RMetS

Royal Meteorological Society

High-resolution precipitation mapping in a mountainous watershed: ground truth for evaluating uncertainty in a national precipitation dataset

Christopher Daly,[a]* Melissa E. Slater,[b] Joshua A. Roberti,[b] Stephanie H. Laseter[c] and Lloyd W. Swift Jr[c]

[a] *Chemical, Biological, and Environmental Engineering, Oregon State University, Corvallis, OR 97331, USA*
[b] *Science Unit, National Ecological Observatory Network (NEON), Boulder, CO 80301, USA*
[c] *US Forest Service, Southern Research Station, Coweeta Hydrologic Lab, Otto, NC 28763, USA*

ABSTRACT: A 69-station, densely spaced rain gauge network was maintained over the period 1951–1958 in the Coweeta Hydrologic Laboratory, located in the southern Appalachians in western North Carolina, USA. This unique dataset was used to develop the first digital seasonal and annual precipitation maps for the Coweeta basin, using elevation regression functions and residual interpolation. It was found that a 10-m elevation grid filtered to an approximately 7-km effective wavelength explained the most variance in precipitation ($R^2 = 0.82–0.95$). A 'dump zone' of locally high precipitation a short distance downwind from the mountain crest marking the southern border of the basin was the main feature that was not explained well by the precipitation–elevation relationship.

These data and maps provided a rare 'ground-truth' for estimating uncertainty in the national-scale Parameter-elevation Relationships on Independent Slopes Model (PRISM) precipitation grids for this location and time period. Differences between PRISM and ground-truth were compared with uncertainty estimates produced by the PRISM model and cross-validation errors. Potential sources of uncertainty in the national PRISM grids were evaluated, including the effects of coarse grid resolution, limited station data, and imprecise station locations.

The PRISM national grids matched closely (within 5%) with the Coweeta dataset. The PRISM regression prediction interval, which includes the influence of stations in an area of tens of kilometres around a given location, overestimated the local error at Coweeta (12–20%). Offsetting biases and generally low error rates made it difficult to isolate major sources of uncertainty in the PRISM grids, but station density and selection, and mislocation of stations were identified as likely sources of error. The methods used in this study can be repeated in other areas where high-density data exist to gain a more comprehensive picture of the uncertainties in national-level datasets, and can be used in network optimization exercises.

KEY WORDS   Coweeta; precipitation mapping; rain gauge; precipitation measurement; uncertainty; PRISM; orography; interpolation

Received 18 August 2016; Revised 5 December 2016; Accepted 6 December 2016

1. Introduction

Spatial climate data sets in digital form are highly valuable, as more and more climate-driven modelling and analysis activities are performed within spatially explicit computing environments. Precipitation is among the most commonly required climate elements, and is used in a wide variety of applications in hydrologic and ecosystem modelling (e.g. Beven, 2004; Elliott *et al.*, 2015), irrigation management (e.g. Allen *et al.*, 1998), climate monitoring (e.g. Alexander *et al.*, 2006), and many others. Precipitation inputs are often provided by interpolated and gridded datasets that cover the conterminous United States at resolutions of about 0.8–5 km. These include long-term climatological averages, as well as monthly and daily time series (Daly *et al.*, 1994, 2002, 2004, 2008; Thornton *et al.*, 1997; Maurer *et al.*, 2002; Hamlet and Lettenmaier, 2005; Livneh *et al.*, 2013, 2015; Newman *et al.*, 2015).

It is difficult to quantify the uncertainties in these interpolated climate data, because the true precipitation fields are usually unknown. One of the most common methods of estimating uncertainty is through cross-validation (C-V; Willmott and Matsuura, 1995; Gyalistras, 2003). In the common practice of jackknife C-V, the process of removal and estimation is performed for each station one at a time, with the station returned to the data set after estimation. Once the process is complete, overall error statistics, such as mean absolute error (MAE), bias, and others are calculated (e.g. Willmott *et al.*, 1985; Legates and McCabe, 1999). The obvious disadvantage to C-V error estimation is that no error information is provided for locations where
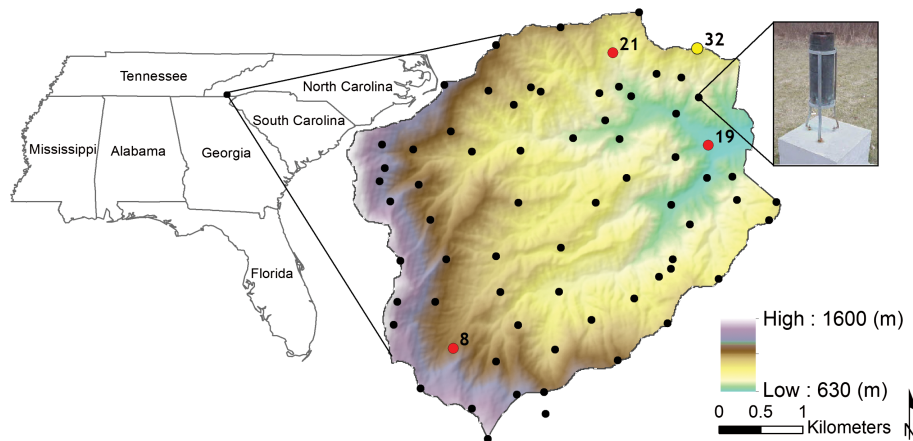
Figure 1. Map of the Coweeta basin showing locations of the 69 rain gauges on the 10-m digital elevation model. Gauges 8, 19, and 21 (red dots), were part of the National Weather Service Cooperative Observer Program (http://cdo.ncdc.noaa.gov/CDO/cdo) during the study period (see Section 4.2). Gauge 32 (northeast boundary of basin, yellow dot) was unusually dry in winter, and is discussed further in Section 3.1.

there are no stations (Daly, 2006). Examples of the use of C-V uncertainty estimation are presented in this article.

Some interpolation models have internal uncertainty estimation methods, such as Kriging's estimation variance (Tabios and Salas, 1985), and Parameter-elevation Regressions on Independent Slopes Model's (PRISM's) regression prediction interval (Daly *et al.*, 2008). These methods are useful in that they provide uncertainty estimates for each interpolated grid cell, not just at station locations. However, these estimates rely on the very same assumptions used in the interpolation process itself, and are therefore, not independent of the method used. In other words, they do not know what they do not know. The usefulness of PRISM's regression prediction interval as an uncertainty estimation method is evaluated in Section 4 of this article.

Meanwhile, hydrologic and ecological investigations have been conducted in small watersheds across the country, several of which are supported by dense rain gauge networks that provide detailed precipitation information at scales finer than those of the national datasets (Renard *et al.*, 1993; Slaughter *et al.*, 2001; Bailey *et al.*, 2003; Laseter *et al.*, 2012). These fine-scale gauge networks have the potential to provide 'ground-truth' for quantifying uncertainty in the national-level datasets, and assessing the sources of that uncertainty. Potential sources of uncertainty include station data (station density, data and metadata quality), interpolation method, and scale issues inherent in coarse-grid products.

This study utilizes a dense network of rain gauges maintained during the 1950s by the US Forest Service, Coweeta Hydrologic Laboratory (CHL) to: (1) develop fine-scale precipitation maps for the Coweeta watersheds, (2) quantify uncertainties in a national-level gridded precipitation product, and (3) attempt to isolate and quantify individual sources of uncertainty. A description of the CHL study area and rain gauge network is given in Section 2. Section 3 describes the development of fine-scale precipitation maps for CHL. Section 4 reports on how the CHL data and maps were used as ground truth in quantifying uncertainty in the PRISM national-scale gridded precipitation dataset.

Section 5 provides an assessment of the potential sources of uncertainty and their relative contributions. Concluding remarks are given in Section 6.

2. Study area

2.1. Study area description

The US Forest Service CHL is located in western North Carolina, USA, within the southern Appalachian Mountains. The 2185-ha laboratory encompasses two east-facing bowl-shaped basins, Coweeta basin and Dryman Fork basin. Elevations range from 675 to 1592 m (Figure 1), climate is classified as maritime, humid temperate (Trewartha, 1954), and long-term mean annual precipitation ranges from approximately 1500 mm in the lower, northeastern portions of the basin to about 2400 mm near the southwestern ridgetops. Watershed experimentation began here in 1934 (Swank and Crossley, 1988) and continues with present day research on the effects of elevation on hydrology, climate and vegetation (Bolstad *et al.*, 1998; Elliott *et al.*, 1999).

2.2. Rain gauge network and data

During the first 30 years of CHL's history, a dense network of standard rain gauges was installed throughout the basin in forest openings. At one time there were more than 120 gauges in place across a gradient of elevation, aspect, land use practices, and forest type, making it the densest mountain rain gauge network documented (Swift *et al.*, 1988). This network was used to test theories of rain gauge placement, (buried *vs* above-ground; tilted normal to the ground slope *vs* upright) in addition to collecting information on storm frequency, duration and intensity in southern Appalachian watersheds.

Data used in this study were obtained from 69 gauges with unbroken records for 8 years between November 1950 and October 1958 (Figure 1) (March *et al.*, 1979; Swift *et al.*, 1988). Precipitation measurements were made with standard 8-inch (20-cm) non-recording National

Weather Service rain gauges (http://www.weather.gov/iwx/coop_8inch). These gauges comprised a 100-in^3 (1645-cm^3) measuring cylinder housed inside an overflow container; a funnel rested on top of the overflow container (NOAA-NWS, 1989). No wind shields accompanied the gauges.

