D2.1 Description of scientific workflows

28 February 2011

This document represents deliverable 2.1 of the ACRID project (http://www.cru.uea.ac.uk/cru/projects/acrid/), prepared by UEA and STFC.

Introduction

This work package analyses the scientific workflows associated with the climate research datasets being considered within the ACRID project. From these workflows, the requirements for capturing software and dataset metadata have been identified and are also represented here. The information flow patterns have been identified in order to construct a general architecture which is applicable to other HEIs. The workflows captured in this document will be refined over the course of the project.

Contents

This document provides an overview of the key concepts associated with the workflows for:

a) CRUTEM3
b) CRU TS 3.0
c) CRU Tree-ring chronologies

Appendix 1: a more detailed text description of the CRUTEM3 workflow (some of which is also applicable to CRU TS)

Appendix 2: flow charts depicting the workflow used in constructing the CRU TS 3.0 dataset

Appendix 3: additional background information for the tree-ring chronologies extracted from the CRU submission to Muir Russell.

a) The Main Concepts of the CRUTEM3 Workflow

This section attempts to identify the key concepts associated with the CRUTEM3 workflow. These concepts are divided into the following two categories:

- The "Dataset" concept represents the different versions/variants of CRUTEM3 and the outcomes of their analyses that are published or referenced in publications (Table 1).
- The "Process" concept captures the information about a process, experiment or analysis that is conducted to construct (or contributes towards the construction of) a dataset (e.g. CRUTEM3). A process may depend on other processes; e.g. the outputs of one process may be used as the inputs of another, where the outputs of the former may or may not include a publishable dataset (Table 2).

1. Datasets

Name/Identifier	**Description**	**Assertions**		**Comments**
		Property	**Value**	
CRUTEM3	CRUTEM3 gridded temperature anomalies	**Constructed By**	Process "Gridding" (see Table 2)	
		Ownership	e.g. Climatic Research Unit, University of East Anglia	
		References	published articles describing the construction of the CRUTEM3 datasets or other related publications	
		Formats	NetCDF, Plain text format	
		Creation date	Day, month, year, time	

		Last modified date	Day, month, year time	NB. CRUTEM terminology was introduced with CRUTEM2 – CRUTEM1 was not actually named.
		Related versions	CRUTEM2, CRUTEM1	
		Data quality information		
		Access rights		
		Download link		
		Grid box Centre	Lat, Lon	
		Coverage period	1850-2010	
		Other information		
CRUTEM3v	a “variance adjusted” version of CRUTEM3	**Constructed by**	Process “Adjusting variance” (see Table2)	
		Ownership	e.g. Climatic Research Unit, University of East Anglia	
		References	published articles describing the construction of the CRUTEM3v datasets or other related publications	
		Formats	NetCDF, Plain text format	
		Creation date	Day, month, year, time	
		Last modified date	Day, month, year time	
		Related versions	CRUTEM2v, CRUTEM3	
		Data quality information		
		Access rights		
		Download link		

		Grid box Centre	Lat, Lon	
		Coverage period	1850-2010	
CRUTEM Station data	Station data	**Ownership**	e.g. Climatic Research Unit, University of East Anglia (note that the National Met Services may still "own" the station data, though CRU/UEA have assembled the dataset of station temperatures).	
		Formats	NetCDF, Plain text format	
		Station name/identifier		
		Station location	Lat, Lon	
		Source info		
		Acquisition date	Day, month, year, time	
		Last modified date	Day, month, year time	
		Data quality/ Other information		
Hemispheric and global means		**Constructed by**	Process "Calculation of hemispheric and global means" (see Table2)	
		Ownership	e.g. Climatic Research Unit, University of East Anglia	
		References	Reference to related publications	
		Formats	NetCDF, Plain text format	
		Creation date	Day, month, year, time	

		Last modified date	Day, month, year time	
		Related versions	CRUTEM3, CRUTEM 2, CRUTEM 1	
		Data quality information		
		Access rights		
		Download link		

2. Processes

Name/Identifier	**Description**	**Assertions**		**Comments**
		Property	**Value**	
Gridding	CRUTEM3 gridded temperature anomalies	**Inputs**	Outputs of process “Removal of outliers”	• Separate notions need to be defined to represent Process Inputs and Outputs • “Software” and “Processor” concepts need to be defined separately
		Outputs	CRUTEM3 The number of station observations within each grid box for each month and year	
		Performed by	Software or human (Processor Information)	
		Performed date	Day, month, year time	
		Algorithms/Method used	The grid-box mean anomaly is the average all the available values. Equation. Could even be a web page	

			describing the algorithm.	
		Other information		
Calculation of "normals"	Calculation of "normals"	**Inputs**	(1….*) Station data (see Table 1)	
		Performed by	Software or human (Processor Information)	
		Performed date	Day, month, year time	
		Algorithms/Method used	Description of the algorithm used. Could even be a web page describing the algorithm.	
		Output	"Monthly Normal" for each Station	
		Other information	I f insufficient data the "normals" may be from other sources	
Calculation of standard deviations	Calculation of standard deviations	**Inputs**	(1….*) Station data (see Table 1)	NB This is an iterative process due to recalculation after outlier removal.
		Performed by	Software or human (Processor Information)	
		Performed date	Day, month, year time	
		Algorithms/Method used	Description of the algorithm used. Could even be a web page describing the algorithm.	
		Output	The monthly temperature standard deviation for each Station	

		Other information		
Conversion to anomalies	deviations from a reference value	**Inputs**	(1….*) Station data (see Table 1) Outputs of Process "Calculation of Normals"	
		Performed by	Software or human (Processor Information)	
		Performed date	Day, month, year time	
		Algorithms/Method used	Description of the algorithm used. Could even be a web page describing the algorithm.	
		Output	Anomalies	
		Other information		
Removal of outliers	Quality control	**Inputs**	Outputs of Process "Conversion of anomalies" (Anomalies)	
		Performed by	Software or human (Processor Information)	
		Performed date	Day, month, year time	
		Algorithms/Method used	Description of the algorithm used. Could even be a web page describing the algorithm.	
		Output	Updated Anomalies	

		Other information		
Calculation of hemisphere and global means	The hemispheric means are averages of the grid boxes for which temperature anomalies are available (this is, of course, time dependent), with weighting according to the area of the grid boxes (proportional to the cosine of their central latitude)	**Inputs**	CRUTEM3	
		Performed by	Software or human (Processor Information)	
		Performed date	Day, month, year time	
		Algorithms/Method used	Description of the algorithm used. Could even be a web page describing the algorithm.	
		Output	hemisphere and global means	
		Other information		
Variance Adjusting	Adjusting the high-frequency temporal variance	**Inputs**	CRUTEM3 The number of station observations within each grid box for each month and year Estimates of average inter-station correlations for each grid box and for each month of the year (RBAR)	The variance adjustment is made direct to the CRUTEM3 gridded data, using information about the number of observations used to calculate each gridded anomaly and an estimate of how well-correlated stations are likely to be.
		Performed by	Software or human (Processor Information)	
		Performed date	Day, month, year time	
		Algorithms/Method used	Description of the algorithm used. Could even be a web page describing	

			the algorithm.	
		Output	CRUTEM3v	
		Other information		

b) The UEA CRU TS 3.0 workflow breakdown:

1. Processes

Step Name	Description	Input	Output	Performed By	Other information	Comments/questions
Construction	Involves the following processes: • Database Update • Data anomalisation.	UEA gridded data-sets of climate observations previously developed (See 3. Datasets).	CRU TS 3.x (See 2. Dataset)	Software	e.g. Purpose: the primary purpose for which this data-set was constructed was to provide environmental modellers with some of the inputs they require to run their models. This purpose governed the choices that were made during the construction of the data-set.	
Database update	Example: the CRU station databases were updated to ensure sufficient station coverage to 2000.		Updated CRU database	Software	Any other relevant information	
Anomalisation .	Example: The station data for 1901–2000 was anomalised relative to 1961–1990.	Station data	Station data	Software		

Analysis	Example: time-series analysis of CRU TS 2.0	Data associated with one or more CRU TS variables (See 3. Variables)	Result of the analysis, i.e. 5. Derived data	Software	Assumptions made, issues identified etc.	
Publishing	Publish an analysis of CRU TS 2.0	N/A	publication	Human(s)	• Publisher info • URL/DOI etc.	

2. Dataset

Identifier	**Name**	**Space**	**Time**	**Variables**	**Dataset Type**	**Reference**	**Ownership**	**Comments/Questions**
A unique identifier for the dataset (E.g. CRU TS 2.0)	A textual name of the dataset	Example: 0.5° globe	Example: 1901-2000	One or more instances of Variable concept defined in Table 3	Example: time-series	Related publications	Example: Dr. T. D. Mitchell and other related information	• Elevation files - on BADC website

3. Dataset Variable

Code/Identifier	**Description**	**Unit**	**Data Files**	**Comments/Questions**
Example: tmp	A textual description of the derived data,	Example: °C * 10	One or more Data File concepts defined in Table 4	

4. Data File

Time	Format	Download link	Comments/Questions
Example: 1901-1911 or all	The format in which the derived data is available. Example: NetCDF.	A web link allowing direct access to the derived data	• If this is defined as an RDF property, it should be restricted to Dataset Variable domain.

6. Dataset Type

Name/code	nature	purpose	Comments/Questions
Example: TS (time-series)	Example: Month-by-month variations in climate over the last century or so. These are high-resolution grids.	Example: Allows the comparison of variations in climate with variations in other phenomena.	

