

A high-resolution data set of surface climate over global land areas

Mark New^{1,*}, David Lister², Mike Hulme³, Ian Makin⁴

¹School of Geography and the Environment, University of Oxford, Mansfield Road, Oxford OX1 3TB, United Kingdom

²Climatic Research Unit, and ³Tyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, United Kingdom

⁴International Water Management Institute, PO Box 2075, Colombo, Sri Lanka

ABSTRACT: We describe the construction of a 10' latitude/longitude data set of mean monthly surface climate over global land areas, excluding Antarctica. The climatology includes 8 climate elements—precipitation, wet-day frequency, temperature, diurnal temperature range, relative humidity, sunshine duration, ground frost frequency and windspeed—and was interpolated from a data set of station means for the period centred on 1961 to 1990. Precipitation was first defined in terms of the parameters of the Gamma distribution, enabling the calculation of monthly precipitation at any given return period. The data are compared to an earlier data set at 0.5° latitude/longitude resolution and show added value over most regions. The data will have many applications in applied climatology, biogeochemical modelling, hydrology and agricultural meteorology and are available through the International Water Management Institute World Water and Climate Atlas (<http://www.iwmi.org>) and the Climatic Research Unit (<http://www.cru.uea.ac.uk>).

KEY WORDS: High resolution data set · Surface climate · Applied climatology · Precipitation probability · IWMI Climate Atlas · CRU

Resale or republication not permitted without written consent of the publisher

1. INTRODUCTION

Spatially complete representations of surface climate are required for many purposes in applied and theoretical environmental science. Examples include biogeochemical modelling (Cramer & Fischer 1996), forestry (Booth & Jones 1998), agriculture (Nicholls 1997, Changnon & Kunkel 1999), hydrology and water resources (Arnell 1999), climate change studies (Hulme & Jenkins 1998, Hulme et al. 1999, Giorgi & Francisco 2000). Typically, the required spatial resolution of climate data increases with the resolution of analysis. Global analyses of climate change impacts and terrestrial biogeochemical cycling historically used data gridded at 30' latitude by longitude (lat/lon) or coarser resolution, due to an absence of higher resolution data that are global in extent as well as the computing penalty invoked if higher resolution input data were used. Regional and country scale analyses tend to

require higher resolution data (grid spacing <25 km) so that climatic and climate-driven differences across regions and within sub-regions can be identified (e.g. Daly et al. 1994, Frei & Schar 1998). However, in many regions of the world such high-resolution data have not been available.

This paper describes the construction of a 10' lat/lon mean monthly climatology of surface climate over global land areas, excluding Antarctica (henceforth NEW01). The data set represents an improvement on an earlier gridded climatology at 30' lat/lon resolution (henceforth NEW99, New et al. 1999) through an increased spatial resolution, the incorporation of additional station data and the inclusion of a description of precipitation variability, enabling the calculation of probability distributions of monthly precipitation. The climatology comprises a suite of variables: mean temperature, diurnal temperature range, relative humidity, sunshine, ground-frost frequency, wet-day frequency, wind speed, and the 2 parameters of the Gamma distribution fitted to monthly precipitation data.

*E-mail: mark.new@geog.ox.ac.uk

The data set was primarily developed for use in the International Water Management Institute's (IWMI) World Water and Climate Atlas, but will also appeal to a wider audience who require climate data at a resolution higher than that of previously available data sets. The IWMI Atlas is designed as a growing collection of data products and analytical tools focused on climate and water resources, assembled in a standardized format that can be quickly and easily analysed using the accompanying software package, the IWMI Synthesizer. Examples of applications of the Atlas data include: identifying areas suitable for rain-fed agriculture, providing inputs for hydrologic modelling of river basins, providing basic climate data for crop modelling, and helping in the projection of water supply and demand nationally and globally. Monthly climate summaries for selected locations and variables are available at the IWMI web site (www.iwmi.org), through a purpose-developed web query engine (CAWQuer) that accesses the full Atlas data set.

The structure of the paper is as follows. In Section 2 the station data and digital elevation data are briefly described, and the Gamma distribution and its use with monthly precipitation is discussed. Section 3 describes the thin-plate spline methodology used to interpolate the station data to produce continuous fields and their associated errors. In Section 4 the derived grids are compared to a previous climatology at 0.5° resolution, while in Section 5 regional examples are used to illustrate some of the characteristics of the gridded data. Section 6 draws the paper together with a summary and some conclusions.

2. DATA

The climate station data used to construct the climatology represent the fruits of a data collation program at the Climatic Research Unit (CRU) extending from the 1980s to the present day (Hulme 1994a,b, Jones 1994, Jones et al. 1999). These data have been comprehensively described by New et al. (1999, 2000), so are only summarised in the sections that follow.

2.1. Sources

The station means used to construct the climatology are predominantly climatological normals¹ for the period 1961 to 1990, collated from a number of sources. These included direct contacts (and indirect contacts

via web-servers) with national meteorological agencies and archive centres, the WMO 1961 to 1990 climatological normals (WMO 1996), the Centro Internacional de Agricultura Tropical (CIAT), as well other published sources and personal contacts, and data sets of monthly station time-series held in CRU. The main sources are detailed below:

- Data supplied by National Meteorological Agencies (NMAs) comprise by far the largest single data source. Most of these data were supplied on diskette, but about 23% were supplied either as published volumes of 1961 to 1990 normals or on NMA paper copies. Data on the latter 2 media were scanned or keyed in, with independent checks against the originals.

- The WMO recently published the 1961 to 1990 climatological normals, which were merged with the CRU data set. This resulted in 690 additional stations that had data for at least 1 variable not previously in the CRU data set.

- Several countries in Africa (e.g. Zaire and Angola) and Southeast Asia (e.g. Cambodia) provided few or no 1961 to 1990 normals to either CRU or the WMO. In these cases data were extracted from 1 of 2 sources: Müller (1982) and FAO (1984). Generally, these means were calculated using data from the period 1931 to 1960. In the case of the FAO publications, the number of years contributing to a mean was unknown. A small number of stations from several other sources were used. These included the USAF Climatological Data Volume (USAF 1987) and a number of personal contacts where data for between one and several tens of stations were obtained.

- Personnel at CIAT have collated several thousand climatological means for South and Central America (P. Jones pers.comm.). Unfortunately, the period each mean represents is unspecified, although the number of years of record contributing to the mean is usually supplied. These were assumed to be broadly representative of the period 1950 to 1990, but were in any case assigned a lower weight in the interpolation scheme.

- Global data sets of monthly time series of precipitation (Eischeid et al. 1991, updated), mean temperature (Jones 1994, updated), maximum and minimum temperature (the Global Historical Climatology Network; Easterling et al. 1997, with additions by CRU) for several thousand stations worldwide were searched for additional stations. CRU also holds smaller data sets of monthly time series of the other variables. These data have been quality-controlled and checked for inhomogeneities. Station means for 1961 to 1990 were calculated from these time series and added to the normals data set.

Despite these data collation efforts, the CRU data in many regions still represent only a sub-set of the potentially available stations.

¹A 'normal' is a climatic average for a defined period, such as the WMO standard normal periods of 1931–1960 and 1961–1990

2.2. Quality control

Data from the WMO collection were subjected to a fairly comprehensive series of quality control (QC) checks by the National Climatic Data Centre (NCDC 1997). Data obtained directly from NMAs were assumed to have been quality-checked at source. Nonetheless, all data were subjected to a 2-stage QC process. In the first stage, prior to interpolation, a standard series of automated tests were performed on individual station normals. These tests were similar to those used by the NCDC during the collation of the WMO 1961 to 1990 climatological normals, namely: (1) internal consistency checks, e.g. ensuring that the monthly means follow a consistent seasonal cycle and that predefined absolute limits are not exceeded; and (2) between-variable consistency tests, e.g. ensuring that monthly minimum, mean and maximum temperatures are consistent and that months with zero precipitation have zero wet-days.

The second stage of QC occurred during the interpolation of station data, where the interpolation diagnostics enabled identification of station-months that had large residuals (see Section 3.2), and were potentially in error. As a general rule, data that failed these QC tests were removed from the interpolation. In some cases, however, the data could be compared and replaced with normals calculated from the CRU monthly station time-series described above.

2.3. Variables

The number of stations for each variable varies markedly (Figs. 1 to 9). Precipitation and temperature are the most widely available, followed by diurnal temperature range (simply the difference between mean maximum and minimum temperature). Windspeed is the least widely reported variable with just over 3950 stations globally (Fig. 1), while precipitation has

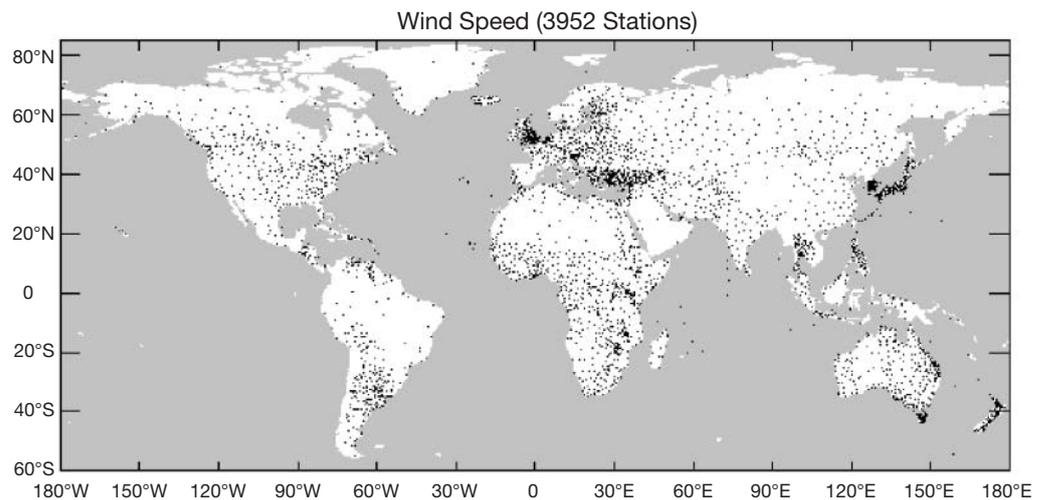


Fig. 1. Distribution of stations for which mean wind speed was available

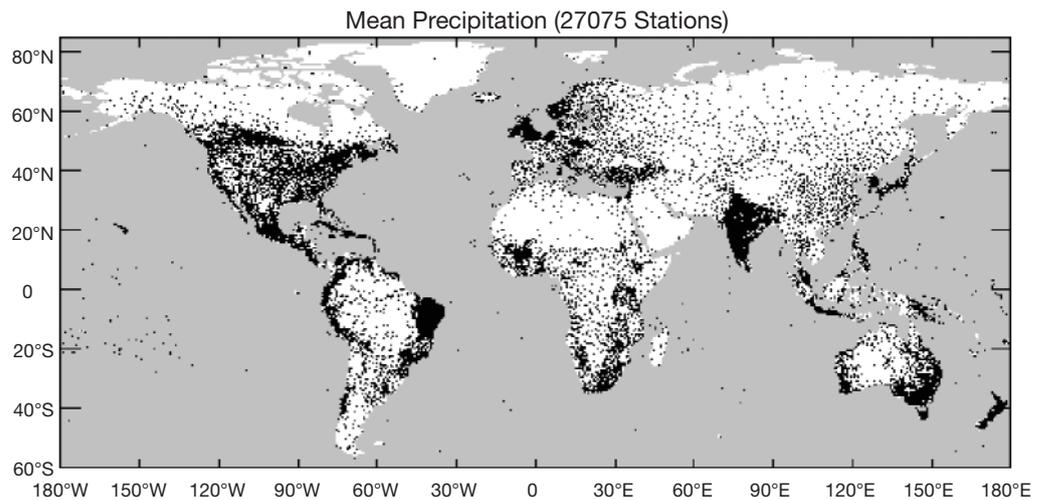


Fig. 2. Distribution of stations for which mean precipitation was available

the most extensive and dense network, with some 27 000 stations for mean precipitation (Fig. 2) and 22 000 stations for the coefficient of variation (CV) of precipitation (Fig. 3).

The ways in which surface climate variables are measured and reported vary both within and between countries. Common differences include gauge/instrument type and height, measurement times, and the units of measurement. For instance, precipitation measurements can be influenced by several factors, most notably gauge type, the ratio of solid to liquid precipitation and wind conditions/turbulence. Various attempts have been made to correct for such biases in precipitation (e.g. Groisman et al. 1991, Legates & DeLiberty 1993, Yang et al. 1999, Groisman & Ranakova 2001). The CRU data over the former USSR include the original data adjusted by Groisman et al. (1991), but we made no attempt to correct precipitation data over other regions; in most cases there was insufficient information to attempt this.

Wind speed is measured at heights above the surface of between 2 and 20 m. Measurement height varies both within and between countries, and in many cases the heights were not specified. Consequently, no corrections were made to wind data. The large majority of known heights were 10 m, and the interpolated wind field should be assumed to represent speed at this height.

Differences in temperature measurement times have been shown to induce disparities of several tenths of a degree Celsius (Karl et al. 1986, Andersson & Mattison 1991) and different countries calculate mean temperature in various ways (Jones et al. 1986). Where possible, mean temperature was defined as the average of mean maximum and minimum temperature, which are measured more uniformly across the world. At stations where only mean temperature was available, these values were used, despite the uncertainty about their derivation.

