
  New tmax database is: tmx.0705162028.dtb
  Discarded tmax stations here: tmx.0702091313.dtb.del
```
<END QUOTE>

The big jump in the number of 'surplus' stations is because we are no longer automatically

```
matching stations with WMO=0.

Here's what happened to the tmin and tmax databases, and the new dtr database:

Old tmin: tmn.0702091139.dtb      Total Records Read:    14309
New tmin: tmn.0705162028.dtb      Total Records Read:    14106
Del tmin: tmn.0702091139.dtb.del  Total Records Read:      203

Old tmax: tmx.0702091313.dtb      Total Records Read:    14315
New tmax: tmx.0705162028.dtb      Total Records Read:    14106
Del tmax: tmx.0702091313.dtb.del  Total Records Read:      209

New dtr:  dtr.0705162028.dtb      Total Records Read:    14107

*sigh* - one record out! Also three header problems:

BLANKS (expected at 8,14,21,26,47,61,66,71,78)
  position   missed
       8        1
      14        1
      21        0
      26        0
      47        1
      61        0
      66        0
      71        0
      78        0

Why?!! Well the sad answer is.. because we've got a date wrong. All three 'header' problems
relate to this line:

6190   94   95   98  100  101  101  102  103  102   97   94   94

..and as we know, this is not a conventional header. Oh bum. But, but.. how? I know we do
muck around with the header and start/end years, but still..

Wrote filtertmm.for, which simply steps through one database (usually tmin) and
looks for a 'perfect' match in another database (usually tmax). 'Perfect' here
means a match of WMO Code, Lat, Lon, Start-Year and End-Year. If a match is
found, both stations are copied to new databases:

<BEGIN QUOTE>
crua6[/cru/cruts/version_3_0/db/dtr] ./filtertmm

  FILTERTMM - Create GOOD tmin/max databases

Please enter the tmin database: tmn.0702091139.dtb

Please enter the tmax database: tmx.0702091313.dtb

  working..

Old tmin database: tmn.0702091139.dtb had   14309 stations
New tmin database: tmn.0705182204.dtb has   13016 stations
Old tmax database: tmx.0702091313.dtb had   14315 stations
New tmax database: tmx.0705182204.dtb has   13016 stations
<END QUOTE>

I am going to *assume* that worked! So now.. to incorporate the Australian
monthly data packs. Ow. Most future-proof strategy is probably to write a
converter that takes one or more of the packs and creates CRU-format databases
of them. Edit: nope, thought some more and the *best* strategy is a program
that takes *pairs* of Aus packs and updates the actual databases. Bearing in
mind that these are trusted updates and won't be used in any other context.

From Dave L - who incorporated the initial Australian dump - for the tmin/tmax bulletins,
he used a threshold of 26 days/month or greater for inclusion.

Obtained two files from Dave - an email that explains some of the Australian
bulletin data/formatting, and a list of Austraian headers matched with their
internal codes (the latter being generated by Dave).

Actually.. although I was going to assume that filtertmm had done the synching job OK, a
brief look at the Australian stations in the databases showed me otherwise. For instance,
I pulled all the headers with 'AUSTRALIA' out of the two 0705182204 databases. Now because
these were produced by filtertmm, we know that the codes (if present), lats, lons and dates
will all match. Any differences will be in altitude and/or name. And so they were:

crua6[/cru/cruts/version_3_0/db/dtr] diff tmn.0705182204.dtb.oz tmx.0705182204.dtb.oz | wc -l
      336

..so roughly 100 don't match. They are mostly altitude discrepancies, though there are an
alarming number of name mismatches too. Examples of both:

74c74
<       0 -3800  14450   11 AVALON AIRPORT           AUSTRALIA 2000 2006   -999  -999.00
---
>       0 -3800  14450    8 AVALON AIRPORT           AUSTRALIA 2000 2006   -999  -999.00

16c16
<       0 -4230  14650  585 TARRALEAH VILLAGE        AUSTRALIA 2000 2006   -999  -999.00
---
>       0 -4230  14650  595 TARRALEAH CHALET         AUSTRALIA 2000 2006   -999  -999.00

Examples of the second kind (name mismatch) are most concerning as they may well be
different stations. Looked for all occurences in all tmin/tmax databases:

crua6[/cru/cruts/version_3_0/db/dtr] grep 'TARRALEAH' *dtb
tmn.0702091139.dtb:      0 -4230  14650  585 TARRALEAH VILLAGE        AUSTRALIA 2000 2006   -999  -999.00
tmn.0702091139.dtb:9597000 -4230  14645  595 TARRALEAH CHALET     AUSTRALIA     1991 2000   -999  -999.00
tmn.0705182204.dtb:      0 -4230  14650  585 TARRALEAH VILLAGE        AUSTRALIA 2000 2006   -999  -999.00
tmn.0705182204.dtb:9597000 -4230  14645  595 TARRALEAH CHALET     AUSTRALIA     1991 2000   -999  -999.00
tmx.0702091313.dtb:      0 -4230  14650  595 TARRALEAH CHALET         AUSTRALIA 2000 2006   -999  -999.00
tmx.0702091313.dtb:9597000 -4230  14645  595 TARRALEAH CHALET     AUSTRALIA     1991 2000   -999  -999.00
tmx.0705182204.dtb:      0 -4230  14650  595 TARRALEAH CHALET         AUSTRALIA 2000 2006   -999  -999.00
tmx.0705182204.dtb:9597000 -4230  14645  595 TARRALEAH CHALET     AUSTRALIA     1991 2000   -999  -999.00

This takes a little sorting out. Well first, recognise that we are dealing with four files: tmin
and tmax, early and late (before and after filtertmm.for). We see there are two TARRALEAH entries
in each of the four files. We see that 'TARRALEAH VILLAGE' only appears in the tmin file. We see,
most importantly perhaps, that they are temporally contiguous - that is, each pair could join with
minimal overlap, as one is 1991-2000 and the other 2000-2006. Also, we note that the 'early' one
of each pair has a slightly different longitude and altitude (the former being the thing that
distinguished the stations in filtertmm.for).

Finally, this, from the tmax.2005120120051231.txt bulletin:

  95018, 051201051231, -42.30, 146.45,    18.0,        00,   31,  31,   585,    TARRALEAH VILLAGE

So we can resolve this case - a single station called TARRALEAH VILLAGE, running from 1991 to 2006.
```

But what about the others?! There are close to 1000 incoming stations in the bulletins, must every one be identified in this way?!! Oh God. There's nothing for it - I'll have to write a prog to find matches for the incoming Australian bulletin stations in the main databases. I'll have to use the databases from before the filtertmm application, so *0705182204.dtb. And it will only need the Australian headers, so I used grep to create *0705182204.dtb.auhead files. The other input is the list of stations taken from the monthly bulletins. Now these have a different number of stations each month, so the prog will build an array of all possible stations based on the files we have. Oh boy. And the program shall be called, 'auminmaxmatch.for'.

Assembled some information:

```
crua6[/cru/cruts/version_3_0/db] wc -l *auhead
      1518 glseries_tmn_final_merged.auhead
      1518 tmn.0611301516.dat.auhead
      1518 tmn.0612081255.dat.auhead
      1518 tmn.0702091139.dtb.auhead
      1518 tmn.0705152339.dtb.auhead
      1426 tmn.0705182204.dtb.auhead
```

(the 'auhead' files were created with <grep 'AUSTRALIA'>)

Actually, stopped work on that. Trying to match over 800 'bulletin' stations against over 3,000 database stations *in two unsynchronised files* was just hurting my brain. The files have to be properly synchronised first, with a more lenient and interactive version of filtertmm. Or... could I use mergedb?! Pretend to merge tmin into tmax and see what pairings it managed? No roll through obviously. Well it's worth a play.

..unfortunately, not. Because when I tried, I got a lot of odd errors followed by a crash. The reason, I eventually deduced, was that I didn't build mergedb with the idea that WMO codes might be zero (many of the australian stations have wmo=0). This means that primary matching on WMO code is impossible. This just gets worse and worse: now it looks as though I'll have to find WMO Codes (or pseudo-codes) for the *3521* stations in the tmin file that don't have one!!!

OK.. let's break the problem down. Firstly, a lot of stations are going to need WMO codes, if available. It shouldn't be too hard to find any matches with the existing WMO coded stations in the other databases (precip, temperature). Secondly, we need to exclude stations that aren't synchronised between the two databases (tmin/tmax). So can mergedb be modified to treat WMO codes of 0 as 'missing'? Had a look, and it does check that the code isn't -999 OR 0.. but not when preallocating flags in subroutine 'countscnd'. Fixed that and tried running it again.. exactly the same result (crash). I can't see anything odd about the station it crashes on:

```
      0 -2810  11790  407 MOUNT MAGNET AERO        AUSTRALIA 2000 2006   -999  -999.00
6190-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
2000  339  344  280  252  214  202  189  196  262  291  316  377
2001  371  311  310  300  235  212  201  217  249  262  314  333
2002-9999-9999  339  297  258  209  205  212  246  299  341  358
2003  365  367  336  296  249  195  193  200  238  287  325  368
2004  395  374  321  284  219  214  173  188  239  309  305  370
2005  389  396  358  315  251  182  189  201  233  267  332  341
2006  366  331  314  246  240-9999-9999-9999-9999-9999-9999-9999
```

.. it's very similar to preceding (and following) stations, and the station before has even less real data (the one before that has none at all and is auto-deleted). The nature of the crash is 'forrtl: error (65): floating invalid' - so a type mismatch possibly. The station has a match in the tmin database (tmn.0702091139.dtb) but the longitude is different:

```
tmn.0702091139.dtb:
      0 -2810  11780  407 MOUNT MAGNET AERO        AUSTRALIA 2000 2006   -999  -999.00
tmx.0702091313.dtb:
      0 -2810  11790  407 MOUNT MAGNET AERO        AUSTRALIA 2000 2006   -999  -999.00
```

It also appears in the tmin/tmax bulletins, eg:

```
 7600, 070401070430, -28.12, 117.84,    16.0,        00,   30,  30,   407,    MOUNT MAGNET AERO
```

Note that the altitude matches (as distinct from the station below).

Naturally, there is a further 'MOUNT MAGNET' station, but it's probably distinct:

```
tmn.0702091139.dtb:
9442800 -2807  11785  427 MOUNT MAGNET (MOUNT  AUSTRALIA     1956 1992   -999  -999.00
tmx.0702091313.dtb:
9442800 -2807  11785  427 MOUNT MAGNET (MOUNT  AUSTRALIA     1957 1992   -999  -999.00
```

I am at a bit of a loss. It will take a very long time to resolve each of these 'rogue' stations. Time I do not have. The only pragmatic thing to do is to dump any stations that are too recent to have normals. They will not, after all, be contributing to the output. So I knocked out 'goodnorm.for', which simply uses the presence of a valid normals line to sort. The results were pretty scary:

<BEGIN QUOTE>

```
crua6[/cru/cruts/version_3_0/db/dtr] ./goodnorm

  GOODNORM: Extract stations with non-missing normals

Please enter the input database name: tmn.0702091139.dtb
The output database will be called:   tmn.0705281724.dtb

(removed stations will be placed in: tmn.0705281724.del)

FINISHED.

Stations retained:   5026
Stations removed:    9283

crua6[/cru/cruts/version_3_0/db/dtr] ./goodnorm

  GOODNORM: Extract stations with non-missing normals

Please enter the input database name: tmx.0702091313.dtb
The output database will be called:   tmx.0705281724.dtb

(removed stations will be placed in: tmx.0705281724.del)

FINISHED.

Stations retained:   4997
Stations removed:    9318
```

<END QUOTE>

Essentially, two thirds of the stations have no normals! Of course, this still leaves us with a lot more stations than we had for tmean (goodnorm reported 3316 saved, 1749 deleted) though still far behind precipitation (goodnorm reported 7910 saved, 8027 deleted).

I suspect the high percentage lost reflects the influx of modern Australian data. Indeed, nearly 3,000 of the 3,500-odd stations with missing WMO codes were excluded by this operation. This means that, for tmn.0702091139.dtb, 1240 Australian stations were lost, leaving only 278.

This is just silly. I can't dump these stations, they are needed to potentially match with the

bulletin stations. I am now going to try the following:

1. Attempt to pair bulletin stations with existing in the tmin database. Mark pairings in the
   database headers and in a new 'Australian Mappings' file. Program auminmatch.for.

2. Run an enhanced filtertmm to synchronise the tmin and tmax databases, but prioritising the
   'paired' stations from step 1 (so they are not lost). Mark the same pairings in the tmax
   headers too, and update the 'Australian Mappings' file.

3. Add the bulletins to the databases.

OK.. step 1. Modified auminmaxmatch.for to produce auminmatch.for. Hit a semi-philosophical
problem: what to do with a positive match between a bulletin station and a zero-wmo database
station? The station must have a real WMO code or it'll be rather hard to describe the match!

Got a list of around 12,000 wmo codes and stations from Dave L; unfortunately there was a problem
with its formatting that I just couldn't resolve.

So.. current thinking is that, if I find a pairing between a bulletin station and a zero-coded
Australain station in the CRU database, I'll give the CRU database station the Australian local
(bulletin) code twice: once at the end of the header, and once as the WMO code *multiplied by -1*
to avoid implying that it's legitimate. Then if a 'proper' code is found or allocated later, the
mapping to the bulletin code will still be there at the end of the header. Of course, an initial
check will ensure that a match can't be found, within the CRU database, between the zero-coded
station and a properly-coded one.

Debated header formats with David. I think we're going to go with (i8,a8) at the end of the header,
though really it's (2x,i6,a8) as I remember the Anders code being i2 and the real start year being
i4 (both from the tmean database). This will mean post-processing existing databases of course,
but that's not a priority.

A brief (hopefully) diversion to get station counts sorted. David needs them so might as well sort
the procedure. In the upside-down world of Mark and Tim, the numbers of stations contributing to
each cell during the gridding operation are calculated not in the IDL gridding program - oh, no! -
but in anomdtb! Yes, the program which reads station data and writes station data has a second,
almost-entirely unrelated function of assessing gridcell contributions. So, to begin with it runs
in the usual way:

```
crua6[/cru/cruts/version_3_0/primaries/precip] ./anomdtb

   > ***** AnomDTB: converts .dtb to anom .txt for gridding *****

   > Enter the suffix of the variable required:
.pre
   > Will calculate percentage anomalies.
   > Select the .cts or .dtb file to load:
pre.0612181221.dtb

   > Specify the start,end of the normals period:
1961,1990
   > Specify the missing percentage permitted:
25
   > Data required for a normal:           23
   > Specify the no. of stdevs at which to reject data:
4
```

But then, we choose a different output, and it all shifts focus and has to ask all the IDL
questions!!

```
   > Select outputs (1=.cts,2=.ann,3=.txt,4=.stn):
4
   > Check for duplicate stns after anomalising? (0=no,>0=km range)
0
   > Select the .stn file to save:
pre.stn
   > Enter the correlation decay distance:
450
   > Submit a grim that contains the appropriate grid.
   > Enter the grim filepath:
clim.6190.lan.pre

   > Grid dimensions and domain size:      720     360   67420
   > Select the first,last years AD to save:
1901,2006
   > Operating...

   > NORMALS            MEAN percent      STDEV percent
   >         .dtb    7315040    73.8
   >         .cts     299359     3.0    7613600    76.8
   > PROCESS        DECISION percent %of-chk
   > no lat/lon       17911     0.2     0.2
   > no normal       2355275    23.8    23.8
   > out-of-range      13253     0.1     0.2
   > accepted        7521013    75.9
   > Calculating station coverages...
```

And then.. it unhelpfully crashes:

```
   > ##### WithinRange: Alloc: DataB #####
forrtl: severe (174): SIGSEGV, segmentation fault occurred
```

Ho hum. I did try this last year which is why I'm not tearing my hair out. The plan is to use the
outputs from the regular anomdtb runs - ie, the monthly files of valid stations. After all we need
to know the station counts on a per month basis. We can use the lat and lon, along with the
correlation decay distance.. shouldn't be too awful. Just even more programming and work. So before
I commit to that, a quick look at the IDL gridding prog to see if it can dump the figures instead:
after all, this is where the actual 'station count' information is assembled and used!!

..well that was, erhhh.. 'interesting'. The IDL gridding program calculates whether or not a
station contributes to a cell, using.. graphics. Yes, it plots the station sphere of influence then
checks for the colour white in the output. So there is no guarantee that the station number files,
which are produced *independently* by anomdtb, will reflect what actually happened!!

Well I've just spent 24 hours trying to get Great Circle Distance calculations working in Fortran,
with precisely no success. I've tried the simple method (as used in Tim O's geodist.pro, and the
more complex and accurate method found elsewhere (wiki and other places). Neither give me results
that are anything near reality. FFS.

Worked out an algorithm from scratch. It seems to give better answers than the others, so we'll go
with that. Also decided that the approach I was taking (pick a gridline of latitude and reverse-
engineer the GCD algorithm so the unknown is the second lon) was overcomplicated, when we don't
need to know where it hits, just that it does. Since for any cell the nearest point to the station
will be a vertex, we can test candidate cells for the distance from the appropriate vertex to the
station. Program is stncounts.for, but is causing immense problems.

The problem is, really, the huge numbers of cells potentially involved in one station, particularly
at high latitudes. Working out the possible bounding box when you're within cdd of a pole (ie, for
tmean with a cdd of 1200, the N-S extent is over 20 cells (10 degs) in each direction. Maybe not a
serious problem for the current datasets but an example of the complexity. Also, deciding on the

potential bounding box is nontrivial, because of cell 'width' changes at high latitudes (at 61 degs North, the half-degree cells are only 27km wide! With a precip cdd of 450 km this means the bounding box is dozens of cells wide - and will be wider at the Northern edge!

Clearly a large number of cells are being marked as covered by each station. So in densely-stationed areas there will be considerable smoothing, and in sparsely-stationed (or empty) areas, there will be possibly untypical data. I might suggest two station counts - one of actual stations contributing from within the cell, one for stations contributing from within the cdd. The former being a subset of the latter, so the latter could be used as the previous release was used.

Well, got stncounts.for working, finally. And, out of malicious interest, I dumped the first station's coverage to a text file and counted up how many cells it 'influenced'. The station was at 10.6E, 61.0N. The total number of cells covered was a staggering 476! Or, if you prefer, 475 indirect and one direct.

Ran for the first month (01/1901). Compared the resulting grid with that from CRU TS 2.1. Seems to compare fine, some higher, some lower. Example:

```
2.10:    139   142   146   154   156   157   165   170
3.00:    141   148   154   153   153   159   163   168
```

(data are on latitude #265 and longitudes #163-170)

Wrote 'makelsmask.for' to, well, make a land-sea mask. It'll work with any gridded data file that uses -999 for sea. The mask is called 'lsmask.halfdeg.dat'. Adapted stncounts.for to read it and use it to mask the output files.

Still a bit disturbed by the large number of cells marked as 'influenced' by a single station. IDL seems to use the inbuilt 'TRIGRID' function to interpolate the grid, so there's no way of getting the station count for a particular cell that way anyway. Not that it would mean much, since there is bound to be some kind of weighting (it's not clear what that weighting is, though, from the IDL website). So the figures in the station count files are really rather loose. What might be useful as a companion dataset would be the ACTUAL station counts. Counts for cells with stations actually INSIDE them. Of course, that might be rather sensitive information..

Managed a full run of stncounts. It took over five and a half hours, which is a bit much!

Back to the gridding. I am seriously worried that our flagship gridded data product is produced by Delaunay triangulation - apparently linear as well. As far as I can see, this renders the station counts totally meaningless. It also means that we cannot say exactly how the gridded data is arrived at from a statistical perspective - since we're using an off-the-shelf product that isn't documented sufficiently to say that. Why this wasn't coded up in Fortran I don't know - time pressures perhaps? Was too much effort expended on homogenisation, that there wasn't enough time to write a gridding procedure? Of course, it's too late for me to fix it too. Meh.

Well, it's been a real day of revelations, never mind the week. This morning I discovered that proper angular weighted interpolation was coded into the IDL routine, but that its use was discouraged because it was slow! Aaarrrgghh. There is even an option to tri-grid at 0.1 degree resolution and then 'rebin' to 720x360 - also deprecated! And now, just before midnight (so it counts!), having gone back to the tmin/tmax work, I've found that most if not all of the Australian bulletin stations have been unceremoniously dumped into the files without the briefest check for existing stations. A classic example would be these 'two' stations:

```
      0 -1570  12870   31 KIMBERLEY RES.STATIO     AUSTRALIA 2000 2006   -999  -999.00
6190-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
2000  245  243  243  232  184  143  138  155  193  231  249  249
2001  245  247  241  216  156  167  163  129  201  238  246  247
2002  244  246  230  208  167  122   92  119  202  217  248  259
2003  253  249  222  220  169  151  144  158  203  216  248  250
2004  252  247  244  209  202  135  129  140  176  230  248  257
2005  245  246  237-9999-9999-9999-9999-9999-9999-9999-9999-9999
2006-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999

      0 -1565  12871   31 KIMBERLEY RES.STATIO AUSTRALIA     1971 2000   -999  -999.00
6190-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1971  254  249  239  218  166  147  142  169  214  246  253  241
1972  246  244  226  198  175  158  126  182  200  222  244  259
1973  255  259  252  232  215  186  171  189  216  240  256  246
1974  247  243  240  217  183  144  134  171  216  247  248  246
1975  239  239  237  216  180  157  168  171  223  233  243  246
1976  235  244  227  190  148  142  142  144  177  236  252  250
1977  253  249  245  218  177  135  130  137  187  226  250  248
1978  247  244  239  199  218  174  162  186  195  233  245  253
1979  247  246  238  217  205  166  147  178  216  234  248  254
1980  249  245  240  221  186  161  141  171  192  241  249  252
1981-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1982-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1983-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1984-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1985-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1986-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1987-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1988-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1989-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1990-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1991  248  244  234  224  169  160  160  140  210  225  252  260
1992  253  251  247  239  206-9999  141  173  218  237  246  260
1993  247-9999  242  225  207  172  149  170  204  237  249  258
1994  253-9999  214  196  171  140  130  141  171  222  248  247
1995  245  249  234  205  186  155  148  151  198  217  245  244
1996  245  238  220  208  159  166  136  161  179  225  233  247
1997  245  243  217  195  186  149  138  156  195  230  242  247
1998  248  250  245  229  188  167  177  158  200  247  253  250
1999  250  245  242  216  144  150  123-9999  188  239  240  251
2000  245  243  243  232  184  143  138  154  194  231  249  249
```

Now, I admit the lats and lons aren't spot on. But c'mon, what are the chances of them being different? The two year 2000s are almost identical. What about:

```
      0 -1550  12450   12 KURI BAY                AUSTRALIA 2000 2006   -999  -999.00
9420800 -1548  12452   29 KURI BAY             AUSTRALIA     1965 1992   -999  -999.00
```

Or:

```
      0 -1550  12810   11 WYNDHAM                 AUSTRALIA 2000 2006   -999  -999.00
      0 -1550  12820    4 WYNDHAM AERO            AUSTRALIA 2000 2006   -999  -999.00
9421400 -1549  12812   11 WYNDHAM POST OFFICE  AUSTRALIA     1968 2000   -999  -999.00
9421401 -1547  12810   20 WYNDHAM (WYNDHAM POR AUSTRALIA     1898 1966   -999  -999.00
```

Come On!! This is one station isn't it.

I'd be content to leave it - but I have to match the bulletins! And I can match to the long, stable series or to the loose, flapping ones put in for the purpose! Meh II.

So.. in the end I matched to the 2000-2006 stations, where they actually did match. Unfortunately the huge bulk of the bulletins still had to have new entries created for them, which is a shame, and begs the question of why the Australian update bulletins don't match the original 'catch-up' block they sent us.

For some reason, the auminmatch program is causing no end of grief. I thought I'd
managed a complete run, and it did produce a good-looking tmin database with lots of
new station stubs tacked on the end:

```
  -1009 -6628  11054   12 KURI BAY                AUSTRALIA 2007 2007    -999    1009
6190-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
2006-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
  -1019 -6628  11054   23 KALUMBURU               AUSTRALIA 2007 2007    -999    1019
6190-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
2006-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
  -1020 -6628  11054   51 TRUSCOTT                AUSTRALIA 2007 2007    -999    1020
6190-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
2006-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
```

However, it doesn't seem to have put the bulletin codes on the (a8) header field, for
some of the matches only!

Not sure why this is yet.. but have found also that there are cases of duplicated lat/lon pairs,
so multiple matches are being made.. argh.. will have to further augment auminmatch. Not happy.

An interesting aside.. David was looking at the v3.00 precip to help National Geographic with
an enquiry. I produced a second 'station' file with the 'honest' counts (see above) and he used
that to mask out cells with a 0 count (ie that only had indirect data from 'nearby' stations).
There were some odd results.. with certain months havign data, and others being missing. After
considerable debate and investigation, it was understood that anomdtb calculates normals on a
monthly basis. So, where there are 7 or 8 missing values in each month (1961-1990), a station
may end up contributing only in certain months of the year, throughout its entire run! This was
noticed in the Seychelles, where only October has real data (the remaining months being relaxed
to the climatology but excluded by David using the 'tight' station mask). There is no easy
solution, because essentially it's an honest result: only October has sufficient values to form
a normal, so only October gets anomalised. It's an unfortunate concidence that it's the only
station in the cell, but it's not the only one. A 'solution' could be for anomdtb to get a bit
more involved in the gridding, to check that if a cell only has one station (for one or more
years) then it's all-or-nothing. Maybe if only one month has a normal then it's dumped and the
whole reverts to climatology. Maybe if 4 or more months have normals.. maybe if >0 months have
normals and the rest can be brought in with a minor relaxation of the '75% rule'.. who knows.

Back to auminmatch.for, and a (philosophical) breakthrough. Built a loop to find 'fuzzy'
matches and group them together. The user then processes one group at a time, pairing up
matches until the potential for further matches is zero (or the user decides it is). Uses a
FSM to work out each chain (all db matches for a bulletin, then all bulletins that match
each of those db stations, then.. etc). To understand it, either read the code (especially
the comments) or just look at this mind-boggling example from the first run of it:

```
-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
User Match Decision(s) Please!
Bulletin stations:    8
   1.    9021 -3193  11598   15 PERTH AIRPORT
   2.    9225 -3192  11587   25 PERTH METRO
   3.    9106 -3205  11598   10 GOSNELLS CITY
   4.    9240 -3201  11614  384 BICKLEY
   5.    9172 -3210  11588   30 JANDAKOT AERO
   6.    9215 -3196  11576   41 SWANBOURNE
   7.    9194 -3222  11581   14 MEDINA RESEARCH CENT
   8.    9256 -3224  11568    6 GARDEN ISLAND HSF
Database stations:   18
   1.       0 -3190  11590   25 PERTH METRO          2000 2006
   2.       0 -3190  11600   15 PERTH AIRPORT        2000 2006
   3. 9461000 -3190  11600   20 PERTH AIRPORT COMPAR 1944 2006
   4. 9461001 -3190  11600   18 PERTH AIRPORT        1944 2004
   5. 9461501 -3198  11607  210 KALAMUNDA (KALAMUNDA 1908 1992
   6.       0 -3200  11580   41 SWANBOURNE          2000 2006
   7.       0 -3200  11580   20 SUBIACO TREATMENT PL 2000 2006
   8.       0 -3196  11579   20 SUBIACO TREATMENT PL 1991 1999
   9. 9460800 -3195  11587   19 PERTH (PERTH REGIONA 1897 1992
  10. 9460801 -3195  11587   19 PERTH-(PERTH-REGIONA 1897 1992
  11.       0 -3210  11590   30 JANDAKOT AERO        2000 2006
  12.       0 -3210  11600   10 GOSNELLS CITY        2000 2006
  13.       0 -3200  11610  384 BICKLEY             2000 2006
  14.       0 -3220  11580   14 MEDINA RESEARCH CENT 2000 2006
  15.       0 -3220  11580    4 KWINANA BP REFINERY  2000 2006
  16.       0 -3223  11576    4 KWINANA BP REFINERY  1961 2000
  17. 9560800 -3222  11581   14 MEDINA RESEARCH CENT 1991 2000
  18.       0 -3220  11570    6 GARDEN ISLAND HSF    2000 2006

Enter a matching pair, (bulletin,database) or 'n' to end: 1,2
Bulletin stations:    7
   2.    9225 -3192  11587   25 PERTH METRO
   3.    9106 -3205  11598   10 GOSNELLS CITY
   4.    9240 -3201  11614  384 BICKLEY
   5.    9172 -3210  11588   30 JANDAKOT AERO
   6.    9215 -3196  11576   41 SWANBOURNE
   7.    9194 -3222  11581   14 MEDINA RESEARCH CENT
   8.    9256 -3224  11568    6 GARDEN ISLAND HSF
Database stations:   17
   1.       0 -3190  11590   25 PERTH METRO          2000 2006
   3. 9461000 -3190  11600   20 PERTH AIRPORT COMPAR 1944 2006
   4. 9461001 -3190  11600   18 PERTH AIRPORT        1944 2004
   5. 9461501 -3198  11607  210 KALAMUNDA (KALAMUNDA 1908 1992
   6.       0 -3200  11580   41 SWANBOURNE          2000 2006
   7.       0 -3200  11580   20 SUBIACO TREATMENT PL 2000 2006
   8.       0 -3196  11579   20 SUBIACO TREATMENT PL 1991 1999
   9. 9460800 -3195  11587   19 PERTH (PERTH REGIONA 1897 1992
  10. 9460801 -3195  11587   19 PERTH-(PERTH-REGIONA 1897 1992
  11.       0 -3210  11590   30 JANDAKOT AERO        2000 2006
  12.       0 -3210  11600   10 GOSNELLS CITY        2000 2006
  13.       0 -3200  11610  384 BICKLEY             2000 2006
  14.       0 -3220  11580   14 MEDINA RESEARCH CENT 2000 2006
  15.       0 -3220  11580    4 KWINANA BP REFINERY  2000 2006
  16.       0 -3223  11576    4 KWINANA BP REFINERY  1961 2000
  17. 9560800 -3222  11581   14 MEDINA RESEARCH CENT 1991 2000
  18.       0 -3220  11570    6 GARDEN ISLAND HSF    2000 2006

Enter a matching pair, (bulletin,database) or 'n' to end: 2,1
Bulletin stations:    6
   3.    9106 -3205  11598   10 GOSNELLS CITY
   4.    9240 -3201  11614  384 BICKLEY
   5.    9172 -3210  11588   30 JANDAKOT AERO
   6.    9215 -3196  11576   41 SWANBOURNE
   7.    9194 -3222  11581   14 MEDINA RESEARCH CENT
   8.    9256 -3224  11568    6 GARDEN ISLAND HSF
Database stations:   16
   3. 9461000 -3190  11600   20 PERTH AIRPORT COMPAR 1944 2006
   4. 9461001 -3190  11600   18 PERTH AIRPORT        1944 2004
   5. 9461501 -3198  11607  210 KALAMUNDA (KALAMUNDA 1908 1992
   6.       0 -3200  11580   41 SWANBOURNE          2000 2006
   7.       0 -3200  11580   20 SUBIACO TREATMENT PL 2000 2006
   8.       0 -3196  11579   20 SUBIACO TREATMENT PL 1991 1999
   9. 9460800 -3195  11587   19 PERTH (PERTH REGIONA 1897 1992
```

```
  10. 9460801 -3195  11587   19 PERTH-(PERTH-REGIONA 1897 1992
  11.       0 -3210  11590   30 JANDAKOT AERO        2000 2006
  12.       0 -3210  11600   10 GOSNELLS CITY        2000 2006
  13.       0 -3200  11610  384 BICKLEY              2000 2006
  14.       0 -3220  11580   14 MEDINA RESEARCH CENT 2000 2006
  15.       0 -3220  11580    4 KWINANA BP REFINERY  2000 2006
  16.       0 -3223  11576    4 KWINANA BP REFINERY  1961 2000
  17. 9560800 -3222  11581   14 MEDINA RESEARCH CENT 1991 2000
  18.       0 -3220  11570    6 GARDEN ISLAND HSF    2000 2006

Enter a matching pair, (bulletin,database) or 'n' to end: 3,12
Bulletin stations:    5
   4.    9240 -3201  11614  384 BICKLEY
   5.    9172 -3210  11588   30 JANDAKOT AERO
   6.    9215 -3196  11576   41 SWANBOURNE
   7.    9194 -3222  11581   14 MEDINA RESEARCH CENT
   8.    9256 -3224  11568    6 GARDEN ISLAND HSF
Database stations:   15
   3. 9461000 -3190  11600   20 PERTH AIRPORT COMPAR 1944 2006
   4. 9461001 -3190  11600   18 PERTH AIRPORT        1944 2004
   5. 9461501 -3198  11607  210 KALAMUNDA (KALAMUNDA 1908 1992
   6.       0 -3200  11580   41 SWANBOURNE           2000 2006
   7.       0 -3200  11580   20 SUBIACO TREATMENT PL 2000 2006
   8.       0 -3196  11579   20 SUBIACO TREATMENT PL 1991 1999
   9. 9460800 -3195  11587   19 PERTH (PERTH REGIONA 1897 1992
  10. 9460801 -3195  11587   19 PERTH-(PERTH-REGIONA 1897 1992
  11.       0 -3210  11590   30 JANDAKOT AERO        2000 2006
  13.       0 -3200  11610  384 BICKLEY              2000 2006
  14.       0 -3220  11580   14 MEDINA RESEARCH CENT 2000 2006
  15.       0 -3220  11580    4 KWINANA BP REFINERY  2000 2006
  16.       0 -3223  11576    4 KWINANA BP REFINERY  1961 2000
  17. 9560800 -3222  11581   14 MEDINA RESEARCH CENT 1991 2000
  18.       0 -3220  11570    6 GARDEN ISLAND HSF    2000 2006

Enter a matching pair, (bulletin,database) or 'n' to end: 4,13
Bulletin stations:    4
   5.    9172 -3210  11588   30 JANDAKOT AERO
   6.    9215 -3196  11576   41 SWANBOURNE
   7.    9194 -3222  11581   14 MEDINA RESEARCH CENT
   8.    9256 -3224  11568    6 GARDEN ISLAND HSF
Database stations:   14
   3. 9461000 -3190  11600   20 PERTH AIRPORT COMPAR 1944 2006
   4. 9461001 -3190  11600   18 PERTH AIRPORT        1944 2004
   5. 9461501 -3198  11607  210 KALAMUNDA (KALAMUNDA 1908 1992
   6.       0 -3200  11580   41 SWANBOURNE           2000 2006
   7.       0 -3200  11580   20 SUBIACO TREATMENT PL 2000 2006
   8.       0 -3196  11579   20 SUBIACO TREATMENT PL 1991 1999
   9. 9460800 -3195  11587   19 PERTH (PERTH REGIONA 1897 1992
  10. 9460801 -3195  11587   19 PERTH-(PERTH-REGIONA 1897 1992
  11.       0 -3210  11590   30 JANDAKOT AERO        2000 2006
  14.       0 -3220  11580   14 MEDINA RESEARCH CENT 2000 2006
  15.       0 -3220  11580    4 KWINANA BP REFINERY  2000 2006
  16.       0 -3223  11576    4 KWINANA BP REFINERY  1961 2000
  17. 9560800 -3222  11581   14 MEDINA RESEARCH CENT 1991 2000
  18.       0 -3220  11570    6 GARDEN ISLAND HSF    2000 2006

Enter a matching pair, (bulletin,database) or 'n' to end: 5,11
Bulletin stations:    3
   6.    9215 -3196  11576   41 SWANBOURNE
   7.    9194 -3222  11581   14 MEDINA RESEARCH CENT
   8.    9256 -3224  11568    6 GARDEN ISLAND HSF
Database stations:   13
   3. 9461000 -3190  11600   20 PERTH AIRPORT COMPAR 1944 2006
   4. 9461001 -3190  11600   18 PERTH AIRPORT        1944 2004
   5. 9461501 -3198  11607  210 KALAMUNDA (KALAMUNDA 1908 1992
   6.       0 -3200  11580   41 SWANBOURNE           2000 2006
   7.       0 -3200  11580   20 SUBIACO TREATMENT PL 2000 2006
   8.       0 -3196  11579   20 SUBIACO TREATMENT PL 1991 1999
   9. 9460800 -3195  11587   19 PERTH (PERTH REGIONA 1897 1992
  10. 9460801 -3195  11587   19 PERTH-(PERTH-REGIONA 1897 1992
  14.       0 -3220  11580   14 MEDINA RESEARCH CENT 2000 2006
  15.       0 -3220  11580    4 KWINANA BP REFINERY  2000 2006
  16.       0 -3223  11576    4 KWINANA BP REFINERY  1961 2000
  17. 9560800 -3222  11581   14 MEDINA RESEARCH CENT 1991 2000
  18.       0 -3220  11570    6 GARDEN ISLAND HSF    2000 2006

Enter a matching pair, (bulletin,database) or 'n' to end: 6,6
Bulletin stations:    2
   7.    9194 -3222  11581   14 MEDINA RESEARCH CENT
   8.    9256 -3224  11568    6 GARDEN ISLAND HSF
Database stations:   12
   3. 9461000 -3190  11600   20 PERTH AIRPORT COMPAR 1944 2006
   4. 9461001 -3190  11600   18 PERTH AIRPORT        1944 2004
   5. 9461501 -3198  11607  210 KALAMUNDA (KALAMUNDA 1908 1992
   7.       0 -3200  11580   20 SUBIACO TREATMENT PL 2000 2006
   8.       0 -3196  11579   20 SUBIACO TREATMENT PL 1991 1999
   9. 9460800 -3195  11587   19 PERTH (PERTH REGIONA 1897 1992
  10. 9460801 -3195  11587   19 PERTH-(PERTH-REGIONA 1897 1992
  14.       0 -3220  11580   14 MEDINA RESEARCH CENT 2000 2006
  15.       0 -3220  11580    4 KWINANA BP REFINERY  2000 2006
  16.       0 -3223  11576    4 KWINANA BP REFINERY  1961 2000
  17. 9560800 -3222  11581   14 MEDINA RESEARCH CENT 1991 2000
  18.       0 -3220  11570    6 GARDEN ISLAND HSF    2000 2006

Enter a matching pair, (bulletin,database) or 'n' to end: 7,14
Bulletin stations:    1
   8.    9256 -3224  11568    6 GARDEN ISLAND HSF
Database stations:   11
   3. 9461000 -3190  11600   20 PERTH AIRPORT COMPAR 1944 2006
   4. 9461001 -3190  11600   18 PERTH AIRPORT        1944 2004
   5. 9461501 -3198  11607  210 KALAMUNDA (KALAMUNDA 1908 1992
   7.       0 -3200  11580   20 SUBIACO TREATMENT PL 2000 2006
   8.       0 -3196  11579   20 SUBIACO TREATMENT PL 1991 1999
   9. 9460800 -3195  11587   19 PERTH (PERTH REGIONA 1897 1992
  10. 9460801 -3195  11587   19 PERTH-(PERTH-REGIONA 1897 1992
  15.       0 -3220  11580    4 KWINANA BP REFINERY  2000 2006
  16.       0 -3223  11576    4 KWINANA BP REFINERY  1961 2000
  17. 9560800 -3222  11581   14 MEDINA RESEARCH CENT 1991 2000
  18.       0 -3220  11570    6 GARDEN ISLAND HSF    2000 2006

Enter a matching pair, (bulletin,database) or 'n' to end: 8,18
-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-

Amazing, huh? Most are actually more like this:

-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
User Match Decision(s) Please!
Bulletin stations:    1
   1.    9053 -3167  11602   40 PEARCE RAAF
Database stations:    2
```

```
   1.       0 -3170  11600   40 PEARCE RAAF          2000 2006
   2. 9461200 -3167  11602   49 BULLSBROOK (PEARCE A 1940 1992

Enter a matching pair, (bulletin,database) or 'n' to end: 1,1
-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-

However.. still teething troubles, with previously-paired stations reappearing
for a second chance sometimes! So more debugging.. fixed. Also added a test
before the user gets a chain, to anticipate what the user (er, I) would do. For
instance, I generally match to a 200-2006, WMO=0 database station if the names
match, as they're the ones David L put in from the Aus update files. I, er, the
user then gets ambiguities and nearby but unconnected stations. Fine, until you
get a nasty surprise like this one:

User Match Decision(s) Please!
Bulletin stations:    2
   1.   58009 -2864  15364   95 BYRON BAY (CAPE BYRO
   2.   58216 -2864  15364   95 BYRON BAY (CAPE BYRO
Database stations:    3
   1.       0 -2860  15360   95 BYRON BAY (CAPE BYRO 2000 2006
   2.       0 -2860  15360   95 BYRON BAY (CAPE BYRO 2000 2006
   3. 9459500 -2863  15363   98 CAPE BYRON           1974 1992

Looking in the files I see that Bulletin 58009 is 'BYRON BAY (CAPE BYRON LIGHTHOUSE)',
and 58216 is 'BYRON BAY (CAPE BYRON AWS)'. But the database stubs that have been
entered have not been intelligently named, just truncated - so I have no way of
knowing which is which! CRU NEEDS A DATA MANAGER. In this case I had to assume that
the updates were processed in .au code order, so 1-1 and 2-2. Argh. A few doubles
found, too:

Bulletin stations:    1
   1.   33106 -2037  14895   59 HAMILTON ISLAND AIRP
Database stations:    3
   1.       0 -2040  14900   23 HAMILTON ISLAND AIRP 2000 2006
   2.       0 -2040  14900   59 HAMILTON ISLAND AIRP 2000 2006
   3. 9436800 -2035  14895   23 HAMILTON ISLAND AIRP 1991 2000

Bulletin stations:    1
   1.   90186 -3829  14245   71 WARRNAMBOOL AIRPORT
Database stations:    4
   1.       0 -3830  14250   71 WARRNAMBOOL AIRPORT  2000 2006
   2.       0 -3830  14240   75 WARRNAMBOOL AIRPORT  2000 2006
   3.       0 -3840  14248   21 WARRNAMBOOL (POST OF 1961 1980
   4.       0 -3828  14243   76 WARRNAMBOOL A        1983 1999

And the results? Strictly average, I thought.. but I'd forgotten to count the extra
'anticipated match' routine achievements! So I grepped the match-by-match file,
matches.0706281447.dat, and got:

crua6[/cru/cruts/version_3_0/db/dtr] grep 'AUTO\:' matches.0706281447.dat |wc -l
       232
crua6[/cru/cruts/version_3_0/db/dtr] grep 'AUTO FROM CHAIN' matches.0706281447.dat | wc -l
       514
crua6[/cru/cruts/version_3_0/db/dtr] grep 'MANUAL' matches.0706281447.dat | wc -l
        12

In other words, all that sweat was worth it - 746 stations matched automatically, and
a further 12 manually! Only (797-758=) 39 bulletins unmatched! Wheeee! And here they are:

  -6072 -2303  11504  111 EMU CREEK STATION        AUSTRALIA 2007 2007    -999    6072
 -12044 -3355  12070  220 MUNGLINUP WEST          AUSTRALIA 2007 2007    -999   12044
 -12241 -2888  12132  370 LEONORA AERO            AUSTRALIA 2007 2007    -999   12241
 -17031 -2965  13806   50 MARREE COMPARISON       AUSTRALIA 2007 2007    -999   17031
 -21118 -3323  13800   10 PORT PIRIE AERODROME    AUSTRALIA 2007 2007    -999   21118
 -22801 -3575  13659  143 CAPE BORDA COMPARISO    AUSTRALIA 2007 2007    -999   22801
 -23122 -3451  13868   65 ROSEWORTHY AWS          AUSTRALIA 2007 2007    -999   23122
 -24521 -3512  13926   33 MURRAY BRIDGE COMPAR    AUSTRALIA 2007 2007    -999   24521
 -25509 -3533  14052   99 LAMEROO COMPARISON      AUSTRALIA 2007 2007    -999   25509
 -26026 -3716  13976    3 ROBE COMPARISON         AUSTRALIA 2007 2007    -999   26026
 -32004 -1826  14602    5 CARDWELL MARINE PDE     AUSTRALIA 2007 2007    -999   32004
 -35019 -2282  14764  260 CLERMONT SIRIUS ST      AUSTRALIA 2007 2007    -999   35019
 -48243 -2943  14797  154 LIGHTNING RIDGE VISI    AUSTRALIA 2007 2007    -999   48243
 -55024 -3103  15027  307 GUNNEDAH RESOURCE CE    AUSTRALIA 2007 2007    -999   55024
 -56037 -3053  15167  987 ARMIDALE (TREE GROUP    AUSTRALIA 2007 2007    -999   56037
 -60013 -3218  15251    4 FORSTER - TUNCURRY R    AUSTRALIA 2007 2007    -999   60013
 -63039 -3371  15031 1015 KATOOMBA (MURRI ST)     AUSTRALIA 2007 2007    -999   63039
 -63226 -3348  15013  900 LITHGOW (COOERWULL)     AUSTRALIA 2007 2007    -999   63226
 -68257 -3406  15077  112 CAMPBELLTOWN (MOUNT     AUSTRALIA 2007 2007    -999   68257
 -70263 -3475  14970  670 GOULBURN TAFE           AUSTRALIA 2007 2007    -999   70263
 -82170 -3655  14600  171 BENALLA AIRPORT         AUSTRALIA 2007 2007    -999   82170
 -84150 -3787  14801    4 LAKES ENTRANCE (EAST    AUSTRALIA 2007 2007    -999   84150
 -85099 -3863  14581    3 POUND CREEK             AUSTRALIA 2007 2007    -999   85099
 -88023 -3723  14591  230 LAKE EILDON             AUSTRALIA 2007 2007    -999   88023
-200001 -2166  15027  209 MIDDLE PERCY ISLAND     AUSTRALIA 2007 2007    -999  200001
-200100 -2066  11558   24 VARANUS ISLAND          AUSTRALIA 2007 2007    -999  200100
-200212 -1061  12598 -999 NORTHERN ENDEAVOUR      AUSTRALIA 2007 2007    -999  200212
-200283 -1629  14997    8 WILLIS ISLAND           AUSTRALIA 2007 2007    -999  200283
-200288 -2904  16794  112 NORFOLK ISLAND AERO     AUSTRALIA 2007 2007    -999  200288
-200731 -1176  13003    7 POINT FAWCETT           AUSTRALIA 2007 2007    -999  200731
-200783 -1772  14845    3 FLINDERS REEF           AUSTRALIA 2007 2007    -999  200783
-200790 -1045  10569  261 CHRISTMAS ISLAND AER    AUSTRALIA 2007 2007    -999  200790
-200824 -1753  21040    2 PAPEETE                 AUSTRALIA 2007 2007    -999  200824
-200838 -3922  14698  116 HOGAN ISLAND           AUSTRALIA 2007 2007    -999  200838
-200851   -52  16692    7 NAURU ARCS-2            AUSTRALIA 2007 2007    -999  200851
-200852  -206  14743    4 MANUS ARCS-1            AUSTRALIA 2007 2007    -999  200852
-300000 -6858   7797   18 DAVIS                   AUSTRALIA 2007 2007    -999  300000
-300001 -6760   6287   10 MAWSON                  AUSTRALIA 2007 2007    -999  300001
-300017 -6628  11054   40 CASEY                   AUSTRALIA 2007 2007    -999  300017

Resultant database: tmn.0707021605.dtb

[edit: found another fault, had to re-run. Headers weren't being modded if the WMO code was
 already there]

32. The next stage *heart falls* will be to synchronise tmax *against* tmin, sweeping
up duplicates in the process. How long's THIS gonna take? Well actually, it might be fairly easy,
if we use a similar approach. We can base it all around the user being given a 'cloud' of
related stations to pick pairs from, only they will be uniquely numbered so that two from the
same database can be selected. The user can in this way 'pair up' stations in groups.

Of course, this comes with the downside of complexity (and therefore bugs). And both databases
will almost certainly have to be preloaded in their entirety because of the need for the user to
be able to confirm header and data precedence info when stations within a database are merged.

Oh - and I'll have to move bloody quick. So more bugs.

Well.. it's written, and debugging. Around 1500 lines of code, or 1000 without all the comments ;-)
It does indeed read in all the data, so has to be compiled on uealogin1 (as crua6 doesn't have
enough memory!). Reusing code from auminmatch.for did speed things up a bit, though two new
subroutines had to be written to carry out checking for merges (within a database) and for
```

matches (between the databases). Also introduced a user decision at the start to allow the TMin
database to take precedence in terms of station metadata. Here's the current state of play:

```
<BEGIN QUOTE>
uealogin1[/cru/cruts/version_3_0/db/dtr] ./auminmaxsync

WELCOME TO THE TMIN/TMAX SYNCHRONISER

Before we get started, an important question: Should TMin header info take precedence over TMax?

This will significantly reduce user decisions later, but is a big step as TMax settings may be silently overridden!

To let TMin header values take precedence over those of TMax, enter 'YES': YES
Please enter the tmin database name: tmn.0707021605.dtb
Please enter the tmax database name: tmx.0702091313.dtb

Reading in both databases..
TMin database stations:    14349
TMax database stations:    14315

Processing one-to-one matches..

Initial scan found:
  one-to-one matches:  7875
  of which confirmed:  7691
  in a station cloud:  6411 (tmin)
  in a station cloud:  6392 (tmax)
         unmatchable:    63 (tmin)
         unmatchable:    48 (tmax)
Processing match clouds..
-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
User Match Decision(s) Please!
TMin stations:    2
   1. -401000  3178   3522  783 JERUSALEM            1863 2000    -999       0
   2. 4018400  3178   3522  809 JERUSALEM            1977 1995    -999       0
TMax stations:    2
   3. -401000  3178   3522  783 JERUSALEM            1863 2000    -999       0
   4. 4018400  3178   3522  809 JERUSALEM            1977 1995    -999       0

*** Remember: Merge first, Match second! ***
Enter ANY pair to match or merge, or 'n' to end:
<END QUOTE>
```

So stats pretty much as expected/hoped. The one-to-one matches should, of course, be 100%.. but as
the databases aren't synchronised, and as there are hundreds of 'duplicate' entries.. only around
50% match straight away. The situation isn't as bleak as it looks, though - there is further
automatching at the beginning of each cloud, so the user can still be spared the obvious. If the
merging gets too onerous, though, I might have to automate that - with associated risks.

And of course - if you look closely - things are still a little offbeam :-/

Found another database bug by chance.. a <tab> instead of a space after 'CRANWELL':

```
-324320  5303    -50   62 CRANWELL                 UK            1961 1995   -999  -999.00
```

Doesn't show up in reads as it's a white space character. Argh. Fixed in tmin & tmax. Now to find
out why some matched stations STILL don't have the backref in the last header field!! ..found it,
not my problem, it's the ones that *pre-existed* in the databases, there's 84 in total I think. So
I can write a proglet to check that any with negative WMO codes have the positive version in that
last field. And I did - 'fixtnxrefs.for'. Fixed:
tmn.0702091139.dtb (84 fixed)
tmn.0707021605.dtb (651 'fixed' - includes all with negative WMOs regardless of end field)
tmx.0702091313.dtb (84 fixed)

So why, when we matched 758 bulletins in the first place, did this program only 'fix' 651, of which
84 were preexisting? Because, of course, the matches only get a negative WMO code if the original
WMO code is missing (zero). The 'missing' stations would be ones that already had a WMO code.

So, try again, and it's looking good!

```
<BEGIN QUOTE>
uealogin1[/cru/cruts/version_3_0/db/dtr] ./auminmaxsync

WELCOME TO THE TMIN/TMAX SYNCHRONISER

Before we get started, an important question: Should TMin header info take precedence over TMax?

This will significantly reduce user decisions later, but is a big step as TMax settings may be silently overridden!

To let TMin header values take precedence over those of TMax, enter 'YES': YES
Please enter the tmin database name: tmn.0702091139.dtb
Please enter the tmax database name: tmx.0702091313.dtb

Reading in both databases..
TMin database stations:    14309
TMax database stations:    14315

Processing one-to-one matches..

Initial scan found:
  one-to-one matches:  7889
  of which confirmed:  7702
  in a station cloud:  6365 (tmin)
  in a station cloud:  6378 (tmax)
         unmatchable:    55 (tmin)
         unmatchable:    48 (tmax)
Processing match clouds..
-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
User Match Decision(s) Please!
TMin stations:    2
   1. -401000  3178   3522  783 JERUSALEM            1863 2000    -999  401000
   2. 4018400  3178   3522  809 JERUSALEM            1977 1995    -999       0
TMax stations:    2
   3. -401000  3178   3522  783 JERUSALEM            1863 2000    -999  401000
   4. 4018400  3178   3522  809 JERUSALEM            1977 1995    -999       0

*** Remember: Merge first, Match second! ***
Enter ANY pair to match or merge, or 'n' to end: 1,2
Merging two stations from the TMin database:
Stn 1:      -401000  3178   3522  783 JERUSALEM            ISRAEL        1863 2000   -999   401000
Stn 2:      -401000  3178   3522  783 JERUSALEM            ISRAEL        1863 2000   -999   401000
Please resolve the following inconsistencies:
 Overlap:  Station A) -401000  3178   3522  783 JERUSALEM            ISRAEL        1863 2000   -999   401000
           Station B) 4018400  3178   3522  809 JERUSALEM            ISRAEL        1977 1995   -999  -999.00

You must decide which station's data takes precedence.
The intercorrelation for the period is:  0.99
Enter A or B, or undo pair(X):

<END QUOTE>
```

Well.. it's kinda working. I found some idiotic bugs, though it is a fearsomely complicated program with
lots of indirect pointers (though I do try and resolve them at the first opportunity). One thing that's
making debugging frustratingly difficult is something that must be a uealogin1 feature, and I haven't seen
it before: the program doesn't actually flush the output channels whenever you write! For example, as I
write this the program has dispensed with auto-matching:

```
Initial scan found:
  one-to-one matches:  7875
  of which confirmed:  7691
  in a station cloud:  6411 (tmin)
  in a station cloud:  6392 (tmax)
         unmatchable:    63 (tmin)
         unmatchable:    48 (tmax)
```

(yes, it's a little tighter now)

Anyway, since then I've merged two pair (JERUSALEM) then paired the remainder. That activity has generated
match reports on channel 31 BUT THEY ARE NOT IN THE FILE YET. Here is the tail of channel 31:

```
crua6[/cru/cruts/version_3_0/db/dtr] tail mat.0707121500.dat
TMax: 9929470  4330   1340  342 MACERATA             ITALY         1953 1975   -999  -999.00
AUTO PAIRING FROM ONE-TO-ONE SCAN:
TMin: 9929480  4030    880  585 MACOMER              ITALY         1952 1978   -999  -999.00
TMax: 9929480  4030    880  585 MACOMER              ITALY         1952 1978   -999  -999.00
AUTO PAIRING FROM ONE-TO-ONE SCAN:
TMin: 9929500  4010   1850   86 PALASCIA AERO        ITALY         1952 1978   -999  -999.00
TMax: 9929500  4010   1850   86 PALASCIA AERO        ITALY         1952 1978   -999  -999.00
AUTO PAIRING FROM ONE-TO-ONE SCAN:
TMin: 9929520  4060   1490   30 PONTECAGNANO         ITALY         1951 1978   -999  -999.00
TMax: 9929520  4060   1490   30 PONTECAGNANO         ITALY         1951 1978   -999  -999.00
```

In addition, the log file is EMPTY, yet at least 416 bytes have been written to it. How the hell can I
debug if I can't monitor what's being written to the log files?!! Of course, once I force-quit the program,
and wait a bit.. the missing info appears. Similarly if I carry on using the program, the files get more
info. It's as if there's a write buffer that runs FIFO. Must look at the 'help'.. why is it that whenever I
crack the programming, the systems themselves step in the screw it up? And computer support is away of course.

Looked at f77 -help.. nothing. well nothing obvious. Anyway, more debugging and..

Seems to be working. But it's going to take ages. Here is an example of the problem:

<BEGIN QUOTE>

```
-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
User Match Decision(s) Please!
TMin stations:    2
   1. -315770  5638   -287   10 LEUCHARS             UK            1959 1995    -999  315770
   2.  317100  5640   -287   12 LEUCHARS             UNITED KINGDO 1997 2006    -999       0
TMax stations:    2
   3. -315770  5638   -287   10 LEUCHARS             UK            1959 1995    -999  315770
   4.  317100  5638   -287   12 LEUCHARS RAF         UK UK         1973 2006    -999       0

*** Remember: Merge first, Match second! ***
Enter ANY pair to match or merge, or 'n' to end:
```

<END QUOTE>

Not only do both databases have unnecessary duplicates, introduced for external mapping purposes
by the look of it, but the 'main' stations (2 and 4) have different station name & country. In fact
one of the country names is illegal! Dealing with things like this cannot be automated as they're
the results of non-automatic decisions.

Something new - a listing of 147 Australian 'bulletin' stations, most of which have mappings to
WMO codes. Decided to xref against the (mapped) TMin database, for a laugh. Then decided to take it
more seriously. Wrote a prog to IMPOSE the mappings onto tmn.0707021605.dtb, overriding existing
mappings as necessary. What a bloody mess.

Decided to be vaguely sensible and let the program, auwmoxref.for, evolve. so to begin with it just
did a scan between the mappings file (au_mapping_to_wmo.dat) and the tmin database with my mappings
in (tmn.0707021605.dtb). Results:

```
crua6[/cru/cruts/version_3_0/db/dtr] ./auwmoxref
```

<BEGIN QUOTE>

```
AUWMOXREF: Check Australian cross-references

 Enter the file of WMO mappings: au_mapping_to_wmo.dat
 115 mappings read

 Enter the mapped TMin database: tmn.0707021605.dtb
 14349 database headers read


RESULTS:

WMO Matches:    92
 (multiples)  (  0)
> Ref matches:  60
> Ref empty:    31
> Ref WRONG:     1

Ref Matches:   114
 (multiples)  (  0)
> WMO matches:  60
> WMO -1*Ref:   41
> WMO WRONG:    13
```

<END QUOTE>

So first the good news - no duplicates. Well there shouldn't have been any anyway of course, but the
way things are going I'm taking nothing for granted. See, I count something turning out as expected
as 'good news'. So anyway.. I also extracted the statistic that 26 mappings matched both Ref and WMO,
but to separate database entries. Thus the 115 mappings are allocated as follows:

```
60  Mapping found to be correctly implemented (over half, excellent)
41  WMO Missing, of which:
    26  WMO found elsewhere (one of which has an unmapped ref attached to it)
    15  WMO not in database (can add wmo codes for these)
13  WMO wrong, of which:
     5  Can be merged with real WMO (effectively same station)
     8  WMO not in database
 1  Completely unmatched (96003 -> 949500)
```

For the purposes of actions to take, the 13 'WMO Wrong' refs can simply be unmapped from their incorrect
mappings and be rolled into the 41 'WMO Missing' refs.

So:

```
60  Mapping found to be correctly implemented (over half, excellent)
54  WMO Missing or wrong, of which:
    31  WMO found elsewhere (one of which has an unmapped ref attached to it)
    23  WMO not in database but pairing made (can add wmo codes for these)
```

```
     8  WMO not in database and no pairing (can add new stations for these)
 1  Completely unmatched (96003 -> 949500)

So, actions to take:

1. For the first 60, no action required.
2. For the 13 with incorrectly-assigned WMOs, disengage and roll into the rest below
3. For the 1 WMO with an unmapped ref attached, disengage and roll into the rest below
3. For the 31 with dislocated WMOs, print a list and ref when doing the tmin/tmax syncing
4. For the 23 with WMO-less stations, add the WMO codes..
5. For the 8 with no WMO found and no pairing found, create new stations.

For the disengagements, decided to work directly with an editor rather than craft another program! So
changes made to tmn.0707021605.dtb (after a suitable backup was made of course!).

The following assignments were disengaged (and replaced with -999.00). Where a WMO code follows in
brackets, the ref was reassigned there.

  1. 9460300 -3200  11550   43 ROTTNEST ISLAND          AUSTRALIA 1898 2006    -999    9193 (9460200)
  2. 9464600 -3090  12810  159 FORREST                  AUSTRALIA 1946 2006    -999   11052 (no)
  3. 9432200 -2020  13000  340 RABBIT FLAT              AUSTRALIA 1969 2006    -999   15666 (no)
  4. 9557400 -2640  15300    6 TEWANTIN RSL PARK        AUSTRALIA 1949 2006    -999   40908 (no)
  5. 9451600 -2810  14860  199 ST GEORGE AIRPORT        AUSTRALIA 1938 2006    -999   43109 (9451700)
  6. 9452700 -2950  14990  213 MOREE AERO              AUSTRALIA 1964 2006    -999   53115 (9552700)
  7. 9454100 -2980  15110  582 INVERELL (RAGLAN ST)     AUSTRALIA 1907 2006    -999   56242 (no)
  8. 9478700 -3140  15290    4 PORT MACQUARIE AIRPO     AUSTRALIA 1907 2006    -999   60139 (no)
  9. 9475800 -3210  15090  216 SCONE SCS               AUSTRALIA 2000 2006    -999   61089 (9473800)
 10. 9494000 -3510  15080   85 JERVIS BAY (POINT PE     AUSTRALIA 1907 2006    -999   68151 (no)
 11. 9491600 -3590  14840 1482 CABRAMURRA SMHEA AWS     AUSTRALIA 1962 2006    -999   72161 (no)
 12. 9482700 -3630  14160  133 NHILL                   AUSTRALIA 1897 2006    -999   78031 (9582900)
 13. 9597900 -4300  14710   63 GROVE (COMPARISON)       AUSTRALIA 1961 2006    -999   94069 (no)

The 'mismatched WMO code' station was disengaged from it's reference and given 48027 instead:
  1. 9471100 -3150  14580  218 COBAR AIRPORT AWS        AUSTRALIA 1962 2006    -999   48237 -> 48027

I mailed BOM as we have 94711 = COBAR AWS but they have *94710* for AWS and 94711 for COBAR MO. The
reply was as follows:

<BEGIN QUOTE>
On 18 Jul 2007, at 8:51, Matthew Bastin wrote:

Hi Ian,

I hope this table helps

Name                  BoM No. WMO No.      Opened           Closed
Cobar Comparison        48244   94711   1/11/1997       15/11/2000
Cobar MO                48027   94711   1/01/1962
Cobar Airport AWS       48237   94710  11/06/1993
Cobar PO                48030            1/1/1881       31/12/1965

The blank in the Closed column means that the site is still open
When Cobar Comparison site closed it transferred its WMO number to Cobar MO
A blank in the WMO No. column means that the site never had a WMO number.

I am not sure of the overlap between the assignment of 94711 between 48244 and 48027. I will find
out and get back to you.
<END QUOTE>

Here are our current 'COBAR' headers:

      0 -3150  14580  260 COBAR COMPARISON     AUSTRALIA     2000 2006    -999 -999.00
      0 -3150  14580  260 COBAR MO             AUSTRALIA     2000 2006    -999 -999.00
      0 -3148  14582  265 COBAR               AUSTRALIA     1962 2004    -999 -999.00
      0 -3150  14580  251 COBAR POST OFFICE    AUSTRALIA     1902 1960    -999 -999.00
9471100 -3150  14580  218 COBAR AIRPORT AWS    AUSTRALIA     1962 2006    -999   48027

Now looking at the dates.. something bad has happened, hasn't it. COBAR AIRPORT AWS cannot start
in 1962, it didn't open until 1993! Looking at the data - the COBAR station 1962-2004 seems to be
an exact copy of the COBAR AIRPORT AWS station 1962-2004, except that the latter has more missing
values. Now, COBAR AIRPORT AWS has 15 months of missing value codes beginning Oct 1993.. coincidence?
No. I think that that series should start there. Furthermore, the overlap between COBAR and COBAR MO
(2000-2004) is *almost* identical:

      0 -3148  14582  265 COBAR               AUSTRALIA     1962 2004    -999 -999.00
2000  177  209  183  135   80   51   45   52  105  122  166  186
2001  223  214  159  126   72   61   43   52  105  110  148  181
2002  195  185  168  148   88   58   49   63  101  128  186  192
2003  222  216  161  137   97   71   56   61   92  113  159  208
2004  207  226  175  141   74   69   46   69   90  136  160  186


      0 -3150  14580  260 COBAR MO             AUSTRALIA     2000 2006    -999 -999.00
2000  178  209  184  136   80   52   45   55  105  122  166  186  (7/12)
2001  223  214  159  126   72   61   43   52  105  110  148  181  (12/12)
2002  195  185  168  148   88   58   49   63  101  128  187  192  (11/12)
2003  222  216  161  137   97   71   56   61   92  113  159  208  (12/12)
2004  207  226  175  141   74   69   46   69   90  136  160  186  (12/12)

I therefore propose to extend COBAR MO using COBAR, and to truncate COBAR AIRPORT AWS at 1993.
All BOM codes will be appended for completeness. So the new headers (with lat/lon from BOM too) are:

      0 -3149  14583  260 COBAR COMPARISON     AUSTRALIA     2000 2006    -999   48244 (closed)
9471100 -3149  14583  260 COBAR MO             AUSTRALIA     1962 2006    -999   48027
      0 -3150  14583  251 COBAR POST OFFICE    AUSTRALIA     1902 1960    -999   48030 (closed)
9471000 -3154  14580  218 COBAR AIRPORT AWS    AUSTRALIA     1995 2006    -999   48237

Deleted:
      0 -3148  14582  265 COBAR               AUSTRALIA     1962 2004    -999 -999.00

The remaining 26 dislocated references were reassigned as for the 13 above. Legitimate mappings:

  1.       3003   9420300
  2.       4032   9431200
  3.       5007   9430200
  4.       7176   9431700
  5.       9021   9461000
  6.      14508   9415000
  7.      14932   9413100
  8.      17031   9448000
  9.      22801   9480500
 10.      26026   9481200
 11.      27045   9417000
 12.      32040   9429400
 13.      40842   9457800
 14.      50052   9470700
 15.      55024   9474000
 16.      67105   9575300
 17.      68072   9475000
 18.      71041   9590800
```

```
 19.      86282   9486600
 20.     200283   9429900
 21.     200288   9499600
 22.     200790   9699500
 23.     200839   9499500
 24.     300000   8957100
 25.     300001   8956400
 26.     300017   8961100

WMO codes were added to these uncoded sites as shown:

  1. 9410000 -1430  12670   23 KALUMBURU                AUSTRALIA 2000 2006    -999    1019
  2. 9562500 -3160  11720  217 CUNDERDIN AIRFIELD       AUSTRALIA 2000 2006    -999   10286
  3. 9564000 -3270  11670  275 WANDERING                AUSTRALIA 2000 2006    -999   10917
  4. 9567000 -3380  13820  109 SNOWTOWN (RAYVILLE P     AUSTRALIA 2000 2006    -999   21133
  5. 9481400 -3530  13890   58 STRATHALBYN RACECOUR     AUSTRALIA 2000 2006    -999   24580
  6. 9548200 -2590  13940   47 BIRDSVILLE AIRPORT       AUSTRALIA 2000 2006    -999   38026
  7. 9552900 -2670  15020  305 MILES CONSTANCE STRE     AUSTRALIA 2000 2006    -999   42112
  8. 9549200 -2800  14380  132 THARGOMINDAH AIRPORT     AUSTRALIA 2000 2006    -999   45025
  9. 9578400 -3190  15250    8 TAREE AIRPORT AWS        AUSTRALIA 2000 2006    -999   60141
 10. 9571900 -3220  14860  284 DUBBO AIRPORT AWS        AUSTRALIA 2000 2006    -999   65070
 11. 9586900 -3560  14500   94 DENILIQUIN AIRPORT A     AUSTRALIA 2000 2006    -999   74258
 12. 9495400 -4070  14470   94 CAPE GRIM BAPS           AUSTRALIA 2000 2006    -999   91245
 13. 9596400 -4110  14680    3 LOW HEAD                 AUSTRALIA 2000 2006    -999   91293
 14. 9595900 -4190  14670 1055 LIAWENEE                 AUSTRALIA 2000 2006    -999   96033

The following was corrected (ref had been mistyped as 78013):
  1. 9582900 -3783  14206  200 HAMILTON RESEARCH ST AUSTRALIA     1971 1998    -999   78031

Now the results look like this:

WMO Matches:   106
> Ref matches: 106
> Ref empty:     0
> Ref WRONG:     0
Ref Matches:   106
> WMO matches: 106
> WMO -1*Ref:    0
> WMO WRONG:     0

In other words, there are (115-106=) 9 mappings unfulfilled. The ref hasn't been matched and
WMO code isn't in the database. However, that didn't mean they weren't in the database with a
missing WMO code, did it? The following were found and augmented with both WMO code and ref.

9457000 -2639  15304    6 TEWANTIN RSL PARK    AUSTRALIA     2000 2004    -999   40908
9594000 -3509  15080   85 JERVIS BAY (PT PERP AWS) AUSTRALIA 2000 2006    -999   68151

The following were added as new station stubs:
9532200 -2018  13001  340 RABBIT FLAT          AUSTRALIA     2007 2007    -999   15666
9554100 -2978  15111  582 INVERELL (RAGLAN ST) AUSTRALIA     2007 2007    -999   56242
9478600 -3143  15287    4 PORT MACQUARIE AIRPT AUSTRALIA     2007 2007    -999   60139
9591600 -3594  14838 1482 CABRAMURRA SMHEA AWS AUSTRALIA     2007 2007    -999   72161
9597100 -4298  14708   63 GROVE (COMPARISON)   AUSTRALIA     2007 2007    -999   94069

The following was complicated by the fact that two versions of the station appear to have been
concatenated. This is the station as it already exists in the TMin database:
9464600 -3085  12811  159 FORREST                  AUSTRALIA 1946 2006    -999 -999.00
However, the current 'live' FORREST station (11052) started in 1993, according to bom.au
records. And wouldn't you know it, the data for this station has missing data between 12/92
and 12/99 inclusive. So I reckon it's the old FORREST AERO station (WMO 9464600, .au ID 11004),
with the new Australian bulletin updates tacked on (hence starting in 2000). Especially as the
old station started in 1946 (http://www.bom.gov.au/climate/averages/tables/cw_011004.shtml).
The trouble is that the bom.au mappings all agree that FORREST is now WMO=9564600. So.. do I
split off the 2000-present data to a new station with the new number, or accept that whoever
joined them (Dave?) looked into it and decided it would be OK? The BOM website says they're
800m apart. Decided to be brave and split the data back into two stations, with both codes
attached (in case we ever get replacement data for the closed station, the site says it went to
1995 after all). So there are now two FORREST stations:

9464600 -3085  12811  159 FORREST AERO              AUSTRALIA 1946 1992    -999   11004
9564600 -3085  12811  159 FORREST                  AUSTRALIA 2000 2006    -999   11052

Hope that's right..

The following mapping was added, though the station does not currently feature in the bulletins.
9495900 -4228  14628 -999 BUTLERS GORGE        AUSTRALIA     2007 2007    -999   96003
6190-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
2007-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999

Also ran a risky search&replace to left-justify the 'AUSTRALIA' in its field, provided the
field wasn't touched by an extended station name. Seems to have been 100% successful.

All 115 refs now matched in the TMin database. Confidence in the fidelity of the Australian
station in the database drastically reduced. Likelihood of invalid merging of Australian
stations high. Let's go..

Well OK, made some final 'improvements' to the syncing program. Now, after it forms a cloud, it
should automatically merge stations provided the criteria are met and no others are possibles.
It also records, in a separate 'action' file (act.*), every relevant action performed during
the run, so that if interrupted I should be able to hack in something to enable a 'resume'. It's
been done a bit hastily so no guarantees that enough information's been saved!

Debugging is still a big issue, unfortunately. It's a complicated program to sort out and the
possibilities for indexing errors are many. In fact, for the first time ever, it's just locked up!
That's a first (it was due to getmos not defaulting to months 1 & 12 if the data was all missing).

Another problem solved - spent ages wondering how the start & end years for a particular station
(WARATAH) were being corrupted. Turns out they weren't - I'd written 'getmos' to trim empty years,
but forgot to check the return flag! Duh.