The data from this network of gauges were tabulated as monthly sums, and also summarized by growing season (April to October) and dormant season (November to March). Water year (annual) totals were calculated from the seasonal averages, with a water year beginning on 1 November and ending on 31 October. Seasonal and annual totals were averaged over the eight available water years to yield 1951–1958 average annual, summer, and winter precipitation totals.

Precipitation measurements (and thus interpolated precipitation data) are subject to numerous systematic uncertainties as a result of gauge design (e.g. Neff, 1977; Groisman and Legates, 1994; Brock and Richardson, 2001). These uncertainties include gauge wetting, evaporative losses, and precipitation undercatch. Measurement uncertainty resulting from gauge wetting is usually negligible (Sevruk, 1982). Evaporative losses are problematic when measuring precipitation in arid climates or when using gauges equipped with inlet heaters (Rasmussen *et al.*, 2012), but neither is the case at CHL. Precipitation undercatch can reach 80% during light rainfall and most snowfall events if wind speeds are greater than 10 m s^{-1} (Brock and Richardson, 2001). Shielding, such as fencing, is commonly placed near the openings of gauges to mitigate wind speeds, and thus undercatch. Although none of the gauges were shielded by fencing, Swift (1968) compared 'unshielded' gauges with co-located, partially buried gauges to quantify catch efficiency within the CHL gauge network. Results showed that differences in catch efficiency were minimal, most likely because the gauges were sited in small forest openings surrounded by continuous tree cover which acted as natural wind breaks. As such, the authors feel confident that precipitation undercatch was minimal.

Elevation data for the study area were provided by a 1/3 arc-second (hereafter termed 10-m) digital elevation model (DEM) from the US Geologic Survey's National Elevation Dataset (http://ned.usgs.gov).

3. Coweeta precipitation mapping

Based on a known strong elevation dependence of precipitation at CHL, the mapping approach taken here was to develop a precipitation–elevation regression function for each 8-year seasonal and annual average, and interpolate the residuals from this function to achieve the final map. The steps used in this process are given below.

3.1. Elevation regression model

As an initial step, the 10-m DEM was lightly smoothed using a low-pass filter to reduce variation by removing small-scale artefacts and noise and averaging high and low
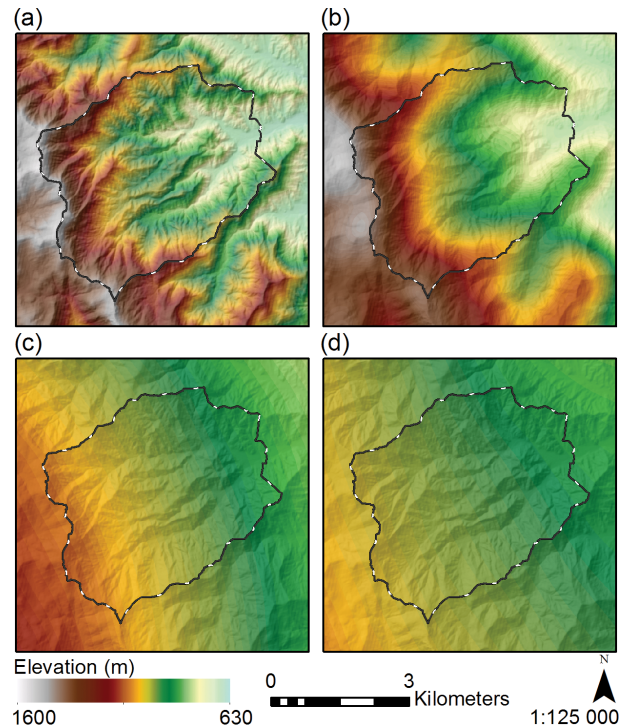

Figure 2. Coweeta 10-m DEM (a) in its unfiltered form, and filtered to a diameter of (b) 1000 m, (c) 7000 m, and (d) 10 000 m.

values to diminish extremes between neighbouring cells (Albani and Klinkenberg, 2003). The next step was to determine the optimal DEM scale for the regression functions. The direct effects of elevation on precipitation have been noted to be most important at scales of 5–10 km, owing to a number of physical mechanisms, including the advective nature of moisture-bearing airflow, the viscosity of the atmosphere, delays between initial uplift and subsequent rainout, and the movement of air around terrain obstacles (Daley 1991; Daly *et al.*, 1994; Funk and Michaelsen, 2004; Sharples *et al.*, 2005). For example, Daly *et al.* (2008) used a 30 arc-seconds (∼800-m) DEM that was low-pass filtered to an effective wavelength of 7 km in their mapping of 1971–2000 precipitation normals with PRISM. To find the optimal elevation scale for this region, the 10-m DEM was filtered to various wavelengths with a focal averaging tool. The tool averaged DEM grid cell values within a circular neighbourhood that varied in size from 100 to 10 000 m diameter (Figure 2). This process retained the original 10-m cell size, while smoothing features based on the diameter of the circular neighbourhood. Elevation values at each of the 69 gauge locations were derived for each DEM filtering diameter.

An Ordinary Least Squares (OLS) linear regression model was applied, using the filtered DEM elevations of the 69 stations as the explanatory variable and their 1951–1958 average annual and seasonal precipitation as the dependent variable in the form of

$$Y = bX + a \qquad (1)$$

Where Y is a station's average seasonal or annual precipitation, X is the filtered DEM elevation, a is the y-axis
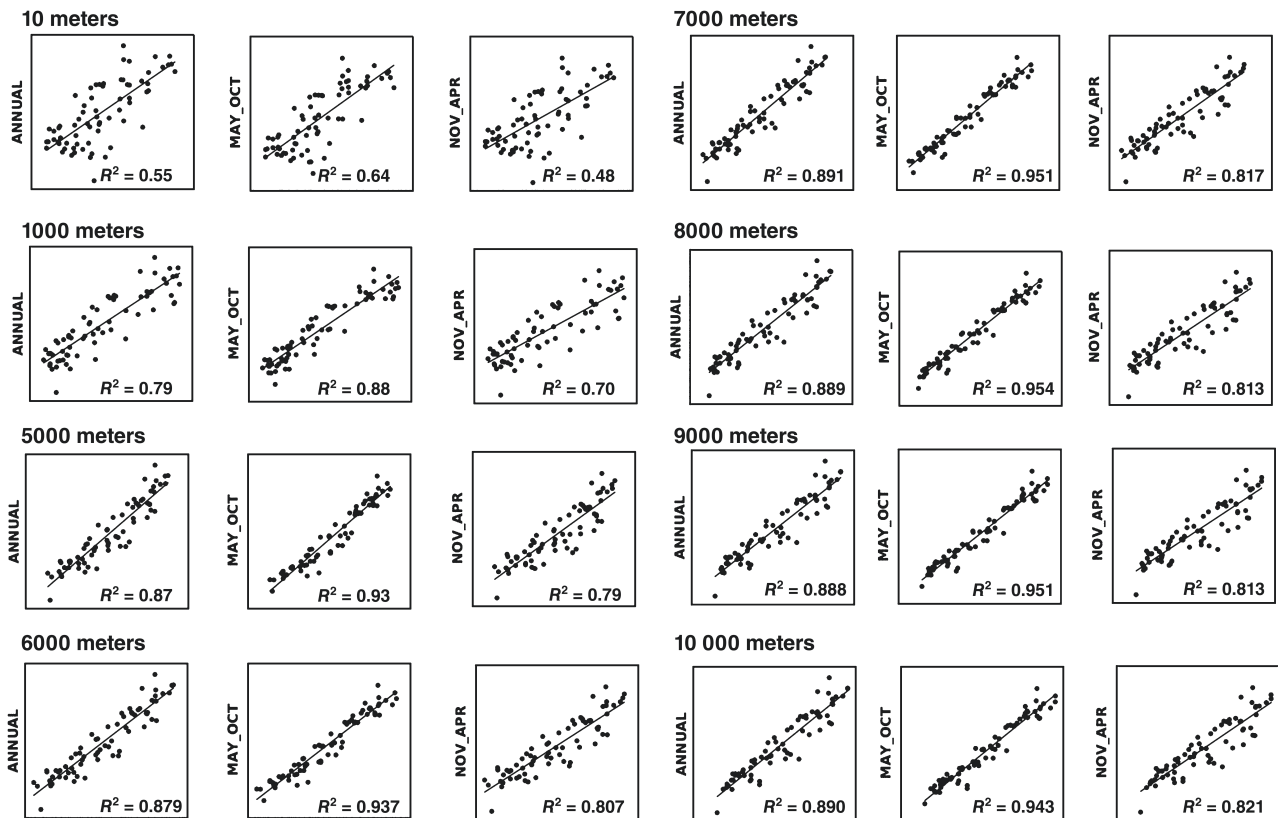
Figure 3. Scatterplots showing the relationship between the 10-m elevation grid, filtered to various wavelengths, and 1951–1958 average annual and seasonal precipitation totals at 69 stations located within the CHL.

Table 1. Parameters used in the precipitation–elevation regression functions for Coweeta 1951–1958 mean annual, winter, and summer precipitation, using a 10-m DEM filtered with 7000-m averaging diameter. The percent slope column expresses the slope of the precipitation–elevation function as percent of the average precipitation at all stations in the basin per kilometre of 7000-m filtered DEM elevation. $N = 69$ stations.

