30-Arcsecond monthly climate surfaces with global land coverage

Citation	Mosier, T. M., Hill, D. F. and Sharp, K. V. (2014). 30-Arcsecond monthly climate surfaces with global land coverage. International Journal of Climatology, 34: 2175–2188. doi:10.1002/joc.3829
DOI	10.1002/joc.3829
Publisher	John Wiley & Sons Ltd.
Version	Version of Record
Terms of Use	http://cdss.library.oregonstate.edu/sa-termsofuse

INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **34**: 2175–2188 (2014)
Published online 18 September 2013 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/joc.3829

RMetS
Royal Meteorological Society

30-Arcsecond monthly climate surfaces with global land coverage

Thomas M. Mosier,[a*] David F. Hill[b] and Kendra V. Sharp[a]
[a] *School of Mechanical, Industrial, and Manufacturing Engineering, Oregon State University, Corvallis, OR, USA*
[b] *School of Civil and Construction Engineering, Oregon State University, Corvallis, OR, USA*

ABSTRACT: Monthly total precipitation and mean temperature climate surfaces, gridded to 30-arcseconds ($\approx$1 km at the equator) and available for all global land areas, are presented. These datasets are generated with a Delta downscaling method, using the 30-arcsecond WorldClim climatologies to scale monthly anomaly grids. For monthly mean temperature, the anomalies are constructed from both the Climate Research Unit (CRU) and Willmott & Matsuura (W&M) 0.5 degree time-series datasets, whereas for monthly precipitation Global Precipitation Climatology Centre (GPCC) data are also used. The 0.5 degree anomalies are then interpolated to the 30-arcsecond resolution. Use of piecewise cubic Hermite interpolating polynomials (PCHIP) to interpolate the anomaly grids results in more physically representative Delta downscaled surfaces, compared to bilinear and cubic spline interpolation. The Delta downscaled products are compared to Global Historical Climatology Network (GHCN) station records for six test regions distributed globally. In this analysis, the Delta grids produced using the W&M time-series dataset perform better than grids produced using GPCC or CRU. Using Oregon, USA as a test region, the Delta downscaled datasets are compared to the Parameter-elevation Regressions on Independent Slopes Model (PRISM) datasets. For monthly precipitation, PRISM performs better than each of the three Delta downscaled datasets, but for mean temperature both Delta downscaled datasets outperform PRISM. Through computing the Pearson product–moment correlation coefficient between GHCN station delineated errors in the WorldClim climatologies and the Delta downscaled W&M data, it is shown that performance of the Delta grids corresponds strongly to performance of the reference climatologies. Therefore, future improvement of the 30-arcsecond Delta grids described in this article is strongly tied to advances in the high-resolution climatological data for all global land surfaces. The Delta downscaled datasets discussed herein are open-source and freely distributed at http://www.globalclimatedata.org.

KEY WORDS downscaling; precipitation; mean temperature; monthly time-series; global climate data; high resolution

Received 30 January 2013; Revised 19 August 2013; Accepted 20 August 2013

1. Introduction

Mean monthly temperature and total monthly precipitation data are commonly used in hydrologic and agricultural studies (Döll *et al.*, 2003; Renard and Freimund, 1994). In North America, several high-spatial resolution (i.e. on the order of 30-arcseconds) monthly datasets exist, such as Parameter-elevation Regressions on Independent Slopes Model (PRISM) (Daly *et al.*, 2002), Daymet (Thornton *et al.*, 1997), and ClimateWNA (Wang *et al.*, 2012). In contrast, for many other regions of the world the highest resolution monthly precipitation and mean temperature datasets currently available have a spatial resolution of 10-arcminutes (approximately 18.5 km at the equator) (Hijmans *et al.*, 2005b), which is 20 times coarser than the Delta downscaled grids presented in this paper. The theoretical temporal coverage of the Delta downscaled datasets presented is from 1900 to 2010; however, in practice grids for the first half of the twentieth century should be used with caution because the reference climatologies utilized in the datasets' construction are for 1950–2000. In addition to this significant temporal coverage, the Delta downscaled monthly precipitation and mean temperature datasets presented herein are gridded to a spatial resolution of 30-arcseconds for all global land surfaces, are open source, and are freely available (at http://www.globalclimatedata.org).

The present datasets are produced from available gridded meteorological data using a Delta downscaling method. The Delta method, as implemented herein, requires a lower resolution monthly times-series and a high-resolution climatology as inputs, where the latter input must contain a physically representative, fine-scale distribution of the meteorological variable over the landscape. The purpose of using the Delta downscaling method, as compared to direct interpolation of low-spatial resolution sources to a higher spatial resolution, is that the Delta downscaling process incorporates high-resolution orographic effects, which are not represented in the low-resolution input grids. In this study, two Delta downscaled datasets are produced for monthly mean

* Correspondence to: Thomas M. Mosier, School of Mechanical, Industrial, and Manufacturing Engineering, Oregon State University, 204 Rogers Hall, Corvallis, OR 97331-6001, USA. E-mail: mosiert@onid.orst.edu

temperature, where one uses the 0.5 degree Climate Research Unit (CRU) (New *et al.*, 2000; Mitchell and Jones, 2005; Harris *et al.*, 2013) time-series and the other uses the analogous Willmott & Matsuura (W&M) product (Matsuura and Willmott, 2012a, 2012b) as the low-resolution time-series input. For monthly precipitation CRU, W&M, and the Global Precipitation Climatology Centre's (GPCC) 0.5 degree Full Data Product (Becker *et al.*, 2013) are used. GPCC is only available for precipitation but is included because it is well regarded (Fekete *et al.*, 2004). In all instances, the WorldClim 30-arcsecond climatologies (Hijmans *et al.*, 2005b) are used as the high-resolution reference needed for the Delta downscaling method.

This implementation of the Delta downscaling method relies entirely on gridded input data. The advantage of using gridded sources instead of beginning from spatially discrete station records is that the gridded approach leverages the work that groups (such as CRU and W&M) have done to collect station records from multiple sources and process the data. This is significant because many sources of station records have copyrights on their data (Becker *et al.*, 2013) and the raw records sometimes contain errors such as incorrect units or spatial position (Hijmans *et al.*, 2005b). While utilizing gridded products does forfeit a degree of control in the downscaling process compared to starting with station records, the decision should be viewed as a tradeoff and likely results in a greater number of station records being incorporated into the presently discussed downscaled products.

There are many sources of 0.5 degree monthly precipitation and mean temperature grids. CRU, W&M, and GPCC are used here because of their global coverage and long temporal ranges, which span almost the entire period from 1900 to the present. For a high-resolution climatology dataset, WorldClim appears to be the only source with a spatial resolution of 30-arcseconds that is available for all global land surfaces. Unlike many other climatologies, WorldClim is a 51-year climate normal, using input data for the period 1950–2000. The next highest resolution climatologies available for all global land surfaces are constructed by New *et al.* (2002) and have a spatial resolution of 10-arcminutes. As noted by Daly *et al.* (2008) WorldClim has deficiencies relative to the PRISM climatologies; most notably, WorldClim appears to overestimate leeward precipitation and underestimate windward precipitation. However, given that PRISM is not available for the majority of the world, WorldClim is, in many regions, the only dataset of 30-arcsecond climatologies.

The quality of the Delta downscaled data is assessed by comparing it to Global Historical Climatology Network (GHCN) station records (Lawrimore *et al.*, 2011; Peterson and Vose, 1997) for six test regions distributed around the world. GHCN is produced by the National Oceanic and Atmospheric Administration, through the US Department of Commerce, and provides a consolidated set of available global weather station time-series data. Various statistics, including mean absolute error (MAE) and weighted mean absolute percent error (WMAPE) are calculated between GHCN records and the corresponding Delta downscaled grid values. As a benchmark of relative quality, the same statistics are also calculated between gridded PRISM data (Daly *et al.*, 2002, 2008) and GHCN records for stations within the state of Oregon, USA. PRISM is chosen for comparative purposes because these grids are the 'official spatial climate data sets of the U.S. Department of Agriculture' (Daly *et al.*, 2008), are well regarded (Adam *et al.*, 2006; Cosgrove *et al.*, 2003), and have a similar resolution to the Delta downscaled data discussed herein.

Within the Oregon test region, use of three separate anomaly interpolation schemes in the Delta downscaling method are compared. Through this, the piecewise cubic Hermite interpolating polynomials (PCHIP) scheme is found to be the anomaly interpolation scheme that results in the most physically representative downscaled grids. This version of the Delta downscaling method is then applied for all six test regions and each of the low-resolution time-series inputs. The statistical correspondence between the GHCN station records and Delta downscaled grids are aggregated over all stations and time-series elements for each test region, and presented herein. This assesses how closely the Delta downscaled grids represent meteorological values at known station locations. As WorldClim is the only source of high-resolution climatologies, it is also important to assess the propagation of error from WorldClim to the Delta downscaled grids. This is done by mapping station delineated WMAPE values for the Pakistan test region and calculating the Pearson correlation coefficient between station WMAPE values for WorldClim and the Delta downscaled W&M data [labelled herein as 'Delta(W&M)']. Together, the analysis presented assesses variations of the Delta downscaling method, choice of input time-series dataset, and the effect of the high-resolution reference climatology on the physical representation of the downscaled data.