7. Derived Data

Description	Source	Format	Download link	Comments/Questions
A textual description of the derived data,	One for more datasets (Table 3)	The format in which the derived data is available. Example: NetCDF.	A web link allowing direct access to the derived data	

c) The UEA tree-ring chronology dataset workflow breakdown:

Step name	Description	Input	Output	Performed by	Other information	Comments
Data acquisition	Acquired from source(s), e.g. ITRDB, other scientists, measured by CRU	N/A	Dated measurement series from many individual tree cores from many sites	Human/software	• Source metadata (source name, etc.) • Site metadata (geospatial information – e.g. Lon/Lat, Elevation) • Tree metadata (e.g. tree genus, sampling date, tree height, pith-offset estimates)	Perhaps store this information (not the data, just the meta-data) using a subset of the new TriDaS standard?
Cross dating	Check and (if needed) correct the dating by cross-dating between tree cores	Dated measurement series	Dated measurement series with possible corrections to the dating	Software	• Software Name, version etc. • SW configuration info • SW runtime parameters? • Information about dating accuracy	
Selection	Select sample of tree cores from one or multiple sites	Correctly dated measurement series	Subset of the dated measurement series	Human/software	• Identifiers of selected subset • Selection criteria	

Step name	Description	Input	Output	Performed by	Other information	Comments
Standardisation	To remove the influence of tree biological age	Subset from the raw measurement database	Tree-ring chronology (and sample counts, standard deviations/uncertainty)	Software	• Software Name, version etc. • SW configuration info • SW runtime parameters	This step is our primary focus in this project – the other steps are included for completeness, ready for future expansion
Analysis	e.g. calibration of tree-ring chronology to represent some climate variable	Tree-ring chronology	Climate reconstructions including uncertainties or confidence intervals	Software	• Software Name, version etc. • SW configuration info • SW runtime parameters	
Publishing	Publish the results of different data analyses	Multiple tree-ring chronologies and/or climate reconstructions derived from tree-ring chronologies	Publication	Human(s)	• Publisher info • URL/DOI etc.	

Further information about each step:

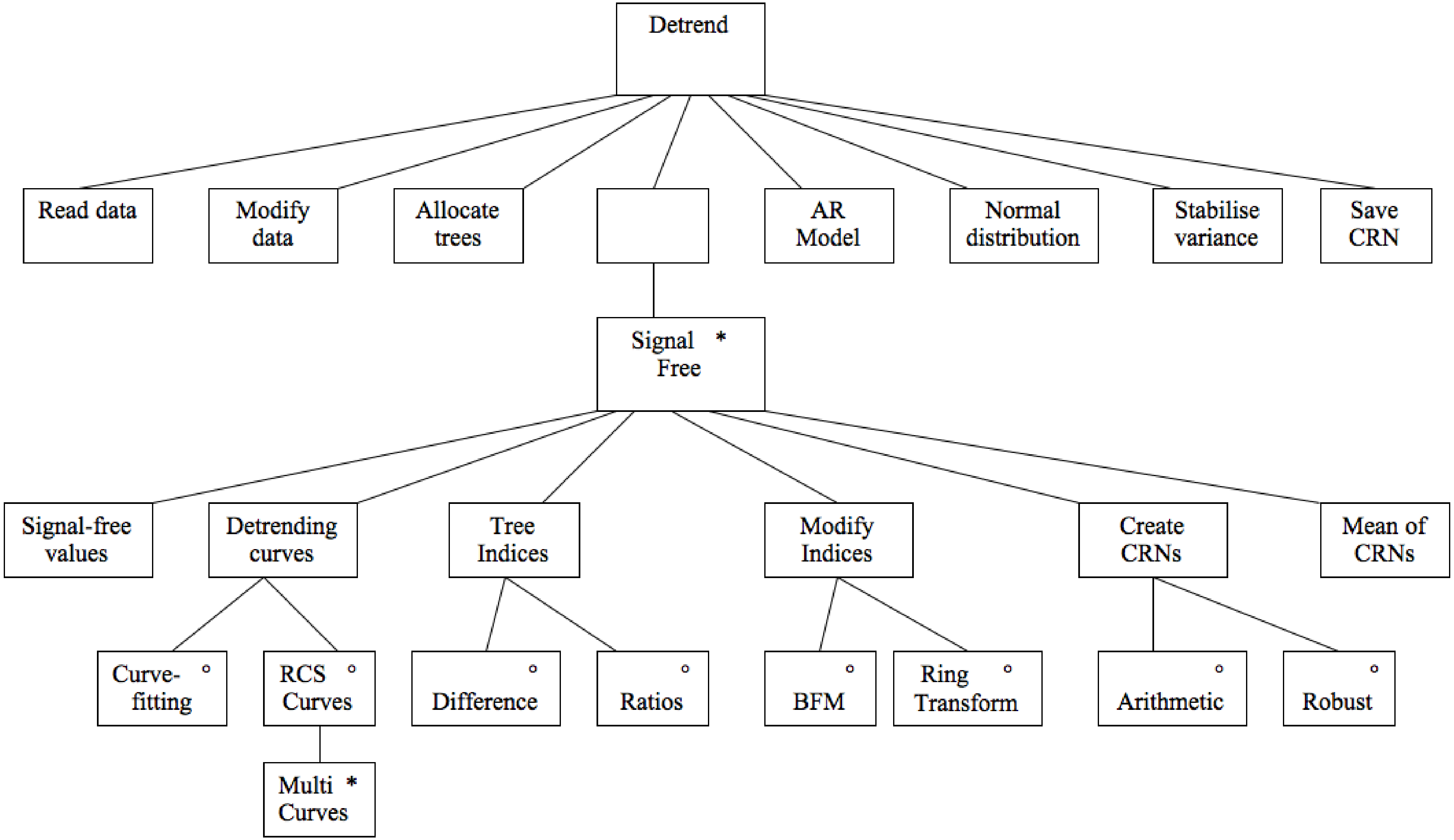
(1) Data acquisition. Do we need to specify the format the data are stored in at this stage?

(2) Selection. We might make several chronologies from different subsets tree-ring data from a particular region, so it would be useful to record the subset used in each case. If each tree-core measurement series is represented/published as a linked-data object ("thing"), can we just point to the URLs of all the tree-cores used in this subset?

(3) Standardisation. This is the crucial step that we want to focus on in this project, using Tom Melvin's software (and hence version number, configuration/runtime parameters need to be recorded). There are various standardisation possibilities, ranging from the overall approach (e.g. RCS = Regional Curve Standardisation versus Curve Fitting to individual tree-core series) through to specific deviations in the overall approach (e.g. traditional versus signal-free). Other options are whether to average multiple tree-core series from one tree (or wood sample) together prior to analysis, or whether to treat them all as individual samples. See diagram below for a flavour of the operations and choices that should be recorded.

(4) Analysis. There are a huge range of analyses that might be undertaken with the chronologies obtained from standardisation. I've listed "calibration", but this is only one possibility. We don't want to do much on this step in this project, but thought we should at least be aware of it for future extension.

Standardisation step: a flavour of the operations and choices:



Appendix 1

CRUTEM scientific and data workflow description

Overview of the workflow

Though this document is specifically about the CRUTEM gridded land temperature dataset, it is worth noting that the steps involved are more generic, being partly common with other land temperature datasets. The sequence of steps from measurement to global gridded temperature is roughly this:

(1) Measurements made at weather stations, recorded (on paper or computer), averaged from daily to monthly, digitised onto computer.

(2) Compile many weather station records into multi-station databases.

(3) Compile a number of multi-station data bases into station databases with good global coverage.

(4) Grid the global station database to make a gridded temperature dataset (and later combine with sea temperatures and make global averages etc.).

Step (3) has been done separately by CRU and GHCN/NCDC, resulting in two step (3) station databases.

Step (4) has been done by:

- by CRU using CRU step (3) database
- by NCDC using GHCN step (3) database
- by GISS using GHCN step (3) database

CRU, NCDC and GISS all use different methods of gridding the data in step (4), which result in small differences in outcome.

For CRUTEM, this document describes the following steps (note that (a) is provided for background only, since the intention in ACRID is not to document the historical development of the dataset, beyond what is already available, but rather to set up a system for the current and future updating):

(a) the station database files and where they originally came from;

(b) how MetOffice-HadleyCentre (MOHC) updates their copy of the station database;

(c) how CRU updates our copy of the station database;

(d) how they are used to make the gridded dataset;

(e) how the gridded dataset is used to make the global & hemispheric means;

(a) Station database files and their origins

As noted above, there are various databases of station temperature data, though the GHCN and CRU databases are most well known – step (3) in the introduction. CRU and GHCN used *similar* step (2) sources but not always the same to obtain their step (3) station databases. Only about ~95% of the weather station records in the CRU database have a record in GHCN, the other ~5% must have come from sources that GHCN did not use. CRU used some sources on the basis that CRU would not pass the data on to others. This is why CRU needs agreement before passing the CRU station database on in full (this is being worked on with the MetOffice).

Even for the 95% of weather stations in the CRU database that have an entry in GHCN, the data may not be identical. It may begin earlier or later, end earlier or later, or the values themselves may be different.

The earlier/later start/end can be related to which step (2) sources CRU or GHCN chose to use or gave priority to.

The reason why the *values* may differ relates to errors and/or inhomogeneities.

- "Errors" may be introduced in any part of step (1) -- simple semi-random mistakes.
- "Inhomogeneities" may be introduced if the weather station moves, thermometer changes, measurement protocol changes (e.g. what time is the measurement made).

Errors can be identified and removed/corrected via quality control (QC). Many inhomogeneities can be identified and partly adjusted via homogeneity checking (HC), but some cannot be (especially smaller ones which may not show up clearly).