So.. perhaps a debugged run through? I'm quickly realising that the Australian stations are in
such a state that I'm having to constantly refer to the station descriptions on the BOM website,
which are individual PDFs:

http://www.bom.gov.au/climate/cdo/metadata/pdf/metadata088110.pdf

It takes time.. time I don't have! Though I'm pleased to see that the second FSM is helpfully
chipping in to pair things up when possible.

getting seriously fed up with the state of the Australian data. so many new stations have been
introduced, so many false references.. so many changes that aren't documented. Every time a
cloud forms I'm presented with a bewildering selection of similar-sounding sites, some with
references, some with WMO codes, and some with both. And if I look up the station metadata with
one of the local references, chances are the WMO code will be wrong (another station will have
it) and the lat/lon will be wrong too. I've been at it for well over an hour, and I've reached
the 294th station in the tmin database. Out of over 14,000. Now even accepting that it will get
easier (as clouds can only be formed of what's ahead of you), it is still very daunting. I go
on leave for 10 days after tomorrow, and if I leave it running it isn't likely to be there when
I return! As to whether my 'action dump' will work (to save repetition).. who knows?
```

Yay! Two-and-a-half hours into the exercise and I'm in Argentina!

Pfft.. and back to Australia almost immediately :-(.. and then Chile. Getting there.

Unfortunately, after around 160 minutes of uninterrupted decision making, my screen has started
to black out for half a second at a time. More video cable problems - but why now?!! The count is
up to 1007 though.

I am very sorry to report that the rest of the databases seem to be in nearly as poor a state as
Australia was. There are hundreds if not thousands of pairs of dummy stations, one with no WMO
and one with, usually overlapping and with the same station name and very similar coordinates. I
know it could be old and new stations, but why such large overlaps if that's the case? Aarrggghhh!
There truly is no end in sight. Look at this:

```
-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
User Match Decision(s) Please!
TMin stations:    4
   1.       0   153  12492   80 MENADO/DR. SA        INDONESIA     1960 1975    -999       0
   2.       0   153  12492   80 MENADO/ SAM RATULANG INDONESIA     1986 2004    -999       0
   4. 9701400   153  12492   80 MENADO/DR. SAM RATUL INDONESIA     1995 2006    -999       0
   5. 9997418   153  12492   81 SAMRATULANGI MENADO  INDONESIA     1973 1989    -999       0
TMax stations:    4
   6.       0   153  12492   80 MAPANGET/MANADO      INDONESIA     1960 1975    -999       0
   7.       0   153  12492   80 MENADO/ SAM RATULANG ID ID         1957 2004    -999       0
   9. 9701400   153  12492   80 MENADO/DR. SAM RATUL INDONESIA     1995 2006    -999       0
  10. 9997418   153  12492   81 SAMRATULANGI MENADO  INDONESIA     1972 1989    -999       0

*** Remember: Merge first, then Match ***
Enter ANY pair to match or merge, 'a' to auto-match (no merges), or 'x' to end:
```

I honestly have no idea what to do here. and there are countless others of equal bafflingness.

I'll have to go home soon, leaving it running and hoping none of the systems die overnight :-(((

.. it survived, thank $deity. And a long run of duplicate stations, each requiring multiple
decisions concerning spatial info, exact names, and data precedence for overlaps. If for any reason
this has to be re-run, it can certainly be speeded up! Some large clouds, too - this one started
with 59 members from each database:

```
-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
User Match Decision(s) Please!
TMin stations:    7
  11. 7101965  4362  -7940   78 TORONTO ISLAND                   1905 1959    -999       0
  14. 7163427  4363  -7940   77 TORONTO ISLAND A     CANADA      1957 1994    -999       0
  23. 7101987  4380  -7955  194 TORONTO MET RES STN              1965 1988    -999       0
  24. 7163434  4380  -7955  194 TORONTO MET RES STN  CANADA      1965 1988    -999       0
  36.       0  4388  -7944  233 RICHMOND HILL                    1959 2003    -999       0
  39. 7163408  4388  -7945  233 RICHMOND HILL        CANADA      1959 1990    -999       0
  40. 7163409  4387  -7943  218 RICHMOND HILL WPCP               1960 1981    -999       0
TMax stations:    8
  70. 7101965  4362  -7940   78 TORONTO ISLAND                   1905 1959    -999       0
  71. 7126500  4363  -7940   77 TORONTO ISLAND A                 1957 1994    -999       0
  73. 7163427  4363  -7940   77 TORONTO ISLAND A     CANADA      1957 1990    -999       0
  82. 7101987  4380  -7955  194 TORONTO MET RES STN              1965 1988    -999       0
  83. 7163434  4380  -7955  194 TORONTO MET RES STN  CANADA      1965 1988    -999       0
  95.       0  4388  -7944  233 RICHMOND HILL                    1959 2003    -999       0
  98. 7163408  4388  -7945  233 RICHMOND HILL        CANADA      1959 1990    -999       0
  99. 7163409  4387  -7943  218 RICHMOND HILL WPCP               1960 1981    -999       0
```

There were even larger clouds later.

One thing that's unsettling is that many of the assigned WMo codes for Canadian stations do
not return any hits with a web search. Usually the country's met office, or at least the
Weather Underground, show up - but for these stations, nothing at all. Makes me wonder if
these are long-discontinued, or were even invented somewhere other than Canada! Examples:

```
7162040 brockville
7163231 brockville
7163229 brockville
7187742 forestburg
7100165 forestburg
```

Here's a heartwarming example of a cloud which self-paired completely (debug ines included):

```
<BEGIN QUOTE>
DBG: cloud formed with ( 6, 6) members
DBG: automerging done, leaving ( 6, 6)
DBG: pot.auto i,j:    1   1
DBG: i,ncs2m,cs2m(1-5):    1   1           1    8578    8582    8596       0
DBG: paired:    1   1  108 MILE HOUSE ABEL

Attempting to pair stations:
From TMin:            0  5170 -12140  994 108 MILE HOUSE ABEL               1987 2002    -999 -999.00
From TMax:            0  5170 -12140  994 108 MILE HOUSE ABEL               1987 2002   -999  -999.00
DBG: AUTOPAIRED:    1   1
DBG: pot.auto i,j:    2   2
DBG: i,ncs2m,cs2m(1-5):    2   1           2    8578    8582    8596       0
DBG: paired:    2   2  100 MILE HOUSE

Attempting to pair stations:
From TMin:      7194273  5165 -12130 1059 100 MILE HOUSE       CANADA        1970 1999    -999 -999.00
From TMax:      7194273  5165 -12130 1059 100 MILE HOUSE       CANADA        1970 1999   -999  -999.00
DBG: AUTOPAIRED:    2   2
DBG: pot.auto i,j:    3   3
DBG: i,ncs2m,cs2m(1-5):    3   1           3    8578    8582    8596       0
DBG: paired:    3   3  HORSE LAKE

Attempting to pair stations:
From TMin:      7103611  5160 -12120  994 HORSE LAKE                        1983 1994    -999 -999.00
From TMax:      7103611  5160 -12120  994 HORSE LAKE                        1983 1994   -999  -999.00
DBG: AUTOPAIRED:    3   3
DBG: pot.auto i,j:    4   4
DBG: i,ncs2m,cs2m(1-5):    4   1           4    8578    8582    8596       0
DBG: paired:    4   4  LONE BUTTE 2

Attempting to pair stations:
From TMin:      7103629  5155 -12120 1145 LONE BUTTE 2                       1981 1991    -999 -999.00
From TMax:      7103629  5155 -12120 1145 LONE BUTTE 2                       1981 1991   -999  -999.00
DBG: AUTOPAIRED:    4   4
DBG: pot.auto i,j:    5   5
DBG: i,ncs2m,cs2m(1-5):    5   1           5    8578    8582    8596       0
DBG: paired:    5   5  100 MILE HOUSE 6NE

Attempting to pair stations:
From TMin:      7103637  5168 -12122  928 100 MILE HOUSE 6NE               1987 2002    -999 -999.00
From TMax:      7103637  5168 -12122  928 100 MILE HOUSE 6NE               1987 2002   -999  -999.00
DBG: AUTOPAIRED:    5   5
DBG: pot.auto i,j:    6   6
DBG: i,ncs2m,cs2m(1-5):    6   1           6    8578    8582    8596       0
```

```
DBG: paired:    6   6  WATCH LAKE NORTH

Attempting to pair stations:
From TMin:      7103660  5147 -12112 1069 WATCH LAKE NORTH                  1987 1996    -999 -999.00
From TMax:      7103660  5147 -12112 1069 WATCH LAKE NORTH                  1987 1996   -999  -999.00
DBG: AUTOPAIRED:    6   6
<END QUOTE>
```

Now arguably, the MILE HOUSE ABEL stations should have rolled into one of the other MILE HOUSE ones with a WMO code.. but the lat/lon/alt aren't close enough. Which is as intended.

*

*

Well, it *kind of* worked. Thought the resultant files aren't exactly what I'd expected:

```
-rw-------   1 f098     cru      12715138 Jul 25 15:25 act.0707241721.dat
-rw-------   1 f098     cru        435839 Jul 25 15:25 log.0707241721.dat
-rw-------   1 f098     cru       4126850 Jul 25 15:25 mat.0707241721.dat
-rw-------   1 f098     cru       6221390 Jul 25 15:25 tmn.0707021605.dtb.lost
-rw-------   1 f098     cru       2962918 Jul 25 15:25 tmn.0707241721.dat
-rw-------   1 f098     cru             0 Jul 25 15:25 tmx.0702091313.dtb.lost
-rw-------   1 f098     cru       2962918 Jul 25 15:25 tmx.0707241721.dat
```

act.0707241721.dat: hopefully-complete record of all activities

log.0707241721.dat: hopefully-useful log of odd happenings (and mergeinfo() trails)

mat.0707241721.dat: hopefully-complete list of all merges and pairings

tmn.0707021605.dtb.lost: too-small collection of unpaired stations

tmn.0707241721.dat: too-small output database

tmx.0702091313.dtb.lost: MUCH too-small collection of unpaired stations!!!

tmx.0707241721.dat: too-small (but hey, the same size as the twin) output database

ANALYSIS

Well, LOL, the reason the output databases are so small is that every station looks like this:

```
9999810  -748  10932  114 SEMPOR              INDONESIA     1971 2000    -999 -999.00
6190-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1971  229  225  225  229  229-9999  223  221  222  225  224-9999
```

Yes - just one line of data. The write loops went from start year to start year. Ho hum :-/

Not as easy to fix as you might think, seeing as the data may well be the result of a merge and so can't just be pasted in from the source database.

As for the 'unbalanced' 'lost' files: well for a start, the same error as above (just one line of data), then on top of that, both sets written to the same file. what time did I write that bit, 3am?!! Ecch.

33. So, as expected.. I'm gonna have to write in clauses to make use of the log, act and mat files. I so do not want to do this.. but not as much as I don't want to do a day's interacting again!!

Got it to work.. sort of. Turns out I had included enough information in the ACT file, and so was able to write auminmaxresync.for. A few teething troubles, but two new databases ('tm[n|x].0707301343.dtb') created with 13654 stations in each. And yes - the headers are identical :-)

[edit: see below - the 'final' databases are tm*.0708071548.dtb]

Here are the header counts, demonstrating that something's still not quite right..

```
Original:
   14355 tmn.0707021605.dtb.heads

New:
   13654 tmn.0707301343.dtb.heads

Lost/merged:
   14318 tmn.0707021605.dtb.lost.heads (should be 14355-13654-37 = 664?)
      37 tmn.0707021605.dtb.merg.heads (seems low)

Original:
   14315 tmx.0702091313.dtb.heads

New:
   13654 tmx.0707301343.dtb.heads

Lost/merged:
   14269 tmx.0702091313.dtb.lost.heads (should be 14315-13654-46 = 615?)
      46 tmx.0702091313.dtb.merg.heads (seems low)
```

In fact, looking at the original ACT file that we used:

```
crua6[/cru/cruts/version_3_0/db/dtr] grep 'usermerg' act.0707241721.dat | wc -l
       258
crua6[/cru/cruts/version_3_0/db/dtr] grep 'automerg' act.0707241721.dat | wc -l
       889
```

..so will have to look at how the db1/2xref arrays are prepped and set in the program. Nonetheless the construction of the new databases looks pretty good. There's aminor problem where the external reference field is sometimes -999.00 and sometimes 0. Not sure which is best, probably 0, as the field will usually be used for reference numbers/characters rather than real data values. Used an inline perl command to fix.

..after some rudimentary corrections:

```
uealogin1[/cru/cruts/version_3_0/db/dtr] wc -l *.heads
   14355 tmn.0707021605.dtb.heads
     122 tmn.0707021605.dtb.lost.heads
     579 tmn.0707021605.dtb.merg.heads
   13654 tmn.0708062250.dtb.heads
   14315 tmx.0702091313.dtb.heads
      93 tmx.0702091313.dtb.lost.heads
     570 tmx.0702091313.dtb.merg.heads
   13654 tmx.0708062250.dtb.heads
```

Almost perfect! But unfortunately, there is a slight discrepancy, and they have a habit of being tips of icebergs. If you add up the header/station counts of the new tmin database, merg and lost files, you get 13654 + 579 + 122 = 14355, the original station count. If you try the same check for tmax, however, you get 13654 + 570 + 93 = 14317, two more than the original count! I suspected a couple of stations were being counted twice, so using 'comm' I looked for identical headers. Unfortunately there weren't any!! So I have invented two stations, hmm. Got the program to investigate, and found two stations in the cross-reference array which had cross refs *and* merge flags:

```
ERROR: db2xref(  126) =      127  -14010 :
```

```
  126> 9596400 -4110  14680    3 LOW HEAD            AUSTRALIA     2000 2006    -999   91293
14010> 9596900 -4170  14710  150 CRESSY RESEARCH STAT AUSTRALIA     1971 2006    -999   91306

and

ERROR: db2xref(13948) =      227    -226 :
13948> 9570600 -3470  14650  145 NARRANDERA AIRPORT   AUSTRALIA     1971 2006    -999       0
  226>       0 -3570  14560  110 FINLEY (CSIRO)       AUSTRALIA     2000 2001    -999       0

So in the first case, LOW HEAD has been merged with another station (#14010) AND paired with #127.
Similarly, NARRANDERA AIRPORT has been mreged with #226 and paired with #227. However, these apparent
merges are false! As we see in the first case, 14010 is not LOW HEAD. Similarly for the second case.

Looking in the relevant match file from the process (mat.0707241721.dat) we find:

AUTO MERGE FROM CHAIN:
TMax Stn 1:       0 -4110  14680    3 LOW HEAD               AUSTRALIA 2000 2006   -999  -999.00
TMax Stn 2:       0 -4105  14678    4 LOW HEAD            AUSTRALIA     2000 2004   -999  -999.00
New Header:       0 -4110  14680    3 LOW HEAD            AUSTRALIA     2000 2006    -999       0
Note: Stn 1 data overwrote Stn 2 data

MANUAL PAIRING FROM CHAIN:
TMin:       9596400 -4110  14680    3 LOW HEAD            AUSTRALIA     2000 2006    -999   91293
TMax:             0 -4110  14680    3 LOW HEAD            AUSTRALIA     2000 2006    -999       0
New Header: 9596400 -4110  14680    3 LOW HEAD            AUSTRALIA     2000 2006    -999   91293

and

AUTO MERGE FROM CHAIN:
TMax Stn 1:       0 -3470  14650  145 NARRANDERA AIRPORT       AUSTRALIA 2000 2006   -999  -999.00
TMax Stn 2: 9570600 -3471  14651  145 NARRANDERA AIRPORT   AUSTRALIA     1972 1980   -999  -999.00
New Header: 9570600 -3470  14650  145 NARRANDERA AIRPORT   AUSTRALIA     1972 2006    -999       0
Note: Stn 2 data overwrote Stn 1 data

MANUAL PAIRING FROM CHAIN:
TMin:       9570600 -3470  14650  145 NARRANDERA AIRPORT   AUSTRALIA     1971 2003    -999       0
TMax:       9570600 -3470  14650  145 NARRANDERA AIRPORT   AUSTRALIA     1972 2006    -999       0
New Header: 9570600 -3470  14650  145 NARRANDERA AIRPORT   AUSTRALIA     1971 2006    -999       0

Found the problem - mistyping of an assignment.. and so:

crua6[/cru/cruts/version_3_0/db/dtr] wc -l *.heads

     14355 tmn.0707021605.dtb.heads
       122 tmn.0707021605.dtb.lost.heads
       579 tmn.0707021605.dtb.merg.heads
     13654 tmn.0708071548.dtb.heads

     14315 tmx.0702091313.dtb.heads
        93 tmx.0702091313.dtb.lost.heads
       568 tmx.0702091313.dtb.merg.heads
     13654 tmx.0708071548.dtb.heads

Phew! Well the headers are identical for the two new databases:

crua6[/cru/cruts/version_3_0/db/dtr] cmp tmn.0708071548.dtb.heads  tmx.0708071548.dtb.heads |wc -l
         0

34. So the to the real test - converting to DTR! Wrote tmnx2dtr.for, which does exactly that. It reported
233 instances where tmin > tmax (all set to missing values) and a handful where tmin == tmax (no prob).
Looking at the 233 illogicals, most of the stations look as though considerable work is needed on them.
This highlights the fact that all I've done is to synchronise the tmin and tmax databases with each
other, and with the Australian stations - there is still a lot of data cleansing to perform at some
stage! But not right now :-)

Input Files
  TMin: tmn.0708071548.dtb
  TMax: tmx.0708071548.dtb

Output file
   DTR: dtr.0708071924.dtb

Cases of identical values:  39
Cases of min > max (BAD!): 233
All illegals written to:   illdtr.0708071924.dat

Example of 'illegal' values to demonstrate quality of station data:

station:  9600100   587   9532  126 SABANG/CUT BAU       ID ID         1984 2006    -999       0
min data: 2006  203 -197  200-9999 -211  207  233-9999-9999-9999-9999-9999
max data: 2006  290 -299  307-9999 -315  309  308-9999-9999-9999-9999-9999

Doesn't look very likely!

Normals added:

crua6[/cru/cruts/version_3_0/db/dtr] ./addnormline

    ****  ADDNORMLINE ****

  Calculates monthly normals
  for 1961-1990, provided at
  least  75% of values are
  present. Results go into a
  normals line coming after
  the header. Operator called
  if different normals exist!

Please enter the input database: dtr.0708071924.dtb

Proposed output database name:   dtr.0708081052.dat

                              ACCEPT/REJECT (A/R): A
Output database name: dtr.0708081052.dat
Derived logfile name: dtr.0708081052.log

So the final DTR database is dtr.0708081052.dtb.

And so to the main process:

<BEGIN QUOTE>
crua6[/cru/cruts/version_3_0/primaries/dtr] ./anomdtb

   > ***** AnomDTB: converts .dtb to anom .txt for gridding *****

   > Enter the suffix of the variable required:
.dtr
   > Select the .cts or .dtb file to load:
dtr.0708081052.dtb
```

```
   > Specify the start,end of the normals period:
1961,1990
   > Specify the missing percentage permitted:
25
   > Data required for a normal:           23
   > Specify the no. of stdevs at which to reject data:
3
   > Select outputs (1=.cts,2=.ann,3=.txt,4=.stn):
3
   > Check for duplicate stns after anomalising? (0=no,>0=km range)
0
   > Select the generic .txt file to save (yy.mm=auto):
dtr.txt
   > Select the first,last years AD to save:
1901,2006
   > Operating...

   > NORMALS            MEAN percent      STDEV percent
   >         .dtb    3746373    65.9
   >         .cts     178161     3.1    3924534    69.0
   > PROCESS        DECISION percent %of-chk
   > no lat/lon          650     0.0     0.0
   > no normal       1763302    31.0    31.0
   > out-of-range         24     0.0     0.0
   > accepted        3924510    69.0
   > Dumping years 1901-2006 to .txt files...

<END QUOTE>

So a lower pewrcentage than last time (69.0 vs. 78.9), but then, more data overall so a better
result (3924510 vs. 3167636).

Gridding:
IDL> quick_interp_tdm2,1901,2006,'dtrglo/dtr.',750,gs=0.5,pts_prefix='dtrtxt/dtr.',dumpglo='dumpglo'

Convert from .glo:
crua6[/cru/cruts/version_3_0/primaries/dtr] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.dtr
Enter a name for the gridded climatology file: clim.6190.lan.dtr.grid
Enter the path and stem of the .glo files: dtrglo/dtr.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files:
Now, CONCENTRATE. Addition or Percentage (A/P)? A
Right, erm.. off I jolly well go!
dtr.01.1901.glo
(etc)
dtr.12.2006.glo

Finally, gridding:
Writing cru_ts_3_00.1901.1910.dtr.dat
Writing cru_ts_3_00.1911.1920.dtr.dat
Writing cru_ts_3_00.1921.1930.dtr.dat
Writing cru_ts_3_00.1931.1940.dtr.dat
Writing cru_ts_3_00.1941.1950.dtr.dat
Writing cru_ts_3_00.1951.1960.dtr.dat
Writing cru_ts_3_00.1961.1970.dtr.dat
Writing cru_ts_3_00.1971.1980.dtr.dat
Writing cru_ts_3_00.1981.1990.dtr.dat
Writing cru_ts_3_00.1991.2000.dtr.dat
Writing cru_ts_3_00.2001.2006.dtr.dat
Writing cru_ts_3_00.1901.2006.dtr.dat


35. Onto the secondaries, working from the rerun methodology (see section 20 above).

Began with temperature, using the anomaly txt files from the half-degree generation:

IDL> quick_interp_tdm2,1901,2006,'tmpbin/tmpbin',1200,gs=2.5,dumpbin='dumpbin',pts_prefix='tmp0km0705101334txt/tmp.'

This produced binaries such as 'tmpbin1901'.

Then precipitation:

IDL> quick_interp_tdm2,1901,2006,'prebin/prebin',450,gs=2.5,dumpbin='dumpbin',pts_prefix='pre0km0612181221txt/pre.'

Finally, dtr:

IDL> quick_interp_tdm2,1901,2006,'dtrbin/dtrbin',50,gs=2.5,dumpbin='dumpbin',pts_prefix='dtrtxt/dtr.'

*** EEEK! Is that '50' a mistype? Meaning that anything using binary DTR will need re-doing? (RAL, Dec 07) ***

And so to the synthetics.

FRS:

IDL> .compile /cru/cruts/fromdpe1a/code/idl/pro/rdbin.pro
% Compiled module: RDBIN.
IDL> .compile /cru/cruts/fromdpe1a/code/idl/pro/frs_gts_tdm.pro
% Compiled module: FRS_GTS.
IDL> frs_gts,dtr_prefix='dtrbin/dtrbin',tmp_prefix='tmpbin/tmpbin',1901,2006,outprefix='frssyn/frssyn'
IDL> quick_interp_tdm2,1901,2006,'frsgrid/frsgrid',750,gs=0.5,dumpglo='dumpglo',nostn=1,synth_prefix='frssyn/frssyn'

crua6[/cru/cruts/version_3_0/secondaries/frs] ../glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.frs
Enter a name for the gridded climatology file: clim.6190.lan.frs.grid
Enter the path and stem of the .glo files: frsgrid/frsgrid.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files:
Now, CONCENTRATE. Addition or Percentage (A/P)? A
Right, erm.. off I jolly well go!
frsgrid.01.1901.glo
(etc)
frsgrid.12.2006.glo

crua6[/cru/cruts/version_3_0/secondaries/frs] ../mergegrids
Welcome! This is the MERGEGRIDS program.
I will create decadal and full gridded files
from the output files of (eg) glo2abs.for.

Enter a gridfile with YYYY for year and MM for month: frsgridabs/frsgrid.MM.YYYY.glo.abs
```

```
Enter Start Year:  1901
Enter Start Month: 01
Enter End Year:    2006
Enter End Month:   12

Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE: cru_ts_3_00.SSSS.EEEE.frs.dat
Writing cru_ts_3_00.1901.1910.frs.dat
Writing cru_ts_3_00.1911.1920.frs.dat
Writing cru_ts_3_00.1921.1930.frs.dat
Writing cru_ts_3_00.1931.1940.frs.dat
Writing cru_ts_3_00.1941.1950.frs.dat
Writing cru_ts_3_00.1951.1960.frs.dat
Writing cru_ts_3_00.1961.1970.frs.dat
Writing cru_ts_3_00.1971.1980.frs.dat
Writing cru_ts_3_00.1981.1990.frs.dat
Writing cru_ts_3_00.1991.2000.frs.dat
Writing cru_ts_3_00.2001.2006.frs.dat

RD0:

IDL> .compile /cru/cruts/fromdpela/code/idl/pro/rdbin.pro
% Compiled module: RDBIN.
IDL> .compile /cru/cruts/fromdpela/code/idl/pro/rd0_gts_tdm.pro
% Compiled module: RD0_GTS.
IDL> rd0_gts,1901,2006,1961,1990,outprefix='rd0syn/rd0syn',pre_prefix='prebin/prebin'
Reading precip and rd0 normals
% Compiled module: STRIP.
yes
filesize=     6220800
gridsize=     0.500000
% Compiled module: DEFXYZ.
yes
filesize=     6220800
gridsize=     0.500000
% Compiled module: DAYS.
Calculating synthetic Rd0 normal
    1961
yes
filesize=      248832
gridsize=      2.50000
% Compiled module: RD0CAL.
    1962
yes

(etc)

    2006
yes
filesize=      248832
gridsize=      2.50000
% Program caused arithmetic error: Floating divide by 0
% Program caused arithmetic error: Floating illegal operand
IDL>

(as before, see section 20.)

IDL> quick_interp_tdm2,1901,2006,'rd0grid/rd0grid',450,gs=0.5,dumpglo='dumpglo',nostn=1,synth_prefix='rd0syn/rd0syn'

crua6[/cru/cruts/version_3_0/secondaries/rd0] ../glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: forrtl: error (69): process interrupted (SIGINT)
crua6[/cru/cruts/version_3_0/secondaries/rd0] mkdir rd0gridabs
crua6[/cru/cruts/version_3_0/secondaries/rd0] ../glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.wet
Enter a name for the gridded climatology file: clim.6190.lan.wet.grid
Enter the path and stem of the .glo files: rd0grid/rd0grid.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: rd0gridabs/
Now, CONCENTRATE. Addition or Percentage (A/P)? A
Right, erm.. off I jolly well go!
rd0grid.01.1901.glo
(etc)
rd0grid.12.2006.glo

crua6[/cru/cruts/version_3_0/secondaries/rd0] ../mergegrids
Welcome! This is the MERGEGRIDS program.
I will create decadal and full gridded files
from the output files of (eg) glo2abs.for.

Enter a gridfile with YYYY for year and MM for month: rd0gridabs/rd0grid.MM.YYYY.glo.abs
Enter Start Year:  1901
Enter Start Month: 01
Enter End Year:    2006
Enter End Month:   12

Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE: cru_ts_3_00.SSSS.EEEE.rd0.dat
Writing cru_ts_3_00.1901.1910.rd0.dat
(etc)


I have to admit, I still don't understand secondary parameter generation. I've read the papers, and the
miniscule amount of 'Read Me' documentation, and it just doesn't make sense. In particular, why use 2.5
degree grids of the primaries instead of 0.5? Why deliberately lose spatial resolution, only to have to
reinterpolate later?

No matter; on to Vapour Pressure. Here's the complete output from the initial binary gridding,using dtr and tmp:

IDL> vap_gts_anom,dtr_prefix='dtrbin/dtrbin',tmp_prefix='tmpbin/tmpbin',1901,2006,outprefix='vapsyn/vapsyn',dumpbin=1
% Compiled module: VAP_GTS_ANOM.
% Compiled module: RDBIN.
% Compiled module: STRIP.
% Compiled module: DEFXYZ.
Land,sea:       56016       68400
Calculating tmn normal
% Compiled module: TVAP.
Calculating synthetic vap normal
% Compiled module: ESAT.
Calculating synthetic anomalies
% Compiled module: MOMENT.
1901 vap (x,s2,<<,>>):  1.61250e-05  6.15570e-06    -0.160607     0.222689
```

```
% Compiled module: WRBIN.
1902 vap (x,s2,<<,>>): -0.000123188  3.46116e-05    -0.268891    0.0261283
1903 vap (x,s2,<<,>>):  6.86689e-05  4.52675e-06    -0.121429     0.123995
1904 vap (x,s2,<<,>>): -1.30788e-05  1.83887e-05    -0.454975    0.0919596
1905 vap (x,s2,<<,>>):  1.94645e-05  1.32224e-05    -0.408679    0.0498396
1906 vap (x,s2,<<,>>):  3.22279e-05  3.74796e-06    -0.178658    0.0261283
1907 vap (x,s2,<<,>>): -2.56545e-05  1.68228e-05    -0.268768    0.0498040
1908 vap (x,s2,<<,>>):  6.39573e-05  3.49149e-06    -0.173230     0.354836
1909 vap (x,s2,<<,>>):  3.50080e-05  3.21530e-06    -0.201157    0.0261283
1910 vap (x,s2,<<,>>):  3.45249e-05  6.15026e-06    -0.130285     0.144744
1911 vap (x,s2,<<,>>):  3.99470e-05  5.85673e-06    -0.360082    0.0261283
1912 vap (x,s2,<<,>>): -7.91931e-06  1.06891e-05    -0.279282    0.0261283
1913 vap (x,s2,<<,>>):  6.07153e-05  7.10663e-07   -0.0148902    0.0261283
1914 vap (x,s2,<<,>>):  7.22507e-05  2.52354e-06    -0.130205     0.124774
1915 vap (x,s2,<<,>>): -2.11176e-05  1.59592e-05    -0.308456    0.0579963
1916 vap (x,s2,<<,>>): -8.95735e-05  2.41852e-05    -0.247123     0.140438
1917 vap (x,s2,<<,>>): -0.000105104  2.43058e-05    -0.229282     0.282290
1918 vap (x,s2,<<,>>):  1.14711e-05  7.76188e-06    -0.248782    0.0261283
1919 vap (x,s2,<<,>>):  2.51597e-05  5.75406e-06    -0.295303     0.215085
1920 vap (x,s2,<<,>>): -2.78549e-06  1.81183e-05    -0.373193    0.0261283
1921 vap (x,s2,<<,>>):  6.07153e-05  7.10663e-07   -0.0148902    0.0261283
1922 vap (x,s2,<<,>>): -1.86602e-05  1.22345e-05    -0.275667    0.0261283
1923 vap (x,s2,<<,>>):  5.76800e-05  1.22728e-06    -0.170021    0.0261283
1924 vap (x,s2,<<,>>):  6.07153e-05  7.10663e-07   -0.0148902    0.0261283
1925 vap (x,s2,<<,>>):  8.32519e-05  5.55618e-06    -0.109315     0.186182
1926 vap (x,s2,<<,>>):  0.000106602  5.15263e-06    -0.105764     0.206929
1927 vap (x,s2,<<,>>):  5.23023e-05  2.64333e-06    -0.194649    0.0498040
1928 vap (x,s2,<<,>>):  5.50934e-05  2.47944e-06    -0.314917    0.0261283
1929 vap (x,s2,<<,>>): -0.000524952  0.000155755    -0.417342     0.215959
1930 vap (x,s2,<<,>>):  8.28323e-05  1.87314e-05    -0.328074     0.193805
1931 vap (x,s2,<<,>>): -7.80687e-05  3.63543e-05    -0.315060     0.215417
1932 vap (x,s2,<<,>>):  5.62579e-05  3.81547e-06    -0.249130     0.120583
1933 vap (x,s2,<<,>>): -3.47433e-05  1.69009e-05    -0.218800     0.148224
1934 vap (x,s2,<<,>>):  0.000156604  1.56121e-05    -0.173230     0.152809
1935 vap (x,s2,<<,>>):  6.69520e-05  4.91451e-06    -0.160529     0.120391
1936 vap (x,s2,<<,>>): -0.000255663  6.63373e-05    -0.398866    0.0261283
1937 vap (x,s2,<<,>>):  6.99402e-05  2.70766e-05    -0.328074     0.201202
1938 vap (x,s2,<<,>>):  5.91796e-05  6.70722e-06    -0.215017     0.155977
1939 vap (x,s2,<<,>>):  4.88266e-05  5.25789e-06    -0.173294    0.0893239
1940 vap (x,s2,<<,>>):  9.63896e-06  7.45103e-06    -0.214763    0.0758103
1941 vap (x,s2,<<,>>):  4.11127e-05  4.15525e-06    -0.234030    0.0261283
1942 vap (x,s2,<<,>>): -9.97969e-05  3.88466e-05    -0.288682     0.148893
1943 vap (x,s2,<<,>>):  8.38607e-05  3.48416e-06   -0.0148902     0.163562
1944 vap (x,s2,<<,>>):  7.96681e-05  7.91305e-06    -0.227413     0.104055
1945 vap (x,s2,<<,>>):  3.37215e-05  3.99524e-06    -0.248782    0.0261283
1946 vap (x,s2,<<,>>):  5.31976e-05  2.63755e-06    -0.128263     0.163584
1947 vap (x,s2,<<,>>):  0.000131113  1.66296e-05    -0.353903     0.193758
1948 vap (x,s2,<<,>>):  6.80941e-05  1.62353e-06   -0.0148902     0.163624
1949 vap (x,s2,<<,>>):  2.47925e-05  2.45819e-05    -0.328074     0.237848
1950 vap (x,s2,<<,>>): -9.57348e-05  7.78468e-05    -0.366764     0.726541
1951 vap (x,s2,<<,>>): -6.54446e-06  1.35656e-05    -0.446058    0.0261283
1952 vap (x,s2,<<,>>): -0.000158974  5.02732e-05    -0.262313     0.193617
1953 vap (x,s2,<<,>>):  1.18525e-05  4.22691e-05    -0.282204     0.230629
1954 vap (x,s2,<<,>>): -0.000151975  6.78713e-05    -0.373235     0.230602
1955 vap (x,s2,<<,>>): -0.000134153  5.23124e-05    -0.298578    0.0841820
1956 vap (x,s2,<<,>>): -9.61671e-05  5.20484e-05    -0.492004    0.0888951
1957 vap (x,s2,<<,>>): -1.18048e-05  1.31769e-05    -0.220902    0.0261283
1958 vap (x,s2,<<,>>): -8.61762e-06  1.12079e-05    -0.207799     0.148170
1959 vap (x,s2,<<,>>):  8.27399e-05  4.88857e-06   -0.0929929     0.170919
1960 vap (x,s2,<<,>>):  3.38773e-05  1.53901e-05    -0.207944     0.155940
1961 vap (x,s2,<<,>>):  5.72571e-05  9.01807e-07   -0.0653905    0.0261283
1962 vap (x,s2,<<,>>):  8.20891e-05  3.78016e-06    -0.240435     0.126662
1963 vap (x,s2,<<,>>): -0.000108489  3.85148e-05    -0.266356    0.0836364
1964 vap (x,s2,<<,>>):  3.02043e-05  6.37207e-06    -0.240547     0.150816
1965 vap (x,s2,<<,>>):  5.76898e-05  2.48022e-06    -0.279282     0.143283
1966 vap (x,s2,<<,>>): -0.000300312  5.32054e-05    -0.622719    0.0261283
1967 vap (x,s2,<<,>>):  6.43500e-05  8.58218e-07   -0.0148902    0.0496181
1968 vap (x,s2,<<,>>): -0.000241750  4.22773e-05    -0.214442     0.271730
1969 vap (x,s2,<<,>>): -0.000568502  9.92260e-05    -0.385322    0.0732047
1970 vap (x,s2,<<,>>):  6.07153e-05  7.10663e-07   -0.0148902    0.0261283
1971 vap (x,s2,<<,>>):  2.15333e-05  4.77100e-06    -0.188071    0.0261283
1972 vap (x,s2,<<,>>): -7.14160e-05  3.56948e-05    -0.365803     0.201611
1973 vap (x,s2,<<,>>):  5.77503e-05  1.17079e-06    -0.160550    0.0261283
1974 vap (x,s2,<<,>>):  3.49354e-05  4.93069e-06    -0.149678     0.144313
1975 vap (x,s2,<<,>>):  6.14429e-05  7.36204e-07   -0.0148902    0.0380432
1976 vap (x,s2,<<,>>):  6.49657e-05  3.25410e-06    -0.266356     0.165472
1977 vap (x,s2,<<,>>):  0.000107180  1.92804e-05    -0.304625     0.208459
1978 vap (x,s2,<<,>>): -4.80106e-05  3.28909e-05    -0.285492     0.105108
1979 vap (x,s2,<<,>>): -0.000102001  2.35900e-05    -0.214390     0.112952
1980 vap (x,s2,<<,>>):  4.16963e-05  2.70211e-06    -0.144913    0.0864268
1981 vap (x,s2,<<,>>):  0.000274196  1.86668e-05   -0.0148902     0.222522
1982 vap (x,s2,<<,>>):  8.57426e-07  7.08135e-06    -0.161781    0.0831981
1983 vap (x,s2,<<,>>): -5.84499e-06  1.76470e-05    -0.234194     0.128289
1984 vap (x,s2,<<,>>): -0.000106476  2.97454e-05    -0.335850     0.150833
1985 vap (x,s2,<<,>>):  9.32757e-06  4.35533e-05    -0.323331     0.222522
1986 vap (x,s2,<<,>>):  7.22110e-05  4.76179e-06    -0.141725     0.185658
1987 vap (x,s2,<<,>>): -2.27107e-05  2.09631e-05    -0.291446     0.103599
1988 vap (x,s2,<<,>>):  6.58090e-05  9.21014e-07   -0.0148902    0.0670816
1989 vap (x,s2,<<,>>):  9.54406e-05  1.72599e-05    -0.266297     0.160293
1990 vap (x,s2,<<,>>):  0.000218826  3.56583e-05    -0.174187     0.236204
1991 vap (x,s2,<<,>>):  5.93288e-05  8.18618e-07   -0.0776650    0.0261283
1992 vap (x,s2,<<,>>):  7.57687e-05  4.27091e-06    -0.174292     0.215085
1993 vap (x,s2,<<,>>): -1.69378e-05  2.36942e-05    -0.314882    0.0420169
1994 vap (x,s2,<<,>>):  6.36348e-05  1.18760e-06   -0.0148902     0.163543
1995 vap (x,s2,<<,>>):  0.000281573  6.09912e-05    -0.463574     0.259426
1996 vap (x,s2,<<,>>):  5.03362e-05  5.47691e-06    -0.224751     0.124774
1997 vap (x,s2,<<,>>):  0.000132649  2.97693e-05    -0.446455     0.281070
1998 vap (x,s2,<<,>>):  5.96544e-07  3.39098e-05    -0.359037     0.201228
1999 vap (x,s2,<<,>>):  5.91499e-05  2.37232e-06    -0.166206     0.215985
2000 vap (x,s2,<<,>>):  4.06034e-05  4.61604e-06   -0.0898572     0.191977
2001 vap (x,s2,<<,>>):  0.000138230  8.53512e-06   -0.0512625     0.206929
2002 vap (x,s2,<<,>>):  0.000218003  4.36873e-05    -0.760830     0.282290
2003 vap (x,s2,<<,>>):  7.00864e-05  7.67472e-06    -0.301868     0.237875
2004 vap (x,s2,<<,>>):  5.49200e-06  2.13246e-05    -0.500544     0.112129
2005 vap (x,s2,<<,>>):  6.05939e-06  5.83817e-05    -0.885566     0.199814
2006 vap (x,s2,<<,>>):  9.02885e-05  3.60834e-05    -0.455230     0.607388

How very useful! No idea what any of that means. although it's heartwarming to see that it's
nothing like the results of the 2.10 rerun, where 1991 looked like this:

1991 vap (x,s2,<<,>>):  0.000493031  0.000742087   -0.0595093      1.86497

Now, of course, it looks like this:

1991 vap (x,s2,<<,>>):  5.93288e-05  8.18618e-07   -0.0776650    0.0261283

From this I can deduce.. err.. umm..

Anyway now I need to use whatever VAP station data we have. And here I'm a little flaky (again),
the vap database hasn't been updated, is it going to be? Asked Dave L and he supplied summaries
```

he'd produced of CLIMAT bulletins from 2000-2006. Slightly odd format but very useful all the
same.

And now, a brief interlude. As we've reached the stage of thinking about secondary variables, I
wondered about the CLIMAT updates, as one of the outstanding work items is to write routines to
convert CLIMAT and MCDW bulletins to CRU format (so that mergedb.for can read them). So I look at
a CLIMAT bulletin, and what's the first thing I notice? It's that there is absolutely no station
identification information apart from the WMO code. None. No lat/lon, no name, no country. Which
means that all the bells and whistles I built into mergedb, (though they were needed for the db
merging of course) are surplus to requirements. The data must simply be added to whichever station
has the same number at the start, and there's no way to check it's right. I don't appear to have a
copy of a MCDW bulletin yet, only a PDF.. I wonder if that's the same? Anyway, back to the main job.

As I was examining the vap database, I noticed there was a 'wet' database. Could I not use that to
assist with rd0 generation? well.. it's not documented, but then, none of the process is so I might
as well bluff my way into it! Units seem to vary:

CLIMAT bulletins have day counts:

```
 SURFACE LAND 'CLIMAT' DATA FOR  2006/10.   MISSING DATA=-32768
  MET OFFICE, HADLEY CENTRE CROWN COPYRIGHT
WMO BLK WMO STN   STNLP    MSLP    TEMP   VAP P DAYS RN  RAIN R   QUINT SUN HRS   SUN %   MIN_T   MAX_T
     01     001   10152   10164      5     52      9       63        2   -32768  -32768     -12      20
```

Dave L's CLIMAT update has days x 10:

```
 100100 7093  -867    9JAN MAYEN(NOR-NAVY) NORWAY       20002006        -7777777
 2000  150  120  180   60  150   20   30  130  120  150   70   70
```

The existing 'wet' database (wet.0311061611.dtb) has days x 100:

```
   10010  7093   -866    9 JAN MAYEN(NOR NAVY)  NORWAY        1990 2003   -999     -999
6190-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1990-9999-9999-9999-9999  400  600  600 1800 1500 1100  800 1800
```

The published climatology has days x 100 as well:

```
Tyndall Centre grim file created on 13.01.2004 at 15:22 by Dr. Tim Mitchell
.wet = wet day frequency (days)
0.5deg lan clim:1961-90 MarkNew but adj so that wet=<pre
[Long=-180.00, 180.00] [Lati= -90.00,  90.00] [Grid X,Y= 720, 360]
[Boxes=   67420] [Years=1975-1975] [Multi=    0.0100] [Missing=-999]
Grid-ref=   1, 148
 1760 1580 1790 1270  890  510  470  290  430  400  590 1160
```

So I guess we go with days x100. Dave's files will have to be reformatted anyway so it's a
negligible overhead. Okaaaay..

Wrote dave2cru.for to convert Dave L's CLIMAT composites to CRU-format files in the appropriate
units. One problem is the significant number of stations without names or countries: they are
simply 'xxxxxxxxxx' and I'm not sure how mergedb is going to take to that! Well only one way to
find out.. so I converted the rain days data:

<BEGIN QUOTE>

```
crua6[/cru/cruts/version_3_0/db] ./dave2cru

DAVE2CRU - convert Dave L CLIMAT composites to dtb files
Enter the CLIMAT composite to be converted: CLIMAT_MCDW_MCDW_rdy_updat_merged

Example data line from that file:
2000  150  120  180   60  150   20   30  130  120  150   70   70

Please enter a factor to apply (or 1): 10
Please enter the 3-ch parameter code: rd0

The output file will be: rd0.0708151122.dtb

 3411 stations written.
```

<END QUOTE>

Then tried to merge that into wet.0311061611.dtb, and immediately hot formatting issues - that pesky last
field has been badly abused here, taking values including:

```
 -999.00
    0.00
  nocode     (yes, really!)
```

Had a quick review of mergedb; it won't be trivial to update it to treat that field as a8. So reluctantly,
changed all the 'nocode' entries to '0':

```
crua6[/cru/cruts/version_3_0/db/rd0] perl -pi -e 's/nocode/     0/g' wet.0311061611.dt*
```

Unfortunately, that didn't solve the problems.. as there are alphanumerics in that field later on:

```
-712356  5492 -11782  665 SPRING CRK WOLVERINE CANADA        1969 1988   -999  307F0P9
```

So.. ***sigh***.. will have to alter mergedb.for to treat that field as alpha. Aaarrgghhh.

Did that. Next problem is best summarised with an example:

```
************************************************
*                                              *
*  OPERATOR DECISION REQUIRED:                 *
*                                              *
 100100  7093   -867    9 JAN MAYEN(NOR-NAVY)  NORWAY        2000 2006    -999       0
*                                              *
*  This incoming station has a possible match in *
*  the current database, but either the WMO code *
*  or the lat/lon values differ.               *
*                                              *
*  Incoming:                                   *
 100100  7093   -867    9 JAN MAYEN(NOR-NAVY)  NORWAY        2000 2006    -999       0
*  Potential match:                            *
  10010  7093   -866    9 JAN MAYEN(NOR NAVY)  NORWAY        1990 2003    -999    -999
```

Yes, the 'wet' database features old-style 5-digit WMO codes. The best approach is probably to alter
mergedb again, to multiply any 5-digit codes by 10. Not sure if there is a similar problem with 7-digit
codes, hopefully not.

Oh, more bloody delays. Modified mergedb to 'adjust' the WMO codes, fine. But then a proper run of it
just demonstrated that it's far too picky. Even a 0.01-degree difference in coordinates required ops
intervention. What we need for updates is an absolute priority for WMO codes, and only a shout if the
name or the spatial coordinates are waaay off. I am seriously considering scrapping mergedb in favour of
a version of auminmaxresync - its cloud-based approach and 'intelligent' matching is far more efficient
than mergedb's brute-force attack, as you'd expect from a program built on top of that knowledge. And it
does save all its actions. But I don't know that I have the wherewithal.. okay, I do.

Derived newmergedb.for from auminmaxresync.for. Should be fairly robust. Doesn't offer as many bells

and whistles as mergedb.for, but should be faster and more helpful all the same.

Well.. it works.. but the data doesn't. It's that old devil called WMO numbering again:

```
Comparing Update:  718000  4868    622  217 NANCY/ESSEY          FRANCE        2001 2002    -999       0
   ..with Master:  718000  4665  -5306   28 CAPE RACE (MARS)     CANADA        1920 1969    -999     -999
```

Now what's happened here? Well the CLIMAT numbering only gives five digits (71 800) and so an extra zero has been added to bring it up to six. Unfortunately, that's the wrong thing to do, because that's the code of CAPE RACE. The six-digit code for NANCY/ESSEY is 071800. Mailed Phil and DL as this could be a big problem - many of the Update stations have no other metadata!

Also noticed that some of the CLIMAT data seemed to be missing, eg for NANCY/ESSEY:

```
 718000 4868   622  217NANCY/ESSEY         FRANCE       20002006        -7777777
 2000-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
 2001-9999  110-9999-9999-9999-9999-9999  120  150  110  130   90
 2002   80  160   70   70   80   30   60  120  100  130  180  140
 2003-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
 2004-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
 2005-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
 2006-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
```

I have the CLIMAT bulletin for 10/2006, which gives data for Rain Days (12 in this case). It doesn't seem likely that nothing was reported after 2002.

I am now wondering whether it would be best to go back to the MCDW and CLIMAT bulletins themselves and work directly from those.

--

Well, information is always useful. And I probably did know this once.. long ago. All official WMO codes are five digits, countrycountrystationstationstation. However, we use seven-digit codes, because when no official code is available we improvise with two extra digits. Now I can't see why we didn't leave the rest at five digits, that would have been clear. I also can't see why, if we had to make them all seven digits, we extended the 'legitimate' five-digit codes by multiplying by 100, instead of adding two numerically-meaningless zeros at the most significant (left) end. But, that's what happened, and like everything else that's the way it's staying.

So - incoming stations with WMO codes can only match stations with codes ending '00'. Put another way, for comparison purposes any 7-digit codes ending '00' should be truncated to five digits.

Also got the locations of the original CLIMAT and MCDW bulletins.

CLIMAT are here:
http://hadobs.metoffice.com/crutem3/data/station_updates/

MCDW are here:
ftp://ftp1.ncdc.noaa.gov/pub/data/mcdw
http://www1.ncdc.noaa.gov/pub/data/mcdw/

Downloaded all CLIMAT and MCDW bulletins (CLIMAT 01/2003 to 07/2007; MCDW 01/2003 to 06/2007 (with a mysterious extra called 'ssm0302.Apr211542' - which turns out to be identical to ssm0302.fin)).

Wrote mcdw2cru.for and climat2cru.for, just guess what they do, go on..

<BEGIN QUOTE>
```
uealogin1[/cru/cruts/version_3_0/incoming/MCDW] ./mcdw2cru

MCDW2CRU: Convert MCDW Bulletins to CRU Format

Enter the earliest MCDW file: ssm0301.fin
Enter the latest MCDW file (or <ret> for single files): ssm0706.fin

All Files Processed
tmp.0709071541.dtb: 2407 stations written
vap.0709071541.dtb: 2398 stations written
pre.0709071541.dtb: 2407 stations written
sun.0709071541.dtb: 1693 stations written

Thanks for playing! Byeee!
```
<END QUOTE>

<BEGIN QUOTE>
```
uealogin1[/cru/cruts/version_3_0/incoming/CLIMAT] ./climat2cru

CLIMAT2CRU: Convert MCDW Bulletins to CRU Format

Enter the earliest CLIMAT file: climat_data_200301.txt
Enter the latest CLIMAT file (or <ret> for single file): climat_data_200707.txt

All Files Processed
tmp.0709071547.dtb: 2881 stations written
vap.0709071547.dtb: 2870 stations written
pre.0709071547.dtb: 2878 stations written
sun.0709071547.dtb: 2020 stations written
tmn.0709071547.dtb: 2800 stations written
tmx.0709071547.dtb: 2800 stations written

Thanks for playing! Byeee!
```
<END QUOTE>

Of course, it wasn't quite that simple. MCDW has an inexplicably complex format, which I'm sure will vary over time and eventually break the converter. For instance, most text is left-justified, except the month names for the overdue data, which are right-justified. Also, there is no missing value code, just blank space if a value is absent. This necessitates reading everything as strings and then testing for content. Oh, and a small amount of rain is marked 'T'.. as are small departures from the mean!!

So moan over, now we have a set of updates for the secondary databases. And, indeed for the primary ones - except that I've already processed those, as updated by Dave L.. er.. ah well. So as I'm running stupidly late anyway - why not find out? It's that Imp of the Perverse on my shoulder again.

Actually as I examined all the databases in the tree to work out what was wheat and what chaff, I had my awful memory jogged quite nastily: WE NEED RAIN DAYS. So both conversion progs will need adjusting and re-running!! Waaaaah! And frankly at 18:45 on a Friday evening.. it's not gonna happen right now.

..okay, a another week, another razorblade to slide down. Modified mcdw2cru to include rain days:

<BEGIN QUOTE>
```
uealogin1[/cru/cruts/version_3_0/incoming/MCDW] ./mcdw2cru

MCDW2CRU: Convert MCDW Bulletins to CRU Format

Enter the earliest MCDW file: ssm0301.fin
Enter the latest MCDW file (or <ret> for single files): ssm0706.fin

All Files Processed
tmp.0709111032.dtb: 2407 stations written
vap.0709111032.dtb: 2398 stations written
rdy.0709111032.dtb: 2407 stations written
```

```
pre.0709111032.dtb: 2407 stations written
sun.0709111032.dtb: 1693 stations written

Thanks for playing! Byeee!
<END QUOTE>

Checked, and the four preexisting databases match perfectly with their counterparts, so I didn't break
anything in the adjustments. and the rdy file looks good too (actually the above is the *final* run;
there were numerous bugs as per).

<BEGIN QUOTE>
uealogin1[/cru/cruts/version_3_0/incoming/CLIMAT] ./climat2cru

CLIMAT2CRU: Convert MCDW Bulletins to CRU Format

Enter the earliest CLIMAT file: climat_data_200301.txt
Enter the latest CLIMAT file (or <ret> for single file): climat_data_200707.txt

All Files Processed
tmp.0709101706.dtb: 2881 stations written
vap.0709101706.dtb: 2870 stations written
rdy.0709101706.dtb: 2876 stations written
pre.0709101706.dtb: 2878 stations written
sun.0709101706.dtb: 2020 stations written
tmn.0709101706.dtb: 2800 stations written
tmx.0709101706.dtb: 2800 stations written

Thanks for playing! Byeee!
<END QUOTE>

Again, existing outputs are unchanged and the new rdy file looks OK (though see bracketed note above for MCDW).

So.. to the incorporation of these updates into the secondary databases. Oh, my.

Beginning with Rain Days, known variously as rd0, rdy, pdy.. this allowed me to modify newmergedb.for to cope
with various 'freedoms' enjoyed by the existing databases (such as six-digit WMO codes). And then, when run,
an unexpected side-effect of my flash correlation display thingy: it shows up existing problems with the data!

Here is the first 'issue' encountered by newmergedb, taken from the top and with my comments in <anglebrackets>:

<BEGIN QUOTE>
uealogin1[/cru/cruts/version_3_0/db/rd0] ./newmergedb

WELCOME TO THE DATABASE UPDATER

Before we get started, an important question:
Should the incoming 'update' header info and data take precedence over the existing database?
Or even vice-versa? This will significantly reduce user decisions later, but is a big step!

Enter 'U' to give Updates precedence, 'M' to give Masters precedence, 'X' for equality: U
Please enter the Master Database name: wet.0311061611.dtb
Please enter the Update Database name: rdy.0709111032.dtb

Reading in both databases..
Master database stations:     4988
Update database stations:     2407

Looking for WMO code matches..

***** OPERATOR ADJUDICATION REQUIRED *****

In attempting to pair two stations, possible data incompatibilities have been found.

MASTER:  221130  6896   3305   51 MURMANSK             EX USSR       1936 2003   -999     -999
UPDATE: 2211300  6858   3303   51 MURMANSK             RUSSIAN FEDER 2003 2007    -999       0

CORRELATION STATISTICS (enter 'C' for more information):
> -0.60 is minimum correlation coeff.
>  0.65 is maximum correlation coeff.
> -0.01 is mean correlation coeff.

Enter 'Y' to allow, 'N' to deny, or an information code letter: C

<OKAY - SO I'VE REQUESTED A DISPLAY OF THE LAGGED CORRELATIONS>

Master Data: Correlation with Update first year aligned to this year -v
 1936  900  600 1000  800 1000  900 1300 1700 2100 1800  900 1000    0.27
 1937  300 1400 1300  800 1400 1800  500 1200 1600 1000 1100 1500    0.15
 1938  900 1000 1500 1800 1200 1500 1200 1700  500  700 1600  700   -0.13
 1939 1500 1300 1100 1400 1200 1200 1000 1300 1800 1600 1100 1300    0.24
 1940 1000 1500 1000 1200 1100 1700 2600 1500 1500 1400 1700 1100    0.15
 1941 1800 1200 1000 1200  900 1100  900 1200 1900 1500 1000 1400    0.48
 1942  900  900 1700  900 1600 1000  600 1100 1400 1300  700  700    0.51
 1943  800 1000 1000 1300  900  800 1500 1600 1400 1500 1300 1200    0.44
 1944 1000  400  900  800 1200  600  900 2000  900 1100 1000  900    0.32
 1945  500  400  700  700  800 1800  900 1100 1200 1100 1300  700    0.19
 1946 1200 1200  100  700  900 1200  400  900  800 1900 1300 1400    0.16
 1947  900 1300 1300 1100 1600 1000  800 1400 1400 1700 2100 1900    0.09
 1948 1100 1400 1400 1200 1300 1800 1200 1700 1500 2200 2100 1900    0.10
 1949 1100 1100  500 1500 1600 1100 1500 1200 2200 2500  900 1600    0.04
 1950 1300  800 1000 1100 1700 1200 1500  800 1100 1300 1500 1400   -0.04
 1951 1100  600 1400 1400 1500 1600 2100 1300 1500 1700 2000 1700   -0.13
 1952 2100  800 1100 1800 1300 1200 2400 2200 1600 1000 1000 2300   -0.23
 1953 2100 1400 2100 1500  900  300 1300 1700 1500  800 1200  800   -0.24
 1954 2100  600 1300 1000 1300 1700 1600 2000 1800 1300 1400 1200   -0.40
 1955 2200 1300  900 1000 1600 2000 1100 1400 1000 2100 2300 1600   -0.20
 1956 1300 1100 1300  400 1600 1300  900 1500 2000 1300 2000 1400   -0.30
 1957 1700 1600 1100 1100 1900 1900 1400 1600 1400 1700 2300 2600   -0.27
 1958 1300 2200 1900  700 1500 1200 2100 1000 1900 1700 1600 1000   -0.21
 1959 2500 1800 1300  900  900 1600 1600 1500 2200 1700 1000  900   -0.33
 1960 1800 1700 1500  400 1300 1500  400 1000 1300 1500 1000 1400   -0.21
 1961 2100 1800 2200 1500  800 1400 1600 1100 1900 1200 1200 2100   -0.59
 1962 2100 1100 1000 1500 1300 1100 1300 1700 1200 2000 1600 2300   -0.37
 1963 2100 2100 2000 1000  700 2000 1400 1800 1400 1600 2000 2400   -0.56
 1964 2400 1100 1000 1700 1100 1400 1400 1400 2000 1200 2100 1800   -0.42
 1965 1400 2100 1300 1000 1700 1700 1400 2400 1300 2100 1900 2100   -0.41
 1966 1600 1600 2000 2000 1700 1200 2000 2500 2500 2700 1600  600   -0.34
 1967 2200 1700 1600 1200 1000 1400 1600 1300 1700 1500 1200 2100   -0.21
 1968 1600 1800 1800 1800 1500 1800 1400 2100 1000 2000 2100 2000   -0.28
 1969 1100  300 1900 1200 1000 1300 1500 1200 1200 2000 1700  800   -0.25
 1970 1900 1400 1200  900  600 1200 1500  700 2300 1700 1700 2100   -0.23
 1971 2000 1300 1600 1600 1200 1100 1400 1800 2000 1600 1700 1500   -0.39
 1972 1300 1200 1300 1200 1700  800 1400 1800 1900 2000 1700 1600   -0.26
 1973 1800 1100 1700  900 1200 1500  500 1800 1200 2000 2100 2100   -0.36
 1974 1100 2400  700 1600 1300 1300 1800 2000 1900 1200 1400 2400   -0.29
 1975 1500 2200 1400 1700 2500 2200 2300 1600 1700 2300 1800 2600   -0.47
 1976 1900  800 1100 1500 1000  900 1300 1800 2200 1600 1400 1600   -0.33
 1977 1800 1400 2200 1200 1600 1900 1300 1500 1500 1900 1500 2000   -0.40
 1978 1500 1800 1400 2100  700 1000 1100 1900 1700 2300 1500 2200   -0.24
 1979 1700 1700 1700 1200 1500 1800  900 1200 1800 1600 1500 2300   -0.39
```

```
 1980 1900 1300 1300 1000 1400  900  700 1100 1300 1600 2200 1700   -0.36
 1981 2600  500 1900 2000  800 1900 1500 2000 1400 1500 1800 1600   -0.46
 1982 2200 1800 1100 1600 1500 2200 1800 1400 1700 1700 1900 1400   -0.60
 1983 2400 1900 1700 1200  800 1500 1200 2000 1400 2100 2000 2500   -0.23
 1984 1900  800 1500 2000 1100 1600 2000 1700 1100 1400 1000 1200
 1985-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
 1986-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
 1987-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
 1988-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
 1989-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999    0.65
 1990-9999-9999-9999-9999-9999  500 1300  900  700  900 1300  700    0.62
 1991-9999  900  500  300  700 1000 1500  700 1700 1000 1300 1300    0.54
 1992  800 1000  600  500  700  900-9999 1300-9999  700  900 1200    0.60
 1993  600  900  400  500  900 1500 1000  800  800 1000  400 1000    0.55
 1994 1300 1000  300  600  700 1000  900  600 1200    0 1400  600    0.43
 1995  900  900  600  700  700  900 1100 1300  600 1800 1300  500    0.61
 1996  500 1100  400  700  700 1200 1200 1100 1100  900 1000 1400    0.54
 1997 1200  800 1300  600  600  100  500 1100  900-9999 1000  900    0.61
 1998 1200 1300  800 1100 1100 1100  800  600 1200 1100  600 1200    0.52
 1999  600  400  600 1000  700  700 1800 1400  700 1600  800 1200    0.62
 2000 1100  600 1500 1700  900 1500  800  800 1000 1000  600  600    0.40
 2001  600  500  700  700  600  500 1200 1200  700 1300  900 1000    0.63
 2002 1000  800 1300  200  900 1100 1400 1200 1400 1800 1100  700
 2003 1100-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
Update Data:
 2003 1100  700  700  500 1000  400  700 1100 1200 2100  800 1900
 2004  900  700  600  600 1300 1200 1000 1200 1400  900 1000 1000
 2005 1000  400  800 1100  900  600 1200 1000 1600 1000 1300 1200
 2006  700  500 1300  400  600 1200 1600  700 1000-9999  600 1500
 2007 1400  400  400 1300 1200 1200-9999-9999-9999-9999-9999-9999

<DO YOU SEE? THERE'S THAT OH-SO FAMILIAR BLOCK OF MISSING CODES IN THE LATE 80S,
 THEN THE DATA PICKS UP AGAIN. BUT LOOK AT THE CORRELATIONS ON THE RIGHT, ALL
 GOOD AFTER THE BREAK, DECIDEDLY DODGY BEFORE IT. THESE ARE TWO DIFFERENT
 STATIONS, AREN'T THEY? AAAARRRGGGHHHHHHH!!!!!>

MASTER:  221130  6896   3305   51 MURMANSK             EX USSR       1936 2003   -999     -999
UPDATE: 2211300  6858   3303   51 MURMANSK             RUSSIAN FEDER 2003 2007    -999       0

CORRELATION STATISTICS (enter 'C' for more information):
> -0.60 is minimum correlation coeff.
>  0.65 is maximum correlation coeff.
> -0.01 is mean correlation coeff.

Enter 'Y' to allow, 'N' to deny, or an information code letter:
<END QUOTE>

So.. should I really go to town (again) and allow the Master database to be 'fixed' by this
program? Quite honestly I don't have time - but it just shows the state our data holdings
have drifted into. Who added those two series together? When? Why? Untraceable, except
anecdotally.

It's the same story for many other Russian stations, unfortunately - meaning that (probably)
there was a full Russian update that did no data integrity checking at all. I just hope it's
restricted to Russia!!

There are, of course, metadata issues too. Take:

<BEGIN QUOTE>
MASTER:  206740  7353   8040   47 DIKSON ISLAND        EX USSR       1936 2003   -999     -999
UPDATE: 2067400  7330   8024   47 OSTROV DIKSON        RUSSIAN FEDER 2003 2007    -999       0

CORRELATION STATISTICS (enter 'C' for more information):
> -0.70 is minimum correlation coeff.
>  0.81 is maximum correlation coeff.
> -0.01 is mean correlation coeff.
<END QUOTE>

This is pretty obviously the same station (well OK.. apart from the duff early period, but I've
got used to that now). But look at the longitude! That's probably 20km! LUckily I selected
'Update wins' and so the metadata aren't compared. This is still going to take ages, because although
I can match WMO codes (or should be able to), I must check that the data correlate adequately - and
for all these stations there will be questions. I don't think it would be a good idea to take the
usual approach of coding to avoid the situation, because (a) it will be non-trivial to code for, and
(b) not all of the situations are the same. But I am beginning to wish I could just blindly merge
based on WMO code.. the trouble is that then I'm continuing the approach that created these broken
databases. Look at this one:

<BEGIN QUOTE>
***** OPERATOR ADJUDICATION REQUIRED *****

In attempting to pair two stations, possible data incompatibilities have been found.

MASTER:  239330  6096   6906   40 HANTY MANSIJSK       EX USSR       1936 1984   -999     -999
UPDATE: 2393300  6101   6902   46 HANTY-MANSIJSK       RUSSIAN FEDER 2003 2007    -999       0

CORRELATION STATISTICS (enter 'C' for more information):
> -0.42 is minimum correlation coeff.
>  0.39 is maximum correlation coeff.
> -0.02 is mean correlation coeff.

Enter 'Y' to allow, 'N' to deny, or an information code letter: C
Master Data: Correlation with Update first year aligned to this year -v
 1936 1400  800 1700  900 1200  800  700  800 1800-9999-9999-9999    0.33
 1937 1400  800  500 1700 1500  800 1200 1000 1700 1300  700 1200    0.32
 1938 1000 1700 1200 1100 1100  800  800 1300 1400 1900 1800 1300    0.04
 1939 1100 1700 1600 1800 1500  800 1500 1900 1700 1800 1300 1300    0.09
 1940 1300  700  900  900 1800 1200  900 1300 1200 2200 1900 1800    0.08
 1941 1400 1100 1800 1000 1400 1900 1400  700 1300 1200 1900 2000    0.02
 1942 1700  900 1600  900 1200 1500 1300 1500 1200 1900 1500 1500   -0.06
 1943 1400 1300 1300  800 1400 1600 1300 1500 1900 2000  700 1900   -0.17
 1944 1900 1500 2000 1100 1200 1300 1500 1700 1800 1200 1500 1900   -0.32
 1945 1300 1000 1400 2100 2000 1100 1700  700 1600 1800 2300 1700   -0.42
 1946 2300 1900 1500 1100 1100 2000 1800 1000 1200 2100 2000 1800   -0.35
 1947 1900 1400 1600 1000 2100 1900 2100 1000 1200 2000 2100 1500   -0.35
 1948 1700 1500 1800  800 1300 1800 1700 1300 1800 2200 2000 2100   -0.15
 1949 2300 2100 1000  700 1600 1400 1200  800 2100 2000 1100 1400   -0.07
 1950 2100 2300 1000 1100 1500 1600 1600 2300 1900 1200 1100 1500    0.00
 1951 1600 1000 1500  800 1500 1400 1200  600 1800 1800 1400 2400   -0.07
 1952 1600  400 1100 1300 1100 1400  800 2000 1500 2300 1300 1600   -0.04
 1953 2000 1200 1500  500 1300 1500 1100 1200 2300 2200 1600 2100   -0.02
 1954 1700 1800  700  700 1000 1300 1200 1600 2000 1800 1800  600    0.01
 1955 2400 1400 1000 1100 1700 1200 1000 1300 1500 1300 2300 1600   -0.08
 1956 1300  800 1000 1100 1000 1000 1400 1800 1900 1900 2600 2000   -0.29
 1957 1900 1200 1700 1000 1100 1100 1100  700  800 2300 1900 2200   -0.18
 1958 1300 1600 1500  400 1500 1100 1300 1400 1900 2400 2000 1600   -0.28
 1959 1700 1600  700 1300 1700 1100 1100 1600 2000 2100 1900 1600   -0.04
 1960 1800 1600-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999    0.24
 1961-9999-9999-9999-9999-9999-9999-9999 1600 1600 1700 1900 1600    0.33
 1962 1700  800 1200  600  400 1100  900 2000 1100 1900 1700 1500    0.25
```

```
 1963 1200 1300 1700  700 1100 1600  900 1000 1100 1400 1800 2000   -0.04
 1964 1900  500 1300 1300 1200 1200 1100 1100 1700 1500 2000 1800    0.13
 1965 1200 1400  700  900 1200 1100 1300 1400 1800 2500 1000 1700    0.23
 1966 1800 1600 2100 1300 1500 2100  900 1800 1500 2400 1900  800    0.11
 1967 1600 1200 1100  600  800 1100 1100  700 1300 1200 1300 1900    0.39
 1968 1600 1400 1600 1200  900 1300 1400 1000 1700 1300 1400 1200    0.24
 1969  900 1000 1100 1500 1700 1700 1000 1800 1200 1400 1900 1300    0.04
 1970 1500 1200 1600 1400  700 1600  700 1600 1000 1500 1900 1600   -0.02
 1971 1700  400 1100 1700 1300 1700  700 2000  900 2100 2000 1900   -0.11
 1972 1200 1500 1400  800 1700 1300 1700 2000 2100 1700 2500 1900   -0.08
 1973 1200 1100 1100  700  800 1300 2100 1000 2400 1900 1800 2300   -0.11
 1974  700 1200 1800 1800 1400 1200 1000 1300 1100 1600 1900  700   -0.14
 1975 2200 1800 1400 1300 1500 1500 1400 1500 1400 2300 1900 2100   -0.15
 1976 2000 1500  600  700 1100 1600 1300 1100 1500 1800 1600 1200   -0.11
 1977 1900 1700 1800 1400 1000 1100 1000 1300 1500 1800 1700 2100   -0.15
 1978 1600 1000  800 1400 1400  800 1600 1600 2300 2200 2200 1800    0.03
 1979 1600 1600 1600  900  900 1900 1200 1700 1200 2100 1600 2000    0.00
 1980 1600 1200  500  800 1500 1100  800 1700 1200  600 2200 2200   -0.05
 1981 2000 1000 1700 1300 1500 1100  800  400 1500  800 1500 1900    0.06
 1982 2400 1800 1100 1200 1200 1100 1000 1700 1200 2100 1800 2000    0.03
 1983 2500 2100 1800 1300 1400 1200 1200 1300 1300 1900 2300 1900    0.10
 1984 1200  700  500 1300  900  800 1100 1000 1700 1600 1600 1300
Update Data:
 2003 1500  900  600  400  900 1200  500  700 1100  600  700 1500
 2004  700  600  700  400  600 1100  500  900  900 1400 1500  600
 2005  700  400  800 1400  300  900  800  800  900  500 1200  600
 2006  800  700  900 1000  800  500 1000  500 1300 1100  700 1600
 2007 1100 1100  900  700 1300 1500-9999-9999-9999-9999-9999-9999
<END QUOTE>

Here, the expected 1990-2003 period is MISSING - so the correlations aren't so hot! Yet
the WMO codes and station names /locations are identical (or close). What the hell is
supposed to happen here? Oh yeah - there is no 'supposed', I can make it up. So I have :-)

If an update station matches a 'master' station by WMO code, but the data is unpalatably
inconsistent, the operator is given three choices:

<BEGIN QUOTE>
You have failed a match despite the WMO codes matching.
This must be resolved!! Please choose one:

1. Match them after all.
2. Leave the existing station alone, and discard the update.
3. Give existing station a false code, and make the update the new WMO station.

Enter 1,2 or 3:
<END QUOTE>

You can't imagine what this has cost me - to actually allow the operator to assign false
WMO codes!! But what else is there in such situations? Especially when dealing with a 'Master'
database of dubious provenance (which, er, they all are and always will be).

False codes will be obtained by multiplying the legitimate code (5 digits) by 100, then adding
1 at a time until a number is found with no matches in the database. THIS IS NOT PERFECT but as
there is no central repository for WMO codes - especially made-up ones - we'll have to chance
duplicating one that's present in one of the other databases. In any case, anyone comparing WMO
codes between databases - something I've studiously avoided doing except for tmin/tmax where I
had to - will be treating the false codes with suspicion anyway. Hopefully.

Of course, option 3 cannot be offered for CLIMAT bulletins, there being no metadata with which
to form a new station.

This still meant an awful lot of encounters with naughty Master stations, when really I suspect
nobody else gives a hoot about. So with a somewhat cynical shrug, I added the nuclear option -
to match every WMO possible, and turn the rest into new stations (er, CLIMAT excepted). In other
words, what CRU usually do. It will allow bad databases to pass unnoticed, and good databases to
become bad, but I really don't think people care enough to fix 'em, and it's the main reason the
project is nearly a year late.

And there are STILL WMO code problems!!! Let's try again with the issue. Let's look at the first
station in most of the databases, JAN MAYEN. Here it is in various recent databases:


dtr.0705152339.dtb: 100100  7093   -867    9 JAN MAYEN            NORWAY        1998 2006   -999  -999.00
pre.0709111032.dtb:0100100  7056   -840    9 JAN MAYEN            NORWAY        2003 2007    -999        0
sun.0709111032.dtb:0100100  7056   -840    9 JAN MAYEN            NORWAY        2003 2007    -999        0
tmn.0702091139.dtb: 100100  7093   -867    9 JAN MAYEN            NORWAY        1998 2006   -999  -999.00
tmn.0705152339.dtb: 100100  7093   -867    9 JAN MAYEN            NORWAY        1998 2006   -999  -999.00
tmp.0709111032.dtb:0100100  7056   -840    9 JAN MAYEN            NORWAY        2003 2007    -999        0
tmx.0702091313.dtb: 100100  7093   -867    9 JAN MAYEN            NORWAY        1998 2006   -999  -999.00
tmx.0705152339.dtb: 100100  7093   -867    9 JAN MAYEN            NORWAY        1998 2006   -999  -999.00
vap.0709111032.dtb:0100100  7056   -840    9 JAN MAYEN            NORWAY        2003 2007    -999        0

As we can see, even I'm cocking it up! Though recoverably. DTR, TMN and TMX need to be written as (i7.7).