2. Input data

2.1. 30-Arcsecond climatological normals

Accounting for orographic effects on precipitation and mean temperature is particularly important at the 30-arcsecond resolution because many grid cells do not contain a meteorological station. In addition, the location of stations tends to be biased towards population centres and arable land, which are primarily located in valleys (Hutchinson and Bischof, 1983). Owing to these meteorological sampling realities, if station records were directly interpolated to a 30-arcsecond grid without utilizing additional parameters (e.g. elevation) as independent variables, the resulting climatologies would inherently be overly smooth and biased towards values in the local valleys.

WorldClim is produced using available station data as input to the ANUSPLIN 4.3 package (Hutchinson, 2004). The ANUSPLIN thin plate smoothing spline algorithm incorporates elevation, longitude, and latitude as independent variables in the process of fitting a climate surface through station records. The parameters related to orographic effects on climate are not limited to those used in ANUSPLIN. PRISM, e.g. includes in its regression algorithm aspect, slope, coastal proximity, and 'orographic effectiveness' (Daly *et al.*, 2008). Thus, it is expected that the PRISM climatologies represent the spatial inhomogeneity of precipitation better than WorldClim, especially in mountainous areas where aspect and slope become more significant. An advantage of WorldClim compared to PRISM, though, is that WorldClim is freely available for all global land surfaces.

Uncertainty is inherent in any procedure where meteorological values are being interpolated to locations without station records. This uncertainty is difficult to quantity precisely because there are no station records. Therefore, in the application of interpolating station records, most measures of uncertainty may better be classified as measures of sensitivity. The two methods presented by WorldClim to assess their products are (1) a procedure called data-partitioning where half the data are used in the interpolation scheme and compared to the remaining half of the data, and (2) a cross-validation technique. For most regions, the data-partitioning procedure yields a mean uncertainty of less than 2°C for mean temperature and 50 mm for total precipitation (Hijmans *et al.*, 2005b). In cross-validation, individual stations are removed in turn, the ANUSPLIN package is run on the remaining data, and the interpolated data at the removed station's location is compared to that station's observed value. Mean cross-validation errors are significantly lower for most regions of the world, with an upper magnitude of 0.4°C for mean temperature and 10 mm for total precipitation. The lower errors found using cross-validation compared to data-partitioning simply reflect the increased amount of information used to produce the climate surfaces in the cross-validation method. While both the cross-validation and data-partitioning errors are presented as absolute uncertainty, normalized uncertainty would be more illuminating, especially for precipitation, as normal precipitation and temperature regimes vary widely by global region.

2.2. 0.5 Degree historical monthly datasets

A wide variety of acceptable low-resolution monthly datasets exist which could be used as input to a downscaling method. For instance, Voisin *et al.* (2010) utilize remotely sensed data from the Tropical Rainfall Measurement Mission (TRMM) to produce high-resolution precipitation grids. TRMM has geographic coverage between 50° South and North, which excludes most of Canada or Russia and was launched in 1997. The short duration of the dataset reduces its utility as an indicator of long-term trends. There is also an ongoing discussion of TRMM's accuracy (Stampoulis and Anagnostou, 2012) and how to transform the TRMM data products to improve their physical representation (Condom *et al.*, 2011). Of particular note, Condom *et al.* (2011) summarize a finding by multiple groups that TRMM underestimates precipitation for mountainous regions. Due to TRMM's limited geographic and temporal coverage, and to avoid the extra computational steps that would be required to transform it, TRMM data are not used in this study.

Reanalysis models, such as ERA (Dee *et al.*, 2011) and National Centers for Environmental Prediction (NCEP) (Kanamitsu *et al.*, 2002), are other possible choices as low-resolution input data. These datasets are derived from physically based climate models, which generally produce multi-layered output for a host of parameters at daily or subdaily time steps. Each reanalysis product has its own strengths and weaknesses, and the quality is generally improving (Sheffield *et al.*, 2004). NCEP has a temporal coverage of 1948 through the present, but is only gridded to a spatial resolution of 2.5 degrees, which is a common resolution for products derived from General Circulation Models (GCM), but is coarser than other available sources. Additionally, the real strength of reanalysis products is for applications within atmospheric studies where forcing data are needed for multiple layers or studies that require derived variables.

Presently, the CRU and W&M 0.5 degree global datasets are used as low-resolution time-series inputs to the Delta downscaling process for both monthly precipitation and mean temperature, and GPCC is utilized as an additional time-series input for monthly precipitation. These three datasets are ranked among those that represent 'our current "state-of-the-art" understanding of global precipitation distribution' (Fekete *et al.*, 2004). The extensive temporal coverages of the CRU, W&M, and GPCC data are also useful because it allows them to be used to evaluate long-term trends.

The W&M and CRU data are comparable in many respects and differ primarily due to the precise list of station records and the interpolation schemes used. The version of the CRU data used herein is CRU TS3.10 which unlike the CRU TS2.10 dataset (Mitchell and Jones, 2005) is not explicitly homogenized by CRU; although, the station data may have been homogenized by the individual meteorological organizations providing it (Harris *et al.*, 2013). The W&M data used for precipitation and mean temperature are versions 3.02 and 3.01, respectively (Matsuura and Willmott, 2012a, 2012b).

Each of these input datasets draws on its own compiled set of station data, which are then interpolated to grids using a methodology specific to the group. Owing to this, the number of stations contributing information to an interpolated cell varies between region and dataset. An example is provided in Figure 1, which shows the average number of stations from the CRU TS3.10

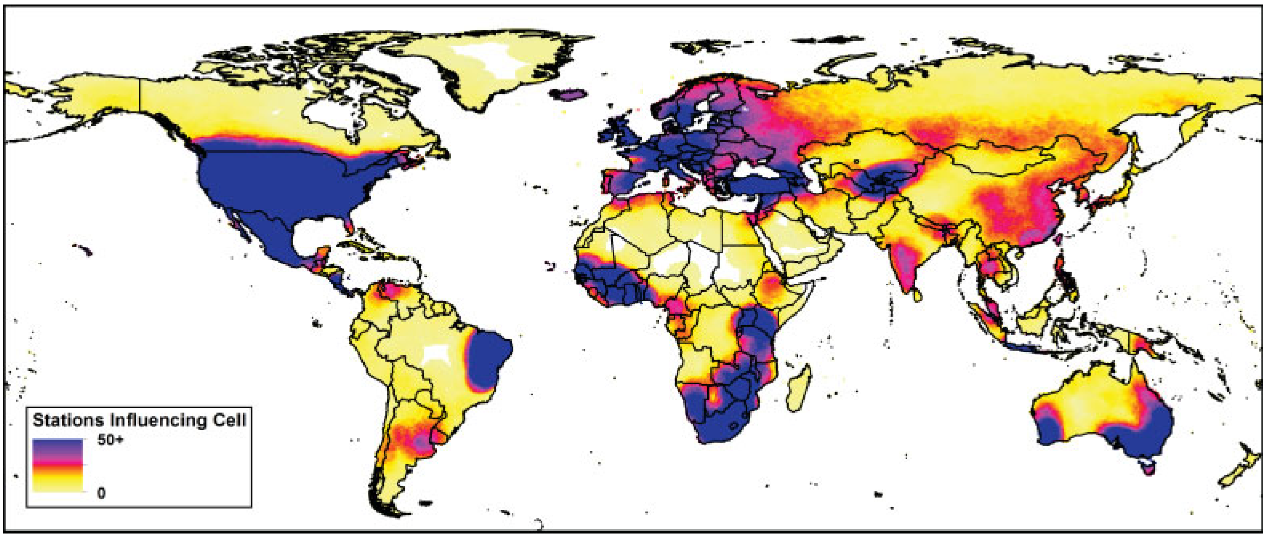

Figure 1. Average number of stations influencing the precipitation value for each cell in the CRU TS3.10 dataset from 1951 to 2000. The colour gradient saturation value is 50, although the largest number of contributing stations for a single cell is 605.

dataset influencing each interpolated cell for the years 1951–2000. From Figure 1, it is apparent that there are significant disparities in station densities between regions. For instance, USA and Europe have relatively high-station densities compared to most of Asia and South America. The trends in regional station densities seen in Figure 1 are qualitatively similar amongst the input datasets used herein (W&M: Matsuura and Willmott, 2012a, 2012b; GPCC: Becker *et al.*, 2013; WorldClim: Hijmans *et al.*, 2005b); however, the number of stations and the networks of station data utilized by each group differ, in some places significantly, as seen through figures included in the preceding dataset citations.

The methodologies used by the CRU and W&M groups to produce their 0.5 degree grids are similar to the Delta method employed herein, except that their methods are designed to extrapolate grids from discrete station data. CRU and W&M first produce 0.5 degree gridded climatologies from all available station records. Each group then calculates anomaly values at cells where station data are available for the specific time-series element being produced. These partially empty anomaly grids are then filled using an interpolation technique specific to the group. Whereas the CRU dataset is created using triangulated linear interpolation for this step (Harris *et al.*, 2013), W&M use inverse-weighting and triangular-decomposition interpolation schemes on a two-dimensional spherical surface (Willmott and Robeson, 1995).