Season	Station average (mm)	Regression slope (mm m^{-1})	Slope (% per km)	Regression Y-intercept (mm)	R^2
Annual	1965	2.63	134	−615	0.89
Summer	776	1.06	107	−259	0.95
Winter	1188	1.58	162	−355	0.82

intercept, and *b* is the slope of the regression line. R^2 values and scatterplots were examined to evaluate the relationship between elevation and precipitation at the various DEM filtering wavelengths. Results from the regression analysis showed that filtering the DEM to higher wavelengths dramatically increased the precipitation variance explained by elevation for both seasons as well as the annual total (Figure 3). This increase reached a plateau at 7000 m, beyond which there was little improvement. At the 7000-m wavelength, chosen as optimal for this study, elevation explained approximately 81–95% of the variation in precipitation (Table 1). This is compared with 48–64% of the variation explained by the original 10 m DEM.

Less of the variation in precipitation was explained by elevation during winter than during summer. This is partly due to the formation of a 'dump zone' precipitation maximum below and to the lee (north) of the mountain ridge forming the southern border of the basin, caused by the acceleration of southwest-to-southeast winds through gaps in the ridge. This effect introduced scatter into the elevation regression function. In addition, winter updrafts on ridgetops at the extreme northern edge of the basin were observed to produce local precipitation minima. One of the gauges affected by these updrafts was standard rain gauge (SRG) 32, labelled in Figure 1, and visible as an outlier at the extreme lower left hand corner of the winter scatterplots in Figure 3. Finally, the lack of wind shielding could have reduced gauge catch efficiency, especially during periods of frozen precipitation. However, snowfall is a minor component of the winter precipitation at CHL (estimated by the authors at less than 10%), and siting of gauges in forest gaps reduced this effect.

The slope of the precipitation increase with elevation, normalized by the average precipitation of all 69 stations, ranged from 107% per km in winter to 162% per km in summer (Table 1). These slopes are more than twice as large as the regional average slopes derived by Daly *et al.* (2008) when mapping 1971–2000 mean annual precipitation in the eastern US with PRISM (see Section 4.1 for a description of PRISM). One reason for this difference may be one of geographic extent. The precipitation–elevation slope values from Daly *et al.*

(2008) were derived from average relationships over large distances, which included various geographic exposures, some with steep precipitation–elevation relationships, and some with little relationship. Coweeta likely represents a locally steep precipitation–elevation relationship, because of its position in a rain shadow on the leeward side of a significant orographic terrain barrier. Precipitation can decrease rapidly on leeward slopes, as evidenced by well-known rain-shadow gradients to the lee of major mountain ranges in the western US (Daly *et al.*, 2008).

3.2. Interpolation of residuals

Inverse distance weighting (IDW) and several Kriging interpolation models were applied to the residuals produced from the 7000-m OLS linear regression function and evaluated both visually and statistically. IDW determines a cell value using a linearly weighted combination of a set of points (Childs, 2004). This method assumes that the surface is a weighted average of scattered points that decrease in influence with distance from the sampled location. The closer a point is to the center of a cell, the more influence, or weight, it has in the averaging process (Li and Heap, 2008). IDW is an exact interpolator, and does not extrapolate beyond the range of the original point data. Therefore, minimum and maximum values can only occur at the point locations, and if data distribution is uneven, the interpolation can result in pits and peaks around outliers and clusters.

The geostatistical prediction technique of Kriging assigns weight based on the spatial covariance values of surrounding data points, and it has been argued that Kriging provides improved estimates over conventional techniques (Tabios and Salas, 1985; Daly *et al.*, 1994; Gotway *et al.*, 1996). Kriging models evaluated in this study included ordinary spherical Kriging, universal Kriging with linear drift, universal Kriging with quadratic drift, first order universal Kriging with and without anisotropy, and second order Kriging with and without anisotropy. Details on these methods can be found in Ver Hoef *et al.* (2003) and de Smith *et al.* (2007). Residual sum, average residual bias and MAE were assessed for each model for the annual and seasonal values. Of the Kriging models, ordinary spherical Kriging produced superior results, both statistically and qualitatively, and was advanced as the best Kriging option.

The IDW and Kriging models were applied to the annual and seasonal residuals with various search settings ranging from 12 to 30 neighbouring points. This was done to analyse the smoothing effect of the models and statistically determine optimal model parameters. The ordinary spherical Kriging model using 24 neighbouring input points exhibited the best statistical performance and produced an interpolated field that balanced smoothness with spatial detail (Figure 4(a)).

As IDW is an exact interpolator, changing the number of neighbouring input points did not significantly affect model fit at the station locations; in this application, 12 neighbour stations were used (Figure 4(b)). IDW was also
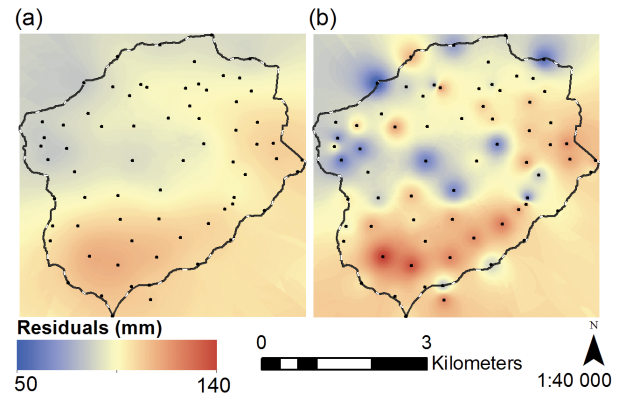

Figure 4. Residual fields of average annual precipitation interpolated with (a) ordinary spherical Kriging using 24 neighbouring stations, and (b) IDW using 12 neighbouring stations.

run with a distance weighting exponent of one (as opposed to the commonly used power of two) to effectively increase the radius of influence on each interpolated point. Results showed that the power setting of two produced slightly higher percent MAEs, but lower residual biases. Based on this information, the final IDW configuration chosen used 12 neighbour stations, distance weighted with an exponent of two.

To determine the overall optimum interpolation configuration for Coweeta precipitation residuals, an omit-one jackknife C-V with replacement was performed, as discussed earlier. This was done for 1951–1958 average annual, winter and summer precipitation, and residual interpolation with the optimum IDW and Kriging methods.

Overall C-V errors were very small, with MAEs ranging from 1.5 to 3% (Table 2), and differences in the errors between the two residual interpolation methods were consequently very small. Kriging MAEs were slightly lower than those of IDW, and root mean squared errors (RMSEs) were substantially lower. The Kriged interpolated fields were more generalized than those generated with IDW. The Kriged residual field exhibited a more realistic and spatially cohesive 'dump zone' in the southern part of the basin (positive residuals), while the IDW field was dominated by a series of bulls eyes in this area (Figures 4 and 5). Overall, Kriging was chosen as the better of the two methods for interpolating residuals from the elevation regression function.

The spatial distributions of C-V absolute errors in the Kriging interpolation of residuals are shown in Figure 5. Errors were highest in winter, but few stations had errors greater than 5%. There were no discernable patterns to the C-V error magnitudes, but as shown in Figure 3, the signs of the residuals were generally positive in the southern and eastern parts of the basin, and generally negative in the western, central, and northern regions.

3.3. Final precipitation maps

The final 1951–1958 mean annual, winter, and summer precipitation maps for the Coweeta basin are shown in Figure 6. Patterns of winter and summer precipitation both

Table 2. Comparison of cross-validation errors for interpolation of 1951–1958 mean annual, summer, and winter precipitation across the Coweeta basin using two methods: elevation regression plus Kriging of the residuals, and elevation regression plus IDW of the residuals. $N = 69$.

		Annual	Summer	Winter
Regression + Kriged residuals	Observed (mm)	1965	776	1188
	Predicted (mm)	1965	778	1188
	MAE (%)	2.20	1.57	2.87
	Bias (mm)	0.18	−0.68	0.31
	RMSE (mm)	26.50	2.69	23.59
Regression + IDW residuals	Predicted (mm)	1968	777	1191
	MAE (%)	2.43	1.60	3.17
	Bias (mm)	−3.08	−0.51	−2.57
	RMSE (mm)	61.54	16.10	48.24

followed the filtered elevation contours closely (Figure 1). The highest amounts were found in the elevated southwestern corner of the basin in both seasons (winter ∼1460 mm, summer ∼950 mm, and annual ∼2400 mm). However, the leeward 'dump zone' near the southern border of the basin was more prominent during the winter, likely owing to more persistent gap winds along the southern ridgeline. The lowest precipitation totals were found at the lower elevations of the northeastern end of the Coweeta basin (winter ∼940 mm, summer ∼625 mm, and annual ∼1560 mm).

4. Quantifying uncertainty in a national dataset

4.1. PRISM precipitation grids

Grids of monthly total precipitation for the Coweeta basin were obtained from the AN81m PRISM dataset (PRISM Climate Group 2015). This dataset spans 1895 to present, covers the conterminous United States at 30 arc-seconds (∼800-m) resolution, and includes total precipitation, average maximum and minimum temperature and vapor pressure deficit, and mean dew point. Mapping of monthly precipitation was performed using PRISM (Daly *et al.*, 1994, 2002, 2003, 2008). For each grid cell, PRISM calculated a local linear regression function between station precipitation and a predictor grid (see Climatologically

aided interpolation below). Nearby stations entering the regression were assigned weights based primarily on the physiographic similarity of the station to the grid cell. Physiographic factors relevant to this study were distance, elevation, topographic facet (hillslope), and effective terrain height (terrain profile). Detailed descriptions of the PRISM model algorithms, structure, input grids, and operation are given in Daly *et al.* (2002, 2008). Specific methods relevant to this study are described below.