Wet-day frequencies were generally expressed as the number of days per month with precipitation >0.1

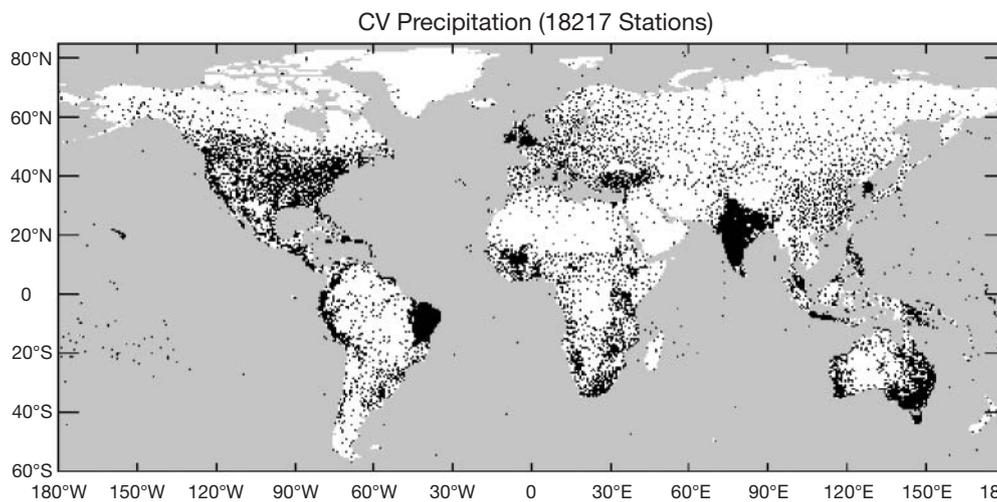


Fig. 3. Distribution of stations for which the coefficient of variation (CV) of monthly precipitation was available

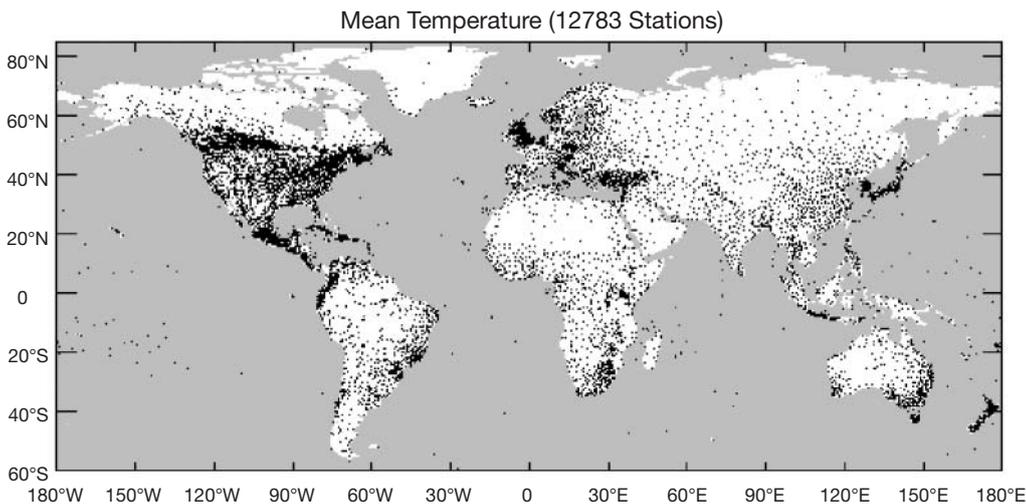


Fig. 4. Distribution of stations for which mean temperature was available

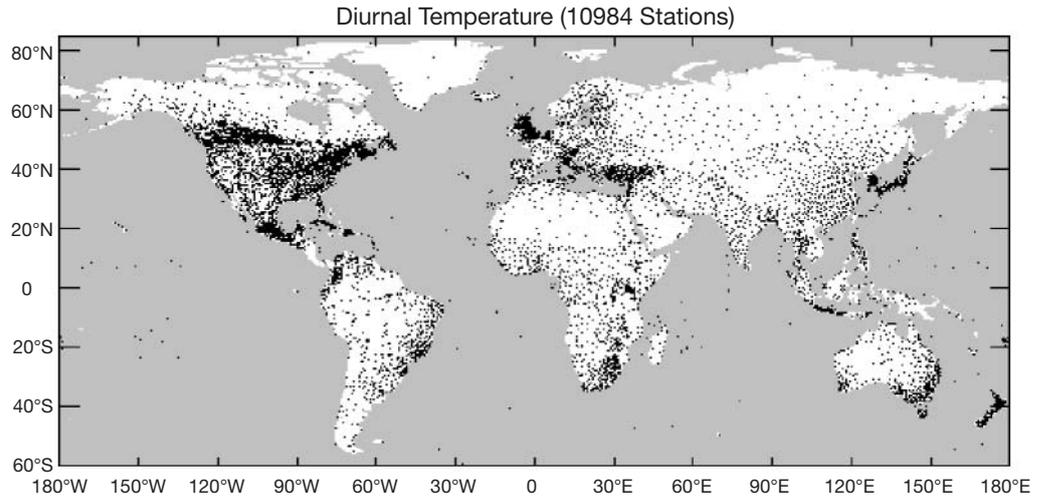


Fig. 5. Distribution of stations for which diurnal temperature range was available

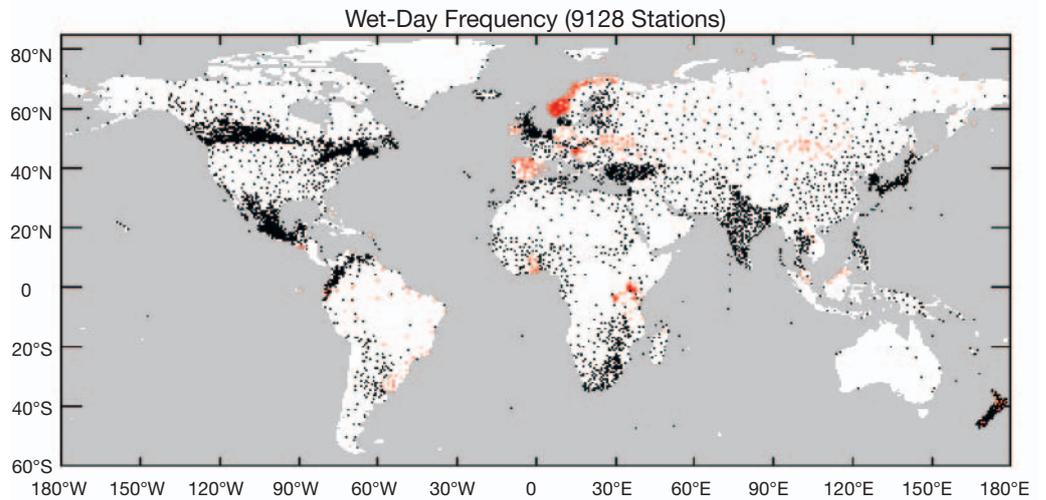


Fig. 6. Distribution of stations for which wet-day frequency was available. Black and red dots represent stations for which wet-day frequency was provided with the 0.1 mm and 1.0 mm thresholds respectively. The latter were converted to the equivalent 0.1 mm wet-day frequency

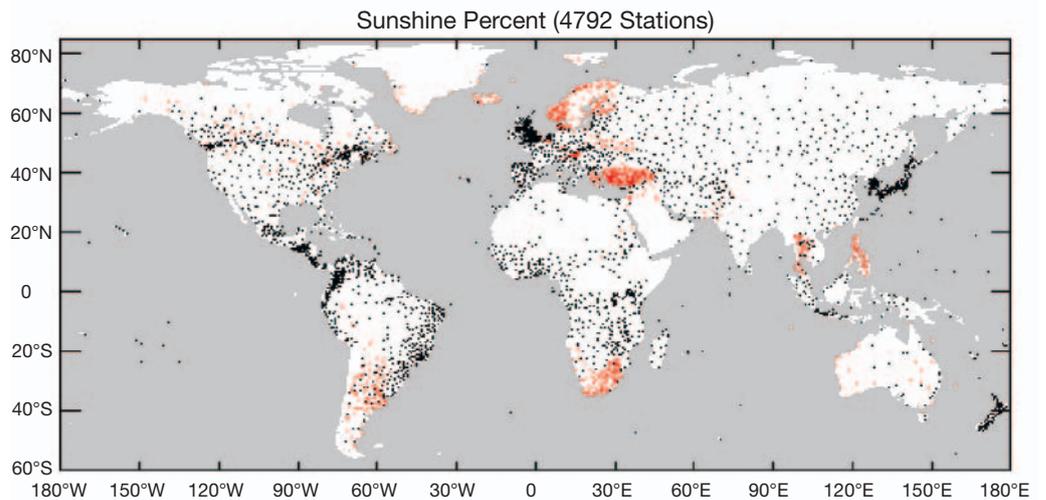


Fig. 7. Distribution of stations for which sunshine (black) and cloud fraction (red) was available. The latter were converted to sunshine percent as described in the text

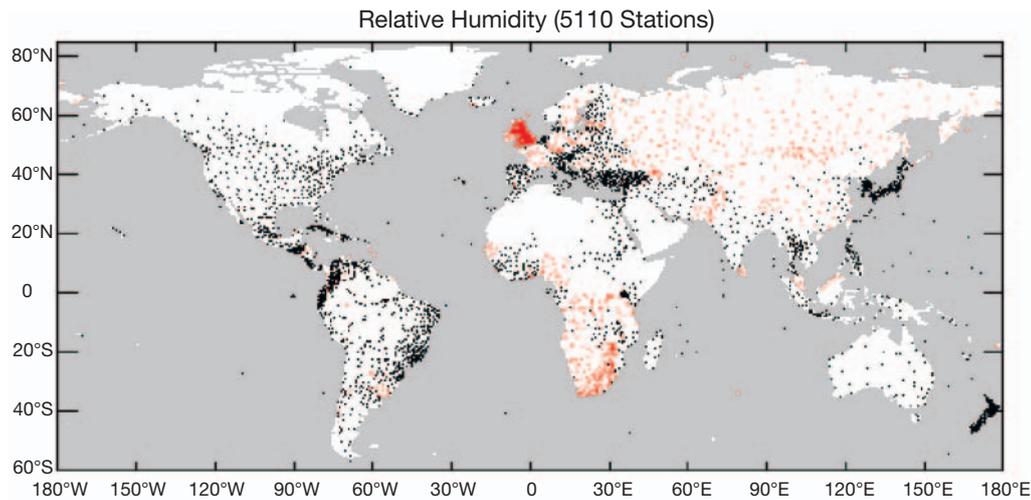


Fig. 8. Distribution of stations for which relative humidity (black) and vapour pressure (red) was available. The latter were converted to relative humidity, as described in the text

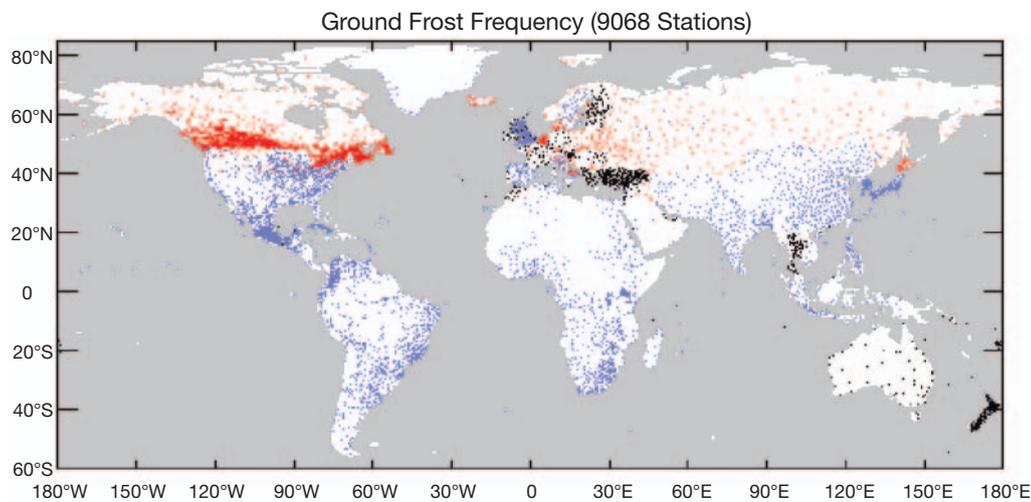


Fig. 9. Distribution of stations for which ground frost frequency (black) and air frost frequency (red) were available. The latter were converted to ground frost frequency, as described in the text. Blue dots represent stations where minimum temperature was used to estimate ground frost frequency

or 1.0 mm. There were approximately twice as many stations reporting wet-day frequency with a 0.1 mm threshold than those with a 1.0 mm threshold; consequently those with a 1.0 mm threshold were converted to a 0.1 mm threshold, using the empirical conversion procedure described by New et al. (1999). A small number of normals (e.g. UK and Australia) used some other threshold, typically 0.2 mm, and no adjustment was made for these more moderate differences.

The data set contains normals for cloud cover and sunshine, mostly from different stations. Sunshine normals were supplied as either mean hours per month or percent of maximum possible bright sunshine. Total cloud cover normals were mostly provided in oktas (eighths) and sometimes in tenths. Normals in units of sunshine hours were converted to percent of possible maximum sunshine, and cloud cover normals were standardised to oktas. For some countries both cloud cover and sunshine data were available, but in most instances either one or the other was provided. To obtain more complete coverage, cloud cover was con-

verted to sunshine using the approach described by New et al. (1999), which is based on the methodology of Doorenbos & Pruitt (1984), but with an additional latitudinal correction.

Humidity normals in the CRU data set comprised roughly equal numbers of relative humidity (RH) and vapour pressure (e). Vapour pressure measurements were converted to RH using the equation:

$$RH = 100 \frac{e}{e_{\text{sat}}} \quad (1)$$

where e_{sat} is the saturated vapour pressure (in hPa) at mean air temperature (T), calculated using the Clausius-Clapeyron equation (Shuttleworth 1992, p. 4.3):

$$e_{\text{sat}} = 6.108 \exp \left[\frac{17.27T}{237.3 + T} \right] \quad (2)$$

This estimation is reliable provided that temperature and relative humidity are measured simultaneously and the temperature is above zero. In some cases the mean temperatures at the times of mea-

surement were available and the estimated RH can be considered reliable. However, in many cases only the mean, or minimum and maximum temperature was available, while RH was the mean of 1 or several daily measurements. Validation of the conversion using stations where both RH and e were available suggest the estimated RH will have errors of up to $\pm 15\%$ (New et al. 1999). Below freezing, 2 issues complicate the measurement of vapour pressure and its conversion to RH. First, saturation vapour pressure over ice is lower than over water, and not all instruments can account for this. Secondly, saturated vapour pressure decreases from ~ 1.0 hPa at -20°C to ~ 0.15 hPa at -40°C ; at these low pressures small absolute measurement errors (~ 0.1 hPa) are large in relative terms, and induce similarly large errors in RH. These problems are revisited in Section 5, where the gridded data are evaluated over selected regions.

Most ground frost normals were defined as the frequency of grass minimum temperatures $< 0^\circ\text{C}$. Some normals, however, were defined as the frequency of minimum air temperatures $< 0^\circ\text{C}$ and these had to be converted to ground frost frequency. As there was no straightforward theoretical basis for this conversion, the empirical formula derived by Hulme et al. (1995) was used. Validation of this formula indicated that the estimates are typically accurate to within 2 d mo^{-1} (New et al. 1999). For much of the tropics, no ground frost or air frost normals were available for the obvious reason that temperatures rarely fall below freezing point. This absence of data presented problems for the interpolation of frost frequency to high elevations in the tropics. In these regions, ground frost frequency was estimated from minimum temperature data using the formulation of New et al. (1999).

The interpolation described in Section 3 makes use of elevation as a co-predictor. The elevation data used to create grids of the interpolated surfaces derives from the GLOBE 1 km elevation database, downloaded from the National Geophysical Data Centre (NGDC 2000). The GLOBE elevation data were resampled to produce a $10'$ elevation grid, simply by calculating the average of all GLOBE elevations in each $10'$ grid box. A $10'$ cell was considered to be ocean only if $< 25\%$ of the constituent GLOBE elevation cells were land. As a result, the elevation of any $10'$ cell designated as land represents the average elevation of at least 100 (maximum 400) GLOBE elevation points.

2.4. Precipitation probabilities

In order to estimate precipitation exceedence probabilities, the Gamma distribution was chosen to model the probability distribution of monthly precipitation. Although other distributions are possible (Stedinger et al. 1992, Hutchinson 1995b), the Gamma distribution can be easily calculated using the 'method of moments' — the mean and standard deviation of the observed data are used to define the 2 parameters of the Gamma distribution (Stedinger et al. 1992):

$$f_x(x) = |\beta|(\beta x)^{\alpha-1} \frac{\exp(-\beta x)}{\Gamma(\alpha)} \quad (3)$$

$$\alpha = \frac{\mu_x^2}{\sigma_x^2} = \frac{1}{CV_x^2} \quad (\text{the 'shape' parameter})$$

$$\beta = \frac{\mu_x}{\sigma_x^2} = \frac{\sqrt{\alpha}}{\sigma_x} \quad (\text{the 'scale' parameter})$$

where μ_x is the arithmetic mean, σ_x is the standard deviation, and CV is the coefficient of variation. The commonly used terms of 'shape' and 'scale' for parameters α and β arise from the effect they have on the Gamma probability density function, illustrated in Fig. 10. α controls the shape (or skewness) of the distribution, as a function of the relative variability, or coefficient of variation (CV). A high CV results in a positively skewed distribution, while a lower CV produces a more normal distribution. In contrast β describes the scale over which the distribution occurs. For any shape, the scale increases (β decreases) as the arithmetic mean of the distribution increases.