GPCC Full Data version 6 (Becker *et al.*, 2013) is used as a third time-series input for producing Delta downscaled precipitation grids. GPCC utilizes the SPHEREMAP method developed by Willmott *et al.* (1985) to interpolate the available station data onto a grid. An apparent strength of GPCC is that their database includes significantly more stations than either CRU or GHCN (Becker *et al.*, 2013).

3. Methods

3.1. Delta downscaling procedure

The Delta downscaling method is used with the data inputs described in Section 2 to produce 30-arcsecond monthly precipitation and mean temperature grids. Figure 2 uses transects at a fixed latitude through Oregon, USA to illustrate the components and steps of the Delta downscaling process for temperature, using the W&M 0.5 degree time-series and WorldClim 30-arcsecond climatology datasets. The first step (Figure 2(a)), is to construct a 0.5 degree climatology for each month from the 0.5 degree time-series dataset. The low-resolution climatology is produced using the years from 1950 to 2000 because this is the range used by WorldClim to construct their climatology data. A 0.5 degree anomaly (blue line of Figure 2(b)) is then calculated. For temperature, the anomaly is the difference between the time-series element and corresponding low-resolution climatology whereas for precipitation, the anomaly is the ratio of the time-series element to climatology. The anomaly is then interpolated to the 30-arcsecond WorldClim grid (red line of Figure 2(b)). The final step of the Delta method (Figure 2(c)), is to transform the high-resolution anomaly back to an absolute surface through scaling it by the WorldClim climatology for the corresponding month. This transformation undoes the creation of the anomaly, and therefore addition is used for temperature, whereas multiplication is used for precipitation.

The step of interpolating the anomaly grid from the original to the high-resolution coordinates, illustrated in Figure 2(b), can be carried out using many interpolation methods. In this study, bilinear interpolation, cubic spline interpolation, and PCHIP are compared. Bilinear interpolation fits a linear function over each interval on the original grid, first in one dimension, then in the other.

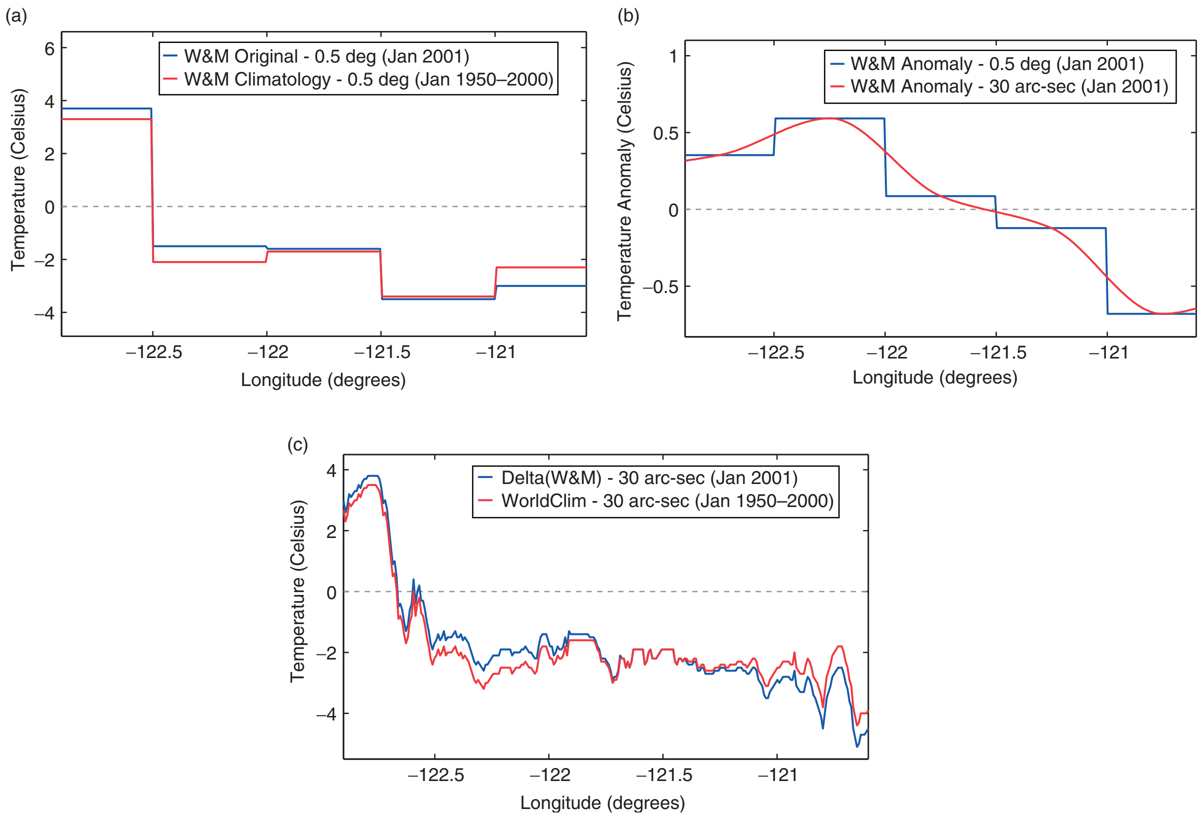

Figure 2. Transects highlighting the Delta downscaling method at latitude of 44.25° North, passing through the state of Oregon, USA. (a) Low-resolution time-series element and associated climatology. (b) Interpolating the anomaly. (c) Scaling high-resolution reference climatology.

The cubic spline interpolation scheme is Matlab's *interp2* function (MATLAB, 2011), which fits a 'natural' spline (i.e. one without a tension parameter) to each interval. The two-dimensional PCHIP function is an extension of the one dimensional form from Fritsch and Carlson (1980), using the method discussed in Press *et al.* (1992) to extend the scheme into two dimensions.

Bilinear interpolation does not allow under- or overshooting (i.e. the interpolated surface cannot have values outside the original range), but is only first order accurate. Cubic spline and PCHIP are both third-order accurate, but cubic spline interpolation allows under- and overshooting, whereas PCHIP constrains the interpolated surface to the original range. In this context, under-shooting can lead to undesirable artefacts such as negative precipitation anomalies, which necessarily produce negative precipitation surfaces and are therefore not physically representative.

It is also important to mention that while the overarching structure of the Delta downscaling method implemented here is similar to those presented in sources such as Hayhoe (2010) and Fowler *et al.* (2007), specific components of the methods are distinct because the applications are different. In both Hayhoe and Fowler's studies, the anomaly grids (also referred to as *change factors*) are calculated as the difference between past and future GCM climatologies. The anomalies are then used scale station observations, resulting in a downscaled dataset that simulates altered future conditions (Hayhoe, 2010). In this study, anomaly grids are calculated as the difference between a historical 0.5 degree time-series climate grid and a corresponding gridded climatology constructed from the same 0.5 degree time-series dataset. These anomaly grids are then scaled by a historical 30-arcsecond climatology containing orographic effects, which results in a downscaled historical time-series dataset that includes high-resolution climatic effects.

A relevant point Fowler raises about the Delta method, though, is that the method assumes 'the spatial pattern of climate will remain constant' (Fowler *et al.*, 2007), where *climate* in this instance is referring to a meteorological variable and not a multi-year climate average. In the current study, this translates to the assumption that the WorldClim climatologies represent high-resolution orographic meteorological effects equally well for all points in time. As all years of data produced using the Delta method rely on the same WorldClim climatology to transmit short-scale orographic effects, there may be instances where this assumption is weak. For example, if regional circulation changes, this may affect the orientation of the windward and leeward sides of a topographic feature.

3.2. Downscaled grid evaluation

A common measure of uncertainty in climate surface assessment is cross-validation. WorldClim and W&M distribute maps of the cross-validation error, which they calculate by systematically withholding a station from the dataset, running the data processing and interpolation algorithms, and calculating the difference between values at the station location for the datasets with the station included and withheld. Cross-validation is a good measure of sensitivity in the downscaled dataset to density and distribution of input stations; however, it requires spatially discrete station data. The input data sources (e.g. CRU and WorldClim) do not release the station records used to produce their grids, primarily to respect copyrights on the station data as discussed in Becker *et al.* (2013). Cross-validation is therefore not applicable because it is not possible to systematically withhold individual stations. As outlined in Section 1, there are multiple benefits to using gridded products in the current study, which, the authors believe, outweigh the disadvantage of not being able to perform cross-validation.

The Delta downscaled data are instead assessed by treating GHCN station data as observed values and comparing them to the corresponding downscaled grid's value for the cell that the GHCN station is within. This method does not independently verify the accuracy of the downscaled grids because all of the gridded input datasets utilized in this work contain GHCN station records as one of their input sources. GHCN is, though, the largest freely available collection of station records with a global spatial distribution. It should be understood, therefore, that utilizing GHCN data to assess the current Delta grids is the most practical option and is a measure of the Delta downscaling method's ability to retain and reproduce observed precipitation and temperature values; this assessment is valuable for understanding the Delta downscaled grids.