4.2. Climatologically aided interpolation

The PRISM AN81m precipitation time series was developed using climatologically aided interpolation (CAI). CAI uses an existing climate grid to improve the interpolation of another climate element for which data may be sparse or intermittent in time (Willmott and Robeson, 1995; Funk *et al.*, 2000; Daly *et al.*, 2004; Hamlet and Lettenmaier 2005; Daly, 2006). This method relies on the assumption that local spatial patterns of the climate element being interpolated closely resemble those of the existing climate grid (called the predictor grid). The use of CAI in mapping monthly precipitation for the AN81m dataset involved using existing PRISM 1981–2010 monthly precipitation normals as the predictor grids (PRISM Climate Group, 2015). For example, interpolating precipitation for a grid cell in January 1951 involved using the January 1981–2010 normal precipitation grid as the independent variable in the local PRISM regression function for a grid cell, and nearby station values of January 1951 total precipitation as the dependent variable. Nationwide, the 1981–2010 normals mapping used approximately 20 000 stations, which is about twice the number of stations available in the 1950s. The 30-year PRISM normals were developed using a 30 arc-seconds DEM filtered to a 3.75 arc-minutes (∼7 km) effective radius as the predictor grid. Details on normals mapping methods are available from Daly *et al.* (2008).

4.3. Station data

In the 1950s, station data available for use in the PRISM precipitation modelling process came primarily from the National Weather Service Cooperative Observer Program

Absolute error (%) < .25 .50 1 2 3 4 5 > 5
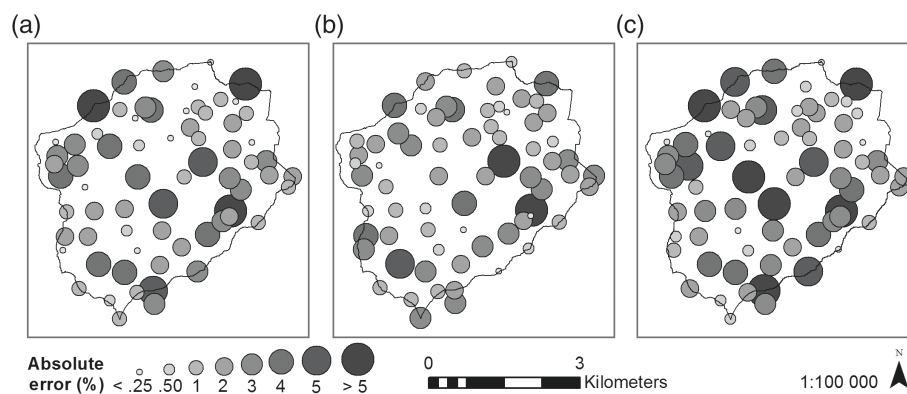
0   3 Kilometers   1:100 000   N

Figure 5. Spatial distribution of absolute cross-validation errors for interpolation of 1951–1958 mean (a) annual, (b) summer, and (c) winter precipitation using elevation regression plus ordinary spherical Kriging.
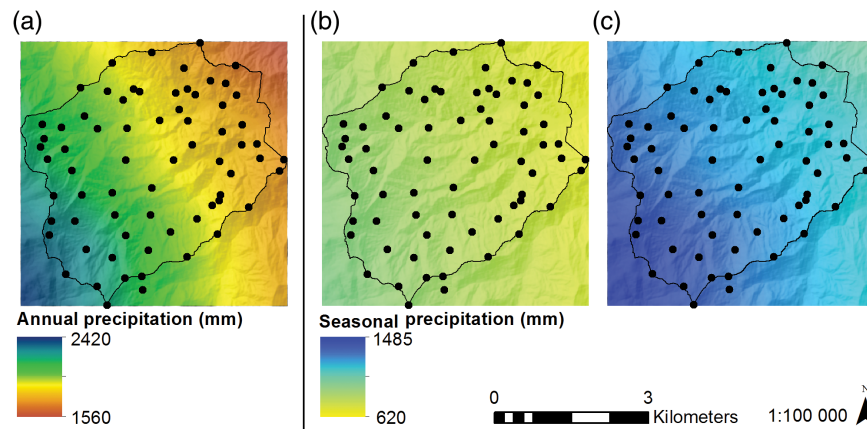
Figure 6. Final 1951–1958 mean (a) annual, (b) summer, and (c) winter precipitation for the Coweeta basin.

(COOP; http://cdo.ncdc.noaa.gov/CDO/cdo). Three stations within the Coweeta basin were reported through the COOP program: SRG 8, 19, and 21 (Figure 7). The COOP identifiers for these stations were 312107, 312102, and 312097, respectively. As will be discussed later, the locations of these stations in the COOP database were not accurate, which was a source of uncertainty in the PRISM grids. The nearest stations outside the basin typically had a small influence on the PRISM interpolated estimates, because they were relatively distant from the basin. However, they did have a large effect on the estimated prediction error (see Section 4.5) and occasionally on the estimates themselves (see Appendix S1, Supporting information). Station data used in the 1981–2010 precipitation normals mapping in the Coweeta area were surprisingly similar to those in the 1950s. In fact, the same three Coweeta stations – gauges 8, 19, and 21 – were included in the normals mapping. This means that, unlike many other parts of the country, the use of CAI to map precipitation at Coweeta during the 1950s did not confer much of an advantage over the conventional method of using a DEM as the predictor grid. The station locations used in the 1981–2010 normals mapping were identical to those used to map precipitation during the 1951–1958 study period.

4.4. Measures of uncertainty in PRISM data

Two methods have been commonly used to estimate uncertainty in the PRISM datasets. The first is the prediction interval, or the degree of scatter, around each grid cell's regression function (Daly *et al.*, 2008). Unlike a confidence interval, the prediction interval takes into account both the variation in the possible location of the expected value of Y for a given X (as the regression parameters must be estimated), and variation of individual values of Y around the expected value (Neter *et al.*, 1989). A 1-α of 0.70 (70%), where α is the significance level, was chosen for the prediction interval because it approximated 1 standard deviation (where 1-$\alpha = 0.67$) around the model prediction. Termed PI70, the 70% prediction interval is evaluated for each grid cell.

The second method is the single-station jackknife C-V error, which has already been introduced here. As C-V errors can only be calculated for stations (Figure 5), they are reported as points or an average of points over an area. Previous work has shown that PI70 values generally agree with C-V MAE values over stations in the same area reasonably well, if summarized over large regions of 200–300 km in extent. However, discrepancies can be large over local areas. Details on the formulation of PI70 and its relationship with C-V errors are given in Daly *et al.* (2008).

4.5. Comparing PRISM to ground truth

PRISM grids of 1951–1958 mean annual, summer, and winter precipitation for the CHL were extracted from the 800-m monthly national grids by: (1) defining a window encompassing the CHL and vicinity from the PRISM 800-m national grid extent, (2) extracting monthly sub-grids defined by this window from the national grids covering November 1950 to October 1958, (3) summing the monthly sub-grids for the eight water years (November to October), summer seasons (April to October), and winter seasons (November to March) in this period, and (4) averaging the eight water-year and seasonal values to obtain 1951–1958 mean annual, summer, and winter precipitation grids.

The PRISM precipitation grids bear a close resemblance qualitatively to the ground truth grids (Figures 6 and 7), except for obvious differences in spatial resolution. The 800-m PRISM grids show the same general southwest-northeast trend in precipitation, with similar minima and maxima. A notable difference is that the precipitation minimum in the winter season (which also appears in the annual) in the northeastern corner of the watershed in the ground truth is shifted to the eastern part of the watershed in the PRISM maps. In addition, the PRISM grids are wetter along the northwestern boundary of the CHL than ground truth.

PRISM and the Coweeta ground-truth precipitation grids were compared quantitatively by subtracting the Coweeta grids from the PRISM grids. Differences listed in Table 3 were calculated after the 800-m PRISM grids
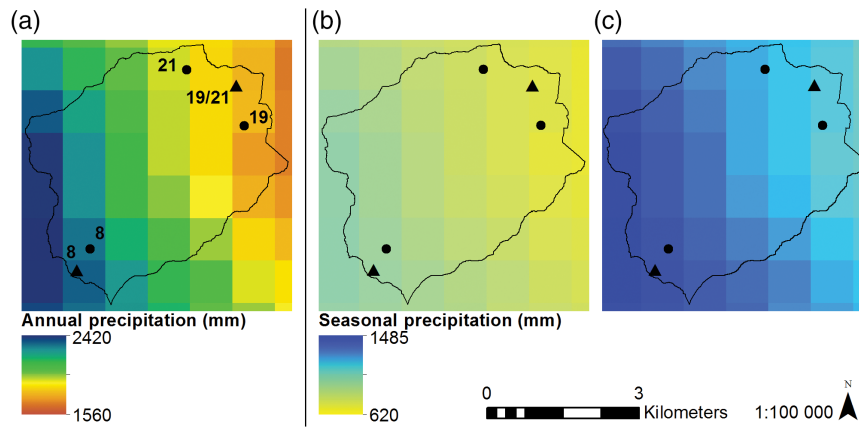
Figure 7. PRISM 1951–1958 mean precipitation grids in the Coweeta area: (a) annual, (b) summer, and (c) winter. Locations of the three Coweeta stations used in PRISM are shown as triangles, and their actual locations are shown as dots.