Anyway, here it is in the problem database:

wet.0311061611.dtb:  10010  7093   -866    9 JAN MAYEN(NOR NAVY)  NORWAY        1990 2003   -999     -999

You see? The leading zero's been lost (presumably through writing as i7) and then a zero has been added at
the trailing end. So it's a 5-digi WMO code BUT NOT THE RIGHT ONE. Aaaarrrgghhhhhh!!!!!!

I think this can only be fixed in one of two ways:

1. By hand.

2. By automatic comparison with other (more reliable) databases.

As usual - I'm going with 2. Hold onto your hats.

Actually, a brief interlude to churn out the tmin & tmax primaries, which got sort-of
forgotten after dtr was done:

<BEGIN ABRIDGED QUOTES (separated by '#####')>
   > ***** AnomDTB: converts .dtb to anom .txt for gridding *****
   > Enter the suffix of the variable required:
.tmn
   > Select the .cts or .dtb file to load:
tmn.0708071548.dtb
   > Specify the start,end of the normals period:
1961,1990
   > Specify the missing percentage permitted:
25
   > Data required for a normal:           23
   > Specify the no. of stdevs at which to reject data:
3
   > Select outputs (1=.cts,2=.ann,3=.txt,4=.stn):
3
   > Check for duplicate stns after anomalising? (0=no,>0=km range)
0
```

```
   > Select the generic .txt file to save (yy.mm=auto):
tmn.txt
   > Select the first,last years AD to save:
1901,2006
   > Operating...
   > NORMALS             MEAN percent      STDEV percent
   >         .dtb    3814210    65.5
   >         .cts     210801     3.6    4025011    69.2
   > PROCESS        DECISION percent %of-chk
   > no lat/lon          650     0.0     0.0
   > no normal       1793923    30.8    30.8
   > out-of-range        976     0.0     0.0
   > accepted        4024035    69.1
   > Dumping years 1901-2006 to .txt files...
#####
IDL> quick_interp_tdm2,1901,2006,'tmnglo/tmn.',750,gs=0.5,pts_prefix='tmntxt/tmn.',dumpglo='dumpglo'
#####
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: gunzip clim.6190.lan.tmn
FILE NOT FOUND - PLEASE TRY AGAIN: clim.6190.lan.tmn
Enter a name for the gridded climatology file: clim.6190.lan.tmn.grid
Enter the path and stem of the .glo files: tmnglo/tmn.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: tmnabs
Now, CONCENTRATE. Addition or Percentage (A/P)? A
Right, erm.. off I jolly well go!
tmn.01.1901.glo
(etc)
tmn.12.2006.glo
#####
Welcome! This is the MERGEGRIDS program.
I will create decadal and full gridded files
from the output files of (eg) glo2abs.for.
Enter a gridfile with YYYY for year and MM for month: tmnabs/tmn.MM.YYYY.glo.abs
Enter Start Year:  1901
Enter Start Month: 01
Enter End Year:    2006
Enter End Month:   12
Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE: cru_ts_3_00.SSSS.EEEE.tmn.dat
Writing cru_ts_3_00.1901.1910.tmn.dat
(etc)
#####
   > ***** AnomDTB: converts .dtb to anom .txt for gridding *****
   > Enter the suffix of the variable required:
.tmx
   > Select the .cts or .dtb file to load:
tmx.0708071548.dtb
   > Specify the start,end of the normals period:
1961,1990
   > Specify the missing percentage permitted:
25
   > Data required for a normal:           23
   > Specify the no. of stdevs at which to reject data:
3
   > Select outputs (1=.cts,2=.ann,3=.txt,4=.stn):
3
   > Check for duplicate stns after anomalising? (0=no,>0=km range)
0
   > Select the generic .txt file to save (yy.mm=auto):
tmx.txt
   > Select the first,last years AD to save:
1901,2006
   > Operating...
   > NORMALS             MEAN percent      STDEV percent
   >         .dtb    3795470    65.4
   >         .cts     205607     3.5    4001077    68.9
   > PROCESS        DECISION percent %of-chk
   > no lat/lon          652     0.0     0.0
   > no normal       1805313    31.1    31.1
   > out-of-range        471     0.0     0.0
   > accepted        4000606    68.9
   > Dumping years 1901-2006 to .txt files...
#####
IDL> quick_interp_tdm2,1901,2006,'tmxglo/tmx.',750,gs=0.5,pts_prefix='tmxtxt/tmx.',dumpglo='dumpglo'
#####
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.tmx
Enter a name for the gridded climatology file: clim.6190.lan.tmx.grid
Enter the path and stem of the .glo files: tmxglo/tmx.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: tmxabs
Now, CONCENTRATE. Addition or Percentage (A/P)? A
Right, erm.. off I jolly well go!
tmx.01.1901.glo
(etc)
tmx.12.2006.glo
#####
Welcome! This is the MERGEGRIDS program.
I will create decadal and full gridded files
from the output files of (eg) glo2abs.for.
Enter a gridfile with YYYY for year and MM for month: tmxabs/tmx.MM.YYYY.glo.abs
Enter Start Year:  1901
Enter Start Month: 01
Enter End Year:    2006
Enter End Month:   12
Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE: cru_ts_3_00.SSSS.EEEE.tmx.dat
Writing cru_ts_3_00.1901.1910.tmx.dat
(etc)
<END ABRIDGED QUOTES>

This took longer than hoped.. running out of disk space again. This is why Tim didn't save more of
the intermediate products - which would have made my detective work easier. The ridiculous process
he adopted - and which we have dutifully followed - creates hundreds of intermediate files at every
stage, none of which are automatically zipped/unzipped. Crazy. I've filled a 100gb disk!

So, anyway, back on Earth I wrote wmocmp.for, a program to - you guessed it - compare WMO codes from
a given set of databases.  Results were, ah.. 'interesting':

<BEGIN QUOTE>
REPORT:
```

```
Database Title                Exact Match  Close Match  Vague Match  Awful Match  Codes Added      WMO = 0
../db/pre/pre.0612181221.dtb          n/a          n/a          n/a          n/a        14397         1540
../db/dtr/tmn.0708071548.dtb         1865         3389           57           77         5747         2519
../db/tmp/tmp.0705101334.dtb            0            4           28          106         4927            0
<END QUOTE>

So the largest database, precip, contained 14397 stations with usable WMO codes (and 1540 without).
The TMin, (and TMax and DTR, which were tested then excluded as they matched TMin 100%) database only agreed
perfectly with precip for 1865 stations, nearby 3389, believable 57, worrying 77. TMean fared worse, with NO
exact matches (WMO misformatting again) and over 100 worrying ones.

The big story is the need to fix the tmean WMO codes. For instance:

  10010   709    -87   10 Jan Mayen            NORWAY        1921 2006 341921  -999.00

is illegal, and needs to become one of:
  01001   709    -87   10 Jan Mayen            NORWAY        1921 2006 341921  -999.00
0001001   709    -87   10 Jan Mayen            NORWAY        1921 2006 341921  -999.00
0100100   709    -87   10 Jan Mayen            NORWAY        1921 2006 341921  -999.00

I favour the first as it's technically accurate. Alternatively we seem to have widely adopted the third, which
at least has the virtue of being consistent. Of course it's the only one that will match the precip:

 100100  7093   -867   10 JAN MAYEN            NORWAY        1921 2006   -999  -999.00

..which itself should be either:

0100100  7093   -867   10 JAN MAYEN            NORWAY        1921 2006   -999  -999.00

or:

  01001  7093   -867   10 JAN MAYEN            NORWAY        1921 2006   -999  -999.00

Aaaaarrrggghhhh!!!!

And the reason this is so important is that the incoming updates will rely PRIMARILY on matching the WMO codes!
In fact CLIMAT bulletins carry no other identification, of course. Clearly I am going to need a reference set
of 'qenuine WMO codes'.. and wouldn't you know it, I've found four!

Location                                              N. Stations      Notes
http://weather.noaa.gov/data/nsd_bbsss.txt            11548            Full country names, ';' delim
http://www.htw-dresden.de/~kleist/wx_stations_ct.html 13000+           *10, leading zeros kept, fmt probs
From Dave Lister                                      13080            *10 and leading zeros lost, country codes
From Philip Brohan                                    11894            2+3, No countries

The strategy is to use Dave Lister's list, grabbing country names from the Dresden list. Wrote
getcountrycodes.for and extracted an imperfect but useful-as-a-reference list. Hopefully in the main the country
will not need fixing or referring to!!

Wrote 'fixwmos.for' - probably not for the first time, but it's the first prog of that name in my repository so I'll
have to hope for the best. After an unreasonable amount of teething troubles (due to my forgetting that the tmp
database stores lats & lons in degs*100 not degs*10, and also to the presence of a '-99999' as the lon for GUATEMALA
in the reference set) I managed to sort-of fix the tmp database:

<BEGIN QUOTE>
crua6[/cru/cruts/version_3_0/db/tmp] ./fixwmos

FIXWMOS - Fix WMO Codes in a Database

Enter the database to be fixed: tmp.0705101334.dtb

The operation completed successfully.

 2263 WMO Codes were 'fixed' and all were rewritten as (i7.7)

The output database is tmp.0709281456.dtb

crua6[/cru/cruts/version_3_0/db/tmp]
<END QUOTE>

The first records have changed as follows:

crua6[/cru/cruts/version_3_0/db/tmp] diff tmp.0705101334.dtb tmp.0709281456.dtb |head -30
1c1
<   10010   709    -87   10 Jan Mayen            NORWAY        1921 2006 341921  -999.00
---
> 0100100   709    -87   10 Jan Mayen            NORWAY        1921 2006 341921  -999.00

So far so good.. but records that weren't matched with the reference set didn't fare so well:

89c89
<   10050   780    142    9 ISFJORD RADIO        NORWAY        1912 1979 101912  -999.00
---
> 0010050   780    142    9 ISFJORD RADIO        NORWAY        1912 1979 101912  -999.00

This is misleading because, although there probably won't BE any incoming updates for ISFJORD RADIO, we can't say for
certain that there will never be updates for any station outside the current reference set. In fact, we can say with
confidence that there will be!

So, what to do? Do we assume a particular factor to adjust ALL codes by, based on the matches? Or do we attempt (note
careful use of verb) to use the country codes database to work out the most significant 'real' digits of these codes?

Well, I fancy the first one. We'll make two passes through the data, the first pass changes nothing but saves counts of
the successful factors in bins: *0.01, *0.1, *1, *10, *100 should do it. I sure hope all the results are in one bin!

It worked. An initial 'verbose' run showed a consistent choice of factor, though it'll exit with an error code if multiple
factors are registered in one database.

<BEGIN QUOTE>
crua6[/cru/cruts/version_3_0/db/tmp] ./fixwmos

FIXWMOS - Fix WMO Codes in a Database

Enter the database to be fixed: tmp.0705101334.dtb
locfac set to: 10
First ref: 0100100

The operation completed successfully.

 2263 WMO Codes were 'matched'
All codes were modified with a factor of  10
Lons/lats were modified with a factor of  10

The output database is tmp.0710011359.dtb

crua6[/cru/cruts/version_3_0/db/tmp]
<END QUOTE>

Example results:
```

```
<BEGIN QUOTE>
crua6[/cru/cruts/version_3_0/db/tmp] diff tmp.0705101334.dtb tmp.0710011359.dtb | head -12
1c1
<   10010   709    -87   10 Jan Mayen            NORWAY        1921 2006 341921  -999.00
---
> 0100100  7090   -870   10 Jan Mayen            NORWAY        1921 2006 341921  -999.00
89c89
<   10050   780    142    9 ISFJORD RADIO        NORWAY        1912 1979 101912  -999.00
---
> 0100500  7800   1420    9 ISFJORD RADIO        NORWAY        1912 1979 101912  -999.00
159c159
<   10080   783    155   28 Svalbard Lufthavn    NORWAY        1911 2006 341911  -999.00
---
> 0100800  7830   1550   28 Svalbard Lufthavn    NORWAY        1911 2006 341911  -999.00
<END QUOTE>

Then.. attacked the wet database! And immediately found this beauty:

      0 -9999 -99999 -999 UNKNOWN              UNKNOWN       1994 2003   -999        0
6190-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1994  500  800  600  400  600  100    0  100  200  400 1000 1300
1995  400  100 1100  900 1200  800  200  100  200  400  800  500
1996  500 1100 1500  600  900-9999    0  300  400  700    0 1100
1997  800 1000  700 1000 1000 1000  200  200  400  700  200 1000
1998  700  700 1000 1000-9999  800  100  100    0  200  400  700
1999  300 1000  800-9999  700  800    0  200-9999  600  400  200
2000 1100  600  900  900 1000  400-9999  100  200  300    0  400
2001    0  800  300  500 1200    0    0    0  200  200  500  800
2002  800  300  600 1300  800  500  400  100  300  400  400  600
2003  300-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999

Gotta love the system! Like this is ever going to be a blind bit of use. Modified the code to
leave such stations unmolested, but identified in a separate file so they can be 'cleansed', it
being a little too risky to auto-cleanse such things.

Hopefully the final attack on 'wet':

<BEGIN QUOTE>
crua6[/cru/cruts/version_3_0/db/rd0] ./fixwmos

FIXWMOS - Fix WMO Codes in a Database

Enter the database to be fixed: wet.0311061611.dtb

The operation completed successfully.

 1920 WMO Codes were 'matched'
All codes were modified with a factor of  10
Lons/lats were modified with a factor of   1

The output database is wet.0710021341.dtb


IMPORTANT: the following WMO codes were not altered:
False codes (wmo<0):          2917
Illegal codes (0<=wmo<1000):     1
  (illegals written to wet.0311061611.bad)
crua6[/cru/cruts/version_3_0/db/rd0]
<END QUOTE>

I then removed the sole illegal (see above) from wet.0710021341.dtb, which becomes the 'new old'
wet/rd0 database.

So.. to incorporate the updates! Finally. First, the MCDW, metadata-rich ones:

<BEGIN QUOTE>
uealogin1[/cru/cruts/version_3_0/db/rd0] ./newmergedb

WELCOME TO THE DATABASE UPDATER

Before we get started, an important question:
If you are merging an update - CLIMAT, MCDW,
ian - do
you want the quick and dirty approach? This will blindly match
on WMO codes alone, ignoring data/metadata checks, and making any
unmatched updates into new stations (metadata permitting)?

Enter 'B' for blind merging, or <ret>: B
Please enter the Master Database name: wet.0710021341.dtb
Please enter the Update Database name: rdy.0709111032.dtb

Reading in both databases..
Master database stations:     4987
Update database stations:     2407

Looking for WMO code matches..
* new header 0100100  7056   -840    9 JAN MAYEN            NORWAY        1990 2007    -999    -999 *
    2 reject(s) from update process 0710041559

Writing wet.0710041559.dtb

OUTPUT(S) WRITTEN

New master database: wet.0710041559.dtb

Update database stations:         2407
 > Matched with Master stations:  1556
                 (automatically:  1556)
                   (by operator:     0)
 > Added as new Master stations:     0
 > Rejected:                         2
   Rejects file:                 rdy.0709111032.dtb.rejected
 Note: IEEE floating-point exception flags raised:
    Inexact;  Invalid Operation;
 See the Numerical Computation Guide, ieee_flags(3M)
uealogin1[/cru/cruts/version_3_0/db/rd0]
<END QUOTE>

(also knocked up rrstats.for at this stage, to analyse replication rates by
 latitude band for a given database - needs a Matlab prog to drive really)

[a bit of debugging here as the last records weren't being written properly,
 filenames adjusted above accordingly]


Then, the CLIMAT, nothing-but-the-code ones:

*WARNING: ignore this, the CLIMAT bulletins were later improved with metadata and newmergedb rerun*

<BEGIN QUOTE>
```

```
uealogin1[/cru/cruts/version_3_0/db/rd0] ./newmergedb

WELCOME TO THE DATABASE UPDATER

Before we get started, an important question:
If you are merging an update - CLIMAT, MCDW, Australian - do
you want the quick and dirty approach? This will blindly match
on WMO codes alone, ignoring data/metadata checks, and making any
unmatched updates into new stations (metadata permitting)?

Enter 'B' for blind merging, or <ret>: B
Please enter the Master Database name: wet.0710041559.dtb
Please enter the Update Database name: rdy.0709101706.dtb

Reading in both databases..
Master database stations:     5836
Update database stations:     2876

Looking for WMO code matches..
  378 reject(s) from update process 0710081508

Writing wet.0710081508.dtb

OUTPUT(S) WRITTEN

New master database: wet.0710081508.dtb

Update database stations:         2876
 > Matched with Master stations:  2498
                 (automatically:  2498)
                   (by operator:     0)
 > Added as new Master stations:     0
 > Rejected:                       378
   Rejects file:                rdy.0709101706.dtb.rejected
 Note: IEEE floating-point exception flags raised:
    Inexact;  Invalid Operation;
 See the Numerical Computation Guide, ieee_flags(3M)
uealogin1[/cru/cruts/version_3_0/db/rd0]
<END QUOTE>

Now of course, we can't add any of the CLIMAT bulletin stations as 'new' stations
because we don't have any metadata! so.. is it worth using the lookup table? Because
although I'm thrilled at the high match rate (87%!), it does seem worse when you
realise that you lost the rest..

* see below, CLIMAT metadata fixed! *

At this stage I knocked up rrstats.for and the visualisation companion tool, cmprr.m. A simple process
to show station counts against time for each 10-degree latitude band (with 20-degree bands at the
North and South extremities). A bit basic and needs more work - but good for a quick & dirty check.

Wrote dllist2headers.for to convert the 'Dave Lister' WMO list to CRU header format - the main difficulty
being the accurate conversion of the two-character 'country codes' - especially since many are actually
state codes for the US! Ended up with wmo.0710151633.dat as our reference WMO set.

Incorporated the reference WMO set into climat2cru.for. Successfully reprocessed the CLIMAT bulletins
into databases with at least SOME metadata:

pre.0710151817.dtb
rdy.0710151817.dtb
sun.0710151817.dtb
tmn.0710151817.dtb
tmp.0710151817.dtb
tmx.0710151817.dtb
vap.0710151817.dtb

In fact, it was far more successful than I expected - only 11 stations out of 2878 without metadata!

Re-ran newmergedb:

<BEGIN QUOTE>
uealogin1[/cru/cruts/version_3_0/db/rd0] ./newmergedb

WELCOME TO THE DATABASE UPDATER

Before we get started, an important question:
If you are merging an update - CLIMAT, MCDW, Australian - do
you want the quick and dirty approach? This will blindly match
on WMO codes alone, ignoring data/metadata checks, and making any
unmatched updates into new stations (metadata permitting)?

Enter 'B' for blind merging, or <ret>: B
Please enter the Master Database name: wet.0710041559.dtb
Please enter the Update Database name: rdy.0710151817.dtb

Reading in both databases..
Master database stations:     5836
Update database stations:     2876

Looking for WMO code matches..
   71 reject(s) from update process 0710161148

Writing wet.0710161148.dtb

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

OUTPUT(S) WRITTEN

New master database: wet.0710161148.dtb

Update database stations:         2876
 > Matched with Master stations:  2498
                 (automatically:  2498)
                   (by operator:     0)
 > Added as new Master stations:   307
 > Rejected:                        71
   Rejects file:                rdy.0710151817.dtb.rejected
 Note: IEEE floating-point exception flags raised:
    Inexact;  Invalid Operation;
 See the Numerical Computation Guide, ieee_flags(3M)
uealogin1[/cru/cruts/version_3_0/db/rd0]
<END QUOTE>

307 stations rescued! and they'll be there in future of course, for metadata-free CLIMAT bulletins
to match with.

So where were we.. Rain Days. Family tree:

wet.0311061611.dtb
```

```
        +
rdy.07091111032.dtb  (MCDW composite)
        +
rdy.0710151817.dtb  (CLIMAT composite with metadata added)
        V
        V
wet.0710161148.dtb
```

Now it gets tough. The current model for a secondary is that it is derived from one or more primaries, plus their normals, plus the normals for the secondary.

The IDL secondary generators do not allow 'genuine' secondary data to be incorporated. This would have been ideal, as the gradual increase in observations would have gradually taken precedence over the primary-derived synthetics.

The current stats for the wet database were derived from the new proglet, dtbstats.for:

<BEGIN QUOTE>

```
crua6[/cru/cruts/version_3_0/secondaries/rd0] ./dtbstat

DTBSTAT: Database Stats Report

Please enter the (18ch.) database name: wet.0710161148.dtb

Report for: wet.0710161148.dtb

Stations in Northern Hemisphere:     5365
Stations in Southern Hemisphere:      778
                          Total:     6143

Maximum Timespan in Northern Hemisphere: 1840 to 2007
Maximum Timespan in Southern Hemisphere: 1943 to 2007
                        Global Timespan: 1840 to 2007

crua6[/cru/cruts/version_3_0/secondaries/rd0]
```

<END QUOTE>

So, without further ado, I treated RD0 as a Primary and derived gridded output from the database:

<BEGIN QUOTE>

```
crua6[/cru/cruts/version_3_0/secondaries/rd0] ./anomdtb

   > ***** AnomDTB: converts .dtb to anom .txt for gridding *****

   > Enter the suffix of the variable required:
.rd0
   > Select the .cts or .dtb file to load:
wet.0710161148.dtb
   > Specify the start,end of the normals period:
1961,1990
   > Specify the missing percentage permitted:
25
   > Data required for a normal:           23
   > Specify the no. of stdevs at which to reject data:
3
   > Select outputs (1=.cts,2=.ann,3=.txt,4=.stn):
3
   > Check for duplicate stns after anomalising? (0=no,>0=km range)
0
   > Select the generic .txt file to save (yy.mm=auto):
rd0.txt
   > Select the first,last years AD to save:
1901,2007
   > Operating...

   > NORMALS            MEAN percent      STDEV percent
   >         .dtb          0     0.0
   >         .cts     731118    45.4     730956    45.4
   > PROCESS        DECISION percent %of-chk
   > no lat/lon            0     0.0     0.0
   > no normal        878015    54.6    54.6
   > out-of-range         56     0.0     0.0
   > accepted         731062    45.4
   > Dumping years 1901-2007 to .txt files...

crua6[/cru/cruts/version_3_0/secondaries/rd0]
```

<END QUOTE>

Not particularly good - the bulk of the data being recent, less than half had valid normals (anomdtb calculates normals on the fly, on a per-month basis). However, this isn't so much of a problem as the plan is to screen it for valid station contributions anyway.

<BEGIN QUOTE>

```
IDL> quick_interp_tdm2,1901,2007,'rd0glo/rd0.',450,gs=0.5,dumpglo='dumpglo',pts_prefix='rd0txt/rd0.'
% Compiled module: QUICK_INTERP_TDM2.
% Compiled module: GLIMIT.
Defaults set
    1901
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: STRIP.
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
    1902
(etc)
    2007
no stations found in: rd0txt/rd0.2007.08.txt
no stations found in: rd0txt/rd0.2007.09.txt
no stations found in: rd0txt/rd0.2007.10.txt
no stations found in: rd0txt/rd0.2007.11.txt
no stations found in: rd0txt/rd0.2007.12.txt
IDL>
```

<END QUOTE>

<BEGIN QUOTE>

```
crua6[/cru/cruts/version_3_0/secondaries/rd0] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.wet
Enter a name for the gridded climatology file: clim.6190.lan.wet.grid2
Enter the path and stem of the .glo files: rd0glo/rd0.
Enter the starting year: 1901
Enter the ending year:   2007
Enter the path (if any) for the output files: rd0abs/
Now, CONCENTRATE. Addition or Percentage (A/P)? A         ! this was a guess! We'll see how the results look
Right, erm.. off I jolly well go!
rd0.01.1901.glo
```

```
(etc)
<END QUOTE>

Then.. wait a minute! I checked back, and sure enough, quick_interp_tdm.pro DOES allow both synthetic and 'real' data
to be included in the gridding. From the program description:

<BEGIN QUOTE>
; TDM: the dummy grid points default to zero, but if the synth_prefix files are present in call,
;  the synthetic data from these grids are read in and used instead
<END QUOTE>

And so.. (after some confusion, and renaming so that anomdtb selects percentage anomalies)..

IDL> quick_interp_tdm2,1901,2006,'rd0pcglo/rd0pc',450,gs=0.5,dumpglo='dumpglo',synth_prefix='rd0syn/rd0syn',pts_prefix='rd0pctxt/rd0pc.'

The trouble is, we won't be able to produce reliable station count files this way. Or can we use the same strategy,
producing station counts from the wet database route, and filling in 'gaps' with the precip station counts? Err.

<BEGIN QUOTE>
crua6[/cru/cruts/version_3_0/secondaries/rd0] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.wet
Enter a name for the gridded climatology file: clim.grid
Enter the path and stem of the .glo files: rd0pcglo/rd0pc.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: rd0pcgloabs/
Now, CONCENTRATE. Addition or Percentage (A/P)? P
Right, erm.. off I jolly well go!
rd0pc.01.1901.glo
(etc)
<END QUOTE>

<BEGIN QUOTE>
crua6[/cru/cruts/version_3_0/secondaries/rd0] ./mergegrids
Welcome! This is the MERGEGRIDS program.
I will create decadal and full gridded files
from the output files of (eg) glo2abs.for.

Enter a gridfile with YYYY for year and MM for month: rd0pcgloabs/rd0pc.MM.YYYY.glo.abs
Enter Start Year:  1901
Enter Start Month: 01
Enter End Year:    2006
Enter End Month:   12

Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE: cru_ts_3_00.SSSS.EEEE.rd0.dat
Writing cru_ts_3_00.1901.1910.rd0.dat
Writing cru_ts_3_00.1911.1920.rd0.dat
Writing cru_ts_3_00.1921.1930.rd0.dat
Writing cru_ts_3_00.1931.1940.rd0.dat
Writing cru_ts_3_00.1941.1950.rd0.dat
Writing cru_ts_3_00.1951.1960.rd0.dat
Writing cru_ts_3_00.1961.1970.rd0.dat
Writing cru_ts_3_00.1971.1980.rd0.dat
Writing cru_ts_3_00.1981.1990.rd0.dat
Writing cru_ts_3_00.1991.2000.rd0.dat
Writing cru_ts_3_00.2001.2006.rd0.dat
crua6[/cru/cruts/version_3_0/secondaries/rd0]
<END QUOTE>

All according to plan.. except the values themselves!

For January, 2001:

Minimum      =      0
Maximum      =  32630
Vals >31000  =      1

For the whole of 2001:

Minimum      =      0
Maximum      =  56763
Vals >31000  =      5

Not good. We're out by a factor of at least 10, though the extremes are few enough to just cap at DiM. So where has
this factor come from?

Well here's the January 2001 climatology:

Minimum      =      0
Maximum      =   3050
Vals >3100   =      0

That all seems fine for a percentage normals set. Not entirly sure about 0 though.

so let's look at the January 2001 gridded anomalies file:

Minimum      =    -48.046
Maximum      =      0.0129

This leads to a show-stopper, I'm afraid. It looks as though the calculation I'm using for percentage anomalies is,
not to put too fine a point on it, cobblers.

This is what I use to build actuals from anomalies in glo2abs.for:

              absgrid(ilon(i),ilat(i)) = nint(normals(i,imo) +
     *                   anoms(ilon(i),ilat(i)) * normals(i,imo) / 100)

or, to put it another way, V = N(A+N)/100

This is what anomdtb.f90 uses to build anomalies from actuals:

                DataA(XAYear,XMonth,XAStn) = nint(1000.0*((real(DataA(XAYear,XMonth,XAStn)) / &
                                             real(NormMean(XMonth,XAStn)))-1.0))

or, in the same terms, A = 1000((V/N)-1)

which reverses to: V = N(A+1000)/1000

This could well explain things. It could also mean that I have to reproduce v3.00 precip AFTER it's been used (against
my wishes) by Dave L and Dimitrious.

Well to start with, I'll try the new calculation in glo2abs to reproduce the rd0 data.

<BEGIN QUOTE>
crua6[/cru/cruts/version_3_0/secondaries/rd0] ./glo2abs
```

```
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.wet
Enter a name for the gridded climatology file: c.grid
Enter the path and stem of the .glo files: rd0pcglo/rd0pc.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: rd0pcgloabs
Now, CONCENTRATE. Addition or Percentage (A/P)? P
Right, erm.. off I jolly well go!
rd0pc.01.1901.glo
(etc)
<END QUOTE>

This *does* improve matters considerably. Now, for January 2001:

Minimum      =      0
Maximum      =   5090  (a little high but not fatal)
Vals >3100   =    556
Vals >3500   =    110
Vals >4000   =      2  (so the bulk of the excessions are only a few days over)

In fact the 2nd highest Max is 4369, well below 5090.

So, good news - but only in the sense that I've found the error. Bad news in that it's a further confirmation that my
abilities are short of what's required here.

Rushed back to precip. Found the .glo files in /cru/cruts/version_3_0/primaries/precip/pre0km0612181221glo/, and
re-ran glo2abs with the revised percentage anomaly equation:

<BEGIN QUOTE>
crua6[/cru/cruts/version_3_0/primaries/precip] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.pre
Enter a name for the gridded climatology file: clim.6190.lan.pre.gridded2
Enter the path and stem of the .glo files: pre0km0612181221glo/pregrid.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: pre0km0612181221abs/
Now, CONCENTRATE. Addition or Percentage (A/P)? P
Right, erm.. off I jolly well go!
pregrid.01.1901.glo
(etc)
<END QUOTE>

<BEGIN QUOTE>
crua6[/cru/cruts/version_3_0/primaries/precip] ./mergegrids
Welcome! This is the MERGEGRIDS program.
I will create decadal and full gridded files
from the output files of (eg) glo2abs.for.

Enter a gridfile with YYYY for year and MM for month: pre0km0612181221abs/pregrid.MM.YYYY.glo.abs
Enter Start Year:  1901
Enter Start Month: 01
Enter End Year:    2006
Enter End Month:   12

Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE: cru_ts_3_00.SSSS.EEEE.pre.dat
Writing cru_ts_3_00.1901.1910.pre.dat
Writing cru_ts_3_00.1911.1920.pre.dat
Writing cru_ts_3_00.1921.1930.pre.dat
Writing cru_ts_3_00.1931.1940.pre.dat
Writing cru_ts_3_00.1941.1950.pre.dat
Writing cru_ts_3_00.1951.1960.pre.dat
Writing cru_ts_3_00.1961.1970.pre.dat
Writing cru_ts_3_00.1971.1980.pre.dat
Writing cru_ts_3_00.1981.1990.pre.dat
Writing cru_ts_3_00.1991.2000.pre.dat
Writing cru_ts_3_00.2001.2006.pre.dat
crua6[/cru/cruts/version_3_0/primaries/precip]
<END QUOTE>

Then back to finish off rd0. Modified glo2abs to allow the operator to set minima and maxima, with a
specific option to set wet day limits (DiM*100):

<BEGIN QUOTE>
crua6[/cru/cruts/version_3_0/secondaries/rd0] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.wet
Enter a name for the gridded climatology file: clim...grid
Enter the path and stem of the .glo files: rd0pcglo/rd0pc.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: rd0pcgloabs/
Now, CONCENTRATE. Addition or Percentage (A/P)? P
Do you wish to limit the output values? (Y/N): Y
1. Set minimum to zero
2. Set a single minimum and maximum
3. Set monthly minima and maxima (for wet/rd0)

Choose: 3
Right, erm.. off I jolly well go!
rd0pc.01.1901.glo
(etc)
<END QUOTE>

Output was checked.. and as expected, January 2001 had 556 values of 3100 :-)

<BEGIN QUOTE>
crua6[/cru/cruts/version_3_0/secondaries/rd0] ./mergegrids
Welcome! This is the MERGEGRIDS program.
I will create decadal and full gridded files
from the output files of (eg) glo2abs.for.

Enter a gridfile with YYYY for year and MM for month: rd0pcgloabs/rd0pc.MM.YYYY.glo.abs
Enter Start Year:  1901
Enter Start Month: 01
Enter End Year:    2006
Enter End Month:   12

Please enter a sample OUTPUT filename, replacing
```

```
start year with SSSS and end year with EEEE: cru_ts_3_00.SSSS.EEEE.rd0.dat
Writing cru_ts_3_00.1901.1910.rd0.dat
Writing cru_ts_3_00.1911.1920.rd0.dat
Writing cru_ts_3_00.1921.1930.rd0.dat
Writing cru_ts_3_00.1931.1940.rd0.dat
Writing cru_ts_3_00.1941.1950.rd0.dat
Writing cru_ts_3_00.1951.1960.rd0.dat
Writing cru_ts_3_00.1961.1970.rd0.dat
Writing cru_ts_3_00.1971.1980.rd0.dat
Writing cru_ts_3_00.1981.1990.rd0.dat
Writing cru_ts_3_00.1991.2000.rd0.dat
Writing cru_ts_3_00.2001.2006.rd0.dat
crua6[/cru/cruts/version_3_0/secondaries/rd0]
<END QUOTE>

Back to where this all started - Vapour Pressure.

We have:

1. 'Master' (ie original) database                    vap.0311181410.dtb
2. MCDW updates database                              vap.0709111032.dtb
3. CLIMAT updates database *with added metadata*      vap.0710151817.dtb

so first we incorporate the MCDW updates..

<BEGIN QUOTE>
uealogin1[/cru/cruts/version_3_0/db/vap] ./newmergedb

WELCOME TO THE DATABASE UPDATER

Before we get started, an important question:
If you are merging an update - CLIMAT, MCDW, Australian - do
you want the quick and dirty approach? This will blindly match
on WMO codes alone, ignoring data/metadata checks, and making any
unmatched updates into new stations (metadata permitting)?

Enter 'B' for blind merging, or <ret>: B
Please enter the Master Database name: vap.0311181410.dtb
Please enter the Update Database name: vap.0709111032.dtb

Reading in both databases..
Master database stations:     7691
Update database stations:     2398

Looking for WMO code matches..
    2 reject(s) from update process 0710241541

Writing vap.0710241541.dtb

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

OUTPUT(S) WRITTEN

New master database: vap.0710241541.dtb

Update database stations:         2398
 > Matched with Master stations:  1847
                 (automatically:  1847)
                   (by operator:     0)
 > Added as new Master stations:   549
 > Rejected:                         2
   Rejects file:                vap.0709111032.dtb.rejected
uealogin1[/cru/cruts/version_3_0/db/vap]
<END QUOTE>

Then, the CLIMAT ones:

<BEGIN QUOTE>
uealogin1[/cru/cruts/version_3_0/db/vap] ./newmergedb

WELCOME TO THE DATABASE UPDATER

Before we get started, an important question:
If you are merging an update - CLIMAT, MCDW, Australian - do
you want the quick and dirty approach? This will blindly match
on WMO codes alone, ignoring data/metadata checks, and making any
unmatched updates into new stations (metadata permitting)?

Enter 'B' for blind merging, or <ret>: B
Please enter the Master Database name: vap.0710241541.dtb
Please enter the Update Database name: vap.0710151817.dtb

Reading in both databases..
Master database stations:     8240
Update database stations:     2870

Looking for WMO code matches..
   68 reject(s) from update process 0710241549

Writing vap.0710241549.dtb

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

OUTPUT(S) WRITTEN

New master database: vap.0710241549.dtb

Update database stations:         2870
 > Matched with Master stations:  2599
                 (automatically:  2599)
                   (by operator:     0)
 > Added as new Master stations:   203
 > Rejected:                        68
   Rejects file:                vap.0710151817.dtb.rejected
uealogin1[/cru/cruts/version_3_0/db/vap]
<END QUOTE>

So, not as good as the MCDW update.. lost 68.. but then of course we are talking about station data that
arrived with NO metadata AT ALL.

So we will try the unaltered rd0 process on vap. It should be the same; a mix of synthetic and observed.

*******************************************************************************
* PRIORITY INTERRUPT * PRIORITY INTERRUPT * PRIORITY INTERRUPT * PRIORITY INTERRUPT *
*******************************************************************************

After an email enquiry from Wladimir J. Alonso (alonsow@mail.nih.gov), in which unusual behaviour of CRU TS 2.10
Vapour Pressure data was observed, I discovered that some of the Wet Days and Vepour Pressure datasets had been
```

swapped!!

The files I was looking at were decadal, 1981-1990.

Vapour Pressure, January: Min 0 Max 310
Vapour Pressure, February: Min 0 Max 280

Wet Days, January: Min 0 Max 3220
Wet days, February: Min 0 Max 3240

So I wrote crutsstats.for, whioch returns monthly and annual minima, maxima and means for any gridded output file.

Tried it on the full runs, and they look OK:

```
crua6[/cru/cruts/vap_wet_investigation] head -90 cru_ts_2_10.1901-2002.vap.grid.stats |tail -10
1981       0     322      82       0     324      84       0     320      90       0     335      99       0     352     111       0     356     130       0     349     144
1982       0     312      80       0     323      83       0     318      88       0     329      98       0     348     111       0     357     126       0     365     143
1983       0     348      82       0     340      85       0     330      90       0     505      99       0     348     112       0     364     130       0     360     145
1984       0     312      80       0     320      82       0     315      89       0     329      97       0     347     112       0     359     130       0     353     144
1985       0     314      80       0     320      81       0     319      88       0     359      98       0     352     111       0     367     128       0     358     141
1986       0     312      81       0     330      83       0     316      89       0     321      99       0     366     112       0     394     129       0     371     143
1987       0     320      81       0     318      85       0     318      88       0     335      98       0     363     112       0     366     130       0     397     147
1988       0     413      83       0     324      84       0     352      90       0     323      99       0     346     113       0     363     131       0     367     148
1989       0     336      80       0     320      83       0     327      90       0     324      98       0     343     112       0     366     130       0     365     145
1990       0     320      83       0     323      85       0     476      92       0     413     101       0     361     113       0     363     132       0     371     146

crua6[/cru/cruts/vap_wet_investigation] head -90 cru_ts_2_10.1901-2002.wet.grid.stats | tail -10
1981       0    3100    1018       0    2800     919       0    3100     980       0    3000     911       0    3100     945       0    3000    1010       0    3100    1051
1982       0    3100     983       0    2800     894       0    3100     967       0    3000     925       0    3100     927       0    3000     941       0    3100     979
1983       0    3100    1035       0    2800     863       0    3100     941       0    3000     919       0    3100     929       0    3000     949       0    3100     990
1984       0    3100     981       0    2900     848       0    3100     920       0    3000     841       0    3100     932       0    3000     973       0    3100    1048
1985       0    3100     969       0    2800     896       0    3100     952       0    3000     896       0    3100     928       0    3000     938       0    3100    1057
1986       0    3100     988       0    2800     908       0    3100     950       0    3000     895       0    3100     922       0    3000     962       0    3100    1022
1987       0    3100    1011       0    2800     909       0    3100     930       0    3000     856       0    3100     954       0    3000     972       0    3100    1021
1988       0    3100    1033       0    2900     924       0    3100     971       0    3000     903       0    3100     938       0    3000     980       0    3100    1039
1989       0    3100    1019       0    2800     936       0    3100    1015       0    3000     892       0    3100     978       0    3000    1020       0    3100    1054
1990       0    3100     996       0    2800     959       0    3100    1011       0    3000     953       0    3100     928       0    3000     907       0    3100     983
```

So the monthly maxima are fine here. But for the decadal files?

```
crua6[/cru/cruts/vap_wet_investigation] cat cru_ts_2_10.1981-1990.vap.grid.stats0
1981       0     310     102       0     280      92       0     310      98       0     300      91       0     310      95       0     300     101       0     310     105
1982       0     310      98       0     280      89       0     310      97       0     300      93       0     310      93       0     300      94       0     310      98
1983       0     310     104       0     280      86       0     310      94       0     300      92       0     310      93       0     300      95       0     310      99
1984       0     310      98       0     290      85       0     310      92       0     300      84       0     310      93       0     300      97       0     310     105
1985       0     310      97       0     280      90       0     310      95       0     300      90       0     310      93       0     300      94       0     310     106
1986       0     310      99       0     280      91       0     310      95       0     300      90       0     310      92       0     300      96       0     310     102
1987       0     310     101       0     280      91       0     310      93       0     300      86       0     310      95       0     300      97       0     310     102
1988       0     310     103       0     290      92       0     310      97       0     300      90       0     310      94       0     300      98       0     310     104
1989       0     310     102       0     280      94       0     310     101       0     300      89       0     310      98       0     300     102       0     310     105
1990       0     310     100       0     280      96       0     310     101       0     300      95       0     310      93       0     300      91       0     310      98

crua6[/cru/cruts/vap_wet_investigation] cat cru_ts_2_10.1981-1990.wet.grid.stats
1981       0    3220     819       0    3240     842       0    3200     903       0    3350     992       0    3520    1113       0    3560    1304       0    3490    1440
1982       0    3120     801       0    3230     827       0    3180     881       0    3290     982       0    3480    1108       0    3570    1264       0    3650    1432
1983       0    3480     820       0    3400     850       0    3300     898       0    5050     993       0    3480    1125       0    3640    1295       0    3600    1451
1984       0    3120     803       0    3200     823       0    3150     887       0    3290     971       0    3470    1124       0    3590    1299       0    3530    1437
1985       0    3140     803       0    3200     815       0    3190     882       0    3590     978       0    3520    1113       0    3670    1277       0    3580    1405
1986       0    3120     809       0    3300     827       0    3160     889       0    3210     990       0    3660    1120       0    3940    1294       0    3710    1428
1987       0    3200     810       0    3180     849       0    3180     880       0    3350     980       0    3630    1124       0    3660    1296       0    3970    1466
1988       0    4130     829       0    3240     835       0    3520     902       0    3230     989       0    3460    1133       0    3630    1311       0    3670    1475
1989       0    3360     804       0    3200     825       0    3270     898       0    3240     978       0    3430    1120       0    3660    1301       0    3650    1447
1990       0    3200     827       0    3230     853       0    4760     918       0    4130    1005       0    3610    1127       0    3630    1322       0    3710    1462
```

Much confusion! The orders of magnitude have changed to reflect the expected ranges - but the data have clearly been swapped!

Another decade:

```
crua6[/cru/cruts/vap_wet_investigation]cat cru_ts_2_10.1921-1930.vap.grid.stats
1921       0     310     102       0     280      89       0     310     100       0     300      88       0     310      95       0     300      97       0     310     101
1922       0     310      95       0     280      93       0     310      97       0     300      89       0     310      95       0     300      98       0     310     105
1923       0     310     100       0     280      88       0     310      97       0     300      90       0     310      97       0     300      98       0     310     101
1924       0     310      97       0     290      89       0     310      95       0     300      90       0     310      91       0     300      97       0     310     100
1925       0     310      98       0     280      89       0     310      98       0     300      87       0     310      90       0     300      96       0     310     101
1926       0     310      99       0     280      87       0     310      95       0     300      87       0     310      95       0     300      93       0     310     103
1927       0     310      96       0     280      87       0     310      96       0     300      89       0     310      94       0     300      97       0     310     103
1928       0     310      97       0     290      89       0     310      91       0     300      88       0     310      90       0     300      96       0     310     101
1929       0     310      95       0     280      84       0     310      95       0     300      86       0     310      91       0     300      95       0     310     100
1930       0     310      98       0     280      88       0     310      97       0     300      88       0     310      93       0     300      93       0     310      99
crua6[/cru/cruts/vap_wet_investigation] cat cru_ts_2_10.1921-1930.wet.grid.stats
1921       0    3120     805       0    3190     814       0    3140     874       0    3210     969       0    3800    1106       0    3590    1289       0    3600    1439
1922       0    3120     794       0    3220     813       0    3140     874       0    3210     971       0    3470    1104       0    3590    1280       0    3560    1420
1923       0    3070     799       0    3140     808       0    3140     871       0    3210     947       0    3460    1082       0    3660    1276       0    3560    1410
1924       0    3270     792       0    3230     817       0    3160     879       0    3340     955       0    3460    1094       0    3710    1264       0    3560    1415
1925       0    3110     786       0    3190     815       0    3140     873       0    3210     966       0    3470    1084       0    3590    1253       0    3560    1408
1926       0    3260     815       0    3290     842       0    3310     889       0    3310     957       0    3460    1085       0    3950    1266       0    3560    1406
1927       0    3120     795       0    3300     822       0    3170     873       0    3360     959       0    3540    1096       0    3610    1271       0    3550    1424
1928       0    3200     809       0    3240     823       0    3140     875       0    3400     963       0    3470    1095       0    3590    1263       0    3560    1425
1929       0    3150     794       0    3190     802       0    3160     867       0    3310     950       0    3600    1084       0    3580    1250       0    3550    1399
1930       0    3190     798       0    3190     824       0    3150     881       0    3210     965       0    3470    1099       0    3590    1276       0    3530    1424
```

The same story. And the final two years:

```
crua6[/cru/cruts/vap_wet_investigation] cat cru_ts_2_10.2001-2002.vap.grid.stats
2001       0     310      87       0     280      84       0     310      90       0     300      81       0     310      87       0     300      93       0     310      95
2002       0     310      91       0     280      85       0     310      92       0     300      83       0     310      88       0     300      89       0     310      93
crua6[/cru/cruts/vap_wet_investigation] cat cru_ts_2_10.2001-2002.wet.grid.stats
2001       0    3320     834       0    3250     841       0    3180     913       0    3490    1010       0    3490    1147       0    4380    1323       0    3660    1487
2002       0    3310     837       0    3390     863       0    3270     918       0    3370    1012       0    3930    1151       0    4140    1339       0    3750    1503
```

It looks like a consistent problem: all the decadal VAp and WET files should be discarded, and only the 'full run' 1901-2002
files used. But my theory that the error occurred when the 1901-2002 files were converted to decadal doesn't sound true now,
because why would the precision levels change? Surely, if the decadal files are derived from the 1901-2002 files, it's just
a case of copying data across?

Let's look at *just* 1981, to try and assess this issue:

```
FULL 1901-2002 FILE
VAP:
1981       0     322      82       0     324      84       0     320      90       0     335      99       0     352     111       0     356     130       0     349     144
WET:
1981       0    3100    1018       0    2800     919       0    3100     980       0    3000     911       0    3100     945       0    3000    1010       0    3100    1051

DECADAL 1981-1990 FILE
VAP:
1981       0     310     102       0     280      92       0     310      98       0     300      91       0     310      95       0     300     101       0     310     105
WET:
1981       0    3220     819       0    3240     842       0    3200     903       0    3350     992       0    3520    1113       0    3560    1304       0    3490    1440
```

It's evident that the data have not only been swapped - they've been scaled too. Aaaarrrgghhhhhh!!!!!

```
*******************************************************************************
* PRIORITY INTERRUPT ENDS * PRIORITY INTERRUPT ENDS * PRIORITY INTERRUPT ENDS *
*******************************************************************************

Now, where were we.. ah yes, Vapour Pressure. So far:

Original:          vap.0311181410.dtb
                           +
MCDW:              vap.0709111032.dtb
                           v
                           v
Intermediate:      vap.0710241541.dtb
                           +
CLIMAT:            vap.0710151817.dtb
                           v
                           v
Final:             vap.0710241549.dtb

Produce anomalies:

<BEGIN_QUOTE>
crua6[/cru/cruts/version_3_0/secondaries/vap] ./anomdtb

   > ***** AnomDTB: converts .dtb to anom .txt for gridding *****

   > Enter the suffix of the variable required:
.vap
   > Select the .cts or .dtb file to load:
vap.0710241549.dtb
   > Specify the start,end of the normals period:
1961,1990
   > Specify the missing percentage permitted:
25
   > Data required for a normal:           23
   > Specify the no. of stdevs at which to reject data:
3
   > Select outputs (1=.cts,2=.ann,3=.txt,4=.stn):
3
   > Check for duplicate stns after anomalising? (0=no,>0=km range)
0
   > Select the generic .txt file to save (yy.mm=auto):
vap.txt
   > Select the first,last years AD to save:
1901,2006
   > Operating...

   > NORMALS            MEAN percent      STDEV percent
   >         .dtb     908812    45.2
   >         .cts      35390     1.8     944202    47.0
   > PROCESS        DECISION percent %of-chk
   > no lat/lon          105     0.0     0.0
   > no normal       1064261    53.0    53.0
   > out-of-range         49     0.0     0.0
   > accepted         944153    47.0
   > Dumping years 1901-2006 to .txt files...

crua6[/cru/cruts/version_3_0/secondaries/vap]
<END_QUOTE>

Well.. 47% accepted, 53% no normals.. pretty much as expected, and unlikely to improve no matter how many new CLIMAT
and MCDW updates there are. We need back data for 1961-1990.

Synthetic production:

<BEGIN_QUOTE>
IDL> vap_gts_anom,dtr_prefix='../dtrbin/dtrbin',tmp_prefix='../tmpbin/tmpbin',1901,2006,outprefix='vapsyn/vapsyn',dumpbin=1
% Compiled module: VAP_GTS_ANOM.
% Compiled module: RDBIN.
% Compiled module: STRIP.
% Compiled module: DEFXYZ.
Land,sea:       56016       68400
Calculating tmn normal
% Compiled module: TVAP.
Calculating synthetic vap normal
% Compiled module: ESAT.
Calculating synthetic anomalies
% Compiled module: MOMENT.
1901 vap (x,s2,<<,>>):  1.61250e-05  6.15570e-06    -0.160607     0.222689
% Compiled module: WRBIN.
1902 vap (x,s2,<<,>>): -0.000123188  3.46116e-05    -0.268891    0.0261283
1903 vap (x,s2,<<,>>):  6.86689e-05  4.52675e-06    -0.121429     0.123995
(etc)
<END_QUOTE>

(also produced, vapsyn/vapsyn1901 .. vapsyn/vapsyn2006)

Gridding with both observed and synthetic data:

IDL> quick_interp_tdm2,1901,2006,'vapglo/vap.',1000,gs=0.5,dumpglo='dumpglo',synth_prefix='vapsyn/vapsyn',pts_prefix='vaptxt/vap.'

Create absolute grids from anomaly grids:

<BEGIN_QUOTE>
crua6[/cru/cruts/version_3_0/secondaries/vap] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.vap
Enter a name for the gridded climatology file: clim.6190.lan.vap.grid
Enter the path and stem of the .glo files: vapglo/vap.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: vapabs/
Now, CONCENTRATE. Addition or Percentage (A/P)? A
Do you wish to limit the output values? (Y/N): Y
1. Set minimum to zero
2. Set a single minimum and maximum
3. Set monthly minima and maxima (for wet/rd0)

Choose: 1
Right, erm.. off I jolly well go!
vap.01.1901.glo
vap.02.1901.glo
(etc)
<END_QUOTE>

and finally, create the output files:
```

<BEGIN_QUOTE>
crua6[/cru/cruts/version_3_0/secondaries/vap] ./mergegrids
Welcome! This is the MERGEGRIDS program.
I will create decadal and full gridded files
from the output files of (eg) glo2abs.for.

Enter a gridfile with YYYY for year and MM for month: vapabs/vap.MM.YYYY.glo.abs
Enter Start Year: 1901
Enter Start Month: 01
Enter End Year: 2006
Enter End Month: 12

Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE: cru_ts_3_00.SSSS.YYYY.vap.dat
Try again.. read instructions this time?

Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE: cru_ts_3_00.SSSS.EEEE.vap.dat
Writing cru_ts_3_00.1901.1910.vap.dat
Writing cru_ts_3_00.1911.1920.vap.dat
Writing cru_ts_3_00.1921.1930.vap.dat
Writing cru_ts_3_00.1931.1940.vap.dat
Writing cru_ts_3_00.1941.1950.vap.dat
Writing cru_ts_3_00.1951.1960.vap.dat
Writing cru_ts_3_00.1961.1970.vap.dat
Writing cru_ts_3_00.1971.1980.vap.dat
Writing cru_ts_3_00.1981.1990.vap.dat
Writing cru_ts_3_00.1991.2000.vap.dat
Writing cru_ts_3_00.2001.2006.vap.dat
<END_QUOTE>

Ah - and I was really hoping this time that it would just WORK. But of course not - nothing works first
time in this project. I ran crutsstats on cru_ts_3_00.1901.2006.vap.dat, and:

<BEGIN_QUOTE>
crua6[/cru/cruts/version_3_0/secondaries/vap] ./crutsstats

CRUTSSTATS: Stats for CRU TS gridded files

Enter the monthly gridded data file: cru_ts_3_00.1901.2006.vap.dat

Please enter the start year: 1901

 106 years from 1901 to 2006

Output file is cru_ts_3_00.1901.2006.vap.dat.stats
1901   1   358   106
1902   1   358   106
1903   1   358   106
1904   1   358   106
1905   1   358   106
(etc)
2002   1   358   106
2003   1   358   106
2004   1   358   106
2005   1   358   106
2006   1   358   106
<END_QUOTE>

What?! Every year has the same min (fine, VAP of 0 is probably impossible), max (I can just about believe,
if there's a cell with no stations inside the cdd and the normal for it happens to be the highest value, and
MEAN (oh no, NO WAY!). What's odder - the .glo files are different:

crua6[/cru/cruts/version_3_0/secondaries/vap/vapabs] diff vap.06.1974.glo.abs.nh vap.06.1975.glo.abs.nh |wc -l
        56

Admittedly, 56 lines different out of 360 isn't hugely different. And looking, they are only slight and
infrequent differences. But the monthly stats are all cloned as well:

1901   1   311   80   1   320   83   1   315   89   1   320   98   1   346   111   1   358   128   1   356   143
1902   1   311   80   1   320   83   1   315   89   1   320   98   1   346   111   1   358   128   1   356   143
1903   1   311   80   1   320   83   1   315   89   1   320   98   1   346   111   1   358   128   1   356   143
1904   1   311   80   1   320   83   1   315   89   1   320   98   1   346   111   1   358   128   1   356   143
1905   1   311   80   1   320   83   1   315   89   1   320   98   1   346   111   1   358   128   1   356   143
1906   1   311   80   1   320   83   1   315   89   1   320   98   1   346   111   1   358   128   1   356   143
1907   1   311   80   1   320   83   1   315   89   1   320   98   1   346   111   1   358   128   1   356   143

Well the first thing to do, after the inevitable wailing and gnashing of teeth, is to re-run glo2abs
without the 'zero minimum' flag (just in case I coded that badly, I was in a hurry):

<BEGIN_QUOTE>
crua6[/cru/cruts/version_3_0/secondaries/vap] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.vap
Enter a name for the gridded climatology file: clim.6190.lan.vap.grid2
Enter the path and stem of the .glo files: vapglo/vap.
Enter the starting year: 1901
Enter the ending year: 2006
Enter the path (if any) for the output files: vapabs/
Now, CONCENTRATE. Addition or Percentage (A/P)? A
Do you wish to limit the output values? (Y/N): N
Right, erm.. off I jolly well go!
vap.01.1901.glo
vap.02.1901.glo
(etc)
<END_QUOTE>

<BEGIN_QUOTE>
crua6[/cru/cruts/version_3_0/secondaries/vap] ./mergegrids
Welcome! This is the MERGEGRIDS program.
I will create decadal and full gridded files
from the output files of (eg) glo2abs.for.

Enter a gridfile with YYYY for year and MM for month: vapabs/vap.MM.YYYY.glo.abs
Enter Start Year: 1901
Enter Start Month: 01
Enter End Year: 2006
Enter End Month: 12

Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE: cru_ts_3_00.SSSS.EEEE.vap.dat
Writing cru_ts_3_00.1901.1910.vap.dat
Writing cru_ts_3_00.1911.1920.vap.dat
Writing cru_ts_3_00.1921.1930.vap.dat
Writing cru_ts_3_00.1931.1940.vap.dat
Writing cru_ts_3_00.1941.1950.vap.dat

```
Writing cru_ts_3_00.1951.1960.vap.dat
Writing cru_ts_3_00.1961.1970.vap.dat
Writing cru_ts_3_00.1971.1980.vap.dat
Writing cru_ts_3_00.1981.1990.vap.dat
Writing cru_ts_3_00.1991.2000.vap.dat
Writing cru_ts_3_00.2001.2006.vap.dat
<END_QUOTE>
```

Sadly, that gave the same result. So what of the published (v2.10) VAP dataset? That looks ~ok:

```
<BEGIN_QUOTE>
crua6[/cru/cruts/version_3_0/secondaries/vap] ./crutsstats

CRUTSSTATS: Stats for CRU TS gridded files

Enter the monthly gridded data file: cru_ts_2_10.1901-2002.vap.grid

Please enter the start year: 1901

 102 years from 1901 to 2002

Output file is cru_ts_2_10.1901-2002.vap.grid.stats
1901       0     411     105
1902       0     413     104
1903       0     465     104
1904       0     359     104
1905       0     383     104
1906       0     376     105
1907       0     387     104
(etc)
<END_QUOTE>
```

Not good at all. Or, rather, good that it must be a solvable problem. Except that it's 10 to 5 on a Sunday afternoon and it's me that's got to solve it.

Where to start? Well, retrace your steps, that's how you get out of a minefield. So first up, to compare similar months in the anomaly files. Though I already know what I'm going to find, don't I? Because glo2abs isn't going to do anything unusual, it just adds the normal and there you go. So if the absolutes are very similar, the anomalies will be, too.. hmm. Well, I *suppose* I could try producing two more copies of the output files - one with just synthetic data and one with just observed data? It's only a couple of re-runs of the quick_interp_tdm2.pro IDL routine..

Started with the synthetic-only run:

```
<BEGIN_QUOTE>
IDL> quick_interp_tdm2,1901,2006,'vapsynglo/vapsyn.',1000,gs=0.5,dumpglo='dumpglo',nostn=1,synth_prefix='vapsyn/vapsyn'

crua6[/cru/cruts/version_3_0/secondaries/vap/syn_only] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: ../clim.6190.lan.vap
Enter a name for the gridded climatology file: clim.6190.lan.vap.grid
Enter the path and stem of the .glo files: vapsynglo/vapsyn.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: vapsynabs/
Now, CONCENTRATE. Addition or Percentage (A/P)? A
Do you wish to limit the output values? (Y/N): N
Right, erm.. off I jolly well go!
vapsyn.01.1901.glo
vapsyn.02.1901.glo
(etc)

crua6[/cru/cruts/version_3_0/secondaries/vap/syn_only] ./mergegrids
Welcome! This is the MERGEGRIDS program.
I will create decadal and full gridded files
from the output files of (eg) glo2abs.for.

Enter a gridfile with YYYY for year and MM for month: vapsynabs/vapsyn.MM.YYYY.glo.abs
Enter Start Year:  1901
Enter Start Month: 01
Enter End Year:    2006
Enter End Month:   12

Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE: cru_ts_3_00.SSSS.EEEE.vap.syn.dat
Writing cru_ts_3_00.1901.1910.vap.syn.dat
Writing cru_ts_3_00.1911.1920.vap.syn.dat
Writing cru_ts_3_00.1921.1930.vap.syn.dat
Writing cru_ts_3_00.1931.1940.vap.syn.dat
Writing cru_ts_3_00.1941.1950.vap.syn.dat
Writing cru_ts_3_00.1951.1960.vap.syn.dat
Writing cru_ts_3_00.1961.1970.vap.syn.dat
Writing cru_ts_3_00.1971.1980.vap.syn.dat
Writing cru_ts_3_00.1981.1990.vap.syn.dat
Writing cru_ts_3_00.1991.2000.vap.syn.dat
Writing cru_ts_3_00.2001.2006.vap.syn.dat
<END_QUOTE>
```

And then the observed-only:

```
<BEGIN_QUOTE>
IDL> quick_interp_tdm2,1901,2006,'vapobsglo/vapobs.',1000,gs=0.5,dumpglo='dumpglo',pts_prefix='vaptxt/vap.'

crua6[/cru/cruts/version_3_0/secondaries/vap/obs_only] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: ../clim.6190.lan.vap
Enter a name for the gridded climatology file: clim.6190.lan.vap.grid
Enter the path and stem of the .glo files: vapobsglo/vapobs.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: vapobsabs/
Now, CONCENTRATE. Addition or Percentage (A/P)? A
Do you wish to limit the output values? (Y/N): N
Right, erm.. off I jolly well go!
vapobs.01.1901.glo
vapobs.02.1901.glo
(etc)

crua6[/cru/cruts/version_3_0/secondaries/vap/obs_only] ./mergegrids
Welcome! This is the MERGEGRIDS program.
I will create decadal and full gridded files
from the output files of (eg) glo2abs.for.

Enter a gridfile with YYYY for year and MM for month: vapobsabs/vapobs.MM.YYYY.glo.abs
```

```
Enter Start Year:  1901
Enter Start Month: 01
Enter End Year:    2006
Enter End Month:   12

Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE: cru_ts_3_00.SSSS.EEEE.vap.obs.dat
Writing cru_ts_3_00.1901.1910.vap.obs.dat
Writing cru_ts_3_00.1911.1920.vap.obs.dat
Writing cru_ts_3_00.1921.1930.vap.obs.dat
Writing cru_ts_3_00.1931.1940.vap.obs.dat
Writing cru_ts_3_00.1941.1950.vap.obs.dat
Writing cru_ts_3_00.1951.1960.vap.obs.dat
Writing cru_ts_3_00.1961.1970.vap.obs.dat
Writing cru_ts_3_00.1971.1980.vap.obs.dat
Writing cru_ts_3_00.1981.1990.vap.obs.dat
Writing cru_ts_3_00.1991.2000.vap.obs.dat
Writing cru_ts_3_00.2001.2006.vap.obs.dat
<END_QUOTE>
```

So.. how do the stats look for these two datasets?

Synthetic-only:

```
<BEGIN_QUOTE>
crua6[/cru/cruts/version_3_0/secondaries/vap/syn_only] ./crutsstats

CRUTSSTATS: Stats for CRU TS gridded files

Enter the monthly gridded data file: cru_ts_3_00.1901.2006.vap.syn.dat

Please enter the start year: 1901

 106 years from 1901 to 2006

Output file is cru_ts_3_00.1901.2006.vap.syn.dat.stats
1901       1     358     106
1902       1     358     106
1903       1     358     106
1904       1     358     106
1905       1     358     106
1906       1     358     106
(etc)
<END_QUOTE>
```

Observed-only:

```
<BEGIN_QUOTE>
crua6[/cru/cruts/version_3_0/secondaries/vap/obs_only] ./crutsstats

CRUTSSTATS: Stats for CRU TS gridded files

Enter the monthly gridded data file: cru_ts_3_00.1901.2006.vap.obs.dat

Please enter the start year: 1901

 106 years from 1901 to 2006

Output file is cru_ts_3_00.1901.2006.vap.obs.dat.stats
1901       1     358     106
1902       1     358     106
1903       1     358     106
1904       1     358     106
1905       1     358     106
1906       1     358     106
(etc)
<END_QUOTE>
```

Oh, GOD. What is going on? Are we data sparse and just looking at the climatology? How can a synthetic dataset derived from tmp and dtr produce the same statistics as an 'real' dataset derived from observations?

Let's be logical. Here are the two 'separated' gridding runs:

```
IDL> quick_interp_tdm2,1901,2006,'vapsynglo/vapsyn.',1000,gs=0.5,dumpglo='dumpglo',nostn=1,synth_prefix='vapsyn/vapsyn'
IDL> quick_interp_tdm2,1901,2006,'vapobsglo/vapobs.',1000,gs=0.5,dumpglo='dumpglo',pts_prefix='vaptxt/vap.'
```

Well they look fine. The synthetic run has no other data inputs ('nostn=1'), and the observed run has no references to the synthetic data. So.. either quick_interp_tdm2.pro is doing something 'unusual', or, or.. hang on, let's try the climatology for stats:

```
1961       1     311      80       1     320      83       1     315      89       1     320      98       1     346     111       1     358     128       1     356     143
```

Ah, Bingo was his name-o! as I was hoping (well OK it's a bad kind of hope), the reason it's all the same is that it is by and large defaulting to the climatology. Which means that not much (any?) data is getting through, no matter if we use synthetic, observed, or both together. What's odd about that conclusion is that the synthetic data is derived from TMP and DTR - two very well-populated datasets! So synthetics alone should pretty much fill the.. hang on, just though of something horrendous.. oh, okay, probably not that. I was wondering if glo2abs.for was factoring the normals so that the anomalies were insignificant, but the equation is:

```
              absgrid(ilon(i),ilat(i)) =
     *          nint(anoms(ilon(i),ilat(i))*10) + normals(i,imo)
```

..so the anomaly is getting the weight! But still - - not a wise thing to leave to automatics. So glo2abs should prompt the user.. but with what? Just one anomaly and normal? Several? The same one from different timesteps? Eeek. Let's look at this actual case.

January 1961, lines 11103, 11104 in the glo file (11099, 11100 without header, putting it on about 33.5 degs N)

```
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  4.7173E-04  4.7224E-03
  5.4273E-03  6.1323E-03  6.8372E-03  7.5422E-03  8.2472E-03  1.9677E-03  0.0000E+00  0.0000E+00
```

Those anomalies are mighty tiny, given that the absolutes are three-digit integers! Hardly surprising they're not really appearing on the radar when added to normals typically two orders of magnitude higher! Even with the *10 in the glo2abs prog, we're still looking at values around 0.06.

Looked at the observed anomalies (output from anomdtb.f90) - here the anomalies are larger! Between -5 and +5, roughly, which is what I'm used to seeing in .txt files.

To investigate the synthetics, I needed to look at re-run vap_gts_tdm.pro. It says,

```
; Note that anomalies are in hPa*10 (bin) or hPa (glo)
```

So the binary file anomaly units - the ones we're using - are in hPa*10. Let's get one o' them synthetic glo files:

```
IDL> vap_gts_anom,dtr_prefix='../dtrbin/dtrbin',tmp_prefix='../tmpbin/tmpbin',1961,1961,outprefix='vapsynglo/vapsyn.',dumpglo=1
Land,sea:       56016       68400
Calculating tmn normal
% Compiled module: TVAP.
Calculating synthetic vap normal
% Compiled module: ESAT.
Calculating synthetic anomalies
```

```
% Compiled module: MOMENT.
1961 vap (x,s2,<<,>>):  5.72571e-05  9.01807e-07   -0.0653905    0.0261283
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
```

For Jan 1961 (may as well stick with it), -999 is the missing value code. The range is -0.0149 to +0.0222 (remember this is an anomaly in hPa according to the program comment). So if it's telling the truth, the binary anomalies presented to quick_interp_tdm2.pro will range from roughly -0.3 to +0.3. still nt going to impinge on normals between 1 and 358, is it?

So, what are the normals in? Well according to clim.6190.lan.vap:

```
crua6[/cru/cruts/version_3_0/secondaries/vap] head -11 clim.6190.lan.vap
Tyndall Centre grim file created on 12.01.2004 at 11:47 by Dr. Tim Mitchell
.vap = vapour pressure (hPa)
0.5deg lan clim:1961-90 MarkNew
[Long=-180.00, 180.00] [Lati= -90.00,  90.00] [Grid X,Y= 720, 360]
[Boxes=   67420] [Years=1975-1975] [Multi=    0.1000] [Missing=-999]
Grid-ref=   1, 148
  291  294  296  293  287  279  265  262  271  279  286  287
Grid-ref=   1, 311
   14   11   13   21   44   69   92   90   65   37   22   14
Grid-ref=   1, 312
   13   10   12   20   43   67   90   87   63   35   21   13
```

That's what I've been missing! D'oh. That '[Multi=   0.1000]'. That would still only give a range of 0.1 to 35.8 hPa, and my anomalies are still around 0.006 (or 0.3 for synthetics).

Two things, then. Firstly to get glo2abs to read the multiplicative factor from the climatology header and impose it on the output. Secondly to work out why all the anomalies have different magnitudes! Or is vapour pressure really so teeny?

Working on glo2abs. Well my theory for additive anomalies is this: I read in the normals, and apply the multiplicative factor in the header (for VAP it's 0.1). I assume the anomalies are already in the relevant units (ie require no factoring). This looks to be the case for .txt files anyway. So I can add the anomaly to the adjusted normal. Then (because I need integer output) I can DIVIDE by the factor (because that got us from integer to real before). Fine in theory but it all depends on the anomalies being in regular 'units' (why wouldn't they be? They're reals!). OK, check from the beginning, obs first:

Database: hPa*10 (typically 3-digit integers)

anomdtb.for calls subroutine CheckVariSuffix, which contains:

<BEGIN_QUOTE>
```
else if (Suffix.EQ.".vap") then
        Variable="vapour pressure (hPa)"
        Factor = 0.1
```
<END_QUOTE>

And how does anomdtb.f90 use the Factor? well in the original version:

<BEGIN_QUOTE>
```
crua6[/cru/cruts/untouched/code/linux/cruts] grep 'Factor' anomdtb.f90
real :: MissThresh,StdevThresh,DistanceThresh,Factor, ExeSpace,WyeSpace
                call CheckVariSuffix (LoadSuffix,Variable,Factor)
                OpTot = OpTot + (real(DataA(XAYear,XMonth,XAStn))/Factor)
                OpTotSq = OpTotSq + ((real(DataA(XAYear,XMonth,XAStn))/Factor) ** 2)
      NormMean  (XMonth,XAStn) = Factor*OpTot/OpEn
      if (OpTotSq.GT.0) NormStdev (XMonth,XAStn) = Factor*sqrt((OpTotSq/OpEn)-((OpTot/OpEn)**2))
                OpTot = OpTot + (real(DataA(XAYear,XMonth,XAStn))/Factor)
                OpTotSq = OpTotSq + ((real(DataA(XAYear,XMonth,XAStn))/Factor) ** 2)
        NormMean  (XMonth,XAStn) = Factor*OpTot/OpEn
          NormStdev (XMonth,XAStn) = Factor*sqrt((OpEn/(OpEn-1))*((OpTotSq/OpEn)-((OpTot/OpEn)**2)))
                            OpTot = OpTot + (DataA(XAYear,XMonth,XAStn)/Factor)
                            OpTotSq = OpTotSq + (DataA(XAYear,XMonth,XAStn)/Factor) ** 2
                            OpStDev = Factor*sqrt((OpEn/(OpEn-1))*((OpTotSq/OpEn)-((OpTot/OpEn)**2)))
                            OpMean  = Factor*(OpTot/OpEn)
           ALat(XAStn),ALon(XAStn),AElv(XAStn),real(DataA(XAYear,XMonth,XAStn))*Factor,AStn(XAStn)
```
<END_QUOTE>

I *think* the factor is being used multiplicatively. I don't understand why it's being used as a divisor though.. I must have understood last December because I managed to rewrite the 'standard deviation' section, also using it as a divisor!

One obvious thing to try is to use the revised glo2abs. That should now be working in 'units' (but saving in whatever range the normals are in). After that I could try comparing the old and 'new' (ie modded by me) versions of anomdtb.f90 to ensure I didn't break something (sure I didn't, but still..)

So, I revised glo2abs. It now reads the 'Multi' factor from the climatology header, and applies it to the normals before they're used.

So, re-ran quick_interp+tdm2.pro:

```
IDL> quick_interp_tdm2,1901,2006,'vapglo/vap.',1000,gs=0.5,dumpglo='dumpglo',synth_prefix='vapsyn/vapsyn',pts_prefix='vaptxt/vap.'
```

A sample of the outputs, vap.12.1962.glo, had a range of values from -2.3006 to +1.8388, with the majority being 0. A total of 56387 cells were nonzero, which given that there are 67420 land cells, isn't too bad. It's a pretty gaussian distribution, too. It still seems like a small variation (typically +/- 0.5). For the cell where I live (Norwich, 363,286), the normals are:

```
Grid-ref= 363, 286
   71   69   76   86  107  129  147  149  135  115   88   77
```

Or in hPa:

```
Grid-ref= 363, 286
    7.1   6.9   7.6   8.6   10.7  12.9  14.7  14.9  13.5  11.5  8.8   7.7
```

The nearest station (well based on a quick search) is LOWESTOFT. Taking 1962 and 1963 and scaling:

```
62  7.6   6.9   6.5   9.2  10.9  12.6  14.4  15.0  13.6  12.3   8.9   6.5
63  5.4   5.5   7.9   9.9  11.1  14.8  15.8  15.1  14.6  11.7  10.3   6.9

The ranges:
    2.2   1.4   1.4   0.7   0.2   2.2   1.4   0.1   1.0   0.6   1.4   0.4
```

Well our sample December 1962 range of anomalies was -2.3006 to +1.8388, and the January range is -3.3640 to +2.1250. So, I have to admit, that's the same order of magnitude for our particular cell, year and month(s).

So, assuming these .glo files are OK, we'll try glo2abs again:

<BEGIN_QUOTE>
```
crua6[/cru/cruts/version_3_0/secondaries/vap] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.vap
Enter a name for the gridded climatology file: deleteme1
Enter the path and stem of the .glo files: vapglo/vap.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: vapabs/
Now, CONCENTRATE. Addition or Percentage (A/P)? A
Do you wish to limit the output values? (Y/N): Y
```

```
1. Set minimum to zero
2. Set a single minimum and maximum
3. Set monthly minima and maxima (for wet/rd0)
Choose: 1
Right, erm.. off I jolly well go!
vap.01.1901.glo
vap.02.1901.glo
(etc)
<END_QUOTE>
```

..and the result.. look good! For (again) December 1962:

```
Min   0 (well I did set that, see above)
Max 315
```

Number of zeros: 1078, perfectly respectable although I do wonder if VAP=0 is illegal.. hmm.. OK, added an option in glo2abs:

```
<BEGIN_QUOTE>
crua6[/cru/cruts/version_3_0/secondaries/vap] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.vap
Enter a name for the gridded climatology file: deleteme3
Enter the path and stem of the .glo files: vapglo/vap.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: vapabs/
Now, CONCENTRATE. Addition or Percentage (A/P)? A
Do you wish to limit the output values? (Y/N): Y
1. Set minimum to zero
2. Set a single minimum and maximum
3. Set monthly minima and maxima (for wet/rd0)
4. Set all values >0, (ie, positive)
Choose: 4
Right, erm.. off I jolly well go!
vap.01.1901.glo
vap.02.1901.glo
(etc)
<END_QUOTE>
```

Result for December 1962: Min 1, Max 315. A good spread of values, without a disproportionate number of '1's, I'm please to say.

So, to generate the output files. Again.