QC and HC have been done for some stations by some national meteorological services, but not for others. Thus some databases at step (2) have such corrections, some do not.

The global station databases obtained at step (3) have also been QC'd and HC'd by CRU and by GHCN. The mix of sources (with/without QC and/or HC) and the use of their own QC and HC means that CRU step (3) station databases and GHCN's can differ in their exact values.

REMEMBER: the step (4) products have relatively small differences between them, especially in terms of the global-mean temperatures, so though these differences exist they are not actually of huge importance at the global scale.

The following table provides some information on the development of the CRUTEM station temperature database during the last 25 years.

Overview of the global land temperature data used for HadCRUT3, the Met-Office-Hadley-Centre/UEA-CRU global temperature data set

Timing	Data / step	Availability	Commentary
Early–mid 1980s	Obtain or digitise temperature measurements recorded by weather stations around the world • Sources include: o World Weather Records o Met Service yearbooks o etc.	Data are still available from these primary sources. Many are also available from a single source, the Global Historical Climatology Network	Sources are documented in TR017
1985 1985/6	The outcome was the "CRU station temperature primary data base" Visual and statistical analysis This led to one of three decisions being made: • The original data values can be used • The original data values can not be used due to obvious problems • The original data values show some obvious problems (perhaps due to a station being moved to a different location) but they can be adjusted by comparison with neighbouring stations The outcome was the "CRU station temperature adjusted data base (1985/6 version)"	The list of stations in this database is recorded in TR022/TR027 The types of analysis undertaken, and the results of these analyses, were reported in TR022/TR027 TR022/027 list the outcomes of the decisions and they also list any adjustments that were made (the years of adjustment and the values that were added or subtracted to the data from those years) The list of stations in this database is recorded in TR022/TR027	Although we call it the "adjusted" data base, the majority of the data were not adjusted because the original data values did not need adjusting
Ongoing work from 1986 to present	Month-by-month updates for those stations that are still recording (and whose data we could easily obtain) were appended to the CRU station temperature adjusted data base. Other sources of additional data obtained, especially from parts of the world with sparse data. Some were supplied with agreement that we would not pass them on to any one else Other sources of replacement data obtained, if national meteorological services undertook their own homogeneity adjustments and created an improved data set for their country. These replaced the data previously in the CRU station temperature adjusted data base.	The month-by-month station temperature updates since 2000 are available at the hadobs.org website These sources of additional and replacement station data are given in various scientific publications (Jones, 1994; Jones and Moberg, 2003; Brohan et al., 2006)	Since 1986 a continuous process of updating the data base has been followed, with new data each year or each month being appended. In recent years, this updating has been undertaken jointly with the Met Office Hadley Centre. In addition, improvements to the older data have been made on an irregular basis, either adding in new data not previously available to us, or replacing our data with improved versions provided by national meteorological services.
Present	The outcome of these updates and improvements is the "CRU station temperature adjusted data base (current version)" This current version of the CRU station temperature adjusted data base is gridded onto a regular grid across the land surface of the Earth (CRUTEM3), and combined with gridded	The current version will be released soon The gridding program (and accompanying data files) to generate the gridded land temperature dataset (CRUTEM3), that is run at the	In recent years, the gridding process and the generation of the global temperature record (HadCRUT3), is undertaken at the Met Office Hadley

	temperatures for the ocean surface (HadSST2), to obtain the global temperature record (HadCRUT3)	Met Office Hadley Centre, will be released soon (**TO BE CONFIRMED!**)	Centre

TR017 – Bradley, R.S., Kelly, P.M., Jones, P.D., Goodess, C.M. and Diaz, H.F., 1985: A Climatic Data Bank for Northern Hemisphere Land Areas, 1851-1980, U.S. Dept. of Energy, Carbon Dioxide Research Division, *Technical Report TRO17*, 335 pp.

TR022 – Jones, P.D., Raper, S.C.B., Santer, B.D., Cherry, B.S.G., Goodess, C.M., Kelly, P.M., Wigley, T.M.L., Bradley, R.S. and Diaz, H.F., 1985: A Grid Point Surface Air Temperature Data Set for the Northern Hemisphere, U.S. Dept. of Energy, Carbon Dioxide Research Division, *Technical Report TRO22*, 251 pp.

TR027 – Jones, P.D., Raper, S.C.B., Cherry, B.S.G., Goodess, C.M. and Wigley, T.M.L., 1986: A Grid Point Surface Air Temperature Data Set for the Southern Hemisphere, 1851-1984, U.S. Dept. of Energy, Carbon Dioxide Research Division, *Technical Report TR027*, 73 pp.

Jones, P.D., 1994: Hemispheric surface air temperature variations: a reanalysis and an update to 1993. *Journal of Climate* **7**, 1794 1802.

Jones, P.D. and Moberg, A., 2003: Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *J. Climate* **16**, 206-223.

Brohan, P., Kennedy, J., Harris, I., Tett, S.F.B. and Jones, P.D., 2006: Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. *J. Geophys. Res.* **111**, D12106, doi:10.1029/2005JD006548.

(b) How MetOffice-HadleyCentre (MOHC) updates their copy of the station database (this is the operational copy of the station database)

The MOHC version of the station database is updated using monthly mean temperatures transmitted over the CLIMAT database. The complete set of CLIMAT records is also made available at the following site http://hadobs.metoffice.com/crutem3/data/station_updates/. Only stations that are already in the CRUTEM database are updated. Updating is undertaken according to the WMO Station Identifier. Additionally some late data comes in for back months via the CLIMAT system. These late data are also incorporated. All CLIMAT data are QC'd before they are put on the web site.

(c) How CRU updates our copy of the station database (this is the developmental copy of the station database, and will not be used operationally until CRUTEM4 is published)

CRU takes the CLIMAT messages from the station database from the MOHC website. The CRU version of the station database is updated in a similar fashion. Additionally CRU adds in data for 42 Canadian stations that are now arriving with different station identifiers (occasionally countries alter their WMO identifiers and these changes are incorporated with every major update). The changes across Canada were extensive so were incorporated to ensure coverage across the country remained good. When this update is made, raw data for many stations for Canada for Quebec eastwards are also modified, so that they are compatible with the historic record for these sites.

CRU also downloads the Monthly Climatic Data for the World (MCDW) record of monthly average temperatures from an ftp at NCDC, Asheville. These data are also included in the update each month, but each MCDW month is between 4 to 8 months behind the latest month. MCDW also includes back data and these are added manually from each MCDW publication. The temperature value for the South Pole is added manually, as it isn't sent in the current CLIMAT format.

Occasionally the British Antarctic Survey (BAS) READER database (http://www.antarctica.ac.uk/met/gjma/) is accessed to improve the amount of data coming in for the Antarctic. This is not undertaken regularly, but roughly once a year.

The result of these additional efforts at CRU means that the MOHC operational and CRU developmental versions of the station database differ. Differences are mostly in Canada, but the back data additions can mean that differences can occur anywhere.

In addition to this semi-regular updating by CRU, David Lister makes occasional alterations to the database when new or replacement data become available – e.g. if some National Met Service(s) have undertaken a homogeneity or digitisation exercise, so that altered or new data are provided.

(d) and (e) Stages in the construction of CRUTEM3 gridded, global and hemispheric temperature anomalies from the CRUTEM3 station temperature database

This outlines the stages in the construction of CRUTEM3 gridded temperature anomalies from the CRUTEM3 station temperature database. These steps have been identified from the published articles describing the construction of the CRUTEM3 datasets (and its previous versions). The page numbers from which the information compiled here was taken is indicated for each of these articles, where **CRUTEM1986NH** is Jones *et al.* (1986a); **CRUTEM1986SH** is Jones *et al.* (1986b); **CRUTEM1** is Jones (1994); **CRUTEM2** is Jones and Moberg (2003); and **CRUTEM3** is Brohan *et al.* (2006).

This note does *not* outline how the station temperature database was assembled (i.e., sources of data and what precedence is given when multiple sources are available), which is described elsewhere (e.g., US DoE TR017, 1985; Jones and Moberg, 2003). It also does not describe how the homogeneity adjustments were determined and applied (to around 10% of the station records), which are instead described in (US DoE TR022, 1985; US DoE TR027, 1986; Jones *et al.*, 1986a,b).