Using parametric estimators also simplifies the interpolation procedure, because only the mean and coefficient

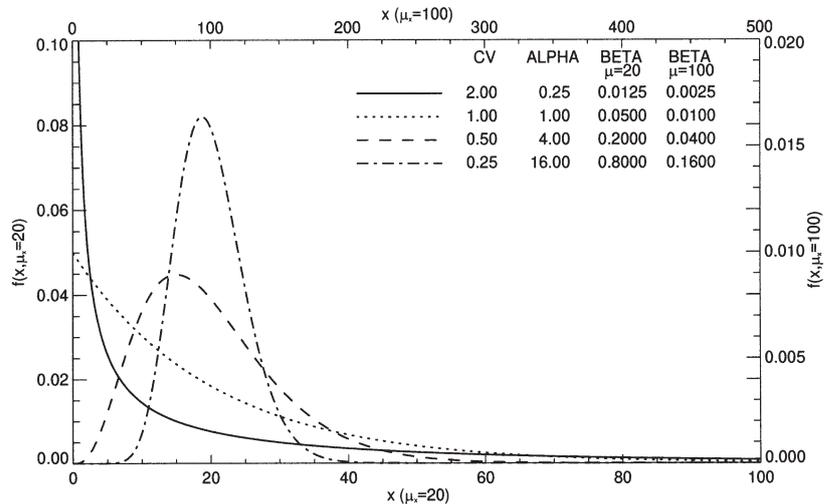


Fig. 10. The probability density function for the Gamma distribution, with α and β defined for a range of coefficients of variation (CV_x) and for data with a mean (μ_x) of 20 and 100

cient of variation of precipitation need to be interpolated. Fields of the parameters of the Gamma distribution can then be calculated from the gridded mean and CV fields. This approach was used because μ and CV are easier to interpret and verify than α and β . The alternative—interpolating α and β directly—would have made intuitive verification of the interpolated fields more difficult.

A further advantage of interpolating the mean and CV of precipitation lay in the larger number of precipitation normals (means) in the CRU data set. In contrast, precipitation CVs had to be calculated from monthly time-series in the CRU data set, which contains far fewer stations (compare Figs. 2 & 3). Interpolation of α and β directly would have limited the number of stations to those in the data set of station time series.

A limitation of the use of the Gamma distribution is its behaviour when fitted to data with many months with zero precipitation, a common occurrence in arid and seasonally arid regions. In such situations, the empirical cumulative distribution function is discontinuous near zero, while the Gamma cumulative distribution function is continuous, and tends to approach zero as x

approaches zero. Because of this the frequency of dry months inferred from the fitted distribution will tend to be underestimated. An alternative procedure would be to estimate the probability of zero rainfall from the data, and then estimate the Gamma distribution only for non-zero values of monthly rainfall. However, this would have required separate interpolation of fields of zero probability and precluded the use of stations for which mean rainfall was available, thus reducing station control. For these reasons, the simpler approach described in the previous paragraph was adopted.

The effect of fitting the Gamma distribution to all data is illustrated in Fig. 11, which shows the fitted and empirical percentiles of monthly precipitation at Niamey (Niger), encompassing a wide range of precipitation regimes. In January, the height of the dry season, only 4 mo with rainfall have occurred in the 93 yr on record; thus the empirical probability of zero rain is ~ 0.96 . In contrast, the fitted distribution predicts the probability of zero rainfall to be lower, at 0.94. A further consequence of this bias is that the predicted rainfall quantiles are overestimated (underestimated) at lower (higher) percentiles.

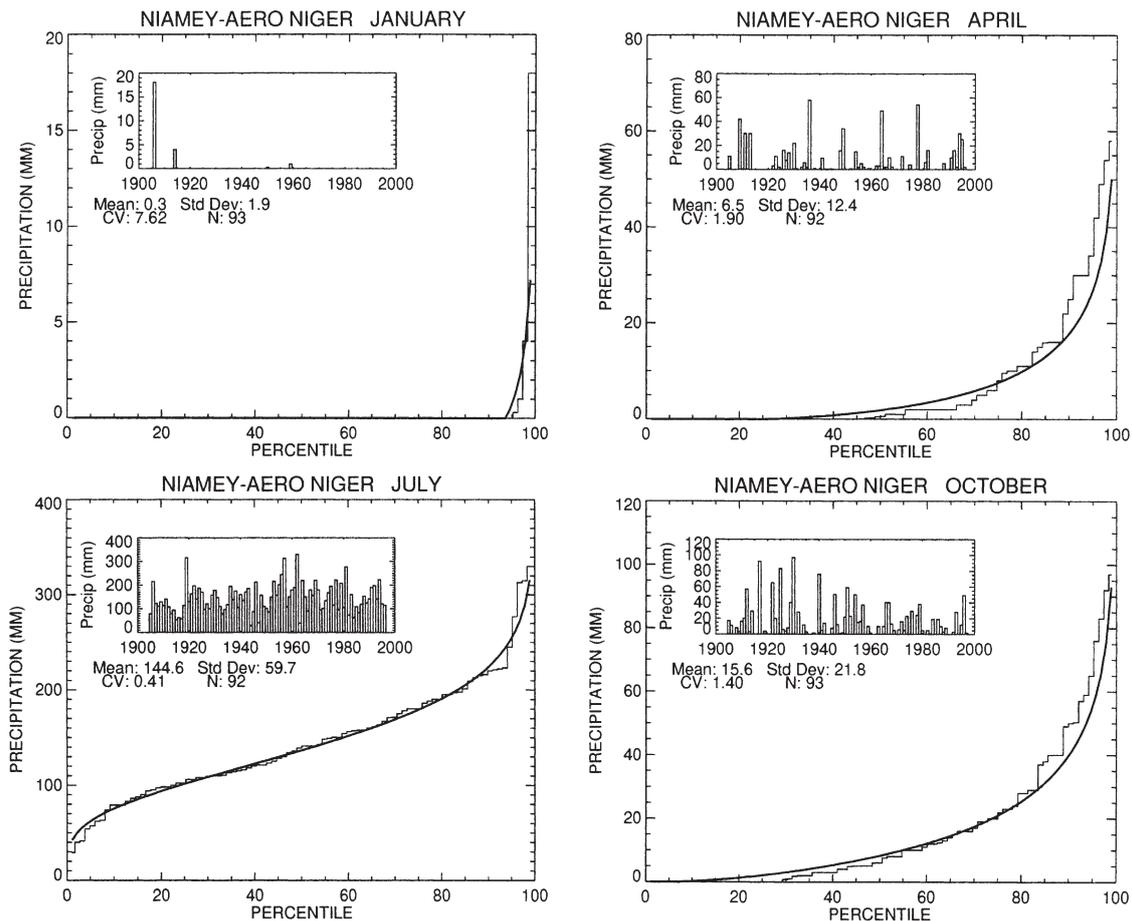


Fig. 11. Fitted (Gamma distribution) and empirical percentiles of monthly precipitation at Niamey, Niger

April and October precipitation at Niamey represent intermediate cases, where there are fewer dry months (30 and 45 % respectively; see Fig. 11). As with January, the probability of zero rainfall is underestimated, this time by ~20 %. Similarly, the lower (higher) precipitation totals are overestimated (underestimated), by a few mm at lower totals and 10 to 20 mm at higher totals. In July, during the West African monsoon, all months in the record have rainfall, with a minimum of 30 mm. Here the fit is much better and the biases at low and high totals are not as marked; nonetheless, there remains over- and underestimation at the lowest and highest totals respectively.

Similar results were found when evaluating the Gamma distribution in a number of climatic environments (United Kingdom, Southern Africa, Brazil and India), but are not shown here. These results all suggest that, apart from the largest percentiles, the estimation error in all months is only a few mm, and is not significant in most practical applications (e.g. crop suitability assessments). At the largest percentiles, precipitation is considerably overestimated and estimates at these percentiles should be treated with caution; indeed it is recommended that precipitation should be estimated only between the 10th and 90th percentiles to avoid the relatively large errors associated with the tails of the fitted distributions.

3. INTERPOLATION

3.1. Thin-plate smoothing splines

The station climate statistics were interpolated using thin-plate smoothing splines (ANUSPLIN) developed by Mike Hutchinson at the Australian National University (Hutchinson 1999). The original thin-plate spline-

fitting technique is described by Wahba (1979), while Hutchinson (1995a) provides a theoretical description of their application to surface climate variables such as precipitation. Spline interpolation is robust in areas with sparse or irregularly spaced data points. Thin-plate splines are defined by minimising the roughness of the interpolated surface, subject to the data having a predefined residual. This is usually accomplished by determining the amount of data smoothing that is required to minimise the generalised cross validation (GCV). The GCV is calculated implicitly and hence without recourse to computationally demanding iterative procedures. The main advantage of splines over many other geostatistical methods is that prior estimation of the spatial auto-covariance structure is not required (Hutchinson 1995a).

Using the Bayesian arguments of Wahba (1983) and Silverman (1985) to define the error covariance matrix of the fitted values, the ANUSPLIN package also enables the calculation of spatially distributed standard errors about the fitted surface. The Bayesian standard error can be interpreted in an analogous manner to the standard error in linear regression. Confidence intervals for the predicted surface can be estimated by multiplying the surface standard error by 1.96, the 95 % 2-sided confidence interval for the standard normal distribution.

3.2. Interpolation of climate variables

The approach to interpolation of all surface climate variables except precipitation CV is the same as that described by New et al. (1999). Trivariate thin-plate spline surfaces were fitted as functions of latitude, longitude and elevation to the station data over several regional domains or 'tiles' (Fig. 12). The inclusion of

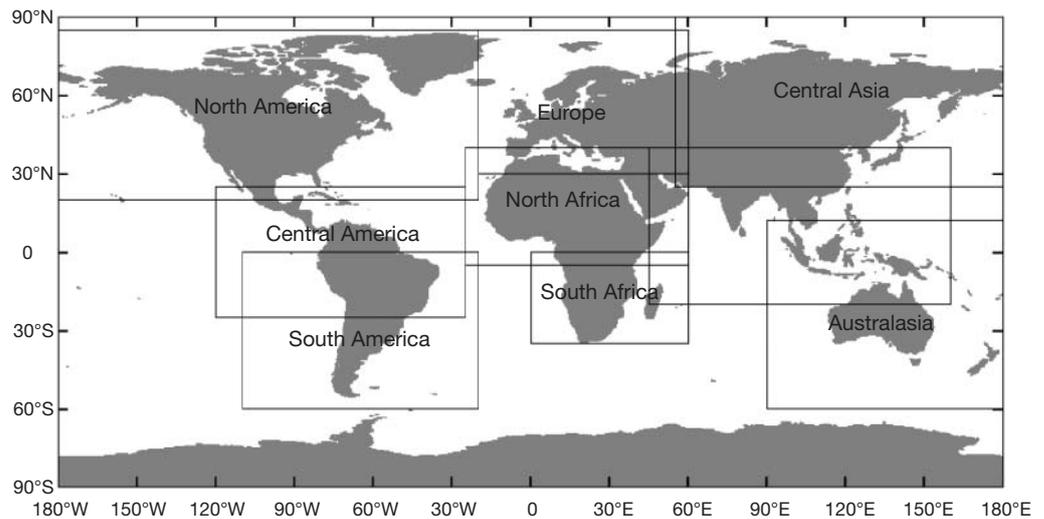


Fig. 12. Spatial domains or 'tiles' over which the interpolations of surface climate data were performed. Note that for some variables, most notably precipitation, the actual surface fitting was undertaken over 2 or more sub-tiles within each domain in order to reduce computational times

elevation as a co-predictor adds considerable skill to the interpolation, enabling topographic controls on climate that are resolved by the station data to be captured. For some of the variables, especially precipitation, station numbers were too large for available computer memory to permit interpolation over an entire domain. In such cases the continental domains were divided into sub-areas that were subsequently merged. Sub-areas were defined with overlaps of several degrees latitude and/or longitude to avoid discontinuities when they were merged.

Precipitation CV was interpolated as a function of latitude, longitude and mean precipitation. There were 2 main reasons for this. First, CV is closely (inversely) related to mean precipitation and we found that precipitation was as good or better as a predictor of CV than elevation in most regions. Second, making the interpolated CV dependent on mean precipitation ensured that the 2 statistics varied together in a consistent manner in regions with poor station control. The alternative, interpolating CV as a function of location and elevation, increased the likelihood of different elevation dependencies arising during the surface fitting process, with the consequence that the mean and CV of precipitation would not necessarily vary together in a consistent manner.

The spline-fitting program provided a list of the stations with the largest residuals from the fitted surfaces. These lists were used to identify and check potentially erroneous stations. In some instances, these stations were deemed to be correct and their positions as outliers were assumed to be due to local climatological variations that could not be resolved with the available network. However, a number of station outliers were found to have identifiable errors—most common mistakes included inaccurate locational and/or elevation information, and typographic errors, where values for a single month did not fit in with the overall seasonal pattern. Stations such as these were corrected or, if this was not possible, excluded from the interpolation data set.

Once spline surfaces had been fitted to the station data, the resultant surface coefficients were used in conjunction with the 10' elevation data to calculate grids of climate variables. In the case of precipitation CV, the (interpolated) grid of mean precipitation was used instead of the elevation to compute the gridded estimates. At this stage, the grids were plotted and inspected visually for errors that might have escaped the earlier QC.

3.3. Interpolation errors

By definition, geostatistical interpolation involves errors, as it is the error structure of the input data that helps to define the fitted surface. For geostatistical

interpolation to be appropriate, the data should have some spatial predictability. Variability in the data that is not predictable (in this case, variability that is not a function of latitude, longitude and elevation) is considered to be noise. Thus local topographic effects such as rain shadows cannot be resolved unless: (1) a predictor that is a proxy for this influence is incorporated in the interpolation, and/or (2) there are sufficient stations to capture this local dependency as a function of latitude, longitude and elevation. In regions with sparse data, the station networks used to create these data sets are clearly unable to capture this sort of detail and the interpolated fields will represent the larger-scale (>25 to 50 km) climatic regime; however the surfaces will nonetheless explicitly incorporate the larger-scale climate lapse rates and these will be reflected in the gridded data set as a function of local elevation in the gridded elevation data. In regions with increasingly dense station data, increasingly finer-scale topographic dependencies will be resolved.

The noisier the data are (in terms of predictability as a function of latitude, longitude and elevation), the greater will be the predictive error as one moves away from control stations. The spline-fitting program provides an estimate of this predictability through the GCV. Although calculated implicitly, the GCV is equivalent to removing each data point in turn and summing, with appropriate weighting, the square of the difference between the omitted point and that predicted by a surface fitted using all the other points.

The square root of the GCV (RTGCV) for each interpolation domain and variable is shown in Tables 1 to 9. As would be expected, the mean cross validation errors are similar to those reported for NEW99 (using a similar set of station data). The variables with the highest relative error are precipitation (Table 1) and wind speed (Table 9). In the case of precipitation this is because of its inherent spatial variability, even though there are >27 000 stations in the data set. RTGCV for precipitation varies from <10% of the domain mean over NE North America to >50% in domains with variable elevation and low station density, such as South America and Central Asia. Wind speed has the sparsest station network, with consequent high cross validation errors; the sparse network is compounded by difficulties in interpolating the rather large gradients in wind speed from coastal to inland sites, the variability in measurement heights (Section 2.3) and the strong control of local topography in modifying the larger-scale wind regime.