```
<BEGIN_QUOTE>
crua6[/cru/cruts/version_3_0/secondaries/vap] ./mergegrids
Welcome! This is the MERGEGRIDS program.
I will create decadal and full gridded files
from the output files of (eg) glo2abs.for.

Enter a gridfile with YYYY for year and MM for month: vapabs/vap.MM.YYYY.glo.abs
Enter Start Year:  1901
Enter Start Month: 01
Enter End Year:    2006
Enter End Month:   12

Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE: cru_ts_3_00.SSSS.EEEE.vap.dat
Writing cru_ts_3_00.1901.1910.vap.dat
Writing cru_ts_3_00.1911.1920.vap.dat
Writing cru_ts_3_00.1921.1930.vap.dat
Writing cru_ts_3_00.1931.1940.vap.dat
Writing cru_ts_3_00.1941.1950.vap.dat
Writing cru_ts_3_00.1951.1960.vap.dat
Writing cru_ts_3_00.1961.1970.vap.dat
Writing cru_ts_3_00.1971.1980.vap.dat
Writing cru_ts_3_00.1981.1990.vap.dat
Writing cru_ts_3_00.1991.2000.vap.dat
Writing cru_ts_3_00.2001.2006.vap.dat
<END_QUOTE>
```

And what of the statistics. Well by now I've realised that we don't have complete coverage! So the normals are bound to poke through quite a bit. In fact, the story is as it was in the beginning! *cries*

```
<BEGIN_QUOTE>
crua6[/cru/cruts/version_3_0/secondaries/vap] ./crutsstats

CRUTSSTATS: Stats for CRU TS gridded files

Enter the monthly gridded data file: cru_ts_3_00.1901.2006.vap.dat

Please enter the start year: 1901

 106 years from 1901 to 2006

Output file is cru_ts_3_00.1901.2006.vap.dat.stats
1901       1     358     106
1902       1     358     106
1903       1     358     106
1904       1     358     106
1905       1     358     106
1906       1     358     106
1907       1     358     106
1908       1     358     106
(etc)
<END_QUOTE>
```

Now admittedly, the 106 mean does vary.. it hioits the dizzying heights of 107 on occasion! With a couple of 105s thrown in to balance the books. Had a look at the stats in detail, compared to those for CRU TS 2.10. And guess what? Yes.. the old stats are better! Here's the first decade:

```
CRU TS 2.10
1901       0     324      79       0     338      82       0     314      88       0     321      97       0     411     110       0     378     128       0     358     143
1902       0     312      80       0     319      82       0     314      87       0     321      96       0     413     109       0     366     125       0     356     141
1903       0     314      79       0     331      82       0     315      88       0     334      95       0     465     109       0     359     125       0     371     141
1904       0     310      78       0     319      81       0     312      86       0     321      95       0     347     109       0     359     126       0     355     140
1905       0     314      79       0     319      79       0     321      86       0     326      95       0     346     109       0     383     127       0     356     142
1906       0     328      80       0     330      81       0     323      87       0     335      98       0     376     111       0     359     128       0     356     142
1907       0     312      79       0     327      80       0     314      87       0     321      94       0     387     106       0     359     125       0     379     140
1908       0     312      79       0     323      81       0     330      86       0     338      95       0     346     109       0     359     127       0     353     142
1909       0     312      79       0     319      81       0     323      87       0     321      94       0     346     107       0     359     125       0     355     141
1910       0     312      80       0     319      82       0     315      86       0     321      95       0     347     109       0     359     126       0     383     142

CRU TS 3.00
1901       1     311      80       1     320      83       1     315      89       1     320      98       1     346     111       1     358     128       1     356     143
1902       1     311      80       1     320      83       1     315      89       1     320      98       1     346     111       1     358     128       1     356     143
1903       1     311      80       1     320      83       1     315      89       1     320      98       1     346     111       1     358     128       1     356     143
```

```
1904       1     311      80       1     320      82       1     315      89       1     320      98       1     346     111       1     358     128       1     356     143
1905       1     311      80       1     320      83       1     315      88       1     320      98       1     346     111       1     358     128       1     356     143
1906       1     311      80       1     320      83       1     315      89       1     320      98       1     346     111       1     358     128       1     356     143
1907       1     311      80       1     320      83       1     315      89       1     320      98       1     346     111       1     358     128       1     356     143
1908       1     311      80       1     320      83       1     315      89       1     320      98       1     346     111       1     358     129       1     356     143
1909       1     311      80       1     320      83       1     315      89       1     320      98       1     346     111       1     358     128       1     356     143
1910       1     311      80       1     320      83       1     315      89       1     320      98       1     346     111       1     358     128       1     356     143

..and here's a more recent decade:

CRU TS 2.10
1991       0     314      82       0     322      84       0     331      90       0     672     100       0     523     113       0     540     134       0     607     147
1992       0     337      82       0     383      84       0     450      90       0     613      98       0     347     112       0     359     128       0     373     140
1993       0     324      81       0     403      83       0     449      90       0     622      98       0     518     113       0     534     131       0     652     147
1994       0     346      82       0     396      82       0     457      90       0     626     100       0     524     113       0     507     132       0     605     146
1995       0     369      83       0     406      86       0     461      90       0     686     100       0     505     114       0     565     134       0     673     146
1996       0     334      81       0     431      83       0     548      88       0     634      97       0     524     113       0     530     131       0     645     147
1997       0     367      82       0     322      84       0     348      90       0     323      99       0     344     113       0     484     133       0     426     147
1998       0     339      84       0     345      89       0     338      92       0     355     104       0     361     116       1     531     137       1     356     152
1999       0     323      83       0     334      86       0     324      90       0     336     100       0     362     113       0     487     132       0     362     148
2000       0     319      82       0     319      85       0     319      91       0     328     102       0     356     114       0     476     133       0     358     146

CRU TS 3.00
1991       1     311      81       1     320      83       1     320      90       1     320     100       1     346     113       1     358     132       1     356     146
1992       1     311      82       1     319      84       1     315      90       1     320      97       1     346     111       1     358     127       1     356     141
1993       1     313      81       1     315      83       1     315      89       1     320      98       1     346     112       1     358     131       1     356     146
1994       1     311      82       1     322      82       1     315      89       1     320      99       1     346     112       1     358     131       1     356     146
1995       1     311      82       1     318      85       1     320      90       1     324      99       1     346     112       1     358     131       1     356     146
1996       1     311      80       1     321      82       1     320      87       1     320      96       1     346     111       1     358     130       1     356     145
1997       1     311      81       1     320      84       1     315      90       1     320      99       1     346     113       1     358     131       1     356     145
1998       1     311      81       1     334      85       1     326      89       1     338     100       1     346     114       1     358     134       1     356     148
1999       1     316      82       1     320      85       1     322      88       1     320      99       1     346     112       1     358     131       1     356     148
2000       1     317      82       1     320      84       1     315      90       1     320     100       1     346     113       1     358     131       1     356     146

I DON'T UNDERSTAND!!!!!

Well, OK - I see that a VAP of zero is acceptable. Though as it's a pressure, I don't believe it! I'll stick with 1.

The issue is that the earlier dataset has a variability (in the maximum) that we just don't have in the new one. And
I feel that I've been through every bloody phase of the process and checked we're doing it right!!!

~~~

Right. Let's look at the distributions of values in each dataset. We'll take Jan 1910 and Jun 2000. And as this is
a textual document, I'll have to describe the results.

Offsets. Well each month has 360 lines, so each year has 4320 lines. So for Jan 1910 we need to skip nine years,
or 38880 lines, then take the next 360. For Jun 2000 we need to skip 99 years, or 427680 lines, then another five
months, or 1800 lines, then take the next 360. So:

head -39240 cru_ts_2.10.1901-2002.vap.dat |tail -360 > cru_ts_2.10.Jan.1910.vap.dat
head -39240 cru_ts_3.00.1901.2006.vap.dat |tail -360 > cru_ts_3.00.Jan.1910.vap.dat

head -428040 cru_ts_2.10.1901-2002.vap.dat |tail -360 > cru_ts_2.10.Jun.2000.vap.dat
head -428040 cru_ts_3_00.1901.2006.vap.dat |tail -360 > cru_ts_3_00.Jun.2000.vap.dat

I loaded the resultant monthly files into Matlab, and played with them mercilessly.

Well to start with, they all look the same. Truly. I've got a 4-plot page with TS 2.10 in the left-hand column,
and TS 3.00 on the right. January 1910 on the top, June 2000 on the bottom. and they look pretty much inseparable,
though if I had to Spot The Difference, the TS 2.10 June 2000 distribution is a little flatter (that is, the
massive spike at the low end is a little shorter, and the rest of the entourage are a little taller.

What are particularly worthy of note are the maximums. Because they don't match those produced by crutsstats.for.

Month        Model    Max (Matlab)   Max (crutsstats)
Jan 1910   TS 2.10             312                312
Jan 1910   TS 3.00             311                311
Jun 2000   TS 2.10             319                476
Jun 2000   TS 3.00             317                358

Not entirely sure why the latter ones would be wrong. But I suspect crutsstats - because otherwise I miscounted
the line numbers to extract June 2000 with! Actually, OK, that does seem more likely.

Let's try it from the 1991-2000 files. The offset will be 9*4320 + 5*360 + 360 = 41040.

gunzip -c /cru/cruts/fromtyn1/data/cru_ts_2.10/newly_gridded/data_dec/cru_ts_2_10.1991-2000.vap.grid.gz | head -41040 | tail -360 > cru_ts_2_10.Jun.2000.vap.dat
gunzip -c cru_ts_3_00.1991.2000.vap.dat.gz | head -41040 | tail -360 > cru_ts_3_00.Jun.2000.vap.dat

Well - looks like I did miscount, because the new files are different! And so are the Maxima:

Month        Model    Max (Matlab)   Max (crutsstats)
Jun 2000   TS 2.10             300                476
Jun 2000   TS 3.00             358                358

..so almost perfect. At least the stats for the file I'm creating match.

And now the June 2000 histograms are much more interesting! And of course (for this is THIS project), much
more worrying. The June 2000 plot for the new data (3.00) shows a fall at VAP ->0. This is in contrast to the
other three, which show a more expotential decline from a high near 0 (though admittedly the 2.10 version does have a second
peak at around 120). In fact, the June 2000 3.00 series has peaks at ~90 and ~300! Oh, help.

The big question must be, why does it have so little representation in the low numbers? Especially given that I'm rounding
erroneous negatives up to 1!!

Oh, sod it. It'll do. I don't think I can justify spending any longer on a dataset, the previous version of which was
completely wrong (misnamed) and nobody noticed for five years.


So.. one week to go before handover, and I'm just STARTING the Sun/Cloud parameter, the one I thought would cause the most
trouble! Oh, boy. Let's try and work out the scenario.

Historically, we've issued Cloud:

crua6[/cru/cruts/fromtyn1/data/cru_ts_2.10/data_all] gunzip -c cru_ts_2_10.1901-2002.cld.Z |head -10
Tyndall Centre grim file created on 22.01.2004 at 13:52 by Dr. Tim Mitchell
.cld = cloud cover (percentage)
CRU TS 2.1
[Long=-180.00, 180.00] [Lati= -90.00,  90.00] [Grid X,Y= 720, 360]
[Boxes=   67420] [Years=1901-2002] [Multi=    0.1000] [Missing=-999]
Grid-ref=   1, 148
  725  750  750  700  638  600  613  613  663  675  713  725

..so data is in % x10.

Then, of course, there's the relevant read_me text (from /cru/cruts/fromdpela/code/idl/pro/read_me_GRIDDING.txt):

"Bear in mind that there is no working synthetic method for cloud, because Mark New
   lost the coefficients file and never found it again (despite searching on tape
   archives at UEA) and never recreated it. This hasn't mattered too much, because
```

```
   the synthetic cloud grids had not been discarded for 1901-95, and after 1995
   sunshine data is used instead of cloud data anyway."

So that's alright then! See also the earlier attempts to recreate TS 2.10 cloud.

The main gridding prog for cloud appears to be cal_cld_gts_tdm.pro:

pro cal_cld_gts_tdm,dtr_prefix,outprefix,year1,year2,info=info

; calculates cld anomalies using relationship with dtr anomalies
; reads coefficients from predefined files (*1000)
; reads DTR data from binary output files from quick_interp_tdm2.pro (binfac=1000)
; creates cld anomaly grids at dtr grid resolution
; output can then be used as dummy input to splining program that also
;  includes real cloud anomaly data

As for converting sun hours to cloud cover.. we only appear to have interactive, file-by-file
programs. Herewith all the relevant progs I can find:

IDL
./idl/pro/cal_cld_gts_tdm.pro        (synthetic cloud from DTR)
./idl/pro/cloudcorr.pro              (construct cloud correlation coefficients with DTR)
./idl/pro/cloudcorrspc.pro           (construct cloud correlation coefficients with sunshine %)
./idl/pro/cloudcorrspcann.pro        (construct cloud correlation coefficients with sunshine %)
./idl/pro/cloudcorrspcann9196.pro    (construct cloud correlation coefficients with sunshine %)

(the 'ann' versions above include the assumption that the relationships remain constant through the year)

F77
./f77/mnew/sh2cld_tdm.for            (this one needs to be modded as for sp2cldp_m.for I think)
./f77/mnew/Hsp2cldp_m.for            (one I wrote last year which seems to almost do what we need)
./f77/mnew/sp2cld_m.for              (this one needs to be modded as for sp2cldp_m.for I think)
./f77/mnew/sh2sp_m.for
./f77/mnew/sh2sp_normal.for
./f77/mnew/sh2sp_tdm.for

Aaaand - another head-banging shocker! The program sh2cld_tdm.for, which describes itself thusly:

      program sunh2cld
c converts sun hours monthly time series to cloud percent (n/N)

Does NO SUCH THING!!! Instead it creates SUN percentages! This is clear from the variable names and
user interactions.

So.. if I add the sunh -> sun% process from sh2cld_tdm.for into Hsp2cldp_m.for, I should end up with a
sun hours to cloud percent convertor. Possibly. Except that the sun% to cld% engine looks like it's
creating oktas instead:

          do im=1,12
            ratio = (real(sunp(im))/100)
            if (ratio.ge.0.95)           cldp(im) = 0
            if (ratio.lt.0.95.and.ratio.ge.0.35)
     *                                  cldp(im) = (0.95-ratio)*100
            if (ratio.lt.0.35.and.ratio.ge.0.15)
     *                                  cldp(im) = ((0.35-ratio)*50)+60
            if (ratio.lt.0.15)           cldp(im) = ((0.15-ratio)*100)+70
            if (cldp(im).gt.80.0)        cldp(im) = 80.0
            if (ratio.lt.0)              cldp(im) = -9999
          enddo

Added the previous '*12.5' mod to approximate true percentages (*10).

Looking back I see we found cloud and sunpercent databases (line counts shown):


    228936 cld.0301081434.dtb
    104448 cld.0312181428.dtb
    111989 combo.cld.dtb
     57395 spc.0301201628.dtb
     51551 spc.0312221624.dtb
     51551 spc.94-00.0312221624.dtb

And agreed a strategy:

<BEGIN_QUOTE>
AGREED APPROACH for cloud (5 Oct 06).

For 1901 to 1995 - stay with published data. No clear way to replicate
process as undocumented.

For 1996 to 2002:
 1. convert sun database to pseudo-cloud using the f77 programs;
 2. anomalise wrt 96-00 with anomdtb.f;
 3. grid using quick_interp_tdm.pro (which will use 6190 norms);
 4. calculate (mean9600 - mean6190) for monthly grids, using the
    published cru_ts_2.0 cloud data;
 5. add to gridded data from step 3.

This should approximate the correction needed.
<END_QUOTE>

This is confusing. I can only use one (observed) cloud database in the final gridding. The above
agreement seems to assume that all data after 1996 will come from sun. But dtbstat.for reports:

<BEGIN_QUOTE>
Report for: spc.0312221624.dtb (it's similar for the other spcs, except the earlier one goes to 2002)

Stations in Northern Hemisphere:     1750
Stations in Southern Hemisphere:      350
                          Total:     2100

Maximum Timespan in Northern Hemisphere: 1889 to 2003
Maximum Timespan in Southern Hemisphere: 1944 to 2003
                         Global Timespan: 1889 to 2003

Minimum Data Value:     0
Maximum Data Value:  1000
<END_QUOTE>

So the Sun Percent databases run for long periods. Similarly, for cloud:

<BEGIN_QUOTE>
Report for: cld.0312181428.dtb

Stations in Northern Hemisphere:     3286
Stations in Southern Hemisphere:      319
                          Total:     3605

Maximum Timespan in Northern Hemisphere: 1905 to 1996
Maximum Timespan in Southern Hemisphere: 1959 to 1996
```

```
                        Global Timespan: 1905 to 1996


Minimum Data Value:     0
Maximum Data Value:  1000
<END_QUOTE>
```

Not as long a run, and it sure ends at 1996! So 1901 to 1995 will, as agreed, remain untouched.

Well.. let's try converting the MCDW and CLIMAT Sun hours to Sun percents, then adding to the SPC database (spc.0312221624.dtb). Modified Hsh2cld .for to save sun percent too. Lots of debugging.. eventually dug out:

Doorenbos, J., Pruitt, W.O., 1977. Guidelines for predicting crop water requirements. FAO irrigation and drainage paper no. 24. Food and Agriculture Organization of the United Nations, Rome.

This was used to inform the Fortran conversion programs by indicating the latitude-potential_sun and sun-to-cloud relationships. It also assisted greatly in understanding what was wrong - Tim was in fact calculating Cloud Percent, despite calling it Sun Percent!! Just awful.

And so..

```
<BEGIN_QUOTE>
crua6[/cru/cruts/version_3_0/db/cld] ./Hsh2cld

Hsh2cld - Convert a Sun Hours database to a Cloud Percent one

Please enter the Sun Hours database: sun.0709111032.dtb
Data Factor detected: *1.000

Completed -     1693 stations converted.

Sun Percentage Database:   spc.0711271420.dtb
Cloud Percentage Database: cld.0711271420.dtb

crua6[/cru/cruts/version_3_0/db/cld] ./Hsh2cld

Hsh2cld - Convert a Sun Hours database to a Cloud Percent one

Please enter the Sun Hours database: sun.0710151817.dtb
Data Factor detected: *0.100

Completed -     2020 stations converted.

Sun Percentage Database:   spc.0711271421.dtb
Cloud Percentage Database: cld.0711271421.dtb

crua6[/cru/cruts/version_3_0/db/cld]
<END_QUOTE>
```

So, now the luxury of a little experiment.. I merged the MCDW and CLIMAT 'spc' databases into the existing one *separately*. Here were the results:

MCDW:

```
<BEGIN_QUOTE>
uealogin1[/cru/cruts/version_3_0/db/cld] ./newmergedb

WELCOME TO THE DATABASE UPDATER

Before we get started, an important question:
If you are merging an update - CLIMAT, MCDW, Australian - do
you want the quick and dirty approach? This will blindly match
on WMO codes alone, ignoring data/metadata checks, and making any
unmatched updates into new stations (metadata permitting)?

Enter 'B' for blind merging, or <ret>: B
Please enter the Master Database name: spc.0312221624.dtb
Please enter the Update Database name: spc.0711271420.dtb

Reading in both databases..
Master database stations:     2100
Update database stations:     1693

New master database: spc.0711271504.dtb

Update database stations:         1693
 > Matched with Master stations:   867
                 (automatically:   867)
                   (by operator:     0)
 > Added as new Master stations:   826
 > Rejected:                         0
<END_QUOTE>
```

CLIMAT:

```
<BEGIN_QUOTE>
Enter 'B' for blind merging, or <ret>: B
Please enter the Master Database name: spc.0312221624.dtb
Please enter the Update Database name: spc.0711271421.dtb

Reading in both databases..
Master database stations:     2100
Update database stations:     2020

   98 reject(s) from update process 0711271505

New master database: spc.0711271505.dtb

Update database stations:         2020
 > Matched with Master stations:   917
                 (automatically:   917)
                   (by operator:     0)
 > Added as new Master stations:  1005
 > Rejected:                        98
   Rejects file:                spc.0711271421.dtb.rejected
<END_QUOTE>
```

So, as expected, a few of the CLIMAT stations couldn't be matched for metadata.. no worries. what's interestng is that roughly the same ratio of stations were matched with existing in both cases (867/1693 vs 917/2020). Slightly better for MCDW though.

Now, as our updates only start in 2003, that means we've just lost between 826 and 1005 sets of data (added as new). We can't be exact as we don't know the overlap between the MCDW and the CLIMAT bulletins.. but we will have a better idea when I try the anomdtb experiment on the combined update. First, add the CLIMAT update again, this time to the MCDW-updated database:

CLIMAT:

```
<BEGIN_QUOTE>
Enter 'B' for blind merging, or <ret>: B
Please enter the Master Database name: spc.0711271504.dtb
Please enter the Update Database name: spc.0711271421.dtb
```

```
Reading in both databases..
Master database stations:     2926
Update database stations:     2020

   38 reject(s) from update process 0711271514

New master database: spc.0711271514.dtb

Update database stations:         2020
 > Matched with Master stations:  1736
                 (automatically:  1736)
                   (by operator:     0)
 > Added as new Master stations:   246
 > Rejected:                        38
   Rejects file:                spc.0711271421.dtb.rejected
<END_QUOTE>
```

Note several bits of good news! Firstly, rejects are down to 38 (60 having matched with MCDW stations). That's not *that* good of course - those will be new and so 2003 onwards only. Similarly, (1005-246=) 759 CLIMAT bulletins matched MCDW ones, they will also be 2003 onwards only. In other words, there were only (1736-759=) 977 updates to existing stations. So.. yes I'm being sidetracked again.. I found and downloaded ALL the MCDW bulletins, back to 1994!

```
<BEGIN_QUOTE>
uealogin1[/cru/cruts/version_3_0/incoming/MCDW] ./mcdw2cru

MCDW2CRU: Convert MCDW Bulletins to CRU Format

Enter the earliest MCDW file: ssm9409.fin
Enter the latest MCDW file (or <ret> for single files): ssm0708.fin

All Files Processed
tmp.0711271645.dtb: 2785 stations written   *** SEE LATER RUNS ***
vap.0711271645.dtb: 2786 stations written   *** SEE LATER RUNS ***
rdy.0711271645.dtb: 2781 stations written   *** SEE LATER RUNS ***
pre.0711271645.dtb: 2791 stations written   *** SEE LATER RUNS ***
sun.0711271645.dtb: 2184 stations written   *** SEE LATER RUNS ***

Thanks for playing! Byeee!
<END_QUOTE>
```

Now I'm not planning to re-run all the previous parameters! Hell, they should have had the older data in already! But for sun/cloud, this could help enormously. Here's the plan:

1. Merge the CLIMAT-sourced database into the new MCDW-sourced database.
2. Convert this modern sun hours database into a modern cloud percent database.
3. Add normals for 95-02.
4. Use the new program 'normshift.for' to calculate 95-02 normals from TS 2.10 CLD.
5. Calculate difference between TS 2.10 6190 normls and the above.
6. Modify the in-database normals (step 3) with the difference (step 5).
7. Carry on as before?

No.. this won't work. anomdtb.for calculates normals on the fly - it would have to know too much.

The next opportunity comes at the output from anomdtb - the normalised values in the *.txt files that the IDL gridder reads. These are just files - one per month - with lists of coordinates and values, so ideal to add normalised values to. Decided that this will be the process:

```
Modern SunH DB  ->  Hsh2cld.for  ->  Modern Cld% DB
Modern Cld% DB  ->  newprog.for  ->  6190anomalies.txt

..meanwhile, as before..

Normal Cld% DB  ->  anomdtb.for  ->  6190anomalies.txt
```

So we then just have to merge the two 6190 anomaly sets! Which could just be a concatenation.

Easy, then.. the only thing we need is the miraculous 'newprog.for'! With three days before delivery.

No, no, no - HANG ON. Let's not try and boil the ocean! How about:

```
1901-2002      Static, as published, leave well alone (or recalculate with better DTR).
2003-2006/7    Calc from modern SunH and use the suggested mods after gridding.
```

This is what was originally intended. But there will be problems:

1. MCDW only goes back to 2006, so what's the data density for 2003-2005? Should this also use synthetic cloud from DTR? I guess yes.

2. No guarantee of continuity from 2002 to 2003. This could be the real stickler. Moving from one system to the other - this is why it might be better to re-run 1901-2002 as well.

OKAY.. normshift.for now creates a gridded set of conversion data between whatever period you choose and 1961-1990. Such that it can be added to the gridded output of the process run with the 'false' normalisation period.

So.. first, merge your bulletins:

Well FIRSTLY, you realise that your databases don't have normals lines, so you modify mcdw2cru.for and climat2cru.for to optionally add them, then you re-run them on the bulletins, ending up with:

```
<BEGIN_QUOTE>
uealogin1[/cru/cruts/version_3_0/incoming/MCDW] ./mcdw2cru

MCDW2CRU: Convert MCDW Bulletins to CRU Format

Enter the earliest MCDW file: ssm9409.fin
Enter the latest MCDW file (or <ret> for single files): ssm0708.fin
Add a dummy normals line? (Y/N): Y

All Files Processed
tmp.0711272156.dtb: 2785 stations written
vap.0711272156.dtb: 2786 stations written
rdy.0711272156.dtb: 2781 stations written
pre.0711272156.dtb: 2791 stations written
sun.0711272156.dtb: 2184 stations written

Thanks for playing! Byeee!
<END_QUOTE>

<BEGIN_QUOTE>
uealogin1[/cru/cruts/version_3_0/incoming/CLIMAT] ./climat2cru

CLIMAT2CRU: Convert MCDW Bulletins to CRU Format

Enter the earliest CLIMAT file: climat_data_200301.txt
Enter the latest CLIMAT file (or <ret> for single file): climat_data_200707.txt
Add a dummy normals line? (Y/N): Y
```

```
All Files Processed
tmp.0711272219.dtb: 2881 stations written
vap.0711272219.dtb: 2870 stations written
rdy.0711272219.dtb: 2876 stations written
pre.0711272219.dtb: 2878 stations written
sun.0711272219.dtb: 2020 stations written
tmn.0711272219.dtb: 2800 stations written
tmx.0711272219.dtb: 2800 stations written

Thanks for playing! Byeee!
<END_QUOTE>

So.. NOW can I merge CLIMAT into MCDW?!

As expected, thank goodness:

<BEGIN_QUOTE>
uealogin1[/cru/cruts/version_3_0/incoming/merge_CLIMAT_into_MCDW] ./newmergedb

WELCOME TO THE DATABASE UPDATER

Before we get started, an important question:
If you are merging an update - CLIMAT, MCDW, Australian - do
you want the quick and dirty approach? This will blindly match
on WMO codes alone, ignoring data/metadata checks, and making any
unmatched updates into new stations (metadata permitting)?

Enter 'B' for blind merging, or <ret>: B
Please enter the Master Database name: sun.0711272156.dtb
Please enter the Update Database name: sun.0711272219.dtb

Reading in both databases..
Master database stations:     2184
Update database stations:     2020

Looking for WMO code matches..
   28 reject(s) from update process 0711272225

Writing sun.0711272225.dtb

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

OUTPUT(S) WRITTEN

New master database: sun.0711272225.dtb

Update database stations:         2020
 > Matched with Master stations:  1775
                 (automatically:  1775)
                   (by operator:     0)
 > Added as new Master stations:   217
 > Rejected:                        28
   Rejects file:                sun.0711272219.dtb.rejected
<END_QUOTE>

Wahey! Lots of stations to play with!

So, next.. convert to cloud!

<BEGIN_QUOTE>
crua6[/cru/cruts/version_3_0/db/cld] ./Hsh2cld

Hsh2cld - Convert a Sun Hours database to a Cloud Percent one

Please enter the Sun Hours database: sun.0711272225.dtb
Data Factor detected: *1.000

Completed -     2401 stations converted.

Sun Percentage Database:   spc.0711272230.dtb
Cloud Percentage Database: cld.0711272230.dtb
<END_QUOTE>

So.. bated breath..

..and yay!

<BEGIN_QUOTE>
crua6[/cru/cruts/version_3_0/secondaries/cld] ./anomdtb

   > ***** AnomDTB: converts .dtb to anom .txt for gridding *****

   > Enter the suffix of the variable required:
.cld
   > Select the .cts or .dtb file to load:
cld.0711272230.dtb

   > Specify the start,end of the normals period:
1995,2002
   > Specify the missing percentage permitted:
12.5
   > Data required for a normal:            7
   > Specify the no. of stdevs at which to reject data:
3
   > Select outputs (1=.cts,2=.ann,3=.txt,4=.stn):
3
   > Check for duplicate stns after anomalising? (0=no,>0=km range)
0
   > Select the generic .txt file to save (yy.mm=auto):
cld.txt
   > Select the first,last years AD to save:
1995,2007
   > Operating...

 /tmp_mnt/cru-auto/cruts/version_3_0/secondaries/cld/cld.0711272230.dts

   > NORMALS            MEAN percent      STDEV percent
   >         .dtb          0      0.0
   >         .cts      83961    49.3      83961    49.3
   > PROCESS        DECISION percent %of-chk
   > no lat/lon           95     0.1     0.1
   > no normal         86174    50.6    50.7
   > out-of-range         28     0.0     0.0
   > accepted          83933    49.3
   > Dumping years 1995-2007 to .txt files...
<END_QUOTE>

Well.. a 'qualified' yay.. only half got normals! But I don't like to raise the 'missing percentage'
limit to 25% because we're only talking about 8 values to begin with!!
```

The output files look OK.. between 400 and 600 values in each, not a lot really but hey, better than nowt. So onto the conversion data (must stop calling 'em factors, they're not multiplicative).

<BEGIN_QUOTE>
```
crua6[/cru/cruts/version_3_0/secondaries/cld] ./normshift

NORMSHIFT - Normals from any period

Please enter the source file:   cru_ts_2_10.1901-2002.cld.grid
Enter the start year of this file:  1901
Enter the end year of this file:    2002
Enter the normal period start year: 1995
Enter the normal period end year:   2002
Enter the 3-character parameter:    cld

Normals file will be: clim.9502.to.6190.grid.cld
```
<END_QUOTE>

So, erm.. now we need to create our synthetic cloud from DTR. Except that's the thing we CAN'T do because pro cal_cld_gts_tdm.pro needs those bloody coefficients (a.25.7190, etc) that went AWOL. Frustratingly we do have some of the outputs from the program (ie, a.25.01.7190.glo), but that's obviously no use.

So, erm. We need synthetic cloud for 2003-2007, or we won't have enough data to run with. And yes it's taken me this long to realise that. Oh, bugger.

Had a detailed search around Mark New's old disk (still online thankfully). Found this:

<BEGIN_QUOTE>
```
crua6[/cru/mark1/markn/gts/cld/val] ls -l
total 7584
lrwxrwxrwx   1 f080     cru           25 Sep 12  2005 c1 -> /cru/u1/f080/isccp/c1_mon
-rw-r--r--   1 f080     cru         1290 Mar 24  1998 cld_corr.j
-rw-r--r--   1 f080     cru          938 Mar 17  1998 cld_scat.j
-rw-r-----   1 f080     cru       922584 Mar 24  1998 cru_hahn_corr.ps
-rw-r-----   1 f080     cru       922588 Mar 24  1998 cru_isccp_corr.ps
-rw-r-----   1 f080     cru          533 Mar 27  1998 cruobs_hahn_corr.j
-rw-r-----   1 f080     cru       868561 Mar 27  1998 cruobs_hahn_corr.ps
-rw-r--r--   1 f080     cru          697 Mar 20  1998 dtr_corr.j
-rw-r-----   1 f080     cru           50 Mar 27  1998 foo
-rw-r-----   1 f080     cru       248832 Mar 27  1998 glo25.cld.1980
-rw-r-----   1 f080     cru       248832 Mar 27  1998 glo25.cld.1981
-rw-r-----   1 f080     cru       248832 Mar 27  1998 glo25.cld.1982
-rw-r-----   1 f080     cru       248832 Mar 27  1998 glo25.cld.1983
-rw-r-----   1 f080     cru       248832 Mar 27  1998 glo25.cld.1984
-rw-r-----   1 f080     cru       248832 Mar 27  1998 glo25.cld.1985
-rw-r-----   1 f080     cru       248832 Mar 27  1998 glo25.cld.1986
-rw-r-----   1 f080     cru       248832 Mar 27  1998 glo25.cld.1987
-rw-r-----   1 f080     cru       248832 Mar 27  1998 glo25.cld.1988
-rw-r-----   1 f080     cru       248832 Mar 27  1998 glo25.cld.1989
-rw-r-----   1 f080     cru       248832 Mar 27  1998 glo25.cld.1990
-rw-r-----   1 f080     cru       248832 Mar 27  1998 glo25.cld.1991
-rw-r-----   1 f080     cru       248832 Mar 27  1998 glo25.cld.1992
-rw-r-----   1 f080     cru       248832 Mar 27  1998 glo25.cld.1993
-rw-r-----   1 f080     cru       248832 Mar 27  1998 glo25.cld.1994
-rw-r-----   1 f080     cru       248832 Mar 27  1998 glo25.cld.1995
-rw-r-----   1 f080     cru       922592 Mar 24  1998 hahn_isccp_corr.ps
-rw-r-----   1 f080     cru         2378 Mar 24  1998 test.j
```
<END_QUOTE>

..which looks to me like the place where he calculated the coefficients. The *.j files are IDL 'Journal' files, so can be run from within IDL. This was my first attempt:

<BEGIN_QUOTE>
```
IDL> .run cld_corr.j
% Compiled module: $MAIN$.
% Compiled module: RD25_GTS.
YEAR:    1981
% Compiled module: RDBIN.
% Compiled module: STRIP.
foo: Permission denied.
foo: Permission denied.
foo: Permission denied.
% OPENR: Error opening file. Unit: 99, File: /home/cru/f098/u1/hahn/hahn25.1981
  No such file or directory
% Execution halted at:  RDBIN             63 /cru/u2/f080/Idl/rdbin.pro
%                       RD25_GTS          11 /cru/u2/f080/Idl/rd25_gts.pro
%                       $MAIN$             1 /tmp_mnt/cru-auto/mark1/f080/gts/cld/val/cld_corr.j
IDL>
```
<END_QUOTE>

I then had to chase around to find three sets of missing files.. to fulfil these five conditions:

```
if keyword_set(hgrid) eq 0 then rd25_gts,$
    hgrid,'~/u1/hahn/hahn25.',1981,1991
if keyword_set(rgrid) eq 0 then rd25_gts,$
    rgrid,'../glo_reg_25/glo.cld.',1981,1991
if keyword_set(hgrid2) eq 0 then rd25_gts,$
    hgrid2,'~/u1/hahn/hahn25.',1983,1991
if keyword_set(igrid) eq 0 then rdisccp_gts,$
    igrid,'c1/isccp.',1983,1991
if keyword_set(rgrid2) eq 0 then rd25_gts,$
    rgrid2,'../glo_reg_25/glo.cld.',1983,1991
```

I managed to find the hahn25 files (on Mark's disk), and some likely-looking isccp files (also on Mark's disk). But although there were plenty of files with 'glo', 'cld' and '25' in them, there were none matching the filename construction above. However, as some of those were in the same directory - I'll take that chance!!

I did try, honestly. Very hard. I found all the files, and put them in directories. I made a local copy of the job file, 'H_cld_corr.j', with the local directory refs in. Hell, I even precompiled the correct version of rdbin!

All for nothing, as usual. It runs quite happily, zipping through things, until:

<BEGIN_QUOTE>
```
% Compiled module: RDISCCP_GTS.
YEAR:    1983
% Compiled module: RDISCCP.
c1/isccp.83.07.72
c1/isccp.83.07.72.Z: No such file or directory
c1/isccp.83.08.72
c1/isccp.83.08.72.Z: No such file or directory
c1/isccp.83.09.72
c1/isccp.83.09.72.Z: No such file or directory
c1/isccp.83.10.72
c1/isccp.83.10.72.Z: No such file or directory
c1/isccp.83.11.72
c1/isccp.83.11.72.Z: No such file or directory
c1/isccp.83.12.72
c1/isccp.83.12.72.Z: No such file or directory
YEAR:    1984
c1/isccp.84.01.72
```
<END_QUOTE>

```
cl/isccp.84.01.72.Z: No such file or directory
(etc)

It isn't seeing the isccp files EVEN THOUGH THEY ARE THERE. Odd. If I create Z files it says they aren't compressed.

It ends with:

YEAR:    1991
yes
filesize=      248832
gridsize=      2.50000
% Compiled module: MARK_CORRELATE.
% Compiled module: CORRELATE.

I have no idea what it's actually done though. It doesn't appear to have produced anything.. ah:


IDL> help
% At  $MAIN$              1 /tmp_mnt/cru-auto/cruts/version_3_0/cloud_synthetics/H_cld_corr.j
CRU_HAHN_CORR   FLOAT     = Array[144, 72, 12]
CRU_ISCCP_CORR  FLOAT     = Array[144, 72, 12]
HGRID           LONG      = Array[144, 72, 12, 11]
HGRID2          LONG      = Array[144, 72, 12, 9]
IGRID           LONG      = Array[144, 72, 12, 9]
ILAT            INT       =       72
ILON            INT       =      144
IM              INT       =       12
ISCCP_HAHN_CORR FLOAT     = Array[144, 72, 12]
N               LONG      = Array[5225]
NN              LONG      =         5225
RGRID           LONG      = Array[144, 72, 12, 11]
RGRID2          LONG      = Array[144, 72, 12, 9]
Compiled Procedures:
    $MAIN$  DEFXYZ      RD25_GTS    RDBIN       RDISCCP     RDISCCP_GTS

Compiled Functions:
 CORRELATE   MARK_CORRELATE          STRIP

IDL>

..so this is one of a set of tools *that you have to know how to use*. All the work's done in the IDl data space.

Well as we don't have any instructions, that's a complete waste of two-and-a-half days' time.

Let's forget about CLD and start worrying about NetCDF.


NETDCF

Well now, we have to make the data available in NetCDF and ASCII grid formats. At the moment, it might be best to
just post-process the final ASCII grids into NetCDF; though more elegant to have mergegrids.for produce both! As it
has the data there anyway.. so I modified mergegrids.for into makegrids.for, with added NetCDF goodness. as
usual, lots of problems getting the syntax right..

*********************************************************************
BADC Work.. at RAL, 3-5 December 2007

Finally got NetCDF & Fortran working on the chosen server here (damp.badc.rl.ac.uk). I am definitely not a
chamaeleonic life form when it comes to unfamiliar computer systems. Shame. The elusive command line compile
statement is:

gfortran -I/usr/local/netcdf-3.6.2/include/ -o fileout.o filein.f /usr/local/netcdf-3.6.2/lib/libnetcdf.a

Hunting for CDDs I found a potential problem with binary DTR (used in the construction of Frost Days, Vapour
Pressure, and (eventually) Cloud. It looks as though there was a mistyping when the 2.5-degree binaries were
constructed:

IDL> quick_interp_tdm2,1901,2006,'dtrbin/dtrbin',50,gs=2.5,dumpbin='dumpbin',pts_prefix='dtrtxt/dtr.'

That '50' should have been.. 750! Oh bugger. Well, might as well see if generation does work here. DTR/bin/2.5:

..er.. hang on while I try and get IDL to recognise a path.. meh. As usual I find this effectively
impossible, so have to issue manual .compile statements. The suite of progs required to compile for
quick_interp_tdm2.pro is:

glimit.pro
area_grid.pro
strip.pro
wrbin.pro

..and, of course, quick_interp_tdm2.pro. Actually, others need others.. so I wrote a
generic IDL script to load all of Satan's little helpers:

IDL> @../../programs/idl/loads4idl.j
% Compiled module: GLIMIT.
% Compiled module: AREA_GRID.
% Compiled module: STRIP.
% Compiled module: WRBIN.
% Compiled module: RDBIN.
% Compiled module: DEFXYZ.
% Compiled module: FRSCAL.
% Compiled module: DAYS.
% Compiled module: RNGE.
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
% Compiled module: TVAP.
% Compiled module: ESAT.
IDL>

This is just because I still don't have IDL_PATH working so
having to issue each of the above as manual compile statements (in that order)
was getting tedious. n00b. [this now fixed - ed] Anyway, here's the corrected
binary DTR production:

IDL> quick_interp_tdm2,1901,2006,'dtrbin/dtrbin',750,gs=2.5,dumpbin='dumpbin',pts_prefix='dtrtxt/dtr.'
Defaults set
    1901
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: MEAN.
% Compiled module: MOMENT.
% Compiled module: STDDEV.
grid 1901 non-zero    0.9415    1.8771    1.8417 cells=     5608
    1902
grid 1902 non-zero    0.8608    1.8713    1.8752 cells=     5569
(etc)

And so to regenerate FRS:

<BEGIN_QUOTE>
```

```
IDL> frs_gts,dtr_prefix='dtrbin/dtrbin',tmp_prefix='tmpbin/tmpbin',1901,2006,outprefix='frssyn/frssyn'
IDL> quick_interp_tdm2,1901,2006,'frsgrid/frsgrid',750,gs=0.5,dumpglo='dumpglo', nostn=1,synth_prefix='frssyn/frssyn'
-bash-3.00$ ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.frs
Enter a name for the gridded climatology file: clim.6190.lan.frs.delme
Enter the path and stem of the .glo files: frsgrid/frs.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: frsabs/
Now, CONCENTRATE. Addition or Percentage (A/P)? A
Do you wish to limit the output values? (Y/N): Y
1. Set minimum to zero
2. Set a single minimum and maximum
3. Set monthly minima and maxima (for wet/rd0)
4. Set all values >0, (ie, positive)
Choose: 3
Right, erm.. off I jolly well go!
frs.01.1901.glo
frs.02.1901.glo
(etc)
<END_QUOTE>

Now looking to get makegrids.for working.. managed to get the data to write by declaring REALs as
DOUBLE PRECISION - later realising I could/should have changed the NetCDF interface calls to REAL
instead! Ah well. Still tussling with the 'time' variable.. not clear how to handle observations.
Luckily, Mike S knew what the standard was:

<BEGIN_QUOTE>
>I need to define the time parameter in the NetCDF version of the CRUTS
>dataset. I suspect I need to use 'Gregorian' (which to all intents and
>porpoises is accurate although it reverts to Julian before xx/yy/1582) but
>I wondered if there was a convention (in CRU, or wider) for allocating
>standard timestamps to observations?

For the gridded temperature we use

         short time(time) ;
                 time:units = "months since 1870-1-1" ;

Remember to start with zero!

Mike
<END_QUOTE>

And that seems to now be working! Here's the run with the compile statement included:

<BEGIN_QUOTE>
-bash-3.00$ gfortran -I/usr/local/netcdf-3.6.2/include/ -o makegrids ../../programs/fortran/makegrids.for /usr/local/netcdf-3.6.2/lib/libnetcdf.a
-bash-3.00$ ./makegrids Welcome! This is the MAKEGRIDS program.
I will create decadal and full gridded files,
in both ASCII text and NetCDF formats, from
the output files of (eg) glo2abs.for.

Writing: cru_ts_3_00.1901.1910.frs.dat
         cru_ts_3_00.1901.1910.frs.nc
Writing: cru_ts_3_00.1911.1920.frs.dat
         cru_ts_3_00.1911.1920.frs.nc
Writing: cru_ts_3_00.1921.1930.frs.dat
         cru_ts_3_00.1921.1930.frs.nc
Writing: cru_ts_3_00.1931.1940.frs.dat
         cru_ts_3_00.1931.1940.frs.nc
Writing: cru_ts_3_00.1941.1950.frs.dat
         cru_ts_3_00.1941.1950.frs.nc
Writing: cru_ts_3_00.1951.1960.frs.dat
         cru_ts_3_00.1951.1960.frs.nc
Writing: cru_ts_3_00.1961.1970.frs.dat
         cru_ts_3_00.1961.1970.frs.nc
Writing: cru_ts_3_00.1971.1980.frs.dat
         cru_ts_3_00.1971.1980.frs.nc
Writing: cru_ts_3_00.1981.1990.frs.dat
         cru_ts_3_00.1981.1990.frs.nc
Writing: cru_ts_3_00.1991.2000.frs.dat
         cru_ts_3_00.1991.2000.frs.nc
Writing: cru_ts_3_00.2001.2006.frs.dat
         cru_ts_3_00.2001.2006.frs.nc
-bash-3.00$
<END_QUOTE>

And here, for a combination of posterity and boredom, is a (curtailed) dump from ncdump:

<BEGIN_QUOTE>
-bash-3.00$ ncdump cru_ts_3_00.1901.2006.frs.nc |head -300
netcdf cru_ts_3_00.1901.2006.frs {
dimensions:
        lon = 720 ;
        lat = 360 ;
        time = UNLIMITED ; // (1272 currently)
variables:
        double lon(lon) ;
                lon:long_name = "longitude" ;
                lon:units = "degrees_east" ;
        double lat(lat) ;
                lat:long_name = "latitude" ;
                lat:units = "degrees_north" ;
        int time(time) ;
                time:long_name = "time" ;
                time:units = "months since 1870-1-1" ;
                time:calendar = "standard" ;
        double frs(time, lat, lon) ;
                frs:long_name = "ground frost frequency" ;
                frs:units = "days" ;
                frs:scale_factor = 0.00999999977648258 ;
                frs:correlation_decay_distance = 750. ;
                frs:_FillValue = -9999. ;
                frs:missing_value = -9999. ;

// global attributes:
                :title = "CRU TS 3.00 Mean Temperature" ;
                :institution = "BADC" ;
                :contact = "BADC <badc@rl.ac.uk>" ;
data:

 lon = -179.75, -179.25, -178.75, -178.25, -177.75, -177.25, -176.75,
    -176.25, -175.75, -175.25,
                                (etc)
                                      170.75, 171.25, 171.75, 172.25, 172.75,
```

```
    173.25, 173.75, 174.25, 174.75, 175.25, 175.75, 176.25, 176.75, 177.25,
    177.75, 178.25, 178.75, 179.25, 179.75 ;

 lat = -89.75, -89.25, -88.75, -88.25, -87.75, -87.25, -86.75, -86.25,
    -85.75, -85.25, -84.75,
                              (etc)
                                      79.75, 80.25, 80.75, 81.25, 81.75,
    82.25, 82.75, 83.25, 83.75, 84.25, 84.75, 85.25, 85.75, 86.25, 86.75,
    87.25, 87.75, 88.25, 88.75, 89.25, 89.75 ;

 time = 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385,
    386, 387, 388, 389, 390,
                              (etc)
                                       1620, 1621, 1622, 1623, 1624, 1625,
    1626, 1627, 1628, 1629, 1630, 1631, 1632, 1633, 1634, 1635, 1636, 1637,
    1638, 1639, 1640, 1641, 1642, 1643 ;

 frs =
  -999, -999, -999, -999, -999, -999, -999, -999, -999, -999, -999, -999,
  (etc - probably some real data there somewhere)

-bash-3.00$
<END_QUOTE>

And VAP:
<BEGIN_QUOTE>
IDL> vap_gts_anom,dtr_prefix='dtrbin/dtrbin',tmp_prefix='tmpbin/tmpbin',1901,2006,outprefix='vapsyn/vapsyn.',dumpbin=1
% Compiled module: VAP_GTS_ANOM.
% Compiled module: RDBIN.
% Compiled module: STRIP.
% Compiled module: DEFXYZ.
Land,sea:       56016       68400
Calculating tmn normal
% Compiled module: TVAP.
Calculating synthetic vap normal
% Compiled module: ESAT.
Calculating synthetic anomalies
% Compiled module: MOMENT.
1901 vap (x,s2,<<,>>):  1.67770e-05  6.23626e-06    -0.160509     0.222689
1902 vap (x,s2,<<,>>): -0.000122533  3.46933e-05    -0.268891    0.0644855
(etc)
<END_QUOTE>

These numbers are different from the original runs - so that was a genuine mistyping. Eek, that's not
very promising, is it?

<BEGIN_QUOTE>
IDL> quick_interp_tdm2,1901,2006,'vapglo/vap.',1000,gs=0.5,dumpglo='dumpglo',synth_prefix='vapsyn/vapsyn.',pts_prefix='vaptxt/vap.'
% Compiled module: QUICK_INTERP_TDM2.
% Compiled module: GLIMIT.
Defaults set
    1901
% Compiled module: RDBIN.
% Compiled module: STRIP.
% Compiled module: DEFXYZ.
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
    1902
(etc)

-bash-3.00$ ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.vap
Enter a name for the gridded climatology file: clim.6190.lan.vap.delme
Enter the path and stem of the .glo files: vapglo/vap.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: vapabs/
Now, CONCENTRATE. Addition or Percentage (A/P)? A
Do you wish to limit the output values? (Y/N): Y
1. Set minimum to zero
2. Set a single minimum and maximum
3. Set monthly minima and maxima (for wet/rd0)
4. Set all values >0, (ie, positive)
Choose: 4
Right, erm.. off I jolly well go!
vap.01.1901.glo
vap.02.1901.glo
(etc)

-bash-3.00$ ./makegrids
Welcome! This is the MAKEGRIDS program.
I will create decadal and full gridded files,
in both ASCII text and NetCDF formats, from
the output files of (eg) glo2abs.for.

Enter a gridfile with YYYY for year and MM for month: vapabs/vap.MM.YYYY.glo.abs
Enter Start Year:  1901
Enter Start Month: 01
Enter End Year:    2006
Enter End Month:   12

Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE, and
ending with '.dat', eg: cru_ts_3_00.SSSS.EEEE.tmp.dat : cru_ts_3_00.SSSS.EEEE.vap.dat

Now please enter the 3-ch parameter code: vap
Enter a generic title for this dataset, eg:
CRU TS 3.00 Mean Temperature : CRU TS 3.00 Vapour Pressure
Writing: cru_ts_3_00.1901.1910.vap.dat
         cru_ts_3_00.1901.1910.vap.nc
Writing: cru_ts_3_00.1911.1920.vap.dat
         cru_ts_3_00.1911.1920.vap.nc
Writing: cru_ts_3_00.1921.1930.vap.dat
         cru_ts_3_00.1921.1930.vap.nc
Writing: cru_ts_3_00.1931.1940.vap.dat
         cru_ts_3_00.1931.1940.vap.nc
Writing: cru_ts_3_00.1941.1950.vap.dat
         cru_ts_3_00.1941.1950.vap.nc
Writing: cru_ts_3_00.1951.1960.vap.dat
         cru_ts_3_00.1951.1960.vap.nc
Writing: cru_ts_3_00.1961.1970.vap.dat
         cru_ts_3_00.1961.1970.vap.nc
Writing: cru_ts_3_00.1971.1980.vap.dat
         cru_ts_3_00.1971.1980.vap.nc
```

```
Writing: cru_ts_3_00.1981.1990.vap.dat
         cru_ts_3_00.1981.1990.vap.nc
Writing: cru_ts_3_00.1991.2000.vap.dat
         cru_ts_3_00.1991.2000.vap.nc
Writing: cru_ts_3_00.2001.2006.vap.dat
         cru_ts_3_00.2001.2006.vap.nc
-bash-3.00$
<END_QUOTE>
```

A quick look at the VAP NetCDF headers & data looked good. So - yay, that's the damage repaired, pity it took over a day of the time at RAL. But I didn't have to fix it now - it was an opportunity to get the process working in this environment.

Next problem - station counts. I had this working fine in CRU - here it's insisting on stopping indefinitely at January 1957. Discovered - after 36 hours of fretting and debugging - that it's popping its clogs on the South Polar base:

```
 890090  -900      0 2853 AMUNDSEN-SCOTT       ANTARCTICA    1957 2006 101957  -999.00
```

And what d'you know, when I debug it, it's as simple as being too close to the pole and not having any loop restrictions in the East and West hunts for valid cells.. just looping forever! Added a few simple conditionals and all seems to run.. but outputs don't look right, the Jan 1957 station counts have missing values for the polar regions.

Managed to get anomdtb compiled with gfortran, after altering a few lines (in anomdtb and its mods) where Tim had shrugged off the surly bounds of strict F90.. it must be compiled in programs/fortran/ though, with the line (embedded in the anomdtb comments too):

```
gfortran -o anomdtb filenames.f90 time.f90 grimfiles.f90 crutsfiles.f90 loadperfiles.f90
                    saveperfiles.f90 annfiles.f90 cetgeneral.f90 basicfun.f90 wmokey.f90
                    gridops.f90 grid.f90 ctyfiles.f90 anomdtb.f90
```

As part of the modifications I removed the unused options - meaning that a .dts file is no longer required (and, of course, neither is 'falsedts.for'). Ran it for the temperature database and got apparently-identical anomaly files to the ones I generated in CRU :-))) Ran quick_interp_tdm2, glo2abs and makegrids, ended up with grids very similar (though sadly not identical) to the originals (could be IDL, could be compiler, could be system).

Scripting. Now this was always going to be the challenge, for a large suite of highly-interactive programs in F77, F90 and IDL which didn't follow universal file naming conventions. So to start with, I thought it might be 'fun' to compile an exhaustive/ing list of the commands needed to make a complete update. Headings begin with *, notes are in brackets

```
* Add MCDW Updates
mcdw2cru (interactive)
newmergedb (per parameter, interactive)
* Add CLIMAT Updates
climat2cru (interactive)
newmergedb (per parameter, interactive)
* Add BOM Updates
au2cru (unfinished, interactive, should do whole job)
* Regenerate DTR Database
tmnx2dtr (interactive)
* Produce Primary Parameters (TMP, TMN, TMX, DTR, PRE)
anomdtb (per parameter, interactive)
quick_interp_tdm2 (per parameter)
glo2abs (per parameter, interactive)
makegrids (per parameter, interactive)
* Prepare Binary Grids (TMP, DTR, PRE) for Synthetics
quick_interp_tdm2 (per parameter)
* Produce Secondary Parameter (FRS, uses TMP,DTR)
frs_gts_tdm
quick_interp_tdm2
glo2abs (interactive)
makegrids (interactive)
* Produce Secondary Parameter (VAP, uses TMP,DTR)
vap_gts_anom
anomdtb (interactive)
quick_interp_tdm2
glo2abs (interactive)
makegrids (interactive)
* Produce Secondary Parameter (WET/RD0, uses PRE)
rd0_gts_anom
anomdtb (interactive)
quick_interp_tdm2
glo2abs (interactive)
makegrids (interactive)
```

*** BACK AT CRU ***

Tried to compile makegrids.for on uealogin1 (as crua6 is being retired). Got an odd error:

```
ld: warning: file /usr/local/netcdf-3.6.1/lib/libnetcdf.a(fort-attio.o): wrong ELF class: ELFCLASS64
```

..which was cured with the addition of '-xarch=native64' to the compile statement:

```
f77 -xarch=native64 -I/usr/local/netcdf-3.6.1/include/ -o makegrids ../../BADC_AREA/programs/fortran/makegrids.for /usr/local/netcdf-3.6.1/lib/libnetcdf.a
```

Then had to play around to try and reduce the size of the NetCDF files - they were bigger than the uncompressedASCII ones! This was because the variable was declared as DOUBLE, which is 64 bits, or 8 bytes, per datum. A waste, since we deal with the data as integers and use factors to restore 'real' values. So redeclared as INT. Considering re-redeclaring as SHORT, which is 16 bits to INT's 32.. however, that only gives me signed -32,768 to 32,767 or unsigned 0 to 65,535. That's enough for our datasets but only if precip has a positive missing value code, which I don't like the sound of.

Reproduced all primaries and secondaries with INT typing for the NetCDF component.

Simultaneously trying to work out why stncounts.for is apparently ignoring the South Pole station (Amundsen-Scott) even though the rest of the output looks fine.. eventually realised that the land/sea mask is blobking it!!

Station counts work continues.. should the NetCDF files be written as INT to match the data files, or SHORT to save a lot of space? In fact, should the station counts be in the same NetCDF files as their data?!!

Finished off the local regenration of VAP and FRS from the corrected dtrbin files:

```
DTR fix:
IDL> quick_interp_tdm2,1901,2006,'dtrbin/dtrbin',750,gs=2.5,dumpbin='dumpbin',pts_prefix='dtrtxt/dtr.'

VAP binary regen:
IDL> vap_gts_anom,dtr_prefix='dtrbin/dtrbin',tmp_prefix='tmpbin/tmpbin',1901,2006,outprefix='vapsyn/vapsyn.',dumpbin=1
```

Note that, as expected, results ARE different for synthetic vap!

First two lines of the CRUA6 run:

```
1901 vap (x,s2,<<,>>):  1.53894e-05  6.16462e-06    -0.160509     0.222662
1902 vap (x,s2,<<,>>): -0.000123921  3.46215e-05    -0.268891    0.0261302
```

First two lines of the damp.badc.rl.ac.uk run:

```
1901 vap (x,s2,<<,>>):  1.67770e-05  6.23626e-06    -0.160509     0.222689
1902 vap (x,s2,<<,>>): -0.000122533  3.46933e-05    -0.268891    0.0644855

Different compilers.. different architectures?.. different OS.. whichever.

IDL> quick_interp_tdm2,1901,2006,'vapglo/vap.',1000,gs=0.5,dumpglo='dumpglo',synth_prefix='vapsyn/vapsyn.',pts_prefix='vaptxt/vap.'

Paused here as waiting for completion of makegrids2.for (includes station counts in NetCDF files). One early ramification of this is
that glo2abs.for now saves filenames with year.month, rather than replicating the loopy month.year that quick_interp_tdm2.pro gives.
This will allow file lists to be compiled with 'ls >tmpfile', so that timespan and missing files can be detected at the start. Have
also had to amend (and re-run at great length!) stnounts.for to do the same.

Just realised I still haven't worked out how to do the station counts for the secondaries. I can't work out how Tim did it either,
since he worked out counts at the same stage I do. Did he let the primary parameter's counts override? Or backfill? Well we could
take the approach that the gridding routine takes, namely to use observed data and only refer to synthetic when that fails (and only
refer to normals when that fails, obviously).

So, stncounts will have to accept TWO sets of .txt files. At each timestep it will have to first count the secondary parameter's
stations, then the primary parameter(s) will be counted and fill in any zeros in the grid. However, we will need different information
for the paper - what use is the effective station count as some is of higher 'quality' than the rest? So will probably need the regular
observed-only count as well.. which could be a separate run of stncounts but faaar more sensible to be a side effect of this run.

Oh Gods.. I've got to modify another program. *cries*

I think the mods to stncounts will be the turning point for all the programs. So far, they have all been generic, but this is not
tenable if the system is to be automated - they need to be aware of the parameters and what they mean. Otherwise stncounts will have
to be told how many primaries produced the synthetic grids, etc, etc - stupid. So I need to devise a directory structure and file
naming schema that will support the entire update process. Eeeeeeeek.



** time passes.. other projects given higher priority.. **



Back to precip, it seems the variability is too low. This points to a problem with the percentage anomaly routines. See earlier
escapades - will the Curse of Tim never be lifted?

A reminder. I started off using a 'conventional' calculation:

              absgrid(ilon(i),ilat(i)) = nint(normals(i,imo) +
     *                                        anoms(ilon(i),ilat(i)) * normals(i,imo) / 100)
            which is: V = N + AN/100

This was shown to be delivering unrealistic values, so I went back to anomdtb to see how the anomalies were contructed in the
first place, and found this:

              DataA(XAYear,XMonth,XAStn) = nint(1000.0*((real(DataA(XAYear,XMonth,XAStn)) / &
                                        real(NormMean(XMonth,XAStn)))-1.0))
            which is: A = 1000((V/N)-1)

So, I reverse engineered that to get this: V = N(A+1000)/1000

And that is apparently also delivering incorrect values. Bwaaaahh!!

Modified anomdtb to dump the precip anomaly calculation. It seems to be working with raw values, eg:

DataA:     1050, NormMean:   712.00, Anom:      475
DataA:      270, NormMean:   712.00, Anom:     -621
DataA:      710, NormMean:   712.00, Anom:       -3
DataA:      430, NormMean:   712.00, Anom:     -396
DataA:      280, NormMean:   712.00, Anom:     -607
DataA:      830, NormMean:   712.00, Anom:      166
DataA:        0, NormMean:   712.00, Anom:    -1000
DataA:      270, NormMean:   712.00, Anom:     -621
DataA:      280, NormMean:   712.00, Anom:     -607
DataA:      450, NormMean:   712.00, Anom:     -368
DataA:      180, NormMean:   712.00, Anom:     -747
DataA:     1380, NormMean:   712.00, Anom:      938

However, that -1000 is interesting for zero precip. It looks as though the anomalies are in mm*10, like
the precip database raw values.. but no! These are dumped from within the program, the output .txt files
have -100 instead of -1000. That's because of the CheckVariSuffix routine, which returns a factor based
on the parameter code, including:

else if (Suffix.EQ.".pre") then
        Variable="precipitation (mm)"
        Factor = 0.1

Factor is then used when the anomaly files are written:

       do XAStn = 1, NAStn
         if (DataA(XAYear,XMonth,XAStn).NE.-9999) write (9,"(2f8.2,f8.1,f13.5,i7)"), &
           ALat(XAStn),ALon(XAStn),AElv(XAStn),real(DataA(XAYear,XMonth,XAStn))*Factor,AStn(XAStn)
       end do

So, the anomalies are in real units (presumably mm/day).

So.. we grid these values. The resulting anomalies (for Jan 1980) look like this:

max = 1612.4
min = -100

These should be applied to the climatology (normals). I think they can be applied to either unscaled 'real'
value normals or to normals which are mm*10. Results will be scaled accordingly. So.. let's look at glo2abs.
Again.

The current formula to convert to real values is:

              absgrid(ilon(i),ilat(i)) = nint((rnormals(i,imo)*
     *                                   (anoms(ilon(i),ilat(i))+1000)/1000)/rmult)
              V = N(A+1000)/1000*rmult

Not happy with that anyway. The multiplicative factor.. should that be there at all?

Now, if these are 'genuine' percentage anomalies - ie, they represent the percentage change from the mean,
then the formula to convert them using unscaled normals would be:

     V = N + N(A/100)

For instance, -100 would give V = 0, and 100 would give V = 2N. Now if the normals are *10, surely the
results will be *10 too? As each term has N as a multiplicative factor anyway. This makes
me wonder about the prevailing theory that the anomalies need to be *10. They are
fractions of the normal, so it shouldn't matter whether the normal is real or *10. The
variability should be the same (ie, the variability of the anomalies).

So, a run (just for 1980) of glo2abs, then, using the following formula:

              absgrid(ilon(i),ilat(i)) = nint(rnormals(i,imo) +
     *                           (anoms(ilon(i),ilat(i)) * rnormals(i,imo)) / 100)
```

```
              V = N + A*N/100
```

This should give integer values of mm*10 (because the normals are mm*10 and uncorrected).

So, working backwards, what *should* the original anomalising routine in anomdtb.f90 look like? It should look like this:

```
              A = 100*(V-N)/N,  or:    A = 100(V/N) - 100, or:    A = 100((V/N)-1)
```

The last version is REMARKABLY similar to the original (A = 1000((V/N)-1)), in fact I think we can call equivalence if V and N are in mm? Complicated. The '100' is not a scaling factor, it's the number that determines a percentage calculation. If we use 1000, what are we saying? The percentage anomalies we would have got are now 10x higher. Where V=0, A will be -1000, a meaningless percentage anomaly for precip. That's where the CheckVariSuffix factor pops up, multiplying by 0.1 and getting us back where we started.

So.. Tim's original looks right, once you understand the correction factor applied later.

In which case, my original algorithm should have worked!!

A slightly Heath-Robinson attempt to verify.. extracted the cell for Wick from the 1980 output. Wick is 58.45N, 3.08W - I reckon that's (297,353). I get:

```
   106
    69
    91
    22
    16
    74
    79
    80
   129
   156
   177
   151
```

The station says:

```
1980  763  582  718  153   95  557  587  658  987 1162 1346 1113
```

sigh Yes, they do follow a similar shape.. it's just that the original station data are much higher. In terms of Up/Down (following direction):

```
Station    DUDDUUUUUUD
Gridded    DUDDUUUUUUD
```

So, once again I don't understand statistics. Quel surprise, given that I haven't had any training in stats in my entire life, unless you count A-level maths.

Normals for the same cell:

```
Grid-ref= 353, 297
 1060  740  870  600  610  670  700  910  990 1070 1200 1110
```

Now, these look like mm*10, the same units as the station data and what I expected. If I apply the anomalies to these normals, it looks like I'll get what I'm after.. the trouble is that glo2abs.for deliberately works with real values and so applies the factor in the clim header before using the values. I just don't think that works with percentages.. hmm.

Actually, it does. I ran through the algorithms and because the normal is multiplicative, you can do the scaling before or after. In other words, if V' is V produced with scaled normals (N*0.1) then we do end up with V = 10V'. So I just need to include the factor in the final equation:

```
              V = (N + A*N/100)/F
```

Ran it, and the results were good. So - as it's the only change - I won't have to regrid precip after all! Just re-run from glo onwards.. did so, then used the old (but working) version of makegrids to produce the gridded ASCII and NetCDF files.

So.. comparisons. Well I want to compare with both 2.0 and 2.1, because they do differ. So I will need to convert 2.0 to regular-gridded, as I did with 2.1. If I could only remember the program I wrote to do it!!

**

Another problem. Apparently I should have derived TMN and TMX from DTR and TMP, as that's what v2.10 did and that's what people expect. I disagree with publishing datasets that are simple arithmetic derivations of other datasets published at the same time, when the real data could be published instead.. but no.

This introduces the problem of derivation. TMN = TMP -DTR/2 and TMX = TMP + DTR/2, but this does not tell us what to do when either or both values (TMP, DTR) are missing. One thing to check is the climatologies. Here are the first two cell normals for all four parameters:

```
.tmp = near-surface temperature (degrees Celsius)
Grid-ref=   1, 148
  270  274  269  265  259  253  246  242  248  252  261  268
Grid-ref=   1, 311
 -187 -213 -195 -137  -30   47   90   81   33  -56 -132 -186

.dtr = diurnal temperature range (degrees Celsius)
Grid-ref=   1, 148
   56   71   54   59   55   50   54   51   57   61   61   73
Grid-ref=   1, 311
   74   71   72   77   59   65   64   56   49   51   63   71

.tmn = near-surface temperature minimum (degrees Celsius)
Grid-ref=   1, 148
  242  238  242  236  232  228  219  217  220  222  230  232
Grid-ref=   1, 311
 -224 -249 -231 -175  -59   15   58   53    8  -81 -163 -222

.tmx = near-surface temperature maximum (degrees Celsius)
Grid-ref=   1, 148
  298  309  296  295  287  278  273  268  277  283  292  305
Grid-ref=   1, 311
 -150 -178 -159  -98    0   80  122  109   58  -31 -100 -151
Grid-ref=   1, 312
```

Well, making allowances for rounding errors, they do seem to hold to the relationship.

Wrote maketmnx.for to derive TMN and TMX from TMP and DTR grids. Works with the output files from glo2abs.for. Ran makegrids to produce .dat and .nc files (still pre-station count inclusion).