Some definitions:

Dimensions

$t = 1 \ldots n_t$; year index

$m = 1 \ldots 12$; month index (1=Jan, 12=Dec)

$i = 1 \ldots n_i$; grid box index in longitude direction, for 5 degree grid $n_i = 72$

$j = 1 \ldots n_j$; grid box index in latitude direction, for 5 degree grid $n_j = 36$

$s = 1 \ldots n_s$; weather station index

Time variables

yr_t ; year value, currently running from yr_1 = 1850 to yr_{nt} = 2009

Gridded variables

x_i ; [1]longitude of centre of grid box *i*, x_1 = -177.5°, x_2 = -172.5°, x_3 = -167.5°, etc., up to x_{ni} = 177.5°
; $x_i < 0$ are west of Greenwich meridian
; $x_i > 0$ are east of Greenwich meridian
; longitude of western edge of grid box *i* is $x_{i\text{-}0.5}$ (i.e. 2.5°W of the centre)
; longitude of eastern edge of grid box *i* is $x_{i+0.5}$ (i.e. 2.5°E of the centre)

y_i ; latitude of centre of grid box *j*, y_1 = -87.5°, y_2 = -82.5°, y_3 = -77.5°, etc., up to y_{nj} = 87.5
; $y_j < 0$ are south of equator
; $y_j > 0$ are north of equator
; latitude of southern edge of grid box *j* is $y_{j\text{-}0.5}$ (i.e. 2.5°S of the centre)
; latitude of northern edge of grid box *j* is $y_{j+0.5}$ (i.e. 2.5°N of the centre)

$G'_{i,j,t,m}$; grid-box temperature anomaly (°C) for grid box *i,j* in year *t* and month *m*

$\Delta_{i,j,t,m}$; a “mask” to indicate which grid-box temperature anomalies are missing and which are available
; $\Delta_{i,j,t,m} = 0$ means that there is no value in year *t* and month *m* for grid box *i,j*
; $\Delta_{i,j,t,m} = 1$ means that there is a value in year *t* and month *m* for grid box *i,j*

Station data

lon_s ; [2]longitude of weather station *s*

lat_s ; latitude of weather station *s*

$T_{s,t,m}$; “raw” monthly-mean station temperature observations (°C) for station *s* in year *t* and month *m*
; note 1 that “raw” means “as received by CRU” – there will have been previous processing (not least the calculation of the monthly means from the daily or sub-daily measurement values, done before we receive the data)

---

[1] Note that the conventions for the ordering of grid boxes (west to east, south to north) are consistent throughout this note, but that these algorithms can be implemented using other conventions provided that they remain consistent throughout. Indeed, some versions of the CRUTEM programs begin the grid at the North Pole rather than the South Pole, or begin the grid at the Greenwich Meridian rather than at the International Dateline. A particular aspect to be aware of is that the longitudes of the weather stations stored in the station data files are recorded under the convention that positive values are *west* of the Greenwich Meridian, not east (which are recorded with negative values). The programs that implement the algorithms described here take this into account.

[2] The convention used within this note is that the longitudes of the weather stations follow the same sign convention as the longitudes of the grid boxes, but note that the station data files currently follow the opposite sign convention and the programs that implement the algorithms described here take this into account (see footnote 1).

; note 2 that they may be received in different units, but this description assumes the units have been converted to °C

$\delta_{s,t,m}$; a "mask" to indicate which station temperature observations are missing (or should not be used) and which are available (and should be used). Some values are available but are considered unreliable or fail the outlier check – for these values, this "mask" is set to zero to prevent them being used.
; $\delta_{s,t,m}$ = 0 means that there is no observational value in year *t* and month *m* for station *s*
; $\delta_{s,t,m}$ = 1 means that there is an observational value in year *t* and month *m* for station *s*

Calculation of "normals" and standard deviations:

[CRUTEM1986NH p167; CRUTEM1986SH p1217; CRUTEM1 p1795; CRUTEM2 p212-213; CRUTEM3 p2,5]

A "normal" is the name used in climatology for the mean value over a reference period (also known as the "base" period or the "normal" period). For each station *s* and each month of the year *m*, the number of values within the reference period is determined:

$$N_{s,m} = \sum_{t=REF1}^{REF2} \delta_{s,t,m} \qquad \text{Equation (1)}$$

where *REF*1 and *REF*2 define the reference period (currently we use yr_{REF1} = 1961 and yr_{REF2} = 1990 **[1951–1970 for CRUTEM1986NH p167; 1951–1970 for CRUTEM1986SH p1217; 1961–1990 for CRUTEM1 p1795; 1961–1990 for CRUTEM2 p212; 1961–1990 for CRUTEM3 p2]**). If there are sufficient values to obtain a reasonable estimate of the mean value (currently we require $N_{s,m}$ >= 15, though we have previously used different criteria, including ones that require a minimum number of values during each decade of the reference period **[>= 15 for CRUTEM1986NH p167; >= 15 for CRUTEM1986SH p1217; >= 21 for CRUTEM1 p1795; >= 20 with >= 4 in each decade for CRUTEM2 p213; >= 15 for CRUTEM3 p2]**) then we estimate the mean value by:

$$\overline{T}_{s,m} = \frac{\sum_{t=REF1}^{REF2} \delta_{s,t,m} T_{s,t,m}}{N_{s,m}} \qquad \text{Equation (2)}$$

Note that we take into account the uncertainty (sampling error) associated with $\overline{T}_{s,m}$, including the greater error when $N_{s,m}$ < 30 (i.e. incomplete data), in the estimation of CRUTEM3 uncertainty ranges (as described in Brohan *et al.*, 2006).

If we have insufficient values to determine a normal (i.e., $N_{s,m}$ < 15, or the different criteria noted above for the earlier versions), then we follow one of three options (in this order):

(i) If a neighbouring station does have sufficient values to determine its normal, then we calculate the mean difference between the temperatures recorded at this neighbouring station and the temperature recorded at the current station over a different period when they both have data (e.g., 1951–1970), and assume that this mean difference still holds during the reference period. The normal for the current station is then calculated as the sum of the normal for the neighbouring station plus the mean difference between the two station's temperatures.

(ii) If the WMO have published a normal for the station and the reference period (perhaps because the National Meteorological Service had calculated it from additional data not available to us), then we use that.

(iii) Otherwise we omit all temperatures recorded at this station for all months of the year, and set all $\delta_{s,t,m} = 0$.

[estimated from neighbours for CRUTEM1986NH p167; estimated from neighbours for CRUTEM1986SH p1217; estimated from gridded data or from neighbours for CRUTEM1 p1795; used WMO normals or estimated from gridded data or from neighbours for CRUTEM2 p213; used WMO normals or estimated from gridded data for CRUTEM3 p2]

The calculation of the station standard deviations is quite similar, though a longer reference period is used because (i) the sampling error associated with standard deviations tends to be relatively greater than that associated with means, and thus a bigger sample is preferred; and (ii) we do not use other sources (e.g. WMO) in cases where we have insufficient data. For each station *s* and each month of the year *m*, the temperature standard deviation is calculated as:

$$\sigma_{s,m} = \sqrt{\frac{\sum_{t=REF3}^{REF4} \delta_{s,t,m} \left(T_{s,t,m} - \overline{T}_{s,m}\right)^2}{\left(\sum_{t=REF3}^{REF4} \delta_{s,t,m}\right) - 1}} \qquad \text{Equation (3)}$$

where $\overline{T}_{s,m}$ is the mean computed over the longer reference period (currently we use yr_{REF3} = 1941 and yr_{REF4} = 1990 **[1941–1990 for CRUTEM1 p1795; 1921–1990 for CRUTEM2 p213]**). Note that for all those stations that are actually used (i.e., that we could obtain a "normal" for, following the steps outline above), this longer reference period yielded at least 15 values from which the standard deviation could be estimated.

Conversion to anomalies (i.e. deviations from a reference value):

Given a "normal" for station *s* and month *m*, $\overline{T}_{s,m}$, all observations are converted to anomalies according to:

where $\delta_{s,t,m} = 1$, $$T'_{s,t,m} = T_{s,t,m} - \overline{T}_{s,m} \qquad \text{Equation (4)}$$

Removal of outliers (quality control):

Given an estimate of the standard deviation of monthly temperature anomalies for station *s* and month *m*, $\sigma_{s,m}$, any anomalies exceeding 5 standard deviations **[>6 for CRUTEM1 p1795; >5 for CRUTEM2 p213; >5 for CRUTEM2 p2]** are removed:

where $\left|T'_{s,t,m}\right| > 5\sigma_{s,m}$ set $\delta_{s,t,m} = 0$ Equation (5)

In fact, this outlier check resulted in some cases where an obvious error could be corrected (e.g., the measurement value was recorded or digitised as ten times too small or too large); in these cases the station temperature files have been corrected. After removal or correction of outliers, the mean (i.e. normal) and standard deviation of the monthly temperature anomalies is recalculated and the outlier check is repeated **[CRUTEM1 p1795-1796; CRUTEM2 p213]**.

Gridding of station anomalies:

This is done separately for each year *t* and each month *m*, and for each grid box *i,j*. The grid-box mean anomaly is the average all the available values **[a different method was used for CRUTEM1986NH p167 and CRUTEM1986SH p1216; the method described here was used for CRUTEM1 p1796-1796; CRUTEM2 p213 and CRUTEM3 p2]**, i.e.:

for all *s* where $x_{i-0.5} <= lon_s < x_{i+0.5}$

and $y_{j-0.5} <= lat_s < y_{j+0.5}$

$$G'_{i,j,t,m} = \frac{\sum_s \delta_{s,t,m} T'_{s,t,m}}{\sum_s \delta_{s,t,n}} \quad \text{and} \quad \Delta_{i,j,t,m} = 1 \qquad \text{Equation (6)}$$

If no *s* match the above criteria, then $\Delta_{i,j,t,m} = 0$

Calculation of hemisphere and global means:

[CRUTEM1986NH p167; CRUTEM1986SH p1217; CRUTEM2 p216; CRUTEM3 p10]

The hemispheric means are averages of the grid boxes for which temperature anomalies are available (this is, of course, time dependent), with weighting according to the area of the grid boxes (proportional to the cosine of their central latitude).

$$NH_{t,m} = \frac{\sum_{j=1+nj/2}^{nj} \sum_{i=1}^{ni} \cos(y_j) \Delta_{i,j,t,m} G'_{i,j,t,m}}{\sum_{j=1+nj/2}^{nj} \sum_{i=1}^{ni} \cos(y_j) \Delta_{i,j,t,m}} \qquad \text{Equation (7)}$$

$$SH_{t,m} = \frac{\sum_{j=1}^{nj/2} \sum_{i=1}^{ni} \cos(y_j) \Delta_{i,j,t,m} G'_{i,j,t,m}}{\sum_{j=1}^{nj/2} \sum_{i=1}^{ni} \cos(y_j) \Delta_{i,j,t,m}} \qquad \text{Equation (8)}$$

The global mean can be calculated in a similar way, or as the simple arithmetic mean of the two hemisphere means.