The relative RTGCVs for precipitation CV lie between 20 and 40% of the domain mean CV, with no region exhibiting relative errors that are notably larger than other regions. This is at least in part because the regions with the most variable CV also have the highest domain-mean CV, so large absolute errors are less

Table 1. Square root of the generalised cross validation (RTGCV; expressed as the percent of the domain-mean) for precipitation over the interpolation domains. Note that in some cases the continental domains were subdivided into smaller interpolation tiles to reduce the computation times where there were a large number of stations present

Region	Domain	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Asia	20° S, 45° E–40° N, 160° E	46	44	41	42	39	35	36	36	34	39	44	43
Australia	60° S, 90° E–12° N, 140° E	32	31	30	29	28	27	30	29	29	29	30	31
Australia	60° S, 140° E–12° N, 180° E	21	22	21	22	23	25	25	26	27	23	22	20
C. Asia	25° N, 55° E–90° N, 180° E	59	54	53	59	47	65	62	54	45	41	21	55
C. America	8° N, 120° W–20° N, 25° W	61	61	56	38	28	23	25	23	21	24	36	54
C. America	0° N, 50° W–20° N, 25° W	17	17	15	18	21	26	36	37	43	38	27	21
C. America	0° N, 120° W–20° N, 80° W	56	49	48	40	31	24	26	26	24	29	44	60
C. America	0° N, 80° W–20° N, 50° W	36	31	29	26	25	26	32	30	27	26	31	39
Europe	30° N, 20° W–85° N, 10° E	20	19	19	16	13	14	15	15	17	18	18	20
Europe	30° N, 10° E–85° N, 60° E	31	30	30	25	21	21	22	23	31	31	29	32
N. Africa	5° S, 25° W–40° N, 60° E	29	28	28	36	40	35	34	31	32	38	38	34
N. America	35° N, 80° W–85° N, 20° W	11	10	9	9	8	8	7	8	9	10	8	9
N. America	20° N, 180° W–35° N, 20° W	26	33	37	39	30	21	23	22	22	27	36	33
N. America	35° N, 180° W–85° N, 105° W	27	27	25	22	17	14	13	13	18	28	27	27
N. America	35° N, 105° W–85° N, 80° W	15	11	10	8	8	8	10	8	10	10	9	11
S. Africa	40° S, 0° E–0° N, 60° E	29	28	27	34	41	35	33	31	32	39	32	29
S. America	60° S, 110° W–15° S, 20° W	17	20	20	26	33	36	38	36	28	23	21	19
S. America	15° S, 110° W–0° N, 50° W	27	25	27	28	30	34	40	40	36	31	31	32
S. America	15° S, 50° W–0° N, 20° W	17	18	16	20	26	33	44	46	44	30	22	19

Table 2. RTGCV as for Table 1, but for the coefficient of variation of precipitation

Region	Domain	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Australia	60° S, 90° E–12° N, 180° E	25	21	21	16	16	18	17	21	19	15	16	20
Asia	20° S, 87° E–40° N, 160° E	25	29	25	23	22	26	24	29	28	28	33	34
Asia	20° S, 45° E–40° N, 87° E	29	28	30	33	33	26	29	30	25	25	28	28
C. Asia	25° N, 55° E–90° N, 180° E	33	30	32	37	35	23	23	23	22	25	30	30
C. America	25° S, 120° W–25° N, 25° W	23	21	22	22	24	29	29	31	28	26	25	22
Europe	30° N, 20° W–85° N, 60° E	26	21	18	22	29	29	34	35	25	26	22	20
N. Africa	5° S, 25° W–40° N, 60° E	26	26	29	32	33	32	33	36	31	28	28	24
N. America	20° N, 180° W–80° N, 95° W	22	20	23	21	24	28	30	25	18	16	18	21
N. America	20° N, 95° W–80° N, 20° W	22	19	20	17	15	17	19	16	14	14	19	21
S. Africa	40° S, 0° E–0° N, 60° E	39	40	33	34	30	37	41	39	34	26	35	37
S. America	60° S, 110° W–0° N, 20° W	20	20	21	22	25	29	29	31	28	27	26	22

Table 3. RTGCV for mean monthly wet-day frequency (expressed in d mo^{-1}) over the interpolation domains

Region	Domain	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Australia	60° S, 90° E–12° N, 180° E	1.8	1.7	1.9	1.8	2.1	2.1	2.1	2.1	2.1	2.1	1.9	2.0
Asia	20° S, 45° E–40° N, 160° E	1.8	1.6	1.6	1.7	1.7	2.0	2.6	2.5	1.9	1.6	1.7	1.8
C. Asia	25° N, 55° E–90° N, 180° E	1.5	1.3	1.3	1.2	1.2	1.7	2.4	2.3	1.6	1.2	1.2	1.5
C. America	0° N, 120° W–20° N, 25° W	2.5	2.1	2.3	2.2	2.5	3.1	3.6	3.5	3.2	2.8	2.4	2.5
Europe	30° N, 20° W–85° N, 60° E	1.8	1.6	1.6	1.4	1.3	1.2	1.3	1.4	1.6	1.6	1.6	1.8
N. Africa	5° S, 25° W–40° N, 60° E	1.8	1.8	1.9	1.8	1.9	1.6	1.8	1.9	1.7	1.9	1.9	1.9
N. America	20° N, 180° W–47° N, 20° W	2.3	1.9	1.8	1.5	1.5	1.6	1.7	1.9	1.8	1.9	1.7	2.2
N. America	47° N, 180° W–85° N, 20° W	2.3	1.9	1.8	1.6	1.7	1.5	1.4	1.5	1.5	1.6	1.8	2.2
S. Africa	40° S, 0° E–0° N, 60° E	2.1	2.3	2.4	2.2	2.3	1.9	1.9	2.0	2.1	2.4	2.5	2.2
S. America	60° S, 110° W–0° N, 20° W	3.5	3.1	3.5	3.4	3.7	4.1	4.3	4.2	3.8	3.4	3.3	3.6

Table 4. RTGCV (expressed in °C) for mean monthly temperature over the interpolation domains

Region	Domain	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Australia	60° S, 90° E–12° N, 180° E	0.9	0.9	0.9	0.9	0.9	1.0	1.0	1.2	1.0	1.0	1.0	0.9
Asia	20° S, 45° E–40° N, 160° E	1.1	1.1	1.1	1.0	1.0	1.2	1.2	1.2	1.2	1.2	1.1	1.1
C. Asia	25° N, 55° E–90° N, 180° E	1.1	1.1	1.0	0.8	0.8	1.0	1.0	1.0	1.0	0.8	1.0	1.0
C. America	0° N, 120° W–20° N, 25° W	1.1	1.1	1.1	1.1	1.1	1.0	1.0	1.0	1.0	1.0	1.1	1.1
Europe	30° N, 20° W–85° N, 60° E	1.1	1.0	0.8	0.8	0.8	0.9	0.9	0.9	0.9	0.9	1.0	1.1
N. Africa	5° S, 25° W–40° N, 60° E	1.8	1.6	1.4	1.4	1.5	1.6	1.7	1.7	1.6	1.5	1.6	1.7
N. America	20° N, 180° W–85° N, 110° W	1.5	0.9	0.9	0.7	0.6	0.7	0.7	0.7	0.7	0.6	0.8	0.9
N. America	20° N, 110° W–85° N, 90° W	1.3	0.9	1.0	0.8	0.8	0.7	0.7	0.7	0.7	0.7	0.8	0.9
N. America	20° N, 90° W–85° N, 75° W	0.6	0.7	0.5	0.6	0.6	0.5	0.5	0.5	0.5	0.6	0.5	0.6
N. America	20° N, 75° W–85° N, 20° W	0.7	0.8	0.6	0.6	0.7	0.6	0.7	0.6	0.6	0.6	0.6	0.7
S. Africa	40° S, 0° E–0° N, 60° E	1.5	1.5	1.4	1.4	1.5	1.7	1.7	1.6	1.5	1.4	1.4	1.4
S. America	60° S, 110° W–0° N, 20° W	0.9	0.9	0.9	0.9	0.9	0.8	0.9	0.9	0.9	0.9	0.9	0.9

Table 5. RTGCV (expressed in °C) for mean monthly diurnal temperature range over the interpolation domains

Region	Domain	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Australia	60° S, 90° E–12° N, 180° E	2.2	2.3	2.3	2.2	2.2	2.2	2.3	2.6	2.3	2.2	2.2	2.1
Asia	20° S, 45° E–40° N, 160° E	2.4	2.3	2.4	2.3	2.4	2.4	2.5	2.5	2.6	2.7	2.5	2.4
C. Asia	25° N, 55° E–90° N, 180° E	1.8	1.4	1.7	1.6	1.6	1.4	1.3	1.3	1.5	1.6	1.7	1.7
C. America	0° N, 120° W–20° N, 25° W	2.5	2.6	2.5	2.6	2.4	2.2	2.1	2.1	2.1	2.2	2.4	2.6
Europe	30° N, 20° W–85° N, 60° E	1.5	1.5	1.7	1.8	2.0	2.1	2.2	2.2	2.1	2.0	1.7	1.5
N. Africa	5° S, 25° W–40° N, 60° E	3.1	3.1	3.2	3.2	3.5	3.8	3.8	3.9	3.8	3.6	3.2	3.0
N. America	20° N, 180° W–85° N, 105° W	1.3	1.4	1.4	1.4	1.4	1.4	1.4	1.5	1.5	1.4	1.3	1.3
N. America	20° N, 105° W–85° N, 85° W	1.4	1.5	1.5	1.5	1.4	1.3	1.3	1.3	1.3	1.3	1.4	1.4
N. America	20° N, 85° W–85° N, 20° W	0.9	1.0	0.9	0.9	1.0	1.0	1.0	1.0	1.0	0.9	0.8	0.9
S. Africa	40° S, 0° E–0° N, 60° E	2.8	2.9	2.8	2.8	3.2	3.7	3.7	3.6	3.2	2.9	2.6	2.7
S. America	60° S, 110° W–0° N, 20° W	1.7	1.7	1.6	1.6	1.7	1.8	1.9	1.9	1.8	1.7	1.7	1.9

Table 6. RTGCV (expressed in d mo^{-1}) for mean monthly ground frost frequency over the interpolation domains

Region	Domain	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Australia	60° S, 90° E–12° N, 180° E	0.8	0.9	1.4	2.7	3.6	4.1	4.2	3.9	3.3	2.8	1.9	1.1
Asia	20° S, 45° E–40° N, 160° E	1.9	1.6	1.8	1.5	1.0	0.7	0.5	0.5	0.9	1.4	2.0	2.0
C. Asia	25° N, 55° E–90° N, 180° E	1.8	1.5	1.8	1.6	1.2	0.8	0.5	0.7	1.1	1.5	2.0	2.1
C. America	0° N, 120° W–20° N, 25° W	2.2	1.9	1.5	1.0	0.7	0.5	0.5	0.5	0.5	0.9	1.5	2.0
Europe	30° N, 20° W–85° N, 60° E	2.5	2.3	2.5	2.3	1.6	0.8	0.4	0.6	1.1	1.9	2.4	2.6
N. Africa	5° S, 25° W–40° N, 60° E	2.5	2.1	2.3	1.7	0.8	0.7	0.7	0.6	0.6	1.5	2.1	2.5
N. America	20° N, 180° W–50° N, 95° W	2.1	1.9	2.0	1.7	1.3	1.1	0.6	0.8	1.3	1.7	1.7	1.9
N. America	50° N, 180° W–85° N, 20° W	1.5	1.3	1.8	1.8	1.8	1.4	0.8	1.2	1.7	2.1	1.4	1.3
N. America	20° N, 95° W–50° N, 20° W	0.4	0.4	0.8	1.4	1.4	1.1	0.3	0.6	1.5	1.7	1.2	0.6
S. Africa	40° S, 0° E–0° N, 60° E	0.5	0.5	0.4	0.5	1.4	2.1	2.3	1.8	0.9	0.5	0.4	0.5
S. America	60° S, 110° W–0° N, 20° W	0.6	0.5	0.7	0.9	1.2	1.3	1.3	1.3	0.9	0.9	0.7	0.8

Table 7. RTGCV for mean monthly relative humidity (%) over the interpolation domains

Region	Domain	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Australia	60° S, 90° E–12° N, 180° E	3.9	3.8	3.6	3.8	3.7	3.5	3.7	3.5	3.4	3.6	3.6	3.7
Asia	20° S, 45° E–40° N, 160° E	5.7	5.4	5.2	4.8	4.8	4.7	4.7	4.7	4.9	4.9	5.4	5.7
C. Asia	25° N, 55° E–90° N, 180° E	6.7	6.4	6.3	5.5	5.4	5.4	5.1	5.1	4.9	5.7	6.0	6.5
C. America	0° N, 120° W–20° N, 25° W	4.8	4.9	4.8	4.8	4.6	4.2	4.4	4.3	4.2	3.8	4.1	4.5
Europe	30° N, 20° W–85° N, 60° E	3.6	3.5	3.4	3.4	3.8	4.0	4.1	3.9	3.7	3.3	3.3	3.5
N. Africa	5° S, 25° W–40° N, 60° E	4.8	4.6	4.7	4.7	5.0	5.4	5.6	5.6	5.2	4.7	4.7	4.8
N. America	20° N, 180° W–85° N, 20° W	4.8	5.1	5.2	5.4	5.1	4.3	4.2	3.9	3.7	3.9	4.3	4.6
S. Africa	40° S, 0° E–0° N, 60° E	5.0	4.7	4.7	5.0	5.2	5.6	5.6	5.5	5.2	5.1	5.3	5.1
S. America	60° S, 110° W–0° N, 20° W	4.4	4.5	4.2	4.1	3.9	4.2	4.5	4.7	4.5	3.8	3.9	4.4

Table 8. RTGCV for mean monthly sunshine fraction (%) over the interpolation domains

Region	Domain	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Australia	60° S, 90° E–12° N, 180° E	5.9	6.2	5.8	6.9	6.2	6.5	6.4	6.1	6.5	6.2	6.0	6.7
Asia	20° S, 45° E–40° N, 160° E	6.6	6.5	5.9	6.0	6.1	6.9	7.5	7.1	6.8	5.8	5.9	6.9
C. Asia	25° N, 55° E–90° N, 180° E	7.3	6.4	5.3	4.8	4.9	5.8	6.4	6.2	5.7	5.1	5.9	7.1
C. America	0° N, 120° W–20° N, 25° W	7.2	7.2	6.6	6.9	7.1	7.1	7.0	6.6	6.2	6.1	6.7	7.3
Europe	30° N, 20° W–85° N, 60° E	3.7	3.7	3.1	3.0	3.3	3.7	3.7	3.4	3.3	2.9	3.0	3.5
N. Africa	5° S, 25° W–40° N, 60° E	5.1	5.0	4.3	4.3	4.5	5.1	5.3	5.3	4.7	4.5	4.3	5.1
N. America	20° N, 180° W–85° N, 20° W	4.8	5.0	5.2	5.3	4.9	4.8	4.5	4.3	4.2	4.3	4.4	5.0
S. Africa	40° S, 0° E–0° N, 60° E	5.2	5.1	5.1	5.5	5.4	5.7	5.8	6.0	5.4	5.5	5.2	5.6
S. America	60° S, 110° W–0° N, 20° W	7.5	7.7	7.6	7.1	7.0	7.7	8.1	8.3	7.6	6.8	7.0	7.9

Table 9. RTGCV (expressed in percent of the domain-mean) for mean monthly wind speed over the interpolation domains

Region	Domain	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Australia	60° S, 90° E–12° N, 180° E	46	46	47	49	51	51	50	48	49	50	49	47
Asia	20° S, 45° E–40° N, 160° E	41	38	35	33	33	34	37	37	37	38	40	42
C. Asia	25° N, 55° E–90° N, 180° E	43	40	35	31	31	33	36	36	37	38	40	43
C. America	0° N, 120° W–20° N, 25° W	41	38	36	34	34	36	37	36	35	36	37	40
Europe	30° N, 20° W–85° N, 60° E	28	27	25	24	24	25	27	28	28	29	29	29
N. Africa	5° S, 25° W–40° N, 60° E	42	40	38	37	37	38	41	42	43	45	45	44
N. America	20° N, 180° W–85° N, 20° W	25	24	21	20	19	20	21	22	22	23	24	26
S. Africa	40° S, 0° E–0° N, 60° E	60	59	60	61	61	61	62	61	61	61	61	62
S. America	60° S, 110° W–0° N, 20° W	34	34	35	36	36	37	34	32	32	31	33	34

significant when expressed in relative terms. Wet-day frequency cross validation errors vary between 1 and 4 d mo⁻¹, with largest errors over South and Central America—again due to topographic complexity—and over Asia and Central Asia in the monsoon period.