Table 3. Summarized differences between the PRISM national precipitation grids and Coweeta ground-truth precipitation grids (PRISM − ground truth), between PRISM grid values for Coweeta stations and the observed values (PRISM estimates – observations), and the PRISM PI70 (70% prediction interval) grids for 1951–1958 mean annual, summer, and winter precipitation. See text for details.

Season	PRISM grid – Coweeta grid ($N = 300\,000$)		PRISM grid – station observations ($N = 69$)		PRISM PI70 grid at station locations ($N = 69$)
	MAE (%)	Bias (%)	MAE (%)	Bias (%)	PI70 (%)
Annual	2.72	1.73	3.78	1.58	15.79
Summer	2.34	−0.01	2.21	−0.13	21.02
Winter	3.92	2.93	5.25	2.74	12.36

were first resampled to 10-m resolution before differencing. In Figure 8, the opposite was done, where the 10-m ground-truth grids were first resampled to 800-m before differencing. Overall, grid-to-grid differences were small, with MAE ranging from 2 to 4% and bias from 0 to 3%. The average MAE was slightly greater at station locations during winter, which contributed to a higher annual value (Figure 8, Table 3). A month-by-month error analysis is summarized in Table S1 and Figure S1 in Appendix S1.

The quantitative patterns of the differences confirm the qualitative differences mentioned above. In winter, the PRISM winter precipitation grids are about 2–3% drier than ground truth in the eastern part of the watershed, and the northeastern edge is about 10% wetter. Along the western boundary, PRISM is about 10% wetter than ground truth. This is likely because the western ridgeline has a similar elevation to the wetter southwestern ridgeline, and with only SRG 8 as guidance, PRISM estimated similar precipitation values along the entire boundary.

The PRISM PI70 grids overestimated the true prediction errors at the Coweeta gauge locations. Errors were estimated at 12% in winter, 21% in summer, and 16% annually, which are several times larger than the actual absolute differences. The main reason for the PI70 error overestimation appears to be that PI70 is generalized in space, and not targeted to a local area. A typical monthly PRISM regression function between 1981 and 2010 average grid values of precipitation and the station data for a grid cell within the basin was weighted towards the three Coweeta stations.

However, a total of 25 stations were used in the regression function, a setting which produced optimal overall performance in the eastern US. (Using this setting, PRISM automatically adjusted its radius of influence to encompass a minimum of 25 stations for each pixel's regression function.) While individual weights of the remaining 22 stations outside the basin were small, the sum total of these weights was not insignificant. Given that these outlying stations represented precipitation regimes that are not necessarily the same as those of Coweeta, the resulting PI70 reflected greater scatter around the regression line than would have been the case if the regression function had used only the three local Coweeta gauges (Figure 9).

PRISM C-V MAE and bias were calculated by omitting each of the three stations used in the interpolation (SRG 8, 19, and 21) one at a time, estimating their observations in each station's absence, then replacing the station. Because C-V errors could only be evaluated for the three stations used in the interpolation, and station data and locations reflected those used in the PRISM modelling process (see Section 5), the C-V errors are not directly comparable to the actual differences between the PRISM grids and the estimates from the other methods. Given these caveats, the C-V MAE and bias are much more in line with the true PRISM-station differences (Table 4). C-V MAEs averaged 1–3% for the three Coweeta stations. These values were slightly lower than the MAEs for the 25 nearest stations used in the example shown in Figure 9. In contrast, the PRISM PI70 values ranged from 11 to 20%.
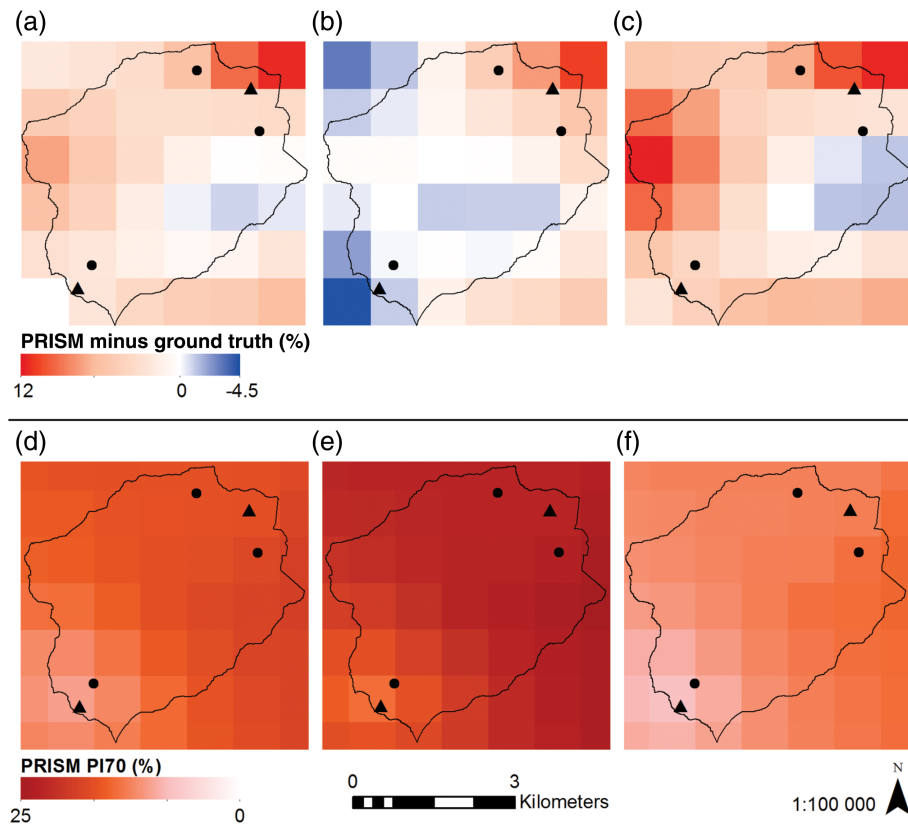
Figure 8. Differences between the Coweeta ground-truth and PRISM precipitation grids (PRISM − ground truth) and accompanying PRISM PI70 grids for: (a, d) annual, (b, e) summer, and (c, f) winter.
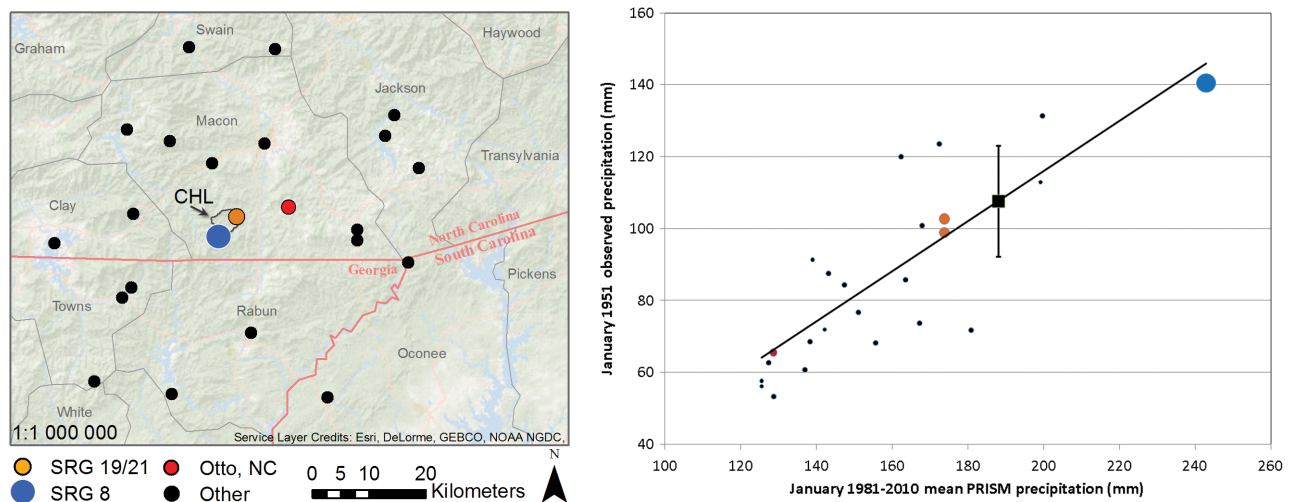

Figure 9. Locations of stations used in the PRISM CAI regression function, and accompanying scatterplot and regression function, for January 1951 at a grid cell in the southeastern portion of the CHL. Each dot represents a station observation, and the size of the dot is proportional to the weight assigned to that station in the regression function. SRG 8 (blue dot) has the highest weight. SRG 19 and 21(orange) are co-located in the PRISM database, resulting in a reduction of their weights by half to prevent over-representation. A station in Otto, NC, 10 km to the northeast (red, lower left of scatterplot), the closest station outside the CHL, has a lower weight. Other, more distant, stations have very low weights. The predicted precipitation value is shown in the scatterplot as a square at (188.1, 107.6). The PI70 70% prediction interval (±15.5 mm) appears as vertical whiskers around the predicted value.