When calculating annual-mean hemisphere-means, or annual-mean global-means, there are two options available: (i) calculate annual-mean anomalies for each grid box first, and then calculate hemispheric or global means of those; or (ii) calculate the hemispheric or global means of the monthly anomalies, and then calculate the annual means of those monthly hemispheric or global means. The Met Office Hadley Centre uses method (i), while CRU uses method (ii).

There are various advantages and disadvantages to the choices in calculating global means and in calculating annual and hemispheric/global means which we do not list here; we just note that small differences in the results can arise and therefore need to be explained.

Adjusting the high-frequency temporal variance:

We also create a "variance adjusted" version of CRUTEM3 (called CRUTEM3v, with an equivalent land and marine version HadCRUT3v) because the variance of the grid-box mean anomalies, $G'_{i,j,t,m}$, varies according to how many stations match the equation (6) criteria. This can cause artificial changes in the variance of $G'_{i,j,t,m}$ between different grid boxes, and also between different time periods for the same grid box when the number of stations varies through time. These artificial changes in the variance can prevent the use of CRUTEM3 for monitoring changes in variability and in the occurrence of extreme values. The "variance adjusted" version, CRUTEM3v, is obtained after the construction of CRUTEM3 is complete by subsequently following these steps:

(i) apply a filter to separate the grid box temperature anomaly time series into "low" and "high" frequency components;

(ii) scale the "high" frequency component with a time-varying scaling factor described in Osborn *et al.* (1997) and Jones *et al.* (2001) to remove the expected artificial time-variations in the variance of the grid-box temperature anomalies;

(iii) combine the adjusted "high" frequency component with the original "low" frequency component to obtained the CRUTEM3v grid-box mean temperature anomaly time series.

Further details of this process are given in Jones *et al.* (2001) and Appendix A of Brohan *et al.* (2006).

References:

Brohan P., Kennedy J.J., Harris I., Tett S.F.B. and Jones P.D., 2006: Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. *Journal of Geophysical Research* **111**, D12106, doi:10.1029/2005JD006548.

Jones P.D., 1994: Hemispheric surface air temperature variations: a reanalysis and an update to 1993. *Journal of Climate* **7**, 1794-1802.

Jones P.D. and Moberg A., 2003: Hemispheric and large-scale surface air temperature variations: an extensive revision and an update to 2001. *Journal of Climate* **16**, 206-223.

Jones P.D., Raper S.C.B., Bradley R.S., Diaz H.F., Kelly P.M. and Wigley T.M.L., 1986a: Northern Hemisphere surface air temperature variations: 1851-1984. *Journal of Climate and Applied Meteorology* **25**, 161-179.

Jones P.D., Raper S.C.B. and Wigley T.M.L., 1986b: Southern Hemisphere surface air temperature variations: 1851-1984. *Journal of Climate and Applied Meteorology* **25**, 1213-1230.

Jones P.D., Osborn T.J., Briffa K.R., Folland C.K., Horton B., Alexander L.V., Parker D.E. and Rayner N.A., 2001: Adjusting for sampling density in grid-box land and ocean surface temperature time series. *Journal of Geophysical Reseach* **106**, 3371-3380.

Osborn T.J., Briffa K.R. and Jones P.D., 1997: Adjusting variance for sample-size in tree-ring chronologies and other regional-mean timeseries. *Dendrochronologia* **15**, 89-99.

US DoE TR017 – Bradley R.S., Kelly P.M., Jones P.D., Goodess C.M. and Diaz H.F., 1985: A climatic data bank for Northern Hemisphere land areas, 1851-1980, U.S. Dept. of Energy, Carbon Dioxide Research Division, *Technical Report TR017*, 335pp.
Available at http://www.cru.uea.ac.uk/st/

US DoE TR022 – Jones P.D., Raper S.C.B., Santer B.D., Cherry B.S.G., Goodess C.M., Kelly P.M., Wigley T.M.L., Bradley R.S. and Diaz H.F., 1985: A grid point surface air temperature data set for the Northern Hemisphere, U.S. Dept. of Energy, Carbon Dioxide Research Division, *Technical Report TRO22*, 251pp.
Available at http://www.cru.uea.ac.uk/st/

US DoE TR027 – Jones P.D., Raper S.C.B., Cherry B.S.G., Goodess C.M. and Wigley T.M.L., 1986: A grid point surface air temperature data set for the Southern Hemisphere, 1851-1984, U.S. Dept. of Energy, Carbon Dioxide Research Division, *Technical Report TR027*, 73pp.
Available at http://www.cru.uea.ac.uk/st/

Appendix 2

CRU TS 3.0 scientific and data workflow description

Figure 1 illustrates the updating procedure using CLIMAT, MCDW and Australian data in near-real time. Each variable is updated in sequence. MCDW is added first, then CLIMAT and finally the Australian BoM data.

Figure 2 shows the procedure for Gridding the station data to the regular latitude/longitude grid.

Figure 3 shows the station counts update process.