Cross validation errors for mean temperature are close to 1°C in all regions (Table 4), with lowest errors again in areas where station coverage is dense relative to the spatial complexity of temperature. Prediction errors for diurnal temperature range are greater, varying between 1 and 4°C. This is partly because diurnal temperature range is a function of both minimum and maximum temperature, which are more difficult to predict than mean temperature, but also because there are fewer diurnal temperature range stations.

Ground frost frequency, relative humidity and sunshine percent RTGCVs (Tables 6 to 8) are between 10 and 20% of the domain-wide means (absolute errors of between 3 and 7% for relative humidity and sunshine percent, and 1 to 3 d for ground frost frequency). Higher errors for ground frost frequency occur where frost is a significant factor, and hence relative errors remain low.

4. COMPARISON TO PREVIOUS CLIMATOLOGIES

As noted in Section 1, the climate data sets differ from an earlier version (NEW99) in 2 main ways: the

data are interpolated onto a finer-resolution grid (10' compared to 30') and the station data set has been expanded to include additional stations for some variables in some areas. Generally, these 2 enhancements interact to produce differences (or 'added value') between the new 10' and old 30' grids. New station data should improve the interpolation skill, especially in areas with more complex mesoscale (100 to 10 000 km²) climate, where a sparser network may have captured only the regional trend surface. A higher-resolution elevation grid enables the surface fitted to such a grid to reflect the influence of local elevation effects not resolved by a coarser grid, provided the surface coefficients incorporate this effect.

A comprehensive evaluation of differences between NEW01 and NEW99 is a non-trivial task that goes beyond the scope of this paper, and so we provide only a summary of 'added value'. As there are no global-scale data sets at this resolution, we do this primarily through comparison of NEW99 and NEW01 to station data. The comparison involves a number of steps. First, the monthly mean climate at each station is estimated from the 4 nearest grid points in NEW99 and NEW01 using an inverse-squared-distance weighting scheme. The absolute difference between each station data point and the monthly estimates are then calculated (i.e. 12 per station). These differences are summarised by computing the mean of all the monthly absolute

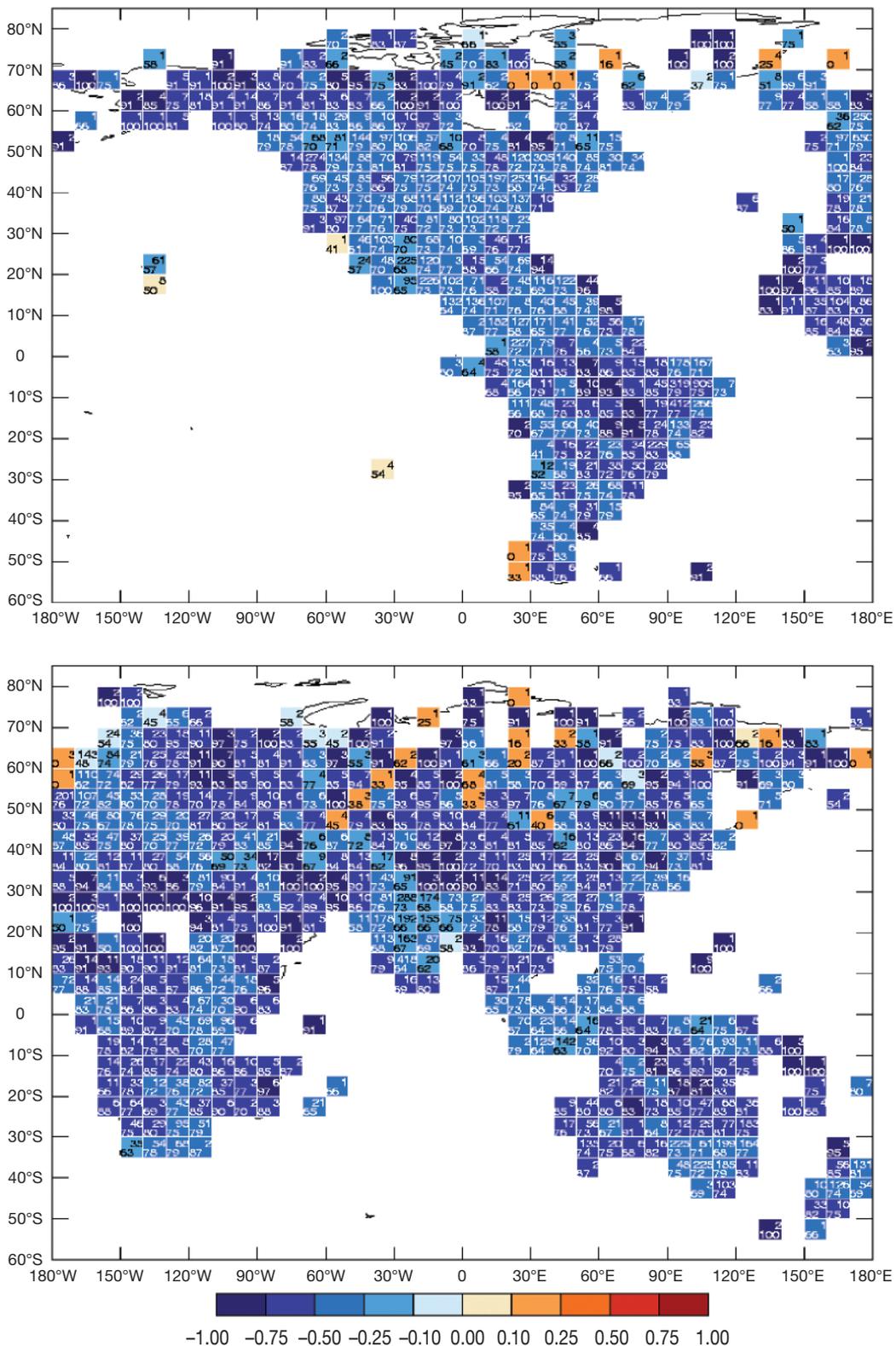


Fig. 13. Fractional change in mean absolute difference between mean monthly precipitation at stations and estimates from NEW01 and NEW99. Colours show fractional change, text in bottom left of each cell shows the percent of individual station-grid differences that show a decrease in error for NEW01 compared to NEW99 (see text for details); text in top-right of each cell shows the number of stations in each 5° grid box

differences (mean absolute difference or MAD) in 5° grid boxes over the globe and plotting the reduction (or increase) in grid box MAD when NEW01 is used instead of NEW99. This change in MAD is expressed as a fractional change for precipitation (for which the size of the absolute difference is closely tied to actual precipitation amount) and as an absolute change for other variables. In addition, we map (1) the percentage of station-estimate absolute differences that are lower for NEW01 compared to NEW99, and (2) the number of stations occurring into each 5° grid box (note that the number of estimates is 12 times the number of stations).

4.1. Precipitation

The change in MAD (as defined above) between precipitation station data and estimates from NEW01 and NEW99 are shown in Fig. 13 and summarised in Table 10. Over nearly all areas the agreement between NEW01 and station data is on average better than for NEW99, and a large proportion of 5° cells show a reduction of >25% in MAD. Moreover >66% of the individual station-to-grid differences are improved in >90% of the 5° grid boxes. There are a few instances where NEW99 is closer to the station estimates than NEW01, but in nearly all of these the MAD is small, with the result that quite small changes in MAD can result in large fractional changes. These locations also tend to have only 1 or a few stations, suggesting that in some data-sparse regions there is little difference between the coarse and fine grids, and that the additional elevational information in the finer grid is not always utilised. However, NEW01 also shows an improvement in as many data-sparse regions, but again the estimates from NEW01 and NEW99 are quite similar.

To investigate the improvement in the precipitation grids compared to NEW99 in more detail, we focus on 2 regions: the Indian sub-continent and the western USA. Over the Indian sub-continent the NEW01 interpolation benefits from ~400 additional stations, as well as the finer resolution elevation grid. The effect of this can be seen in Fig. 14, which compares July precipitation for NEW99, NEW01 fitted to the 0.5° (NEW99) elevation grid, NEW01 and the raw station data. NEW99 fails to

capture the steep increase in monsoon rainfall from the coastal plain to the western Ghats, and then the sharp drop to arid conditions in the rain shadow on the eastern (lee) side of the Ghats. This is primarily due to the sparser station network in NEW99, but is also a function of the under-representation of topography in the coarser grid. This can be seen from Fig. 14b which shows the NEW01 interpolation fitted to the coarser (NEW99) elevation grid; although the precipitation is higher along Ghats, it is still under-estimated relative to both the 10' grid and the station data (Fig. 14c,d). However, the NEW01 10' grid does not capture the full variance of rainfall along the coast exhibited by station data, with some smoothing of the steep spatial gradients and under-estimation of the highest precipitation.

Over the western USA, we adopt a similar comparison to that employed over the Indian sub-continent, but we are also able to compare NEW01 to a 'state of the art' regional climatology, the 10' version of the PRISM data set (Daly et al. 1994), downloaded from the USDA Natural Resources Conservation Services (<http://www.ftw.nrcs.usda.gov>). Note also that in this example we do not compare NEW99 to NEW01 fitted to the 0.5° grid; however, the 2 grids are very similar, suggesting that in both cases the same larger-scale climate trends are well captured.

Fig. 15 compares the grids for January. Both NEW99 and NEW01 capture the main features of precipitation over the region when compared to PRISM, including the higher rainfall over the Coast Range, the Cascades

Table 10. Summary, for all variables, of analyses undertaken to produce a diagram such as Fig. 16. Results for all 5° cells were sorted and the cell statistics at various percentiles extracted. The first figure for each percentile is the percent of NEW01 estimates of station values that are better than NEW99 estimates (i.e. bottom-left values from each cell in Fig. 16). Values in brackets are the reduction in cell MADs corresponding to that percentile (corresponding to cell colours in Fig. 16). For example, in the case of precipitation, half the 5° cells have 78% or more NEW01 estimates that are better than the NEW99 estimates, and half the cells exhibit a fractional decrease in MAD of at least 0.52. Note that for all variables except precipitation, changes in MAD are in absolute units rather than fractional changes

Variable	Percentile				
	10	25	50	75	90
Precipitation	95 (-0.8)	86 (-0.7)	78 (-0.5)	71 (-0.4)	64 (-0.2)
Wet-day frequency	100 (-1.5)	95 (-0.8)	88 (-0.4)	79 (-0.2)	69 (-0.1)
Mean temperature	100 (-0.8)	91 (-0.6)	83 (-0.4)	72 (-0.3)	60 (-0.1)
Diurnal temperature range	100 (-1.0)	95 (-0.6)	86 (-0.4)	77 (-0.2)	66 (-0.1)
Frost frequency	100 (-0.9)	94 (-0.5)	85 (-0.3)	75 (-0.1)	58 (0.0)
Relative humidity ^a	100 (-0.9)	100 (-0.5)	91 (-0.3)	83 (-0.2)	75 (-0.1)
Sunshine	95 (-2.2)	87 (-1.2)	73 (-0.6)	58 (-0.2)	35 (0.6)
Wind speed	100 (-0.7)	97 (-0.6)	84 (-0.4)	66 (-0.1)	41 (0.2)

^aRelative humidity is compared to NEW01 fitted to the 0.5° elevation grid as this variable does not form part of NEW99, which instead used vapour pressure

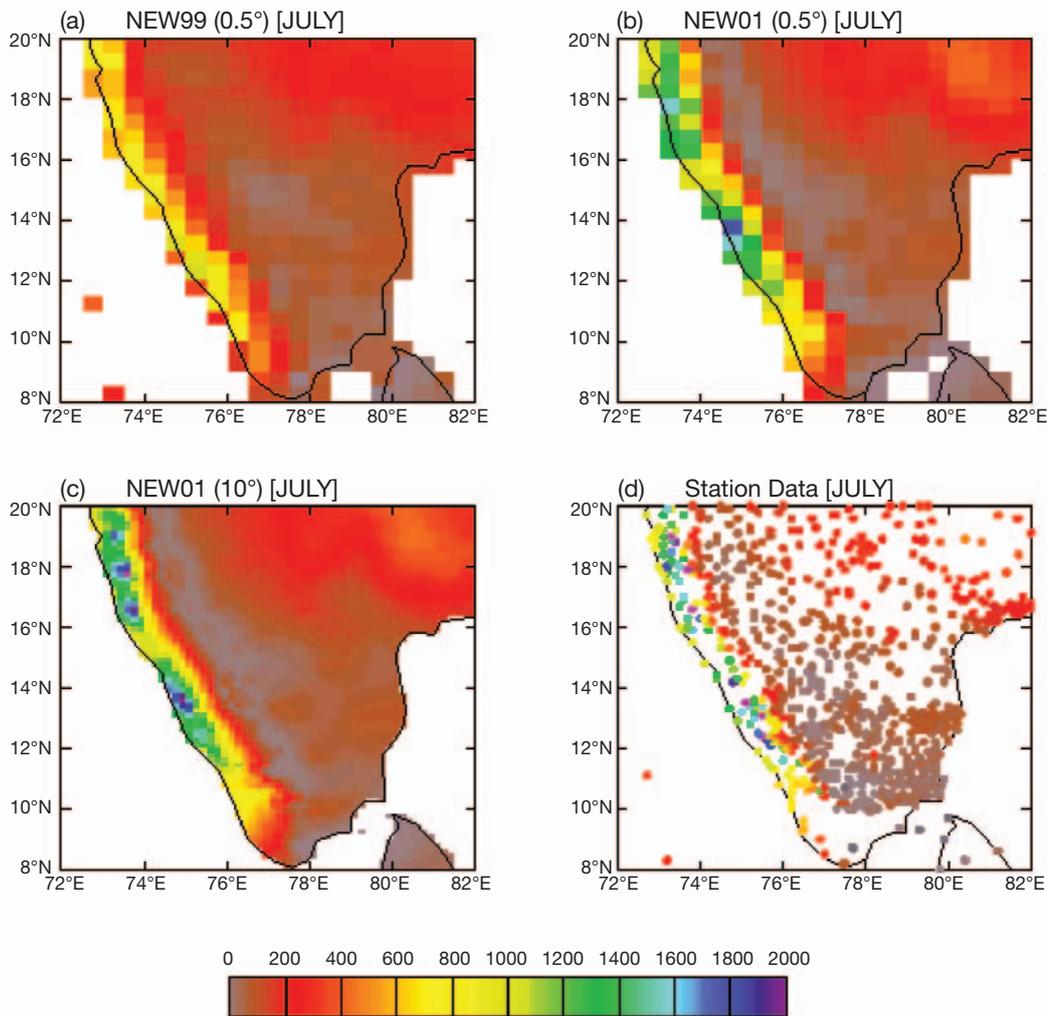


Fig. 14. Comparison of precipitation grids (mm mo⁻¹) for July over the Indian subcontinent. (a) Original 0.5° grid from NEW99; (b) NEW01 fitted to the NEW99 0.5° elevation grid; (c) NEW01 at 10°; (d) precipitation station data used to interpolate NEW01, with the colours corresponding to the July mean precipitation at these stations

and Sierra Nevada, the intervening rain shadows, and then the gradual decrease in rainfall over the Rocky Mountains and into the Great Plains. In other months these larger features are also present in all 3 data sets (not shown).