On to precip problems. Tim O ran some comparisons between 2.10 and 3.00, in general things are much improved but there are a few hair raisers (asterisked for special concern):

```
Cape Verde Isl, MAM & DJF, P
Galapagos, All, T&P
Guinea, MAM 1901, P
Bangladesh, All 1991-2000, P **
Bhutan, DJF 1939 & 1945, P
Laos/Vietnam, DJF 1991, P **

Looked at Bangladesh first. Here, the 1990s show a sudden drop that really can only be some
stations having data a factor of 10 too low. This ties in with the WWR station data that DL
added for 1991-2000, which aprently was prone to scaling issues. Wrote stnx10.for to scale
a file of WWR Bangladesh records, then manually C&P'd the decade over the erroneous ones in
the database. Also fixed country name from 'BNGLADESH'!

Then Laos/Vietnam. Here we have an anomalously high peak for 1991 DJF. Used getllstations.for
to extract all stations in a box around Laos & Vietnam (8 to 25N, 100 to 110E), a total of 96
stations from Thailand, Vietnam, Laos, Kampuchea, and China. Eeeek. Tim O's program only worked
with boxes though. Also, I'm not 100% sure which year DJF belongs to in Tim's world.. hopefully
it's the December year (as it was the fourth column in his plot table). However.. plotted *all*
the data as overlapping years, and there is no trace of a spike in DJF. Uh-oh.

I'm not actually convinced that the 'country box' approach is much cop. Better to examine each
land cell and automagically mark any with excessions? Say 5 SD to begin with. Could then be
extra clever and pull the relevant stations and find the source of the excession? Of course, this
shouldn't happen, since there is a 4SD limit imposed by anomdtb.f90 for precip (3SD for others).

Wrote vietlaos.for to run through the lists of Vietnam and Laos cells (provided by Tim O) and
extract the DJF precip values for each (from the 1901-2006 gridded file). It then calculates the
standard deviation of each series, normalises, and notes any values over 6.0 SDs (1991 onwards).

Result.. some very high values (up to 11.3 standard deviations!) in 1991/2. Worst cells:
     Row  Column   Index  StdDev
     212     571     273   11.21
     213     571     273   11.30
     214     571     273   10.15
     212     572     273   11.11
     213     572     273   11.20
     214     572     273   10.58
     212     573     273   10.84
     213     573     273   11.10
     212     574     273   10.76
     215     572     273   10.06
     214     573     273   10.53
     213     574     273   10.94
     214     574     273   10.44
     212     575     273   10.65
     213     575     273   10.66
     211     576     273   10.96
     212     576     273   10.51

Index 273 can be related to time as follows. The series begins in 1901 and we take three values
per year (J,F,D). So 1990 would be the 90th year and the 268th-270th values. Thus 273 = Dec 1991.

The cells are all contiguous, implying a single station's influence via the gridding process:

      570   571   572   573   574   575   576
211   n/a   n/a  6.37  7.56  8.36  9.71 10.96
212   n/a 11.21 11.11 10.84 10.76 10.65 10.51
213  5.52 11.30 11.20 11.10 10.94 10.66   n/a
214  5.34 10.15 10.58 10.53 10.44   n/a   n/a
215  4.37  9.97 10.06   n/a   n/a   n/a   n/a

'n/a' means the cell isn't in the Laos or Vietnam areas.

The 'epicentre' of the anomaly looks to be cell (213,571), which is in the Laos file:
     Box  Column     Row     Lon     Lat
  205773     571     213  105.75   16.75

So we're looking for stations in the vicinity of 105.75E, 16.75N. Well the precip database has a
total of EIGHT Laos stations, so that should be straightforward:

4893000  1990  10210  304 LUANG PRABANG        LAOS          1951 2006   -999  -999.00
4893800  1920  10170  323 SAYABOURY            LAOS          1969 2006   -999  -999.00
4894000  1800  10260  170 VIENTIANE            LAOS          1941 2006   -999  -999.00
4894600  1738  10465  152 THAKHEK              LAOS          1989 2006   -999  -999.00
4894700  1660  10480  155 SAVANNAKHET          LAOS          1970 2006   -999  -999.00
4894800  1670  10500  184 SENO                 LAOS          1951 2006   -999  -999.00
4895200  1568  10643  168 SARAVANE             LAOS          1989 2006   -999  -999.00
4895500  1510  10580   93 PAKSE                LAOS          1968 2006   -999  -999.00

Well, SENO has to be the prime candidate. Unfortunately, this is from SENO:

4894800  1670  10500  184 SENO                 LAOS          1951 2006   -999  -999.00
<snip>
1989    0    0-9999 1910-9999 1010 4450 2690 2880 1340    0    0
1990   60 1560  150  420 1110 4830 3620 3690-9999  780   30    0
1991    0    0  400    0  690 1907 1890 5308 3238  805    0  366
1992  488  280   50   80 1883 2503 2644 2935 2039  131    0   89
1993    0    0  139  280 2324 1163 1949 4460 2145    0   29    0

A most undistinguished set. So, the net widens:

4894700  1660  10480  155 SAVANNAKHET          LAOS          1970 2006   -999  -999.00
<snip>
1989-9999    0-9999-9999 1080-9999-9999-9999-9999 1490-9999-9999
1990   30-9999  240-9999-9999-9999 1920-9999-9999-9999-9999    0
1991    0    0  127   49  952 2508 1681 4034 4006 1690    0  338
1992  324  338   93  691 1932 2344 2048 4464  756  607    0  197
1993    0    5  335  263 2665  921 2884 2204 1834   17   23    0

..nope..

4894600  1738  10465  152 THAKHEK              LAOS          1989 2006   -999  -999.00
<snip>
1989-9999    0-9999-9999 2030-9999-9999-9999-9999 1490-9999-9999
1990   10-9999  520-9999-9999-9999-9999-9999-9999-9999    0
1991    0    0  905  119  861 6058 3578 7092 2417  373    0  324
1992  105  318  125  456 2140 2978 4623 4595 3376  425    0  854
1993    0  108   52 1343 5835 2999 6285 4375 1017  467    8    0

..nope.. unless these values *are* unusual? Let's look at the highest two Decembers from
each station:

4893000  1990  10210  304 LUANG PRABANG        LAOS          1951 2006   -999  -999.00
1992  193  911    0  497  657 1246 2971 2837  929  584   95 1372
1994    0   54 1107  291 1702 2436 2025 3636 1516  316  185  816

4893800  1920  10170  323 SAYABOURY            LAOS          1969 2006   -999  -999.00
1992  411  719    0  816  754 1252 2573 1671 1686  991  351  879
1994    0  208 1695  503 2262 1607 1743 2562 3205  118  193  454
```

```
4894000  1800  10260  170 VIENTIANE            LAOS          1941 2006   -999  -999.00
1971    0   70  140  340 2940 2750 2890 2260 1630 1030    0  180
1992  381  273   11  424 2372 4878 4381 3676 3091  630    0  212
1994    0  300  921  322 2685 2725 4698 1932 4000 3031 1016  166 (inc for comparison with previous)

4894600  1738  10465  152 THAKHEK              LAOS          1989 2006   -999  -999.00
1991    0    0  905  119  861 6058 3578 7092 2417  373    0  324
1992  105  318  125  456 2140 2978 4623 4595 3376  425    0  854
1994    0  612  952  558 1697 7092 5121 4276 2428  486   20    2 (inc for comparison with previous)

4894700  1660  10480  155 SAVANNAKHET          LAOS          1970 2006   -999  -999.00
1991    0    0  127   49  952 2508 1681 4034 4006 1690    0  338
1992  324  338   93  691 1932 2344 2048 4464  756  607    0  197
1994    0  734  390  494 1381 3377 1525 5651 1881  600    0    0 (inc for comparison with previous)

4894800  1670  10500  184 SENO                 LAOS          1951 2006   -999  -999.00
1971    0  880  130  370 1270 4010 2200 2860 1930  410    0  140
1991    0    0  400    0  690 1907 1890 5308 3238  805    0  366
1992  488  280   50   80 1883 2503 2644 2935 2039  131    0   89 (inc for comparison with previous)
1994    0  532  318  969 2065 1937 1197 4552 1934  197    0    0 (inc for comparison with previous)

4895200  1568  10643  168 SARAVANE             LAOS          1989 2006   -999  -999.00
1992  287   33   52  222 1072 5444 2998 8899 2243 1070    0    0 (inc for comparison with previous)
1994    0   10  354  686 1743 3387 5829 3254 4219  408   41    4 (inc for comparison with previous)
1998   26  619    0  574 2386 2871 1530 2308 2680  913  463   73
2005    0    0  120 1230 1990 2860 4350 8060 3770  280  140   70

4895500  1510  10580   93 PAKSE                 LAOS          1968 2006   -999  -999.00
1972-9999-9999-9999-9999-9999-9999-9999-9999 2610  870  280  140
1992  166  101    0  210  665 1898 2574 6448 2942  648   10   31 (inc for comparison with previous)
1994    0    0  134  220 2537 3596 5161 5384 7693 1513  236   94
```

Summary: LUANG PRABANG shows a significant anomaly of 1372 for Dec 1992. Unfortunately, this finds echoes both temporal (1994 has 816) and spatial (SAYABOURY's 1992 is 879). So, if these values are causing the spike, it's genuine (if exaggerated in a way yet to be determined).

Wrote vietlaos2, to gather data from the cells AND stations. It also gets the climatology. Initially it only gathered 13 stations with data in 1991/2, using 'VIETNAM' and 'LAOS' to select on country name. However, taking the cell [214,574] in December 1991 as the peak incident, we can use those coordinates (17.25N, 107.25E) to centre a bounding box for station selection. A box 10degs square yields only 17 stations, none of which have anything remotely spikey in Dec 1991. A box 20degs square (some would say unfeasibly large) yields 98 stations, one of which does have a bit of a spike in Dec 91.. not impressively so though, and it's a long way away:

```
4855200   853   9993    3 NAKHON SI THAMMARAT  THAILAND      1912 2000   -999  -999.00
```

Over 10.5 degrees South and over 7 degrees West of the target cell. Not very convincing, especially as closer stations are bound to have masked it.

One FINAL try with vietlaos3.for. Just looking at December, now, and getting the original station normals as well as the climatological ones. The whole chain. This proves to be surprisingly complicated.

On a parallel track (this would really have been better as a blog), Tim O has found that the binary grids of primary vars (used in synthetic production of secondary parameters) should be produced with 'binfac' set to 10 for TMP and DTR. This may explain the poor performance and coverage of VAP in particular.

Back to VietLaos.. the station output from vietlaos3.for had a couple of stations with missing anomaly values:

```
       LAT       LON       ALT      NORM       VAL      ANOM
     17.15    104.13    171.00     29.00     62.00  -9999.00
     15.80    102.03    182.00     45.00     40.70  -9999.00
```

I eventually worked out that I hadn't collapsed a universal probability, it was just the 4 standard deviation screen in anomdtb (4 for precip, 3 for temp). To confirm, I did a short anomdtb run (just for 1991) with the sd limit set to 10, and sure enough:

```
   17.15  104.13   171.0   2037.900024835600
   15.80  102.03   182.0    804.400024840300
```

They both look high enough to trigger the 4sd cap. However, since the spike we're investigating is from a regular process run, where that limit was in place, we can't use those values. Program is thus amended to omit any stations without anomalies (for Dec 1991)

Next issue is to make sense of the output. The first line from the station file is (headings added):

```
       LAT       LON       ALT      NORM       VAL      ANOM
     22.60    114.10     25.00     29.00     21.60    -25.50
```

Remembering it's percentage anomalies! So 25.5% of 29 is 29*.255 = 7.395. Add that to 21.6? 29.0 :-)

By contract, the cell file looks like this:

```
     ROW     COL     LAT     LON     VAL    NORM
     220     561   20.25  100.75   12.90   15.00
```

There are 63 stations and 204 cells (196 when missing values (sea) eliminated). I guess one approach would be to grid the anomalies, to see if a peak is visible. I did. It is. The simple interpolation in Matlab puts the peak at 17.25N, 105.25E - matches the grid peak for lat and a little west for lon. The nearest high station anomaly is 2369.2, that's from:

```
4838300  1653  10472  138 MUKDAHAN             THAILAND      1934 2000   -999  -999.00
6190   34  127  290  907 1813 2900 2271 3353 2596  886   84   13
1934    0  100    0  500 3150 2940 2460 2980 3320  350    0   20
1935    0    0    0  440 1920 1560 3220  580 1770    0  170    0
1936    0    0  820    0 2320 1900 3460  810 2120    0    0    0
1937    0    0  350  660 3640  740 1920 2890 4470  330    0    0
1938    0    0  550 1300  730 1720 2340  400 2030  810    0    0
1939    0  280  700  230 1320  420 2480 4190 2130    0    0    0
1940-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1941    0    0  530  590 1800 2710  420 2790  650  750    0    0
1942    0    0  540  660 1650 3200 1200 1730 1990    0    0    0
1943    0    0 1600 1300 1960 1880 2000 2200 2600    0    0    0
1944    0-9999    0  320 2210 1040 1700 2500 2150  820   50    0
1945   70  600    0  340    0 2470 3400 2780 1620   20  330  500
1946    0    0 1360    0-9999 1720 1070 3330 2870 1260    0    0
1947    0  180   50 1390 3200 1530 3520 1150-9999   50    0    0
1948-9999-9999    0 1630 3520 1040 2900 3980 2160  380   10    0
1949    0  200  170  470 3000 2720 3110 4920  360  690  90-9999
1950    0   70    0  250 1610 2090 1040 1390 3500 1960   20    0
1951    0  340  770 1380  530 3380 1590 1950 3580 1430   20    0
1952    0    0 1170  660 1640 3160 2320 4150 3510  860   30    0
1953  260  110  430  630 1010 2200 1480 2780 1180  310   10    0
1954  460   10   30 1100 1950 2870 1120 2640 4220  620    0    0
1955    0    0  280  580 2180 4100 2900 2570 1810  270   20    0
1956    0  420  150 1000 3000 2930 3980 3840 2020  220    0    0
1957    0   30 1210  630 1690 2130 2090 3030 3240  460    0    0
```

```
1958   70   50   10 1090 1060 1690 2670  910 2750  700    0    0
1959    0  430  730  290 2300 1540 2080 2030 3910  280    0    0
1960    0  190  550  650 1230 1750 3750 5090 2190  700   90    0
1961    0    0  590  660 2190 5880 2150 4310 4030 1140    0    0
1962    0   20  120  880 2200 2690 2780 4770-9999  360   50    0
1963    0    0  610  600 1010 3480 2130 3410 1250  220  150    0
1964    0    0  740  910 3370 2600  630 2050 4700 1120   20    0
1965    0  250  660  780 2120 2700 2110 2810 2210 1350    0    0
1966    0  610  310  730 3340 1370 3100 4010 2020  510   40  110
1967    0    0   50  870 1810 1800 1540 1960 3270  150  130    0
1968   10   70   50  170 1770 2320 1140 2360 5140  700    0    0
1969    0   20  290  260 1850 1990 3430 2060 3470  290   30    0
1970    0   10   90 1150 2210 3620 2610 4310 1290  340   10    0
1971    0  730  570  740 2130 3580 4060 2100 3240  510   30  170
1972    0  550  460 1280 1040 3470 3250 3640 2980 2340   20    0
1973    0    0   10 1080 2650 1990 2460 2050 2090  190    0    0
1974  400    0   60 2160 1160 2520 3070 6110 1920  570  260    0
1975   20  350  360  410 2200 3340 3230 3560  940  520   20   40
1976    0  210  380 1700 1160 1460 2430 3720 3250  780   60    0
1977   20   10   90 1000  620  620 1470 3980 4010  100    0    0
1978    0  100  920  650 1710 3690 2960 4420 2190  110    0    0
1979   10  140   50  900 2000 4230 1230 2540 2910    0    0    0
1980    0   50  190 1040 1440 3490 3310  930 6130 1830  170   20
1981    0  210  220  720 2630 4730 2490 1750  610 1260   90    0
1982    0    0  290  840 1330 2160  390 4390 3400 1720  370    0
1983   90   30    0  550 1360 2700  830 5200 1380 1680    0    0
1984   10    0  350 1270 2030 2290 2900 3880 2130 1380  650    0
1985  380   50  170  860 1100 4270 1580 3350  900 1170    0    0
1986    0    0  120 1650 2120 2210 1830 2980 1760 1700  240   10
1987    0  110  290  360 1090 4210 2670 3640 2140 1040   20    0
1988    0   30   90 1170 1130 1790 1800 3580  740 1730    0    0
1989-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1990-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1991    0-9999  105  226 1370 2079 1452 4190 3799 1610-9999  321
1992  328  314  150  637 1968 1906 2366 4973 1287  238    0  216
1993-9999    5  476  768 2438 1169 3671 2463 2215   13   22-9999
1994    0  781  274  409 1837 2297 1625 5755 1709  216-9999-9999
1995    0  140  834  672 1556 1606 4439 2848  681  857   69    0
1996   12    2  660 2394 1566 1526 1960 3350 4843  724  476-9999
1997   35  321  458  642 1154 2832 1197 4071 1722 1800    0-9999
1998    0  346  154  241 2174 3348  813 2153 2231  276   85   37
1999   63    0  182 1025 3449 1207 3681 1570 2628  299  109-9999
2000-9999   95   48 2742 2816 1852 1725 1903 2903  391    0    0
```

Note that the Dec 1991 value is anomalous, but not as extreme as the 1945 datum,
which would get the same treatment with normals and climatologies, so should
produce an even bigger spike for 1945 DJF! Unless of course it's screened out by
the 4SD rule.. which it is! Well - no value in pre.1945.12.txt for this location.

Anyway.. this is the highest value in the Vietnam/Laos cells for Dec 1991:

```
     ROW     COL     LAT     LON     VAL    NORM
     198     571    9.25  105.75   63.50  130.00
```

With a normal of 130, that makes the anomaly -48.85. Now I'm confused. How can
an anomalously high value be well below the 61-90 mean? Aaarrgghhhh. Perhaps I
should look at the highest anomaly. That turns out to be 80, from here:

```
     216     563   18.25  101.75    1.80    1.00
```

Not exactly a show stopper. Time to look at the .glo files, which glo2abs processes
into absolutes. Here's a Far-Eastern region with a spike:

```
>> glod3(210:216,567:573)

             567          568          569          570          571          572          573
216       1393.6       1791.6       1757.4       1723.2       1674.5       1553.2       1431.9
215       1501.7       1899.8       1927.3       1893.1       1786.3         1665       1505.3
214       1609.9       2007.9       2097.2       2019.5       1885.4       1712.8       1540.2
213       1359.4       2116.1       2252.6       2092.9       1920.3       1747.7       1575.1
212       80.145       1195.5       1796.1         1882       1955.2       1782.6         1610
211       -6.125      -36.614       563.99       649.87       735.75       821.63        907.5
210      -59.833      -90.333      -89.649      -83.283      -76.929      -70.576      -64.223
```

The spike is at [213,569]. Yes, I know, it's the n-th set of coordinates. You should see the
plots! But looking at the anomalies is the closest we'll get to what Tim's program was doing,
ie, calculating DJF standard deviations. Or something. Now, the coordinates are 16.75N, 104.75E.
And wouldn't you know it, our prime suspect (see above) is on top of it:

```
4838300  1653  10472  138 MUKDAHAN             THAILAND      1934 2000   -999  -999.00
```

So OK, here we go with the full run-down for December 1991, in the 16.75N,105.75E region:

```
TYPE                VALUE     COMMENT
Raw data:             321     Highest unscreened December for this station (67 years)
Normal:                13     Looks right - of course, very low for the target data!
Anomaly:            2369.2    Correctly calculated
Gridded anomaly:    2252.6    Believable interpolation
Gridded actual:     er...     Strangely, it seems to be 0.
```

Ah well - had enough. It looks like it's an extreme but believable event in a Thai station, let's
leave it like that. Re-running precip, with the new updated database pre.0803271802.dtb:

```
<BEGIN_QUOTE>
crua6[/cru/cruts/version_3_0/primaries/precip] ./anomdtb

   > ***** AnomDTB: converts .dtb to anom .txt for gridding *****

   > Enter the suffix of the variable required:
.pre
   > Will calculate percentage anomalies.
   > Select the .cts or .dtb file to load:
pre.0803271802.dtb
   > Specify the start,end of the normals period:
1961,1990
   > Specify the missing percentage permitted:
25
   > Data required for a normal:           23
   > Specify the no. of stdevs at which to reject data:
4
   > Select outputs (1=.cts,2=.ann,3=.txt,4=.stn):
3
   > Check for duplicate stns after anomalising? (0=no,>0=km range)
0
   > Select the generic .txt file to save (yy.mm=auto):
pre.txt
   > Select the first,last years AD to save:
1901,2006
   > Operating...
 /tmp_mnt/cru-auto/cruts/version_3_0/primaries/precip/pre.0803271802.dtb
```

```
   > NORMALS            MEAN percent      STDEV percent
   >         .dtb    7315040    73.8
   >         .cts     299359     3.0    7613600    76.8
   > PROCESS        DECISION percent %of-chk
   > no lat/lon        17911     0.2     0.2
   > no normal       2355275    23.8    23.8
   > out-of-range      13249     0.1     0.2
   > accepted        7521017    75.9
   > Dumping years 1901-2006 to .txt files...

IDL> quick_interp_tdm2,1901,2006,'preglo/pregrid.',450,gs=0.5,dumpglo='dumpglo',pts_prefix='pretxt/pre.'
% Compiled module: QUICK_INTERP_TDM2.
% Compiled module: GLIMIT.
Defaults set
    1901
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: STRIP.
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
    1902
(etc)
    2006
IDL>

crua6[/cru/cruts/version_3_0/primaries/precip] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.pre
Enter a name for the gridded climatology file: clim.6190.lan.pre.grid4
Enter the path and stem of the .glo files: preglo/pregrid.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: preabs/
Now, CONCENTRATE. Addition or Percentage (A/P)? P
Do you wish to limit the output values? (Y/N): Y
1. Set minimum to zero
2. Set single minimum and maximum values
3. Set minima and maxima based on days in month
4. Set integer values >=1, (ie, positive)
5. Changed my mind, no limits
Choose: 1
Right, erm.. off I jolly well go!
pregrid.01.1901.glo
(etc)
pregrid.12.2006.glo

uealogin1[/cru/cruts/version_3_0/primaries/precip] ./makegrids
Welcome! This is the MAKEGRIDS program.
I will create decadal and full gridded files,
in both ASCII text and NetCDF formats, from
the output files of (eg) glo2abs.for.

Enter a gridfile with YYYY for year and MM for month: preabs/pregrid.YYYY.MM.glo.abs
Enter Start Year:  1901
Enter Start Month: 01
Enter End Year:    2006
Enter End Month:   12

Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE, and
ending with '.dat', eg: cru_ts_3_00.SSSS.EEEE.tmp.dat : cru_ts_3_00.SSSS.EEEE.pre.dat

Now please enter the 3-ch parameter code: pre
Enter a generic title for this dataset, eg:
CRU TS 3.00 Mean Temperature : CRU TS 3.00 Precipitation
Writing: cru_ts_3_00.1901.1910.pre.dat
         cru_ts_3_00.1901.1910.pre.nc
Writing: cru_ts_3_00.1911.1920.pre.dat
         cru_ts_3_00.1911.1920.pre.nc
Writing: cru_ts_3_00.1921.1930.pre.dat
         cru_ts_3_00.1921.1930.pre.nc
Writing: cru_ts_3_00.1931.1940.pre.dat
         cru_ts_3_00.1931.1940.pre.nc
Writing: cru_ts_3_00.1941.1950.pre.dat
         cru_ts_3_00.1941.1950.pre.nc
Writing: cru_ts_3_00.1951.1960.pre.dat
         cru_ts_3_00.1951.1960.pre.nc
Writing: cru_ts_3_00.1961.1970.pre.dat
         cru_ts_3_00.1961.1970.pre.nc
Writing: cru_ts_3_00.1971.1980.pre.dat
         cru_ts_3_00.1971.1980.pre.nc
Writing: cru_ts_3_00.1981.1990.pre.dat
         cru_ts_3_00.1981.1990.pre.nc
Writing: cru_ts_3_00.1991.2000.pre.dat
         cru_ts_3_00.1991.2000.pre.nc
Writing: cru_ts_3_00.2001.2006.pre.dat
         cru_ts_3_00.2001.2006.pre.nc
<END_QUOTE>

On to the reproduction of binaries for TMP and DTR, and subsequent regeneration of VAP and FRS.

TMP Binaries:

IDL> quick_interp_tdm2,1901,2006,'tmpbin/tmpbin',1200,gs=2.5,dumpbin='dumpbin',binfac=10,pts_prefix='tmp0km0705101334txt/tmp.'
Defaults set
    1901
% Compiled module: MEAN.
% Compiled module: MOMENT.
% Compiled module: STDDEV.
grid 1901 non-zero   -0.1127    0.9472    1.3993 cells=    46444
% Compiled module: STRIP.
% Compiled module: WRBIN.
    1902
grid 1902 non-zero   -0.4378    1.0267    1.4712 cells=    48321
    1903
grid 1903 non-zero   -0.3462    0.9586    1.3704 cells=    48803
    1904
grid 1904 non-zero   -0.4288    0.9753    1.3866 cells=    49493
    1905
grid 1905 non-zero   -0.2732    0.9620    1.3722 cells=    49845
    1906
grid 1906 non-zero   -0.1910    0.9176    1.3323 cells=    49568
    1907
grid 1907 non-zero   -0.5290    1.0338    1.4636 cells=    50458
    1908
grid 1908 non-zero   -0.3546    0.9165    1.2607 cells=    50449
    1909
```

```
grid 1909 non-zero   -0.4005    0.9998    1.4222 cells=    50163
    1910
grid 1910 non-zero   -0.3404    0.9459    1.3879 cells=    50489
    1911
grid 1911 non-zero   -0.3868    0.9445    1.3154 cells=    50752
    1912
grid 1912 non-zero   -0.4441    1.0156    1.4580 cells=    51731
    1913
grid 1913 non-zero   -0.3315    0.9125    1.2647 cells=    51634
    1914
grid 1914 non-zero   -0.1896    0.9655    1.4205 cells=    50734
    1915
grid 1915 non-zero   -0.1935    1.0501    1.5537 cells=    51773
    1916
grid 1916 non-zero   -0.3523    1.0092    1.4885 cells=    52042
    1917
grid 1917 non-zero   -0.5407    1.1470    1.6613 cells=    54151
    1918
grid 1918 non-zero   -0.4379    1.0228    1.4534 cells=    52579
    1919
grid 1919 non-zero   -0.3663    1.0327    1.5279 cells=    51872
    1920
grid 1920 non-zero   -0.2173    1.0092    1.4856 cells=    50540
    1921
grid 1921 non-zero   -0.1261    0.9004    1.2767 cells=    52468
    1922
grid 1922 non-zero   -0.2355    0.9258    1.3934 cells=    54202
    1923
grid 1923 non-zero   -0.1942    1.0040    1.4733 cells=    54975
    1924
grid 1924 non-zero   -0.1418    0.9416    1.3491 cells=    55664
    1925
grid 1925 non-zero   -0.1033    1.0121    1.5038 cells=    55677
    1926
grid 1926 non-zero    0.0471    0.9750    1.4242 cells=    55826
    1927
grid 1927 non-zero   -0.1290    0.9839    1.4396 cells=    57033
    1928
grid 1928 non-zero   -0.0254    0.9581    1.3929 cells=    56950
    1929
grid 1929 non-zero   -0.2651    1.1120    1.7327 cells=    58284
    1930
grid 1930 non-zero   -0.0233    1.0157    1.5554 cells=    57481
    1931
grid 1931 non-zero    0.0072    1.0705    1.6009 cells=    57932
    1932
grid 1932 non-zero    0.0407    1.0426    1.5664 cells=    57752
    1933
grid 1933 non-zero   -0.2517    1.1010    1.6794 cells=    59297
    1934
grid 1934 non-zero    0.0858    1.0705    1.6510 cells=    58932
    1935
grid 1935 non-zero   -0.0383    1.0498    1.5969 cells=    59316
    1936
grid 1936 non-zero   -0.0118    1.0867    1.6457 cells=    59676
    1937
grid 1937 non-zero    0.1841    1.0572    1.6419 cells=    59702
    1938
grid 1938 non-zero    0.2843    1.0094    1.4853 cells=    59478
    1939
grid 1939 non-zero    0.0828    1.0270    1.5633 cells=    60643
    1940
grid 1940 non-zero    0.1223    1.0033    1.5251 cells=    60381
    1941
grid 1941 non-zero    0.0049    1.0253    1.4988 cells=    63950
    1942
grid 1942 non-zero    0.0486    1.0061    1.5799 cells=    61984
    1943
grid 1943 non-zero    0.1795    1.0288    1.5243 cells=    63082
    1944
grid 1944 non-zero    0.1993    0.9783    1.4922 cells=    62327
    1945
grid 1945 non-zero   -0.0306    1.0840    1.5827 cells=    62977
    1946
grid 1946 non-zero   -0.0376    1.0094    1.4989 cells=    63193
    1947
grid 1947 non-zero    0.1326    1.0977    1.7075 cells=    64854
    1948
grid 1948 non-zero    0.0276    0.9783    1.4466 cells=    66490
    1949
grid 1949 non-zero   -0.0873    1.0422    1.5665 cells=    68159
    1950
grid 1950 non-zero   -0.2032    1.0344    1.5841 cells=    67736
    1951
grid 1951 non-zero   -0.0537    0.9777    1.4482 cells=    70202
    1952
grid 1952 non-zero   -0.0112    0.9952    1.5704 cells=    69668
    1953
grid 1953 non-zero    0.2020    1.0140    1.5735 cells=    70734
    1954
grid 1954 non-zero   -0.0062    1.0381    1.6387 cells=    71309
    1955
grid 1955 non-zero   -0.1527    1.0281    1.6167 cells=    73181
    1956
grid 1956 non-zero   -0.2183    1.0542    1.5788 cells=    77564
    1957
grid 1957 non-zero   -0.0142    0.9929    1.5067 cells=    77649
    1958
grid 1958 non-zero   -0.0257    1.0166    1.5491 cells=    80430
    1959
grid 1959 non-zero    0.0019    1.0058    1.5493 cells=    79903
    1960
grid 1960 non-zero   -0.0661    0.9628    1.4564 cells=    80353
    1961
grid 1961 non-zero    0.0016    0.9440    1.3960 cells=    81093
    1962
grid 1962 non-zero    0.0660    0.9249    1.4273 cells=    77733
    1963
grid 1963 non-zero   -0.0850    1.0164    1.5649 cells=    80869
    1964
grid 1964 non-zero   -0.3518    0.9639    1.4553 cells=    82284
    1965
grid 1965 non-zero   -0.2396    0.9097    1.3315 cells=    82512
    1966
grid 1966 non-zero   -0.2748    1.0171    1.5915 cells=    81405
    1967
grid 1967 non-zero   -0.0372    0.9385    1.4324 cells=    81573
    1968
grid 1968 non-zero   -0.2106    0.9665    1.5115 cells=    81706
    1969
```

```
grid 1969 non-zero   -0.1505    1.0751    1.7496 cells=    81490
    1970
grid 1970 non-zero   -0.0771    0.8111    1.1569 cells=    80462
    1971
grid 1971 non-zero   -0.1102    0.9128    1.3577 cells=    82451
    1972
grid 1972 non-zero   -0.1992    1.0147    1.5390 cells=    82070
    1973
grid 1973 non-zero    0.0914    0.9087    1.3303 cells=    81625
    1974
grid 1974 non-zero   -0.1369    0.9896    1.5273 cells=    81687
    1975
grid 1975 non-zero   -0.0400    0.9258    1.3720 cells=    81390
    1976
grid 1976 non-zero   -0.2595    0.9596    1.4088 cells=    82439
    1977
grid 1977 non-zero    0.0718    0.9855    1.5405 cells=    80143
    1978
grid 1978 non-zero   -0.0233    0.9729    1.5545 cells=    80118
    1979
grid 1979 non-zero   -0.0921    1.0054    1.6126 cells=    79714
    1980
grid 1980 non-zero    0.1600    0.9471    1.4078 cells=    80417
    1981
grid 1981 non-zero    0.4437    1.0207    1.5695 cells=    81226
    1982
grid 1982 non-zero   -0.0664    0.9502    1.4287 cells=    80230
    1983
grid 1983 non-zero    0.2325    0.9886    1.4907 cells=    82258
    1984
grid 1984 non-zero    0.0904    1.0216    1.6368 cells=    81431
    1985
grid 1985 non-zero    0.0625    0.9590    1.5123 cells=    81731
    1986
grid 1986 non-zero    0.1007    0.8952    1.3674 cells=    81016
    1987
grid 1987 non-zero    0.1116    0.9654    1.4412 cells=    84529
    1988
grid 1988 non-zero    0.3365    0.9242    1.3069 cells=    82070
    1989
grid 1989 non-zero    0.2451    1.0428    1.6170 cells=    79951
    1990
grid 1990 non-zero    0.4275    1.0581    1.5828 cells=    82418
    1991
grid 1991 non-zero    0.4279    0.9504    1.3005 cells=    80068
    1992
grid 1992 non-zero    0.0265    0.9827    1.4543 cells=    80204
    1993
grid 1993 non-zero    0.1614    0.9820    1.4965 cells=    78945
    1994
grid 1994 non-zero    0.2598    0.9699    1.3958 cells=    77509
    1995
grid 1995 non-zero    0.5222    1.0500    1.5943 cells=    80001
    1996
grid 1996 non-zero    0.3733    0.9745    1.4383 cells=    78304
    1997
grid 1997 non-zero    0.3947    1.0250    1.4473 cells=    80450
    1998
grid 1998 non-zero    0.6010    1.1707    1.5968 cells=    82794
    1999
grid 1999 non-zero    0.4249    1.0485    1.5105 cells=    80640
    2000
grid 2000 non-zero    0.4574    1.0256    1.4596 cells=    78883
    2001
grid 2001 non-zero    0.5695    1.0596    1.4876 cells=    78391
    2002
grid 2002 non-zero    0.6281    1.1662    1.6404 cells=    80200
    2003
grid 2003 non-zero    0.6715    1.0531    1.3475 cells=    77636
    2004
grid 2004 non-zero    0.5362    0.9762    1.2947 cells=    79600
    2005
grid 2005 non-zero    0.8050    1.1350    1.4605 cells=    80465
    2006
no stations found in: tmp0km0705101334txt/tmp.2006.11.txt
no stations found in: tmp0km0705101334txt/tmp.2006.12.txt
grid 2006 non-zero    0.6621    1.1017    1.5361 cells=    65396
IDL>
```

DTR Binaries:

```
IDL> quick_interp_tdm2,1901,2006,'dtrbin/dtrbin',750,gs=2.5,dumpbin='dumpbin',binfac=10,pts_prefix='dtrtxt/dtr.'
% Compiled module: QUICK_INTERP_TDM2.
% Compiled module: GLIMIT.
Defaults set
    1901
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: MEAN.
% Compiled module: MOMENT.
% Compiled module: STDDEV.
grid 1901 non-zero    0.3357    0.8741    1.1669 cells=    18526
% Compiled module: STRIP.
% Compiled module: WRBIN.
    1902
grid 1902 non-zero    0.2560    0.8530    1.1640 cells=    19088
    1903
grid 1903 non-zero    0.2015    0.8514    1.1841 cells=    19063
    1904
grid 1904 non-zero    0.2647    0.8584    1.1827 cells=    19155
    1905
grid 1905 non-zero    0.1753    0.9056    1.2595 cells=    20808
    1906
grid 1906 non-zero    0.2458    0.9003    1.2127 cells=    20892
    1907
grid 1907 non-zero    0.2658    0.9124    1.2370 cells=    21621
    1908
grid 1908 non-zero    0.3003    0.8911    1.1912 cells=    22028
    1909
grid 1909 non-zero    0.2063    0.8791    1.1952 cells=    22253
    1910
grid 1910 non-zero    0.2524    0.8563    1.1561 cells=    23297
    1911
grid 1911 non-zero    0.2317    0.8808    1.2201 cells=    24153
    1912
grid 1912 non-zero    0.1748    0.9544    1.3196 cells=    24284
    1913
grid 1913 non-zero    0.1990    0.9047    1.2463 cells=    24366
    1914
grid 1914 non-zero    0.1395    0.9195    1.2747 cells=    24431
```

```
    1915
grid 1915 non-zero    0.0185    0.9405    1.2969 cells=    26178
    1916
grid 1916 non-zero    0.0178    0.8761    1.1904 cells=    27095
    1917
grid 1917 non-zero    0.1518    0.9108    1.2619 cells=    26614
    1918
grid 1918 non-zero    0.1303    0.9134    1.2533 cells=    26447
    1919
grid 1919 non-zero    0.1300    0.8856    1.2029 cells=    25701
    1920
grid 1920 non-zero    0.0500    0.8480    1.1650 cells=    26563
    1921
grid 1921 non-zero    0.1660    0.8213    1.1246 cells=    26549
    1922
grid 1922 non-zero    0.1133    0.8468    1.1707 cells=    26701
    1923
grid 1923 non-zero    0.1794    0.8962    1.2293 cells=    27643
    1924
grid 1924 non-zero    0.1309    0.8549    1.1583 cells=    28642
    1925
grid 1925 non-zero    0.2252    0.9167    1.2254 cells=    28841
    1926
grid 1926 non-zero    0.1076    0.8559    1.1687 cells=    30671
    1927
grid 1927 non-zero    0.1285    0.8715    1.1766 cells=    30962
    1928
grid 1928 non-zero    0.1279    0.8576    1.1773 cells=    31156
    1929
grid 1929 non-zero    0.1784    0.8826    1.1974 cells=    32021
    1930
grid 1930 non-zero    0.1344    0.8711    1.1830 cells=    33360
    1931
grid 1931 non-zero    0.0238    0.8470    1.1548 cells=    32726
    1932
grid 1932 non-zero    0.0872    0.8489    1.1546 cells=    33396
    1933
grid 1933 non-zero    0.1012    0.8560    1.1597 cells=    34574
    1934
grid 1934 non-zero    0.0295    0.8591    1.1676 cells=    34203
    1935
grid 1935 non-zero    0.1092    0.8682    1.1665 cells=    34561
    1936
grid 1936 non-zero    0.2106    0.8947    1.2055 cells=    35632
    1937
grid 1937 non-zero    0.1686    0.8463    1.1182 cells=    35921
    1938
grid 1938 non-zero    0.0845    0.8242    1.1125 cells=    35718
    1939
grid 1939 non-zero    0.1428    0.8362    1.1277 cells=    37525
    1940
grid 1940 non-zero    0.1891    0.8662    1.1493 cells=    38227
    1941
grid 1941 non-zero    0.1334    0.8502    1.1411 cells=    38486
    1942
grid 1942 non-zero    0.1176    0.8344    1.1001 cells=    39312
    1943
grid 1943 non-zero    0.1844    0.8476    1.1191 cells=    40361
    1944
grid 1944 non-zero    0.1564    0.8195    1.0904 cells=    40010
    1945
grid 1945 non-zero    0.1566    0.8164    1.0740 cells=    40172
    1946
grid 1946 non-zero    0.1648    0.8529    1.1288 cells=    40305
    1947
grid 1947 non-zero    0.0668    0.8305    1.1166 cells=    41045
    1948
grid 1948 non-zero    0.2204    0.8015    1.0473 cells=    41004
    1949
grid 1949 non-zero    0.1192    0.8185    1.1029 cells=    41329
    1950
grid 1950 non-zero    0.1349    0.7915    1.0675 cells=    42803
    1951
grid 1951 non-zero    0.1881    0.8003    1.0647 cells=    47064
    1952
grid 1952 non-zero    0.1751    0.7776    1.0556 cells=    46868
    1953
grid 1953 non-zero    0.1331    0.7763    1.0518 cells=    48129
    1954
grid 1954 non-zero    0.0824    0.7668    1.0349 cells=    47912
    1955
grid 1955 non-zero    0.0963    0.7733    1.0468 cells=    49105
    1956
grid 1956 non-zero    0.0901    0.7695    1.0440 cells=    50637
    1957
grid 1957 non-zero    0.0689    0.7536    1.0279 cells=    50456
    1958
grid 1958 non-zero    0.0050    0.7504    1.0291 cells=    52013
    1959
grid 1959 non-zero    0.0439    0.7329    1.0020 cells=    52162
    1960
grid 1960 non-zero    0.0674    0.7049    0.9542 cells=    52787
    1961
grid 1961 non-zero    0.0445    0.6810    0.9111 cells=    56188
    1962
grid 1962 non-zero    0.1297    0.6877    0.9156 cells=    54897
    1963
grid 1963 non-zero    0.1449    0.7088    0.9661 cells=    55755
    1964
grid 1964 non-zero    0.0955    0.6719    0.9029 cells=    54909
    1965
grid 1965 non-zero    0.0913    0.6638    0.8950 cells=    54906
    1966
grid 1966 non-zero    0.0878    0.6566    0.8813 cells=    54751
    1967
grid 1967 non-zero    0.0805    0.6626    0.8876 cells=    54393
    1968
grid 1968 non-zero    0.0826    0.6611    0.8923 cells=    54602
    1969
grid 1969 non-zero    0.0253    0.6787    0.9309 cells=    55176
    1970
grid 1970 non-zero    0.0576    0.6232    0.8301 cells=    55444
    1971
grid 1971 non-zero    0.0987    0.6340    0.8426 cells=    54610
    1972
grid 1972 non-zero    0.0472    0.6631    0.8979 cells=    55812
    1973
grid 1973 non-zero    0.0287    0.6424    0.8741 cells=    54755
    1974
grid 1974 non-zero   -0.0119    0.6782    0.9289 cells=    56105
```

```
    1975
grid 1975 non-zero    0.0287    0.6259    0.8458 cells=    54696
    1976
grid 1976 non-zero    0.0740    0.6565    0.8966 cells=    55239
    1977
grid 1977 non-zero   -0.0454    0.6600    0.9026 cells=    54227
    1978
grid 1978 non-zero   -0.1045    0.6529    0.8775 cells=    55036
    1979
grid 1979 non-zero   -0.0749    0.6510    0.8753 cells=    54990
    1980
grid 1980 non-zero   -0.0565    0.6269    0.8300 cells=    55430
    1981
grid 1981 non-zero   -0.0498    0.6704    0.8970 cells=    55023
    1982
grid 1982 non-zero   -0.0828    0.6622    0.8874 cells=    56028
    1983
grid 1983 non-zero   -0.1380    0.6808    0.9142 cells=    55854
    1984
grid 1984 non-zero   -0.1530    0.6675    0.8850 cells=    55751
    1985
grid 1985 non-zero   -0.1156    0.6359    0.8438 cells=    55066
    1986
grid 1986 non-zero   -0.0822    0.6412    0.8583 cells=    55111
    1987
grid 1987 non-zero   -0.0579    0.6599    0.8919 cells=    55607
    1988
grid 1988 non-zero   -0.0691    0.6615    0.8752 cells=    54503
    1989
grid 1989 non-zero   -0.0165    0.6685    0.9064 cells=    55615
    1990
grid 1990 non-zero   -0.0655    0.6734    0.9057 cells=    54497
    1991
grid 1991 non-zero   -0.0899    0.6602    0.8722 cells=    53533
    1992
grid 1992 non-zero   -0.1171    0.6951    0.9349 cells=    53739
    1993
grid 1993 non-zero   -0.1314    0.6882    0.9188 cells=    51271
    1994
grid 1994 non-zero   -0.0480    0.6934    0.9282 cells=    50211
    1995
grid 1995 non-zero   -0.0357    0.7596    1.2498 cells=    52366
    1996
grid 1996 non-zero   -0.0723    0.7666    1.1579 cells=    51667
    1997
grid 1997 non-zero   -0.1340    0.7854    1.1636 cells=    52195
    1998
grid 1998 non-zero   -0.1627    0.8328    1.1992 cells=    52907
    1999
grid 1999 non-zero   -0.1237    0.7134    0.9716 cells=    51050
    2000
grid 2000 non-zero   -0.1765    0.7575    1.0412 cells=    51137
    2001
grid 2001 non-zero   -0.1305    0.7564    1.0658 cells=    49146
    2002
grid 2002 non-zero   -0.0984    0.7549    1.0967 cells=    46178
    2003
grid 2003 non-zero   -0.1072    0.7128    1.0087 cells=    46904
    2004
grid 2004 non-zero   -0.1628    0.8113    1.2574 cells=    47399
    2005
grid 2005 non-zero   -0.1150    0.8546    1.3540 cells=    43715
    2006
no stations found in: dtrtxt/dtr.2006.09.txt
no stations found in: dtrtxt/dtr.2006.10.txt
no stations found in: dtrtxt/dtr.2006.11.txt
no stations found in: dtrtxt/dtr.2006.12.txt
grid 2006 non-zero   -0.0087    0.9041    1.6077 cells=    28592
IDL>

VAP synthetics:

IDL> vap_gts_anom,dtr_prefix='../dtrbin/dtrbin',tmp_prefix='../tmpbin/tmpbin',1901,2006,outprefix='vapsyn/vapsyn.',dumpbin=1
% Compiled module: VAP_GTS_ANOM.
% Compiled module: RDBIN.
% Compiled module: STRIP.
% Compiled module: DEFXYZ.
Land,sea:       56016       68400
Calculating tmn normal
% Compiled module: TVAP.
Calculating synthetic vap normal
% Compiled module: ESAT.
Calculating synthetic anomalies
% Compiled module: MOMENT.
1901 vap (x,s2,<<,>>):   -0.0450600     0.230521     -4.50663      4.92927
% Compiled module: WRBIN.
1902 vap (x,s2,<<,>>):    -0.102633     0.271200     -4.14974      4.92341
1903 vap (x,s2,<<,>>):    -0.107597     0.242152     -5.74305      5.58190
1904 vap (x,s2,<<,>>):    -0.123137     0.221801     -4.30042      3.71240
1905 vap (x,s2,<<,>>):   -0.0799978     0.267905     -4.54584      6.04190
1906 vap (x,s2,<<,>>):   -0.0343380     0.240282     -4.86007      6.45160
1907 vap (x,s2,<<,>>):    -0.137421     0.284412     -5.06625      6.02255
1908 vap (x,s2,<<,>>):    -0.105214     0.234139     -6.99258      5.15916
1909 vap (x,s2,<<,>>):    -0.103285     0.252210     -5.48791      5.12214
1910 vap (x,s2,<<,>>):    -0.104377     0.225462     -4.64360      7.19087
1911 vap (x,s2,<<,>>):    -0.109270     0.255412     -4.35549      6.99751
1912 vap (x,s2,<<,>>):    -0.127306     0.287857     -4.63037      5.84822
1913 vap (x,s2,<<,>>):    -0.107747     0.271437     -4.73097      4.98110
1914 vap (x,s2,<<,>>):   -0.0481493     0.274125     -4.43114      6.19318
1915 vap (x,s2,<<,>>):   -0.0343964     0.332448     -5.31914      6.04190
1916 vap (x,s2,<<,>>):   -0.0947238     0.293859     -6.31417      4.63692
1917 vap (x,s2,<<,>>):    -0.170714     0.396007     -8.69476      3.86362
1918 vap (x,s2,<<,>>):    -0.133214     0.311174     -5.48035      6.19711
1919 vap (x,s2,<<,>>):   -0.0687798     0.307508     -6.14869      9.28416
1920 vap (x,s2,<<,>>):   -0.0619862     0.262944     -5.76398      4.10598
1921 vap (x,s2,<<,>>):   -0.0319013     0.301257     -5.73133      7.07691
1922 vap (x,s2,<<,>>):   -0.0621843     0.231719     -5.03104      4.18061
1923 vap (x,s2,<<,>>):   -0.0626035     0.285980     -5.70781      7.32851
1924 vap (x,s2,<<,>>):   -0.0720660     0.281999     -4.88423      7.32851
1925 vap (x,s2,<<,>>):   -0.0673457     0.292218     -5.70781      8.66163
1926 vap (x,s2,<<,>>):   -0.0264075     0.297545     -4.88423      6.07568
1927 vap (x,s2,<<,>>):   -0.0400396     0.277274     -4.26864      9.46017
1928 vap (x,s2,<<,>>):   -0.0376556     0.250591     -4.88423      4.32885
1929 vap (x,s2,<<,>>):   -0.0885709     0.313796     -5.52846      5.63952
1930 vap (x,s2,<<,>>):   -0.0261154     0.252065     -4.93317      5.76016
1931 vap (x,s2,<<,>>):   0.00818994     0.340814     -5.13082      6.25571
1932 vap (x,s2,<<,>>): -0.000556678     0.290796     -4.36993      6.39834
1933 vap (x,s2,<<,>>):   -0.0732750     0.299654     -4.85280      5.68388
1934 vap (x,s2,<<,>>):   -0.0230768     0.306463     -4.27863      6.18267
1935 vap (x,s2,<<,>>):   -0.0294056     0.256452     -4.08567      5.28612
```

```
1936 vap (x,s2,<<,>>):   -0.0166144     0.312561     -4.76875      6.14991
1937 vap (x,s2,<<,>>):   0.00628125     0.273150     -4.94706      8.10473
1938 vap (x,s2,<<,>>):    0.0538196     0.277858     -3.83216      5.16185
1939 vap (x,s2,<<,>>):   -0.0101522     0.261258     -4.26058      6.09505
1940 vap (x,s2,<<,>>):   -0.0340492     0.270330     -5.28004      4.99802
1941 vap (x,s2,<<,>>):  -0.00272590     0.359517     -4.30990      7.86635
1942 vap (x,s2,<<,>>):   -0.0107379     0.256225     -4.31849      5.04972
1943 vap (x,s2,<<,>>):  -0.00847952     0.306959     -4.05096      5.09181
1944 vap (x,s2,<<,>>):    0.0135632     0.286842     -6.32264      5.44941
1945 vap (x,s2,<<,>>):   -0.0348352     0.339675     -5.77718      6.11337
1946 vap (x,s2,<<,>>):   -0.0308651     0.304056     -6.39576      6.89301
1947 vap (x,s2,<<,>>):    0.0119494     0.347709     -4.52023      7.91578
1948 vap (x,s2,<<,>>):  -0.00279501     0.269904     -4.20986      8.68827
1949 vap (x,s2,<<,>>):   -0.0392110     0.341484     -5.71766      6.29158
1950 vap (x,s2,<<,>>):   -0.0805553     0.315878     -4.60512      9.89166
1951 vap (x,s2,<<,>>):   -0.0455985     0.286749     -5.21167      5.09294
1952 vap (x,s2,<<,>>):   -0.0285279     0.310278     -6.19114      5.91316
1953 vap (x,s2,<<,>>):    0.0164626     0.328090     -5.10127      6.91089
1954 vap (x,s2,<<,>>):   -0.0270003     0.325794     -4.60512      5.86057
1955 vap (x,s2,<<,>>):   -0.0427618     0.309960     -6.32375      5.67290
1956 vap (x,s2,<<,>>):    -0.141589     0.324435     -5.33854      4.30108
1957 vap (x,s2,<<,>>):   -0.0322534     0.290902     -6.51552      5.47146
1958 vap (x,s2,<<,>>):  -0.00254215     0.284311     -4.56239      5.24883
1959 vap (x,s2,<<,>>):   -0.0136386     0.273182     -4.56765      7.10322
1960 vap (x,s2,<<,>>):   -0.0268859     0.251590     -4.74454      5.61688
1961 vap (x,s2,<<,>>):   0.00428266     0.256116     -3.58335      5.09660
1962 vap (x,s2,<<,>>):   -0.0169736     0.241635     -5.13969      4.88660
1963 vap (x,s2,<<,>>):   -0.0132022     0.280225     -4.86808      6.25239
1964 vap (x,s2,<<,>>):   -0.0801179     0.230050     -4.37340      5.66511
1965 vap (x,s2,<<,>>):   -0.0818651     0.255070     -4.92358      5.56393
1966 vap (x,s2,<<,>>):   -0.0280720     0.231274     -5.94548      5.55636
1967 vap (x,s2,<<,>>):   -0.0181908     0.260895     -4.50197      5.60940
1968 vap (x,s2,<<,>>):   -0.0724171     0.263899     -7.34842      4.57032
1969 vap (x,s2,<<,>>):   -0.0251546     0.277463     -5.45023      5.64616
1970 vap (x,s2,<<,>>):   -0.0310875     0.183015     -5.20684      4.72542
1971 vap (x,s2,<<,>>):   -0.0584375     0.247623     -4.25373      5.45411
1972 vap (x,s2,<<,>>):   -0.0703370     0.291676     -4.58312      5.34509
1973 vap (x,s2,<<,>>):    0.0333626     0.249423     -4.92304      4.40218
1974 vap (x,s2,<<,>>):   -0.0588555     0.269075     -5.21312      7.79163
1975 vap (x,s2,<<,>>):   -0.0141657     0.248957     -5.68768      6.99932
1976 vap (x,s2,<<,>>):    -0.122256     0.269952     -6.63680      3.87281
1977 vap (x,s2,<<,>>):    0.0307466     0.242061     -4.42668      4.59490
1978 vap (x,s2,<<,>>):   -0.0234150     0.225811     -5.53217      5.65761
1979 vap (x,s2,<<,>>):    0.0121905     0.234581     -3.81796      8.11386
1980 vap (x,s2,<<,>>):    0.0243552     0.225838     -4.95252      10.0661
1981 vap (x,s2,<<,>>):    0.0730062     0.273207     -6.21452      5.02121
1982 vap (x,s2,<<,>>):   0.00410785     0.225016     -4.32441      5.82923
1983 vap (x,s2,<<,>>):    0.0990013     0.349267     -5.66664      7.11964
1984 vap (x,s2,<<,>>):    0.0240819     0.257495     -4.72244      5.45411
1985 vap (x,s2,<<,>>):   0.00794266     0.232222     -4.08511      5.98364
1986 vap (x,s2,<<,>>):    0.0247494     0.226368     -4.53393      6.49533
1987 vap (x,s2,<<,>>):    0.0702606     0.326509     -6.49176      5.65761
1988 vap (x,s2,<<,>>):     0.108340     0.262510     -4.71399      5.24402
1989 vap (x,s2,<<,>>):    0.0558202     0.259338     -5.43051      5.26850
1990 vap (x,s2,<<,>>):     0.142205     0.288092     -5.08627      4.54985
1991 vap (x,s2,<<,>>):     0.123237     0.286160     -4.73092      7.81615
1992 vap (x,s2,<<,>>):   0.00923000     0.309573     -4.60065      7.87882
1993 vap (x,s2,<<,>>):    0.0506631     0.270131     -4.38424      5.70831
1994 vap (x,s2,<<,>>):     0.119075     0.311620     -3.41067      5.64707
1995 vap (x,s2,<<,>>):     0.153732     0.311293     -5.12451      6.16339
1996 vap (x,s2,<<,>>):    0.0641272     0.281972     -7.07159      5.34453
1997 vap (x,s2,<<,>>):     0.171295     0.432609     -6.05247      7.83205
1998 vap (x,s2,<<,>>):     0.316044     0.580134     -5.70816      6.81637
1999 vap (x,s2,<<,>>):     0.154989     0.339337     -5.84341      5.00767
2000 vap (x,s2,<<,>>):     0.151733     0.320928     -5.67183      5.19904
2001 vap (x,s2,<<,>>):     0.207467     0.427043     -5.60383      10.6589
2002 vap (x,s2,<<,>>):     0.224772     0.487041     -10.9627      10.6589
2003 vap (x,s2,<<,>>):     0.230009     0.420027     -4.31180      6.60522
2004 vap (x,s2,<<,>>):     0.196860     0.370951     -6.88193      9.91745
2005 vap (x,s2,<<,>>):     0.294053     0.485357     -10.8553      7.02645
2006 vap (x,s2,<<,>>):     0.193683     0.411812     -6.28115      9.19209
IDL>
```

VAP Gridding:

```
IDL> quick_interp_tdm2,1901,2006,'vapglo/vap.',1000,gs=0.5,dumpglo='dumpglo',synth_prefix='vapsyn/vapsyn.',pts_prefix='vaptxt/vap.'
% Compiled module: QUICK_INTERP_TDM2.
% Compiled module: GLIMIT.
Defaults set
    1901
% Compiled module: RDBIN.
% Compiled module: STRIP.
% Compiled module: DEFXYZ.
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
    1902
(etc)
    2006
IDL>
```

VAP Gridded Absolutes:

```
crua6[/cru/cruts/version_3_0/secondaries/vap] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.vap
Enter a name for the gridded climatology file: clim.6190.lan.vap.grid4
Enter the path and stem of the .glo files: vapglo/vap.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: vapabs/
Now, CONCENTRATE. Addition or Percentage (A/P)? A
Do you wish to limit the output values? (Y/N): Y
1. Set minimum to zero
2. Set single minimum and maximum values
3. Set minima and maxima based on days in month
4. Set integer values >=1, (ie, positive)
5. Changed my mind, no limits
Choose: 4
Right, erm.. off I jolly well go!
vap.01.1901.glo
(etc)
vap.12.2006.glo
```

VAP Output Files:

```
uealogin1[/cru/cruts/version_3_0/secondaries/vap] ./makegrids
Welcome! This is the MAKEGRIDS program.
I will create decadal and full gridded files,
in both ASCII text and NetCDF formats, from
the output files of (eg) glo2abs.for.

Enter a gridfile with YYYY for year and MM for month: vapabs/vap.YYYY.MM.glo.abs
Enter Start Year:  1901
Enter Start Month: 01
Enter End Year:    2006
Enter End Month:   12

Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE, and
ending with '.dat', eg: cru_ts_3_00.SSSS.EEEE.tmp.dat : cru_ts_3_00.SSSS.EEEE.vap.dat

Now please enter the 3-ch parameter code: vap
Enter a generic title for this dataset, eg:
CRU TS 3.00 Mean Temperature : CRU TS 3.00 Vapour Pressure
Writing: cru_ts_3_00.1901.1910.vap.dat
         cru_ts_3_00.1901.1910.vap.nc
Writing: cru_ts_3_00.1911.1920.vap.dat
         cru_ts_3_00.1911.1920.vap.nc
Writing: cru_ts_3_00.1921.1930.vap.dat
         cru_ts_3_00.1921.1930.vap.nc
Writing: cru_ts_3_00.1931.1940.vap.dat
         cru_ts_3_00.1931.1940.vap.nc
Writing: cru_ts_3_00.1941.1950.vap.dat
         cru_ts_3_00.1941.1950.vap.nc
Writing: cru_ts_3_00.1951.1960.vap.dat
         cru_ts_3_00.1951.1960.vap.nc
Writing: cru_ts_3_00.1961.1970.vap.dat
         cru_ts_3_00.1961.1970.vap.nc
Writing: cru_ts_3_00.1971.1980.vap.dat
         cru_ts_3_00.1971.1980.vap.nc
Writing: cru_ts_3_00.1981.1990.vap.dat
         cru_ts_3_00.1981.1990.vap.nc
Writing: cru_ts_3_00.1991.2000.vap.dat
         cru_ts_3_00.1991.2000.vap.nc
Writing: cru_ts_3_00.2001.2006.vap.dat
         cru_ts_3_00.2001.2006.vap.nc

FRS synthetics:

IDL> frs_gts,dtr_prefix='../dtrbin/dtrbin',tmp_prefix='../tmpbin/tmpbin',1901,2006,outprefix='frssyn/frssyn'
% Compiled module: RDBIN.
% Compiled module: STRIP.
filesize=     6220800
gridsize=     0.500000
% Compiled module: DEFXYZ.
Calculating synthetic frs normal
    1961
filesize=      248832
gridsize=      2.50000
(etc)
    1990
filesize=      248832
gridsize=      2.50000
% Compiled module: DAYS.
Calculating synthetic anomalies
    1901
filesize=      248832
gridsize=      2.50000
(etc)
    2006
filesize=      248832
gridsize=      2.50000
IDL>

FRS gridding:

IDL> quick_interp_tdm2,1901,2006,'frsgrid/frsgrid',750,gs=0.5,dumpglo='dumpglo',nostn=1,synth_prefix='frssyn/frssyn'
% Compiled module: QUICK_INTERP_TDM2.
% Compiled module: GLIMIT.
Defaults set
    1901
% Compiled module: RDBIN.
% Compiled module: STRIP.
ls: frssyn/frssyn1901 not found
ls: frssyn/frssyn1901.Z not found
found: frssyn/frssyn1901.gz
% Compiled module: DEFXYZ.
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
    1902
ls: frssyn/frssyn1902 not found
ls: frssyn/frssyn1902.Z not found
found: frssyn/frssyn1902.gz
(etc)
    2006
ls: frssyn/frssyn2006 not found
ls: frssyn/frssyn2006.Z not found
found: frssyn/frssyn2006.gz
IDL>



VAP Problems again.. this time it's too high **sigh**

Tim suggested the 'synthfac' parameter in quick_interp_tdm2. The note for it says:

; multi factor to obtain synth file actual values

..so I reckon it should be 0.1 - but was wrong. The note is misleading at best, since the actual code
looks like this:

   dummygrid=dummygrid/synthfac

..so the factor should be 10:

IDL> quick_interp_tdm2,1901,2006,'vapglo/vap.',1000,gs=0.5,dumpglo='dumpglo',synthfac=10,synth_prefix='vapsyn/vapsyn.',pts_prefix='vaptxt/vap.'
% Compiled module: QUICK_INTERP_TDM2.
% Compiled module: GLIMIT.
Defaults set
    1901
% Compiled module: RDBIN.
% Compiled module: STRIP.
% Compiled module: DEFXYZ.
```

```
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
    1902
(etc)
    2006
IDL>

crua6[/cru/cruts/version_3_0/secondaries/vap] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.vap
Enter a name for the gridded climatology file: clim.6190.lan.vap.grid9
Enter the path and stem of the .glo files: vapglo/vap.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: vapabs/
Now, CONCENTRATE. Addition or Percentage (A/P)? A
Do you wish to limit the output values? (Y/N): Y
1. Set minimum to zero
2. Set single minimum and maximum values
3. Set minima and maxima based on days in month
4. Set integer values >=1, (ie, positive)
5. Changed my mind, no limits
Choose: 1
Right, erm.. off I jolly well go!
vap.1901.01.glo
(etc)
vap.2006.12.glo

uealogin1[/cru/cruts/version_3_0/secondaries/vap] ./makegrids
Welcome! This is the MAKEGRIDS program.
I will create decadal and full gridded files,
in both ASCII text and NetCDF formats, from
the output files of (eg) glo2abs.for.

Enter a gridfile with YYYY for year and MM for month: vapabs/vap.YYYY.MM.glo.abs
Enter Start Year:  1901
Enter Start Month: 01
Enter End Year:    2006
Enter End Month:   12

Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE, and
ending with '.dat', eg: cru_ts_3_00.SSSS.EEEE.tmp.dat : cru_ts_3_00.SSSS.EEEE.vap.dat

Now please enter the 3-ch parameter code: vap
Enter a generic title for this dataset, eg:
CRU TS 3.00 Mean Temperature : CRU TS 3.00 Vapour Pressure
Writing: cru_ts_3_00.1901.1910.vap.dat
         cru_ts_3_00.1901.1910.vap.nc
Writing: cru_ts_3_00.1911.1920.vap.dat
         cru_ts_3_00.1911.1920.vap.nc
Writing: cru_ts_3_00.1921.1930.vap.dat
         cru_ts_3_00.1921.1930.vap.nc
Writing: cru_ts_3_00.1931.1940.vap.dat
         cru_ts_3_00.1931.1940.vap.nc
Writing: cru_ts_3_00.1941.1950.vap.dat
         cru_ts_3_00.1941.1950.vap.nc
Writing: cru_ts_3_00.1951.1960.vap.dat
         cru_ts_3_00.1951.1960.vap.nc
Writing: cru_ts_3_00.1961.1970.vap.dat
         cru_ts_3_00.1961.1970.vap.nc
Writing: cru_ts_3_00.1971.1980.vap.dat
         cru_ts_3_00.1971.1980.vap.nc
Writing: cru_ts_3_00.1981.1990.vap.dat
         cru_ts_3_00.1981.1990.vap.nc
Writing: cru_ts_3_00.1991.2000.vap.dat
         cru_ts_3_00.1991.2000.vap.nc
Writing: cru_ts_3_00.2001.2006.vap.dat
         cru_ts_3_00.2001.2006.vap.nc


Re-doing FRS as well:

IDL> quick_interp_tdm2,1901,2006,'frsgrid/frsgrid',750,gs=0.5,dumpglo='dumpglo',nostn=1,synthfac=10,synth_prefix='frssyn/frssyn'
% Compiled module: QUICK_INTERP_TDM2.
% Compiled module: GLIMIT.
Defaults set
    1901
% Compiled module: RDBIN.
% Compiled module: STRIP.
% Compiled module: DEFXYZ.
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
    1902
(etc)
    2006
IDL>



Also re-doing WET/RD0:

IDL> quick_interp_tdm2,1901,2006,'prebin/prebin',450,gs=2.5,dumpbin='dumpbin',binfac=10,pts_prefix='pretxt/pre.'
% Compiled module: QUICK_INTERP_TDM2.
% Compiled module: GLIMIT.
Defaults set
    1901
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: MEAN.
% Compiled module: MOMENT.
% Compiled module: STDDEV.
grid 1901 non-zero   -0.2703   38.0775   72.6735 cells=    36537
% Compiled module: STRIP.
% Compiled module: WRBIN.
    1902
grid 1902 non-zero   -1.8369   38.8160  105.0312 cells=    37346
(etc)
    2006
no stations found in: pretxt/pre.2006.09.txt
no stations found in: pretxt/pre.2006.10.txt
no stations found in: pretxt/pre.2006.11.txt
```

```
no stations found in: pretxt/pre.2006.12.txt
grid 2006 non-zero    2.2381   38.8997   70.0935 cells=    32663
IDL>

There then followed a produciton run for WET, resulting in unrealistic, banded output.
This was tracked down to the sythetic gridder, rd0_gts_tdm.pro, using half-degree
normals with a 2.5-degree output. So it was modified to read the 2.5 normals, and rerun:

IDL> .compile /cru/cruts/version_3_0/BADC_AREA/programs/idl/rd0_gts_tdm.pro
% Compiled module: RD0_GTS.
IDL> rd0_gts,1901,2006,1961,1990,outprefix='rd0syn/rd0syn',pre_prefix='../prebin/prebin'
Reading precip and rd0 normals
% Compiled module: RDBIN.
% Compiled module: STRIP.
yes
filesize=      248832
gridsize=      2.50000
% Compiled module: DEFXYZ.
yes
filesize=      248832
gridsize=      2.50000
% Compiled module: DAYS.
Calculating synthetic Rd0 normal
    1961
filesize=      248832
gridsize=      2.50000
(etc)
    1990
filesize=      248832
gridsize=      2.50000
Calculating synthetic anomalies
    1901
filesize=      248832
gridsize=      2.50000
% Compiled module: WRBIN.
    1902
(etc)
    2006
filesize=      248832
gridsize=      2.50000
% Program caused arithmetic error: Floating divide by 0
% Program caused arithmetic error: Floating illegal operand
IDL>

..which is what happened last time. And, again - all synthetics produced, apparently
OK. I think it's just the last few empty months of 2006..

IDL> quick_interp_tdm2,1901,2006,'rd0pcglo/rd0pc',450,gs=0.5,dumpglo='dumpglo',synth_prefix='rd0syn/rd0syn',synthfac=10,pts_prefix='rd0pctxt/rd0pc.'
% Compiled module: QUICK_INTERP_TDM2.
% Compiled module: GLIMIT.
Defaults set
    1901
% Compiled module: RDBIN.
% Compiled module: STRIP.
% Compiled module: DEFXYZ.
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
    1902
(etc)
    2006
IDL>

crua6[/cru/cruts/version_3_0/secondaries/wet] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.wet
Enter a name for the gridded climatology file: clim.6190.lan.wet.grid4
Enter the path and stem of the .glo files: rd0pcglo/rd0pc.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: rd0pcabs
Now, CONCENTRATE. Addition or Percentage (A/P)? P
Do you wish to limit the output values? (Y/N): Y
1. Set minimum to zero
2. Set single minimum and maximum values
3. Set minima and maxima based on days in month
4. Set integer values >=1, (ie, positive)
5. Changed my mind, no limits
Choose: 3
Right, erm.. off I jolly well go!
rd0pcglo/rd0pc.01.1901.glo
rd0pcglo/rd0pc.1901.01.glo
rd0pc.1901.01.glo
(etc)
rd0pc.2006.12.glo

uealogin1[/cru/cruts/version_3_0/secondaries/wet] ./makegrids
Welcome! This is the MAKEGRIDS program.
I will create decadal and full gridded files,
in both ASCII text and NetCDF formats, from
the output files of (eg) glo2abs.for.

Enter a gridfile with YYYY for year and MM for month: rd0pcabs/rd0pc.YYYY.MM.glo.abs
Enter Start Year:  1901
Enter Start Month: 01
Enter End Year:    2006
Enter End Month:   12

Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE, and
ending with '.dat', eg: cru_ts_3_00.SSSS.EEEE.tmp.dat : cru_ts_3_00.SSSS.EEEE.wet.dat

Now please enter the 3-ch parameter code: wet
Enter a generic title for this dataset, eg:
CRU TS 3.00 Mean Temperature : CRU TS 3.00 Rain Days
Writing: cru_ts_3_00.1901.1910.wet.dat
         cru_ts_3_00.1901.1910.wet.nc
Writing: cru_ts_3_00.1911.1920.wet.dat
         cru_ts_3_00.1911.1920.wet.nc
Writing: cru_ts_3_00.1921.1930.wet.dat
         cru_ts_3_00.1921.1930.wet.nc
Writing: cru_ts_3_00.1931.1940.wet.dat
         cru_ts_3_00.1931.1940.wet.nc
Writing: cru_ts_3_00.1941.1950.wet.dat
         cru_ts_3_00.1941.1950.wet.nc
Writing: cru_ts_3_00.1951.1960.wet.dat
```

```
         cru_ts_3_00.1951.1960.wet.nc
Writing: cru_ts_3_00.1961.1970.wet.dat
         cru_ts_3_00.1961.1970.wet.nc
Writing: cru_ts_3_00.1971.1980.wet.dat
         cru_ts_3_00.1971.1980.wet.nc
Writing: cru_ts_3_00.1981.1990.wet.dat
         cru_ts_3_00.1981.1990.wet.nc
Writing: cru_ts_3_00.1991.2000.wet.dat
         cru_ts_3_00.1991.2000.wet.nc
Writing: cru_ts_3_00.2001.2006.wet.dat
         cru_ts_3_00.2001.2006.wet.nc



VAP - three stations deleted with unbelievable data:
6450000    45    942   15 LIBREVILLE/LEON MBA  GABON         1979 2007    -999    -999
6451000   208   1148  599 BITAM                GABON         1971 2007    -999    -999
6275000  1400   3233  378 ED DUEIM             SUDAN         1971 2007    -999    -999

VAP then re-run with new database vap.0804231150.dtb:

crua6[/cru/cruts/version_3_0/secondaries/vap] ./anomdtb

   > ***** AnomDTB: converts .dtb to anom .txt for gridding *****

   > Enter the suffix of the variable required:
.vap
   > Select the .cts or .dtb file to load:
vap.0804231150.dtb
   > Specify the start,end of the normals period:
1961,1990
   > Specify the missing percentage permitted:
25
   > Data required for a normal:           23
   > Specify the no. of stdevs at which to reject data:
3
   > Select outputs (1=.cts,2=.ann,3=.txt,4=.stn):
3
   > Check for duplicate stns after anomalising? (0=no,>0=km range)
0
   > Select the generic .txt file to save (yy.mm=auto):
vap.txt
   > Select the first,last years AD to save:
1901,2006
   > Operating...

   > NORMALS            MEAN percent      STDEV percent
   >         .dtb     907894    45.2
   >         .cts      35390     1.8     943284    47.0
   > PROCESS        DECISION percent %of-chk
   > no lat/lon          105     0.0     0.0
   > no normal       1064261    53.0    53.0
   > out-of-range         49     0.0     0.0
   > accepted         943235    47.0
   > Dumping years 1901-2006 to .txt files...

IDL> quick_interp_tdm2,1901,2006,'vapglo/vap.',1000,gs=0.5,dumpglo='dumpglo',synthfac=10,synth_prefix='vapsyn/vapsyn.',pts_prefix='vaptxt/vap.'
% Compiled module: QUICK_INTERP_TDM2.
% Compiled module: GLIMIT.
Defaults set
    1901
% Compiled module: RDBIN.
% Compiled module: STRIP.
% Compiled module: DEFXYZ.
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
    2006
IDL>

crua6[/cru/cruts/version_3_0/secondaries/vap] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.vap
Enter a name for the gridded climatology file: clim.6190.lan.vap.grid2
Enter the path and stem of the .glo files: vapglo/vap.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: vapabs/
Now, CONCENTRATE. Addition or Percentage (A/P)? A
Do you wish to limit the output values? (Y/N): Y
1. Set minimum to zero
2. Set single minimum and maximum values
3. Set minima and maxima based on days in month
4. Set integer values >=1, (ie, positive)
5. Changed my mind, no limits
Choose: 1
Right, erm.. off I jolly well go!
vapglo/vap.01.1901.glo
vapglo/vap.1901.01.glo
vap.1901.01.glo
(etc)
vap.2006.12.glo

uealogin1[/cru/cruts/version_3_0/secondaries/vap] ./makegrids
Welcome! This is the MAKEGRIDS program.
I will create decadal and full gridded files,
in both ASCII text and NetCDF formats, from
the output files of (eg) glo2abs.for.

Enter a gridfile with YYYY for year and MM for month: vapabs/vap.YYYY.NN.glo.abs
Enter a gridfile with YYYY for year and MM for month: vapabs/vap.YYYY.MM.glo.abs
Enter Start Year:  1901
Enter Start Month: 01
Enter End Year:    2006
Enter End Month:   12

Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE, and
ending with '.dat', eg: cru_ts_3_00.SSSS.EEEE.tmp.dat : cru_ts_3_00.SSSS.EEEE.vap.dat

Now please enter the 3-ch parameter code: vap
Enter a generic title for this dataset, eg:
CRU TS 3.00 Mean Temperature : CRU TS 3.00 Vapour Pressure
Writing: cru_ts_3_00.1901.1910.vap.dat
         cru_ts_3_00.1901.1910.vap.nc
Writing: cru_ts_3_00.1911.1920.vap.dat
```

```
         cru_ts_3_00.1911.1920.vap.nc
Writing: cru_ts_3_00.1921.1930.vap.dat
         cru_ts_3_00.1921.1930.vap.nc
Writing: cru_ts_3_00.1931.1940.vap.dat
         cru_ts_3_00.1931.1940.vap.nc
Writing: cru_ts_3_00.1941.1950.vap.dat
         cru_ts_3_00.1941.1950.vap.nc
Writing: cru_ts_3_00.1951.1960.vap.dat
         cru_ts_3_00.1951.1960.vap.nc
Writing: cru_ts_3_00.1961.1970.vap.dat
         cru_ts_3_00.1961.1970.vap.nc
Writing: cru_ts_3_00.1971.1980.vap.dat
         cru_ts_3_00.1971.1980.vap.nc
Writing: cru_ts_3_00.1981.1990.vap.dat
         cru_ts_3_00.1981.1990.vap.nc
Writing: cru_ts_3_00.1991.2000.vap.dat
         cru_ts_3_00.1991.2000.vap.nc
Writing: cru_ts_3_00.2001.2006.vap.dat
         cru_ts_3_00.2001.2006.vap.nc


Next round of problems..

Well, VAP finally looks OK, which is good news.

WET looks better, but variability is still too low. It's complicated by the synthetic elements in
combination with percentage anomalies!

I found that I could examine the 'mark New binary files' with Matlab, using (ie):

>> fid = fopen('glo25.rd0.6190','r');
>> [d,c] = fread(fid,inf,'int16');
>> whos
  Name      Size         Bytes  Class

  c         1x1              8  double array
  d    124416x1         995328  double array
  fid       1x1              8  double array

Grand total is 124418 elements using 995344 bytes

>> c

c =

      124416

>> min(d)

ans =

     0

>> max(d)

ans =

   303

>> hmean(d)

ans =

   123.7939

So we can deduce that the rd0 2.5 degree normals are in days*10. Similarly for the others of interest:

FILE                  MIN            MAX            MEAN            UNITS
glo25.rd0.6190          0            303             106          days*10
glo25.pre.6190          0            391              21              mm?
glo.rd0.norm            0            310             124          days*10
glo.pre.norm            0           1244              58              mm
clim.6190.lan.wet.grid  0           3090            1018         days*100
clim.6190.lan.pre.grid  0          12430             574            mm*10

My guess is that glo25.pre.6190 has a lower max because the wider coverage of each cell is squashing
the peaks. So.. an experimental run with half-degree synthetics!

First, generate half-degree binaries for pre:

IDL> quick_interp_tdm2,1901,2006,'prebin05/prebin05.',450,gs=0.5,dumpbin='dumpbin',binfac=10,pts_prefix='pretxt/pre.'
% Compiled module: QUICK_INTERP_TDM2.
% Compiled module: GLIMIT.
Defaults set
    1901
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: MEAN.
% Compiled module: MOMENT.
% Compiled module: STDDEV.
grid 1901 non-zero   -0.0524   38.0853   75.9241 cells=   921834
(etc)

Then re-generate synthetic rd0 on a half-degree grid:

IDL> rd0_gts,1901,2006,1961,1990,outprefix='rd0syn05/rd0syn.',pre_prefix='../prebin05/prebin05.'
Reading precip and rd0 normals
yes
filesize=      248832
gridsize=      2.50000
yes
filesize=      248832
gridsize=      2.50000
Calculating synthetic Rd0 normal
    1961
yes
filesize=     6220800
gridsize=     0.500000
(etc.. as before but with appropriately larger numbers)
    2006
filesize=     6220800
gridsize=     0.500000
% Program caused arithmetic error: Floating divide by 0
% Program caused arithmetic error: Floating illegal operand
IDL>

Then the gridding, using the new half-degree synthetics:

IDL> quick_interp_tdm2,1901,2006,'rd0pcglo05/rd0pc05.',450,gs=0.5,dumpglo='dumpglo',synth_prefix='rd0syn05/rd0syn.',synthfac=10,pts_prefix='rd0pctxt/rd0pc.'
% Compiled module: QUICK_INTERP_TDM2.
```

```
% Compiled module: GLIMIT.
Defaults set
    1901
% Compiled module: RDBIN.
% Compiled module: STRIP.
% Compiled module: DEFXYZ.
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
    1902
(etc)

RIGHT, stop all that.. Tim O has recalculated the 2.5-degree binary normals for PRE (from half degree)
and WET (from TS 2.1). So.. time to try out the OTHER synthetic rd0 generator, the one that reads
precip anomalies:

IDL> .compile ../../BADC_AREA/programs/idl/rd0_gts_anom.pro
% Compiled module: RD0_GTS_ANOM.
IDL> .compile ../../BADC_AREA/programs/idl/
% Error opening file. File: ../../BADC_AREA/programs/idl/
  No such file or directory
IDL> .compile ../../BADC_AREA/programs/idl/rdbin.pro
% Compiled module: RDBIN.
IDL> rd0_gts_anom,1901,2006,1961,1990,outprefix='rd0syn/rd0syn.',pre_prefix='../prebin/prebin'
Reading precip and rd0 normals
% Compiled module: STRIP.
% Compiled module: DEFXYZ.
% Compiled module: DAYS.
Calculating synthetic rd0 normal
% Compiled module: RD0CAL.
Calculating synthetic rd0 anomalies
IDL>

IDL> quick_interp_tdm2,1901,2006,'rd0glo/rd0.',450,gs=0.5,dumpglo='dumpglo',synth_prefix='rd0syn/rd0syn.',synthfac=10,pts_prefix='rd0pctxt/rd0pc.'
% Compiled module: QUICK_INTERP_TDM2.
% Compiled module: GLIMIT.
Defaults set
    1901
% Compiled module: RDBIN.
% Compiled module: STRIP.
% Compiled module: DEFXYZ.
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
    1902
(etc)
    2006
IDL>

crua6[/cru/cruts/version_3_0/secondaries/wet] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.wet
Enter a name for the gridded climatology file: clim.6190.lan.wet.grid7
Enter the path and stem of the .glo files: rd0glo/rd0.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: rd0abs/
Now, CONCENTRATE. Addition or Percentage (A/P)? P
Do you wish to limit the output values? (Y/N): Y
1. Set minimum to zero
2. Set single minimum and maximum values
3. Set minima and maxima based on days in month
4. Set integer values >=1, (ie, positive)
5. Changed my mind, no limits
Choose: 3
Right, erm.. off I jolly well go!
rd0glo/rd0.01.1901.glo
rd0glo/rd0.1901.01.glo
rd0.01.1901.glo
(etc)
rd0.12.2006.glo

uealogin1[/cru/cruts/version_3_0/secondaries/wet] ./makegrids
Welcome! This is the MAKEGRIDS program.
I will create decadal and full gridded files,
in both ASCII text and NetCDF formats, from
the output files of (eg) glo2abs.for.

Enter a gridfile with YYYY for year and MM for month: rd0abs/rd0.YYYY.MM.glo.abs
Enter Start Year:  1901
Enter Start Month: 01
Enter End Year:    2006
Enter End Month:   12

Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE, and
ending with '.dat', eg: cru_ts_3_00.SSSS.EEEE.tmp.dat : cru_ts_3_00.SSSS.EEEE.wet.dat

Now please enter the 3-ch parameter code: wet
Enter a generic title for this dataset, eg:
CRU TS 3.00 Mean Temperature : CRU TS 3.00 Wet Days
Writing: cru_ts_3_00.1901.1910.wet.dat
         cru_ts_3_00.1901.1910.wet.nc
Writing: cru_ts_3_00.1911.1920.wet.dat
         cru_ts_3_00.1911.1920.wet.nc
Writing: cru_ts_3_00.1921.1930.wet.dat
         cru_ts_3_00.1921.1930.wet.nc
Writing: cru_ts_3_00.1931.1940.wet.dat
         cru_ts_3_00.1931.1940.wet.nc
Writing: cru_ts_3_00.1941.1950.wet.dat
         cru_ts_3_00.1941.1950.wet.nc
Writing: cru_ts_3_00.1951.1960.wet.dat
         cru_ts_3_00.1951.1960.wet.nc
Writing: cru_ts_3_00.1961.1970.wet.dat
         cru_ts_3_00.1961.1970.wet.nc
Writing: cru_ts_3_00.1971.1980.wet.dat
         cru_ts_3_00.1971.1980.wet.nc
Writing: cru_ts_3_00.1981.1990.wet.dat
         cru_ts_3_00.1981.1990.wet.nc
Writing: cru_ts_3_00.1991.2000.wet.dat
         cru_ts_3_00.1991.2000.wet.nc
Writing: cru_ts_3_00.2001.2006.wet.dat
         cru_ts_3_00.2001.2006.wet.nc

Wrong again! The saga continues.. actually I'm beginning to wonder if it'll still be going when I JOIN SAGA.
```

```
This time, the 'real' areas have variability 10x too low, and the 'synthetic' areas have variability sqrt(10)
too low. The latter can be explained by the binary precip being in %age anoms *10, so rd0_gts_anom.pro modified
to divide by 1000 when calculating (instead of 100). Example (from the normals calculation):

Before:
 pregrd(nland)=((pregrd(nland)/100.0)+1.0)*prenorm(nland) ; make pre anom into abs

After:
 pregrd(nland)=((pregrd(nland)/1000.0)+1.0)*prenorm(nland) ; make pre anom into abs (mm)

'Synthfac=10' will also not be needed in the final gridding, that should take care of the 'real' area variability.

So..

IDL> .compile ../../BADC_AREA/programs/idl/rd0_gts_anom.pro
% Compiled module: RD0_GTS_ANOM.
IDL> .compile ../../BADC_AREA/programs/idl/rdbin.pro
% Compiled module: RDBIN.
IDL> rd0_gts_anom,1901,2006,1961,1990,outprefix='rd0syn/rd0syn.',pre_prefix='../prebin/prebin'
Reading precip and rd0 normals
% Compiled module: STRIP.
% Compiled module: DEFXYZ.
% Compiled module: DAYS.
Calculating synthetic rd0 normal
% Compiled module: RD0CAL.
Calculating synthetic rd0 anomalies
% Compiled module: WRBIN.
IDL>

IDL> quick_interp_tdm2,1901,2006,'rd0glo/rd0.',450,gs=0.5,dumpglo='dumpglo',synth_prefix='rd0syn/rd0syn.',pts_prefix='rd0pctxt/rd0pc.'

<snip!>

That didn't work, real areas 10x too small (synth areas OK though). So..

IDL> quick_interp_tdm2,1901,2006,'rd0glo/rd0.',anomfac=10,450,gs=0.5,dumpglo='dumpglo',synth_prefix='rd0syn/rd0syn.',pts_prefix='rd0pctxt/rd0pc.'
% Compiled module: QUICK_INTERP_TDM2.
% Compiled module: GLIMIT.
Defaults set
    1901
% Compiled module: RDBIN.
% Compiled module: STRIP.
% Compiled module: DEFXYZ.
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
    1902
(etc)
    2006
IDL>

crua6[/cru/cruts/version_3_0/secondaries/wet] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.wet
Enter a name for the gridded climatology file: clim.6190.lan.wet.grid9
Enter the path and stem of the .glo files: rd0glo/rd0.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: rd0abs/
Now, CONCENTRATE. Addition or Percentage (A/P)? P
Do you wish to limit the output values? (Y/N): Y
1. Set minimum to zero
2. Set single minimum and maximum values
3. Set minima and maxima based on days in month
4. Set integer values >=1, (ie, positive)
5. Changed my mind, no limits
Choose: 3
Right, erm.. off I jolly well go!
rd0glo/rd0.01.1901.glo
rd0glo/rd0.1901.01.glo
rd0.1901.01.glo
(etc)
rd0.2006.12.glo

uealogin1[/cru/cruts/version_3_0/secondaries/wet] ./makegrids
Welcome! This is the MAKEGRIDS program.
I will create decadal and full gridded files,
in both ASCII text and NetCDF formats, from
the output files of (eg) glo2abs.for.

Enter a gridfile with YYYY for year and MM for month: rd0abs/rd0.YYYY.MM.glo.abs
Enter Start Year:  1901
Enter Start Month: 01
Enter End Year:    2006
Enter End Month:   12

Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE, and
ending with '.dat', eg: cru_ts_3_00.SSSS.EEEE.tmp.dat : cru_ts_3_00.SSSS.EEEE.wet.dat

Now please enter the 3-ch parameter code: wet
Enter a generic title for this dataset, eg:
CRU TS 3.00 Mean Temperature : CRU TS 3.00 Rain Days
Writing: cru_ts_3_00.1901.1910.wet.dat
         cru_ts_3_00.1901.1910.wet.nc
Writing: cru_ts_3_00.1911.1920.wet.dat
         cru_ts_3_00.1911.1920.wet.nc
Writing: cru_ts_3_00.1921.1930.wet.dat
         cru_ts_3_00.1921.1930.wet.nc
Writing: cru_ts_3_00.1931.1940.wet.dat
         cru_ts_3_00.1931.1940.wet.nc
Writing: cru_ts_3_00.1941.1950.wet.dat
         cru_ts_3_00.1941.1950.wet.nc
Writing: cru_ts_3_00.1951.1960.wet.dat
         cru_ts_3_00.1951.1960.wet.nc
Writing: cru_ts_3_00.1961.1970.wet.dat
         cru_ts_3_00.1961.1970.wet.nc
Writing: cru_ts_3_00.1971.1980.wet.dat
         cru_ts_3_00.1971.1980.wet.nc
Writing: cru_ts_3_00.1981.1990.wet.dat
         cru_ts_3_00.1981.1990.wet.nc
Writing: cru_ts_3_00.1991.2000.wet.dat
         cru_ts_3_00.1991.2000.wet.nc
Writing: cru_ts_3_00.2001.2006.wet.dat
         cru_ts_3_00.2001.2006.wet.nc
```