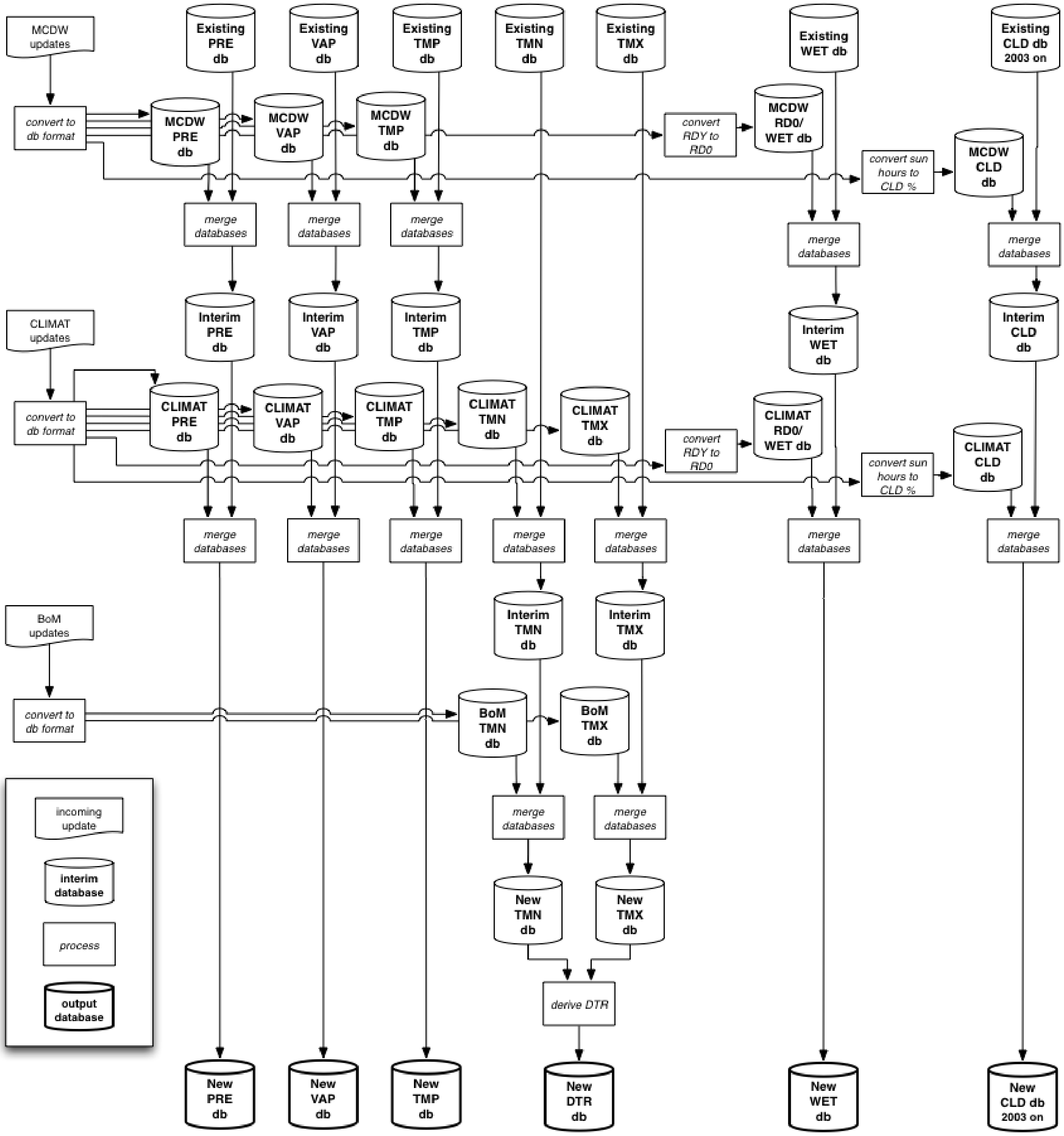


Figure 1: Database update flowchart

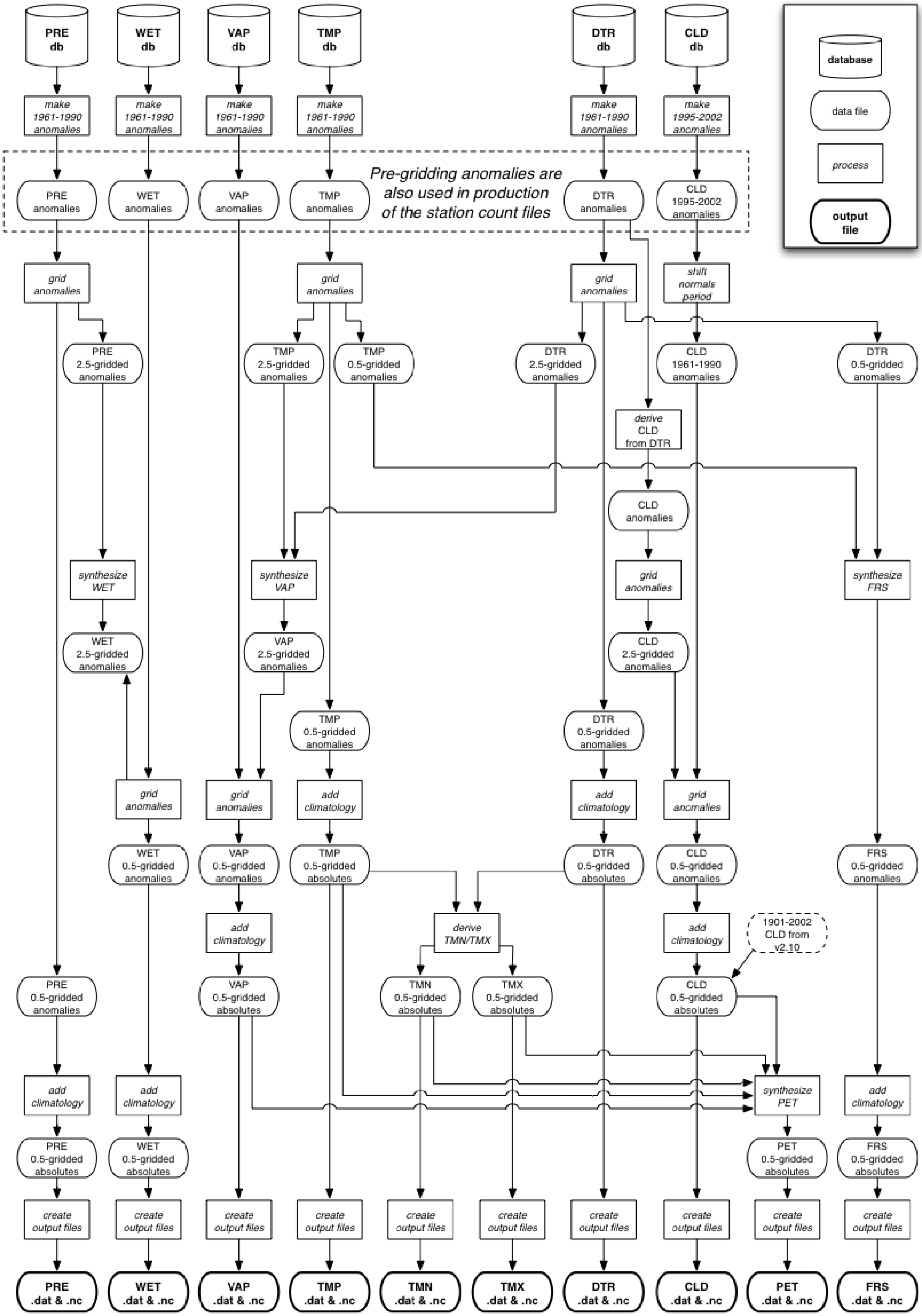


Figure 2: Dataset Update (gridding) Process

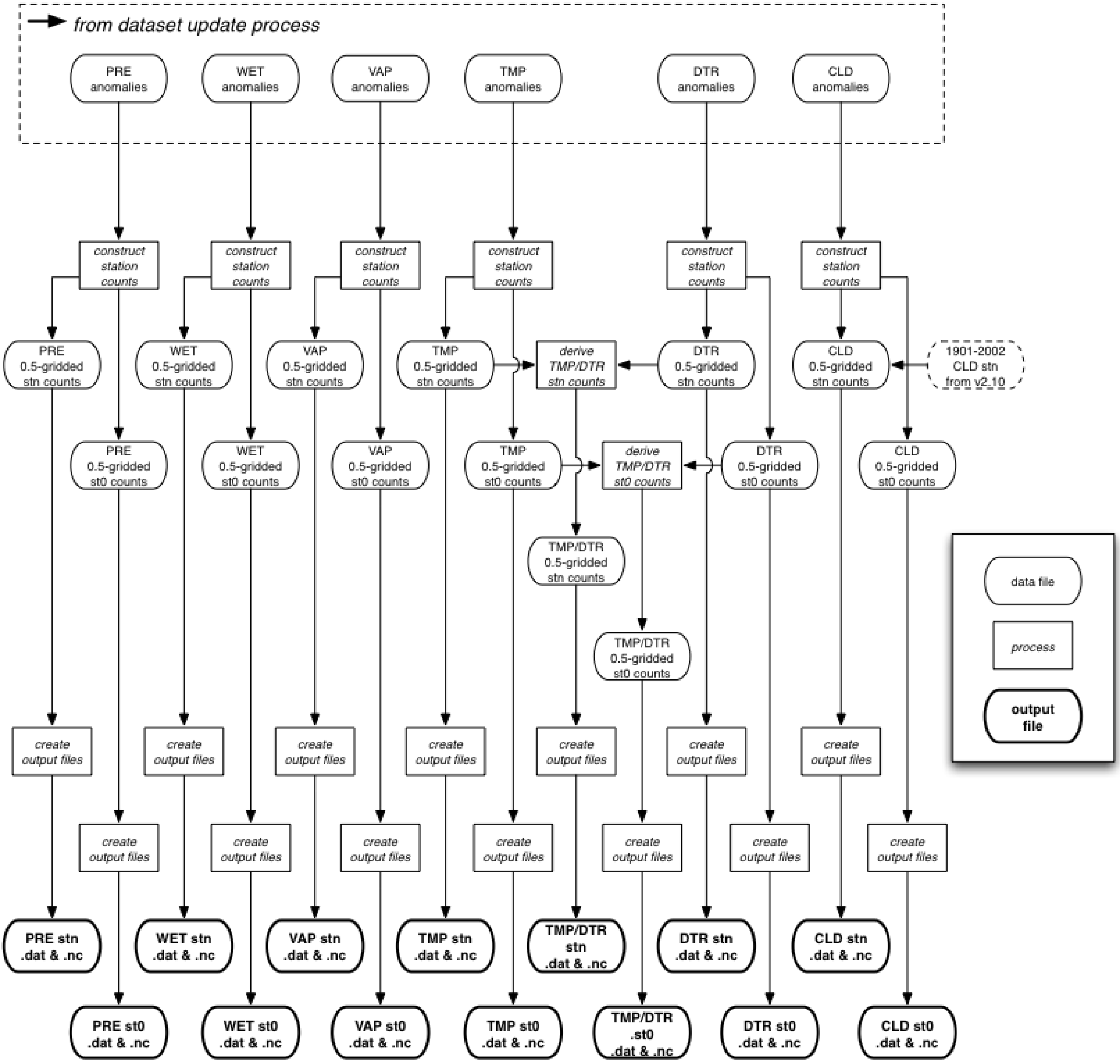


Figure 3: Station Counts Update (gridding) Process

Appendix 3

Background information about the overall workflow associated with the development of tree-ring chronologies

This text is extracted from CRU submission to Muir Russell which may be found at:
http://www.cce-review.org/evidence/Climatic_Research_Unit.pdf

In common with prior research practice in dendroclimatology, CRU has adopted an empirical approach to the interpretation of tree-ring-derived data as evidence for past climate variability. General texts describing tree-ring methods include Fritts (1976), Cook and Kairiukstis (1990) and Briffa (1995). Axiomatic in this approach is a clear separation between the processes of constructing a tree-ring chronology or chronologies (i.e. timeseries of averaged tree-growth records), as distinct from making inferences about the variability of some specific climate parameter based on the evidence provided by the chronology data. These inferences are dependent on the nature and strength of the statistically-assessed association between tree growth and climate, established through a comparison of overlapping tree-ring chronology data and instrumental climate records. We now describe CRU work with respect to these two categories.

Chronology Construction

Our intention when building a tree-ring chronology is to try to produce a statistically robust, year-by-year record of changing net tree productivity, representative of either a local site or a wider geographical region, depending on the scale and required spatial resolution of the study in question.

Where the ultimate aim is to reconstruct some aspect of changing temperature conditions, tree-ring sampling will be focussed on areas and sites where prior knowledge of the ecological and physiological controls on tree growth suggest potential for a strong temperature limitation of annual growth, i.e. in areas where trees are growing near high-latitude or high-elevation tree lines.

In the field, the usual procedure is to take one or more radial core samples, within trees and between trees to provide a representative sample of yearly growth across the local tree population. Sampling multiple trees provides the replication of data required to allow rigorous cross-comparisons of the high-frequency variability, in all series of measurements of the relevant extracted parameters (e.g. total ring width or maximum-latewood density – MXD). This systematic inter-series comparison is routine practice in dendroclimatology, used to ensure accuracy in the dating of all ring values, known as "crossdating". It establishes an absolute calendrical time frame for all data (Wigley et al. 1987).

By virtue of the accuracy of sample dates, multiple series can be averaged together to provide an annually-resolved chronology. The underlying, common pattern of tree growth variability through time is increasingly better expressed because random deviations, apparent in different trees across the site, cancel in proportion to the number of data series incorporated in the chronology (Wigley et al. 1984, Briffa and Jones 1990).