When NEW01 and PRISM are compared in more detail, it is clear that the spatial variance in the former is reduced over regions of complex topography, being underestimated in many high elevation regions. This indicates that NEW01 does not capture the full topographic control on precipitation. One reason for this lies with the station data used to fit the NEW01 surface (Fig. 15d), which also do not capture most of the high precipitation peaks in PRISM—the problem of ‘topographic bias’ of input data identified by Briggs & Cogley (1996). Moreover, if the climate-elevation relationship is non-linear, or 2 stations with similar elevation on either side of a topographic barrier have high (windward) and low (leeward) precipitation, values interpolated as a function of location and elevation (as

in NEW01) will tend to be underestimated. PRISM overcomes this problem by including (1) more stations and (2) additional predictors—such as aspect—but at the expense of computational effort. Thus in regions with dense station networks that can resolve some of the topographic complexities in precipitation, an interpolation technique like PRISM is able to produce superior results. However, on a global scale few regions have the required station density, and the computational and station-input demands of an approach such as PRISM are unsustainable. The simpler technique employed to construct NEW01 is therefore more appropriate for such global-scale interpolations.

4.2. Mean temperature

Fig. 16 shows the change in the mean station-grid differences for mean temperature when NEW01 and NEW99 are compared to station data. As with precipi-

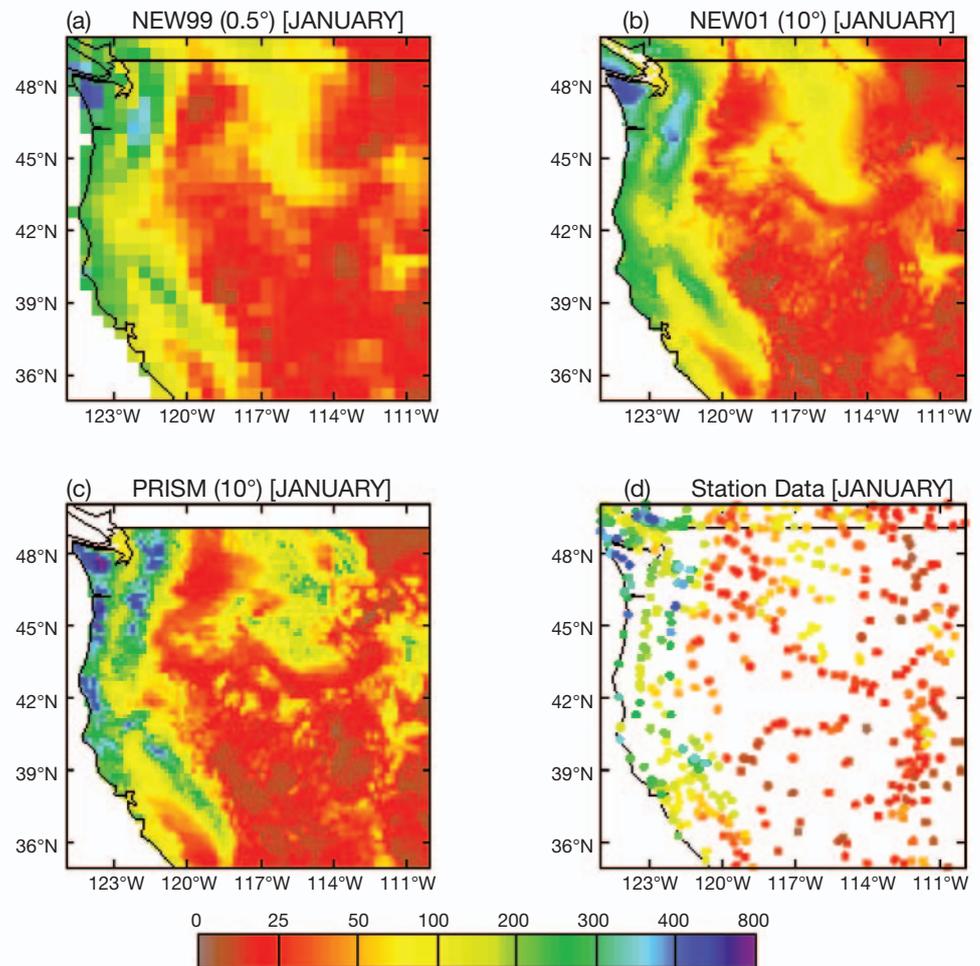


Fig. 15. Comparison of precipitation grids (mm mo^{-1}) for January over the western USA. (a) Original 0.5° grid from NEW99; (b) NEW01 at $10'$; (c) the PRISM precipitation grid at $10'$; (d) precipitation station data used to interpolate NEW01, with the colours corresponding to the January mean precipitation at these stations

tation, a large proportion of stations show closer agreement with NEW01, with only 7% of grid boxes showing an increase in MAD, and over 75% of grid-boxes showing a decrease in MAD of $>0.2^\circ\text{C}$. In contrast, the greatest increase in MAD is just $<0.2^\circ\text{C}$, and only a few 5° cells show any increase in MAD. Thus, as for precipitation, the $10'$ grids described here show an overall improvement in accuracy when compared to station observations.

4.3. Wind speed

The final detailed comparison between NEW01 and NEW99 is wind speed, which is, after precipitation, the most difficult variable to interpolate (see Table 9) and suffers from a particularly sparse station network. This variable therefore has lower potential 'added value' when interpolated to finer resolution, because the station network may not capture spatial variability at this resolution. Over much of North America, Asia and

Australia the MAD of wind speed at stations (Fig. 17) decreases when NEW01 is used (by between 0.2 and 1.00 m s^{-1} ; note that typical wind speeds vary from 2 to 8 m s^{-1} globally). This is mainly because: (1) the $10'$ grid points are closer to the station locations than the 0.5° points and will simply, by their proximity, produce a better station estimate; and (2) there is some additional skill derived from the improved elevation grid. Over Australia the decrease in MAD is quite large (generally 0.5 to 1.0 m s^{-1}), predominantly due to improved station data, with the addition of ~ 200 stations, whereas over much of Asia the improvement is a result of the finer grid resolution, as no additional stations were added in this domain. Over large regions of Africa and South America, the MAD actually increases, by up to 0.2 m s^{-1} ; these increases are relatively small and are in part due to the fitted surface being slightly smoother (lower roughness coefficient and larger station error assumed in the statistical model). Overall, wind speed in NEW01 shows an improvement in some regions and slightly less skill in others.

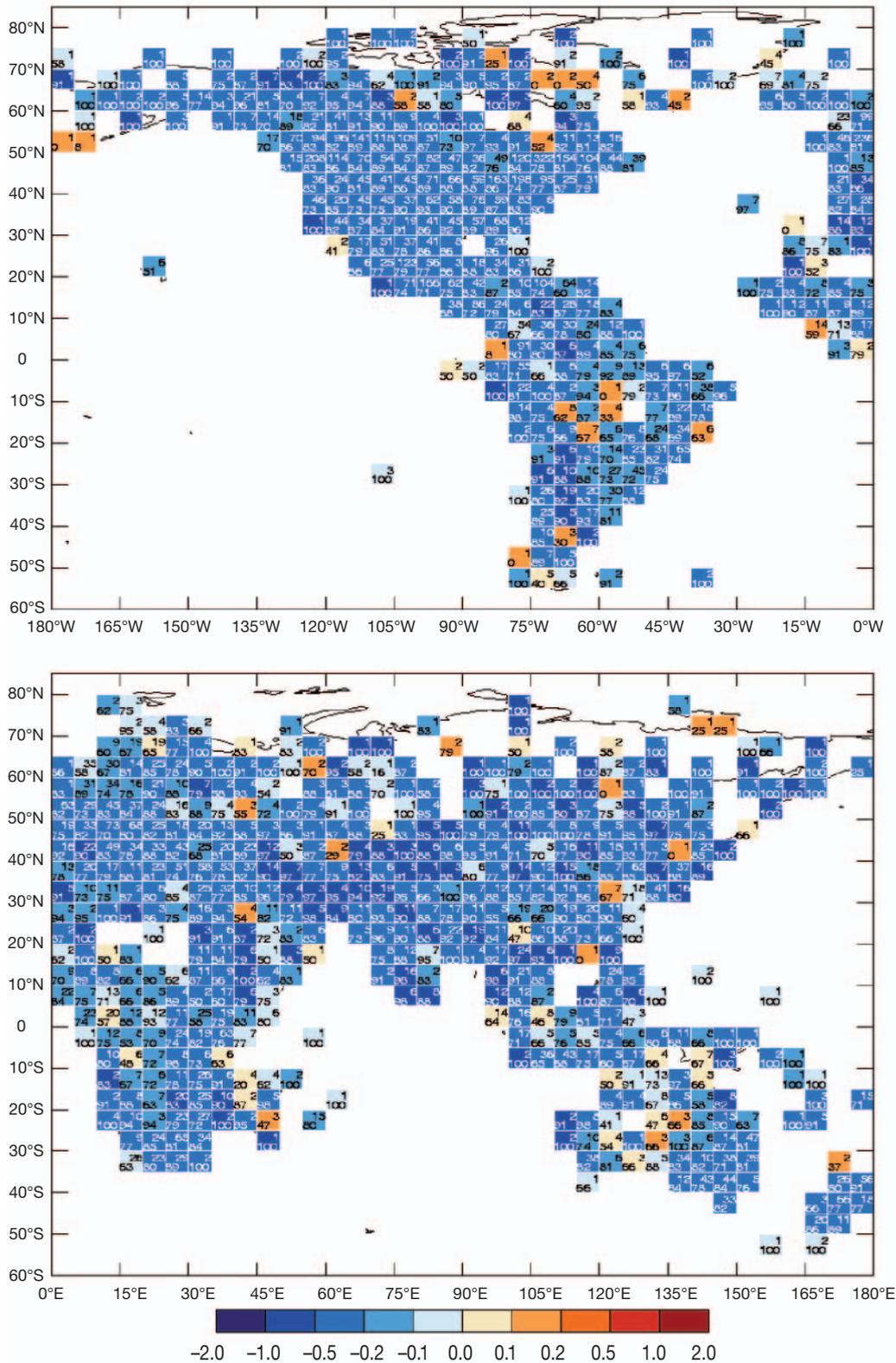


Fig. 16. Change in mean absolute difference (MAD) between mean monthly temperature from station data and estimates from NEW01 and NEW99. Colours show the change in MAD ($^{\circ}\text{C}$); text in bottom left of each cell shows the percent of individual station-grid differences that show a decrease in error for NEW01 compared to NEW99 (see text for details); text in top-right of each cell show the number of stations in each 5° grid box

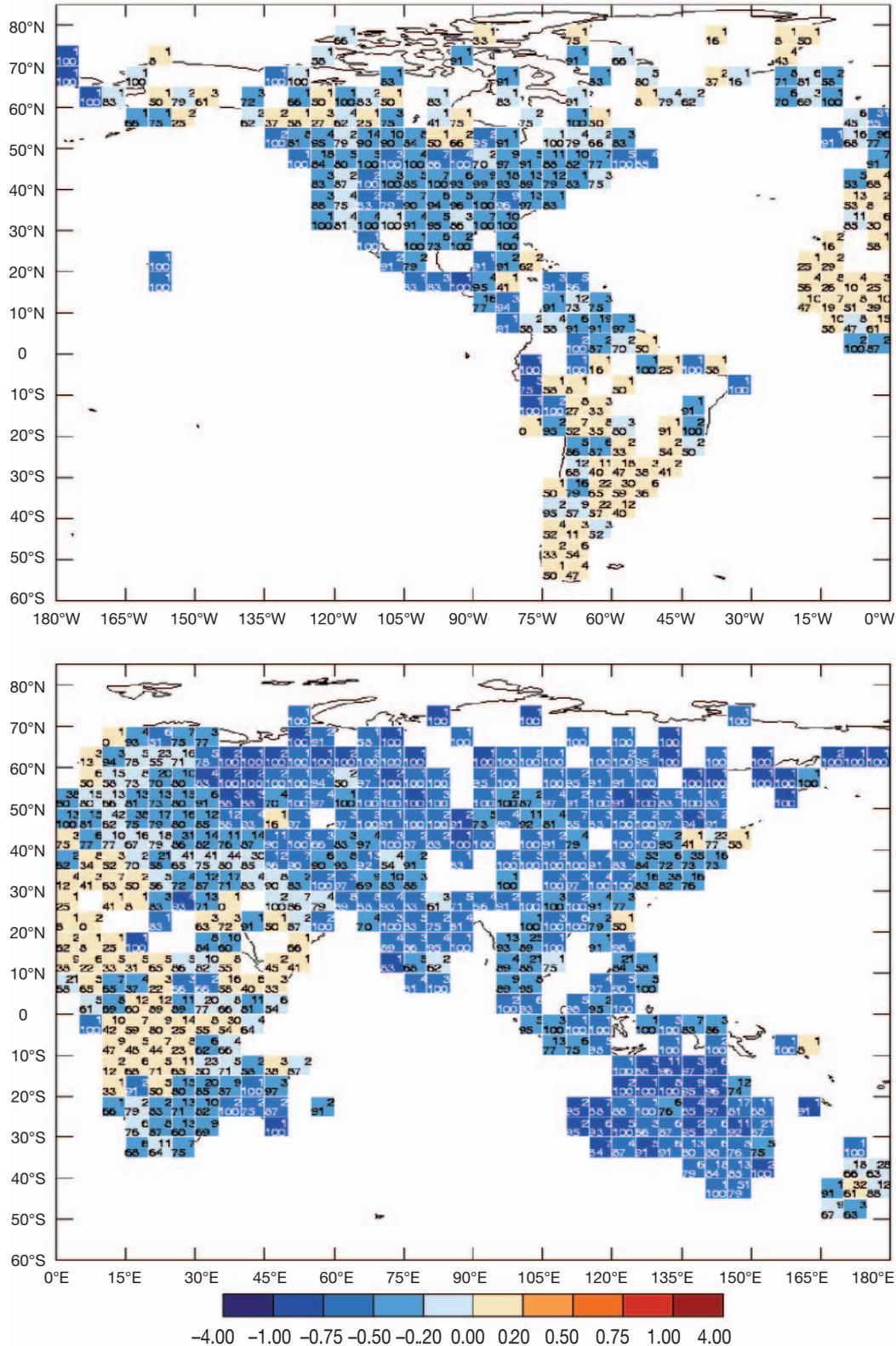


Fig. 17. Change in mean absolute difference (MAD) between wind speed station data and estimates from NEW01 and NEW99. Colours show change in m s^{-1} , text in bottom left of each cell shows the percent of individual station-grid differences that show a decrease in error for NEW01 compared to NEW99 (see text for details); text in top-right of each cell show the number of stations in each 5° grid box

4.4. Other variables

Space limitations preclude the inclusion of maps showing the changes in MAD for the other variables. However, the results of all analyses are summarised in Table 10. In all cases except wind speed and sunshine, >90% of 5° grid cells show a decrease in MAD and >50% show a decrease that is non-negligible (e.g. 50% of diurnal temperature range cells show a decrease of 0.4°C). Similarly, depending on the variable, at between 75 and 90% of cells (second-last and last columns), more than half the individual station esti-

mates derived from NEW01 are better than those derived from NEW99.