Hmmm.. still some problems. In several areas, including a swathe of Russia, the mean values drop
around 1991 - just when MCDW comes in.

Targeted the area and found several candidates:

uealogin1[/cru/cruts/version_3_0/db/rd0] ./getllstations

GETCELLSTATIONS

Extracts stations from a CRU TS Database.

Please define the bounding box in tenths of a
degree, ie 55 degrees North would be 550:
  Southern Edge (-900 to 899): 570
  Northern Edge (570 to 900): 630
  Western Edge (-1800 to 1799): 900
  Eastern Edge (900 to 1800): 1050

Enter the CRU TS Database file: wet.0710161148.dtb

And finally, an output filename: wet.0710161148.57_63N.90_105E.dat

Done. Found   7 matching stations out of   6143

2388400 6160   9000   63 BOR   RUSSIA (ASIA) 1936 2007   -999   -999
2389100 6167   9637 261 BAJKIT   RUSSIA (ASIA) 1936 2007   -999   -999
2490800 6033 10227 260 VANAVARA   RUSSIA (ASIA) 1936 2007   -999   -999
2926300 5845   9215   78 JENISEJSK   RUSSIA (ASIA) 1936 2007   -999   -999
2928200 5842   9740 134 BOGUCANY   RUSSIA (ASIA) 1936 2007   -999   -999
2398600 6047   9302 521 SEVERO-JENISEJSK   RUSSIA (ASIA) 2004 2007   -999   0
2937900 5720   9488 147 TASEJEVA RIVER   RUSSIA (ASIA) 2004 2007   -999   0

The last two are too short to have any meaning. The second and third have missing data over
the entire period of concern. That leaves BOR, JENISEJSK and BOGUCANY, the latter of which
we'll examine closer. Here's the series, lifted directly from wet.0710161148.dtb:

2928200 5842   9740 134 BOGUCANY   RUSSIA (ASIA) 1936 2007   -999   -999
6190-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1936 1200 500 800 500 1100 1400 900 1200 400 1300 1600 1700
1937 1400 1000 1300 400 800 1200 1100 1300 1800 2000 1300 1000
1938 1100 1100 500 1000 1000 800 600 1300 1400 2300 1800 1500
1939 1200 700 900 700 1500 1300 1100 1600 800 1900 1800 1900
1940 800 1500 500 1400 1400 1300 1400 1600 1500 2400 2000 1600
1941 2200 1500 1800 1100 1400 1300 1700 1600 900 1600 1900 1700
1942 1100 1600 1500 1100 1000 1900 600 1600 500 2000 1700 1900
1943 1000 1000 800   0 1500 1100 1000 1100 1200 2000 2100 2100
1944 1300 1100 1100 400 1400 1200-9999 1100 1500 1600 2000 2900
1945 2100 900 1000-9999 900 2100 1700 1200 900 1700 1900 1600
1946 2400 1400 1500 1000 1800 1700 1300 1800 700 1700 2200 2300
1947 1400 1300 1700 1500 900 600 1000 700 2000 600 1300 1200
1948 1700 1100 900 1100 1100 1100 1100 1900 1400 1300 1200 1500
1949 2100 1100 1000 700 1900 1600 800 1500 1600 1100 1600 1200
1950 1500 1100 800 800 1400 600 600 800 1600 1500 1900 2500
1951 1100 1200 1400 500 1000 1400 1200 2000 1100 1400 1100 1400
1952 1500 1200 1100 700 1600 1100 1300 1200 1200 2100 1200 1300
1953 1200 600 1300 700 800 800 1100 1100 1400 2100 1500 2100
1954 1900 1300 1100 1100 800 400 1700 1100 1300 1800 2000 1500
1955 1100 1600 1100 1000 1500 1400 1000 1100 1500 1400 1600 2000
1956 1800 1200 1000 1200 800 900 1900 800 1100 2100 800 1200
1957 1200 300 700 1200 1300 900 1300 1200 1700 1700 1900 2200
1958 2000 1000 1200 900 1400 1100 800 1000 1200 2000 2200 1900
1959 2100 1000 900 1400 1800 700 1600 1300 1300 1600 2300 1900
1960 1700 1800 1000 1600 1000 1500 1400 1500 2300 1100 1900 1200
1961 1700 1400 600 500 1000 1400 1400 1700 1800 1500 1600 1800
1962 1200 1300 700 700 800 1000 1300 1200 1200 900 2100 2000
1963 900 600 700 800 1300 1000 1300 1300 1400 1300 1100 2100
1964 1000 800 1200 500 800 1400 1000 800 700 1900 1000 2300
1965 1500 1100 800 500 1000 1100 800 1500 1900 1200 1600 1400
1966 2100 2100 1400 1000 1800 1200 1000 1000 1200 2100 2700 2400
1967 1800 1700 900 900 1200 900 1100 1100 800 1100 1300 1700
1968 1000 1500 800 900 800 1700 600 1200 1600 800 2200 1400
1969 1400 1500 1000 1500 1300 1300 1100 1100 1400 1200 1500 1700
1970 2000 1300 200 1000 1100 900 1100 1700 900 1500 1500 1800
1971 2100 900 1300 700 1200 500 1000 1600 2000 1400 1600 1600
1972 1600 1600 800 500 1400 1200 600 1700 1600 1500 2400 2300
1973 1400 1200 1600 1600 1200 900 1000 1400 600 800 1700 1800
1974 1800 1400 900 1100 1500 1400 1000 1500 1500 1700 2100 2200
1975 2600 1500 1100 1000 1200 1200 1100 1700 1300 1200 2200 1200
1976 1300 1400 1000 600 1300 600 700 1300 1800 1900 1900 1800
1977 1800 1100 2000 900 1400 1400 900 1700 1000 1600 2500 1600
1978 2500 1100 1000 1300 900 1400 1100 1300 1300 1500 1600 2000
1979 2100 1700 1900 1200 1200 600 600 1500 1000 2200 2400 1500
1980 1700 1100 600 700 900 1100 1300 800 1300 1500 2100 2200
1981 1600 1800 1100 1200 1300 700 1100 1200 1200 1300 1300 1800
1982 1300 1000 1500 800 1200 1300 1700 1400 1700 1800 2800 2500
1983 2200 1200 1400 1700 1200 1000 700 1600 1200 1200 1900 2400
1984 1400 1900 1000 800 1300 700 700 200 400 1500 2400 2100
1985-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1986-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1987-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1988-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1989-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1990-9999-9999-9999-9999 1000 1100 300 700 1000 700 800 400
1991-9999 400 400 600 800 900 1200 600 800 700 300 700
1992 600 100 200 800 500-9999-9999 700 500 200 1000 600
1993 100 200 200 200 700 1300 100 1100 900 800 300 900
1994 700 200 300 500 1000 700 300 600 1400 900 300 900
1995 900 900 600 900 500 1100 800   0 1000 100 1100 400
1996 600 100 300 1000 600 300 200 1100 1100 600 1000 1500
1997 700 500 600 400 600 1200 500 1500 700 1100 900 1000
1998 700 500   0 1000 1100 1000-9999 900 1300 1500 900 1600
1999 900 400 200 700 200 900 700 900 600 1000 800 700
2000 400 700 600 500 1400 1000 700 600 900 1000 900 1200
2001 400 700 700 300 1100 1000 1300 400 1000 900 900 1000
2002 800 1100 600 600 400 1500-9999 1100 900 700 1100 500
2003 800 800 400 600 500 200 400 900 1200 900 1100 500
2004 500 900 1000 600 800 900 800 1000 1500 1200 900 400
2005 700 300 100 900 600 1500 900 600 900 1600 800 500
2006 600 600 900 400 700 500 500 1000 900 1100 800 1300
2007 500 1000 900 200 1200 900 700-9999-9999-9999-9999-9999

You can see that the data after 1990 are for some months significantly lower than the
period before.. which would be the period the normals would be based on! I used Matlab
to calculate the normals for this series:

6190 1667 1342 1063 933 1172 1080 980 1288 1272 1412 1852 1840

I then look in the 1995 anomaly files, rd0pctxt/rd0pc.1995.[01-12].txt.

Tabulated process for 1995:

raw	norm	anom
900	1666.7	-4.60
900	1341.7	-3.29
600	1062.5	-4.35
900	933.33	-0.36
500	1172	-5.73
1100	1080	0.19
800	980	-1.84
0	1288	-10.00
1000	1272	-2.14
100	1412	-9.29
1100	1852	-4.06
400	1840	-7.83

They aren't percentage anomalies! They are percentage anomalies /10. This could explain why
the real data areas had variability 10x too low. BUT it shouldn't be - they should be
regular percentage anomalies! This whole process is too convoluted and created myriad
problems of this kind. I really think we should change it.

Back on the case. I need to find where the post-1990 data came from for these three stations. I already know
the geneaology of the database:

```
wet.0311061611.dtb
        +
rdy.0709111032.dtb  (MCDW composite)
        +
rdy.0710151817.dtb  (CLIMAT composite with metadata added)
        V
        V
wet.0710161148.dtb
```

I was going to do further backtracing, but it's been revealed that the same issues were in 2.1 - meaning that
I didn't add the duff data. The suggested way forward is to not use any observations after 1989, but to allow
synthetics to take over. I'm not keen on this approach as it's likely (imo) to introduce visible jumps at 1990,
since we're effectively introducing a change of data source just after calculating the normals. My compromise is
to try it - but to also try a straight derivation from half-degree synthetics.

So, first, we need synthetic-only from 1990 onwards, that can be married with the existing glos from pre-1990.

Actually, we might as well produce a full series of gridded syn-only rd0. Hell, we can do both options in one go!

No point in using the final gridding routine, rd0_gts_anom can produce glo files itself, let's give it a go.

Well - not straightforward. rd0_gts_anom.pro is quite resistant to the idea that it might produce half-degree
synthetics, to the point where I'm really not sure what's left to modify! Eventually found it.. the .glo saving
routine takes a second argument which is a code for the grid size. Because just giving it the grid size just
wouldn't be the same, would it? Here it is:

```
     SaveGlo,23,rd0month,CallFile=Savefile,CallTitle=SaveTitle
```

Now that 23 is the key, but you have to look in quick_interp_tdm2.pro to decode it:

```
    if (gs[0] eq 0.5) then SaveGrid=12
    if (gs[0] eq 2.5) then SaveGrid=22
    if (gs[0] eq 5.0) then SaveGrid=23
```

So actually, this was saving with a gridsize of 5 degrees! Disquietingly, this isn't born out by the file sizes,
but we'll gloss over that. So, with '23' changed to '12', we have rd0_gts_anom_05.pro.

Produced half-degree synthetic-only WET:

```
IDL> .compile ../../BADC_AREA/programs/idl/rd0_gts_anom_05.pro
% Compiled module: RD0_GTS_ANOM.
IDL> .compile ../../BADC_AREA/programs/idl/rdbin.pro
% Compiled module: RDBIN.
IDL> rd0_gts_anom,1901,2006,1961,1990,outprefix='rd0syn05glo/rd0syn05.',pre_prefix='../prebin05/prebin05.'
Reading precip and rd0 normals
% Compiled module: STRIP.
% Compiled module: DEFXYZ.
% Compiled module: DAYS.
Calculating synthetic rd0 normal
% Compiled module: RD0CAL.
Calculating synthetic rd0 anomalies
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
IDL>

crua6[/cru/cruts/version_3_0/secondaries/wet] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.wet
Enter a name for the gridded climatology file: clim.6190.lan.wet.grid14
Enter the path and stem of the .glo files: rd0syn05glo/rd0syn05.
Enter the starting year: 1901
Enter the ending year:   2006
Enter the path (if any) for the output files: rd0syn05abs/
Now, CONCENTRATE. Addition or Percentage (A/P)? P
Do you wish to limit the output values? (Y/N): Y
1. Set minimum to zero
2. Set single minimum and maximum values
3. Set minima and maxima based on days in month
4. Set integer values >=1, (ie, positive)
5. Changed my mind, no limits
Choose: 3
Right, erm.. off I jolly well go!
rd0syn05glo/rd0syn05.01.1901.glo
rd0syn05glo/rd0syn05.1901.01.glo
rd0syn05.1901.01.glo
(etc)
rd0syn05.2006.12.glo
```

There was then some copying around of decades' worth chunks of .abs files, to make a set with obs/syn to 1989
and syn from 1990 onwards. Then:

```
uealogin1[/cru/cruts/version_3_0/secondaries/wet] ./makegrids
Welcome! This is the MAKEGRIDS program.
I will create decadal and full gridded files,
in both ASCII text and NetCDF formats, from
the output files of (eg) glo2abs.for.

Enter a gridfile with YYYY for year and MM for month: rd0abs_mixed/rd0.YYYY.MM.glo.abs.gz
Enter Start Year:  1901
Enter Start Month: 01
Enter End Year:    2006

Please enter a sample OUTPUT filename, replacing
```

```
start year with SSSS and end year with EEEE, and
ending with '.dat', eg: cru_ts_3_00.SSSS.EEEE.tmp.dat : cru_ts_3_00.SSSS.EEEE.wet.dat

Now please enter the 3-ch parameter code: wet
Enter a generic title for this dataset, eg:
CRU TS 3.00 Mean Temperature : CRU TS 3.00 Rain Days synth 1990 on
Writing: cru_ts_3_00.1901.1910.wet.dat
         cru_ts_3_00.1901.1910.wet.nc
Writing: cru_ts_3_00.1911.1920.wet.dat
         cru_ts_3_00.1911.1920.wet.nc
Writing: cru_ts_3_00.1921.1930.wet.dat
         cru_ts_3_00.1921.1930.wet.nc
Writing: cru_ts_3_00.1931.1940.wet.dat
         cru_ts_3_00.1931.1940.wet.nc
Writing: cru_ts_3_00.1941.1950.wet.dat
         cru_ts_3_00.1941.1950.wet.nc
Writing: cru_ts_3_00.1951.1960.wet.dat
         cru_ts_3_00.1951.1960.wet.nc
Writing: cru_ts_3_00.1961.1970.wet.dat
         cru_ts_3_00.1961.1970.wet.nc
Writing: cru_ts_3_00.1971.1980.wet.dat
         cru_ts_3_00.1971.1980.wet.nc
Writing: cru_ts_3_00.1981.1990.wet.dat
         cru_ts_3_00.1981.1990.wet.nc
Writing: cru_ts_3_00.1991.2000.wet.dat
         cru_ts_3_00.1991.2000.wet.nc
Writing: cru_ts_3_00.2001.2006.wet.dat
         cru_ts_3_00.2001.2006.wet.nc

This is the dataset we're going with.


Now, we're on to STATISTICS. Specifically, the ones needed for the paper. They include:

Cell coverage - the percentage of cells in each region that have stations, are near stations, or are not covered;

Station counts - the number of contributing stations in each region;

Data sources - the observed/synthetic split for secondary parameters.

All statistics will be required at a yearly timestep and broken down by region (ie, Europe, Africa, Asia..).


Cell coverage would ideally come from the IDL gridder - however, as we know, the approach of quick_interp_tdm2.pro
does not lend itself to revealing such information! The best solution will be to use the paired cell/cdd station
counts, as produced by stncounts.for.

Station counts should be straightforward to derive from the anomaly files (.txt), as output by anomdtb.f90. This,
however, will only work for Primary parameters, since Secondaries are driven from synthetic data as well. Further,
the synthetic element in this is usually at 2.5 degrees, so a direct relationship with half-degree coverage will
be hard to establish.

Data sources will not be easy (see Station counts above). One approach could be to analyse the anomaly files for
the Primary parameter(s), and make the assumption that their half-degree coverage will carry through (via the
2.5-degree synthetic stage and the gridding) to the final gridded data.

Actually, I think the most logical approach is to produce secondary station files that just record the observed
contributions (as opposed to the derived ones). Users will be free to use these in tandem with the appropriate
primary counts, which they can assume will have 'contributed' to the unfilled cells but to a less reliable extent
(both because of the indirect derivation and the lower resolution).

So - about time we had a drains-up on the update procedures. I've already established a logical hierarchy:

/cru/cruts/final_structure
/cru/cruts/final_structure/database                  * Repository for databases, might need subdirs
/cru/cruts/final_structure/incoming
/cru/cruts/final_structure/incoming/BOM
/cru/cruts/final_structure/incoming/CLIMAT
/cru/cruts/final_structure/incoming/MCDW
/cru/cruts/final_structure/incoming/other
/cru/cruts/final_structure/primary
/cru/cruts/final_structure/primary/tmp
/cru/cruts/final_structure/primary/tmp/txt
/cru/cruts/final_structure/primary/tmp/glo
/cru/cruts/final_structure/primary/tmp/abs
/cru/cruts/final_structure/primary/tmp/stn
/cru/cruts/final_structure/primary/tmp/stn/cdd0
/cru/cruts/final_structure/primary/tmp/stn/cddn
/cru/cruts/final_structure/primary/pre
/cru/cruts/final_structure/primary/pre/txt
/cru/cruts/final_structure/primary/pre/glo
/cru/cruts/final_structure/primary/pre/abs
/cru/cruts/final_structure/primary/pre/stn
/cru/cruts/final_structure/primary/pre/stn/cdd0
/cru/cruts/final_structure/primary/pre/stn/cddn
/cru/cruts/final_structure/primary/tmn
/cru/cruts/final_structure/primary/tmn/txt
/cru/cruts/final_structure/primary/tmn/glo
/cru/cruts/final_structure/primary/tmn/abs
/cru/cruts/final_structure/primary/tmn/stn
/cru/cruts/final_structure/primary/tmn/stn/cdd0
/cru/cruts/final_structure/primary/tmn/stn/cddn
/cru/cruts/final_structure/primary/tmx
/cru/cruts/final_structure/primary/tmx/txt
/cru/cruts/final_structure/primary/tmx/glo
/cru/cruts/final_structure/primary/tmx/abs
/cru/cruts/final_structure/primary/tmx/stn
/cru/cruts/final_structure/primary/tmx/stn/cdd0
/cru/cruts/final_structure/primary/tmx/stn/cddn
/cru/cruts/final_structure/primary/dtr
/cru/cruts/final_structure/primary/dtr/txt
/cru/cruts/final_structure/primary/dtr/glo
/cru/cruts/final_structure/primary/dtr/abs
/cru/cruts/final_structure/primary/dtr/stn
/cru/cruts/final_structure/primary/dtr/stn/cdd0
/cru/cruts/final_structure/primary/dtr/stn/cddn
/cru/cruts/final_structure/secondary
/cru/cruts/final_structure/secondary/vap
/cru/cruts/final_structure/secondary/vap/syn
/cru/cruts/final_structure/secondary/vap/txt
/cru/cruts/final_structure/secondary/vap/glo
/cru/cruts/final_structure/secondary/vap/abs
/cru/cruts/final_structure/secondary/vap/stn
/cru/cruts/final_structure/secondary/vap/stn/observed_only
/cru/cruts/final_structure/secondary/vap/stn/observed_only/cdd0
/cru/cruts/final_structure/secondary/vap/stn/observed_only/cddn
/cru/cruts/final_structure/secondary/vap/stn/all                     * might not do these
/cru/cruts/final_structure/secondary/vap/stn/all/cdd0
/cru/cruts/final_structure/secondary/vap/stn/all/cddn
/cru/cruts/final_structure/secondary/wet
```

```
/cru/cruts/final_structure/secondary/wet/syn
/cru/cruts/final_structure/secondary/wet/txt
/cru/cruts/final_structure/secondary/wet/glo
/cru/cruts/final_structure/secondary/wet/abs
/cru/cruts/final_structure/secondary/wet/stn
/cru/cruts/final_structure/secondary/wet/stn/observed_only
/cru/cruts/final_structure/secondary/wet/stn/observed_only/cdd0
/cru/cruts/final_structure/secondary/wet/stn/observed_only/cddn
/cru/cruts/final_structure/secondary/wet/stn/all                        * might not do these
/cru/cruts/final_structure/secondary/wet/stn/all/cdd0
/cru/cruts/final_structure/secondary/wet/stn/all/cddn
/cru/cruts/final_structure/secondary/frs
/cru/cruts/final_structure/secondary/frs/syn
/cru/cruts/final_structure/secondary/frs/txt
/cru/cruts/final_structure/secondary/frs/glo
/cru/cruts/final_structure/secondary/frs/abs
/cru/cruts/final_structure/secondary/frs/stn
/cru/cruts/final_structure/secondary/frs/stn/observed_only
/cru/cruts/final_structure/secondary/frs/stn/observed_only/cdd0
/cru/cruts/final_structure/secondary/frs/stn/observed_only/cddn
/cru/cruts/final_structure/secondary/frs/stn/all                        * might not do these
/cru/cruts/final_structure/secondary/frs/stn/all/cdd0
/cru/cruts/final_structure/secondary/frs/stn/all/cddn
/cru/cruts/final_structure/secondary/cld
/cru/cruts/final_structure/secondary/cld/syn
/cru/cruts/final_structure/secondary/cld/txt
/cru/cruts/final_structure/secondary/cld/glo
/cru/cruts/final_structure/secondary/cld/abs
/cru/cruts/final_structure/secondary/cld/stn
/cru/cruts/final_structure/secondary/cld/stn/observed_only
/cru/cruts/final_structure/secondary/cld/stn/observed_only/cdd0
/cru/cruts/final_structure/secondary/cld/stn/observed_only/cddn
/cru/cruts/final_structure/secondary/cld/stn/all                        * might not do these
/cru/cruts/final_structure/secondary/cld/stn/all/cdd0
/cru/cruts/final_structure/secondary/cld/stn/all/cddn
/cru/cruts/final_structure/static
/cru/cruts/final_structure/static/climatology
/cru/cruts/final_structure/static/mask


Then, there's the list of procedures (which probably need checking):

* Add MCDW Updates
mcdw2cru (interactive)
newmergedb (per parameter, interactive)
* Add CLIMAT Updates
climat2cru (interactive)
newmergedb (per parameter, interactive)
* Add BOM Updates
au2cru (unfinished, interactive, should do whole job)
* Regenerate DTR Database
tmnx2dtr (interactive)
* Produce Primary Parameters (TMP, TMN, TMX, DTR, PRE)
anomdtb (per parameter, interactive)
quick_interp_tdm2 (per parameter)
glo2abs (per parameter, interactive)
makegrids (per parameter, interactive)
* Prepare Binary Grids (TMP, DTR, PRE) for Synthetics
quick_interp_tdm2 (per parameter)
* Produce Secondary Parameter (FRS, uses TMP,DTR)
frs_gts_tdm
quick_interp_tdm2
glo2abs (interactive)
makegrids (interactive)
* Produce Secondary Parameter (VAP, uses TMP,DTR)
vap_gts_anom
anomdtb (interactive)
quick_interp_tdm2
glo2abs (interactive)
makegrids (interactive)
* Produce Secondary Parameter (WET/RD0, uses PRE)
rd0_gts_anom
anomdtb (interactive)
quick_interp_tdm2
glo2abs (interactive)
makegrids (interactive)

Now, in terms of methodology, we obviously need this to be portable. So either a system parameter or a text file
with local info is going to be needed. Since reading system parameters is less than easy (syntax differs between
platforms) a text file might be the way forward. At a pinch, all it would need to contain would be the root to
the hierarchy (ie, '/cru/cruts/final_structure/' in the above example). Might be an idea to provide a sandbox for
users to compile into.. and a file of compile lines? I wonder how far off I am from a makefile? That would help with
the frightening anomdtb.f linkages. Tried 'make' with anomdtb and it doesn't automatically find the includes, even
when they're in the same directory!

I guess I need to finish the fortran gridder program. That would allow steamlining. Notes on that
work are mainly in the file 'gridder.sandpit'. Suffice to say, it works :-) Needs tweaking, and
a few philosophical questions resolving, but apart from that..

So, you release a dataset that people have been clamouring for, and the buggers only start
using it! And finding problems. For instance:

<QUOTE>
Hi Tim (good start! -ed)

I realise you are likely to be very busy at the moment, but we have come across something in
the CRU TS 3.0 data set which I hope you can help out with.

We have been looking at the monthly precipitation totals over southern Africa (Angola, to be
precise), and have found some rather large differences between precipitation as specified in
the TS 2.1 data set, and the new TS 3.0 version. Specifically, April 1967 for the cell 12.75
south, 16.25 east, the monthly total in the TS 2.1 data set is 251mm, whereas in TS 3.0 it is
476mm. The anomaly does not only appear in this cell, but also in a number of neighbouring
cells. This is quite a large difference, and the new TS 3.0 value doesn't entirely tie in
with what we might have expected from the station-based precip data we have for this area.
Would it be possible for you could have a quick look into this issue?

Many thanks,

Daniel.

-------------------------------------------------------------
Dr Daniel Kingston
Post Doctoral Research Associate
Department of Geography
University College London
Gower Street
London
WC1E 6BT
UK
```

Email d.kingston@ucl.ac.uk
Tel. +44 (0)20 7679 0510
 <END>

Well, it's a good question! And it took over two weeks to answer. I wrote angola.m, which pretty much established that three local stations had been augmented for 3.0, and that April 1967 was anomalously wet. Lots of non-reporting stations (ie too few years to form normals) also had high values. As part of this, I also wrote angola3.m, which added two rather interesting plots: the climatology, and the output from the Fortran gridder I'd just completed. This raised a couple of points of interest:

1. The 2.10 output doesn't look like the climatology, despite there being no stations in the area. It ought to have simply relaxed to the clim, instead it's wetter.

2. The gridder output is lower than 3.0, and much lower than the stations!

I asked Tim and Phil about 1., they couldn't give a definitive opinion. As for 2., their guesses were correct, I needed to mod the distance weighting. As usual, see gridder.sandpit for the full info.

So to CLOUD. For over a year, rumours have been circulating that money had been found to pay somebody for a month to recreate Mark New's coefficients. But it never quite gelled. Now, at last, someone's producing them! Unfortunately.. it's me.

The idea is to derive the coefficients (for the regressing of cloud against DTR) using the published 2.10 data. We'll use 5-degree blocks and years 1951-2002, then produce coefficients for each 5-degree latitude band and month. Finally, we'll interpolate to get half-degree coefficients. Apparently.

Lots of 'issues'. We need to exclude 'background' stations - those that were relaxed to the climatology. This is hard to detect because the climatology consists of valid values, so testing for equivalence isn't enough. It might have to be the station files *shudder*.

Using station files was OK, actually. A bigger problem was the inclusion of strings of consecutive, identical values (for cloud and/or dtr). Not sure what the source is, as they are not == to the climatology (ie the anoms are not 0). Discussed with Phil - decided to try excluding any cell with a string like that of >10 values. Cloud only for now. The result of that was, unfortunately, the loss of several output values, ie:

```
lat band: 19   month:  7
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -36.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -37.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -41.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -43.00
    3.00  -38.00
    3.00  -38.00
    3.00  -41.00
    3.00  -38.00
    3.00  -39.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -43.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -44.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00
    3.00  -38.00

Results (n=  52):          nan         nan
```

As can be seen, neither the dtr (left) nor the cloud (right) look 'sensible', even as anomalies. Several other months in lat band #19 are either nan or -999 (count=0).

However, if we push the duplicates limit up to, say, 20, we get:

```
lat band: 19   month:  7
   13.50  -50.00
   13.50  -50.00
   13.50  -50.00
   13.50  -50.00
   13.50  -50.00
   13.50  -50.00
   13.50  -50.00
   13.50  -50.00
   13.50  -53.50
   13.50  -50.00
   13.50  -50.00
   13.50  -50.00
   13.50  -50.00
   13.50  -50.00
   13.50  -50.00
   13.50  -50.00
   13.50  -50.00
   13.50  -50.00
```

```
   13.50  -50.00
   13.50  -50.00
   13.50  -50.00
   13.50  -50.00
   13.50  -50.00
   13.50  -50.00
   13.50  -50.00
   17.00  -65.00
    8.50  -42.00
   11.50  -49.00
   18.00  -71.00

////////////////////

    1.00   33.50
    1.00   40.00
    1.00   32.00
    1.00   42.50
    1.00   38.00
    1.00   38.00
    1.00   32.50
    1.00   52.50
    1.00   44.00
    1.00   36.50
    1.00   41.00
    1.00   30.50
    1.00   38.00
    1.00   36.00
    1.00   38.00
    1.00   38.50
    1.00   39.00
    1.00   31.50
    1.00   40.00
    1.00   38.00
    1.00   31.00
    1.00   44.00
    1.00   43.00
    1.00   37.00
    1.00   31.00
    1.00   31.00
    1.00   30.50

Results (n= 988):     37.59707    -6.05338

So, we can have a proper result, but only by including a load of garbage! In fact,
I might print out this cell as an example. Let's see:

limit   nvals     factor   intercept
   10      52        nan         nan
   20     728   -7.68581    33.30551
 none    1716   -8.32450    34.28972

Hmm.. also tried just removing duplicate strings (rather than whole cells):

limit   nvals    factor    intercept
   10    1160  -6.99748     26.31960

This 'looks' better - not so steep, and the intercept is a shade closer to 0. The
Matlab script plotcld.m allows comparison of scatter diagrams, these are fed from
example data files manually extracted from the cloudreg.log file after varying the
duplicate limit and/or strategy.

Showed Phil - and now sidetracked into producing global mean series from the 3.0
parameters (DTR first).

OK, got cloud working, have to generate it now.. but distracted by starting on the
mythical 'Update' program. As usual, it's much more complicated than it seems. So,
let's work out the order of events.

Well first, some ground rules: should this be 'dumb'? Should the operator say what
they want to happen, and walk away, coming back later to check it worked? Or should
it be interactive, to the extent of the operator deciding on station matches and so
forth? At the moment, the introduction of new data (MCDW, CLIMAT, BOM) is highly
interactive, and, though BOM should be fully automatic in the future, the same
cannot be said for MCDW and CLIMAT. Hmmm. well I guess there are two possibilities:

1. Operator selects 'interactive' additions. Script proceeds, calling merge programs
   as necessary, some of which may ask the operator to decide on matches. This could
   take hours, or even days, depending on the quality of the incoming metadata.

2. Operator selects 'automatic' additions. Script proceeds, calling special versions
   of the merge programs. These have a fixed threshold of confidence for adding new
   data to exisitng databases. When the threshold is crossed, the data is not added
   but stored in a new database, which might of course be later added under option 1.
   Note that the threshold would be higher than that in 1. that initiates operator
   involvement.

Is this sufficient? It certainly means more coding, but not a huge amount. In a worst
case scenario (where the operator always chooses '2.'), we still have the unused data
updates that can be interactively merged in at any time (even yaesr in the future).

This all avoids the big questions, of course. When do updates happen, and how far back
do they go? For instance, let's say there are six-month published updates. So say the
full 1901-present files are published yearly, with six-month update files as interims.
What happens in any of the following circumstances?

A. Updates for, say, 1965, are available.

B. The data used in the January-to-June update is further updated after publication and
   is present in the next 'full' release (so that the early Jan-Jun grids differ from
   those in the 1901-present publication).

(in both A. and B., it would usually be MCDW updates that carried retrospective data,
this is marked as 'OVERDUE').


Luckily, this isn't really up to me. Or.. is it? If the operator specifies a time period
to update, it ought to warn if it finds earlier updates in those files. So further mods
to mcdw2cruauto are required.. its results file must list extras. Or - ooh! How about a
SECOND output database for the MCDW updates, containing just the OVERDUE stuff?

Back.. think.. even more complicated. My head hurts. No, it actually does. And I ought
to be on my way home. But look, we create a new master database (for each parameter)
every time we update, don't we? What we ought to do is provide a log file for each
new database, identifying which data have been added. Oh, God. OK, let's go..

NEW DATA PROCESS

1. Ops runs 'Update', and chooses 'New Data'.
```

2. Ops selects MCDW, CLIMAT, and/or BOM data and gives update dates

3. Ops selects 'interactive' or 'automatic' database merging.

4. Update checks source files are present and initiates conversion to CRU format.

5. Update runs the merging program to join the new data to the existing databases,
   creating new databases. If data for previous periods is included in the update
   files, it will be included.

5a. If Ops selected 'automatic', merging program asks for decisions on 'difficult'
    matches. These are all logged of course.

6. Merge program creates log of changes between old databases and new ones, inc.
   source of the data.

UPDATE PROCESS

1. Ops runs 'Update', and chooses 'Update'. Yes, I know.

2. Ops gives parameter(s) and time period to update.

3. Ops specifies six-month interim or full update.

4. Update provides candidate databases for the update, Ops chooses.

5. Update runs the anomaly and gridding programs for the specified period.

Note. The following system command will find the number of stations reporting in a
given year from a given database.

```
grep '^2006 ' tmp/tmp.0710011359.dtb | grep -v '\-9999\-9999\-9999\-9999\-9999\-9999\-9999\-9999\-9999\-9999\-9999\-9999' | wc -l
```


Discovered ('remembered' would be better; sadly I didn't) that I never got round to
writing a BOM-to-CRU converter. It got overtaken by the drastic need to get the tmin
and tmax databases synchronised (see above, somewhere). There was a barely-started
thing, so I cannibalised it for bom2cruauto.for, which eventually worked. In fact, it
was a good entry into the fraught world of automatic, script-fed programs.

Got bom2cruauto.for working, then climat2cruauto.for and mcdw2cruauto.for in quick
succession (the latter two having their output databases compared successfully with
those generated in Nov 2007).

Next, I suppose it's the next in the sequence: mergedb. This is where I'm anxious: I
want it all to be plain sailing and automatic, but I don't think there's any practical
way to obviate the operator from the need to make judgements on the possible mapping
of stations.

Back to get CLD sorted out. Need a break from the updater! Though much the same
difficulties, trying to work out the process (it's anything but straightforward for
cloud, seeing as the incoming updates are in Sun Hours, and we have to apply our
own regressions to DTR). Knocked up a subroutine, 'sh2cp', to convert sun hours to
cloud percentage on the fly, and added it to mcdw2cruauto and climat2cruauto. Of course,
one of the problems is that you need a latitude value to perform the conversion - so
the CLIMAT bulletins lose the value if they can't be matched in the WMO list! Not much
I can do about that, and let's face it those stations are going to end up as 'new'
stations with no possibility of a 61-90 normal.

So.. using the new converters (which are built to be driven by the update program), I
set about converting MCDW and CLIMAT bulletins:

```
uealogin1[/cru/cruts/version_3_0/update_top] ./mcdw2cruauto
uealogin1[/cru/cruts/version_3_0/update_top] cat results/results.0901101032/mcdw.0901101032.res
OK
uealogin1[/cru/cruts/version_3_0/update_top]

uealogin1[/cru/cruts/version_3_0/update_top] ./climat2cruauto
uealogin1[/cru/cruts/version_3_0/update_top] cat results/results.0901101032/climat.0901101032.res
OK
uealogin1[/cru/cruts/version_3_0/update_top]
```

Gotta love silent running :=)

The output cld databases both look OK, and pretty much equivalent except that MCDW goes back
further (to 1994). CLIMAT is 2000 onwards because that's what's on Phil Brohan's website.

Now we have to merge. This is where it gets hairy.

Looking at existing cld databases, we have the originals (cld.0301081434.dtb and
cld.0312181428.dtb), and the 2007 version (cld.0711272230.dtb). Looking back through
the notes, this was the product of processing the MCDW and CLIMAT bulletins into
sun hours databases (sun.0711272156.dtb and sun.0711272219.dtb respectiverly), then
merging those to form sun.0711272225.dtb, then converting to cloud using Hsp2cld.for,
giving us cld.0711272230.dtb. So the new cloud databases I've just produced should
be, if not identical, very similar? Oh, dear. There is a passing similarity, though
this seems to break down in Winter. I don't have time to do detailed comparisons, of
course, so we'll just run with the new one. After all, I have tested the conversion
for the latest programs, I'm not sure how much testing was done last time.

The procedure last time - that is, when I was trying to re-produce TS 2.10, we have
no idea what the procedure was for its initial production! - was to incorporate the
sun percent data from the bulletins into the existing sun percent db (spc.0312221624.dtb).
The trouble is, the existing cloud dbs are bigger. They stop at 1996, but that's no
problem, since we have MCDW from then.

```
    228936 cld.0301081434.dtb
    104448 cld.0312181428.dtb
    111989 combo.cld.dtb
     57395 spc.0301201628.dtb
     51551 spc.0312221624.dtb
     51551 spc.94-00.0312221624.dtb
```

So, how about merging our new MCDW cloud database into cld.0312181428.dtb, then merging
the CLIMAT one into that? The logic here is that the cloud must be from observations,
because the sun databases are much smaller. Well, the ones we know about! It might be
worth checking the station numbers for each year though. Unfortunately, we don't have a
lot of luck merging MCDW updates into the Dec 2003 CLD database:

```
uealogin1[/cru/cruts/version_3_0/db/cld] ./newmergedb

WELCOME TO THE DATABASE UPDATER

Before we get started, an important question:
If you are merging an update - CLIMAT, MCDW, Australian - do
```

```
you want the quick and dirty approach? This will blindly match
on WMO codes alone, ignoring data/metadata checks, and making any
unmatched updates into new stations (metadata permitting)?

Enter 'B' for blind merging, or <ret>: B
Please enter the Master Database name: cld.0312181428.dtb
Please enter the Update Database name: mcdw.cld.0901101032.dtb

Reading in both databases..
Master database stations:     3605
Update database stations:     2204

Looking for WMO code matches..
    5 reject(s) from update process 0902101404

Writing cld.0902101404.dtb

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

OUTPUT(S) WRITTEN

New master database: cld.0902101404.dtb

Update database stations:         2204
 > Matched with Master stations:   858
                (automatically:   858)
                  (by operator:     0)
 > Added as new Master stations:  1341
 > Rejected:                         5
   Rejects file:                mcdw.cld.0901101032.dtb.rejected
uealogin1[/cru/cruts/version_3_0/db/cld]

Of course, as we are only generating from 1996 onwards, this probably isn't
of much significance. Luckily.

Next, merge CLIMAT into the new database. well of course this is much more
satisfactory, because it matches with the MCDW stations:

uealogin1[/cru/cruts/version_3_0/db/cld] ./newmergedb

WELCOME TO THE DATABASE UPDATER

Before we get started, an important question:
If you are merging an update - CLIMAT, MCDW, Australian - do
you want the quick and dirty approach? This will blindly match
on WMO codes alone, ignoring data/metadata checks, and making any
unmatched updates into new stations (metadata permitting)?

Enter 'B' for blind merging, or <ret>: B
Please enter the Master Database name: cld.0902101404.dtb
Please enter the Update Database name: climat.cld.0901101032.dtb

Reading in both databases..
Master database stations:     4946
Update database stations:     2038

Looking for WMO code matches..
    2 reject(s) from update process 0902101409

Writing cld.0902101409.dtb

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

OUTPUT(S) WRITTEN

New master database: cld.0902101409.dtb

Update database stations:         2038
 > Matched with Master stations:  1858
                (automatically:  1858)
                  (by operator:     0)
 > Added as new Master stations:   178
 > Rejected:                         2
   Rejects file:                climat.cld.0901101032.dtb.rejected
uealogin1[/cru/cruts/version_3_0/db/cld]

So we now have cld.0902101409.dtb, a database consisting of cld.0312181428.dtb,
updated first with derived-cloud data from MCDW (1994-2008), then with
derived-cloud data from CLIMAT (2000-2008). This should be anomalised, then
fed into quick_interp_tdm2.pro along with synthetic, DTR-derived cloud. So that
is the next hurdle.

Well, we have the program. And we've played with it, but forgot to c&p those runs
into here (well they were only a few days ago!) so here they are now:

crua6[/cru/cruts/version_3_0/secondaries/cld/cldfromdtrtxt] ./dtr2cld
Please enter the path/file of the FIRST dtr txt file: /cru/cruts/version_3_0/primaries/dtr/dtrtxt/dtr.1901.01.txt
Please enter the path/file of the LAST dtr txt file:  /cru/cruts/version_3_0/primaries/dtr/dtrtxt/dtr.2006.12.txt
crua6[/cru/cruts/version_3_0/secondaries/cld/cldfromdtrtxt]

Then an experimental IDL gridding using half degree and glo output. It was late at night,
I think I had an idea of visualising or comparing.. no such luck.

crua6[/cru/cruts/version_3_0/secondaries/cld] idl
IDL Version 5.4 (OSF alpha). (c) 2000, Research Systems, Inc.
Installation number: 66286.
Licensed for use by: Climatic Research Unit

IDL> quick_interp_tdm2,1995,2006,'cld.',750,gs=0.5,dumpglo='dumpglo',pts_prefix='cldfromdtrtxt/cld.'
% Compiled module: QUICK_INTERP_TDM2.
% Compiled module: GLIMIT.
Defaults set
    1995
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: STRIP.
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
    1996
    1997
    1998
    1999
    2000
    2001
    2002
    2003
    2004
    2005
```

```
    2006
no stations found in: cldfromdtrtxt/cld.2006.09.txt
no stations found in: cldfromdtrtxt/cld.2006.10.txt
no stations found in: cldfromdtrtxt/cld.2006.11.txt
no stations found in: cldfromdtrtxt/cld.2006.12.txt
IDL>
```

So now we need to try that last step again - this time going for 2.5-degree binary outputs, suitable for feeding back into it for the full cloud gridding. Oh, my.

```
IDL> quick_interp_tdm2,1996,2006,'cldfromdtr25bin/cld.',750,gs=2.5,dumpbin='dumpbin',pts_prefix='cldfromdtrtxt/cld.'

<output removed as re-done below with CDD=600>
```

Okay, that's the synthetic binary cloud ready. Now we need to run anomdtb on the 'observed' cloud database:

```
crua6[/cru/cruts/version_3_0/secondaries/cld] ./anomdtb

   > ***** AnomDTB: converts .dtb to anom .txt for gridding *****

   > Enter the suffix of the variable required (eg, .tmp):
.cld
   > Select the .dtb file to load:
cld.0902101409.dtb
 cld.0902101409.dtb


 /tmp_mnt/cru-auto/cruts/version_3_0/secondaries/cld/cld.0902101409.dtb


   > Specify the start,end of the normals period:
1961,1990
   > Specify the missing percentage permitted:
25
   > Data required for a normal:           23
   > Specify the no. of stdevs at which to reject data:
3
The output selection is tied to 3 (ungridded anomalies)
   > Check for duplicate stns after anomalising? (0=no,>0=km range)
0
   > Select the generic .txt file to save (yy.mm=auto):
cldnew.txt
   > Select the first,last years AD to save:
1996,2006
   > Operating...
 /tmp_mnt/cru-auto/cruts/version_3_0/secondaries/cld/cld.0902101409.dtb


 /tmp_mnt/cru-auto/cruts/version_3_0/secondaries/cld/cld.0902101409.dtb


   > NORMALS            MEAN percent      STDEV percent
   >         .dtb          0     0.0
   >         .cts     288328    22.2     288328    22.2
   > PROCESS        DECISION percent %of-chk
   > no lat/lon            0     0.0     0.0
   > no normal       1010030    77.8    77.8
   > out-of-range         24     0.0     0.0
   > accepted         288304    22.2
   > Dumping years 1996-2006 to .txt files...

crua6[/cru/cruts/version_3_0/secondaries/cld]
```

Unfortunately, that isn't working. Too many stations outside the usual normals period (1961-1990). My notes from the last attempt are less than inspiring.. it looks as though we need the program 'normshift.for', and normalise 95-02. So:

```
crua6[/cru/cruts/version_3_0/secondaries/cld] ./anomdtb

   > ***** AnomDTB: converts .dtb to anom .txt for gridding *****

   > Enter the suffix of the variable required (eg, .tmp):
.cld
   > Select the .dtb file to load:
cld.0902101409.dtb
 cld.0902101409.dtb


 /tmp_mnt/cru-auto/cruts/version_3_0/secondaries/cld/cld.0902101409.dtb


   > Specify the start,end of the normals period:
1995,2002
   > Specify the missing percentage permitted:
25
   > Data required for a normal:            6
   > Specify the no. of stdevs at which to reject data:
3
The output selection is tied to 3 (ungridded anomalies)
   > Check for duplicate stns after anomalising? (0=no,>0=km range)
0
   > Select the generic .txt file to save (yy.mm=auto):
cldupdate.txt
   > Select the first,last years AD to save:
1996,2006
   > Operating...
 /tmp_mnt/cru-auto/cruts/version_3_0/secondaries/cld/cld.0902101409.dtb


 /tmp_mnt/cru-auto/cruts/version_3_0/secondaries/cld/cld.0902101409.dtb


   > NORMALS            MEAN percent      STDEV percent
   >         .dtb          0     0.0
   >         .cts     271495    20.9     271495    20.9
   > PROCESS        DECISION percent %of-chk
   > no lat/lon            0     0.0     0.0
   > no normal       1026863    79.1    79.1
   > out-of-range         474     0.0     0.2
   > accepted         271021    20.9
   > Dumping years 1996-2006 to .txt files...

crua6[/cru/cruts/version_3_0/secondaries/cld]
```

Hmm.. that's giving us between 670 and 790 stations per month.. not too bad I suppose seeing as it's a secondary parameter. Now for normshift, which has already been run (search back in this file), producing clim.9502.to.6190.grid.cld, which is your standard 12-grids-360rx720c, giving the diffs between 1995-2002 normals and 1961-1990 normals. So after gridding we could add these.. except that

```
after gridding we'll have incorporated the DTR_derived synthetic cloud, which is
of course based on the 1961-1990 normals as it's derived from DTR!! Arrrrggghh.

So.. {sigh}.. another problem. Well we can't change the updates side, that has to
use 1995-2002 normals. But maybe we'll have to adjust the station anomalies, prior
to gridding? I don't see an alternative.

Wrote movenorms.for, using the engine of dtr2cld (as it's processing the same kind
of files and also needs to map stations to cells). However we quickly hit a problem:

crua6[/cru/cruts/version_3_0/secondaries/cld] ./movenorms

Please enter the adjustment file: clim.9502.to.6190.grid.cld

Please enter a generic source file with MM for month and YYYY for year: cldupdatetxt/cldupdate.YYYY.MM.txt
Start YEAR:  1996
Start MONTH: 01
End YEAR:    2006
End MONTH:   12

Please enter a generic destination file with MM for month and YYYY for year: cldupdate6190/cldupdate6190.YYYY.MM.txt

ERROR. Station in sea:
File: cldupdate6190/cldupdate6190.1996.01.txt
Offending line:    18.54   72.49    11.0     -6.600004305700
Resulting indices ilat,ilon:      218     505
crua6[/cru/cruts/version_3_0/secondaries/cld]

This is a station on the West coast of India; probably Mumbai. Unfortunately, as a coastal
station it runs the risk of missing the nearest land cell. The simple movenorms program is
about to become less simple.. but was do-able. The log file was empty at the end, indicating
that all 'damp' stations had found dry land:

crua6[/cru/cruts/version_3_0/secondaries/cld] ./movenorms

Please enter the adjustment file: clim.9502.to.6190.grid.cld

Please enter a generic source file with MM for month and YYYY for year: cldupdatetxt/cldupdate.YYYY.MM.txt
Start YEAR:  1996
Start MONTH: 1
End YEAR:    2006
End MONTH:   12

Please enter a generic destination file with MM for month and YYYY for year: cldupdate6190/cldupdate6190.YYYY.MM.txt
crua6[/cru/cruts/version_3_0/secondaries/cld] wc -l movenorms.log
          0 movenorms.log
crua6[/cru/cruts/version_3_0/secondaries/cld]

So.. now I should be able to do the final gridding of cloud for 1996-2006.

IDL> quick_interp_tdm2,1996,2006,'cloudcomboglo/cld.',750,gs=0.5,dumpglo='dumpglo',synth_prefix='cldfromdtr25bin/cld.',pts_prefix='cldupdate6190/cldupdate6190.'

<output removed as re-done below with CDD=600>

crua6[/cru/cruts/version_3_0/secondaries/cld] ./glo2abs

<output removed as re-done below with CDD=600>

uealogin1[/cru/cruts/version_3_0/secondaries/cld] ./makegrids

<output removed as re-done below with CDD=600>

All files look alright. BUT. The NetCDF attributes (which are still bad) do say that
the CDD for cloud is 600. If it is, I will eat my screen, because I'll have to do all
the gridding ops again :-(((  and.. it is :-o

So.. binary re-production:


IDL> quick_interp_tdm2,1996,2006,'cldfromdtr25bin/cld.',600,gs=2.5,dumpbin='dumpbin',pts_prefix='cldfromdtrtxt/cld.'
% Compiled module: QUICK_INTERP_TDM2.
% Compiled module: GLIMIT.
Defaults set
    1996
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: MEAN.
% Compiled module: MOMENT.
% Compiled module: STDDEV.
grid 1996 non-zero  -22.2967   46.5119  134.6274 cells=    53787
% Compiled module: STRIP.
% Compiled module: WRBIN.
    1997
grid 1997 non-zero  -22.7474   47.1535  131.6472 cells=    53374
    1998
grid 1998 non-zero  -24.3090   50.5343  155.2392 cells=    53557
    1999
grid 1999 non-zero  -21.0658   46.2280  127.9565 cells=    52391
    2000
grid 2000 non-zero  -23.3182   50.4612  154.3142 cells=    51948
    2001
grid 2001 non-zero  -25.4712   50.3292  147.5332 cells=    50464
    2002
grid 2002 non-zero  -22.6252   49.8823  153.3252 cells=    46980
    2003
grid 2003 non-zero  -21.8382   48.3537  136.5305 cells=    48279
    2004
grid 2004 non-zero  -23.9221   47.9721  145.4819 cells=    48179
    2005
grid 2005 non-zero  -26.0049   48.8422  145.3165 cells=    44448
    2006
no stations found in: cldfromdtrtxt/cld.2006.09.txt
no stations found in: cldfromdtrtxt/cld.2006.10.txt
no stations found in: cldfromdtrtxt/cld.2006.11.txt
no stations found in: cldfromdtrtxt/cld.2006.12.txt
grid 2006 non-zero  -17.3353   45.5259  135.7803 cells=    29194
IDL>

..and finals re-produced:

IDL> quick_interp_tdm2,1996,2006,'cloudcomboglo/cld.',600,gs=0.5,dumpglo='dumpglo',synth_prefix='cldfromdtr25bin/cld.',pts_prefix='cldupdate6190/cldupdate6190.'
Defaults set
    1996
% Compiled module: RDBIN.
% Compiled module: DEFXYZ.
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
    1997
    1998
    1999
```

```
    2000
    2001
    2002
    2003
    2004
    2005
    2006
IDL>

crua6[/cru/cruts/version_3_0/secondaries/cld] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.cld
Enter a name for the gridded climatology file: clim.6190.lan.cld.toothbrush
Enter the path and stem of the .glo files: cloudcomboglo/cld.
Enter the starting year: 1996
Enter the ending year:   2006
Enter the path (if any) for the output files: cloudcomboabs/
Now, CONCENTRATE. Addition or Percentage (A/P)? A
Do you wish to limit the output values? (Y/N): Y
1. Set minimum to zero
2. Set single minimum and maximum values
3. Set minima and maxima based on days in month
4. Set integer values >=1, (ie, positive)
5. Changed my mind, no limits
Choose: 2

Enter minimum value: 0

Enter maximum value: 1000
Right, erm.. off I jolly well go!
cloudcomboglo/cld.01.1996.glo
cloudcomboglo/cld.1996.01.glo
cld.1996.01.glo
cld.1996.02.glo

(etc)


cld.2006.11.glo
cld.2006.12.glo
crua6[/cru/cruts/version_3_0/secondaries/cld]

uealogin1[/cru/cruts/version_3_0/secondaries/cld] ./makegrids
Welcome! This is the MAKEGRIDS program.
I will create decadal and full gridded files,
in both ASCII text and NetCDF formats, from
the output files of (eg) glo2abs.for.

Enter a gridfile with YYYY for year and MM for month: cloudcomboabs/cld.YYYY.MM.glo.abs
Enter Start Year:  1996
Enter Start Month: 01
Enter End Year:    2006
Enter End Month:   12

Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE, and
ending with '.dat', eg: cru_ts_3_00.SSSS.EEEE.tmp.dat : cru_ts_3_00.SSSS.EEEE.cld.dat

Now please enter the 3-ch parameter code: cld
Enter a generic title for this dataset, eg:
CRU TS 3.00 Mean Temperature : CRU TS 3.00 Percentage Cloud Cover
Writing: cru_ts_3_00.1996.2000.cld.dat
         cru_ts_3_00.1996.2000.cld.nc
Writing: cru_ts_3_00.2001.2006.cld.dat
         cru_ts_3_00.2001.2006.cld.nc
uealogin1[/cru/cruts/version_3_0/secondaries/cld]

The question is, IS THIS ANY GOOD? Well, we currently have published cloud data
to 2002. So we can make comparisons between 1996 and 2002. Oh, my. I am sure
I've written plenty of comparison routines, but as to their location or name..ah.
There is cmpmgrids.m, which I modified away from its precipitation-only mentality.
I used mmeangrid.for to calculate monthly mean fields (1996-2002) for both 2.10 and
3.00 cloud. The resulting mean files, cru_ts_2_10.1996.2002.cld.dat.mmeans and
cru_ts_3_00.1996.2002.cld.dat.mmeans, were fed into cmpmgrids.m. The results were
less than ideal, though they could have been much worse. Essentially, North America
is totally different - cloudier in Feb/Mar/Apr, sunnier the rest of the year. There
are other differences, particularly in Northern Asia, but these are oatchier and
don't extend throughout the year. So.. the obvious cause would be the inclusion of
DTR-derived cloud, since that would have significant station counts in North America
compared to CLD? Also, there seems to be horizontal banding.. not a good sign given
the nature of the DTR-to-CLD conversion! Naturally, the way to test this is to make
comparisons between FIVE different datasets:

1. CRU TS 2.10 1996-2002 monthly means
2. CRU TS 3.00 1996-2002 monthly means
3. CRU TS 3.00 1996-2002 (synthetic only) monthly means
4. CRU TS 3.00 1996-2002 (observed only) monthly means
5. CLD Climatology

The inclusion of 5 will show the extent of missing data, perhaps.. so I'm suggesting
the following tests:

2-1   Basic comparison of old and new (already done)
3-1   How the DTR-derived synthetic CLD relates to 2.1
4-1   How the sun-hours-derived observed CLD relates to 2.1
3-2   How the DTR-derived synthetic CLD relates to the 'combo' CLD
1-5   How 2.1 relates to the climatology
2-5   How 3.0 relates to the climatology

So, to making datasets 3 and 4:

IDL> quick_interp_tdm2,1996,2002,'cldfromdtrglo05/cld.',600,gs=0.5,dumpglo='dumpglo',pts_prefix='cldfromdtrtxt/cld.'
% Compiled module: QUICK_INTERP_TDM2.
% Compiled module: GLIMIT.
Defaults set
    1996
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: STRIP.
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
    1997
    1998
    1999
    2000
    2001
    2002
```

```
    2003
    2004
    2005
    2006
no stations found in: cldfromdtrtxt/cld.2006.09.txt
no stations found in: cldfromdtrtxt/cld.2006.10.txt
no stations found in: cldfromdtrtxt/cld.2006.11.txt
no stations found in: cldfromdtrtxt/cld.2006.12.txt
IDL> quick_interp_tdm2,1996,2002,'cldfromupdate6190glo/cld.',600,gs=0.5,dumpglo='dumpglo',pts_prefix='cldupdate6190/cldupdate6190.'
Defaults set
    1996
    1997
    1998
    1999
    2000
    2001
    2002
    2003
    2004
    2005
    2006
IDL>

crua6[/cru/cruts/version_3_0/secondaries/cld] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.cld
Enter a name for the gridded climatology file: clim.6190.lan.cld.dunconvertin
Enter the path and stem of the .glo files: cldfromdtrglo05/cld.
Enter the starting year: 1996
Enter the ending year:   2002
Enter the path (if any) for the output files: cldfromdtrglo05abs/
Now, CONCENTRATE. Addition or Percentage (A/P)? A
Do you wish to limit the output values? (Y/N): Y
1. Set minimum to zero
2. Set single minimum and maximum values
3. Set minima and maxima based on days in month
4. Set integer values >=1, (ie, positive)
5. Changed my mind, no limits
Choose: 2

Enter minimum value: 0

Enter maximum value: 1000
Right, erm.. off I jolly well go!
cldfromdtrglo05/cld.01.1996.glo
cldfromdtrglo05/cld.1996.01.glo
cld.1996.01.glo
cld.1996.02.glo

(etc)

cld.2002.11.glo
cld.2002.12.glo
crua6[/cru/cruts/version_3_0/secondaries/cld] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.cld
Enter a name for the gridded climatology file: clim.6190.lan.cld.notuagain
Enter the path and stem of the .glo files: cldfromupdate6190glo/cld.
Enter the starting year: 1996
Enter the ending year:   2002
Enter the path (if any) for the output files: cldfromupdate6190gloabs/
Now, CONCENTRATE. Addition or Percentage (A/P)? A
Do you wish to limit the output values? (Y/N): Y
1. Set minimum to zero
2. Set single minimum and maximum values
3. Set minima and maxima based on days in month
4. Set integer values >=1, (ie, positive)
5. Changed my mind, no limits
Choose: 2

Enter minimum value: 0

Enter maximum value: 1000
Right, erm.. off I jolly well go!
cldfromupdate6190glo/cld.01.1996.glo
cldfromupdate6190glo/cld.1996.01.glo
cld.1996.01.glo
cld.1996.02.glo

(etc)

cld.2002.11.glo
cld.2002.12.glo
crua6[/cru/cruts/version_3_0/secondaries/cld]

uealogin1[/cru/cruts/version_3_0/secondaries/cld] ./makegrids
Welcome! This is the MAKEGRIDS program.
I will create decadal and full gridded files,
in both ASCII text and NetCDF formats, from
the output files of (eg) glo2abs.for.

Enter a gridfile with YYYY for year and MM for month: cldfromdtrglo05abs/cld.YYYY.MM.glo.abs
Enter Start Year:  1996
Enter Start Month: 01
Enter End Year:    2002
Enter End Month:   12

Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE, and
ending with '.dat', eg: cru_ts_3_00.SSSS.EEEE.tmp.dat : cru_ts_3_00.SSSS.EEEE.cld_from_dtr_only.dat

Now please enter the 3-ch parameter code: cld
Enter a generic title for this dataset, eg:
CRU TS 3.00 Mean Temperature : CRU TS 3.00 Percentage Cloud Cover from DTR only
Writing: cru_ts_3_00.1996.2000.cld_from_dtr_only.dat
         cru_ts_3_00.1996.2000.cld_from_dtr_only.nc
Writing: cru_ts_3_00.2001.2002.cld_from_dtr_only.dat
         cru_ts_3_00.2001.2002.cld_from_dtr_only.nc
uealogin1[/cru/cruts/version_3_0/secondaries/cld] ./makegrids
Welcome! This is the MAKEGRIDS program.
I will create decadal and full gridded files,
in both ASCII text and NetCDF formats, from
the output files of (eg) glo2abs.for.
```

```
Enter a gridfile with YYYY for year and MM for month: cldfromupdate6190gloabs/cld.YYYY.MM.glo.abs
Enter Start Year:  1996
Enter Start Month: 01
Enter End Year:    2002
Enter End Month:   12

Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE, and
ending with '.dat', eg: cru_ts_3_00.SSSS.EEEE.tmp.dat : cru_ts_3_00.SSSS.EEEE.cld_from_sunobs_only.dat

Now please enter the 3-ch parameter code: cld
Enter a generic title for this dataset, eg:
CRU TS 3.00 Mean Temperature : CRU TS 3.00 Percentage Cloud Cover from SUN obs only
Writing: cru_ts_3_00.1996.2000.cld_from_sunobs_only.dat
         cru_ts_3_00.1996.2000.cld_from_sunobs_only.nc
Writing: cru_ts_3_00.2001.2002.cld_from_sunobs_only.dat
         cru_ts_3_00.2001.2002.cld_from_sunobs_only.nc
uealogin1[/cru/cruts/version_3_0/secondaries/cld] ./mmeangrid

MMEANGRID - Calculate monthly means for CRU_TS grids

Please enter the gridded data filename: cru_ts_3_00.1996.2002.cld_from_dtr_only.dat

Writing monthly mean grids to: cru_ts_3_00.1996.2002.cld_from_dtr_only.dat.mmeans

uealogin1[/cru/cruts/version_3_0/secondaries/cld] ./mmeangrid

MMEANGRID - Calculate monthly means for CRU_TS grids

Please enter the gridded data filename: cru_ts_3_00.1996.2002.cld_from_sunobs_only.dat

Writing monthly mean grids to: cru_ts_3_00.1996.2002.cld_from_sunobs_only.dat.mmeans

uealogin1[/cru/cruts/version_3_0/secondaries/cld]

Giving us (numbering from before):

1. cru_ts_2_10.1996.2002.cld.dat.mmeans
2. cru_ts_3_00.1996.2002.cld.dat.mmeans
3. cru_ts_3_00.1996.2002.cld_from_dtr_only.dat.mmeans
4. cru_ts_3_00.1996.2002.cld_from_sunobs_only.dat.mmeans
5. clim.6190.lan.cld.grid

And here are our target comparisons again, this time with notes:

2-1   Basic comparison of old and new (already done)

   -> Major diffs in N America, all months (lat. striping)

3-1   How the DTR-derived synthetic CLD relates to 2.1

   -> major diffs globally, all months (lat. striping)

4-1   How the sun-hours-derived observed CLD relates to 2.1

   -> minor patchy diffs globally, all months

3-2   How the DTR-derived synthetic CLD relates to the climatology

   -> major diffs globally, all months (lat. striping) c/w 3-1

1-5   How 2.1 relates to the climatology

   -> minor patchy diffs globally - equiv for

2-5   How 3.0 relates to the climatology

   -> Pretty much as for 2-1

The deduction so far is that the DTR-derived CLD is waaay off. The DTR looks OK, well
OK in the sense that it doesn;t have prominent bands! So it's either the factors and
offsets from the regression, or the way they've been applied in dtr2cld.

Well, dtr2cld is not the world's most complicated program. Wheras cloudreg is, and I
immediately found a mistake! Scanning forward to 1951 was done with a loop that, for
completely unfathomable reasons, didn't include months! So we read 50 grids instead
of 600!!! That may have had something to do with it. I also noticed, as I was correcting
THAT, that I reopened the DTR and CLD data files when I should have been opening the
bloody station files!! I can only assume that I was being interrupted continually when
I was writing this thing. Running with those bits fixed improved matters somewhat,
though now there's a problem in that one 5-degree band (10S to 5S) has no stations! This
will be due to low station counts in that region, plus removal of duplicate values.

Had a think. Phil advised averaging the bands either side to fill the gap, but yuk! And
also the band to the North (ie, 5S to equator) is noticeably lower (extreme, even). So
after some investigation I found that, well, here's the email:

<MAIL QUOTE>
Phil,

I've looked at why we're getting low counts for valid cloud cells in certain 5-degree latitude bands.

The filtering algorithm omits any cell values where the station count is zero, for either CLD or DTR. In general, it's the CLD counts that are zero and losing us the data.

However, in many cases, the cloud value in that cell on that month is not equal to the climatology. And there is plenty of DTR data. So I'm wondering how accurate the statio

(all values are x10)
  CLD-------------------  DTR--------------------
     val     stn    anom     val     stn    anom
  553.00    0.00  -10.00  134.00   20.00   -1.00
  558.00    0.00  -17.00  139.00   20.00    2.00
  565.00    0.00  -23.00  137.00   20.00    5.00
  581.00    0.00  -32.00  139.00   16.00    8.00
  587.00    0.00  -38.00  137.00   16.00    9.00
  567.00    0.00  -46.00  127.00   15.00    6.00
  564.00    0.00  -49.00  120.00   14.00    3.00
  552.00    0.00  -48.00  111.00   12.00    0.00
  543.00    0.00  -45.00  105.00   12.00   -1.00
  535.00    0.00  -40.00   99.00   10.00   -1.00

So, I'm proposing to filter on only the DTR counts, on the assumption that PRE was probably available if DTR was, so synthesis of CLD was likely to have happened, just not s
<END MAIL QUOTE>

I didn't get an email back but he did verbally consent. So away we go!

Running with a DTR-station-only screening gives us lots of station values, even with
duplicate filtering turned back on. Niiice. It's still not exactly smooth, but it
might be enough to 'fix' the synthetic cloud.

So, moved all existing directories to <name>_old_badcoeffs, and..
```