Comparison of 30-arcsecond Delta downscaled data at all global GHCN locations would be impractical (due to computational time), therefore the southwest portion of the Yukon Territory in Canada, Altai Republic of Russia, Germany, Oregon in the USA, Pakistan, and central Argentina are used as test regions (Figure 3). Comparison of Delta downscaled data to GHCN station records within each of these regions assesses how well the grids produced herein correspond to known values. This enables assessment of sensitivity to anomaly interpolation method, time-series input, and WorldClim climatologies, and ensures that the final product performs well.

GHCN monthly precipitation and mean temperature station data, in adjusted or non-adjusted form, are available for almost all global land surfaces. The GHCN adjustment process removes apparent shifts in the measurements that are unrelated to "true climatic variations," as described in detail by Menne and Williams (2009) and Enloe (2012). As Figure 3 shows, adjusted temperature data (version 3) are available for the entire globe, whereas adjusted precipitation data are mainly available for northern latitude regions (North America, Northern Europe, and parts of Asia). In assessing the Delta downscaled temperature grids, the adjusted GHCN records are always used. With precipitation, adjusted data are used when present; where partially available, analysis of precipitation based upon both adjusted and non-adjusted data are provided.

The earliest temperature record in the GHCN records is from 1701 (Peterson and Vose, 1997), but in this study the Delta downscaled data are compared for all months of the years beginning with 1981 and continuing through 2009. The range 1981–2009 is chosen because it is long enough to capture a significant degree of inter-annual variability and corresponds to a procured set of PRISM data. Comparison is made between the PRISM and Delta downscaled grids for the Oregon, USA test region, producing an instructive benchmark of relative quality for the high-resolution time-series datasets. The PRISM data are available at a resolution of 1.25-arcminutes, which is 2.5 times coarser than the Delta downscaled data;

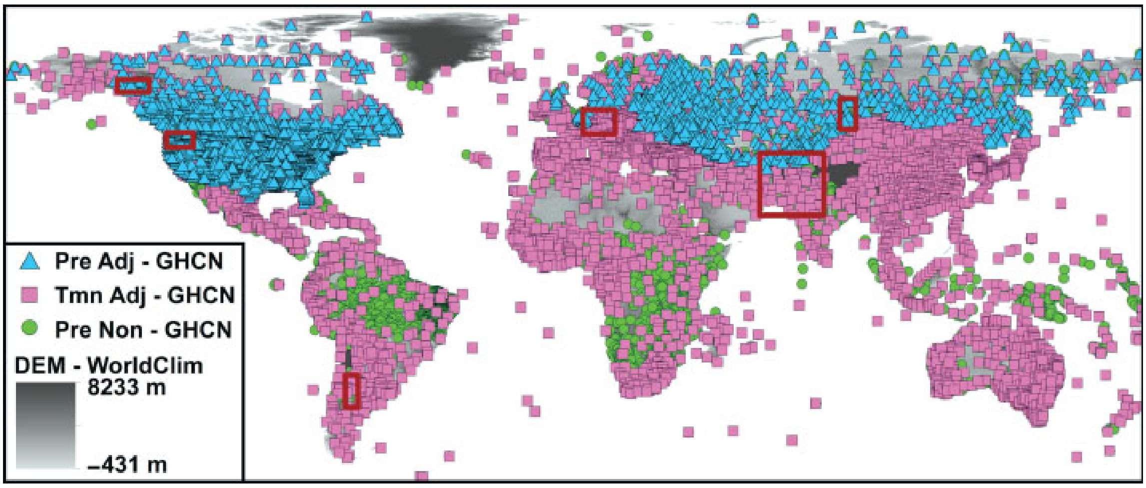

Figure 3. Global Historical Climatology Network (GHCN) station distribution and Delta data test regions (delineated with red boxes); 'Adj' refers to adjusted GHCN station records and 'Non' to non-adjusted GHCN records. The top-most layer is stations with adjusted precipitation (pre) records because these are the scarcest, followed by adjusted mean temperature (tmn), and with non-adjusted precipitation as the bottom-most layer. Due to the layering, many of the 'Tmn Adj' and 'Pre Non' locations are not visible.

Table 1. Reported statistics, where P refers to the predicted value, O to the observed value, and n is the number of elements.

Statistic	Definition
Bias	$\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)$
Mean absolute error (MAE)	$\frac{1}{n}\sum_{i=1}^{n}\lvert P_i - O_i\rvert$
Root mean square error (RMSE)	$\sqrt{\frac{\sum_{i=1}^{n}(P_i-O_i)^2}{n}}$
Mean absolute % error (MAPE)	$\frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{P_i - O_i}{O_i}\right\rvert \times 100\%$
Weighted MAPE (WMAPE)	$\frac{\sum_{i=1}^{n}\lvert O_i\rvert\times\left\lvert\frac{P_i-O_i}{O_i}\right\rvert}{\sum_{j=1}^{n}\lvert O_j\rvert} \times 100\%$
Nash–Sutcliffe efficiency (NSE)	$1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{j=1}^{n}\left(O_j - \overline{O}\right)^2}$

neither dataset is resampled though, to preserve their intended structures. A visual comparison is also carried out between PRISM and Delta downscaled time-series transects, allowing for a qualitative assessment of their respective meteorological parameter representations.

For the Oregon, USA test region, the Delta downscaled data are also compared to 30-arcsecond grids which are directly interpolated from the 0.5 degree W&M data [labelled 'DI(W&M)']. PCHIP interpolation is used to produce the DI(W&M) grids because of the advantages of PCHIP relative to bilinear or cubic spline interpolation outlined in Section 3.1. It is expected that the directly interpolated grids will perform worse than their downscaled equivalents because direct interpolation ignores all fine-scale topographic influences on precipitation and temperature. Still, it is an instructive exercise to quantify how much increased value is added through utilization of the Delta downscaling process.

3.3. Statistical formulations

Six statistics, defined mathematically in Table 1, are used to compare the performance of Delta downscaled grids relative to GHCN station records for each of the six test regions; three of the statistics are percent statistics and three are dimensional statistics. Five of the statistics are reported for precipitation whereas only the dimensional statistics are applicable for temperature due to the arbitrarily defined zero point in most temperature scales.

The first dimensional statistic reported is bias, which simply quantifies the net average difference between the modelled and observed variable. MAE is similar to bias except that MAE is the mean of the absolute errors. Also reported is the root mean square error (RMSE), which is a measure of deviation within the set. Reviews on the usefulness of RMSE as a metric are mixed, with Willmott *et al.* (2009) stating that RMSE's interpretation is conflated because its value is affected by both the mean in the error and the variability. The essential argument is that RMSE is affected by the distribution of errors within the set, the size of the set, and the average error, while the MAE is directly a measure of average model performance (Willmott and Matsuura, 2005). Nonetheless, RMSE is a common statistic across a number of disparate fields and is useful to those who are familiar with it.

MAPE uses an absolute error normalized by the observed value to form a percentage. The MAPE normalization factor causes errors in smaller observed values to be reflected more heavily than similar magnitude errors in larger observed values. A method of correcting for this is to form the WMAPE, which is constructed from MAPE by replacing the averaging factor ($1/n$) with weighting factors ($\lvert O_i\rvert$ in the numerator and $\sum\lvert O_j\rvert$ in the denominator). The formulation of WMAPE provided in Table 1 highlights the weighting factor in the definition, although for computational purposes the WMAPE formula can be simplified.

MAPE and WMAPE are equivalent in the special case when all observed values are equal and all associated errors are also equal. WMAPE will be similar to or larger than MAPE when errors occur for the wettest precipitation measurements or more extreme temperature regimes; WMAPE will be smaller than MAPE if the converse is true or if the errors are equal but the observations vary. An advantage of MAPE and WMAPE over non-normalized variants, such as MAE, is that they are more consistent across regions with different climate regimes.

The Nash–Sutcliffe efficiency (NSE), like RMSE, is a squared statistic, but with more common application within hydrology. The NSE's range is negative infinity to positive one, where a value of zero indicates that the mean of the observed data is as good as a predictor of the modelled values. NSE values less than zero indicate the modelled predictions are a worse fit than the mean observed value and an NSE of one indicates an exact fit. Legates and McCabe (1999) and Willmott *et al.* (2011) note that because the NSE is a squared metric, it is more sensitive to a fewer number of extreme errors than to small, consistent errors. Additionally, the effect of the denominator is to bias the NSE towards values closer to one in situations where variations in the observed parameter (i.e. the $\sum\left(O_i - \overline{O}\right)^2$ factor) are large compared to when the observed parameter is relatively static.

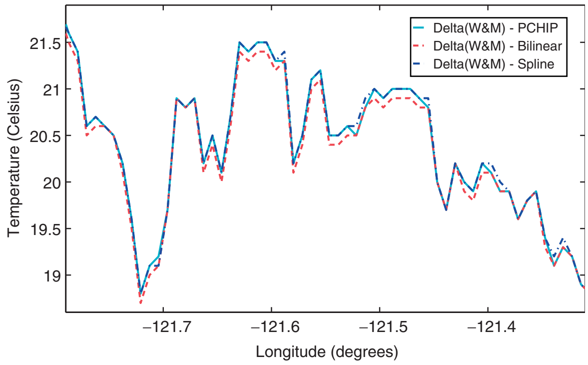

Figure 4. Delta downscaled products for July 1996 at latitude of 42.25°, comparing use of bilinear, spline, and PCHIP interpolation as described in Section 3.1.