5. REGIONAL EXAMPLES

In this section, a selection of regional examples is presented to provide a flavour of the derived climate fields. A complete graphical presentation is beyond the scope of this paper because of the sheer number of climate fields, but readers are referred to the IWMI World Water and Climate Atlas website (<http://www.>

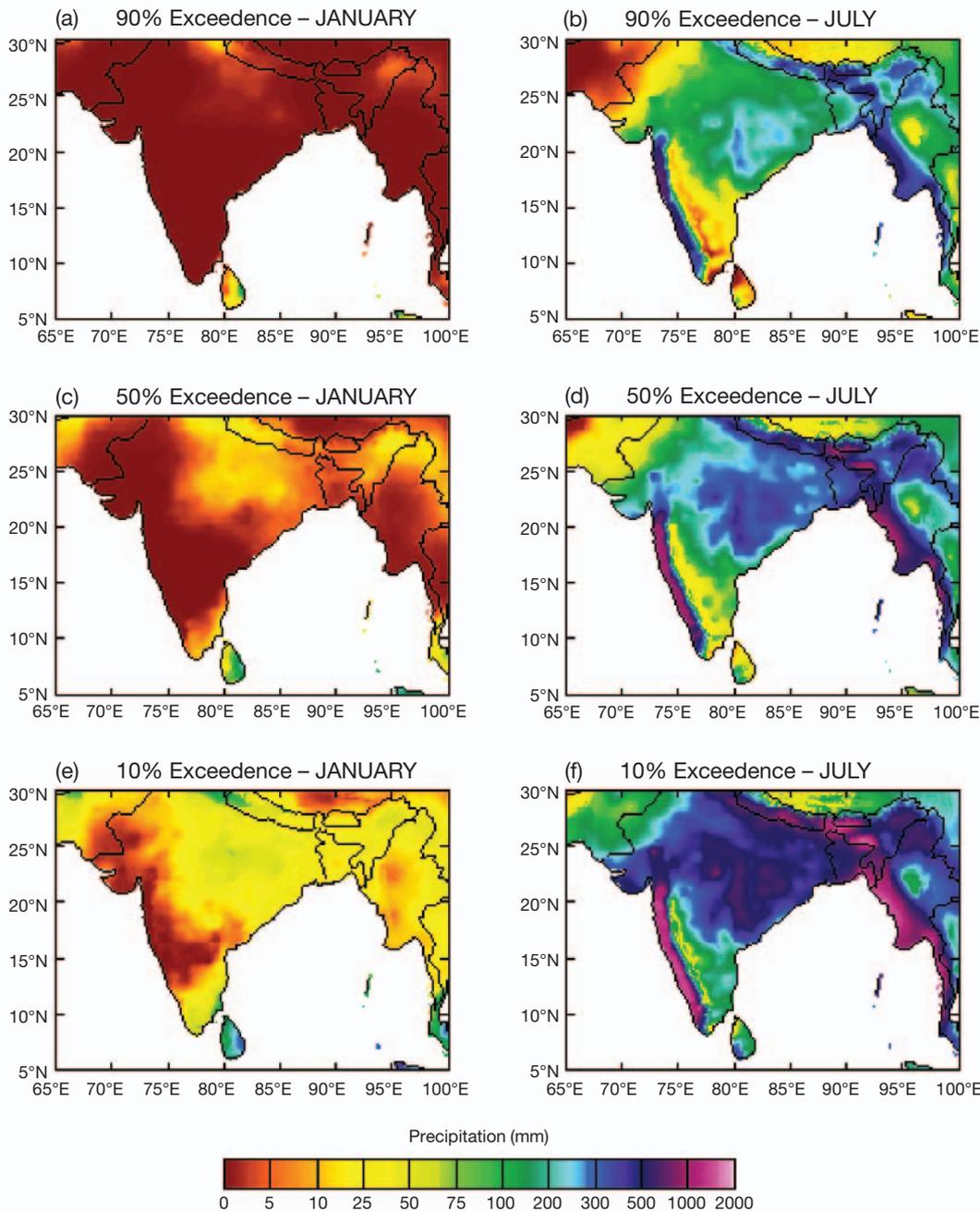


Fig. 18. Southern Asian precipitation (January and July) exceeded 10, 50 and 90% of the time (i.e. the 90th, 50th and 10th percentiles), derived using the interpolated parameters of the Gamma distribution

cgiaar.org/iwmi/watlas/atlas.htm) at which data can be visualised and downloaded (in various formats) free of charge, or the Climatic Research Unit website (<http://www.cru.uea.ac.uk>) where the gridded data are also available in raw ASCII format.

5.1. Precipitation over southern Asia

We illustrate the use of the parameters of the Gamma distribution by deriving fields of precipitation exceeded 10, 50 and 90 percent of the time (i.e. the 90th, 50th and 10th percentiles) for January and July over the south-central portion of the Asian domain. The percentiles were determined on a grid-point-by-grid-point basis using a numerical inversion of the Gamma distribution (available in most graphics and numerical packages; alternatively there are numerous approximations in the literature, e.g. Stedinger et al. 1992).

The resultant fields (Fig. 18) demonstrate the change in monthly precipitation depth at these different exceedence probabilities. In January, at the height of the dry season, the 90 % exceedence precipitation is effectively zero over much of the region, except for parts of the Himalayas. The median precipitation in January (50 % exceedence) is similar to the mean precipitation maps most readers will be familiar with, and shows the

well-known dry period over large parts of the Indian sub-continent and Myanmar, with wetter conditions over Sri Lanka and SE Asia. For precipitation with 10 % exceedence probabilities, only the western part of India and parts of Tibetan China have zero precipitation; other regions have significant non-zero precipitation, up to 250 mm in Sri Lanka and Indonesia.

In July, which is in the Asian monsoon season, precipitation is much higher than in January. The 90 % exceedence precipitation is close to zero only in Pakistan and the extreme SE (NW) of India (Sri Lanka). Over the western coastal regions of India and Myanmar the July precipitation exceeds ~300 mm 90 % of the time, the median precipitation is ~1000 mm, and the 10 % exceedence is close to 2000 mm in places. Highest precipitation occurs over the highlands around Shillong, India. The 50 % exceedence is very similar to the mean precipitation field in Fig. 14c.

5.2. Wet-day frequency over southern Asia

The interpolated mean wet-day frequency over southern Asia in January, April, July and October is shown in Fig. 19. The picture for January shows a similar pattern of spatial variability to the median precipitation (Fig. 18), with zero or only a few wet-days per

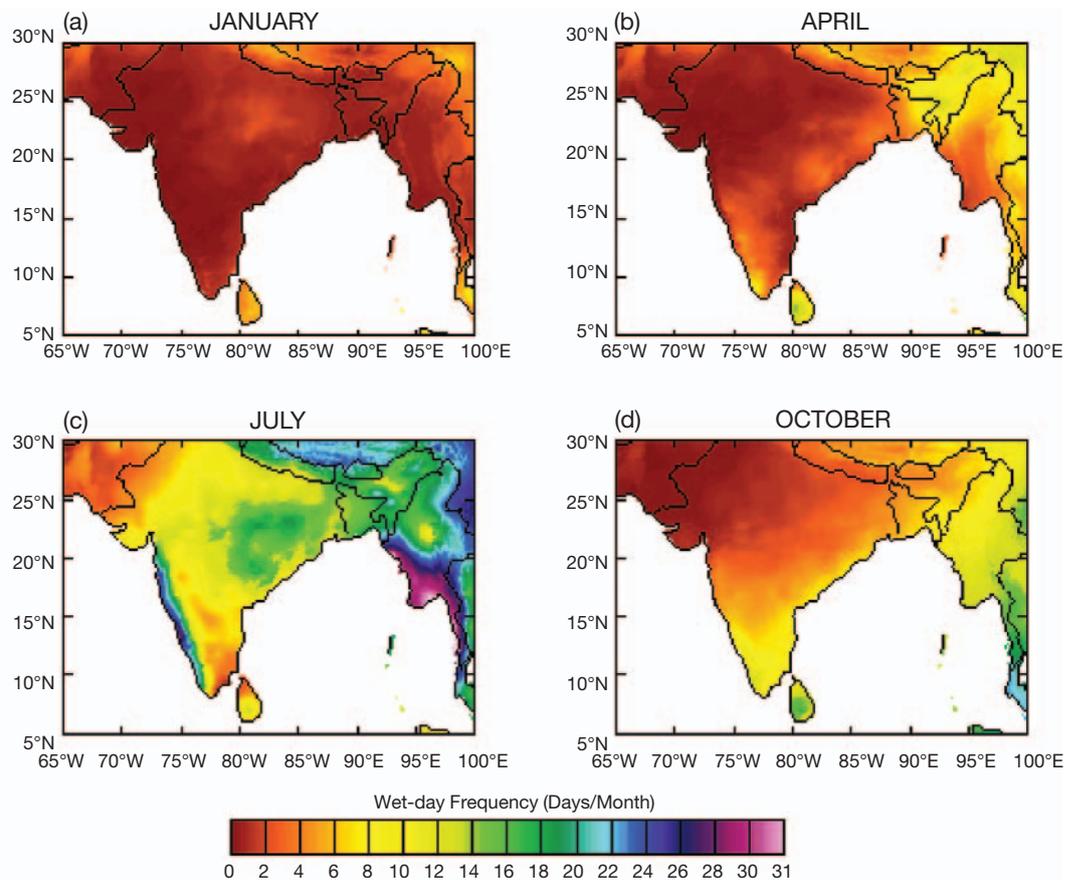


Fig. 19. Interpolated wet-day frequency over the southern Asian region

month over most of the region during this dry period. In contrast, July has higher mean wet-day frequency, up to ~20 d over western India, and up to 31 d over Myanmar. Wet-day frequency in April and October reflects the patterns in precipitation that are associated with the advance and retreat of the Asian monsoon.

These wet-day frequency fields can be used in combination with the mean precipitation fields (calculated using a simple back-computation with the parameters of the Gamma distribution—see Section 2.4) to determine mean monthly precipitation intensity in mm per wet-day. There is also the potential to use the empirical relationship between wet-day frequency and monthly precipitation derived by New et al. (2000) to calculate the number of wet-days that will typically occur in a month with precipitation of a given exceedance probability.

5.3. Mean temperature and humidity over Europe

We make use of part of the European domain to illustrate the interpolated mean temperature and relative humidity fields. The station networks used to derive these fields are shown in Figs. 4 and 8. The temperature fields (Fig. 20) show the expected seasonal progression of temperatures, with warmest temperatures of 25°C in central Spain in July, and coldest tempera-

tures of -18°C in northern Scandinavia in January. The effect of continentality is clearly visible, with coastal areas being warmer than their adjacent interiors. The strong elevation dependence of mean temperature is also portrayed, with high elevation regions such as the Alps and Pyrenees being cooler than surrounding lowlands. Elevation lapse rates are calculated implicitly during the interpolation because of the inclusion of station elevation as an independent predictor: the lapse rate therefore varies across the domain as a function of spatial variations in this elevation dependence.

The interpolated fields of monthly relative humidity over Europe (Fig. 21) show the change from relatively moist conditions in winter to drier conditions in summer. The latitudinal gradient in atmospheric moisture is also evident, as are moister conditions in western coastal regions. In spring and summer (April and July), the Alps also have higher relative humidity compared to surrounding areas; this increase with elevation is reversed in autumn and winter (October and January).

Fig. 21 also highlights some problems arising from the merging of raw relative humidity and vapour pressure station data to produce a single humidity variable (in this case relative humidity, see Section 2.3). In winter, there are marked differences in the interpolated relative humidity over Scandinavia, with values of 80 to 90% in Norway and Finland, and 95 to 100% in Sweden. The humidity data for both Finland and Norway were sup-