5. Evaluating sources of uncertainty

This study provided an opportunity to evaluate the sources of uncertainty that may have contributed to the differences between the CHL grids and the PRISM national grids. A series of sensitivity tests involving six uncertainty scenarios was prepared, starting with the original 'best map' Coweeta mapping methodology as Scenario 1, and ending with a scenario that approximated the PRISM modelling situation. Each scenario added a potential source of uncertainty encountered in the preparation of the PRISM

Table 4. Average PRISM C-V errors, differences between PRISM grid values for Coweeta stations and the observed values (PRISM estimates − observations), and PRISM PI70 grid values at the locations of SRG 8, 19, and 21, for 1951–1958 mean annual, summer, and winter precipitation. C-V errors used PRISM station locations and COOP observations; others used Coweeta station locations and observations.

Season	PRISM C-V errors Coweeta stations ($N = 3$)		PRISM C-V errors Coweeta and vicinity stations ($N = 25$)		PRISM ppt grid − station ppt observations ($N = 3$)		PRISM PI70 grid at station locations ($N = 3$)
	MAE (%)	Bias (%)	MAE (%)	Bias (%)	MAE (%)	Bias (%)	PI70 (%)
Annual	1.17	−1.17	2.59	−0.16	1.47	0.32	14.90
Summer	3.30	−3.04	3.18	−2.41	2.37	−0.49	20.02
Winter	1.41	−0.08	3.35	1.58	2.04	0.85	11.55

Table 5. Summary of uncertainties in mapping 1950–1958 mean precipitation in the CHL under various data and methodological scenarios. MAEs and biases are expressed as prediction − observation differences averaged over the full 69-station dataset. Basin averages refer to the areal average precipitation over the CHL. N is the number of stations used in the scenario. Values are given for annual (A), summer (S), and winter (W).

Scenario	N		MAE		Bias		Basin average precipitation	
			Value (%)	Diff from best (%)	Value (%)	Diff from best (%)	Value (mm)	Diff from best (%)
1. Best map: Elevation regression, Kriged residuals, all stations, 10-m DEM filtered to 7 km	69	A	1.75	0.00	0.07	0.00	1970	0.00
		S	1.55	0.00	−0.04	0.00	779	0.00
		W	2.24	0.00	0.11	0.00	1192	0.00
2. Same as Scenario 1, except using Coweeta precipitation values at PRISM COOP stations (SRG 8, 19, and 21)	3	A	4.95	3.20	4.74	4.67	2097	6.42
		S	4.16	2.61	4.11	4.15	813	4.38
		W	5.57	3.33	5.19	5.08	1256	5.42
3. Same as Scenario 2, except using PRISM COOP station values	3	A	3.81	2.06	3.28	3.21	2040	3.51
		S	3.47	1.92	−3.12	−3.08	756	−2.88
		W	7.54	5.30	7.51	7.40	1283	7.72
4. Same as Scenario 3, except using PRISM station locations, with averaged precipitation values from SRG 19 and 21	3[a]	A	3.58	1.83	3.27	3.20	2034	3.24
		S	1.64	0.09	0.79	0.83	785	0.78
		W	5.18	2.94	4.94	4.83	1250	4.89
5. Same as Scenario 4, except using PRISM 800-m DEM filtered to 3.75 min as predictor grid	3[a]	A	5.27	3.52	4.23	4.16	2039	3.48
		S	3.84	2.29	3.30	3.34	799	2.62
		W	6.28	4.04	4.90	4.79	1240	4.07
6. Same as Scenario 5, except using PRISM 1981–2010 ppt normals as predictor grids	3[a]	A	5.22	3.47	3.38	3.31	2023	2.67
		S	3.98	2.43	2.48	2.52	793	1.85
		W	6.06	3.82	4.01	3.90	1230	3.23
7. Same as Scenario 1, except using current 12-station Coweeta network	12	A	2.56	0.81	1.96	1.89	2006	1.83
		S	6.84	5.29	6.84	6.88	832	6.92
		W	3.28	1.04	−1.24	−1.35	1174	−1.49
8. Same as Scenario 3, except using 'worst-case' COOP station locations	3	A	8.81	7.06	8.81	8.74	2147	8.95
		S	7.99	6.44	7.99	8.03	841	8.08
		W	9.40	7.16	9.40	9.29	1305	9.55
9. PRISM national grids	25[b]	A	3.78	2.03	1.58	1.51	2009	1.96
		S	2.21	0.66	−0.13	−0.09	780	0.18
		W	5.25	3.01	2.74	2.63	1229	3.15

[a]Due to station location errors in the COOP database, SRG 19 and 21 were incorrectly placed at the same location. Thus, precipitation data from these two stations were averaged.
[b]PRISM used 25 stations in its regression functions, but the three COOP stations within the Coweeta Basin were the most highly weighted.

grids. These included reducing the number of stations from 69 to 3, altering the station data and locations from their original values to match those reported to PRISM through the COOP program, coarsening the grid resolution from 10 to 800 m, and using different predictor grids. To provide more targeted comparisons, three additional scenarios were added, the first reducing the station network to the 12 stations currently in operation at CHL, and the second moving station locations to 'worst case' locations reported through official channels that could have been used for modelling and analysis. Finally a scenario representing the final PRISM grids themselves was added for reference.

Descriptions of the scenarios and results are summarized in Table 5 and Figure 10. MAE, bias, and basin average precipitation were calculated for each scenario. MAE and bias were determined by averaging prediction-minus-observation differences over the original
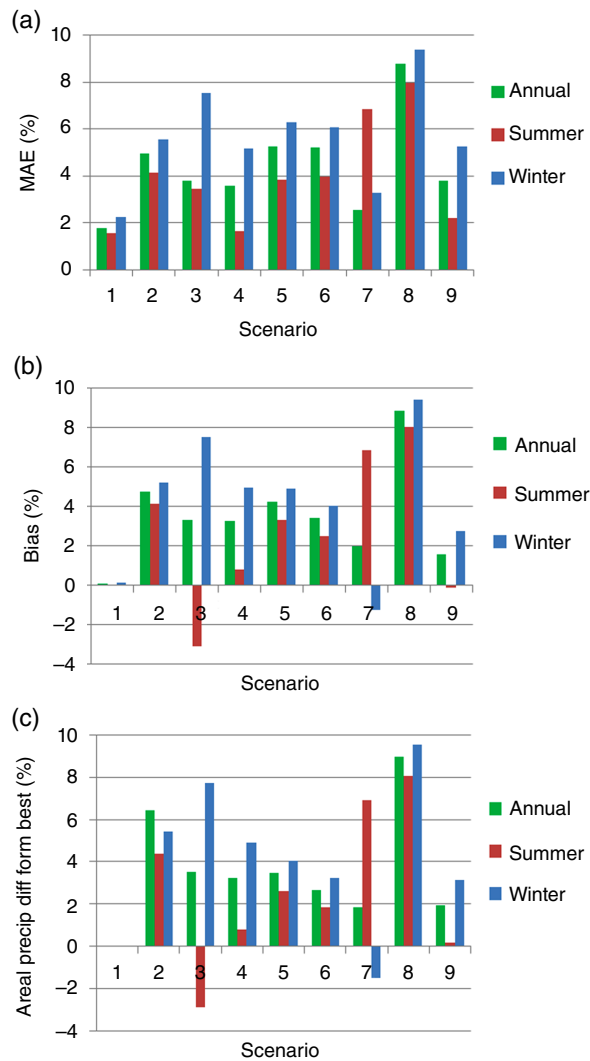
Figure 10. Uncertainty scenarios and associated error statistics: (a) MAE, (b), bias, and (c) average precipitation over the Coweeta basin. MAE and bias were determined by averaging prediction-minus-observation differences over the original 69-station dataset. See Table 5 and text for scenario descriptions.

69-station dataset. In Scenario 2, all but the three stations available to PRISM through the COOP database (SRG 8, 19, and 21) were omitted. This more than doubled the MAE from 1.5–2.3% to 4.1–5.6% over the 'best map' Scenario 1, and produced an overall wet bias of 4–5%. Average basin precipitation also increased 4–6% over the best map. In Scenario 3, precipitation data from these three stations were modified from their original values to those reported through the COOP database. These modifications were small: −0.30, −0.65, and −0.07% for annual, summer, and winter, respectively, at SRG 8, +0.69, +0.57, and +0.77% at SRG 19, and −0.32, −0.27, and −0.35% at SRG 21. However, these small changes had relatively large effects on the mapping process. The 4% positive bias in summer precipitation in Scenario 2 was reduced to a 3% negative bias in Scenario 3. The positive winter precipitation bias from Scenario 2 increased from 5% to over 7% in Scenario 3. On an annual basis, altering the

precipitation values reduced Scenario 2's 6.4% positive bias in basin average precipitation by half.

In Scenario 4, the three station locations were moved from their actual locations to the COOP locations used in PRISM. During most of the 20th century, it was not uncommon for COOP station locations to be roughly located based on driving directions or positions on maps, and often expressed only to the nearest minute latitude/longitude. These positions were manually entered, and errors sometimes occurred, such as transposed digits, rounding error, or inaccurate copying of coordinates. The locational history of these three stations was studied, and the results are summarized in Table S2 and Figure S2 in Appendix S1. To match the locations used by PRISM, SRG 8 was moved 0.49 km southwest of the actual location, SRG 19 was moved 0.76 km to the northwest, and SRG 21 was moved 1.09 km to the east. Interestingly, the moves of SRG19 and 21 put them in exactly the same location, likely due to truncating of the coordinates to the nearest minute latitude/longitude. In order to perform the mapping process, the precipitation values from SRG 19 and 21 were averaged and combined into one station. The elevation regression function was then performed with only two stations. Kriging of the residuals was not performed, because a minimum of three unique station locations are required. MAEs for annual precipitation remained about the same, but those for winter and summer decreased about 2% points. This was caused by an increase in summer precipitation, compensated for by a decrease in winter precipitation.

