

Procedure for updating the databases underlying the CRU high-resolution grids.

Tim Mitchell, 25.06.03, revised 30.3.04

1. **Transform additional station data into CRU time-series**
(.cts) format. The .cts subroutines are in crutsfiles.f90.
The .cts files should be stored under ~/data/stnmon
The .f90 programs described are under ~/code/linux/cruts or
~/code/alpha/cruts . The programs should be easily portable from
the Alphas to Linux or vice versa.
If the additional data is in the style of ...
 - (a) GHCNv2 or CLIMAT (the Phil Jones format)
i.e. one file per time interval,
use (or modify) makecruts.f90.
 - (b) MCDW or CLIMAT (original) or CLIMAT (AOPC-Offenbach) format
i.e. one file per year and month,
use or modify reformat.f90. The MCDW data must have already
gone through a two stage process - see the readme file in
~/data/stnmon/mcdw/_raw
 - (c) Jian's Chinese data from Excel
i.e. a single ASCII table per variable, with one
line per station/year, use (or modify) fromexcel.f90.
 - (d) the CRU time-series file format (but not quite right)**
it may be easily convertible using option 1
in oprcuts.f90.
2. **The size of the arrays required in the entire procedure can be substantially reduced by subdividing the additional station data by continent at this stage. Simply subdivide the initial raw data into a set of raw files, by continent.** Reduced array sizes mean programs that run more quickly and reliably.
3. Clean up the metadata in the .cts headers. This is done using the information in the master metadata file, which is the most recently dated file in /cru/tyn1/f709762/cruts/master. Run cleanmeta.f90 on the transformed CRU ts file. The sole purpose of this program is to make the header line as accurate as possible, without adding new information. Thus the following steps are included:
 - (A) **The original station code is stored** as a 7-digit code in both the main and 'old' code columns.
 - (B) The station and country labels are made all-caps and any hyphens (etc.) are removed.
 - (C) Impossible lat/lon/elv values are setting to missing.
 - (D) The country label is checked, and made consistent, using the master country list.
 - (E) The lat/lon are checked to ensure that they are

reasonable, using the country information. Each country is given a centroid and a 3-sigma distance. Stns lying outside this radius are flagged.

(F) If a corresponding source code file (.src) is available, it too is checked, else one is created using information about the source supplied by the user.

4. Check the homogeneity of the additional .cts file. This is done using reference time-series. This is far too complex to describe here. Run homogiter.f90 on the cleaned .cts file.

The program runs iteratively, to maximise the proportion of the original data that can be checked and placed in homog.

A .cts (and .src) file with best estimates is stored in ~/cruts/homog, and stns that could not be checked are stored in ~/cruts/retry

This program may fire up 2 subsidiary xterms on start-up. (If given the option, decline.)

These can be killed forcibly if necessary once the program has finished executing, but not before. One xterm is simply a view of the log file, which provides a rough progress meter. The other is the idl window, running an IDL program that awaits for prompts through a pipe (stored in /cru/scratch2/f709762) to tell it when and where there is a data file for it to read and plot, in (f).

5. Merge the additional .cts file with the existing database (.dtb and .dts). The most recent version of each database is in ~/data/cruts/database, the latest version of the master station code file is in ~/data/cruts/master, and the latest accessions file is in ~/data/cruts/accession

Use updateddtb.f90 to merge the new file in ~/cruts/homog with the existing database. Do this immediately, before going through the whole process for a new region, to ensure that as much info as possible is available for creating reference series for the new region.

6. OPTIONAL AT THIS STAGE

Add normals to the database file. The best time to do this is when all the new information has been absorbed into the database (steps 1-5) and the database is about to be used for gridding. The program addnorm.f90 is used to add the 1961-90 normal in a header line in the .dtb files. This is then used by the program that calculates anomalies.

Where possible, a normal is calculated from the station series itself. Where this is not possible through insufficient data, an attempt is made to estimate what the normal would have been if it had been measured. This estimate is made using the reference series construction software used in

step 4. **Neighbouring stations are used to construct a reference time-series that includes 1961-90 wherever possible, and this reference time-series is used to calculate a 1961-90 normal for that stn**, and stores it in the 'normal' line in the database file.

7. ADDITIONAL CAPABILITY, REQUIRED FOR GRIDGING

Transform the database file (with normals added under step 6, if possible) from absolute values to anomaly values, prior to gridding, using anomdtb.f90. The key output option from this program is (3), which dumps the anomalies to a set of .txt files that can be read by the idl gridding software. The other output options provide info - (1) produces the same data in the .cts format, (2) only summarises the outputs through data counts, (4) summarises the original data discarded as duplicates.

The usual options to select are:

normal period = 1961 - 1990
missing percent permitted = 25
stdevs to reject = 3
duplicate stns = 8km

8. ADDITIONAL CAPABILITY, USE AS REQUIRED

The program opcruts.f90 is the home for all the little useful routines for manipulating the .cts and .dtb (and .src and .dts) files. Option 1 can be used to convert from one of these formats to another. The other options can be explored to find out what they do.