However, before averaging them, the 'raw' tree-ring measurement series are "standardised". Effectively this involves a form of statistical detrending to remove possible biasing effects of naturally reducing width and density apparent in measurements from the inner to outer rings of a tree. These growth trends, related to a tree's age and size, are a consequence of the geometric, mechanical and physiological constraints on the processes that allocate new material to radial stem production.

In our early work (during the 1980s and 1990s), in common with other researchers, we frequently used data-adaptive (essentially curve-fitting) standardisation methods to produce dimensionless series of tree indices, with zero slope over the lifespan of an individual tree. This type of detrending removes evidence of climate variability on long timescales in chronologies produced using this approach (i.e. typically losing information on timescales longer than the average age of individual trees).

However, because the detrending algorithm is also unable to distinguish between local variance associated with climate change and variance associated with age-related bias, some additional evidence of climate variability on shorter timescales (on the order of decades) may also be removed, or the local trend distorted, particularly near the recent (i.e. modern) end of chronologies (Briffa 1995, Cook et al. 1995, Melvin and Briffa 2008, Briffa and Melvin 2010).

To mitigate the time-scale limitation that arises in traditional forms of standardisation, we proposed the use of a technique that had been used in some early forestry studies but had not been widely used in dendroclimatology. Now generally termed Regional Curve Standardisation (RCS), this method is based on constructing a single, 'best-estimate' statistical model of the expected change in ring width or density as a function of tree age, for all trees in a region (Briffa et al. 1992a). By expressing the measured ring-growth for each tree in the form of deviations from this expectation, the age biasing effect is removed from the raw data, but because this model is not fit directly to each measurement series, the resulting mean index series for each tree is not constrained to be 1.0 nor is its slope constrained to be zero as in earlier, curve-fitting methods. This means that chronologies produced using the RCS approach are able to preserve the evidence of long-timescale climate change. However, we recognise that the RCS approach is generally only applicable if the tree samples span a very long period of time and are well distributed, overlapping each other, through time (Briffa et al. 1992a, Briffa et al. 1996, Melvin 2004, Briffa and Melvin 2010). If the sample trees have largely parallel life spans, the standardisation model itself will be biased by common climate forcing influencing the expected growth value at particular ages where the climate exerts a common influence on all trees. To date, this problem has largely restricted the use of RCS to the development of long chronologies spanning many hundreds or preferably thousands of years, in those few locations where old sub-fossil tree remnants exist.

We have recently developed a "signal-free" technique of standardisation, in which common chronology variance is removed from each of the sample measurement series in an iterative separation of the potential climate and non-climate variance. This approach overcomes the local distortion problem but if applied to the traditional (i.e. curve fitting or data adaptive) methods we refer to above, it still cannot overcome the loss of long-timescale climate information. We believe that using the 'signal-free' approach in conjunction with the RCS, should allow us to mitigate bias effects in the RCS and enable its use to standardise shorter, living-tree chronologies (Melvin 2004, Melvin and Briffa 2008, Briffa and Melvin 2010).

It is important to understand that the construction of tree-ring chronologies is achieved entirely independently of any knowledge about the behaviour of climate parameters that we aim to reconstruct. Within the obvious constraints of the availability of samples, the decision on which, or how many, tree data series to incorporate in a chronology is made solely on the basis of analysis of the common tree growth signal, with the intention of achieving an optimum and consistent expression of this signal throughout the length of the chronology. Depending on the focus of the study, a chronology may be processed to emphasise some particular timescale of information or alternatively to investigate information in some predefined region, but at no time have the constituent tree-ring data been chosen by CRU on the basis of comparison with climate data. The interpretation of the climate signal contained in a chronology is a subsequent and separate process.

Inferring past temperature variability from tree-ring chronologies

Whatever the expectation of climate control over tree-ring growth, it is common practice to demonstrate the actual character and strength of the relationship between tree-ring chronology and climate variables and to formalise this relationship in a way that allows past estimates of the tree-ring variability to be interpreted as evidence of past climate change. This is most often achieved using linear regression analysis. Establishing the nature of the climate influence on tree growth involves regression where the dependent (tree-growth) series is expressed as a function of the variability in an ensemble of climatic series (e.g. a number of monthly mean temperature and precipitation series). This is known as calculating a 'response function': a set of regression coefficients whose signs and magnitudes suggest the character of seasonal climate influence on tree growth and indicate the climate variable (e.g. mean June-August temperature) that might be optimally reconstructed using these tree-ring data (Briffa and Cook 1990).

A 'transfer-function' is subsequently developed. Again this is a set of regression coefficients but, in this case, one which when multiplied by past tree-growth variability provides estimates of past climate (Briffa et al. 1983, Cook et al. 1994, Osborn and Briffa 2000).

In practice, depending on the nature of the problem and the availability of data, a range of possible regression scaling techniques might be used to produce climate reconstructions: from simple least-squares linear regression, involving one predictand and a single predictor, up to the most complex spatial problem where multiple predictand series (such as in a network or grid of temperature data) are estimated simultaneously using a set of multiple predictors. To reduce the effective number of predictors in such regressions, some form of principal components regression may be employed, such as Orthogonal Spatial Regression (Briffa et al. 1986, Cook et al. 1994).

Where multiple regression is used to derive a transfer function (i.e. 'calibrate' the reconstruction equation) there is always potential for 'over fitting' in estimating the regression coefficients, particularly where the ratio of dependent to independent data is low. To assess this possibility it is common practice to 'verify' the regression equation by using it to produce estimates of climate data that are independent of those used in the calibration procedure. Statistical metrics of the goodness of fit between actual and estimated data can provide a more reliable indication of likely reconstruction fidelity than might be assumed on the basis of the calibration-period comparison alone (Briffa et al. 1986, Cook et al. 1994).

Having identified and assessed the strength of statistical association represented by the regression equations, past tree-ring data are used to produce estimates of past climate over the longer period represented by the tree-ring chronology. Importantly, these estimates can be presented along with information about their statistical confidence. This is ultimately a function of both the statistical quality of the chronology data and the strength of the match with the particular climate variable being reconstructed. It is important to note that both of these may, and should ideally, be expressed in a way which makes clear their time-scale and time dependence (Briffa et al. 2001, Briffa et al. 2002a, Jones et al. 2009).

The large majority of data used in CRU work aimed at reconstructing different aspects of temperature history were acquired directly from international collaborators: the most significant being ring-width and wood-densiometric measurements from a network of living-tree sites circling the Northern Hemisphere produced by Fritz Schweingruber and colleagues at the Swiss Federal Institute for Forest, Snow and Landscape Research in Birmendsdorf, Switzerland (WSL). These data provided the basis for CRU work in reconstructing detailed, spatially-explicit patterns of summer temperature variability, expressed on regional grids covering much of North America and Europe, and extending over three to six centuries (Briffa et al. 1986, Briffa et al. 1988, 1992b) and a wider expanse of the northern hemisphere (Briffa et al. 2002b). Reconstructions, using these data, have been made of regional-average temperatures across sub-continental areas of high-latitude or high-elevation regions of the Northern Hemisphere. These have also been aggregated and scaled to represent a pseudo-hemispheric average reconstruction running from 1402 to 1994 (Briffa et al. 2001, Briffa et al. 2002a).

WSL also provided longer densiometric data sets extracted from both living trees and remnant (sub-fossil) wood in both Sweden and the Polar Ural mountains in Russia that CRU used to produce millennial-length temperature reconstructions in these areas. Stepan Shiyatov and Rashit Hantemirov (Ural Division of the Russian Academy of Science, Ekaterinburg) and Eugene Vaganov and Muchtar Naurzbaev (Siberian Federal University, Krasnoyarsk) made available tree-ring width data from the Yamal and Taimyr Peninsular regions of Siberia; Matti Eronen and colleagues (University of Helsinki) provided data from northern Finland; and Wibjörn Karlén and Håken Grudd (Stockholm University) provided data from northern Sweden.

In collaboration with all of the above colleagues, CRU has also produced longer, multi-millennial chronologies and inferred histories of regional summer temperature changes at discrete locations stretching from west to east across much of northern Eurasia (for details and references see Table 1.2, Section 1.6 reproduced below).

Table 1.2: A summary of CRU-related publications that describe the production of tree-ring chronologies or temperature reconstructions that extend back to medieval times. (a) lists work concerned with tree-based data and (b) lists publications exploring 'large-scale' temperature changes based on the evidence of multiple proxies.

(a)

Location	**Period**	**Variable**[1]	**Standardisation**[2]	**Scale**	**Scale Reference**
N. Sweden	443–1981	TRW, MXD	CF	Local	Briffa et al. (1990)
N. Sweden	443–1981	TRW, MXD	CF & RCS	Local	Briffa et al. (1992a)
N. Sweden	443–1981	TRW, MXD	CF & RCS	Local	Briffa et al. (1996)
Sweden & Finland	1–1997	TRW	RCS	Region	Briffa et al. (2008)
Polar Urals	914–1990	TRW, MXD	CF & RCS	Local	Briffa et al. (1995)
Polar Urals	914–1990	TRW, MXD	CF & RCS	Local	Briffa et al. (1996)
Yamal	1–1996	TRW	RCS	Local	Briffa (2000)
Yamal	1–1996	TRW	RCS	Local	Briffa et al. (2008)
Taimyr	1–2003	TRW	RCS	Region	Briffa et al. (2008)
Arctic	1–1993	TRW	RCS	Region	Briffa (2000)
Eurasian Arctic	1–2003	TRW	RCS	Region	Briffa et al. (2008)

[1]TRW = Tree Ring Width; MXD = Maximum Latewood Density
[2]CF = Curve Fitting standardisation; RCS = Regional Curve Standardisation

(b)

Region	Period	Season	Resolution[1]	Proxies[2]	Reference
NH	1000–1991	Summer	1-year	5xT 5xN 0xM	Jones et al. (1998)
Eurasian Arctic	1–1993	Summer	1-year	3xT 0xN 0xM	Briffa and Osborn (1999), Briffa et al. (2004)
NH, SH, Globe	200–1980	Annual	10-year	3xT 4xN 1xM	Mann and Jones (2003)
NH	800–1995	Annual	20-year	10xT 3xN 1xM	Osborn and Briffa (2006)
NH	1000–1980	Annual	1-year	6xT 6xN 1xM	Juckes et al. (2007)
Arctic	5–1995	Summer	10-year	4xT 19xN 0xM	Kaufman et al. (2009)

[1]This is the nominal time resolution of reconstruction; some individual proxy records may have higher or lower resolution
[2]Number of tree-ring (T), non-tree-ring (N) and mixed (M) proxy records used in the reconstruction; some records are themselves composites of multiple underlying proxy records; in most cases these records do not all span the reconstruction period and there are fewer records in the earliest periods of some reconstructions.