..reproduced the synthetic cloud:

```
crua6[/cru/cruts/version_3_0/secondaries/cld] ./dtr2cld
Please enter the path/file of the FIRST dtr txt file: /cru/cruts/version_3_0/primaries/dtr/dtrtxt/dtr.1996.01.txt
Please enter the path/file of the LAST dtr txt file:  /cru/cruts/version_3_0/primaries/dtr/dtrtxt/dtr.2006.12.txt
crua6[/cru/cruts/version_3_0/secondaries/cld]
```

Binary 2.5 grid production:

```
IDL> quick_interp_tdm2,1996,2006,'cldfromdtr25bin/cld.',600,gs=2.5,dumpbin='dumpbin',pts_prefix='cldfromdtrtxt/cld.'
Defaults set
    1996
grid 1996 non-zero  -28.0336   35.2211  155.5659 cells=    35824
    1997
grid 1997 non-zero  -28.7817   35.5387  155.2398 cells=    36027
    1998
grid 1998 non-zero  -31.6090   41.7502  182.8359 cells=    36481
    1999
grid 1999 non-zero  -26.8934   35.6123  151.3076 cells=    34127
    2000
grid 2000 non-zero  -28.3765   41.1417  180.7935 cells=    34846
    2001
grid 2001 non-zero  -31.7763   40.9542  172.0433 cells=    33724
    2002
grid 2002 non-zero  -29.6498   42.6404  182.7383 cells=    31683
    2003
grid 2003 non-zero  -27.8828   38.8903  161.9822 cells=    32227
    2004
grid 2004 non-zero  -30.5593   38.4767  174.0231 cells=    32315
    2005
grid 2005 non-zero  -33.2088   40.1841  170.8421 cells=    30951
    2006
no stations found in: cldfromdtrtxt/cld.2006.09.txt
no stations found in: cldfromdtrtxt/cld.2006.10.txt
no stations found in: cldfromdtrtxt/cld.2006.11.txt
no stations found in: cldfromdtrtxt/cld.2006.12.txt
grid 2006 non-zero  -27.2999   36.2585  161.9338 cells=    20383
IDL>
```

Final gridding with obs as well:

```
IDL> quick_interp_tdm2,1996,2006,'cloudcomboglo/cld.',600,gs=0.5,dumpglo='dumpglo',synth_prefix='cldfromdtr25bin/cld.',pts_prefix='cldupdate6190/cldupdate6190.'
Defaults set
    1996
    1997
    1998
    1999
    2000
    2001
    2002
    2003
    2004
    2005
    2006
IDL>

crua6[/cru/cruts/version_3_0/secondaries/cld] ./glo2abs
Welcome! This is the GLO2ABS program.
I will create a set of absolute grids from
a set of anomaly grids (in .glo format), also
a gridded version of the climatology.
Enter the path and name of the normals file: clim.6190.lan.cld
Enter a name for the gridded climatology file: clim.6190.lan.cld.speeeeew
Enter the path and stem of the .glo files: cloudcomboglo/cld.
Enter the starting year: 1996
Enter the ending year:   2006
Enter the path (if any) for the output files: cloudcomboabs/
Now, CONCENTRATE. Addition or Percentage (A/P)? A
Do you wish to limit the output values? (Y/N): Y
1. Set minimum to zero
2. Set single minimum and maximum values
3. Set minima and maxima based on days in month
4. Set integer values >=1, (ie, positive)
5. Changed my mind, no limits
Choose: 2

Enter minimum value: 0

Enter maximum value: 1000
Right, erm.. off I jolly well go!
cloudcomboglo/cld.01.1996.glo
cloudcomboglo/cld.1996.01.glo
cld.1996.01.glo
cld.1996.02.glo
(etc)
cld.2006.11.glo
cld.2006.12.glo
crua6[/cru/cruts/version_3_0/secondaries/cld]

uealogin1[/cru/cruts/version_3_0/secondaries/cld] ./makegrids
Welcome! This is the MAKEGRIDS program.
I will create decadal and full gridded files,
in both ASCII text and NetCDF formats, from
the output files of (eg) glo2abs.for.

Enter a gridfile with YYYY for year and MM for month: cloudcomboabs/cld.YYYY.MM.glo.abs
Enter Start Year:  1996
Enter Start Month: 01
Enter End Year:    2006
Enter End Month:   12

Please enter a sample OUTPUT filename, replacing
start year with SSSS and end year with EEEE, and
ending with '.dat', eg: cru_ts_3_00.SSSS.EEEE.tmp.dat : cru_ts_3_00.SSSS.EEEE.cld.dat

Now please enter the 3-ch parameter code: cld
Enter a generic title for this dataset, eg:
CRU TS 3.00 Mean Temperature : CRU TS 3.00 Percentage Cloud Cover
Writing: cru_ts_3_00.1996.2000.cld.dat
         cru_ts_3_00.1996.2000.cld.nc
Writing: cru_ts_3_00.2001.2006.cld.dat
         cru_ts_3_00.2001.2006.cld.nc
uealogin1[/cru/cruts/version_3_0/secondaries/cld]

uealogin1[/cru/cruts/version_3_0/secondaries/cld] head -30240 cru_ts_3_00.1996.2006.cld.dat >cru_ts_3_00.1996.2002.cld.dat
uealogin1[/cru/cruts/version_3_0/secondaries/cld] ./mmeangrid

MMEANGRID - Calculate monthly means for CRU_TS grids
```

```
Please enter the gridded data filename: cru_ts_3_00.1996.2002.cld.dat

Writing monthly mean grids to: cru_ts_3_00.1996.2002.cld.dat.mmeans

uealogin1[/cru/cruts/version_3_0/secondaries/cld]

Back with cmpmgrids.m.. and things look MUCH better. Differences with the climatology,
or with the 2.10 release, are patchy and generally below 30%. Of course it would be
nice if the differences with the 2.10 release were negligable, since our regression
coefficients were based on 2.10 DTR and CLD.. though of course the sun hours component
is an unknown there, as is the fact that 2.10 used PRE as well as DTR for the synthetics.
Anyway it gets the thumbs-up. The strategy will be to just produce it for 2003-2006.06,
to tie in with the rest of the 3.00 release. So I just need to.. argh. I don't have any
way to create NetCDF files 1901-2006 without the .glo.abs files to work from! I'd have
to specially code a version that swallowed the existing 1901-2002 then added ours. Meh.
Well it's no problem to release a concatenated ASCII file, so I'll do that:

crua6[/cru/cruts/version_3_0/secondaries/cld/cld_final] ls -l
total 2414884
-rw-------   1 f098     cru      1904005440 Feb 13 10:35 cru_ts_2_10.1901-2002.cld.grid
-rw-------   1 f098     cru         7542905 Feb 12 17:50 cru_ts_3_00.1996.2000.cld.dat.gz
-rw-------   1 f098     cru        62217668 Feb 12 17:50 cru_ts_3_00.1996.2000.cld.nc
-rw-------   1 f098     cru       273762720 Feb 12 17:51 cru_ts_3_00.1996.2006.cld.dat
-rw-------   1 f098     cru       136867556 Feb 12 17:51 cru_ts_3_00.1996.2006.cld.nc
-rw-------   1 f098     cru         8893264 Feb 12 17:51 cru_ts_3_00.2001.2006.cld.dat.gz
-rw-------   1 f098     cru        74659316 Feb 12 17:51 cru_ts_3_00.2001.2006.cld.nc
crua6[/cru/cruts/version_3_0/secondaries/cld/cld_final] gunzip cru_ts_3_00.2001.2006.cld.dat.gz
crua6[/cru/cruts/version_3_0/secondaries/cld/cld_final] wc -l cru_ts_3_00.2001.2006.cld.dat
     25920 cru_ts_3_00.2001.2006.cld.dat
crua6[/cru/cruts/version_3_0/secondaries/cld/cld_final] tail -17280 cru_ts_3_00.2001.2006.cld.dat >cru_ts_3_00.2003.2006.cld.dat
crua6[/cru/cruts/version_3_0/secondaries/cld/cld_final] wc -l cru_ts_3_00.2003.2006.cld.dat
     17280 cru_ts_3_00.2003.2006.cld.dat
crua6[/cru/cruts/version_3_0/secondaries/cld/cld_final] mv cru_ts_2_10.1901-2002.cld.grid cru_ts_3.00.1901.2006.cld.dat
crua6[/cru/cruts/version_3_0/secondaries/cld/cld_final] cat cru_ts_3_00.2003.2006.cld.dat >>cru_ts_3.00.1901.2006.cld.dat
crua6[/cru/cruts/version_3_0/secondaries/cld/cld_final] wc -l cru_ts_3.00.1901.2006.cld.dat
    457920 cru_ts_3.00.1901.2006.cld.dat
crua6[/cru/cruts/version_3_0/secondaries/cld/cld_final] cmp ../cru_ts_2_10.1901-2002.cld.grid cru_ts_3.00.1901.2006.cld.dat
cmp: EOF on ../cru_ts_2_10.1901-2002.cld.grid
crua6[/cru/cruts/version_3_0/secondaries/cld/cld_final]

Done and tested.

So.. back to the update process :-(

Well to take a slightly different tack, I thought I'd look at the gridding end of
things. Specifically, how to run IDL in batch mode. I think I've got it: you create
a batch file with the command(s) in, then setenv IDL_STARTUP [name of batch file].
When you type 'idl' it runs the batch file, unfortunately it doesn't quit afterwards,
though adding an 'exit' line to the batch file does the trick! Of course, there is no
easy way to check it's working properly, since the random element (used when relaxing
to the climatology) ensures that each run gives different results:

crua6[/cru/cruts/version_3_0/secondaries/cld] cmp testglo/cld.2004.11.glo testglo2/cld.2004.11.glo
testglo/cld.2004.11.glo testglo2/cld.2004.11.glo differ: char 9863, line 104
crua6[/cru/cruts/version_3_0/secondaries/cld]

Still, the mechanism is so similar to that used to run other Fortran progs that we
can carry on, I guess. Naturally I would prefer to use the gridder I wrote, partly
because it does a much better, *documentable* job, but mainly because I don'y want
all that effort wasted!

Also looked at NetCDF production, as it's still looming. ncgen looks quite good, it
can work from a 'CDL' file (format is the same as the output from ncdump). It can
even produce fortran code to reproduce the file!!

Ah well. Back to the 'incoming data' process. The fact that the mcdw2cruauto and
climat2cruauto programs worked fine for CLD is a big bonus, they read their runs
and date files andthey wrote their results. Though the results didn't include the
names of the output databases, I've had second thoughts about that. I want the
update program to be in charge, so it should know what files have been produced
(assuming the result is 'OK'). If the conversion program sends back a list, then
the update program will have to parse it to find out which parameter is which,
and that's silly when it should know anyway!! The situation is different for
merging. I don't have a full strategy for file naming yet. Let's look at a typical
process for an unnamed (not tmn or tmx) primary parameter, ie simple case:

File(s)                     Process
mcdw update(s)
                            convert mcdw
mcdw db
current db
                            merge mcdw into current
current+mcdw db

climat update(s)
                            convert climat
climat db
                            merge climat into current+mcdw
current+mcdw+climat db
                            anomalise
anomaly files
                            grid
gridded anomalies
climatology
                            actualise
gridded actuals
                            reformat into .dat and .nc
final output files

So, naming. Well the governing principle of the update process is that all files
have the same 10-digit datestamp. So the run can be uniquely identified, as can
all its files (data, log, etc). I am NOT changing that! A main problem is that
we will have to depart from the rigid database naming schema ('tla.datestr.dtb')
because we will have lots of databases in a single run. In the above example,
four databases will all have the same datestamp. Here's a possible name system:

mcdw db                     mcdw.tla.datestr.dtb
current+mcdw db             int1.tla.datestr.dtb
climat db                   clmt.tla.datestr.dtb
current+mcdw+climat db      int2.tla.datestr.dtb

The final db would then be copied or renamed to:

tla.datestr.dtb
pre.0902161401.dtb

For secondary parameters it's even worse! I'm not super-keen on the use of 'int1'
('interim 1') and so on.. they give no useful information. But a more complicated
schema isn't going to be uderstood by anyone else anyway! And we should have the
Database Master List to refer to at all times.. okay. All interim databases will
```

be labeled 'int1', 'int2', and so forth. The update program will have to keep
track of numbering. And, of course - it will have to tell the merging program
what to call the output database! Bah.

It gets WORSE. The update program has to know which 'Master' database to pass to
the merge program. For MCDW, it's going to be the 'current' database for that
parameter. But for CLIMAT and BOM, it depends on whether MCDW or CLIMAT
(respectively) merges have gone before. And only for those parameters that are
precursored! More complexity. Well, I suppose I can take one of two approaches:

1. Test at each stage for each parameter (ie for BOM, test whether CLIMAT tmx/tmn
   have just been done). This could be done by testing for the filenames or by
   setting flags.
2. Maintain a list in memory of 'latest' databases for each parameter. A bit less
   elegant, but easier to understand and use.

Well, as we already HAVE (2), we'll go with that one ;0).

Okay. Because it is so complicated (well, for my brain anyway), I'm going to write
out the filenames that update is using and expecting, so I can check that the
conversion and merging programs tie in.

```
INITS/ASSUMPTIONS
dtstr = 0902161655
par = TMP
source = MCDW
prev db = db/tmp/tmp.0809111204.dtb

CONVERSION
runs/runs.0902161655/conv.mcdw.0902161655.dat           Run information
updates/MCDW/db/db.0902161655                           Dir for output dbs
results/results.0902161655/conv.mcdw.0902161655.res     Expected results file
updates/MCDW/db/db.0902161655/mcdw.tmp.0902161655.dtb   Expected output db
logs/logs.0902161655/conv.mcdw.0902161655.log           Expected log file

MERGING
db/tmp/tmp.0809111204.dtb                               Current/latest db
updates/MCDW/db/db.0902161655/mcdw.tmp.0902161655.dtb   New db to be merged in
updates/MCDW/db/db.0902161655/int1.tmp.0902161655.dtb   Interim output db
runfile.latest.dat                                      Contains name of current run file
runs/runs.0902161655/merg.mcdw.0902161655.dat           Run information (read from above)
results/results.0902161655/merg.mcdw.0902161655.res     Expected results file
updates/MCDW/db/db.0902161655/int1.tmp.0902161655.dtb   Expected output db
logs/logs.0902161655/merg.mcdw.0902161655.log           Expected log file
```

These all seem to match up with the respective programs! Not sure that all
the necessary directories are being created yet, though.. they are now. Some
modifications to the above have been made (and retrospectively updated).

So, with half of the update program written, I got it all compiled, reset all
the incoming data to 'unprocessed', and.. got it working!

Of course, I immediately realised that I'd missed out the DTR conversion at the end.
And that.. didn't go any better than the rest of it, despite a quick conversion of
tmnx2dtrauto.for.

Well, keen-eyed viewers will remember that all the tmin/tmax/dtr/back-to-tmin-and-tmax
stuff revolves around the tmin and tmax databases being kept in absolute step. That is,
same stations, same coordinates and names, same data spans. Otherwise the job of
synching, and of converting to DTR, becomes horrendous. But look at what happens to the
line counts of the databases as they're mangled through the system:

```
originals ** identical metadata **
    606244 tmn/tmn.0708071548.dtb
    606244 tmx/tmx.0708071548.dtb

climat conversions
     27090 climat.tmn.0902192248.dtb
     27080 climat.tmx.0902192248.dtb

climat merged interims
    607692 int2.tmn.0902192248.dtb
    604993 int2.tmx.0902192248.dtb

bom conversions ** identical metadata **
      5388 bom.tmn.0902192248.dtb
      5388 bom.tmx.0902192248.dtb

bom merged (into climat interims) interims
    607692 int3.tmn.0902192248.dtb
    604993 int3.tmx.0902192248.dtb
```

Sometimes life is just too hard. It's after midnight - again. And I'm doing all this
over VNC in 256 colours, which hurts. Anyway, the above line counts. I don't know
which is the more worrying - the fact that adding the CLIMAT updates lost us 1251
lines from tmax but gained us 1448 for tmin, or that the BOM additions added sod all.
And yes - I've checked, the int2 and int3 databases are IDENTICAL. Aaaarrgghhhhh.

I guess.. I am going to need one of those programs I wrote to sync the tmin and tmax
databases, aren't I?

Actually, it's worse than that. The CLIMAT merges for TMN and TMX look very similar:

```
<QUOTE CLIMAT TMN MERGE INTO LATEST DB>
New master database: updates/CLIMAT/db/db.0902192248/int2.tmn.0902192248.dtb

Update database stations:         2922
 > Matched with Master stations:  2227
                 (automatically:  2227)
                   (by operator:     0)
 > Added as new Master stations:   566
 > Rejected:                       129
   Rejects file:                 updates/CLIMAT/db/db.0902192248/climat.tmn.0902192248.dtb.rejected
<END QUOTE>

<QUOTE CLIMAT TMX MERGE INTO LATEST DB>
New master database: updates/CLIMAT/db/db.0902192248/int2.tmx.0902192248.dtb

Update database stations:         2921
 > Matched with Master stations:  2226
                 (automatically:  2226)
                   (by operator:     0)
 > Added as new Master stations:   566
 > Rejected:                       129
   Rejects file:                 updates/CLIMAT/db/db.0902192248/climat.tmx.0902192248.dtb.rejected
<END QUOTE>
```

I don't see how we end up with such drastic differences in line counts!!

Well the first thing to do was to fix climat2cruauto so that it treated tmin and tmax as

```
inseparable. Thus the CLIMAT databases for these two should be identical (um, apart from
the data values).

OK, this is getting SILLY. Now the BOM and CLIMAT conversions are in sync, and the original
databases are in synch, yet the processing creates massive divergence!!

originals
    606244 db/tmn/tmn.0708071548.dtb
    606244 db/tmx/tmx.0708071548.dtb

climat conversions
     27080 updates/CLIMAT/db/db.0902201023/climat.tmn.0902201023.dtb
     27080 updates/CLIMAT/db/db.0902201023/climat.tmx.0902201023.dtb

climat merged interims
    607687 updates/CLIMAT/db/db.0902201023/int2.tmn.0902201023.dtb
    604987 updates/CLIMAT/db/db.0902201023/int2.tmx.0902201023.dtb

bom conversions ** identical metadata **
      5388 updates/BOM/db/db.0902201023/bom.tmn.0902201023.dtb
      5388 updates/BOM/db/db.0902201023/bom.tmx.0902201023.dtb

bom merged (into climat interims) interims
    607687 updates/BOM/db/db.0902201023/int3.tmn.0902201023.dtb
    604987 updates/BOM/db/db.0902201023/int3.tmx.0902201023.dtb

So the behaviour of newmergedbauto is, for want of a better word, unpredictable. Oh, joy.
And, as indicated, the BOM updates are totally rejected:

<QUOTE BOM TMN MERGE INTO INTERIM DB>
New master database: updates/BOM/db/db.0902201023/int3.tmn.0902201023.dtb

Update database stations:          898
 > Matched with Master stations:     0
                 (automatically:     0)
                   (by operator:     0)
 > Added as new Master stations:     0
 > Rejected:                       898
   Rejects file:                updates/BOM/db/db.0902201023/bom.tmn.0902201023.dtb.rejected
<END QUOTE>

<QUOTE BOM TMX MERGE INTO INTERIM DB>
Update database stations:          898
 > Matched with Master stations:     0
                 (automatically:     0)
                   (by operator:     0)
 > Added as new Master stations:     0
 > Rejected:                       898
   Rejects file:                updates/BOM/db/db.0902201023/bom.tmx.0902201023.dtb.rejected
<END QUOTE>

I really thought I was cracking this project. But every time, it ends up worse than before.

OK, let's try and work out the order of events. I'm using getheads to look at metadata only.

1. CLIMAT conversions. These seem to be working fine:

crua6[/cru/cruts/..CLIMAT/db/db.0902201023] cmp climat.tmn.0902201023.hds climat.tmx.0902201023.hds
crua6[/cru/cruts/..CLIMAT/db/db.0902201023]

2. Original databases. They look OK:

crua6[/cru/cruts/version_3_0/update_top/db] cmp tmn/tmn.0708071548.hds tmx/tmx.0708071548.hds
crua6[/cru/cruts/version_3_0/update_top/db]

3. CLIMAT merging into original databases. Bad, bad, bad.

crua6[/cru/cruts/..CLIMAT/db/db.0902201023] diff int2.tmn.0902201023.hds int2.tmx.0902201023.hds |wc -l
       4848
crua6[/cru/cruts/../CLIMAT/db/db.0902201023]

Something is very poorly. It's my programming skills, isn't it.

Looking at the log files for the CLIMAT merging, they give identical stats! what differ are
the dates, ie:

<QUOTE DIFFS BETWEEN CLIMAT MERGE LOGS>
crua6[/cru/cruts/version_3_0/update_top/logs/logs.0902201023] diff merg.climat.tmn.0902201023.log merg.climat.tmx.0902201023.log |more
1,2c1,2
< Master file: db/tmn/tmn.0708071548.dtb
< Update file: updates/CLIMAT/db/db.0902201023/climat.tmn.0902201023.dtb
---
> Master file: db/tmx/tmx.0708071548.dtb
> Update file: updates/CLIMAT/db/db.0902201023/climat.tmx.0902201023.dtb
281c281
< code match with: 1033800  5247    970   55 HANNOVER             DL GM         1927 2006    -999       0
---
> code match with: 1033800  5247    970   55 HANNOVER             DL GM         1930 2006    -999       0
287c287
< code match with: 1038400  5247   1340   49 BERLIN-TEMPELHOF     GERMANY       1991 2006    -999       0
---
> code match with: 1038400  5247   1340   49 BERLIN-TEMPELHOF     GERMANY       1929 2006    -999       0
<END QUOTE>

..and so on. What's got me stumped is that the headers of both pairs of input databases
are IDENTICAL. These dates are spurious! Look:

crua6[/cru/cruts/version_3_0/update_top/db] grep '55 HANNOVER' tmn/tmn.0708071548.dtb
1033800  5247    970   55 HANNOVER             DL GM         1927 2006    -999       0
crua6[/cru/cruts/version_3_0/update_top/db] grep '55 HANNOVER' tmx/tmx.0708071548.dtb
1033800  5247    970   55 HANNOVER             DL GM         1927 2006    -999       0
crua6[/cru/cruts/version_3_0/update_top/db] grep '49 BERLIN-TEMPELHOF' tmn/tmn.0708071548.dtb
1038400  5247   1340   49 BERLIN-TEMPELHOF     GERMANY       1929 2006    -999       0
crua6[/cru/cruts/version_3_0/update_top/db] grep '49 BERLIN-TEMPELHOF' tmx/tmx.0708071548.dtb
1038400  5247   1340   49 BERLIN-TEMPELHOF     GERMANY       1929 2006    -999       0
crua6[/cru/cruts/version_3_0/update_top/db]

You see? The HANNOVER 1930 date, and the BERLIN-TEMPELHOF 1991 date, are wrong!! Christ.
That's not even consistent, one's supposedly in the tmin file, the other, the tmax one.

So, an apparently-random pollution of the start dates. And.. FOUND IT! As usual, the program is
doing exactly what I asked it to do. When I wrote it I simply didn't consider the possibility
of tmin and tmax needing to sync. So one of the first things it does, when reading in the
exisitng database, is to truncate station data series where whole years are missing values. And
for HANNOVER, tmax has 1927-1929 missing, but tmin has (some) data in those years. A-ha!

What to do.. I guess the logical thing to do is to not truncate for tmin and tmax! So I added a
flag to newmergedbauto, that it passes to the 'getmos' subroutine, that stops it from replacing
start and end years, and.. it worked!! Hurrah! Or, well.. it ran without giving any errors or
crashing horribly. Yes, that's it. And here are all the 142 files (and directories) it created:
```

```
crua6[/cru/cruts/version_3_0/update_top] find . -name '*0902201545*'
./results/results.0902201545
./results/results.0902201545/conv.mcdw.0902201545.res
./results/results.0902201545/merg.mcdw.tmp.0902201545.res
./results/results.0902201545/merg.mcdw.pre.0902201545.res
./results/results.0902201545/merg.mcdw.vap.0902201545.res
./results/results.0902201545/merg.mcdw.wet.0902201545.res
./results/results.0902201545/merg.mcdw.cld.0902201545.res
./results/results.0902201545/conv.climat.0902201545.res
./results/results.0902201545/merg.climat.tmp.0902201545.res
./results/results.0902201545/merg.climat.vap.0902201545.res
./results/results.0902201545/merg.climat.wet.0902201545.res
./results/results.0902201545/merg.climat.pre.0902201545.res
./results/results.0902201545/merg.climat.cld.0902201545.res
./results/results.0902201545/merg.climat.tmn.0902201545.res
./results/results.0902201545/merg.climat.tmx.0902201545.res
./results/results.0902201545/conv.bom.0902201545.res
./results/results.0902201545/merg.bom.tmn.0902201545.res
./results/results.0902201545/merg.bom.tmx.0902201545.res
./results/results.0902201545/mdtr.0902201545.res
./runs/runs.0902201545
./runs/runs.0902201545/conv.mcdw.0902201545.dat
./runs/runs.0902201545/merg.mcdw.0902201545.dat
./runs/runs.0902201545/conv.climat.0902201545.dat
./runs/runs.0902201545/merg.climat.0902201545.dat
./runs/runs.0902201545/conv.bom.0902201545.dat
./runs/runs.0902201545/merg.bom.0902201545.dat
./runs/runs.0902201545/mdtr.0902201545.dat
./db/tmp/tmp.0902201545.dtb
./db/tmn/tmn.0902201545.dtb
./db/tmx/tmx.0902201545.dtb
./db/dtr/dtr.0902201545.dtb
./db/pre/pre.0902201545.dtb
./db/vap/vap.0902201545.dtb
./db/wet/wet.0902201545.dtb
./db/cld/cld.0902201545.dtb
./updates/BOM/db/db.0902201545
./updates/BOM/db/db.0902201545/bom.tmn.0902201545.dtb
./updates/BOM/db/db.0902201545/bom.tmx.0902201545.dtb
./updates/BOM/db/db.0902201545/int3.tmn.0902201545.dtb
./updates/BOM/db/db.0902201545/bom.tmn.0902201545.dtb.rejected
./updates/BOM/db/db.0902201545/int3.tmx.0902201545.dtb
./updates/BOM/db/db.0902201545/bom.tmx.0902201545.dtb.rejected
./updates/BOM/db/db.0902201545/int3.dtr.0902201545.dtb
./updates/BOM/mergefiles/merg.bom.tmn.0902201545.mat
./updates/BOM/mergefiles/merg.bom.tmn.0902201545.act
./updates/BOM/mergefiles/merg.bom.tmn.0902201545.xrf
./updates/BOM/mergefiles/merg.bom.tmx.0902201545.mat
./updates/BOM/mergefiles/merg.bom.tmx.0902201545.act
./updates/BOM/mergefiles/merg.bom.tmx.0902201545.xrf
./updates/CLIMAT/mergefiles/merg.climat.tmp.0902201545.mat
./updates/CLIMAT/mergefiles/merg.climat.tmp.0902201545.act
./updates/CLIMAT/mergefiles/merg.climat.tmp.0902201545.xrf
./updates/CLIMAT/mergefiles/merg.climat.vap.0902201545.mat
./updates/CLIMAT/mergefiles/merg.climat.vap.0902201545.act
./updates/CLIMAT/mergefiles/merg.climat.vap.0902201545.xrf
./updates/CLIMAT/mergefiles/merg.climat.wet.0902201545.mat
./updates/CLIMAT/mergefiles/merg.climat.wet.0902201545.act
./updates/CLIMAT/mergefiles/merg.climat.wet.0902201545.xrf
./updates/CLIMAT/mergefiles/merg.climat.pre.0902201545.mat
./updates/CLIMAT/mergefiles/merg.climat.pre.0902201545.act
./updates/CLIMAT/mergefiles/merg.climat.pre.0902201545.xrf
./updates/CLIMAT/mergefiles/merg.climat.cld.0902201545.mat
./updates/CLIMAT/mergefiles/merg.climat.cld.0902201545.act
./updates/CLIMAT/mergefiles/merg.climat.cld.0902201545.xrf
./updates/CLIMAT/mergefiles/merg.climat.tmn.0902201545.mat
./updates/CLIMAT/mergefiles/merg.climat.tmn.0902201545.act
./updates/CLIMAT/mergefiles/merg.climat.tmn.0902201545.xrf
./updates/CLIMAT/mergefiles/merg.climat.tmx.0902201545.mat
./updates/CLIMAT/mergefiles/merg.climat.tmx.0902201545.act
./updates/CLIMAT/mergefiles/merg.climat.tmx.0902201545.xrf
./updates/CLIMAT/db/db.0902201545
./updates/CLIMAT/db/db.0902201545/climat.tmp.0902201545.dtb
./updates/CLIMAT/db/db.0902201545/climat.vap.0902201545.dtb
./updates/CLIMAT/db/db.0902201545/climat.wet.0902201545.dtb
./updates/CLIMAT/db/db.0902201545/climat.pre.0902201545.dtb
./updates/CLIMAT/db/db.0902201545/climat.cld.0902201545.dtb
./updates/CLIMAT/db/db.0902201545/climat.tmn.0902201545.dtb
./updates/CLIMAT/db/db.0902201545/climat.tmx.0902201545.dtb
./updates/CLIMAT/db/db.0902201545/int2.tmp.0902201545.dtb
./updates/CLIMAT/db/db.0902201545/climat.tmp.0902201545.dtb.rejected
./updates/CLIMAT/db/db.0902201545/int2.vap.0902201545.dtb
./updates/CLIMAT/db/db.0902201545/climat.vap.0902201545.dtb.rejected
./updates/CLIMAT/db/db.0902201545/int2.wet.0902201545.dtb
./updates/CLIMAT/db/db.0902201545/climat.wet.0902201545.dtb.rejected
./updates/CLIMAT/db/db.0902201545/int2.pre.0902201545.dtb
./updates/CLIMAT/db/db.0902201545/climat.pre.0902201545.dtb.rejected
./updates/CLIMAT/db/db.0902201545/int2.cld.0902201545.dtb
./updates/CLIMAT/db/db.0902201545/climat.cld.0902201545.dtb.rejected
./updates/CLIMAT/db/db.0902201545/int2.tmn.0902201545.dtb
./updates/CLIMAT/db/db.0902201545/climat.tmn.0902201545.dtb.rejected
./updates/CLIMAT/db/db.0902201545/int2.tmx.0902201545.dtb
./updates/CLIMAT/db/db.0902201545/climat.tmx.0902201545.dtb.rejected
./updates/MCDW/mergefiles/merg.mcdw.tmp.0902201545.mat
./updates/MCDW/mergefiles/merg.mcdw.tmp.0902201545.act
./updates/MCDW/mergefiles/merg.mcdw.tmp.0902201545.xrf
./updates/MCDW/mergefiles/merg.mcdw.pre.0902201545.mat
./updates/MCDW/mergefiles/merg.mcdw.pre.0902201545.act
./updates/MCDW/mergefiles/merg.mcdw.pre.0902201545.xrf
./updates/MCDW/mergefiles/merg.mcdw.vap.0902201545.mat
./updates/MCDW/mergefiles/merg.mcdw.vap.0902201545.act
./updates/MCDW/mergefiles/merg.mcdw.vap.0902201545.xrf
./updates/MCDW/mergefiles/merg.mcdw.wet.0902201545.mat
./updates/MCDW/mergefiles/merg.mcdw.wet.0902201545.act
./updates/MCDW/mergefiles/merg.mcdw.wet.0902201545.xrf
./updates/MCDW/mergefiles/merg.mcdw.cld.0902201545.mat
./updates/MCDW/mergefiles/merg.mcdw.cld.0902201545.act
./updates/MCDW/mergefiles/merg.mcdw.cld.0902201545.xrf
./updates/MCDW/db/db.0902201545
./updates/MCDW/db/db.0902201545/mcdw.tmp.0902201545.dtb
./updates/MCDW/db/db.0902201545/mcdw.vap.0902201545.dtb
./updates/MCDW/db/db.0902201545/mcdw.wet.0902201545.dtb
./updates/MCDW/db/db.0902201545/mcdw.pre.0902201545.dtb
./updates/MCDW/db/db.0902201545/mcdw.sun.0902201545.dtb
./updates/MCDW/db/db.0902201545/mcdw.cld.0902201545.dtb
./updates/MCDW/db/db.0902201545/int1.tmp.0902201545.dtb
./updates/MCDW/db/db.0902201545/mcdw.tmp.0902201545.dtb.rejected
./updates/MCDW/db/db.0902201545/int1.pre.0902201545.dtb
./updates/MCDW/db/db.0902201545/int1.vap.0902201545.dtb
```

```
./updates/MCDW/db/db.0902201545/mcdw.vap.0902201545.dtb.rejected
./updates/MCDW/db/db.0902201545/int1.wet.0902201545.dtb
./updates/MCDW/db/db.0902201545/mcdw.wet.0902201545.dtb.rejected
./updates/MCDW/db/db.0902201545/int1.cld.0902201545.dtb
./updates/MCDW/db/db.0902201545/mcdw.cld.0902201545.dtb.rejected
./logs/logs.0902201545
./logs/logs.0902201545/conv.mcdw.0902201545.log
./logs/logs.0902201545/merg.mcdw.tmp.0902201545.log
./logs/logs.0902201545/merg.mcdw.pre.0902201545.log
./logs/logs.0902201545/merg.mcdw.vap.0902201545.log
./logs/logs.0902201545/merg.mcdw.wet.0902201545.log
./logs/logs.0902201545/merg.mcdw.cld.0902201545.log
./logs/logs.0902201545/conv.climat.0902201545.log
./logs/logs.0902201545/merg.climat.tmp.0902201545.log
./logs/logs.0902201545/merg.climat.vap.0902201545.log
./logs/logs.0902201545/merg.climat.wet.0902201545.log
./logs/logs.0902201545/merg.climat.pre.0902201545.log
./logs/logs.0902201545/merg.climat.cld.0902201545.log
./logs/logs.0902201545/merg.climat.tmn.0902201545.log
./logs/logs.0902201545/merg.climat.tmx.0902201545.log
./logs/logs.0902201545/conv.bom.0902201545.log
./logs/logs.0902201545/merg.bom.tmn.0902201545.log
./logs/logs.0902201545/merg.bom.tmx.0902201545.log
./logs/logs.0902201545/mdtr.0902201545.log
crua6[/cru/cruts/version_3_0/update_top]
```

So, this leaves the new databases in the db/xxx/ directories, and db/latest.versions.dat telling us which ones they are. Which should be all the next suite of programs needs to create the final output files. Eeeeeeeek.

Well for this 'half' of the process it's going to be 90% planning and strategy - because that's how the first half ended up.

Let's revisit the process list from earlier - just the database-onwards bits and interactivity removed:

```
* Produce Primary Parameters (TMP, TMN, TMX, DTR, PRE)
anomdtb (per parameter)
quick_interp_tdm2 (per parameter)
glo2abs (per parameter)
makegrids (per parameter)
* Prepare Binary Grids (TMP, DTR, PRE) for Synthetics
quick_interp_tdm2 (per parameter)
* Produce Secondary Parameter (FRS, uses TMP,DTR)
frs_gts_tdm
quick_interp_tdm2
glo2abs
makegrids
* Produce Secondary Parameter (VAP, uses TMP,DTR)
vap_gts_anom
anomdtb
quick_interp_tdm2
glo2abs
makegrids
* Produce Secondary Parameter (WET/RD0, uses PRE)
rd0_gts_anom
anomdtb
quick_interp_tdm2
glo2abs
makegrids
* Produce Secondary Parameter (CLD, uses DTR)
anomdtb (95-02 norm period)
movenorms
dtr2cld
quick_interp_tdm2
glo2abs
makegrids
```

Having drawn out the process flowchart, I wondered if quick_interp_tdm2.pro would be kind enough to output both .glo and binary gridded files, simultaneously? This would simplify and speed things up a bit. So, with absolutely no alarm bells ringing at all, I decided to make a sample run for DTR, just for 2006, to compare simultaneous outputs with the original ones. You idiot.

```
IDL> quick_interp_tdm2,2006,2006,'testdtrglo/dtr.',750,gs=0.5,pts_prefix='dtrtxt/dtr.',dumpglo='dumpglo',dumpbin='dumpbin'
% Compiled module: QUICK_INTERP_TDM2.
% Compiled module: GLIMIT.
Defaults set
    2006
% Compiled module: MAP_SET.
% Compiled module: CROSSP.
% Compiled module: STRIP.
% Compiled module: SAVEGLO.
% Compiled module: SELECTMODEL.
no stations found in: dtrtxt/dtr.2006.09.txt
no stations found in: dtrtxt/dtr.2006.10.txt
no stations found in: dtrtxt/dtr.2006.11.txt
no stations found in: dtrtxt/dtr.2006.12.txt
% Compiled module: MEAN.
% Compiled module: MOMENT.
% Compiled module: STDDEV.
grid 2006 non-zero    0.1125    2.1122    2.9219 cells=   202010
% Compiled module: WRBIN.
IDL> exit
crua6[/cru/cruts/version_3_0/primaries/dtr] ls -l testdtrglo/
total 43048
-rw-------   1 f098     cru      6220800 Feb 23 11:06 dtr.2006
-rw-------   1 f098     cru      3142986 Feb 23 11:06 dtr.2006.01.glo
-rw-------   1 f098     cru      3142986 Feb 23 11:06 dtr.2006.02.glo
-rw-------   1 f098     cru      3142986 Feb 23 11:06 dtr.2006.03.glo
-rw-------   1 f098     cru      3142986 Feb 23 11:06 dtr.2006.04.glo
-rw-------   1 f098     cru      3142986 Feb 23 11:06 dtr.2006.05.glo
-rw-------   1 f098     cru      3142986 Feb 23 11:06 dtr.2006.06.glo
-rw-------   1 f098     cru      3142986 Feb 23 11:06 dtr.2006.07.glo
-rw-------   1 f098     cru      3142986 Feb 23 11:06 dtr.2006.08.glo
-rw-------   1 f098     cru      3142986 Feb 23 11:06 dtr.2006.09.glo
-rw-------   1 f098     cru      3142986 Feb 23 11:06 dtr.2006.10.glo
-rw-------   1 f098     cru      3142986 Feb 23 11:06 dtr.2006.11.glo
-rw-------   1 f098     cru      3142986 Feb 23 11:06 dtr.2006.12.glo
crua6[/cru/cruts/version_3_0/primaries/dtr]
```

So there, as hoped-for, binary and text output files. BUT. Comparisons with earlier versions from the same database.. are depressingly awful:

```
crua6[/cru/cruts/version_3_0/primaries/dtr] diff dtr.2006.01.glo testdtrglo/dtr.2006.01.glo |wc -l
     33484
crua6[/cru/cruts/version_3_0/primaries/dtr]
```

Sample comparison of lines 700-710 from old and new glo files:

```
crua6[/cru/cruts/version_3_0/primaries/dtr] head -710 dtr.2006.01.glo |tail -11
```

```
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  3.9705E-04
  1.1257E-02  2.2117E-02  3.1641E-02  8.9739E-03  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
crua6[/cru/cruts/version_3_0/primaries/dtr] head -710 testdtrglo/dtr.2006.01.glo | tail -11
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  1.7088E-03  8.5614E-04
  3.4384E-06  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
crua6[/cru/cruts/version_3_0/primaries/dtr]
```

They're NOTHING LIKE EACH OTHER. I really do hate this whole project. Ran the gridder again, just for text output.. and..

```
crua6[/cru/cruts/version_3_0/primaries/dtr] head -710 testdtrglo2/dtr.2006.01.glo | tail -11
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  3.1268E-03
  6.5528E-03  9.9787E-03  1.3405E-02  1.6831E-02  9.7796E-03  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00  0.0000E+00
crua6[/cru/cruts/version_3_0/primaries/dtr]
```

Different again! Can this just be the random seed used in the gridding algorithm? If so, why aren't we seeing a consistent pattern of 0.0 vs non-0.0 values? Another reason - if one were needed - why we should dump this gridding approach altogether. But, er, not yet! No time to finish and test the fortran gridder, which will doubtless sink to some depth and never be seen again, we'll carry on with this mediocre approach.

Spent a whole day knocking up an anomaly program - as I felt anomdtb was vastly overweight and supremely complicated to compile. Unfortunately, I got stuck trying to work out data and latlon factors for different parameters, (argh! why?), and what percentage anomalies really were, and in the end GAVE UP and now I have to modify anomdtb after all. Actually - that looked even worse, so went back to anomauto and finished it off. And.. it works. Actually, a bit too well. For example, when deriving anomalies from the CLD database, this was the original (a few weeks ago!):

```
uealogin1[/cru/cruts/version_3_0/update_top] wc -l cld.2000.11.txt
     606 cld.2000.11.txt
```

..and this is the new one, from the same source database of course:

```
uealogin1[/cru/cruts/version_3_0/update_top] wc -l interim_data/anoms/anoms.0902201545/cld/cld.2000.11.txt
    1282 interim_data/anoms/anoms.0902201545/cld/cld.2000.11.txt
```

..so, um - more than twice as many got through! Erk. Screening not tough enough! Results also not exactly identical (> indicates potential match):

```
OLD:
uealogin1[/cru/cruts/version_3_0/update_top] head -10 cld.2000.11.txt
 > 68.27   22.30   327.0    -21.20000 208000
 > 65.83   24.15     6.0    -41.00000 219600
 > 63.18   14.50   370.0    -45.30000 222600
 > 59.37   13.47    55.0    -43.90000 241800
 > 57.78   11.88    53.0    -35.50000 251200
 > 57.67   18.35    47.0    -42.10000 259000
   69.75   27.03   101.0    -23.80000 280500
   67.37   26.65   179.0    -24.30000 283600
   64.93   25.37    15.0    -33.40000 287500
   62.40   25.68   145.0     -0.80000 293500

NEW:
uealogin1[/cru/cruts/version_3_0/update_top/interim_data/anoms/anoms.0902201545/cld] head -10 cld.2000.11.txt
   67.27   14.37    13.0      4.85715 115200
 > 68.27   22.30   327.0      8.56250 208000
 > 65.83   24.15     6.0     16.59999 219600
 > 63.18   14.50   370.0      3.26250 222600
 > 59.37   13.47    55.0     15.33749 241800
 > 57.78   11.88    53.0      8.93749 251200
 > 57.67   18.35    47.0      6.58749 259000
   60.13   -1.18    84.0     -7.02500 300500
   58.22   -6.32    13.0     -1.22501 302600
   57.20   -2.22    65.0     -7.80000 309100
```

OK, let's look at the means being used. Here's an example:

```
     lat     lon     alt         anom    wmo    mean
   68.27   22.30   327.0      8.56250 208000   90.14
```

and the actual Nov 2000 value for this station (KARESUANDO, SWEDEN) is 987:

```
2000  887  800  900-9999-9999  812  762  825  625  825  987-9999
```

OK. So we read in 987. Then we multiply by the factor, which should be 0.1, giving us 98.7.

Then we subtract the mean, giving us 98.7-90.14 = 8.56, which is what we're getting. So no mismatches between data, time, and metadata. Good. and the 95/02 mean is right, too (90.1375).

So, er. AH! solved it. Looking at the wrong 'old' cloud text files. tadaa:

```
OLD BUT CORRECT:
crua6[/cru/cruts/version_3_0/update_top] head -10 ../secondaries/cld/cldupdatetxt/cldupdate.2000.11.txt
   68.27   22.30   327.0      8.50000 208000
   65.83   24.15     6.0     16.60000 219600
   63.18   14.50   370.0      3.20000 222600
   59.37   13.47    55.0     15.30000 241800
   57.78   11.88    53.0      8.90000 251200
   57.67   18.35    47.0      6.50000 259000
   69.75   27.03   101.0      8.90000 280500
   67.37   26.65   179.0      9.40000 283600
   64.93   25.37    15.0     11.20000 287500
   62.40   25.68   145.0      5.90000 293500
```

Hurrah. Now I need to know why I'm producing too many. It's not as bad, though:

OLD BUT CORRECT:
```
crua6[/cru/cruts/version_3_0/update_top] wc -l ../secondaries/cld/cldupdatetxt/cldupdate.2000.11.txt
     760 ../secondaries/cld/cldupdatetxt/cldupdate.2000.11.txt
```

NEW:
```
uealogin1[/cru/cruts/version_3_0/update_top] wc -l interim_data/anoms/anoms.0902201545/cld/cld.2000.11.txt
    1282 interim_data/anoms/anoms.0902201545/cld/cld.2000.11.txt
```

Let's look at the first example, a station we let through that anomdtb kicked back:

```
0115200  6727   1437   13 BODO VI (CIV/MIL)    NORWAY        1995 2008    -999       0
6190-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1995-9999-9999-9999-9999-9999-9999-9999-9999-9999  875-9999-9999
1996-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1997-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
1998-9999-9999-9999-9999  575  675  762  762  675  837  775-9999
1999 1000  812  762  750  550  750  862  775  637  825 1000-9999
2000 1000  912  800  750  812  850  737  825  700  737  862-9999
2001  875  750  475  650  775  775  825  825  750  900 1000-9999
2002  800  862  750  737  612  612-9999  562  800  462  762-9999
2003  850  825  862  550  712-9999  525  775  762  750  825-9999
2004  937  875  762  525  637  725  787  675  837  750 1000-9999
2005 1000  812  762  700  737  775  687  800  850  850-9999-9999
2006-9999  850  500  612-9999-9999  800  575  812  750  962-9999
2007 1000  712  750  837  762  687  675  812  850  975  950-9999
2008 1000  887  687-9999  750  775  675  612  725  887-9999-9999
```

Now, our limit for a valid normal is 75%, which for 1995-2002 should mean 6.

BODO VI has five valid values in November. So our limit is either wrong, or not being applied.

..yup:

```
uealogin1[/cru/cruts/version_3_0/update_top] ./anomauto
minn calculated as        7
```

Ho hum. Recalculated it to 6 (whilst checking that 1961-1990 still gave 23). Re-ran.

To my horror - if not surprise - that let EVEN MORE IN! Well of course it did you silly sausage.
This still doesn't explain how BODO VI gets in with 5 values:

```
uealogin1[/cru/cruts/version_3_0/update_top] wc -l interim_data/anoms/anoms.0902201545/cld/cld.2000.11.txt
    1404 interim_data/anoms/anoms.0902201545/cld/cld.2000.11.txt
```

Aha. I wonder if I'm initialising the onestn() array in the wrong place? Because data is
only added if not -9999, so it has to be prefilled with -9999 *every time*.. dammit. If
I fix that, I get:

```
uealogin1[/cru/cruts/version_3_0/update_top] wc -l interim_data/anoms/anoms.0902201545/cld/cld.2000.11.txt
     746 interim_data/anoms/anoms.0902201545/cld/cld.2000.11.txt
```

14 stations LESS than the previous exercise. That'll do, surely?

OLD RELIABLE:
```
crua6[/cru/cruts/version_3_0/update_top] head -10 ../secondaries/cld/cldupdatetxt/cldupdate.2000.11.txt
   68.27   22.30   327.0      8.50000 208000
   65.83   24.15     6.0     16.60000 219600
   63.18   14.50   370.0      3.20000 222600
   59.37   13.47    55.0     15.30000 241800
   57.78   11.88    53.0      8.90000 251200
   57.67   18.35    47.0      6.50000 259000
   69.75   27.03   101.0      8.90000 280500
   67.37   26.65   179.0      9.40000 283600
   64.93   25.37    15.0     11.20000 287500
   62.40   25.68   145.0      5.90000 293500
```

NEW LATEST:
```
uealogin1[/cru/cruts/version_3_0/update_top] head interim_data/anoms/anoms.0902201545/cld/cld.2000.11.txt
   68.27   22.30   327.0      8.56250 208000   90.14
   65.83   24.15     6.0     16.59999 219600   83.40
   63.18   14.50   370.0      3.26250 222600   85.44
   59.37   13.47    55.0     15.33749 241800   84.66
   57.78   11.88    53.0      8.95714 251200   86.04
   57.67   18.35    47.0      6.58749 259000   85.91
   69.75   27.03   101.0      8.95714 280500   91.04
   67.37   26.65   179.0      9.47143 283600   90.53
   64.93   25.37    15.0     11.27142 287500   88.73
   62.40   25.68   145.0      5.90000 293500   94.10
```

It's not going to be easy to find 14 missing stations, is it? Since the anomalies aren't exactly the same.

Should I be worried about 14 lost series? Less than 2%. Actually, I noticed something interesting.. look
at the anomalies. The anomdtb ones aren't *rounded* to 1dp, they're *truncated*! So, er - wrong!

So let's say, anomalies are done. Hurrah. Onwards, plenty more to do!

Got the gridding working, I think. IDL of course. I modified quick_interp_tdm2.pro to accept
start and end months, otherwise it just produces whole years with files of zeros for months with
no anomaly file. And errors. And since this is likely to be a six-month update..

Re-planned the program layout. Not a major exercise, just putting different loops in to speed up and
simplify operations. It now runs as follows (note this is simplified!!):

1. User chooses update databases or update datasets. Dates, parameters, etc.

2. Update Databases
2.1 Convert any MCDW bulletins to CRU format; merge into existing databases
2.2 Convert any CLIMAT bulletins to CRU format; merge into databases from 2.1
2.3 Convert any BOM bulletins to CRU format; merge into databases from 2.2

3. Update datasets
3.1 Convert databases to anomalies
3.2 Grid primary parameters
3.3 Generate synthetic secondary parameters
3.4 Grid secondary parameters
3.5 Convert gridded anomalies to actuals
3.6 Produce final datasets

1876 lines including subrotuines and notes. Ten Fortran and four IDL programs (plus indirect ones). All
Fortran programs are mine, now. Top-level listing:

```
drwx------ 10 f098 cru   4096 Feb 19 20:55 db
drwx------  3 f098 cru   4096 Feb 28 17:01 reference
drwx------  3 f098 cru   4096 Mar  1 15:41 runs
drwx------  4 f098 cru   4096 Feb 23 12:15 gridded_finals
drwx------  4 f098 cru   4096 Feb 27 17:56 results
```

```
drwx------  5 f098 cru   4096 Mar  1 15:40 logs
drwx------  6 f098 cru   4096 Dec 18 11:00 updates
drwx------  8 f098 cru   4096 Feb 28 16:15 interim_data
-rw-------  1 f098 cru     11 Feb 27 17:48 newdata.latest.date
-rwxr-xr-x  1 f098 cru 132425 Mar  1 14:41 update
-rwxr-xr-x  1 f098 cru  16465 Mar  1 14:41 dtr2cldauto
-rwxr-xr-x  1 f098 cru  17990 Mar  1 14:55 tmnx2dtrauto
-rwxr-xr-x  1 f098 cru  19427 Mar  1 15:43 glo2absauto
-rwxr-xr-x  1 f098 cru  20929 Mar  1 14:42 movenormsuato
-rwxr-xr-x  1 f098 cru  23350 Mar  1 15:42 anomauto
-rwxr-xr-x  1 f098 cru  29076 Mar  1 14:50 climat2cruauto
-rwxr-xr-x  1 f098 cru  29481 Mar  1 14:50 bom2cruauto
-rwxr-xr-x  1 f098 cru  29867 Mar  1 14:49 mcdw2cruauto
-rwxr-xr-x  1 f098 cru 323870 Mar  1 15:52 makegridsauto
-rwxr-xr-x  1 f098 cru  89515 Mar  1 16:10 newmergedbauto
```

So, to station counts. These will have to mirror section 3 above. Coverage of secondary parameters is particularly difficult - what is the best approach? To include synthetic coverage, when it's only at 2.5-degree?

No. I'm going to back my previous decision - all station count files reflect actualy obs for that parameter only. So for secondaries, you get actual obs of that parameter (ie naff all for FRS). You get the info about synthetics that enables you to use the relevant primary counts if you want to. Of course, I'm going to have to provide a combined TMP and DTR station count to satisfy VAP & FRS users. The problem is that the synthetics are incorporated at 2.5-degrees, NO IDEA why, so saying they affect particular 0.5-degree cells is harder than it should be. So we'll just gloss over that entirely ;0)

ARGH. Just went back to check on synthetic production. Apparently - I have no memory of this at all - we're not doing observed rain days! It's all synthetic from 1990 onwards. So I'm going to need conditionals in the update program to handle that. And separate gridding before 1989. And what TF happens to station counts?

OH FUCK THIS. It's Sunday evening, I've worked all weekend, and just when I thought it was done I'm hitting yet another problem that's based on the hopeless state of our databases. There is no uniform data integrity, it's just a catalogue of issues that continues to grow as they're found.

rd0_gts_anom_05 will produce half-degree .glo files from gridded pre anoms. So if we call that, we can use it, and stncounts for PRE will be authentic (as it's the sole input). Final decision: coded update.for to produce WET from obs+syn until 12/1989, syn only thereafter. WET station counts only produced until 1989, PRE must be used (with caveats) after that point.

Wrote tmpdtrstnsauto.for to produce tmp.and.dtr station counts (ie you only get a count when both parameters have a count, and even then it's the min()). The resulting counts are the effective FRS counts, and the synthetic VAP counts.

Onto PET. Tracked down the PET program from Dimitrios, way back in 2007! It uses TMP, TMN, TMX, VAP, CLD and WND (the latter as 61-90 normals from IPCC). Converted to f77 'automatic' (makepetauto.for).

Onto whole runs of the update program. With a lot of debugging!

Discovered that WMO codes are still a pain in the arse. And that I'd forgotten to match Australian updates by BOM code (last field in header) instead of WMO code - so I had to modify newmergedbauto. Also found that running fixwmos.for was less than successful on VAP, because it's already screwed:

```
uealogin1[/cru/cruts/version_3_0/update_top/db/vap] grep -i 'jan mayen' vap.0804231150.dtb
0100100  7093   -867    9 JAN MAYEN(NOR-NAVY)  NORWAY        2003 2007    -999       0
1001000  7093   -866    9 JAN MAYEN(NOR NAVY)  NORWAY        1971 2003    -999     -999
uealogin1[/cru/cruts/version_3_0/update_top/db/vap]
```

Started work on fixdupes.for, to cleanse a given database of obvious duplicate stations, but self-diverted back onto getting the whole update process compiled and running end to end. Almost immediately found that match rated in the merging were mixed. Added a section to newmergedbauto that did a quick matchmaking exercise on any update stations that failed the code matching. Just lat/lon and character fields really. Didn't seem to make a lot of difference. Here are the merge results for all updates and parameters, in the order they would have happened:

```
uealogin1[/cru/cruts/version_3_0/update_top/logs/logs.0903091631] tail merg.mcdw.tmp.0903091631.log
OUTPUT(S) WRITTEN

New master database: updates/MCDW/db/db.0903091631/int1.tmp.0903091631.dtb

Update database stations:         2802
 > Matched with Master stations:  1759
                 (automatically:  1759)
                   (by operator:     0)
 > Added as new Master stations:  1043
 > Rejected:                         0
uealogin1[/cru/cruts/version_3_0/update_top/logs/logs.0903091631] tail merg.mcdw.pre.0903091631.log
OUTPUT(S) WRITTEN

New master database: updates/MCDW/db/db.0903091631/int1.pre.0903091631.dtb

Update database stations:         2807
 > Matched with Master stations:  2783
                 (automatically:  2783)
                   (by operator:     0)
 > Added as new Master stations:    24
 > Rejected:                         0
uealogin1[/cru/cruts/version_3_0/update_top/logs/logs.0903091631] tail merg.mcdw.vap.0903091631.log

New master database: updates/MCDW/db/db.0903091631/int1.vap.0903091631.dtb

Update database stations:         2804
 > Matched with Master stations:  2677
                 (automatically:  2677)
                   (by operator:     0)
 > Added as new Master stations:   124
 > Rejected:                         3
   Rejects file:                updates/MCDW/db/db.0903091631/mcdw.vap.0903091631.dtb.rejected
uealogin1[/cru/cruts/version_3_0/update_top/logs/logs.0903091631] tail merg.mcdw.wet.0903091631.log

New master database: updates/MCDW/db/db.0903091631/int1.wet.0903091631.dtb

Update database stations:         2801
 > Matched with Master stations:  2634
                 (automatically:  2634)
                   (by operator:     0)
 > Added as new Master stations:   163
 > Rejected:                         4
   Rejects file:                updates/MCDW/db/db.0903091631/mcdw.wet.0903091631.dtb.rejected
uealogin1[/cru/cruts/version_3_0/update_top/logs/logs.0903091631] tail merg.mcdw.cld.0903091631.log

New master database: updates/MCDW/db/db.0903091631/int1.cld.0903091631.dtb

Update database stations:         2204
 > Matched with Master stations:  2199
                 (automatically:  2199)
                   (by operator:     0)
 > Added as new Master stations:     0
```

```
 > Rejected:                          5
   Rejects file:                updates/MCDW/db/db.0903091631/mcdw.cld.0903091631.dtb.rejected
uealogin1[/cru/cruts/version_3_0/update_top/logs/logs.0903091631] tail merg.climat.tmp.0903091631.log

New master database: updates/CLIMAT/db/db.0903091631/int2.tmp.0903091631.dtb

Update database stations:         3065
 > Matched with Master stations:  2629
                (automatically:  2629)
                  (by operator:     0)
 > Added as new Master stations:   345
 > Rejected:                        91
   Rejects file:                updates/CLIMAT/db/db.0903091631/climat.tmp.0903091631.dtb.rejected
uealogin1[/cru/cruts/version_3_0/update_top/logs/logs.0903091631] tail merg.climat.vap.0903091631.log

New master database: updates/CLIMAT/db/db.0903091631/int2.vap.0903091631.dtb

Update database stations:         3039
 > Matched with Master stations:  2912
                (automatically:  2912)
                  (by operator:     0)
 > Added as new Master stations:    38
 > Rejected:                        89
   Rejects file:                updates/CLIMAT/db/db.0903091631/climat.vap.0903091631.dtb.rejected
uealogin1[/cru/cruts/version_3_0/update_top/logs/logs.0903091631] tail merg.climat.wet.0903091631.log

New master database: updates/CLIMAT/db/db.0903091631/int2.wet.0903091631.dtb

Update database stations:         3047
 > Matched with Master stations:  2718
                (automatically:  2718)
                  (by operator:     0)
 > Added as new Master stations:   232
 > Rejected:                        97
   Rejects file:                updates/CLIMAT/db/db.0903091631/climat.wet.0903091631.dtb.rejected
uealogin1[/cru/cruts/version_3_0/update_top/logs/logs.0903091631] tail merg.climat.pre.0903091631.log

New master database: updates/CLIMAT/db/db.0903091631/int2.pre.0903091631.dtb

Update database stations:         3054
 > Matched with Master stations:  2801
                (automatically:  2801)
                  (by operator:     0)
 > Added as new Master stations:   229
 > Rejected:                        24
   Rejects file:                updates/CLIMAT/db/db.0903091631/climat.pre.0903091631.dtb.rejected
uealogin1[/cru/cruts/version_3_0/update_top/logs/logs.0903091631] tail merg.climat.cld.0903091631.log

New master database: updates/CLIMAT/db/db.0903091631/int2.cld.0903091631.dtb

Update database stations:         2038
 > Matched with Master stations:  1964
                (automatically:  1964)
                  (by operator:     0)
 > Added as new Master stations:    71
 > Rejected:                         3
   Rejects file:                updates/CLIMAT/db/db.0903091631/climat.cld.0903091631.dtb.rejected
uealogin1[/cru/cruts/version_3_0/update_top/logs/logs.0903091631] tail merg.climat.tmn.0903091631.log

New master database: updates/CLIMAT/db/db.0903091631/int2.tmn.0903091631.dtb

Update database stations:         2921
 > Matched with Master stations:  2406
                (automatically:  2406)
                  (by operator:     0)
 > Added as new Master stations:   387
 > Rejected:                       128
   Rejects file:                updates/CLIMAT/db/db.0903091631/climat.tmn.0903091631.dtb.rejected
uealogin1[/cru/cruts/version_3_0/update_top/logs/logs.0903091631] tail merg.climat.tmx.0903091631.log

New master database: updates/CLIMAT/db/db.0903091631/int2.tmx.0903091631.dtb

Update database stations:         2921
 > Matched with Master stations:  2406
                (automatically:  2406)
                  (by operator:     0)
 > Added as new Master stations:   387
 > Rejected:                       128
   Rejects file:                updates/CLIMAT/db/db.0903091631/climat.tmx.0903091631.dtb.rejected
uealogin1[/cru/cruts/version_3_0/update_top/logs/logs.0903091631] tail merg.bom.tmn.0903091631.log

New master database: updates/BOM/db/db.0903091631/int3.tmn.0903091631.dtb

Update database stations:          906
 > Matched with Master stations:   783
                (automatically:   783)
                  (by operator:     0)
 > Added as new Master stations:   120
 > Rejected:                         3
   Rejects file:                updates/BOM/db/db.0903091631/bom.tmn.0903091631.dtb.rejected
uealogin1[/cru/cruts/version_3_0/update_top/logs/logs.0903091631] tail merg.bom.tmx.0903091631.log

New master database: updates/BOM/db/db.0903091631/int3.tmx.0903091631.dtb

Update database stations:          906
 > Matched with Master stations:   783
                (automatically:   783)
                  (by operator:     0)
 > Added as new Master stations:   120
 > Rejected:                         3
   Rejects file:                updates/BOM/db/db.0903091631/bom.tmx.0903091631.dtb.rejected
uealogin1[/cru/cruts/version_3_0/update_top/logs/logs.0903091631]


Probably the worst story is temperature, particularly for MCDW. Over 1000 new stations! Highly
unlikely. I am tempted to blame the different lat/lon scale, but for now it will have to rest.

Still hitting the problem with TMP lats and lons being a mix of deg*10 and deg*100, it's screwing
up the station counts work (of course). Unfortunately, I did some tests and the 'original' TMP
database has the trouble, it's not my update suite :-(((

Then.. I worked it out. Sample headers from the 'original' TMP db tmp.0705101334.dtb:

  10010   709    -87   10 Jan Mayen            NORWAY        1921 2006 341921  -999.00
  10050   780    142    9 ISFJORD RADIO        NORWAY        1912 1979 101912  -999.00
  10080   783    155   28 Svalbard Lufthavn    NORWAY        1911 2006 341911  -999.00
  10100   693    162 -999 ANDENES                             1868 1955 101868  -999.00
  10250   697    189   10 TROMSO/LANGNES       NORWAY        1949 2006 101949  -999.00
  10260   697    189  100 Tromsoe              NORWAY        1890 2006 341890  -999.00
  10280   745    190   16 Bjoernoeya           NORWAY        1920 2006 341920  -999.00
```

And from the fixwmos-wmo-fixed version tmp.0903081416.dtb:

```
0100100  7090   -870   10 Jan Mayen            NORWAY        1921 2006 341921  -999.00
0100500  7800   1420    9 ISFJORD RADIO        NORWAY        1912 1979 101912  -999.00
0100800  7830   1550   28 Svalbard Lufthavn    NORWAY        1911 2006 341911  -999.00
0101000  6930   1620 -999 ANDENES                             1868 1955 101868  -999.00
0102500  6970   1890   10 TROMSO/LANGNES       NORWAY        1949 2006 101949  -999.00
0102600  6970   1890  100 Tromsoe              NORWAY        1890 2006 341890  -999.00
0102800  7450   1900   16 Bjoernoeya           NORWAY        1920 2006 341920  -999.00
```

FM! fixwmos fixed coordinates as well! Here's the snippet:

```
      locfac = 10    ! init location factor assuming lat & lon are in degs*10
      do i=1,10000
        read(10,'(a86)',end=12)buffy
        read(buffy,fmt)wmo,lat,lon,alt,sname,ctry,sy,ey,flag,extref
        if (lat.gt.maxlat) maxlat = lat
        if (lat.gt.900) goto 12
        do j=sy,ey+norml
          read(10,'()')
        enddo
      enddo
   12 if (maxlat.gt.900) locfac = 1     ! lat & lon are in degs *100
```

So it was written with TMP in mind! Oh, for a memory. So we don't need to fret about TMP lat/lon being *10 any more!!

And it's taken me until NOW to realise that the IDL synthetic generators (vap_gts_anom, frs_gts_tdm, rd0_gts_anom) all need to calculate 1961-1990 normals! So they will need TMP, DTR and/or PRE binary normals for 1961 to 1990. Which means anomalies will have to be automatically generated for that period regardless of the requested period!!! *Cries*

Introduced suitable conditionals to ensure that 61-90 anomalies and gridded binaries are automatically produced if the relevant secondary parameters are requested.

More run-time issues, VAP is still giving an apparently non-fatal error:

```
% Program caused arithmetic error: Floating underflow
% Program caused arithmetic error: Floating overflow
% Program caused arithmetic error: Floating illegal operand
```

(this only appears once the vap_gts_anom.pro program has finished, so can't be identified)

Stuck on WET production, getting an error from rd0_gts_anom_05.pro, from the same bit of code that works fine in rd0_gts_anom.pro! The error:

```
% Attempt to subscript RD0SYN with NSEA is out of range.
% Execution halted at: RD0_GTS_ANOM_05    34 /cru-auto/cruts/version_3_0/BADC_AREA/programs/idl/rd0_gts_anom_05.pro
%                      $MAIN$
```

Line 34:

```
rd0syn=float(prenorm)*0.0 & rd0syn(nsea)=-9999
```

Do you know, I actually worked this one out by myself. Preen. It turned out that nsea was -1, which meant that it was finding no hits here:

```
nsea =where(rd0norm eq -9999 or  prenorm eq -9999)
```

When I looked - the min and max of rd0norm and prenorm were:

```
  -32768   32514
```

..and I thought, what a coincidence, that's 2^16. Aha! Must be an Endian problem. Looked it up on the web, abd the IDL ref manual, and found that adding:

```
,/swap_if_big_endian
```

..to the end of the openr statements in rdbin.pro, it all worked! :-)))

and then, of course, another problem I should have anticipated: half-degree gridding of synthetics needs half-degree primary binaries. So the precip binaries must be half-degree for WET (after 1989) and the usual 2.5-degrees earlier. More modifications to update.for!! And it took me a further 24 hours to cotton on that I'd need half-degree TMP and DTR binaries for FRS. VAP won't mind as it's using the synthetics as an adjunct to the observations - the exceptions are those secondaries where no observations can be used. WET after 1989, FRS, and CLD after 2002 (but CLD considerately works from DTR anomalies, ungridded).

So I'm going to have to produce half-degree gridded binary TMP and DTR anomalies, adding HALF AN HOUR to the run time. Bollocks. Though I could be clever and save it.. then I'd have to monitor when 1961-1990 databases were altered, and compare, and.. wibble.

Got that done. Then got it all working (though outputs not tested). Whooo. Now for BADC.

...