4. Results

Results are primarily summarized as tables of statistics between the Delta downscaled grids and GHCN station records, aggregated over all station locations and all time steps for the period of years beginning with 1981 and extending through 2009. Delta(W&M) is used to denote 30-arcsecond Delta downscaled data produced using W&M as the low-resolution time-series whereas Delta(CRU) and Delta(GPCC) refer to the analogous products produced using CRU and GPCC, respectively.

Section 4.1 presents the performance of the Delta downscaled data produced using the three anomaly interpolation methods described in Section 3.1 and PRISM data for the Oregon, USA test region. Sections 4.2 and 4.3 then present the main results which are the quality of the Delta(W&M), Delta(CRU), and Delta(GPCC) monthly precipitation and mean temperature grids for the six globally-distributed test regions.

4.1. Anomaly interpolation method for Oregon

It is first necessary to establish which of the anomaly interpolation methods discussed in Section 3.1 produces the most accurate downscaled surfaces. Figure 4 shows transects of Delta downscaled temperature grids for July 1996, passing through Oregon, using bilinear, cubic spline, and PCHIP interpolation. Tables 2 and 3 compare the associated aggregated GHCN metrics for Oregon precipitation and mean temperature data produced using the Delta downscaling method, direct interpolation, and PRISM data. It is evident in both Tables 2 and 3 that the Delta downscaling method is not particularly sensitive to bilinear, cubic spline, or PCHIP interpolation. For the majority of statistics though, the Delta downscaled data produced using PCHIP interpolation perform slightly better than those produced using either bilinear or cubic spline interpolation. Therefore, for the remainder of the paper, the Delta downscaled data presented and discussed are produced using PCHIP interpolation.

For both precipitation and mean temperature, direct interpolation performs worse than each of the Delta(W&M) and PRISM datasets, which is expected per the discussion in Section 3.2. It is also evident in Table 3 that the Delta(W&M) mean temperature datasets outperform the PRISM data, which is an unexpected result. In the case of precipitation, though, the PRISM data perform better than the Delta(W&M) data.

Table 3. Analogous to Table 2 except for mean temperature grids instead of precipitation and 46 GHCN stations.

Product	Interp	Bias (°C)	MAE (°C)	RMSE (°C)
PRISM	N/A	−0.187	0.747	1.402
DI(W&M)	PCHIP	−0.974	1.358	1.682
Delta(W&M)	PCHIP	−0.175	0.611	0.783
	Bilinear	−0.179	0.616	0.787
	Spline	−0.175	0.612	0.792

4.2. 30-Arcsecond monthly precipitation surfaces

Table 4 provides aggregated statistics for each of the six test regions delineated in Figure 3 for the Delta(W&M), Delta(CRU), and Delta(GPCC) precipitation grids. GHCN adjusted precipitation records (as defined in Section 3.2) are available for the entire area of all test regions except Pakistan and Argentina. For the Pakistan test region, both adjusted and non-adjusted statistics are given because adjusted GHCN records are only available for a northern band of the region whereas for Argentina only non-adjusted GHCN records are available.

The Delta(W&M) precipitation dataset outperforms both the Delta(CRU) and Delta(GPCC) datasets on aggregated statistics for every region except Germany, where Delta(CRU) performs better than either of the others. Assessing inter-regional performance within the

Table 2. Correspondence between high-resolution precipitation grids and GHCN station records for PRISM data, directly interpolated W&M data, and Delta downscaled W&M data produced with PCHIP, bilinear, and cubic spline anomaly interpolation for the Oregon, USA test region. Statistics are aggregated over available GHCN records between the years 1981 and 2009 for 38 stations. The Interp column denotes the anomaly interpolation method used, as described in Section 3.1.

Product	Interp	MAE (mm)	RMSE (mm)	MAPE (%)	WMAPE (%)	NSE
PRISM	N/A	5.305	12.906	14.072	8.552	0.972
DI(W&M)	PCHIP	10.295	19.007	36.101	16.597	0.939
Delta(W&M)	PCHIP	8.212	15.576	30.786	13.151	0.959
	Bilinear	8.535	15.887	33.118	13.662	0.957
	Spline	8.429	15.962	32.523	13.483	0.957

Table 4. Aggregated statistics between the Delta downscaled monthly precipitation grids and GHCN station records for all data available in the range of years 1981–2009, inclusive. Only Delta statistics relative to adjusted (ADJ) GHCN records are given if adjusted data exists for the entire region. For Pakistan, there is partial adjusted coverage and for Argentina there is no adjusted coverage.

Low-res source	Region	Type	# stns	MAE (mm)	RMSE (mm)	MAPE (%)	WMAPE (%)	NSE
Willmott and Matsuura	Canada	ADJ	13	4.447	6.291	26.652	16.566	0.914
	Russia	ADJ	11	7.997	14.351	37.197	21.402	0.822
	Germany	ADJ	9	8.572	11.911	14.220	11.187	0.919
	Oregon	ADJ	38	8.212	15.576	30.786	13.151	0.959
	Pakistan	ADJ	17	5.620	11.893	36.247	19.695	0.921
		NON	97	7.549	18.551	41.964	18.016	0.946
	Argentina	NON	14	6.624	14.206	48.103	19.178	0.901
CRU	Canada	ADJ	13	10.179	14.667	92.884	37.649	0.535
	Russia	ADJ	11	10.024	16.558	66.807	26.672	0.763
	Germany	ADJ	9	6.009	10.953	8.570	7.842	0.932
	Oregon	ADJ	38	13.094	23.612	70.731	21.108	0.906
	Pakistan	ADJ	17	8.467	16.866	85.496	28.757	0.840
		NON	97	12.171	28.378	105.265	27.408	0.874
	Argentina	NON	14	9.451	20.435	110.570	27.361	0.794
GPCC	Canada	ADJ	13	6.102	8.905	38.552	22.839	0.829
	Russia	ADJ	11	9.286	15.999	44.238	25.018	0.779
	Germany	ADJ	9	12.130	16.828	19.910	15.833	0.839
	Oregon	ADJ	38	11.491	20.390	45.528	18.525	0.930
	Pakistan	ADJ	17	7.553	14.875	50.838	26.989	0.876
		NON	97	13.299	39.532	72.382	33.109	0.753
	Argentina	NON	14	8.455	18.125	52.232	24.479	0.838

Table 5. Aggregated statistics between the Delta downscaled monthly mean temperature grids and GHCN station records for all data points available in the range of years 1981–2009, inclusive.

Low-res source	Region	#stns	Bias (°C)	MAE (°C)	RMSE (°C)
Willmott and Matsuura	Canada	17	−0.142	0.557	0.815
	Russia	4	−0.214	0.553	0.814
	Germany	62	−0.130	0.642	1.423
	Oregon	46	−0.175	0.611	0.783
	Pakistan	57	−0.097	0.588	0.873
	Argentina	10	−0.521	0.950	1.555
CRU	Canada	17	−0.075	0.693	1.075
	Russia	4	−0.198	0.502	0.704
	Germany	62	−0.003	0.661	1.457
	Oregon	46	−0.105	0.659	0.849
	Pakistan	57	−0.049	0.613	0.919
	Argentina	10	−0.536	0.975	1.575

Delta(W&M) dataset requires identifying which metric best reflects a given project's needs. For example, if regional performance within the Delta(W&M) precipitation dataset is ranked by WMAPE, the Germany test region has the lowest value (11%) and the Russia region the highest (21%); whereas under inspection of the NSE the Oregon, USA region has the best value (0.959) although the Russia region's is still the worst (0.822).

4.3. 30-Arcsecond monthly mean temperature surfaces

With version 3 of the GHCN station records, adjusted temperature data are available for all regions of the world. Table 5 provides a comparison of the aggregated statistics for each of the six test regions. For monthly mean temperature performance between Delta(W&M) and Delta(CRU) as well as between regions is similar relative to performance of the analogous precipitation data. For MAE, Delta(CRU) only outperforms Delta(W&M) in the Russian test region. Within the Delta(W&M) dataset, the MAE is almost twice as large for Argentina as for Canada, Russia, or Pakistan, but is still $<1^{\circ}$C.