In Scenario 5, the 10-m DEM filtered to 7 km, used to develop the precipitation–elevation regression function, was replaced by the DEM used by PRISM (800-m resolution, filtered to 7 km). In Scenario 6, the DEM was replaced by the PRISM 1981–2010 precipitation normal grid corresponding to the appropriate season (the PRISM CAI method used this as the predictor grid). Scenario 6 approximated the actual PRISM modelling situation most closely. The coarsening of the DEM resolution in Scenario 5 increased the MAEs 1–2% across the board, but improved the winter basin average slightly. Bringing in the PRISM normal grids in Scenario 6 improved performance slightly over that of Scenario 5. While errors associated with the actual PRISM grids (Scenario 9) were similar to those of Scenario 6, PRISM did a somewhat better job of approximating the basin average precipitation, coming within 3% of the actual. This may have been due to the influence of stations outside the basin.

The results of uncertainty scenarios above can best be described as mixed, with offsetting biases making it difficult to see a clear trend towards increasing uncertainty as the scenarios progress from the best map towards the PRISM modelling situation. Part of the reason for the lack of a clear trend is that the PRISM errors were very small to begin with, meaning that individual contributions could have been within the noise level of the analysis. To help clarify the picture, we created Scenarios 7 and 8. In Scenario 7, the best mapping method was used, but the station dataset was reduced to the 12-station network that

is being maintained at CHL today. The network reduction was done carefully by CHL personnel in an attempt to minimize biases in the basin average precipitation compared with that offered by the more complete network. Compared with the best map Scenario 1, the network reduction maintained the basin-wide average within 2% for annual and winter. However, using the reduced network resulted in a 7% increase in summer precipitation compared with the full network (Table 5), demonstrating that network design can have significant effects on the estimation of basin-wide precipitation inputs.

In Scenario 8, the three COOP stations were moved to officially documented (but incorrect), locations that were expected to cause the greatest change in the precipitation map. This involved moving SRG 8 0.76 km to the northeast, SRG 19 2.87 km to the northeast, and SRG 21 1.09 km to the east (Table S2 in Appendix S1). Compared with Scenario 3, where only the three COOP stations were used, but at the correct locations, moving the stations resulted in an increase in MAE from 3–4% to 8–9% for annual and summer, and from 7.5 to 9.4% in winter. Biases and CHL areal average precipitation deviations were all strongly positive, while those of Scenario 3 were both positive and negative.

6. Conclusions

This study took advantage of a dense, 69-station rain gauge network operating in the CHL from 1951 to 1958 to produce seasonal and annual average precipitation maps over this period. These maps were used as ground-truth to evaluate the uncertainty in a national precipitation dataset. The ground-truth precipitation maps were produced by developing regression functions between a 10-m DEM and 1951–1958 mean annual (November to October), summer (April to October), and winter (November to March) precipitation, and interpolating the residuals with ordinary spherical Kriging. It was found that the 10-m DEM filtered to an approximately 7-km effective wavelength explained the most variance in precipitation ($R^2 = 0.89$ annual, 0.95 summer, and 0.82 winter). The maps showed the highest precipitation amounts in the elevated southwestern corner of the basin in both seasons (annual ~2400 mm, summer ~950 mm, and winter ~1460 mm). However, there was a noticeable leeward 'dump zone' downwind of the mountain crest making up the southern border of the basin. This dump zone was apparently caused by carry-over of hydrometeors by winds accelerating through gaps in the southern ridgeline, and was most noticeable during winter. The Kriged residuals from the elevation regression function performed reasonably well in adding this feature back into the maps, but the true dump zone was likely narrower and more sharply defined than that depicted in the maps. The lowest precipitation totals were found at the lower elevations of the northeastern end of the basin (winter ~940 mm, summer ~625 mm, and annual ~1560 mm). Given the density of the CHL rain gauge network, the station data and resulting annual and seasonal precipitation maps were considered a close approximation

of ground-truth – that is, coming close to depicting the actual precipitation fields in the basin. This provided a unique opportunity to quantify the actual uncertainty of PRISM national-level datasets for the same time period, evaluate how well the model estimated its own uncertainty, and explore the potential sources of uncertainty and their contributions. It was a rare circumstance to be able to directly evaluate the accuracy of the model and its estimates of uncertainty.

The PRISM grids were found to match ground-truth closely (less than 5% average difference), aided by the fact that three stations in the basin (SRG 8, 19, and 21), were reported through the National Weather Service COOP program, and thus available to PRISM. PRISM uses PI70 and C-V to estimate interpolation errors. While the C-V errors at the three stations were similar to the differences between the PRISM and CHL grids, PI70 error estimates were much larger, ranging from 12 to 20%. PI70 is derived from the scatter of the data points in the PRISM regression functions, which include 25 stations surrounding the modelled grid cell, not just the closest 3. Thus, the PI70 became larger when other, more distant, stations were taken into account. This finding is consistent with a previous analysis that found that C-V errors become similar to PI70 values only when averaged over large, regional blocks.

This study provided an opportunity to evaluate the sources of uncertainty that may have contributed to the differences between the CHL grids and the PRISM national grids. A series of sensitivity tests involving six uncertainty scenarios was prepared, with each scenario adding a potential source of uncertainty encountered in the preparation of the PRISM grids. These included reducing the number of stations from 69 to 3, altering the station data and locations from their original values to match those reported to PRISM through the COOP program, coarsening the grid resolution from 10 to 800 m, and using different predictor grids. An increase in error was noted when the station network was reduced from 69 gauges to 3. Beyond that, no clear trend in error magnitude was found through the progression of scenarios. The strong ability of elevation, filtered to 7-km wavelength, to predict precipitation over the watershed resulted in very small PRISM interpolation errors. This made it difficult to tease out individual contributions to these errors. In addition, offsetting biases (compensating errors) sometimes reduced the apparent error rates from one scenario to the next.

In an attempt to isolate the contributions of network density to the overall uncertainty, a scenario was developed that reduced the station network from 69 stations to the 12 stations currently operating at CHL. The scenario revealed that if the mapping process had used this reduced network only, the difference would have been minimal for the annual and winter maps, but the reduced network would have overestimated summer precipitation by 7% on a basin-wide basis.

In an attempt to bracket how much uncertainty might be introduced through station mislocations, which is a common problem nationwide, a final scenario was developed that moved the locations of the three COOP stations, which

were mislocated in many of the official metadata documents, to 'worst case' locations (see Appendix S1). The station mislocation scenario resulted in a doubling of the MAE for annual and summer precipitation, confirming that the effects were substantial. The effects could have been much worse if precipitation patterns had been more closely tied to fine-scale (rather than 7-km effective scale) terrain features, as is often the case with temperature, for example. A location error of just a few hundred metres in complex terrain can have large effects on temperature if the associated elevations are altered significantly.

Analyses like the ones conducted here can be expanded to determine optimal network design. One possible exercise would analyse all station combinations to find the network configuration that minimizes the number of stations to be maintained and at the same time introduces the least bias in precipitation amounts. Such an exercise would be useful if the CHL precipitation maps are updated to a more recent climatological period; a task which will necessarily involve fewer stations.

The CHL rain gauge network maintained in the early and middle part of the 20th century is not the only high-density network in the United States. Repeating this study in other basins where dense rain gauge networks exist (or have existed) would be useful to gain a more comprehensive view of how national precipitation products such as PRISM compare to ground truth, and better interpret the error estimates in these models.

Acknowledgements

Supporting information

The following supporting information is available as part of the online article:
Appendix S1. Monthly comparison of PRISM grids and CHL station data, and location history of CHL COOP stations.

References

Albani M, Klinkenberg B. 2003. A spatial filter for the removal of striping artifacts in Digital Elevation Models. *Photogramm. Eng. Remote Sensing* **69**: 755–765, doi: 10.14358/PERS.69.7.755..

Alexander LV, Zhang X, Peterson TC, Caesar J, Gleason B, Klein Tank AMG, Haylock M, Collins D, Trewin B, Rahimzadeh F, Tagipour A, Ambenje P, Rupa Kumar K, Revadekar J, Griffiths G. 2006. Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res.* **111**: D05109, doi: 10.1029/2005JD006290.

Allen RG, Pereira LS, Raes D, Smith M. 1998. Crop evapotranspiration – Guidelines for computing crop water requirements – FAO Irrigation and Drainage Paper 56. Food and Agriculture Organization of the United Nations, Rome.

Bailey AS, Hornbeck JW, Campbell JL, Eagar C. 2003. Hydrometeorological database for Hubbard Brook Experimental Forest: 1955–2000. General Technical Report NE-305. 36 pp. Newtown Square, PA, USDA Forest Service Northeastern Research Station.

Beven KJ. 2004. *Rainfall-runoff Modelling: The Primer*, 2nd edn. Wiley: Chichester, UK.

Bolstad PV, Swift L, Collins F, Regniere J. 1998. Measured and predicted air temperatures at basin to regional scales in the southern Appalachian Mountains. *Agric. For. Meteorol.* **91**: 161–176, doi: 10.1016/s0168-1923(98)00076-8.

Brock FV, Richardson SJ. 2001. *Meteorological Measurement Systems*. Oxford University Press: Oxford, UK.