References

Briffa, K. R. 1995. Interpreting high resolution proxy climate data - The example of dendroclimatology. Pages 77-94 *in* H. von Storch and A. Navarra, editors. Analysis of Climate Variability: Applications of Statistical Techniques.

Briffa, K. R. 2000. Annual climate variability in the Holocene: interpreting the message of ancient trees. Quaternary Science Reviews **19**:87-105.

Briffa, K. R., and E. R. Cook. 1990. Methods of Response Function Analysis. Pages 240-246 *in* E. R. Cook and A. Kairiukstis, editors. Methods of Dendrochronology. Institute for Applied Systems Analysis, Kluwer Academic Publishers, Dordrecht.

Briffa, K. R., and P. D. Jones. 1990. Basic Chronology Statistics and Assessment. Pages 137-152 *in* E. R. Cook and L. A. Kairiukstis, editors. Methods of Dendrochronology. Kluwer Academic Publishers.

Briffa, K. R., P. D. Jones, T. S. Bartholin, D. Eckstein, F. H. Schweingruber, W. Karlén, P. Zetterberg, and M. Eronen. 1992a. Fennoscandian Summers from AD 500: temperature changes on short and long timescales. Climate Dynamics **7**:111-119.

Briffa, K. R., P. D. Jones, and F. H. Schweingruber. 1988. Summer Temperature Patterns over Europe - a Reconstruction from 1750-Ad Based on Maximum Latewood Density Indexes of Conifers. Quaternary Research **30**:36-52.

Briffa, K. R., P. D. Jones, and F. H. Schweingruber. 1992b. Tree-Ring Density Reconstructions of Summer Temperature Patterns across Western North-America since 1600. Journal of Climate **5**:735-754.

Briffa, K. R., P. D. Jones, F. H. Schweingruber, W. Karlén, and S. G. Shiyatov. 1996. Treering variables as proxy-climate indicators: Problems with low frequency signals. Pages 9-41 *in* P. D. Jones, R. S. Bradley, and J. Jouzel, editors. Climatic Variations and Forcing Mechanisms of the Last 2000 Years. Springer-Verlag, Berlin.

Briffa, K. R., P. D. Jones, T. M. L. Wigley, J. R. Pilcher, and M. G. L. Baillie. 1983. Climate Reconstruction from Tree Rings .1. Basic Methodology and Preliminary-Results for England. Journal of Climatology **3**:233-242.

Briffa, K. R., P. D. Jones, T. M. L. Wigley, J. R. Pilcher, and M. G. L. Baillie. 1986. Climate Reconstruction from Tree Rings .2. Spatial Reconstruction of Summer Mean Sea-Level Pressure Patterns over Great-Britain. Journal of Climatology **6**:1-15.

Briffa, K. R., P. D. Jones, F. H. Schweingruber, S. G. Shiyatov, and E. R. Cook. 1995. Unusual 20th-Century Summer Warmth in a 1,000-Year Temperature Record from Siberia. Nature **376**:156-159.

Briffa, K. R., and T. M. Melvin. 2010. A closer look at Regional Curve Standardisation of tree-ring records: justification of the need, a warning of some pitfalls, and suggested improvements in its application. Pages (in press) *in* M. K. Hughes, H. F. Diaz, and T. W. Swetnam, editors. Dendroclimatology: Progress and Prospects. Springer Verlag.

Briffa, K. R., and T. J. Osborn. 1999. Seeing the wood from the trees. Science **284**:926-927.

Briffa, K. R., T. J. Osborn, and F. H. Schweingruber. 2004. Large-scale temperature inferences from tree rings: a review. Global and Planetary Change **40**:11-26.

Briffa, K. R., T. J. Osborn, F. H. Schweingruber, I. C. Harris, P. D. Jones, S. G. Shiyatov, and E. A. Vaganov. 2001. Low-frequency temperature variations from a northern tree ring density network. Journal of Geophysical Research-Atmospheres **106**:2929-2941.

Briffa, K. R., T. J. Osborn, F. H. Schweingruber, P. D. Jones, S. G. Shiyatov, and E. A. Vaganov. 2002a. Tree-ring width and density data around the Northern Hemisphere: Part 1, local and regional climate signals. Holocene **12**:737-757.

Briffa, K. R., T. J. Osborn, F. H. Schweingruber, P. D. Jones, S. G. Shiyatov, and E. A. Vaganov. 2002b. Tree-ring width and density data around the Northern Hemisphere: Part 2, spatio-temporal variability and associated climate patterns. Holocene **12**:759-789.

Briffa, K. R., V. V. Shishov, T. M. Melvin, E. A. Vaganov, H. Grudd, R. M. Hantemirov, M. Eronen, and M. M. Naurzbaev. 2008. Trends in recent temperature and radial tree growth spanning 2000 years across northwest Eurasia. Philosophical Transactions of the Royal Society B-Biological Sciences **363**:2271-2284.

Cook, E. R., K. R. Briffa, and P. D. Jones. 1994. Spatial Regression Methods in Dendroclimatology - a Review and Comparison of 2 Techniques. International Journal of Climatology **14**:379-402.

Cook, E. R., K. R. Briffa, D. M. Meko, D. A. Graybill, and G. Funkhouser. 1995. The Segment Length Curse in Long Tree-Ring Chronology Development for Paleoclimatic Studies. Holocene **5**:229-237.

Cook, E. R., and L. A. Kairiukstis. 1990. Methods of Dendrochronology. International Institute for Applied Systems Analysis, Kluwer Academic Publishers, Dordrecht.

Fritts, H. C. 1976. Tree Rings and Climate. Academic Press, London.

Jones, P. D., K. R. Briffa, T. P. Barnett, and S. F. B. Tett. 1998. High-resolution palaeoclimatic records for the last millennium: interpretation, integration and comparison with General Circulation Model control-run temperatures. Holocene **8**:455-471.

Jones, P. D., K. R. Briffa, T. J. Osborn, J. M. Lough, T. D. van Ommen, B. M. Vinther, J. Luterbacher, E. R. Wahl, F. W. Zwiers, M. E. Mann, G. A. Schmidt, C. M. Ammann, B. M. Buckley, K. M. Cobb, J. Esper, H. Goosse, N.

Graham, E. Jansen, T. Kiefer, C. Kull, M. Kuttel, E. Mosley-Thompson, J. T. Overpeck, N. Riedwyl, M. Schulz, A. W. Tudhope, R. Villalba, H. Wanner, E. Wolff, and E. Xoplaki. 2009. High-resolution palaeoclimatology of the last millennium: a review of current status and future prospects. The Holocene **19**:3-49.

Juckes, M. N., M. R. Allen, K. R. Briffa, J. Esper, G. C. Hegerl, A. Moberg, T. J. Osborn, and S. L. Weber. 2007. Millennial temperature reconstruction intercomparison and evaluation. Climate of the Past **3**:591-609.

Kaufman, D. S., D. P. Schneider, N. P. McKay, C. M. Ammann, R. S. Bradley, K. R. Briffa, G. H. Miller, B. L. Otto-Bliesner, J. T. Overpeck, and B. M. Vinther. 2009. Recent Warming Reverses Long-Term Arctic Cooling. Science **325**:1236-1239.

Mann, M. E., and P. D. Jones. 2003. Global surface temperatures over the past two millennia. Geophysical Research Letters **30**:DOI: 10.1029/2003GL017814

Melvin, T. M. 2004. Historical growth rates and changing climatic sensitivity of boreal conifers. Thesis. (http://www.cru.uea.ac.uk/cru/pubs/thesis/2004-melvin/) University of East Anglia, Norwich.

Melvin, T. M., and K. R. Briffa. 2008. A "Signal-Free" approach to Dendroclimatic Standardisation. Dendrochronologia **26**:71-86.

Osborn, T. J., and K. R. Briffa. 2000. Revisiting timescale-dependent reconstruction of climate from tree-ring chronologies. Dendrochronologia **18**:9-25.

Osborn, T. J., and K. R. Briffa. 2006. The spatial extent of 20th-century warmth in the context of the past 1200 years (vol 311, pg 841, 2006). Science **312**:697-697.

Wigley, T. M. L., K. R. Briffa, and P. D. Jones. 1984. On the average value of correlated time series with applications in dendroclimatology and hydrometeorology. Journal of Climate & Applied Meteorology **23**:201-213.

Wigley, T. M. L., P. D. Jones, and K. R. Briffa. 1987. Cross-Dating Methods in Dendrochronology. Journal of Archaeological Science **14**:51-64.