5. Discussion

The three possible sources of error in the Delta downscaled data are the Delta procedure, the low-resolution time-series dataset, and high-resolution reference climatologies. Section 4.1 of the results finds that error incurred through the Delta method is reduced slightly by using PCHIP for interpolating the anomaly grids, which does

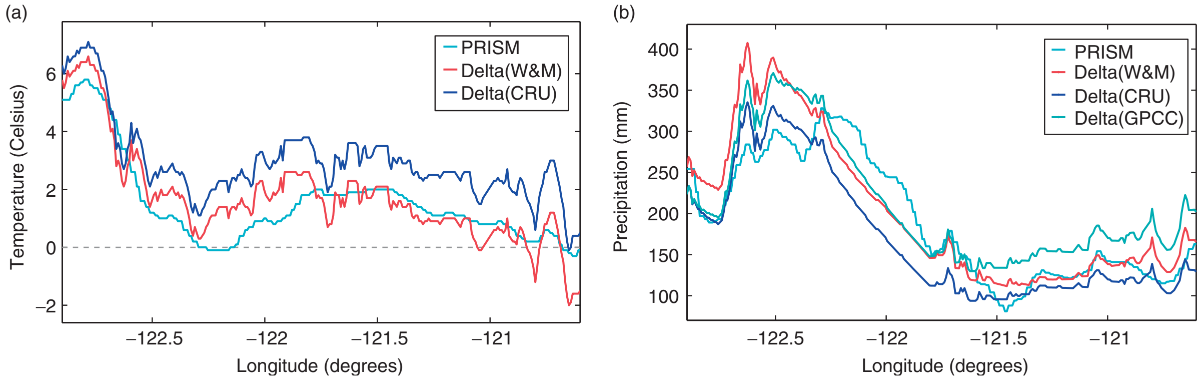

Figure 5. PRISM, Delta(W&M), Delta(CRU), and Delta(GPCC) (GPCC for precipitation only) transects for November 1998, at latitude of 42.25°, passing through Oregon, USA. The Delta data are produced using the method described in Section 3.1 and have a spatial resolution of 30-arcseconds, whereas the PRISM data are gridded at 1.25-arcminutes. (a) Monthly mean temperature. (b) Monthly total precipitation.

not require further exploration. Section 5.1 of the discussion compares Delta downscaled grid performance relative to PRISM for the Oregon, USA test region, exploring performance differences between Delta grids created using the 0.5 degree W&M, CRU, and GPCC datasets. Assessment of the Delta grid statistics relative to PRISM generally validates the Delta method and establishes a benchmark of performance. Lastly, impact of the 30-arcsecond WorldClim climatologies on Delta downscaled grids is examined by first computing WorldClim statistics relative to GHCN station records, then calculating the correlation between the station delineated WorldClim WMAPEs and corresponding Delta(W&M) WMAPEs.

5.1. Delta and PRISM performance for Oregon

Assessing the performance of PRISM grids in the Oregon test region is useful as a benchmark for the Delta downscaled data in other regions. Example transects of all high-resolution time-series products are provided in Figure 5. It is readily apparent that while the values of the Delta(W&M), Delta(CRU), and Delta(GPCC) datasets vary substantially, the shapes of the distributions are very similar. It should be recognized based upon the description of the Delta downscaling method in Section 3.1 that this is a property of the Delta downscaling method and that the quickly varying spatial structure results from the spatial distribution of the WorldClim dataset.

5.1.1. Monthly mean temperature for Oregon

Table 6 summarizes the aggregated statistics for Delta(W&M), Delta(CRU), and PRISM in the Oregon, USA test region. For RMSE, the Delta downscaled mean temperature data perform almost twice as well as PRISM. The precise reason for the Delta datasets' superior representation of monthly mean temperature is unknown. It could simply be related to differences in input data or could result from PRISM's choice of independent variables used to distribute and grid their temperature input data.

Table 6. Aggregated statistics between high-resolution time-series mean temperature grids [PRISM, Delta(W&M), and Delta(CRU)] and GHCN records for the Oregon, USA test region. Forty-six GHCN stations are used in formulating these aggregated statistics.

Metric	PRISM	Delta(W&M)	Delta(CRU)
Bias (°C)	−0.187	−0.175	−0.105
MAE (°C)	0.747	0.611	0.659
RMSE (°C)	1.402	0.783	0.849

Delta(W&M) performs better than Delta(CRU) for MAE and RMSE, although Delta(CRU) has a slightly lower bias. The difference in bias is evident in Figure 5a, which shows the Delta(CRU) temperatures to be consistently warmer than those in the Delta(W&M) dataset. As the two Delta datasets are identical except for the input time-series, this difference in bias necessarily results from differences in the W&M and CRU datasets.

5.1.2. Monthly precipitation for Oregon

Table 7 summarizes the statistical performance between high-resolution precipitation grids and GHCN station records for the Oregon, USA test region. The disparity between PRISM, Delta(W&M), Delta(CRU), and Delta(GPCC) is greater for precipitation than mean temperature. Of note, the ratio between MAPE and WMAPE varies significantly by dataset. For example, the ratio of MAPE to WMAPE for Delta(CRU) is 3.4 compared to 1.6 for PRISM. As mentioned in Section 3.3, this suggests that the PRISM errors tend to be larger for wetter months rather than drier ones whereas the errors in Delta(CRU) are relatively more even between wet and dry months.

Better performance by PRISM is expected because the PRISM algorithm accounts for aspect and slope, among other variables, whereas the ANUSPLIN package used to construct WorldClim only accounts for elevation, longitude, and latitude as independent variables. The stark contrast between Delta(W&M), Delta(CRU), and

Table 7. Aggregated statistics between high-resolution time-series precipitation grids [PRISM, Delta(W&M), Delta(CRU), and Delta(GPCC)] and GHCN records for the Oregon, USA test region. Thirty-eight GHCN stations are used in formulating these aggregated statistics.

Metric	PRISM	Delta (W&M)	Delta (CRU)	Delta (GPCC)
MAE (mm)	5.305	8.212	13.093	11.491
MAPE (%)	14.072	30.786	70.731	45.528
WMAPE (%)	8.553	13.238	21.108	18.525
RMSE (mm)	12.906	15.576	23.612	20.390
NSE (unitless)	0.972	0.959	0.906	0.930

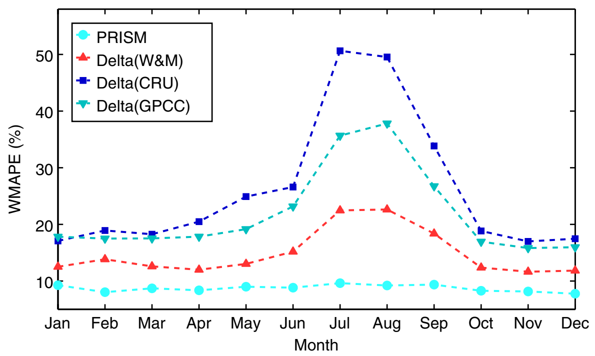

Figure 6. Monthly WMAPE for PRISM, Delta(W&M), Delta(CRU), and Delta(GPCC) precipitation grids in the Oregon, USA test region, using all GHCN station records for the years 1981–2009, inclusive.

Delta(GPCC) is not expected though, especially because Oregon has the highest precipitation gauge density of all six regions.

Figure 6 compares monthly WMAPE values aggregated over all stations for the years 1981 through 2009. It is evident that the Delta downscaled products' largest WMAPE values occur from June through September, which are the driest months in Oregon, USA. Reviewing the WMAPE formula in Table 1, greater WMAPE values for the Delta data during these months indicate that the size of the errors may remain constant whereas the sum of the observed values for the month (i.e. the WMAPE's denominator) decreases. The difference in monthly WMAPE values between the Delta downscaled datasets highlights that choice of low-resolution dataset in the Delta downscaling procedure significantly affects the quality of the resulting dataset.

Although the summer errors in Oregon are significantly higher than the winter errors, the effect of these errors on a hydrological model may be less significant. Leibowitz *et al.* (2011) and McCabe and Clark (2005) show that precipitation in many regions of Oregon is an order of magnitude greater for January than it is for July. Thus, summer runoff is largely driven by winter precipitation and spring–summer temperatures (Serreze *et al.*, 1999; Leibowitz *et al.*, 2011).

Table 8. Aggregated statistics between WorldClim precipitation grids and GHCN records for the Pakistan test region. GHCN stations with non-adjusted (Non) records are distributed throughout the region whereas those with adjusted data (Adj) are only in the northern portion of the region, as shown in Figure 3.

Metric	Adj	Non
# Stns	14	41
MAE (mm)	4.862	5.059
RMSE (mm)	10.349	13.705
MAPE (%)	41.230	26.118
WMAPE (%)	19.624	13.000
NSE	0.883	0.954

Table 9. Aggregated statistics between WorldClim mean temperature grids and the 40 GHCN stations with adjusted records within the Pakistan test region.

Bias (°C)	MAE (°C)	RMSE (°C)
−0.115	0.662	0.969

5.2. Evaluation of WorldClim for Pakistan

Pakistan is chosen as the region to evaluate WorldClim because it straddles the boundary where adjusted GHCN precipitation records exist and because it is a region of the world where WorldClim is the only source of 30-arcsecond climatologies. Section 5.2.1 establishes the aggregated degree of correspondence between WorldClim and GHCN records for the region. It is then shown in Section 5.2.2 that the station delineated WMAPE values between WorldClim and Delta(W&M) grids for the region are strongly correlated.

5.2.1. WorldClim correspondence to GHCN

Tables 8 and 9 provide aggregated statistics highlighting the similarity between the WorldClim 30-arcsecond reference climatologies and the GHCN station records for the Pakistan region. The statistical formulas are the same as those used to assess the Delta time-series grids (Sections 4.2 and 4.3), but the methodology for calculating the statistics is different because the WorldClim grids are climatologies instead of time-series. In calculating the statistics, a given GHCN station is only used if it has data for at least 75% of the time steps used to produce the corresponding WorldClim climatology. If this requirement is met for a given month of a station's records, the month's records are averaged over the years between 1950 and 2000, inclusive. Statistics are then calculated between this spatially discrete GHCN climatology and the relevant WorldClim climatology. The initial number of non-adjusted GHCN stations in the Pakistan region is 97, which is reduced to 41 through the 75% time-step threshold criteria.