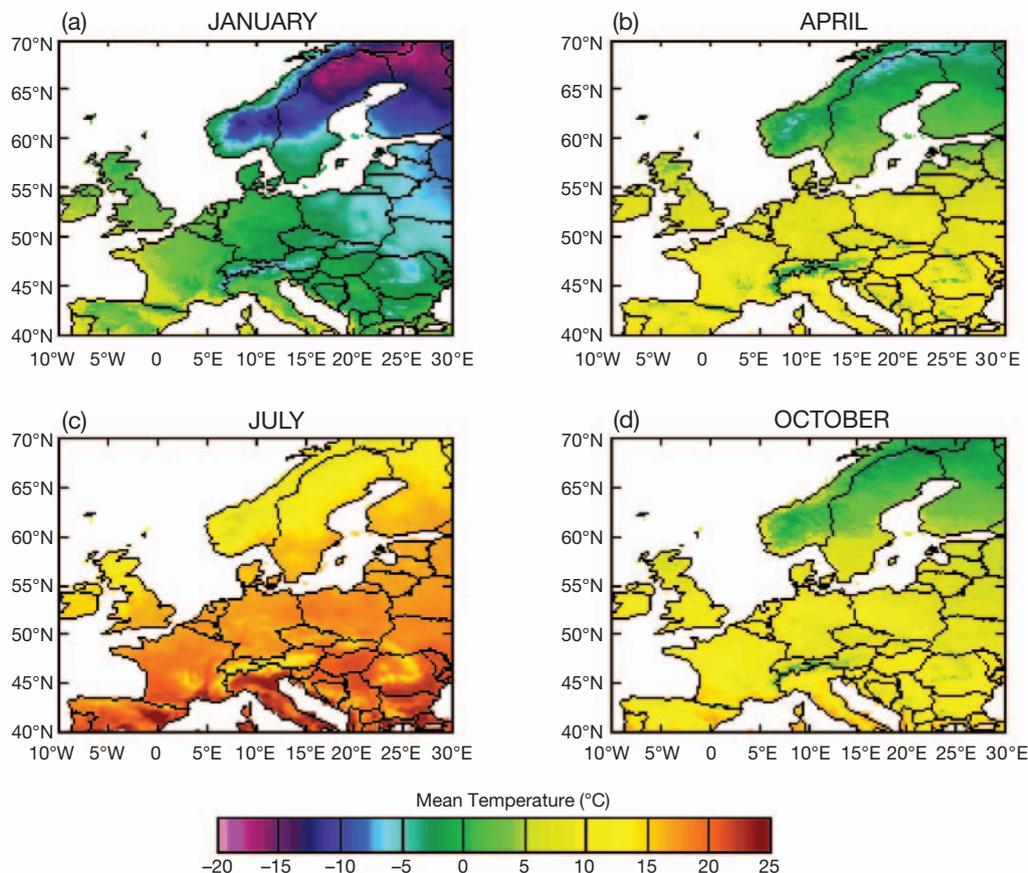


Fig. 20. Interpolated mean monthly temperature over Western Europe

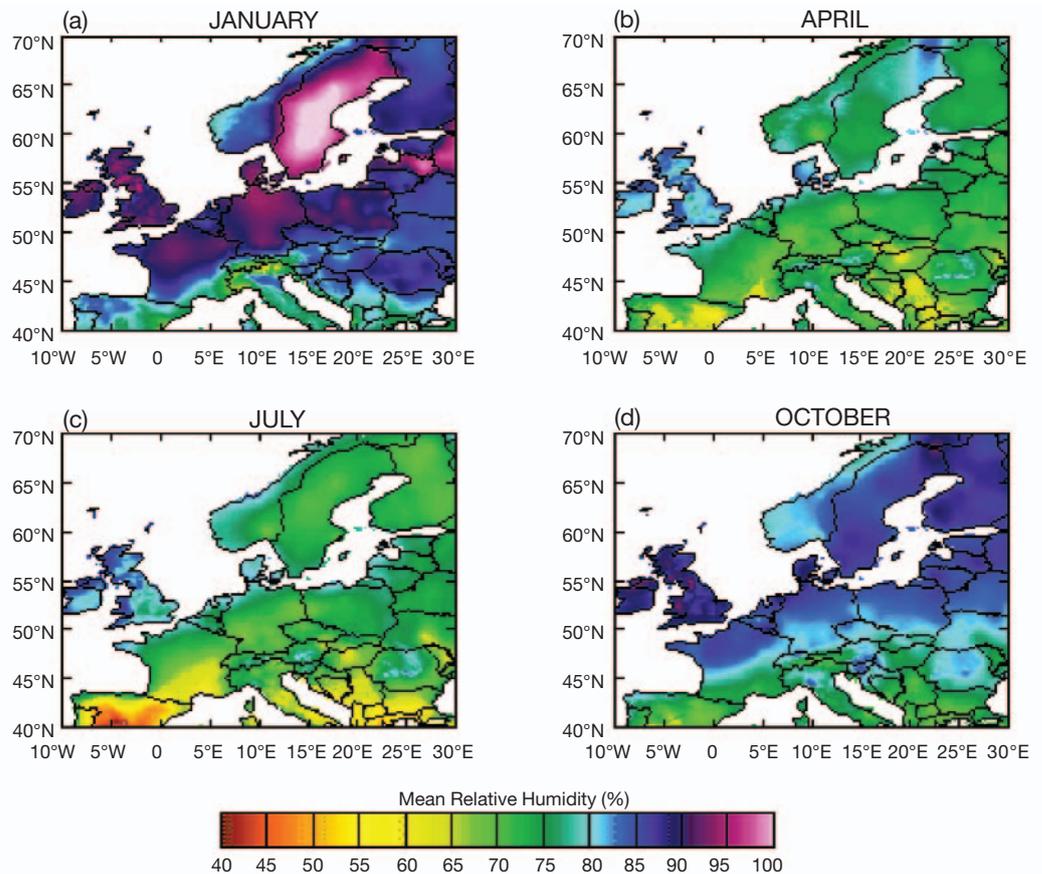


Fig. 21. Interpolated mean monthly relative humidity over Western Europe

plied as relative humidity, most likely an average of daily or sub-daily measurements. In contrast, the Swedish humidity data were provided as vapour pressure measurements and were converted to relative humidity using Eq. (1) prior to the interpolation. The higher values of calculated relative humidity for Sweden probably arise because of mismatches in the timing of raw vapour pressure (typically once or twice a day) and mean temperature (either average of maximum and minimum temperature, or the average of several readings through the day). Several reasons may act to exaggerate the biases in winter—an increased diurnality in humidity and/or temperature, and the fact that ambient temperatures are below freezing for much of the time.

5.4. Mean windspeed over Australia

Some 280 wind speed stations over mainland Australia and coastal islands contributed to the interpolation over the Australasian domain, providing reasonably good station density compared to some other regions (e.g. Russia). The interpolation has captured several large-scale features of mean wind speed in Australia, namely the southwest to northeast decrease in wind speed and the higher wind speed along the coast (Fig. 22).

6. CONCLUSIONS

We have documented the construction of a high-resolution mean monthly climatology of global land areas excluding Antarctica. The data set has a spatial resolution of 10' lat/lon, which approximates to about 18 km × 18 km at the equator, with the east-west dimension decreasing to ~16 and ~9 km at 30 and 60° N and S, respectively. This research builds on earlier work which resulted in a 30' lat/lon data set over the same domain (New et al. 1999) by including additional station data in some data-sparse areas, making use of an improved topographic data set and expressing precipitation in terms of the parameters of the Gamma distribution. The latter enables the calculation of precipitation at any exceedence probability (i.e. the monthly precipitation exceeded a given percentage of the time), which has a number of potential applications in applied hydroclimatology and agricultural meteorology. The data set comprises 8 surface climate variables—precipitation, wet-day frequency, mean temperature, diurnal temperature range, relative humidity, sunshine duration, ground frost frequency and wind speed.

The climate surfaces were interpolated from networks of station observations using thin-plate smoothing splines, with elevation, latitude and longitude as independent predictors. The density of stations varies

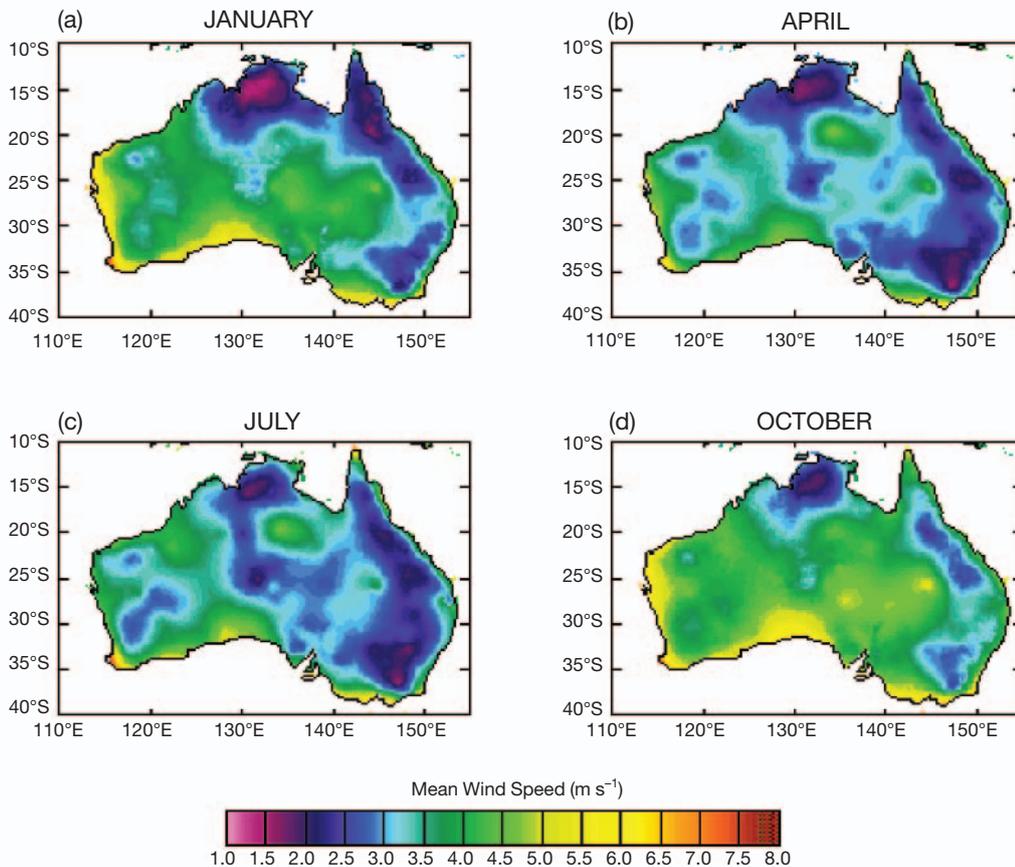


Fig. 22. Interpolated mean wind speed over Australia

both spatially and according to the climate element being mapped. In all cases, station networks are sparsest over cold, dry and mountainous regions, with concomitant increased interpolation errors. Mountainous regions are particularly prone to interpolation errors because of the more complex topographic forcing in such regions and a low-elevation (valley bottom) bias to the station network (see e.g. Briggs & Cogley 1996). Many regions of the tropics also have relatively sparse station networks and larger interpolation errors, particularly for the less commonly reported variables (sunshine, relative humidity and wind speed).

Although the data set has a spatial resolution of $10'$, and has been shown to benefit from the improved elevation information at this resolution, small-scale variability in mean climate that is not captured by the spatial and elevation dependencies in the station data will likewise not be reflected in the interpolated fields. For example, local orographic forcing over a mountain range would not be captured unless there are stations at several elevations; in the absence of such stations, elevation relationships from the larger-scale station network would be used in defining the spline surface. Similarly, any other forcings on local climate that are not reflected in elevation and lat/lon—such as aspect—

will be ignored. Thus the full benefit of the higher resolution will be realised only where the station data permit and it is important therefore that users make use of the figures showing the station networks and the cross-validation statistics to obtain a qualitative idea of the accuracy of the interpolation in different regions. Data files listing the locations of all stations used in the interpolation are also available through the IWMI World Water and Climate Atlas website (<http://www.cgiar.org/iwmi/watlas/atlas.htm>) and the CRU website (<http://www.cru.uea.ac.uk>). These can be used to determine the proximity of any particular grid point(s) to the station controls.

Acknowledgements. The development of the data sets was commissioned by IWMI with financial support provided through the United States Assistance International Development (USAID) and the Official Development Assistance of the Government of Japan. The support of these 2 donor agencies is gratefully acknowledged. The station data used in the data set have been collated over many years at the Climatic Research Unit during research funded by (among others) the UK Natural Environment Research Council, the UK Department of Environment Transport and Regions and the US Department of Energy. Data were supplied by various National Meteorological Agencies, research organisations and private individuals, whose generosity made this work possible.

LITERATURE CITED

- Andersson T, Mattison I (1991) A field test of thermometer screens. Swedish Meteorological and Hydrological Institute, Oslo
- Arnell NW (1999) A simple water balance model for the simulation of streamflow over a large geographic domain. *J Hydrol* 217:314–335
- Booth TH, Jones PG (1998) Identifying climatically suitable areas for growing particular trees in Latin America. *Forest Ecology and Management* 108:167–173
- Briggs PR, Cogley JG (1996) Topographic bias in mesoscale precipitation networks. *J Climate* 9:205–218
- Changnon SA, Kunkel KE (1999) Rapidly expanding uses of climate data and information in agriculture and water resources: causes and characteristics of new applications. *Bull Amer Meteor Soc* 80:821–830
- Cramer W, Fischer A (1996) Data requirements for global terrestrial ecosystem modelling. In: Walker B, Steffen W (eds) *Global change and terrestrial ecosystems*. Cambridge University Press, Cambridge, MA, p 530–565
- Daly C, Neilson RP, Phillips DL (1994) A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J Appl Meteor* 33:140–158
- Doorenbos J, Pruitt WO (1984) Guidelines for predicting crop water requirements. Food and Agriculture Organisation of the United Nations, Rome
- Easterling DR, Horton B, Jones PD, Peterson TC and 7 others (1997) Maximum and minimum temperature trends for the globe. *Science* 277:364–367
- Eischeid JK, Diaz HF, Bradley RS, Jones PD (1991) A comprehensive precipitation data set for global land areas. US Department of Energy, Washington, DC
- FAO (1984) Agroclimatological data for Africa. Food and Agriculture Organisation of the United Nations, Rome
- Frei C, Schar C (1998) A precipitation climatology of the Alps from high-resolution rain-gauge observations. *Int J Climatol* 18:873–900
- Giorgi F, Francisco R (2000) Uncertainties in regional climate change prediction: a regional analysis of ensemble simulations with the HadCM2 coupled AOGCM. *Clim Dyn* 16: 169–182
- Groisman PY, Rankova EY (2001) Precipitation trends over the Russian permafrost-free zone: removing the artefacts of pre-processing. *Int J Climatol* 21:657–678
- Groisman PY, Koknaeva VV, Belokrylova TA, Karl TR (1991) Overcoming biases of precipitation measurement: a history of the USSR experience. *Bull Am Meteorol Soc* 72: 1725–1733
- Hulme M (1994a) The cost of climate data—a European experience. *Weather* 49:168–175
- Hulme M (1994b) Validation of large-scale precipitation fields in General Circulation Models. In: Desbois M, Desalmand F (eds) *Global precipitation and climate change*. Springer, Berlin, p 387–405
- Hulme M, Jenkins GJ (1998) Climate change scenarios for the UK: scientific report. Climatic Research Unit, Norwich, UK
- Hulme M, Conway D, Jones PD, Jiang T, Barrow EM, Turney C (1995) Construction of a 1961–1990 European climatology for climate change modelling and impact applications. *Int J Climatol* 15:1333–1363
- Hulme M, Mitchell JFB, Jenkins J, Gregory JM, New M, Viner D (1999) Global climate scenarios for fast-track impacts studies. *Glob Env Change Suppl Iss*:S3–S19
- Hutchinson MF (1995a) Interpolating mean precipitation using thin plate smoothing splines. *Int J Geog Inf Sys* 9:385–403
- Hutchinson MF (1995b) Stochastic space-time weather models from ground-based data. *Agric Forest Meteor* 73: 237–264
- Hutchinson MF (1999) ANUSPLIN Version 4.0 user guide. Centre for Resources and Environmental Studies, Australian National University, Canberra ACT 0200, Australia
- Jones PD (1994) Hemispheric surface air temperature variability—a reanalysis and update to 1993. *J Climate* 7: 1794–1802
- Jones PD, Raper SCB, Bradley R, Diaz HF, Kelly PM, Wigley TML (1986) Northern Hemisphere surface air temperature variations: 1851–1984. *J Climate Appl Meteorol* 25:161–179
- Jones PD, New M, Parker DE, Martin S, Rigor IG (1999) Surface air temperature and its changes over the past 150 years. *Rev Geophys* 37:173–199
- Karl TR, Williams CN, Young PJ (1986) A model to estimate the time of observation bias associated with monthly mean maximum, minimum and mean temperatures for the United States. *J Climate Appl Meteorol* 25:145–160
- Legates DR, DeLiberty TL (1993) Precipitation measurement biases in the United States. *Wat Resour Bull* 29:855–861
- Müller MJ (1982) Selected climatic data for a global set of standard stations for vegetation science. W. Junk, The Hague
- NCDC (1997) Data documentation for data set World Meteorological Organisation 1961–1990 global standard normals. National Climate Data Centre, Ashville, NC
- New M, Hulme M, Jones P (1999) Representing twentieth-century space-time climate variability. Part I: Development of a 1961–90 mean monthly terrestrial climatology. *J Climate* 12:829–856
- New M, Hulme M, Jones PD (2000) Representing twentieth-century space-time climate variability. Part II: Development of 1901–1996 monthly grids of terrestrial surface climate. *J Climate* 13:2217–2238
- NGDC (2000) Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Data: Version 1.0. Available on-line from: <http://www.ngdc.noaa.gov/seg/fliers/globe-dem.shtml>
- Nicholls N (1997) Increased Australian wheat yield due to recent climate trends. *Nature* 387:484–485
- Shuttleworth, WJ (1992) Evaporation. In: Maidment DR (ed) *Handbook of Hydrology*. McGraw Hill, New York, p 4.1–4.53
- Silverman BW (1985) Some aspects of the spline smoothing approach to nonparametric regression curve fitting (with discussion). *J R Statist Soc B* 47:1–52
- Stedinger JR, Vogel RM, Foufoula-Georgiou E (1992) Frequency analysis of extreme events. In: Maidment DR (ed) *Handbook of Hydrology*. McGraw-Hill, New York, p 18.1–18.66
- USAF (1987) Station climatic summaries: Europe. United States Air Force, Environmental Technical Applications Centre
- Wahba G (1979) How to smooth curves and surfaces with splines and cross-validation. 24th Conference on the Design of Experiments. US Army Research Office, Research Triangle Park, NC, p 167–192
- Wahba G (1983) Bayesian confidence intervals for the cross-validating smoothing spline. *J R Statist Soc B* 45:133–150
- WMO (1996) Climatological normals (CLINO) for the period 1961–1990. World Meteorological Organisation, Geneva
- Yang DQ, Ishida S, Goodison BE, Gunther T (1999) Bias correction of daily precipitation measurements for Greenland. *J Geophys Res Atmos* 104:6171–6181