Actually, BADC wasn't too bad. Took a day or so to get everything to compile, mainly having to shift to gfortran rather than f77, and also to use -w to suppress warnings. Discovered that the IDL there didn't look at IDL_STARTUP, bah, but then found a way to specify a startup file on the command line, ie:

```
            call system('idl -e @runs/runs.'//dtstr//'/'//
     *       'idl.syn1.wet.txt.'//dtstr//'.dat -quiet')                        ! idl no startup
```

..so that's all right then. Got it all running without errors at BADC. Well, I say that, I'm still getting this for VAP:

```
gridding PRE anomalies at 0.5 for synthetics
Producing secondary: VAP
% Program caused arithmetic error: Floating overflow
% Program caused arithmetic error: Floating illegal operand
gridding VAP anomalies and synthetics
Producing secondary: WET
```

I haven't been able to identify what's causing that. Um.

Anyway the next items are the tricky saving of 2.5 and 0.5 binaries for 1961-1990, only regenerating then if the dbs have been altered. Requires multi-process cooperation, since we can't tell from the database timestamps which years were potentially changed. Admittedly, with this system that only accepts MCDW/CLIMAT/BOM updates, a pre-1991 change is all but impossible, but build for the case you can't anticipate..! Also up next is the deconstruction of the early cloud data (ie to 2002) so we can generate NetCDF files for the whole shebang. degroupcld.for will do the honours.

Did CLD first. Having reverse-engineered gabs ('gridded absolute') files for 1901-2002, I then modified update (extensively) to skip anything to do with CLD (including station counts) before 2003. Then, at the anoms-to-absolutes stage, unzipped and copied over any pre-2003 CLD gabs files from the reference repository.

I suppose I'll have to do CLD station counts (just 'n' obviously) at some stage, too.

Ran update, just for CLD, just for 1901-06/2006. Realised halfway through that I'd really have to do station counts as well because update does 'em for DTR anyway! That ought to cut out but doesn't at the moment.

It's getting faster.. implementing the 'saved binaries' was easier than I thought as well. Lots to change but straightforward. Now the IDL synthetics generators will always look in the reference area for 1961-1990 gridded binaries, whether 2.5-degree or 0.5-degree. And those datasets *should* be regenerated if flags are set that 1961-1990 data has been changed in the databases.

Then, a big problem. Lots of stars ('*********') in the PET gridded absolutes. Wrote sidebyside.m to display the five input parameters; VAP looks like being the culprit, with unfeasibly large values (up to 10034 in fact). And that's after the standard /10. So, erm.. a drains-up on VAP is now required. Oh, joy. And CLD also looks unacceptable, despite all that work - big patches of 100% and 0% dominate, no doubt a result of clipping by glo2absauto. The clipping is necessary, but shouldn't be needed so often!

Reassuringly, the 3_00 VAP and CLD that are published look fine, so it's soemthing I've done in the automation process. Mis-scaling is most likely.

Started chaining back through (initially) VAP. The gabs files were identical to the finals (now, if that had failed it would have been a problem!). The gridded anomaly files were a lot more interesting, because although they looked just as bad, their max values were exactly 9999. That ain't no coincidence!

Trailing fiurther back.. VAP anoms are OK, so suspicion falls on the synthetics. And lo and behold, re-running the TMP and DTR 2.5-grid binary productions with quick_interp_tdm2 gives:

```
IDL> quick_interp_tdm2,2006,2006,'interim_data/gbins/gbins.0903201540/tmp/tmp.',1200,gs=2.5,pts_prefix='interim_data/anoms/anoms.0903201540/tmp/tmp.',dumpbin='dumpbin',start
Defaults set
    2006
grid 2006 non-zero-8341.4424 8341.4424 3712.6726 cells=    74509
IDL> quick_interp_tdm2,2006,2006,'interim_data/gbins/gbins.0903201540/dtr/dtr.', 750,gs=2.5,pts_prefix='interim_data/anoms/anoms.0903201540/dtr/dtr.',dumpbin='dumpbin',start
Defaults set
    2006
grid 2006 non-zero-9116.9639 9116.9639 2825.0928 cells=    68171
IDL>
```

Those strings of numbers? They're supposed to be mean, average magnitude, and std dev! Should look something like this:

```
IDL> quick_interp_tdm2,2006,2006,'testdtrglo/dtr.',750,gs=0.5,pts_prefix='dtrtxt/dtr.',dumpglo='dumpglo',dumpbin='dumpbin'
Defaults set
    2006
grid 2006 non-zero    0.1125    2.1122    2.9219 cells=   202010
```

If I run with info='info', I get:

```
                        MEAN    AV MAG   STD DEV
    2006
data 2006 month  7    0.0981    1.1909    1.7665
data 2006 month  8    0.3129    1.3504    1.9677
data 2006 month  9    0.2413    1.2774    2.0954
data 2006 month 10   -0.0024    1.3375    2.0739
data 2006 month 11    0.2594    1.1632    2.0542
data 2006 month 12    0.0874    1.3236    2.2353
```

..confirming that the DTR (in this case) incoming anomalies are all within expected tolerances.

Ooh! Just found this a few thousand lines back, which may be relevant:

<QUOTE>
On a parallel track (this would really have been better as a blog), Tim O has found that the binary grids of primary vars (used in synthetic production of secondary parameters) should be produced with 'binfac' set to 10 for TMP and DTR. This may explain the poor performance and coverage of VAP in particular.
<END_QUOTE>

Did that help? Not much at all, unfortunately. This is frustrating, I can't see what's different. I even enabled a commented-out line that prints the ranges of pts2(*,2) and r, and they look OK:

```
IDL> quick_interp_tdm2,2006,2006,'interim_data/gbins/gbins.0903201540/dtr/dtr.', 750,gs=2.5,pts_prefix='interim_data/anoms/anoms.0903201540/dtr/dtr.',dumpbin='dumpbin',start
                        MEAN    AV MAG   STD DEV
    2006
data 2006 month  7    0.0981    1.1909    1.7665
     -10.4367      17.5867
     -7.15403      14.5088
data 2006 month  8    0.3129    1.3504    1.9677
     -7.00800      25.6929
     -4.89867      15.7781
data 2006 month  9    0.2413    1.2774    2.0954
     -18.4621      26.2400
     -15.1905      22.7162
data 2006 month 10   -0.0024    1.3375    2.0739
     -8.61333      18.5400
     -6.05684      15.7678
data 2006 month 11    0.2594    1.1632    2.0542
     -6.91852      33.3200
     -5.10848      28.5915
data 2006 month 12    0.0874    1.3236    2.2353
     -8.76667      24.4500
     -6.03609      21.9419
grid 2006 non-zero-9124.9951 9124.9951 2813.1790 cells=    68111
IDL>
```

OK, I *think* I've got it. It's the fact that we're writing a yearly binary file but only have data for the second half of that year:

```
minmax Jan-Jun:      -9999.00     -9999.00
minmax Jul-Dec:      -15.1905      28.5915
```

Now, I don't see how we get:

```
grid 2006 non-zero-9125.1289 9125.1289 2813.0332 cells=    68110
```

..but I do see how vap_gts_anom might just read the *first* six months, which would all be -9999.

So, we need to be able to write six-month binaries. Oh, my giddy aunt. What a crap crap system. We'll have to switch to monthly binaries, it's the only unambiguous way. Meaning major modifications to numerous IDL proglets. Fuck. Everything from the main progs (vap_gts_anom, quick_interp_tdm2, etc) to the supporting ones (rdbin for one).

After HOURS, I think I've sussed it, at least for VAP. The incoming *integer* binaries had to be constructed with binfac=10, because otherwise the integer bit renders most anomalies 0. Then, in the vap_gts_anom script, the values have to be divided by 100, to give degrees. any other combination of scaling factors throws the sat vap pressure calculations into the weeds. Of course, monthly binaries are still required. Ho hum.

Modified quick_interp_tdm2 to take another additional command-line option: dumpmobin, which if set will see that binaries are saved monthly, not yearly. Of course, the 2.5-degree TMP and DTR binary grids are only used by VAP - FRS uses 0.5-degree.

So, rather than carry on with mods, I thought I'd mod update enough to fix VAP, then run it all again.

Well, it ran. Until PET production, where it crashed with the same (understandable) read error as before ('********' not being an integer). However, when I invoked the Matlab sidebyside proglet to examine the VAP, it was much improved on the previous VAP. The max was still 10000, just a shade too high, but the actual spatial pollution was much reduced. There's hope! I think this all stems from the sensitivity of the saturated vapour pressure calculations, where a factor of 10 error in an input can make a factor of 1000 difference to the output.

Had to briefly divert to trick makegridsauto into thinking it was in the middle of a full 1901-2006 update, to get CLD NetCDF files produced for the whole period to June '06. Kept some important users in Bristol happy.

So, back to VAP. Tried dividing the incoming TMP 7 DTR binaries by 1000! Still no joy. Then had the bright idea of imposing a threshold on the 3.00 vap in the Matlab program. The result was that quite a lot of data was lost from 3.00, but what remained was a very good match for the 2.10 data (on which the thresholds were based).

I think I've got it! Hey - I might be home by 11. I got quick_interp_tdm2 to dump a min/max for the synthetic grids. Guess what? Our old friend 32767 is here again, otherwise known as big-endian trauma. And sure enough, the 0.5 and 2.5 binary normals (which I inherited, I've never produced them), both need to be opened for reading with:

```
  openr,lun,fname,/swap_if_big_endian
```

..so I added that as an argument to rdbin, and used it wherever rdbin is called to open these normals.

So, I went through all the IDL routines. I added an integer-to-float conversion on all binary reads, and generally spruced things up. Also went through the parameters one by one and fixed (hopefully) their scaling factors at each stage. What a minefield!

The PET problem, or unwriteable numbers, was solved by this tightening of secondaries, particularly VAP, and also putting in a clause to abs() any negative values from the wind climatology. I really don't think there should be any, but there are!

Finally I'm able to get a run of all ten parameters. The results, compared to 2.10 with sidebyside3col.m, are pretty good on the whole. Not really happy with FRS (range OK but mysterious banding in Southern Hemisphere), or PET:

```
pet
range210 = 0   573
range300 = 0   17.5000
```

So I've ended up with a range that doesn't scale simply to the 2.10 range. I also have no idea what the actual range ought to be. And they said PET would be easy. Next step has to be a comparison of max/min values of PET precursors vs. PET actuals for the two sources. Did that. No significant differences, except that of course the 2.10 PET was produced with uncorrected wind. When I took out the correction for 3.00, it shot up to even higher levels, so we'll just have to ignore 2.10 comparisons with PET.

Still, a top whack of 17.5 isn't too good for PET. Printed out the ranges of the precursors:

```
PET precursor parameters: ranges

tm    -49.40     39.20
tn    -52.80     39.50
tx    -45.10     59.80
vp      0.00     36.60
wn      0.00     29.00
cl      0.00      1.00
```

So the temps are in degs C, vapour pressure's in hPa, wind's in m/s and cloud's fractional.

Then I thought about it. 17.5mm/day is pretty good - especially as it looks to be Eastern Sahara.

As for FRS.. with those odd longitudinal stripes - I just tidied the IDL prog up and it, er..

..went away. How very comforting.

Did a complete run for 7/06 to 12/06, ran the Matlab visuals, all params looked OK (if not special).

FTP'd the program suite and reference tree to BADC, replacing the existing ones, and tried the same full run there.

Well the first thing I noticed was how slow it was! Ooops. Maybe 3x slower than uealogin1. Then, lots of error messages (see below). I had wondered whether the big endian scene was going to show, maybe this is it. Anyway, it finished! Here's the screen dump:

```
<QUOTE>
date25:   0903270742
date05:   0903270742
last6190: 0901010001
Producing anomalies
Producing station counts
Gridding primary parameters
Producing gridded binaries for synthetics
gridding TMP binary anomalies for secondary support
% Program caused arithmetic error: Floating illegal operand
gridding DTR binary anomalies for secondary support
% Program caused arithmetic error: Floating illegal operand
gridding TMP anomalies at 0.5 for synthetics
gridding DTR anomalies at 0.5 for synthetics
gridding PRE anomalies at 0.5 for synthetics
% Program caused arithmetic error: Floating illegal operand
Producing secondary: VAP
% Program caused arithmetic error: Floating divide by 0
% Program caused arithmetic error: Floating underflow
% Program caused arithmetic error: Floating overflow
% Program caused arithmetic error: Floating illegal operand
gridding VAP anomalies and synthetics
Producing secondary: WET
Producing secondary: CLD
Making synthetic CLD from DTR anomalies
gridding CLD anomalies and synthetics
Producing secondary: FRS
Converting anomalies to absolutes
Deriving PET
Creating output data and station files
creating final n-station tmpdtr files
creating final 0-station tmpdtr files

All work completed satisfactorarily
see: logs/completion/infolog.0904010108.dat
and: logs/logs.0904010108/update.0904010108.log
```

```
-bash-3.00$
<END_QUOTE>
```

Pulled back the output files and ran the sidebyside3col Matlab script to compare
with ours. Interesting. Here are the ranges:

```
tmp: BADC 300 m/m: -49.4         39.2, CRU 300 m/m: -49.4         39.2
tmn: BADC 300 m/m: -52.8         39.5, CRU 300 m/m: -52.8         39.5
tmx: BADC 300 m/m: -45.1         59.8, CRU 300 m/m: -45.1         59.8
dtr: BADC 300 m/m:   1           39.2, CRU 300 m/m:   1           39.2
pre: BADC 300 m/m:   0         4573,   CRU 300 m/m:   0         4573
vap: BADC 300 m/m:   0           47.9, CRU 300 m/m:   0           36.3
wet: BADC 300 m/m:   0          310,   CRU 300 m/m:   0          309.5
cld: BADC 300 m/m:   0           99.9, CRU 300 m/m:   0          100
frs: BADC 300 m/m:   0          310, CRU 300 m/m:     0          310
pet: BADC 300 m/m:   0           17.1, CRU 300 m/m:   0           17.5
```

I don't know which is more worrying - the VAP discrepancy or the fact that the
minimum DTR is 1 degree (for both!), the maximum BADC CLD is 99.9%, and the
maximum CRU WET is 30.95 days! Well I guess the VAP issue is the show-stopper, and
must be related to those errors:

```
Producing secondary: VAP
% Program caused arithmetic error: Floating divide by 0
% Program caused arithmetic error: Floating underflow
% Program caused arithmetic error: Floating overflow
% Program caused arithmetic error: Floating illegal operand
```

Now, these are IDL errors, and probably from our old pal vap_gts_anom_m.pro. So,
the established procedure is to re-run just that program, with all the info
turned on. Oh, my:

```
IDL> !path = 'programs/idl:' + !path
IDL> vap_gts_anom_m,2006,2006,dtr_prefix='interim_data/gbins/gbins.0904010108/dtr/dtr.',tmp_prefix='interim_data/gbins/gbins.0904010108/tmp/tmp.',outprefix='interim_data/syn
% Compiled module: VAP_GTS_ANOM_M.
Land,sea:       56016       68400
Calculating tmn normal
Calculating synthetic vap normal
Calculating synthetic anomalies
200607 vap (x,s2,<<,>>):         -Inf         -NaN         -Inf      12.1306
200608 vap (x,s2,<<,>>):         -Inf         -NaN         -Inf      15.4191
200609 vap (x,s2,<<,>>):         -Inf         -NaN         -Inf      23.2317
200610 vap (x,s2,<<,>>):         -Inf         -NaN         -Inf      22.4792
200611 vap (x,s2,<<,>>):         -Inf         -NaN         -Inf      15.7444
200612 vap (x,s2,<<,>>):         -Inf         -NaN         -Inf      11.2271
% Program caused arithmetic error: Floating divide by 0
% Program caused arithmetic error: Floating underflow
% Program caused arithmetic error: Floating overflow
% Program caused arithmetic error: Floating illegal operand
IDL>
```

Yes, it's back. Right back where we started with VAP at CRU, all those, er, days ago.
Well last time it was big endian stuff, wasn't it? And presumably the little Linux
box at BADC is big endian. So I might try changing those rdbin calls, just to see..

..that didn't seem to help. Here's a dump of key array ranges, just before the main
loop kicks in:

```
norgrd min/max:      -68.9000      36.6000
tadj min/max:      -715.433      445.602
tmpgrd min/max:      -3165.30      3153.60
dtrgrd min/max:      -1868.90      1612.80
vapsyn min/max:    0.01000000          Inf
v min/max:       0.00000          Inf
```

So tmpgrd and dtrgrd look waaay too high, though could just be *100. v and vapsyn are shot.

This does look like scaling. Boo hoo. I *fixed* that!! These are the ranges on UEALOGIN1:

```
norgrd min/max:      -68.9000      36.6000
tadj min/max:      -46.7671      25.8468
tmpgrd min/max:      -68.9000      37.6000
dtrgrd min/max:      -7.40000      6.30000
vapsyn min/max:     0.0100000      33.4119
v min/max:    0.00521439      41.9604
```

I wonder if I need to reverse the rdbin logic? reverse_if_little_endian on the non-clim calls?
Since normals are reading OK without it (CRU version has bigend=1, BADC version doesn't). Let's
try with just the tmpgrd & dtrgrd reads.. ooh:

```
norgrd min/max:      -68.9000      36.6000
tadj min/max:      -715.433      445.602
tmpgrd min/max:      -68.9000      37.6000
dtrgrd min/max:      -7.40000      6.30000
vapsyn min/max:    0.01000000      59.3142
v min/max:    0.00521439      61.0593
200607 vap (x,s2,<<,>>):     0.614392      1.57785     -5.33517      10.2892
200608 vap (x,s2,<<,>>):     0.593988      1.69958     -3.55033      12.3374
200609 vap (x,s2,<<,>>):     0.448958     0.793516     -5.20586      10.4787
200610 vap (x,s2,<<,>>):     0.525755      1.15223     -2.83614      8.92979
200611 vap (x,s2,<<,>>):     0.243011     0.939122     -4.66185      21.2776
200612 vap (x,s2,<<,>>):     0.302154     0.628504     -5.05943      5.84549
% Program caused arithmetic error: Floating underflow
% Program caused arithmetic error: Floating overflow
IDL>
```

So, just tadj to 'fix', then? Though surely I should read the 2006 tmp & dtr the same way.
Or is it that I copied the 61-90 over from here, but generated the 2006 there. Ah. Should
probably regenerate the 61-90 binaries at BADC? Yes. Anyway, found the 'other' tmp/dtr
reads and adjusted those, and behold:

```
norgrd min/max:      -68.9000      36.6000
tadj min/max:      -46.7671      25.8468
tmpgrd min/max:      -68.9000      37.6000
dtrgrd min/max:      -7.40000      6.30000
vapsyn min/max:    0.01000000      33.4119
v min/max:    0.00521439      41.9604
200607 vap (x,s2,<<,>>):     0.493936      1.15553     -6.14130      5.82037
200608 vap (x,s2,<<,>>):     0.460588     0.994776     -4.42765      5.69217
200609 vap (x,s2,<<,>>):     0.381708     0.674991     -4.56722      8.27954
200610 vap (x,s2,<<,>>):     0.506931     0.895855     -4.37382      5.13692
200611 vap (x,s2,<<,>>):     0.263565     0.656560     -3.84704      11.5171
200612 vap (x,s2,<<,>>):     0.374605     0.791715     -3.15873      6.41023
```

The CRU version has the same ranges, but some month stats differ:

```
norgrd min/max:      -68.9000      36.6000
tadj min/max:      -46.7671      25.8468
tmpgrd min/max:      -68.9000      37.6000
dtrgrd min/max:      -7.40000      6.30000
```

```
vapsyn min/max:     0.0100000      33.4119
v min/max:    0.00521439      41.9604
200607 vap (x,s2,<<,>>):     0.462426      1.11003     -6.14130      4.69486
200608 vap (x,s2,<<,>>):     0.428764     0.999707     -4.42765      6.48673
200609 vap (x,s2,<<,>>):     0.342277     0.642287     -4.56722      8.27954
200610 vap (x,s2,<<,>>):     0.485457     0.858445     -4.37382      5.13692
200611 vap (x,s2,<<,>>):     0.276767     0.685921     -3.72504      11.5171
200612 vap (x,s2,<<,>>):     0.373327     0.862642     -3.15873      15.3975

December in particular has quite a drift! No idea why, since the data going in
should be the same.

So, another full run, with regeneration of binary reference grids enforced:

tmp: BADC 300 m/m:    -49.4     39.2, CRU 300 m/m:    -49.4         39.2
tmn: BADC 300 m/m:    -52.8     39.5, CRU 300 m/m:    -52.8         39.5
tmx: BADC 300 m/m:    -45.1     59.8, CRU 300 m/m:    -45.1         59.8
dtr: BADC 300 m/m:      1       39.2, CRU 300 m/m:      1           39.2
pre: BADC 300 m/m:      0     4573,   CRU 300 m/m:      0         4573
vap: BADC 300 m/m:      0       36.5, CRU 300 m/m:      0           36.3
wet: BADC 300 m/m:      0      309.3, CRU 300 m/m:      0          309.5
cld: BADC 300 m/m:      0       99.9, CRU 300 m/m:      0          100
frs: BADC 300 m/m:      0      310,   CRU 300 m/m:      0          310
pet: BADC 300 m/m:      0       17.5, CRU 300 m/m:      0           17.5

I honestly don't think it'll get closer. So, I guess I'll clear out and reset
the BADC process, and let Kevin loose on it.

Well, BADC have had it for a good while, without actually doing anything. what
a surprise. It's lucky actually, as I've ironed out a few bugs (including PET
being garbage). One bug is eluding me, however - I can't get a full 1901-2008
run to complete! It gets stuck after producing the final TMP files (data plus
stations), it just seems to sit there indefinitely. So I tried different periods.

1901-2008 failed
1901 only worked
2008 only worked
1901-1910 worked
1901-1950 worked
1951-2008 worked
1901-2008 failed

**sigh** WHAT THE HELL'S GOING ON?! Well, time to ask the compiler. So I recompiled
as follows:

g77 -o update -Wall -Wsurprising -fbounds-check programs/fortran/update.for

Then, I re-ran. This time I got an error almost immediately:

Producing anomalies
Subscript out of range on file line 1011, procedure programs/fortran/update.for/MAIN.
Attempt to access the 6-th element of variable dobin25[subscript-1-of-2].
Abort (core dumped)

Hurrah! In a way.. thyat bug was easy enough, I'd just forgotten to put an extra
test (ipar.le.5) in the test for binary production, so as it was in a 1..8 loop,
there was bound (ho ho) to be trouble. There was a second, identical, instance.

After all that - final success:

date25:   0903270742
date05:   0903270742
last6190: 0901010001
Producing anomalies
Producing station counts
Gridding primary parameters
Producing gridded binaries for synthetics
gridding TMP binary anomalies for secondary support
gridding DTR binary anomalies for secondary support
gridding PRE binary anomalies for secondary support
gridding TMP anomalies at 0.5 for synthetics
gridding DTR anomalies at 0.5 for synthetics
gridding PRE anomalies at 0.5 for synthetics
Producing secondary: VAP
gridding VAP anomalies and synthetics
Producing secondary: WET
gridding WET anomalies and synthetics
Producing secondary: CLD
Making synthetic CLD from DTR anomalies
gridding CLD anomalies and synthetics
Producing secondary: FRS
Converting anomalies to absolutes
Deriving PET
Creating output data and station files
creating final n-station tmpdtr files
creating final 0-station tmpdtr files

All work completed satisfactorarily
see: logs/completion/infolog.0905070939.dat
and: logs/logs.0905070939/update.0905070939.log


uealogin1[/esdata/cru/f098/update_top]

..and in terms of disk usage (um, remember it's not *that* reliable):

uealogin1[/esdata/cru/f098/update_top] du -ks *
64      anomauto
32      batchdel
64      bom2cruauto
64      climat2cruauto
32      compile_all
629856  db
32      dtr2cldauto
64      glo2absauto
16108896        gridded_finals
13822176        interim_data
18368   logs
416     makegridsauto
64      makepetauto
64      mcdw2cruauto
32      movenormsauto
32      newdata.latest.date
288     newmergedbauto
2368    programs
1101088 reference
2848    results
3008    runs
32      saved_timings_090420_1716
64      stncountsauto
```

```
64      stncountsauto_safe
704     timings
32      tmnx2dtrauto
32      tmpdtrstnsauto
352     update
638432  updates
uealogin1[/esdata/cru/f098/update_top]
```

Meaning that a complete 1901-2008 run will need about 14gb of working data and the resulting files will need approximately 16gb. All gzipped!!

Then, of course (or 'at last', depending on your perspective), Tim O had a look at the data with that analytical brain thingy he's got. Oooops. Lots of wild values, even for TMP and PRE - and that's compared to the previous output!! Yes, this is comparing the automated 1901-2008 files with the 1901-June2006 files, not with CRu TS 2.1. So, you guessed it, bonnet up again.

First investigation was WET, where variance was far too low - usually indicative of a scaling issue, and thus it was. Despite having had a drains-up on scaling, WET seems to have escaped completely. The initial gridding (to binary) outputs at x10, which is absolutely fine. But the PRE-to-WET converters are not so simple. The 2.5-degree converter (rd0_gts_anom_m.pro) has reasonable output values (five monthly examples):

```
minmax rd0 2.5 binaries:      -100.000      357.327
minmax rd0 2.5 binaries:      -94.2232      250.621
minmax rd0 2.5 binaries:      -93.0808      512.557
minmax rd0 2.5 binaries:      -100.000      623.526
minmax rd0 2.5 binaries:      -95.1105      521.668
```

The trouble is, when written to binary, these will be rounded to integer and a degree of accuracy will be lost. They should be x10. Then there's the 0.5-degree converter (rd0_gts_anom_05_m.pro), which has indescribably awful output values:

```
minmax rd0 0.5 binaries:      -1.00000      8.33519
minmax rd0 0.5 binaries:     -0.970328      8.13772
minmax rd0 0.5 binaries:     -0.951749      4.33032
minmax rd0 0.5 binaries:      -1.00000      9.26219
minmax rd0 0.5 binaries:     -0.960226      3.80590
```

These are basically 1000 times too small!!! How did this happen when I specifically had a complete review of scaling factors?! FFS.

Aha. Not so silly. The 0.5 grids are saved as .glo files (because after 1989 it's all synthetic). So they're not rounded. On the other hand, they are still 100x too low for percentage anomalies. and the 2.5 grids are sent to the gridder as 'synthfac=100'!! When currently it's 1! So.. some changes to make :P

The 2.5-degree PRE/WET path is now at x10 all the way to the final gridding. The 0.5-degree PRE/WET path is at x10 until the production of the synthetic WET, at which point it has to be x1 to line up with the pre-1990 output from the gridder (the gridder outputs .glo files as x1 only, we haven't used the 'actfac' parameter yet and we're not going to start!!).

Got all that fixed. Then onto the excessions Tim found - quite a lot that really should have triggered the 3/4 sd cutoff in anomauto.for. Wrote 'retrace.for', a proglet I've been looking for an excuse to write. It takes a country or individual cell, along with dates and a run ID, and preforms a reverse trace from final output files to database. It's not complete yet but it already gives extremely helpful information - I was able to look at the first problem (Guatemala in Autumn 1995 has a massive spike) and find that a station in Mexico has a temperature of 78 degrees in November 1995! This gave a local anomaly of 53.23 (which would have been 'lost' amongst the rest of Mexico as Tim just did country averages) and an anomaly in Guatemala of 24.08 (which gave us the spike):

```
7674100  1808  -9425   22 COATZACOALCOS, VER.  MEXICO        1951 2009  101951 -999.00

1994  188-9999  244-9999-9999  286  281  275  274  274  262-9999
1995  237-9999-9999-9999  300-9999  281  283  272-9999  780  239
1996  219  232  235  256  285  276  280  226  285  260  247  235
```

Now, this is a clear indication that the standard deviation limits are not being applied. Which is extremely bad news. So I had a drains-up on anomauto.for.. and.. yup, my awful programming strikes again. Because I copied the anomdtb.f90 process, I failed to notice an extra section where the limit was applied to the whole station - I was only applying it to the normals period (1961-90)! So I fixed that and re-ran. Here are the before and after outputs from trace.for:

```
Trace on Mexico, 11/1995, run #0905070939:
crua6[/cru/cruts/version_3_0/fixing_tmp_and_pre] cat retrace.Mexico.tmp.1995.11.stat
    1995      11   21.41    9.80     239     145   72.80     216     173
    1995      11   21.41    9.80     239     145   72.80     216     173
    1995      11   18.33    8.70     239     145   23.80     216     173
    1995      11    3.08    1.13     239     145   49.04     216     173
    1995      11    3.23   -0.57     238     148    2232   53.23     217     172    2244
    1995      11   22.39   12.80     243     148 7227000   78.00     217     172 7674100

Trace on Mexico, 11/1995, run #0907031504:
crua6[/cru/cruts/version_3_0/fixing_tmp_and_pre] cat retrace.Mexico.tmp.1995.11.stat
    1995      11   19.51    9.80     239     145   28.90     216     156
    1995      11   19.51    9.80     239     145   28.90     216     156
    1995      11   18.33    8.70     239     145   28.50     216     156
    1995      11    1.18    1.13     239     145    0.36     216     156
    1995      11    0.73   -0.57     238     148    2227    1.82     227     148    2231
    1995      11   22.39   12.80     243     148 7227000   78.00     217     172 7674100

The column to be looking at is this one ---------------------^

Because it's a traceability program, it works backwards in time:
Row 1  Final gridded output
Row 2  Gridded absolutes (should == Row 1)
Row 3  Climatology
Row 4  Gridded anomalies
Row 5  Anomalies
Row 6  Actual station values

Columns are as follows:
Cols 1,2  Year, Month
Col 3     Mean
Cols 4-6  Min with cell indices
Cols 7-9  Max with cell indices

Row 5:
Cols 4-7  Min with cell indices and line # in anoms file
Cols 8-11 Max with cell indices and line # in anoms file

Row 6:
Cols 4-7  Min with cell indices and WMO code in database
Cols 8-11 Max with cell indices and WMO code in database

In this case, the erroneous value of 78 degrees has been counted in the earlier run,
giving an anomaly of 53.23. In the later run, it hasn't - the anomaly of 1.82 is from a
```

different cell (227,148 instead of 217,172).

So, re-running improved matters. The extremes have vanished. But the means are still out, sometimes significantly.

Stand by, we're about to go down the rabbit-hole again!

I took the twelve 1990 anomaly files from the original 1901-2006 run (that was done with some flavour of anomdtb.f90). They were here:

/cru/cruts/version_3_0/primaries/tmp/tmptxt/*1990*

Then I modified the update 'latest databases' file to say that tmp.0705101334.dtb was the current database, and made a limited run of the update program for tmp only, killing it once it had produced the anomaly files. The run was #0908181048.

So, under /cru/cruts/version_3_0/fixing_tmp_and_pre/custom_anom_comparisons, we have a 'manual' directory and an 'automatic' directory, each with twelve 1990 anomaly files. And how do they compare? NOT AT ALL!!!!!!!!!

Example from January:

```
crua6[/cru/cruts/version_3_0/fixing_tmp_and_pre/custom_anom_comparisons] head manual/tmp.1990.01.txt
   70.90   -8.70    10.0      4.20000  10010
   78.30   15.50    28.0      9.10000  10080
   69.70   18.90    10.0      0.90000  10250
   69.70   18.90   100.0      1.30000  10260
   74.50   19.00    16.0      5.40000  10280
   69.50   25.50   129.0      0.30000  10650
   70.40   31.10    14.0     -0.40000  10980
   66.00    2.00     0.0      1.70000  11000
   67.30   14.40    13.0      0.80000  11520
   66.80   14.00    39.0      2.50000  11530
crua6[/cru/cruts/version_3_0/fixing_tmp_and_pre/custom_anom_comparisons] head automatic/tmp.1990.01.txt
    7.09   -0.87    10.0      2.97558  10010
    7.83    1.55    28.0      7.87895  10080
    6.97    1.89    10.0      0.50690  10250
    6.97    1.89   100.0      0.49478  10260
    7.45    1.90    16.0      3.88554  10280
    6.95    2.55   129.0     -1.48960  10650
    7.04    3.11    14.0     -0.46391  10980
    6.60    0.20     0.0      1.63333  11000
    6.73    1.44    13.0      0.12662  11520
    6.68    1.40    39.0      2.03333  11530
```

The numbers of values in each pair are not always identical, but are within 2 or 3 so that's not too worrying. Worrying, yes, just not fatal.

There are a number of things going on. The lats and lons are the same, just scaled (because originally the TMP coordinates were x10 not x100). We can ignore that problem. The real problem is the completely different results from the automated system - I don't understand this because I painstakingly chcked the anomauto.for file to ensure it was doing the right job!! The overall pattern of anomalies is roughly the same - it's just that the actual values differ. Not always lower - sometime higher. Could be a rounding error..

Got anomauto to dump the first month of the first station (10010). The clue to the problem is in the first lines - we're only getting the full-length mean (used for SD calculations) and not the 61-90 mean. nv should be <= 30.

```
WMO =    10010, im = 01
(nv.gt.5) sums =  -384.90, nv =       86, ave =    -4.48
onestn(0490,01) =    -1.50
d(      1,0490,01) =     2.98
```

Aaaaand.. FOUND IT! What happened was this: in the original anomdtb.f90 program, there's a test for existing normals (in the header of each station). If they are present, then SD is calculated (to allow excessions to be screened out). If not, SD is calculated and then used to screen excessions, then a 61-90 normal is built provided there are enough values after the screening. However, in my version, I followed the same process - but crucially, I wasn't using the same variable to store the existing normals and the calculated ones!! So we were ending up with the 'full length' normal (n=86 in the above example) instead. All I had to add was a single line:

```
            ave = xstnrms(im)*fac  ! NEW - actually *use* existing normals!!
```

We then get:

```
onestn(0490,01) =    -1.50
d(      1,0490,01) =     4.20
```

Which is what we want. So, a complete re-run (just tmp) for 1990, still using the old db.

Tadaa:

```
uealogin1[/cru/cruts/version_3_0/fixing_tmp_and_pre/custom_anom_comparisons/new_automatic] head tmp.1990.01.txt
    7.09   -0.87    10.0      4.20000  10010
    7.83    1.55    28.0      9.10000  10080
    6.97    1.89    10.0      0.90000  10250
    6.97    1.89   100.0      1.30000  10260
    7.45    1.90    16.0      5.40000  10280
    6.95    2.55   129.0      0.30000  10650
    7.04    3.11    14.0     -0.40000  10980
    6.60    0.20     0.0      1.70000  11000
    6.73    1.44    13.0      0.80000  11520
    6.68    1.40    39.0      2.50000  11530
```

Spot on! A 100% match with the previously-generated anomalies. Phew!

Of course, now we have to do a proper run with the latest db..

..that very nearly worked :(

Mostly the same, but one noticeable exception is the hot 2003 JJA in Europe - it's much less extreme in the automated version. So I ran with the original database again. Thought I'd see if there were different station counts:

For the original anomdtb and original June 2006 db, tmp.0705101334.dtb):

```
    1259 tmp.2003.06.txt
    1216 tmp.2003.07.txt
    1223 tmp.2003.08.txt
```

For 0909041051 (fixed anomauto and original June 2006 db, tmp.0705101334.dtb):

```
    1210 tmp.2003.06.txt (-49)
    1201 tmp.2003.07.txt (-15)
    1178 tmp.2003.08.txt (-45)
```

For 0909021348 (the 'fixed' anomauto and the latest db, tmp.0904151410.dtb):

```
    1246 tmp.2003.06.txt (-13)
    1250 tmp.2003.07.txt (+34)
    1228 tmp.2003.08.txt (+ 5)
```

```
Ran retrace (because I might as well use it!).

For 0909041051, original db:
    2003       6   16.13    6.90     273     375   20.60     266     380
    2003       6   16.13    6.90     273     375   20.60     266     380
    2003       6   16.13    6.90     273     375   20.60     266     380
    2003       6    0.00    0.00     273     375    0.00     266     380
    2003       6 -999.00 -999.00    -999    -999    -999 -999.00    -999    -999    -999
    2003       6 -999.00 -999.00    -999    -999    -999 -999.00    -999    -999    -999
crua6[/cru/cruts/version_3_0/fixing_tmp_and_pre] cat retrace.France.0909041051.tmp.2003.07.stat
    2003       7   18.58    9.40     273     375   23.80     265     380
    2003       7   18.58    9.40     273     375   23.80     265     380
    2003       7   18.58    9.40     273     375   23.80     265     380
    2003       7    0.00    0.00     273     375    0.00     265     380
    2003       7 -999.00 -999.00    -999    -999    -999 -999.00    -999    -999    -999
    2003       7 -999.00 -999.00    -999    -999    -999 -999.00    -999    -999    -999
crua6[/cru/cruts/version_3_0/fixing_tmp_and_pre] cat retrace.France.0909041051.tmp.2003.08.stat
    2003       8   18.25    8.90     273     375   23.80     265     380
    2003       8   18.25    8.90     273     375   23.80     265     380
    2003       8   18.25    8.90     273     375   23.80     265     380
    2003       8    0.00    0.00     273     375    0.00     265     380
    2003       8 -999.00 -999.00    -999    -999    -999 -999.00    -999    -999    -999
    2003       8 -999.00 -999.00    -999    -999    -999 -999.00    -999    -999    -999

For 0909021348, latest db:
    2003       6   19.12    8.50     273     375   23.60     267     367
    2003       6   19.12    8.50     273     375   23.60     267     367
    2003       6   16.13    6.90     273     375   20.10     267     367
    2003       6    2.99    1.63     273     375    3.45     267     367
    2003       6    3.40    1.90     277     352      66    5.10     270     360      68
    2003       6   20.94   10.60     272     375 6717000   26.10     267     371 7650000
    2003       7   20.61   12.20     273     375   26.70     266     380
    2003       7   20.61   12.20     273     375   26.70     266     380
    2003       7   18.58    9.40     273     375   23.80     266     380
    2003       7    2.03    2.80     273     375    2.87     266     380
    2003       7    2.18    1.10     275     358      69    3.20     273     373      65
    2003       7   20.70    9.20     272     375 6717000   26.40     266     379 7790000
    2003       8   21.04   11.90     273     375   26.60     266     380
    2003       8   21.04   11.90     273     375   26.60     266     380
    2003       8   18.25    8.90     273     375   23.80     266     380
    2003       8    2.79    2.97     273     375    2.83     266     380
    2003       8    2.90    2.80     277     352      62    3.00     281     369      61
    2003       8   23.15   11.80     272     375 6717000   28.20     266     379 7790000

Well the differences certainly show up! And it looks like a database change. So.. I
guess I need to look at changes in French stations. Argh. And that 'argh' was prescient,
since, when I ran getcountry to extract the French stations from each database, I found:

crua6[/cru/cruts/version_3_0/fixing_tmp_and_pre] ./getcountry
Enter the database to search: ../update_top/db/tmp/tmp.0705101334.dtb

Enter the country name to extract: FRANCE
   33 stations written to ../update_top/db/tmp/tmp.0705101334.dtb.FRANCE

crua6[/cru/cruts/version_3_0/fixing_tmp_and_pre] ./getcountry
Enter the database to search: ../update_top/db/tmp/tmp.0904151410.dtb

Enter the country name to extract: FRANCE
  104 stations written to ../update_top/db/tmp/tmp.0904151410.dtb.FRANCE

Somehow, I've added 71 new French stations?! Surely I'd remember that. Especially
as they'd have had to have arrived with the MCDW/CLIMAT bulletins. Sizes:

crua6[/cru/cruts/version_3_0/fixing_tmp_and_pre] wc -l *FRANCE
      2725 tmp.0705101334.dtb.FRANCE
      3700 tmp.0904151410.dtb.FRANCE

That's not so bad. Well the ratio's improved. Could be a lot of unmatched incoming
stations?

Oh, ****. It's the bloody WMO codes again. **** these bloody non-standard, ambiguous,
illogical systems. Amateur hour again.

First example, the beautiful city of Lille. Here are the appropriate headers:

tmp.0705101334.dtb.FRANCE:
  70150   506     31   47 LILLE                FRANCE        1851 2006 101851  -999.00

tmp.0904151410.dtb.FRANCE:
 701500  5034    306   52 LILLE                FRANCE        2000 2009    -999       0
7015000  5034    306   52 LILLE                FRANCE        1851 2008  101851 -999.00

So.. just what I was secretly hoping for (not!) - a drains-up on the CLIMAT and MCDW
programs, otherwise known as climat2cruauto.for and mcdw2cruauto.for, as well as the
merging program, mergedbauto.for.

Some random, manual traceability that may be useful, or not:

Considering the MCDW update run numbered 0904151410:

cat /cru/cruts/version_3_0/update_top/runs/runs.0904151410/merg.mcdw.0904151410.dat
db/cld/cld.0904021239.dtb
updates/MCDW/db/db.0904151410/mcdw.cld.0904151410.dtb
updates/MCDW/db/db.0904151410/int1.cld.0904151410.dtb
blind
U

..is for cld, but indicates that the input database was tmp.0904021239.dtb.

The MCDW database was mcdw.tmp.0904151410.dtb.

The output database was, of course, tmp.0904151410.dtb.

I wonder what they all have for LILLE?

tmp.0904021239.dtb:
0701500  5057    310   52 LILLE/LESQUIN        FRANCE        2000 2009    -999       0
7015000  5034    306   52 LILLE                FRANCE        1851 2008  101851 -999.00

mcdw.tmp.0904151410.dtb:
0701500  5034    306   52 LILLE                FRANCE        2009 2009    -999       0

tmp.┐.dtb:
 701500  5034    306   52 LILLE                FRANCE        2000 2009    -999       0
7015000  5034    306   52 LILLE                FRANCE        1851 2008  101851 -999.00

I'll bet this just updated the 'false' LILLE with another month or something. In fact:

conv.mcdw.0904151410.dat:
       1    2009       1    2009
```

```
Before (tmp.0904021239.dtb):
0701500  5057    310   52 LILLE/LESQUIN        FRANCE        2000 2009    -999       0
6190-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
2000   42   63   76   99  147  166  159  182  164  116   78   57
2001   37-9999-9999-9999  144  158  187  193  135  145   69   30
2002   45   77   80  103  133  167  175  186  148  109   87   51
2003   30   29   87  106  137  186  195  208  158   85   81   45
2004   39   53   65  106  127  166  176  192  165  120   71   34
2005   54   30   71  105  129  179  182  170  165  144   65   37
2006   20   28   49   93  137  170  226  166  185  142   90   57
2007   74   67   79  136  142  168  173  173  145  107   69   40
2008   65   59   67   93  166  159  182  176  140  102   72   30
2009    7   37-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999

After (tmp.0904151410.dtb):
 701500  5034    306   52 LILLE                FRANCE        2000 2009    -999       0
6190-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999
2000   42   63   76   99  147  166  159  182  164  116   78   57
2001   37-9999-9999-9999  144  158  187  193  135  145   69   30
2002   45   77   80  103  133  167  175  186  148  109   87   51
2003   30   29   87  106  137  186  195  208  158   85   81   45
2004   39   53   65  106  127  166  176  192  165  120   71   34
2005   54   30   71  105  129  179  182  170  165  144   65   37
2006   20   28   49   93  137  170  226  166  185  142   90   57
2007   74   67   79  136  142  168  173  173  145  107   69   40
2008   65   59   67   93  166  159  182  176  140  102   72   30
2009    7   37-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999

The data is i-dentical. But ooh, lookit the WMO codes! Oh, Lordy! The update has changed
the code from 7-digit to 6-digit!

Let's look at all available headers for this LILLE station (ie the short modern one).

tmp.0904021106.dtb:0701500  5057    310   52 LILLE/LESQUIN        FRANCE        2000 2009    -999       0
tmp.0904021239.dtb:0701500  5057    310   52 LILLE/LESQUIN        FRANCE        2000 2009    -999       0
tmp.0904151410.dtb: 701500  5034    306   52 LILLE                FRANCE        2000 2009    -999       0

Aha - it's not the same station! ;0) An input has updated the details, istr that can happen.

Let's look at the MCDW update for Jan 2009, ssm0901.fin:
07015 LILLE                       5034N 00306E   52    1007.8 1013.7      .7        5.8        7   63         75 145

And a previous update, ssm0812.fin:
07015 LILLE                       5034N 00306E   52    1012.7 1018.6     3.0        6.9        7   32      1  65 140

Well that's no help. Why did it change? CLIMAT? But CLIMAT doesn't have station names!
Let's try and work out how the updates happened:

MCDW is first:
drwx------   2 f098     cru         4096 Mar  5  2009 db.0903051342
drwx------   2 f098     cru         4096 Mar  5  2009 db.0903051442
drwx------   2 f098     cru         4096 Mar  5  2009 db.0903051448
drwx------   2 f098     cru         4096 Mar  9  2009 db.0903091631
drwx------   2 f098     cru         4096 Apr  2 11:25 db.0904021106
drwx------   2 f098     cru         4096 Apr  2 12:57 db.0904021239
drwx------   2 f098     cru         4096 Apr 15 14:16 db.0904151410

Then CLIMAT:
drwx------   2 f098     cru         4096 Mar  5  2009 db.0903051342
drwx------   2 f098     cru         4096 Mar  5  2009 db.0903051448
drwx------   2 f098     cru         4096 Mar  9  2009 db.0903091631
drwx------   2 f098     cru         4096 Apr  2 11:26 db.0904021106
drwx------   2 f098     cru         4096 Apr  2 12:59 db.0904021239

We won't bother with BOM, I don't think they stretch to Lille..

I'm hoping I didn't do any on the escluster! *checks*.. nope. Just copied.

So this looks like the sequence:

0903051342  MCDW  CLIMAT
0903051442  MCDW only
0903051448  MCDW  CLIMAT
0903091631  MCDW  CLIMAT
0904021106  MCDW  CLIMAT
0904021239  MCDW  CLIMAT
0904151410  MCDW only

Interestingly, we only seem to have the last three tmp databases, at least in terms of
having the short LILLE station.

This is so hard because I cannot remember the process. Have to dig some more..

OK, I think the key update is 0903091631. Here's the CLIMAT runfile:
conv.climat.0903091631.dat:
       1    2000      12    2008

MCDW was even further-ranging: conv.mcdw.0903091631.dat:
       9    1994      11    2008

Still not there. One issue is that for some reason I didn't give the merg runfiles
individual names for each parameter! So I might mod the update program to do that.
Then re-run all updates.

The problem with re-running all updates, of course, is that I also fixed WMO codes. And,
(though my memory is extremely flaky), probably corrected some extreme values detected
by Tim. Oh bugger.

Well, WMO code fixing is identifiable because you get a log file, ie, here's the tmp dir:

-r--------   1 f098     cru      25936385 Aug 18 10:48 tmp.0705101334.dtb
-r--r--r--   1 f098     cru       7278548 Feb 17  2009 tmp.0705101334.dtb.gz
-rw-r--r--   1 f098     cru      25936385 Mar  8  2009 tmp.0903081416.dtb
-rw-r--r--   1 f098     cru           408 Mar  8  2009 tmp.0903081416.log
-rw-r--r--   1 f098     cru      27131064 Apr  2 11:15 tmp.0904021106.dtb
-rw-r--r--   1 f098     cru      27131064 Apr  2 12:48 tmp.0904021239.dtb
-rw-r--r--   1 f098     cru      27151999 Apr 15 14:11 tmp.0904151410.dtb

tmp.0705101334.dtb had its WMO codes fixed and became tmp.0903081416.dtb, which has the
accompanying log file (tmp.0903081416.log) to prove it:

Program fixwmos was run:  0903081416
Reference WMO List:       /cru/cruts/version_3_0/WMO/from_dave_lister_wmo_list
Country Codes (unused):   /cru/cruts/version_3_0/WMO/wmo_country_codes.dat
Input database:           tmp.0705101334.dtb
Total stations:            5065
WMOs matched:              2263
False WMOs (<0):              0
Bad WMOs (>=0 & <1000):       0
```

```
Output Database:         tmp.0903081416.dtb

Unfortunately, only tmp and pre have such log files. Here's the one for pre:

Program fixwmos was run:  0903051740
Reference WMO List:       /cru/cruts/version_3_0/WMO/from_dave_lister_wmo_list
Country Codes (unused):   /cru/cruts/version_3_0/WMO/wmo_country_codes.dat
Input database:           pre.0803271802.dtb
Total stations:           15937
WMOs matched:              4196
False WMOs (<0):             88
Bad WMOs (>=0 & <1000):    1540
Bad WMO heads written to: pre.0803271802.bad
Output Database:         pre.0903051740.dtb

So.. I guess I will use tmp.0903081416.dtb, pre.0903051740.dtb, and the earliest available
from the other parameters. In other words:

cld.0902101409.dtb
dtr.0708081052.dtb
pre.0903051740.dtb
tmn.0708071548.dtb
tmx.0708071548.dtb
tmp.0903081416.dtb
vap.0804231150.dtb
wet.0710161148.dtb

Well it's worth a try. Actually, let's compare those eight databases - assuming we can find
at least some common stations!

Oh, boy:

cld.0902101409.dtb: 7015000  5056    310   52 LILLE/LESQUIN        FRANCE        1971 1996   -999     -999
cld.0902101409.dtb: 0701500  5057    310   52 LILLE/LESQUIN        FRANCE        2000 2008    -999       0
dtr.0708081052.dtb:  701500  5057    310   52 LILLE                FRANCE        1973 2006    -999       0
pre.0903051740.dtb: 0701500  5060    310   47 LILLE                FRANCE        1784 2006   -999  -999.00
tmn.0708071548.dtb:  701500  5057    310   52 LILLE                FRANCE        1973 2006    -999       0
tmp.0903081416.dtb: 0701500  5060    310   47 LILLE                FRANCE        1851 2006 101851  -999.00
tmx.0708071548.dtb:  701500  5057    310   52 LILLE                FRANCE        1973 2006    -999       0
vap.0804231150.dtb: 0701500  5057    310   52 LILLE/LESQUIN        FRANCE        2003 2007    -999       0
vap.0804231150.dtb: 1378000  6108   1048  271 LILLEHAMMER SAETHER  NORWAY        1972 1994   -999     -999
vap.0804231150.dtb: 7015000  5056    310   52 LILLE/LESQUIN        FRANCE        1971 2003   -999     -999
wet.0710161148.dtb: 0701500  5057    310   52 LILLE/LESQUIN        FRANCE        1996 2007    -999    -999

This whole project is SUCH A MESS. No wonder I needed therapy!!

So, cld already has the problem, and it's the earliest version in the archive. Also vap.

Well, looking back (er, up ^) we know what happened to cld - it was updated with newmergedb
before it went 'auto':

  'So we now have cld.0902101409.dtb, a database consisting of cld.0312181428.dtb,
   updated first with derived-cloud data from MCDW (1994-2008), then with
   derived-cloud data from CLIMAT (2000-2008).'

And, finding cld.0312181428.dtb, does it have the 'Lille problem'? No!

  70150  5056    310   52 LILLE/LESQUIN        FRANCE        1971 1996   -999     -999

So cld.0312181428.dtb is our new starting point I think.

VAP - oh, dear. Again, from above:

  'Discovered that WMO codes are still a pain in the arse. And that I'd forgotten to match Australian
   updates by BOM code (last field in header) instead of WMO code - so I had to modify newmergedbauto.
   Also found that running fixwmos.for was less than successful on VAP, because it's already screwed:

   uealogin1[/cru/cruts/version_3_0/update_top/db/vap] grep -i 'jan mayen' vap.0804231150.dtb
   0100100  7093   -867    9 JAN MAYEN(NOR-NAVY)  NORWAY        2003 2007    -999       0
   1001000  7093   -866    9 JAN MAYEN(NOR NAVY)  NORWAY        1971 2003   -999     -999'

Ulp!

I am seriously close to giving up, again. The history of this is so complex that I can't get far enough
into it before by head hurts and I have to stop. Each parameter has a tortuous history of manual and
semi-automated interventions that I simply cannot just go back to early versions and run the update prog.
I could be throwing away all kinds of corrections - to lat/lons, to WMOs (yes!), and more.

So what the hell can I do about all these duplicate stations? Well, how about fixdupes.for? That would
be perfect - except that I never finished it, I was diverted off to fight some other fire. Aarrgghhh.

I - need - a - database - cleaner.

What about the ones I used for the CRUTEM3 work with Phil Brohan? Can't find the bugger!! Looked everywhere,
Matlab scripts aplenty but not the one that produced the plots I used in my CRU presentation in 2005. Oh,
FUCK IT. Sorry. I will have to WRITE a program to find potential duplicates. It can show me pairs of headers,
and correlations between the data, and I can say 'yay' or 'nay'. There is the finddupes.for program, though
I think the comment for *this* program sums it up nicely:

  '      program postprocdupes2
   c Further post-processing of the duplicates file - just to show how crap the
   c program that produced it was! Well - not so much that but that once it was
   c running, it took 2 days to finish so I couldn't really reset it to improve
   c things. Anyway, *this* version does the following useful stuff:
   c (1) Removes and squirrels away all segments where dates don't match;
   c (2) Marks segments >5 where dates don't match;
   c (3) Groups segments from the same pair of stations;
   c (4) Sorts based on total segment length for each station pair'

You see how messy it gets when you actually examine the problem?

This time around, (dedupedb.for), I took as simple an approach as possible - and almost immediately hit a
problem that's generic but which doesn't seem to get much attention: what's the minimum n for a reliable
standard deviation?

I wrote a quick Matlab proglet, stdevtest2.m, which takes a 12-column matrix of values and, for each month,
calculates standard deviations using sliding windows of increasing size - finishing with the whole vector
and what's taken to be *the* standard deviation.

The results are depressing. For Paris, with 237 years, +/- 20% of the real value was possible with even 40
values. Windter months were more variable than Summer ones of course. What we really need, and I don't think
it'll happen of course, is a set of metrics (by latitude band perhaps) so that we have a broad measure of
the acceptable minimum value count for a given month and location. Even better, a confidence figure that
allowed the actual standard deviation comparison to be made with a looseness proportional to the sample size.

All that's beyond me - statistically and in terms of time. I'm going to have to say '30'.. it's pretty good
apart from DJF. For the one station I've looked at.

Back to the actual database issues - I need a day or two to think about the duplicate finder.
```

```
Let's just look at the year 2003, for all the French stations in each database! Duh.

Original db extraction: tmp.0705101334.dtb.FRANCE:
2003   30   29   87  106  137  186  195  208  158   85   81   45
2003   59   64   96  108  123  159  174  188  155  110  103   75
2003-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999 x
2003-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999 x
2003   10    2   83  110  162  226  210  239  159   81   67   26
2003   42   54  108  122  144  199  200  232  173  109  101   65
2003   24   34  102  120  149  221  213  252  168   99   82   45
2003   12   15   86  109  154  232  220  251  161   88   67   31
2003   20   33  102  114  137  212  200  240  162  101   90   49
2003-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999 x
2003   51   58  123  142  160  227  219  252  190  129  109   77
2003   45   56  114  133  165  242  243  267  197  133  109   72
2003-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999 x
2003   54   63  111  141  194  261  263  279  201  153  121   85
2003   85   79  115  137  191  248  253  274  209  156  129   96
2003   73   73  118  140  181  247  262  275  205  155  125   94
2003   86   73  105  131  184  241  251  265  207  166  136   93

New db extraction: tmp.0904151410.dtb.FRANCE:
2003   30   29   87  106  137  186  195  208  158   85   81   45 *
2003   59   64   96  108  123  159  174  188  155  110  103   75 *
2003   10    2   83  110  162  226  210  239  159   81   67   26 *
2003   42   54  108  122  144  199  200  232  173  109  101   65 *
2003   24   34  102  120  149  221  213  252  168   99   82   45 *
2003   12   15   86  109  154  232  220  251  161   88   67   31 *
2003   20   33  102  114  137  212  200  240  162  101   90   49 *
2003   51   58  123  142  160  227  219  252  190  129  109   77 *
2003   45   56  114  133  165  242  243  267  197  133  109   72 *
2003   85   79  115  137  191  248  253  274  209  156  129   96 *
2003   73   73  118  140  181  247  262  275  205  155  125   94 *
2003   86   73  105  131  184  241  251  265  207  166  136   93 *
2003   30   29   87  106  137  186  195  208  158   85   81   45  Do
2003   27   32   88  101  129  181  186  209  153   87   79   45  DA
2003   59   64   96  108  123  159  174  188  155  110  103   75  Do
2003   10    2   83  110  162  226  210  239  159   81   67   26  Do
2003   42   54  108  122  144  199  200  232  173  109  101   65  Do
2003   24   34  102  120  149  221  213  252  168   99   82   45  Do
2003   12   15   86  109  154  232  220  251  161   88   67   31  Do
2003   20   33  102  114  137  212  200  240  162  101   90   49  Do
2003   18   32   93  116  157  232  217  242  160  104   83   43  DB
2003   51   58  123  142  160  227  219  252  190  129  109   77  Do
2003   45   56  114  133  165  242  243  267  197  133  109   72  Do
2003   57   60  110  133  184  245  256  267  196  147  113   81  DC
2003   54   63  111  141  194  261  263  279  201  153  121   85 *
2003   85   79  115  137  191  248  253  274  209  156  129   96  Do
2003   73   73  118  140  181  247  262  275  205  155  125   94  Do
2003   86   73  105  131  184  241  251  265  207  166  136   93  Do
2003-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999-9999  -
2003   27   32   88  101  129  181  186  209  153   87   79   45  DA
2003   18   32   93  116  157  232  217  242  160  104   83   43  DB
2003   57   60  110  133  184  245  256  267  196  147  113   81  DC
2003   35   35   85  101  128  170  183  200  152   86   83   46
2003   73   69   88  102  122  156  174  185  169  129  116   84
2003   43   49   90  105  130  178  186  204  154   99   91   55
2003   17   18   83  100  143  195  199  218  148   83   77   41
2003   69   70   95  107  124  160  177  184  164  123  109   82
2003   45   53  106  120  142  195  198  225  174  112  101   64
2003   30   39   93  107  134  191  191  221  155   91   83   47
2003   67   75  105  116  136  179  189  210  173  125  120   90
2003   28   39  102  115  144  209  205  243  167   97   85   51
2003   10   -3   78  102  156  230  210  235  150   74   62   22
2003   61   67  113  128  151  197  207  229  188  136  120   89
2003   30   43  101  122  144  214  206  243  168  104   89   54
2003   -7   -7   52   77  125  202  188  206  127   73   54   13
2003   20   28  100  128  171  254  233  261  179  107   86   42
2003   33   40  112  129  154  233  220  255  176  111   98   60
2003  -29  -40   25   32   79  158  151  189   96   39   42   -4
2003   39   48  107  131  183  254  251  268  182  126  100   61
2003    4    6   79   99  148  215  223  229  158   90   68   31
2003   48   52  113  135  159  229  224  255  184  128   99   65
2003   43   42  105  120  146  214  209  241  178  116   92   60
2003   41   44  101  117  145  216  212  237  173  116   93   55
2003   81   71  111  133  183  244  246  272  200  149  128  101
2003   93   69  111  137  191  254  264  282  212  168  137  103
2003   17   24   88  105  146  208  207  234  158   91   81   42
2003    1    3   84   99  143  208  202  233  153   77   70   29

In the original db, I've x'd those lines missing in the new one. Just missing vals.
In the new db, I've asterisked all the lines matching the old one, with duplicate
matches labeled 'Do'. Any other duplicates are marked Da, DB, DC. We can see that
all the original 2003 lines are included, *and replicated*. Three new lines are also
replicated. A further 25 lines are apparently new (though could well have parents
in the original db). This implies that very little matching is being performed!!

Had a look at the .act (Action) files. Interesting..

crua6[/cru/cruts/version_3_0/update_top] gunzip -c updates/MCDW/mergefiles/merg.mcdw.tmp.0904021106.act.gz  | grep 'LILLE'
Master:  701500  5034    306   52 LILLE                FRANCE        1851 2008  101851 -999.00
Update: 0701500  5034    306   52 LILLE                FRANCE        2001 2008    -999       0
NewH:    701500  5034    306   52 LILLE                FRANCE        1851 2008  101851 -999.00
crua6[/cru/cruts/version_3_0/update_top] gunzip -c updates/MCDW/mergefiles/merg.mcdw.tmp.0904021239.act.gz  | grep 'LILLE'
Master:  701500  5034    306   52 LILLE                FRANCE        1851 2008  101851 -999.00
Update: 0701500  5034    306   52 LILLE                FRANCE        2001 2008    -999       0
NewH:    701500  5034    306   52 LILLE                FRANCE        1851 2008  101851 -999.00
crua6[/cru/cruts/version_3_0/update_top] gunzip -c updates/MCDW/mergefiles/merg.mcdw.tmp.0904151410.act.gz  | grep 'LILLE'
Master:  701500  5034    306   52 LILLE                FRANCE        2000 2009    -999       0
Update: 0701500  5034    306   52 LILLE                FRANCE        2009 2009    -999       0
NewH:    701500  5034    306   52 LILLE                FRANCE        2000 2009    -999       0

So it worked fine until the 0904151410 run, when it went crazee.
So.. what happened? Why did it behave differently? No idea. It was the same for pre though!

crua6[/cru/cruts/version_3_0/update_top] gunzip -c updates/MCDW/mergefiles/merg.mcdw.pre.0904021239.act.gz  | grep 'LILLE'
Master:  701500  5034    306   52 LILLE                FRANCE        1784 2008    -999 -999.00
Update: 0701500  5034    306   52 LILLE                FRANCE        2001 2008    -999       0
NewH:    701500  5034    306   52 LILLE                FRANCE        1784 2008    -999 -999.00
crua6[/cru/cruts/version_3_0/update_top] gunzip -c updates/MCDW/mergefiles/merg.mcdw.pre.0904151410.act.gz  | grep 'LILLE'
Master:  701500  5034    306   52 LILLE                FRANCE        2000 2009    -999       0
Update: 0701500  5034    306   52 LILLE                FRANCE        2009 2009    -999       0
NewH:    701500  5034    306   52 LILLE                FRANCE        2000 2009    -999       0

There was something very fishy about that run. Of course it was a single month - I wonder if that made a difference?

crua6[/cru/cruts/version_3_0/update_top] cat runs/runs.0904021239/conv.mcdw.0904021239.dat
       9    1994      12    2008
crua6[/cru/cruts/version_3_0/update_top] cat runs/runs.0904151410/conv.mcdw.0904151410.dat
       1    2009       1    2009
```

```
Also of interest - how did the program find a 2000-2009 station when the previous update was to 2008?

Aha:

crua6[/cru/cruts/version_3_0/update_top] cat runs/runs.0904021239/conv.climat.0904021239.dat
       1    2000       2    2009

The CLIMAT update did it!! It's that bloody no-metadata problem!! So I should be looking at the
CLIMAT process for 0904021239, not the MCDW one. Duhh. So, the merge run:

crua6[/cru/cruts/version_3_0/update_top] cat runs/runs.0904021239/merg.climat.0904021239.dat
db/tmx/tmx.0708071548.dtb
updates/CLIMAT/db/db.0904021239/climat.tmx.0904021239.dtb
updates/CLIMAT/db/db.0904021239/int2.tmx.0904021239.dtb
blind
M
crua6[/cru/cruts/version_3_0/update_top]

Looking at the CLIMAT dtb conversion...

The converted CLIMAT bulletins:
crua6[/cru/cruts/version_3_0/update_top] gunzip -c updates/CLIMAT/db/db.0904021239/climat.tmp.0904021239.dtb.gz |grep -i 'lille'
0701500  5057    310   52 LILLE/LESQUIN        FRANCE        2000 2009    -999       0

The merged bulletins in tmp.0904021239.dtb (er, presumably)
crua6[/cru/cruts/version_3_0/update_top] gunzip -c updates/CLIMAT/db/db.0904021239/int2.tmp.0904021239.dtb.gz |grep -i 'lille'
0701500  5057    310   52 LILLE/LESQUIN        FRANCE        2000 2009    -999       0
7015000  5034    306   52 LILLE                FRANCE        1851 2008  101851 -999.00

crua6[/cru/cruts/version_3_0/update_top] grep -i 'lille' db/tmp/tmp.0*dtb
db/tmp/tmp.0705101334.dtb:  70150   506     31   47 LILLE                FRANCE        1851 2006 101851  -999.00

db/tmp/tmp.0903081416.dtb:0701500  5060    310   47 LILLE                FRANCE        1851 2006 101851  -999.00

db/tmp/tmp.0904021106.dtb:0701500  5057    310   52 LILLE/LESQUIN        FRANCE        2000 2009    -999       0
db/tmp/tmp.0904021106.dtb:7015000  5034    306   52 LILLE                FRANCE        1851 2008  101851 -999.00

db/tmp/tmp.0904021239.dtb:0701500  5057    310   52 LILLE/LESQUIN        FRANCE        2000 2009    -999       0
db/tmp/tmp.0904021239.dtb:7015000  5034    306   52 LILLE                FRANCE        1851 2008  101851 -999.00

db/tmp/tmp.0904151410.dtb: 701500  5034    306   52 LILLE                FRANCE        2000 2009    -999       0
db/tmp/tmp.0904151410.dtb:7015000  5034    306   52 LILLE                FRANCE        1851 2008  101851 -999.00

OK.. let's be absolutely clear about the process.

STEP 1 - convert MCDW bulletins (09/94 - 12/08) to produce mcdw.tmp.0904021106.dtb:
0701500  5034    306   52 LILLE                FRANCE        2001 2008    -999       0

STEP 2 - Merge mcdw.tmp.0904021106.dtb into tmp.0903081416.dtb to produce int1.tmp.0904021106.dtb:
 701500  5034    306   52 LILLE                FRANCE        1851 2008  101851 -999.00

STEP 3 - convert CLIMAT bulletins (01/00 - 02/09) to climat.tmp.0904021106.dtb:
0701500  5057    310   52 LILLE/LESQUIN        FRANCE        2000 2009    -999       0

STEP 4 - Merge climat.tmp.0904021239.dtb into int1.tmp.0904021106.dtb to produce int2.tmp.0904021106.dtb:
0701500  5057    310   52 LILLE/LESQUIN        FRANCE        2000 2009    -999       0
7015000  5034    306   52 LILLE                FRANCE        1851 2008  101851 -999.00

(there's then the BOM section but it's all over by now)

So, the merging of climat.tmp.0904021106.dtb into int1.tmp.0904021106.dtb FAILED. WHY?

Well, the WMO codes are the same as for MCDW: 0701500. So it can't be that. The lat and lon
are ~slightly~ different, though. Remember, the DATABASE entry was originally (tmp.0903081416.dtb):

0701500  5060    310   47 LILLE                FRANCE        1851 2006 101851  -999.00

After the MCDW (09/94 - 12/08) merge, it became (int1.tmp.0904021106.dtb):

 701500  5034    306   52 LILLE                FRANCE        1851 2008  101851 -999.00

After the CLIMAT (01/00 - 02/09) merge (int2.tmp.0904021106.dtb):

0701500  5057    310   52 LILLE/LESQUIN        FRANCE        2000 2009    -999       0
7015000  5034    306   52 LILLE                FRANCE        1851 2008  101851 -999.00

Now, the 'LILLE/LESQUIN' station header comes from the CLIMAT bulletins, ie, from the WMO
reference file wmo.0710151633.dat. But it should have matched with the existing LILLE -
the problem looks like the latitude shift (from 50.60 to 50.34) introduced by MCDW did
the damage. Obviously, if we are going to trust MCDW metadata as being valid corrections,
then the WMO reference file needs to be updated at the same time!! So, we'll need:

1. A file called 'wmoref.latest.dat' that contains the name of the latest WMO reference file.
  - DONE

2. A hook in newmergedbauto that flags when a header is being changed by an MCDW/BOM bulletin.
  - EXTREMELY COMPLICATED

3. A routine to write a new WMO reference and to archive the old one.
  - EXTREMELY COMPLICATED

4. A record of every iteration of the db/latest.versions.dat file (in db/previous.latest/).
  - DONE

As part of the investigations, I found that I wasn't close()-ing off channel 10 when I used
it in update.for. Now I'm pretty sure that F77 follows the convention that an OPEN on an
open channel initiates an initial CLOSE automatically, but who wants to take that chance with
the variety of compilers we're subject to? So I went through and inserted an indecent number
of close(10)s.

Point 2 - flagging changes to the metadata.

Well, the merged database is written principally from dbm*, with dbu* chipping in 'new' stations.
I guess that new stations should be added to the wmo reference file? They are pan-parameter (well
the MCDW ones are) but I have an eerie feeling that I won't experience joy when headers are
compared between parameters :/

Wrote metacmp.for. It accepts a list of parameter databases (by default, latest.versions.dat) and
compares headers when WMO codes match. If all WMO matches amongst the databases share common
metadata (lat, lon, alt, name, country) then the successful header is written to a file. If,
however, any one of the WMO matches fails on any metadata - even slightly! - the gaggle of
disjointed headers is written to a second file. I know that leeway should be given, particularly
with lats & lons, but as a first stab I just need to know how bad things are. Well, I got that:

crua6[/cru/cruts/version_3_0/update_top] ./metacmp
METACMP - compare parameter database metadata
RESULTS:

Matched/unopposed:  2435
```

```
Clashed horribly:   4077

Ouch! Though actually, far, far better than expected. As for the disport of those 2435:

crua6[/cru/cruts/version_3_0/update_top] grep '^1' report.0909181759.metacmp.wmo | wc -l
      1250
crua6[/cru/cruts/version_3_0/update_top] grep '^2' report.0909181759.metacmp.wmo | wc -l
       279
crua6[/cru/cruts/version_3_0/update_top] grep '^3' report.0909181759.metacmp.wmo | wc -l
        41
crua6[/cru/cruts/version_3_0/update_top] grep '^4' report.0909181759.metacmp.wmo | wc -l
        92
crua6[/cru/cruts/version_3_0/update_top] grep '^5' report.0909181759.metacmp.wmo | wc -l
        83
crua6[/cru/cruts/version_3_0/update_top] grep '^6' report.0909181759.metacmp.wmo | wc -l
         9
crua6[/cru/cruts/version_3_0/update_top] grep '^7' report.0909181759.metacmp.wmo | wc -l
       129
crua6[/cru/cruts/version_3_0/update_top] grep '^8' report.0909181759.metacmp.wmo | wc -l
       552

Interesting, but not astounding. Roughly half are unpaired stations, with an impressive
23% showing a perfect match across all eight databases.

Analysis of the 4000+ bad matches will be more complicated unfortunately. An initial
re-run looking for lat/lon within half a degree, and/or station partial, will be useful.

No, hang on. Easier to analyse the output from metacmp! And so.. postmetacmp.for:

Stats report for: report.0909181759.metacmp.bad

  Overall distribution of group sizes:
  2 in group:  642
  3 in group:   71
  4 in group:  188
  5 in group:  625
  6 in group:  183
  7 in group:  411
  8 in group: 1957

LAT:
  Number of diffs within a group:
  1.    0
  2. 3059
  3.  276
  4.   15
  5.    0
  6.    0
  7.    0
  8.    0
  Maximum differences:
  <0.1: 1233
  <0.2:  726
  <0.5: 1225
  <1.0:   15
  1.0+:  151

LON:
  Number of diffs within a group:
  1.    0
  2. 2996
  3.  339
  4.   30
  5.    1
  6.    0
  7.    0
  8.    0
  Maximum differences:
  <0.1: 1195
  <0.2:  722
  <0.5: 1242
  <1.0:   30
  1.0+:  177

ALT:
  Number of diffs within a group:
  1.    0
  2. 2035
  3.  237
  4.   17
  5.    0
  6.    0
  7.    0
  8.    0
  Maximum differences:
  <50m : 1767
  <100m:   75
  <500m:  121
  <1km :   36
  1km+ :  290

STATION NAME:
  Number of diffs within a group:
  1.    0
  2. 2167
  3.  365
  4.   43
  5.    0
  6.    0
  7.    0
  8.    0
  Worst percentage matches:
  <25% :  281
  <50% :  385
  <75% :  770
  <100%:  276
  100% :  863

COUNTRY NAME:
  Total groups with country mismatches: 1698
  Number of diffs within a group:
  1.    0
  2. 1475
  3.  182
  4.   41
  5.    0
  6.    0
  7.    0
  8.    0
```

Hmmm.. lots of groups that could be eliminated if we incorporated the WMO reference list, because then we could allow an element of 'drift' from a reference point.

Decided to make it a bit quicker and easier as well, by removing tmn/tmx and letting dtr take the strain - they should all be identical anyway.