'gu23wld0098.dat', Version 1.0, March 1999

1. The Dataset

An historical monthly precipitation dataset for global land areas from 1900 to 1998, gridded at 2.5° latitude by 3.75° longitude resolution (a 5° latitude/longitude resolution version is also available: 'g55wld0098.dat'), has been constructed and is available for use in scientific research. This work has been supported by the UK Department of the Environment, Transport and the Regions (Contract EPG 1/1/85). The following credit should be used in reports or publications, etc.:

"'gu23wld0098.dat' (Version 1.0) constructed and supplied by Dr Mike Hulme at the Climatic Research Unit, University of East Anglia, Norwich, UK. This work has been supported by the UK Department of the Environment, Transport and the Regions (Contract EPG 1/1/85)."

The appropriate scientific papers to reference are as follows:

Hulme,M. (1992) A 1951-80 global land precipitation climatology for the evaluation of General Circulation Models **Climate Dynamics**, 7, 57-72

Hulme,M. (1994) Validation of large-scale precipitation fields in General Circulation Models pp.387-406 in, **Global precipitations and climate change** (eds.) Desbois,M. and Desalmand,F., NATO ASI Series, Springer-Verlag, Berlin, 466pp.

Hulme,M., Osborn,T.J. and T.C.Johns (1998) Precipitation sensitivity to global warming: Comparison of observations with HadCM2 simulations **Geophys. Res. Letts**., 25, 3379-3382.

The station dataset from which this gridded dataset has been constructed is an extension of the original CRU/US DoE data described in Eischeid *et al*. (1991). Substantial additional work in extending these station time series and increasing the network has been undertaken by the Climatic Research Unit in recent years. A total of over 11,880 station time series now exist. For access to these station data one should approach Russ Vose working on the Global Historical Climatology Network (GHCN Version 2) at Arizona State University, USA (Email: rvose@smtpl.asu.edu).

2. Obtaining the Gridded Dataset

'gu23wld0098.dat' is about 10.8MB (3.7MB zipped) in size and can be obtained via FTP. The gzipped (binary) version of the file may be found in the following location and unzipped using 'gunzip' under UNIX:

ftp	ftp.cru.uea.ac.uk
login	'anonymous'
password	your email address

cd people/mikehulme/outgoing/griddedprecip/gu23wld0098.gz

A documentation file - 'gu23wld0098_doc.pdf' - should also be copied from the same location as it contains details about the dataset, data format and conditions of use.

Once transferred by FTP please email: m.hulme@uea.ac.uk to say that you have received the dataset. 'gu23wld0098.dat' (or earlier versions of it: gu23wld0094.dat, gu23wld0092.dat, etc.) has been used in over 100 research institutes worldwide.

3. Gridding Method

Thiessen polygon weights were used to average gauge data within each gridbox. Where a monthly station value was missing an estimate was obtained by calculating the mean anomaly for that location derived from surrounding stations. This anomaly interpolation method required the station values to be converted into percentage anomalies from some reference period. These standard anomalies were then interpolated onto the missing station location using an inverse distance (with spherical adjustment), angular weighted method similar to that described in Shepherd (1984) and Legates and Willmott (1990). For this interpolation, a maximum percent anomaly value of 500 per cent was imposed. The interpolated percent anomaly was then converted back into a station mm estimate using that station's mean monthly precipitation total for the reference period. This mean anomaly interpolation was only performed for a missing station value where two stations within a 600km radius possessed valid data (for 1997 and 1998 this search radius was reduced to 400km to minimise instabilities in the resulting gridbox estimates). Otherwise the station value remained missing and hence the gridbox average could not be calculated for that month. A maximum of the 50 nearest stations could contribute to the interpolation.

Two gridded datasets were initially calculated based on two different reference periods: 1931-70 and 1951-90. Stations could only contribute to these climatologies if they possessed 75% or more valid monthly measurements in the reference period. For 1931-70, 5986 stations were used resulting in historical gridded time series for 1277 2.5° by 3.75° gridboxes. The maximum number of stations per gridbox was 51. For 1951-90, 6655 stations were included generating time series for 1418 2.5° by 3.75° gridboxes. The maximum number of stations per gridbox was 45. These two datasets were then combined using the common 1951-70 period to blend the data on the basis of mean monthly values and their variance. When merged, a total of 1520 gridboxes possessed time series of which 726 had complete data between 1900 and 1998. The period of most complete coverage was from 1952-75; the year 1959 possessed *no* missing data in any of the 1520 gridboxes.