In the case of non-adjusted GHCN precipitation records, WorldClim performs slightly better than the Delta(W&M) dataset (see Tables 8 and 4, respectively).

In contrast, when compared to the adjusted GHCN records, Delta(W&M) and WorldClim perform similarly for both precipitation (see Tables 4 and 8, respectively) and mean temperature (see Tables 5 and 9, respectively). The fact that the aggregated errors for WorldClim and the Delta grids are generally similar begs the question of whether the discrepancies between WorldClim and GHCN are directly correlated to the differences between the Delta grids and GHCN data, which is discussed in Section 5.2.2.

5.2.2. *Correlation of WMAPE between WorldClim and Delta(W&M)*

The distribution of station delineated WMAPE values for WorldClim and Delta(W&M) in the Pakistan region are shown in Figure 7. The stations included for WorldClim WMAPEs (Figure 7a) are those meeting the 75% criteria described in Section 5.2.1. In the map of WMAPE distribution for Delta(W&M) (Figure 7b) all stations with records for the period 1981–2009 are included, as these are the time-series elements reported in Section 4.2. The WMAPE bins used in these maps are not uniform, proceeding in increments of 8 up to 32%, with a large fifth bin extending between 32 and 145%. This scheme is chosen because the majority of station WMAPE values are less than 32%, with a few extreme values of up to 145%.

WMAPEs in the 32–145% range are present for both the adjusted and non-adjusted GHCN station records and for both the WorldClim and Delta(W&M) datasets. The majority of these large WMAPEs occur in northern Pakistan where the terrain is very mountainous and partially glaciated. There are also a few large WMAPE values at station locations away from the Himalayas in relatively topographically homogeneous areas. Although the origin of these outlying WMAPEs is not known, two plausible explanations are that they are artefacts of the WorldClim data cleaning and adjustment process or that they are related to the version of GHCN data used by WorldClim. The paper describing WorldClim was published in 2005 and states that version 2 of the GHCN data are used. The current GHCN temperature dataset is version 3.2 although version 2 is still current for precipitation. As the aggregated statistical differences between WorldClim and adjusted GHCN data are similar for both precipitation and mean temperature, it seems that the version of GHCN data used does not entirely explain the difference between the two datasets. Regardless of which explanation accounts for the differences between the WorldClim and GHCN climatologies, the two datasets have a similar spatial distribution, especially for stations with WMAPE values of 32–145% (Figure 7(a) compared to Figure 7(b)).

The hypothesis that there is a strong correspondence between magnitude of station WMAPEs in the WorldClim and Delta(W&M) datasets is confirmed by calculating the Pearson product–moment correlation coefficient (often referred to as the r value) (Lee Rodgers and Nicewander, 1988), between station WMAPEs present in both datasets. The r value is defined as

$$r = \frac{\sum_{i}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{j}^{n}\left(X_j - \overline{X}\right)^2 \sum_{k}^{n}\left(Y_k - \overline{Y}\right)^2}}, \qquad (1)$$

where n is the total number of elements being compared, the two ordered sets are denoted by X and Y, and the range of r is -1 to 1. All station delineated WMAPE values in Figure 7(a) are compared to WMAPEs in Figure 7(b), in cases where the station IDs match. The resulting r value is 0.939, suggesting that, at least for the Pakistan region, the spatial distribution of WMAPE values present in the WorldClim climatologies are passed on the Delta(W&M) data.

Evidence of a correlation between WorldClim and the resulting Delta downscaled grids does not determine whether it is desirable or undesirable for WorldClim to deviate from GHCN records at specific station locations. It does confirm, though, that the climate distributions present in the high-resolution reference climatologies are transferred to the Delta downscaled dataset. For regions such as North America where there are multiple sources of high-resolution climatologies, it is valuable to ensure a physically accurate climatology is used. For other regions, where WorldClim is the only option, it should at least be noted that the effects present in WorldClim, both good and bad, will significantly impact the resulting Delta downscaled grids.

From inspecting the figures of station distribution and cross-validation errors for WorldClim in Hijmans *et al.* (2005b), it is evident that there is not an obvious relationship between station density and dataset performance. The reason is that other factors such as elevation, aspect, and proximity to a large body of water also affect local climate (Daly *et al.*, 2008). Therefore, while maps of relative station density (such as Figure 1) may help inform a dataset user, other metrics such as cross-validation of the dataset are also worth considering. Maps of cross-validation are available for WorldClim and the gridded time-series datasets utilized as inputs herein, but are not produced for the resulting Delta data because cross-validation is primarily applicable when gridded data are being derived from station data.

6. Conclusion

The 30-arcsecond Delta(W&M), Delta(CRU), and Delta(GPCC) monthly precipitation and mean temperature datasets presented here are open source and available for all global land surfaces, with a usable temporal resolution extending back to at least 1950. Considering the set of six test regions, the Delta(W&M) dataset outperforms both Delta(CRU) and Delta(GPCC) in most cases. For the case of temperature, both Delta downscaled datasets perform well for all regions, with

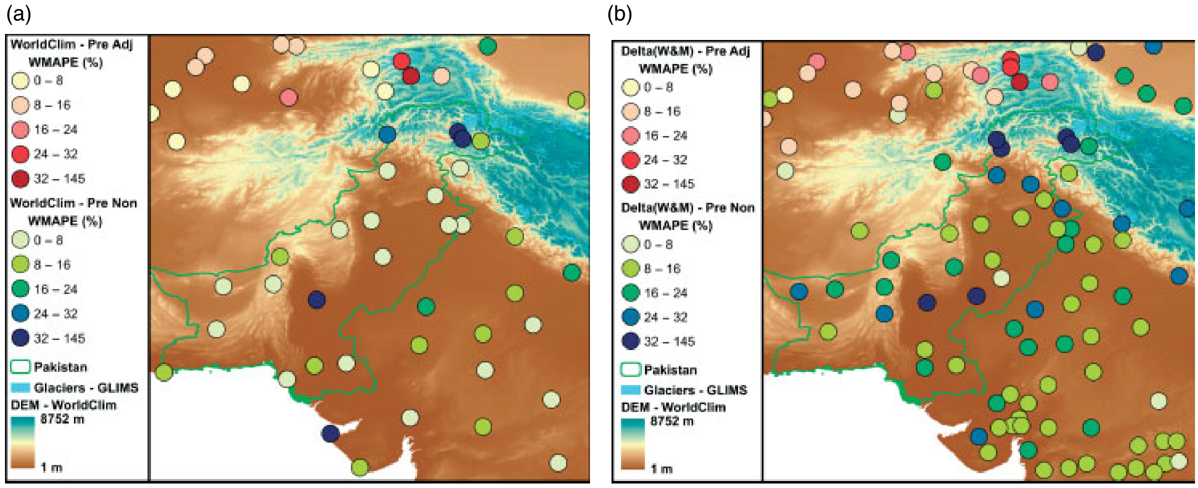

Figure 7. Station delineated WMAPE for precipitation in the Pakistan test region. (a) WorldClim WMAPE at GHCN stations for which at least 75% of the time-series records used in WorldClim are present. (b) Delta(W&M) WMAPE at all stations where any GHCN records are present. Glacier outlines are from GLIMS (Armstrong *et al.*, 2012; Raup *et al.*, 2007) and the Digital Elevation Model (DEM) is from the WorldClim data portal (Hijmans *et al.*, 2005a).

MAE values less than 0.7°C for all regions except Argentina. The regional statistics validate the Delta downscaling method as a tool for increasing the spatial resolution of gridded meteorological data while maintaining the desirable features of both the low-resolution time-series and high-resolution reference climatology inputs. In addition to considering the performance of the Delta grids relative to GHCN station data, users should also be aware of the relative station density and physiographic heterogeneities within their region of interest.

Improvements to the Delta method are of course possible and do impact the resulting data's performance. For example, using Oregon, USA as a test region it is shown that interpolating the anomaly grid with PCHIP results in better performance than using either bilinear or cubic spline interpolation. All of the Delta downscaling variations, though, are shown to be superior to directly interpolating the 0.5 degree resolution time-series data to the 30-arcsecond grid. For mean monthly temperature, both the Delta(W&M) and Delta(CRU) datasets are shown to perform slightly better than PRISM when compared to GHCN station records. In the case of precipitation, Delta(W&M) performs significantly better than Delta(CRU) and Delta(GPCC) but worse than PRISM.

The spatial distribution of WMAPE values for the WorldClim grids corresponds strongly to WMAPE values for the resulting Delta(W&M) grids in Pakistan. WorldClim is the only dataset of 30-arcsecond climatologies available for all global land surfaces; yet, the WorldClim climatologies are purported by Daly *et al.* (2008) to inaccurately represent windward and leeward precipitation. Thus, a significant limitation to improving the Delta downscaled precipitation data performance may be the availability of a more physical representative set of 30-arcsecond reference climatologies. Without WorldClim though, the Delta downscaled datasets discussed herein would not be possible for much of the Earth's land area. If another set of 30-arcsecond global climatologies becomes available, it will be straightforward to incorporate the new (or updated) climatologies into the Delta downscaling tool described herein.

Existence of a single source of 30-arcsecond precipitation and mean temperature grids is a significant addition to the research community because it allows groups to focus more resources on uses of the data rather than on producing the data themselves. The programme and inputs to create the Delta downscaled datasets described herein are freely available at http://www.globalclimatedata.org. The authors' hope is that as improvements to the dataset become possible, they will be implemented and the work will continue to be freely distributed; however, the Delta(W&M) dataset is already a strong resource for parties whose work requires high-spatial resolution monthly precipitation and mean temperature data.

Acknowledgements

The authors would like to thank the National Science Foundation (award No. 1137272) and the Glumac Faculty Fellowship for contributing funding to this project.

References

Adam JC, Clark EA, Lettenmaier DP, Wood EF. 2006. Correction of global precipitation products for orographic effects. *J. Climate* **19**(1): 15–38.

Armstrong R, Raup B, Khalsa S, Barry R, Kargel J, Helm C, Kieffer H. 2012. *GLIMS glacier database. Technical report*. National Snow and Ice Data Center: Colorado.

Becker A, Finger P, Meyer-Christoffer A, Rudolf B, Schamm K, Schneider U, Ziese M. 2013. A description of the global land-surface precipitation data products of the global precipitation climatology centre with sample applications including centennial (trend) analysis from 1901-present. *Earth Syst. Sci. Data* **5**(1): 71–99.

Condom T, Rau P, Espinoza JC. 2011. Correction of TRMM 3B43 monthly precipitation data over the mountainous areas of Peru during the period 1998–2007. *Hydrol. Processes* **25**(12): 1924–1933.

Cosgrove BA, Lohmann D, Mitchell KE, Houser PR, Wood EF, Schaake JC, Robock A, Marshall C, Sheffield J, Duan Q. 2003. Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *J. Geophys. Res.* **108**(D22): 8842.

Daly C, Gibson WP, Taylor GH, Johnson GL, Pasteris P. 2002. A knowledge-based approach to the statistical mapping of climate. *Climate Res.* **22**(2): 99–113.

Daly C, Halbleib M, Smith JI, Gibson WP, Doggett MK, Taylor GH, Curtis J, Pasteris PP. 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol.* **28**(15): 2031–2064.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P. 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. Roy. Meteorol. Soc.* **137**(656): 553–597.

Döll P, Kaspar F, Lehner B. 2003. A global hydrological model for deriving water availability indicators model tuning and validation. *J. Hydrology* **270**(1–2): 105–134.

Enloe J.2012. GHCN-Monthly version 3. http://www.ncdc.noaa.gov/ghcnm/v3.php.

Fekete BM, Vörösmarty CJ, Roads JO, Willmott CJ. 2004. Uncertainties in precipitation and their impacts on runoff estimates. *J. Climate* **17**(2): 294–304.

Fowler HJ, Blenkinsop S, Tebaldi C. 2007. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. *Int. J. Climatol.* **27**(12): 1547–1578.

Fritsch FN, Carlson RE. 1980. Monotone piecewise cubic interpolation. *SIAM J. Numer. Anal.* **17**(2): 238–246.

Harris I, Jones P, Osborn T, Lister D. 2013. Updated high-resolution grids of monthly climatic observations the CRU TS3.10 dataset. *Int. J. Climatol.*, DOI: 10.1002/joc.3711

Hayhoe K. 2010. *A Standardized Framework for Evaluating The Skill of Statistical Downscaling Methods*, PhD thesis, University of Illinois at Urbana-Champaign.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005a. Data for current conditions (∼1950–2000) WorldClim – global climate data. http://www.worldclim.org/current.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005b. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**(15): 1965–1978.

Hutchinson MF. 2004. *ANUSPLIN version 4.3, user guide*. The Australian National University, Centre for Resources and Environmental Studies: Canberra, Australia.

Hutchinson MF, Bischof RJ. 1983. A new method for estimating the spatial distribution of mean seasonal and annual rainfall applied to the Hunter Valley, New South Wales. *Aust. Meteor. Mag.* **31**(3): 179–184.

Kanamitsu M, Ebisuzaki W, Woollen J, Yang SK, Hnilo JJ, Fiorino M, Potter GL. 2002. NCEP-DOE AMIP-II reanalysis (R-2). *Bull. Am. Meteor. Soc.* **83**(11): 1631–1644.

Lawrimore JH, Menne MJ, Gleason BE, Williams CN, Wuertz DB, Vose RS, Rennie J. 2011. An overview of the global historical climatology network monthly mean temperature data set, version 3. *J. Geophys. Res.* **116**(D19).

Lee Rodgers J, Nicewander WA. 1988. Thirteen ways to look at the correlation coefficient. *Am. Stat.* **42**(1): 59–66.

Legates DR, McCabe GJ Jr. 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **35**(1): 233–241.

Leibowitz SG, Wigington PJ Jr, Comeleo RL, Ebersole JL. 2011. A temperature-precipitation-based model of thirty-year mean snowpack accumulation and melt in Oregon, USA. *Hydrol. Processes* **26**(5): 741–759.

MATLAB. 2011. *version 7.13.0 (R2011b)*. The MathWorks Inc.: Natick, MA.

Matsuura K, Willmott CJ. 2012a. *Terrestrial Air Temperature: 1900–2010 Gridded Monthly Time Series (version 3.01)*. http://climate.geog.udel.edu/ climate/.

Matsuura K, Willmott CJ. 2012b. *Terrestrial Precipitation: 1900–2010 Gridded Monthly Time Series (version 3.02)*. http://climate.geog.udel.edu/ climate/.

McCabe GJ, Clark MP. 2005. Trends and variability in snowmelt runoff in the western United States. *J. Hydrometeorol.* **6**(4): 476–482.

Menne MJ, Williams CN. 2009. Homogenization of temperature series via pairwise comparisons. *J. Climate* **22**(7): 1700–1717.

Mitchell TD, Jones PD. 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.* **25**(6): 693–712.

New M, Hulme M, Jones P. 2000. Representing twentieth-century space-time climate variability. part II: development of 1901–96 monthly grids of terrestrial surface climate. *J. Climate* **13**(13): 2217–2238.

New M, Lister D, Hulme M, Makin I. 2002. A high-resolution data set of surface climate over global land areas. *Climate Res.* **21**(1): 1–25.

Peterson TC, Vose RS. 1997. An overview of the global historical climatology network temperature database. *Bull. Am. Meteorol. Soc.* **78**(12): 2837–2849.

Press WH, Flannery BP, Teukolsky SA, Vetterling WT. 1992. Numerical recipes in FORTRAN 77: *The art of Scientific Computing*, Vol. **1**. Press Syndicate of the University of Cambridge: New York, NY.

Raup B, Racoviteanu A, Khalsa SJS, Helm C, Armstrong R, Arnaud Y. 2007. The GLIMS geospatial glacier database: a new tool for studying glacier change. *Global Planet. Change* **56**(1–2): 101–110.

Renard KG, Freimund JR. 1994. Using monthly precipitation data to estimate the r-factor in the revised USLE. *J. Hydrol.* **157**(1–4): 287–306.

Serreze MC, Clark MP, Armstrong RL, McGinnis DA, Pulwarty RS. 1999. Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data. *Water Resour. Res.* **35**(7): 2145–2160.

Sheffield J, Ziegler AD, Wood EF, Chen Y. 2004. Correction of the high-latitude rain day anomaly in the NCEP-NCAR reanalysis for land surface hydrological modeling. *J. Climate* **17**(19): 3814–3828.

Stampoulis D, Anagnostou EN. 2012. Evaluation of global satellite rainfall products over continental Europe. *J. Hydrometeorol.* **13**(2): 588–603.

Thornton PE, Running SW, White MA. 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. *J. Hydrol.* **190**(3): 214–251.

Voisin N, Schaake JC, Lettenmaier DP. 2010. Calibration and downscaling methods for quantitative ensemble precipitation forecasts. *Weather Forecasting* **25**(6): 1603–1627.

Wang T, Hamann A, Spittlehouse DL, Murdock TQ. 2012. ClimateWNA–high-resolution spatial climate data for western north america. *J. Appl. Meteorol. Climatol.* **51**(1): 16–29.

Willmott CJ, Matsuura K. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Res.* **30**(1): 79.

Willmott CJ, Robeson SM. 1995. Climatologically aided interpolation (CAI) of terrestrial air temperature. *Int. J. Climatol.* **15**(2): 221–229.

Willmott CJ, Rowe CM, Philpot WD. 1985. Small-scale climate maps: a sensitivity analysis of some common assumptions associated with grid-point interpolation and contouring. *Am. Cartogr.* **12**(1): 5–16.

Willmott CJ, Matsuura K, Robeson SM. 2009. Ambiguities inherent in sums-of-squares-based error statistics. *Atmos. Environ.* **43**(3): 749–752.

Willmott CJ, Robeson SM, Matsuura K. 2011. A refined index of model performance. *Int. J. Climatol.* **32**(13): 2088–2094.