4. Some Notes About Reliability

All station data have been screened for gross outliers and typographical errors using a number of semi-automated techniques. Owing to the large spatial variability of precipitation these methods, however, are not foolproof. The GHCN has considered and implemented further improvements to these screening methods (Easterling and Peterson, 1995; Easterling *et al*., 1996).

No corrections for gauge undercatch have been applied to the station data (cf. Sevruk, 1982; Legates and Willmott, 1990). A spatially varying, but temporally constant, correction could be applied to the estimates derived from Legates and Willmott

(1990), although this would not alter the trends in the data. Applying time-dependent corrections to gauge time series on a global scale is a gigantic undertaking which may well not be either feasible or justifiable.

A number of Northern Hemisphere high latitude time series contain inhomogeneities due to varying sensitivities to snowcatch of different gauge designs and mountings. These have been well documented for certain countries (e.g. Russia; Groisman *et al*., 1991; Scandinavian countries - the North Atlantic Climatological Dataset) and work is underway to "clean" other country datasets (e.g. Canada). Groisman's 'adjusted' data have now been added to the master station dataset used here, as has the NACD archive, but further improvements in the reliability of this gridded dataset over high latitudes will follow. For the present, the user should be cautious about the precise interpretation of high latitude precipitation trends outside Russia and Scandinavia, especially in winter.

No topographic weighting has been applied to the interpolation scheme. A number of different methods exist for incorporating the effects of topography on precipitation (e.g. the PRISM and AURELHY methods and the spline algorithms of Hutchinson, 1995). However, the dependence of precipitation anomalies on elevation is much smaller and more ambiguous. Since the method used here only interpolates anomalies, and not precipitation values themselves, excluding the effects of elevation is reasonable. There are, however, other problems associated with using precipitation anomalies in a gridding algorithm like this and these are discussed by Hulme and New (1997). Further discussion and applications of these, and other gridded, precipitation datasets can be found in the following publications:

Hulme,M. (1991) An intercomparison of model and observed global precipitation climatologies **Geophys. Res. Lett**., 18, 1715-1718

Hulme,M. (1992) A 1951-80 global land precipitation climatology for the evaluation of General Circulation Models **Climate Dynamics** 7, 57-72.

Hulme,M., Marsh,R. and Jones,P.D. (1992) Global changes in a humidity index between 1931-60 and 1961-90 **Climate Research**, 2, 1-22.

Hulme,M. (1992) Rainfall changes in Africa: 1931-60 to 1961-90 **Int. J. Climatol**., 12, 685-699.

Hulme,M. and Jones,P.D. (1993) A historical monthly precipitation dataset for global land areas: applications for climate monitoring and climate model evaluation pp. A/14-A/17 in, **Analysis methods of precipitation on a global scale** Report of a GEWEX Workshop, 14-17 September 1992, Koblenz, Germany, WMO/TD-No.558, Geneva

Hulme,M. (1994) Validation of large-scale precipitation fields in General Circulation Models pp.387-406 in, **Global precipitations and climate change** (eds.) Desbois,M. and Desalmand,F., NATO ASI Series, Springer-Verlag, Berlin, 466pp.

Hulme,M. (1995) Estimating global changes in precipitation **Weather**, 50, 34-42.

Hulme,M. (1996) Recent climate change in the world's drylands **Geophys. Res. Letts.**, 23, 61-64

Jones,P.D. and Hulme,M. (1996) Calculating regional climatic time series for temperature and precipitation: methods and illustrations **Int. J. Climatol**., 16, 361-377

Hulme,M. and New,M. (1997) The dependence of large-scale precipitation climatologies on temporal and spatial gauge sampling **J.Climate** 10, 1099-1113.

Hulme,M., Osborn,T.J. and T.C.Johns (1998) Precipitation sensitivity to global warming: Comparison of observations with HadCM2 simulations **Geophys. Res. Letts**., 25, 3379-3382.

Doherty,R.M., Hulme,M. and Jones,C.G. (1999) A gridded reconstruction of land and ocean precipitation for the extended Tropics from 1974-1994 **Int. J. Climatol**., 19, 119-142.

New,M., Hulme,M. and Jones,P.D. (1999) Representing twentieth century space-time climate variability. Part 1: development of a 1961-90 mean monthly terrestrial climatology **J.Climate**, 12, 829-856.

see also:

Legates,D.R. (1995) Global and terrestrial precipitation: a comparative assessment of existing climatologies **Int. J. Climatol**., 15, 236-258.

5. References

Easterling,D.R. and Peterson,T.C. (1995) A new method for detecting undocumented discontinuities in climatological time series **Int. J. Climatol**., 15, 369-378

Easterling,D.R., Peterson,T.C. and Karl,T.R. (1996) On the development and use of homogenized climate data sets **J.Climate**, 9, 1429-1434.

Groisman,P.Ya., Koknaeva,V.V., Belokrylova,T.A. and Karl,T.R. (1991) Overcoming biases of precipitation measurement: a history of the USSR experience **Bull. Amer. Met. Soc**., 72, 1725-1733.

Hutchinson,M.F. (1995) Interpolating mean rainfall using thin-plate smoothing splines **Int. J. Geographical Inf. Systems**, 9, 385-403.

Legates,D.R. and Willmott,C.J. (1990) Mean seasonal and spatial variability in gauge-corrected, global precipitation **Int. J. Climatol**., 10, 111-128.

Shepherd,D. (1984) Computer mapping: the SYMAP interpolation algorithm in, **Spatial statistics and models** (eds.) Gaile,G.L. and Willmott,C.J., D.Reidel Publishing, Dordrecht, 133pp.

DATA file format

The data file is organised by gridbox with each gridbox entry consisting of a header line followed by n lines of monthly precipitation totals, one line per year.

HEADER line: the FORTRAN format for this line is:
(I7,I5,I6,I5,A15,I4,A14,2I4,I7,I9). The information contained is as follows:

FIELD 1. Seven digit number: this is the global gridbox number which at a 2.5° by 3.75° resolution varies between 1 and 6816. The gridbox numbering starts at 87.5° S, 180°W (centre of box number 1), proceeds first eastward, then northward and ends at 87.5°N, 176.25°E (centre of box number 6912). This gridding geography follows that of the Hadley Centre's current Unified Model (post-1992 version). [NB. this grid is slightly different from previous grids since half a gridbox at the Poles has been lost]. Only those gridboxes which have valid data are included in the data file; thus 1520 gridbox entries exist in the current version of gu23wld0096.dat.

FIELD 2. Five digit number: this is the latitude of the centre of the gridbox to the nearest 0.01. Latitude has been decimalised and multiplied by 100 (e.g. -8750 means 87°30’S). North is positive and south is negative.

FIELD 3. Five digit number: this is the longitude of the centre of the gridbox to the nearest 0.01 . Longitude has been decimalised and multiplied by 100 (e.g. -17625 means 176°15'W). West is negative and east is positive.

FIELD 4. Five digit number: this is the mean altitude, in metres above sea level, of the stations which have contributed to the gridbox estimate. Where altitude is not known '-999' is entered. This is not the `true' elevation of the gridbox.

FIELD 5. Fifteen character country name: this is the dominant country among the stations which have contributed to the gridbox estimate. It is not necessarily the country in which the largest portion of the gridbox falls. A full list of countries worldwide is available on a separate sheet.

FIELD 6. Four digit integer: this is the number of stations which have contributed to the gridbox estimate.

FIELD 7. Fourteen character string which contains diagnostic information about the gridding technique. Can be ignored.

FIELD 8. Two four digit numbers: these represent the start and end years respectively of the data held for that gridbox in the data file. These years do not necessarily imply that all monthly totals for these years are present.

FIELD 9. Same as FIELD 1.

FIELD 10. Nine digit integer: a diagnostic which can be ignored.

DATA lines: each line contains 14 numbers in format (I4,12I5,I6). Number 1: is the year; Numbers 2-13: are the twelve monthly precipitation totals in mms*10 (e.g. 246 is 24.6mm). Where a monthly total is missing '-10' is entered. Trace rainfalls are entered as '0'; Number 14: is the annual precipitation total in mms*10 (e.g. 15467 is 1546.7mm). This annual total will be the sum of the 12 monthly totals. Where one or more monthly totals are missing, no annual total is entered and the missing code '-10' is used.

***** No liability is accepted for errors in the dataset *****

For further information about these gridded datasets contact:

Dr Mike Hulme

Climatic Research Unit
School of Environmental Sciences
University of East Anglia
Norwich NR4 7TJ, UK

tel: +1603 593162; fax: +1603 507784
email <m.hulme@uea.ac.uk>
web site: http://www.cru.uea.ac.uk/~mikeh