Childs C. 2004. Interpolating surfaces in ArcGIS spatial analyst. ARCUser July-September 2004. http://www.esri.com/news/arcuser/0704/files/interpolating.pdf (accessed 10 September 2015)

Daley R. 1991. *Atmospheric Data Analysis*. Cambridge University Press: Cambridge, UK.

Daly C. 2006. Guidelines for assessing the suitability of spatial climate data sets. *Int. J. Climatol.* **26**: 707–721, doi: 10.1002/joc.1322.

Daly C, Neilson RP, Phillips DJ. 1994. A Statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteorol.* **33**: 140–158, doi: 10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2.

Daly C, Gibson WP, Taylor GH, Johnson GL, Pasteris P. 2002. A knowledge-based approach to the statistical mapping of climate. *Clim. Res.* **22**: 99–113, doi: 10.3354/cr022099.

Daly C, Helmer EH, Quinones M. 2003. Mapping the climate of Puerto Rico, Vieques, and Culebra. *Int. J. Climatol.* **23**: 1359–1381, doi: 10.1002/joc.937.

Daly C, Gibson WP, Doggett M, Smith J, Taylor G. 2004. Up-to-date monthly climate maps for the conterminous United States. *Proc., 14th AMS Conf. on Applied Climatology, 84th AMS Annual Meeting Combined Preprints, Amer. Meteorological Soc.*, Seattle, WA, January 13–16, Paper P5.1. http://ams.confex.com/ams/pdfpapers/71444.pdf. (accessed 25 July 2016).

Daly C, Halbleib M, Smith JI, Gibson WP, Doggett MK, Taylor GH, Curtis J, Pasteris PA. 2008. Physiographically-sensitive mapping of temperature and precipitation across the conterminous United States. *Int. J. Climatol.* **28**: 2031–2064, doi: 10.1002/joc.1688.

Elliott KJ, Vose JM, Swank WT, Bolstad PV. 1999. Long-term patterns in vegetation-site relationships in a southern Appalachian forest. *J. Torrey Bot. Soc.* **126**: 320–334.

Elliott KJ, Ford Miniat C, Pederson N, Laseter SH. 2015. Forest tree growth response to hydroclimate variability in the southern Appalachians. *Glob. Change Biol.* **21**: 4627–4641, doi: 10.1111/gcb.13045.

Funk C, Michaelsen J. 2004. A simplified diagnostic model of orographic rainfall for enhancing satellite-based rainfall estimates in data-poor regions. *J. Appl. Meteorol. Clim.* **43**: 1366–1378, doi: 10.1175/jam2138.1.

Funk C, Michaelsen J, Verdin J, Artan G, Husak G, Senay G, Gadain H, Magadazire T. 2000. The collaborative historical African rainfall model: description and evaluation. *Int. J. Climatol.* **1**: 47–66, doi: 10.1002/joc.866.

Gotway CA, Ferguson RB, Hergert GW, Peterson TA. 1996. Comparison of kriging and inverse-distance methods for mapping soil parameters. *Soil Sci. Soc. Am. J.* **60**: 1237–1247, doi: 10.2136/sssaj1996.03615995006000040040x.

Groisman PY, Legates DR. 1994. The accuracy of United States precipitation data. *Bull. Am. Meteorol. Soc.* **75**: 215–227, doi: 10.1175/1520-0477(1994)075<0215:TAOUSP>2.0.CO;2.

Gyalistras D. 2003. Development and validation of a high-resolution monthly gridded temperature and precipitation data set for Switzerland (1951–2000). *Clim. Res.* **25**: 55–83, doi: 10.3354/cr025055.

Hamlet AF, Lettenmaier DP. 2005. Production of temporally consistent gridded precipitation and temperature fields for the continental U.S. *J. Hydrometeorol.* **6**: 330–336, doi: 10.1175/jhm420.1.

Laseter SH, Ford CR, Vose JM, Swift LW. 2012. Long-term temperature and precipitation trends at the Coweeta Hydrologic Laboratory, Otto, North Carolina, USA. *Hydrol. Res.* **43**: 890–901, doi: 10.2166/nh.2012.067.

Legates DR, McCabe GJ. 1999. Evaluating the use of "goodness of fit" measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **35**: 233–241, doi: 10.1029/1998wr900018.

Li J, Heap AD. 2008. A review of spatial interpolation methods for environmental scientists. Geoscience Australia, Record 2008/23, 137 pp.

Livneh B, Rosenburg EA, Lin C, Nijssen B, Mishra V, Andreadis KM, Maurer EP, Lettenmaier DP. 2013. A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: update and extensions. *J. Clim.* **26**: 9384–9392, doi: 10.1175/jcli-d-12-00508.1.

Livneh B, Bohn TJ, Pierce DW, Munoz-Arriola F, Nijssen B, Vose R, Cayan DR, Brekke L. 2015. A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950–2013. *Sci. Data* **2**: 150042, doi: 10.1038/sdata.2015.42.

March WJ, Wallace JR, Swift LW Jr. 1979. An investigation into the effect of storm type on precipitation in a small mountain watershed. *Water Resour. Res.* **15**: 298–304, doi: 10.1029/wr015i002p00298.

Maurer EP, Wood AW, Adam JC, Lettenmaier DP, Nijssen B. 2002. A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *J. Clim.* **15**: 3237–3251, doi: 10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2.

Neff EL. 1977. How much rain does a rain gauge gauge? *J. Hydrol.* **35**: 213–220, doi: 10.1016/0022-1694(77)90001-4.

Neter J, Wasserman W, Kutner MH. 1989. *Applied Linear Regression Models*, 2nd edn. Richard D. Irwin, Inc.: Boston, MA.

Newman AJ, Clark MP, Craig J, Nijssen B, Wood A, Gutmann E, Mizukami N, Brekke L, Arnold JR. 2015. Gridded ensemble precipitation and temperature estimates for the contiguous United States. *J. Hydrometeorol.* **16**: 2481–2500, doi: 10.1175/JHM-D-15-0026.1.

NOAA National Weather Service (NWS). 1989. *Cooperative Station Observations. NWS Observing Handbook No. 2*. U.S. Department of Commerce: Silver Spring, MD.

PRISM Climate Group, Oregon State University [internet]. Descriptions of PRISM spatial climate datasets for the conterminous United States. 2015. http://www.prism.oregonstate.edu/documents/PRISM_datasets .pdf. (accessed 9 October 2015)

Rasmussen R, Baker B, Kochendorfer J, Meyers T, Landolt S, Fischer AP, Black J, Thériault JM, Kucera P, Gochis D, Smith C, Nitu R, Hall M, Ikeda K, Gutmann E. 2012. How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation test bed. *Bull. Am. Meteorol. Soc.* **93**: 811–829, doi: 10.1175/bams-d-11-00052.1.

Renard KG, Lane LJ, Simanton JR, Emmerich WE, Stone JJ, Weltz MA, Goodrich DC, Yakowitz DS. 1993. Agricultural impacts in an arid environment: Walnut Gulch case study. *Hydrol. Sci. Technol.* **9**: 145–190.

Sevruk B. 1982. Method of Correction for Systematic Error in Point Precipitation Measurement Gauges for Operational Use. Operational Hydrology Report No. 21 (WMO-589), Geneva, Switzerland.

Sharples JJ, Hutchinson MF, Jellet DR. 2005. On the horizontal scale of elevation dependence of Australian monthly precipitation. *J. Appl. Meteorol.* **44**: 1850–1865, doi: 10.1175/jam2289.1.

Slaughter CW, Marks D, Flerchinger GN, Van Vactor SS, Burgess MD. 2001. Thirty-five years of research data collection, Reynolds Creek Experimental Watershed, Idaho, USA. *Water Resour. Res.* **37**: 2825–2830, doi: 10.1029/2001wr000413.

de Smith MJ, Goodchild MF, Longley PA. 2007. *Geospatial Analysis – A Comprehensive Guide to Principles, Techniques and Software Tools*. Matador: Leicester, UK, 398.

Swank WT, Crossley DA. 1988. Introduction and site description. In *Ecological Studies: Forest Hydrology and Ecology at Coweeta*, Swank WT, Crossley DA (eds). Vol. **66**. Springer-Verlag: New York, NY, 3–16.

Swift Jr L. W., 1968. Comparison of methods for estimating areal precipitation totals for a mountain watershed. Presented at the Conference and Workshop on Applied Climatology, Amer. Meteorol. Soc., Asheville, NC, October 29–31, 1968. Abstract in *Bull. Amer. Meteorol. Soc.*, **49**: 782–783.

Swift LW Jr, Cunningham GB, Douglass JE. 1988. Climatology and hydrology. In *Ecological Studies: Forest Hydrology and Ecology at Coweeta*, Vol. **66**. Springer-Verlag: New York, NY, 35–55.

Tabios GQ, Salas JD. 1985. A comparative analysis of techniques for spatial interpolation of precipitation. *J. Am. Water Resour. Assoc.* **21**: 365–380, doi: 10.1111/j.1752-1688.1985.tb00147.x.

Thornton PE, Running SW, White MA. 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. *J. Hydrol.* **190**: 214–251, doi: 10.1016/s0022-1694(96)03128-9.

Trewartha GT. 1954. *An Introduction to Climate*. Mc-Graw Hill: New York, NY.

Ver Hoef JM, Krivoruchko K, Lucas N. 2003. *Using ArcGIS Geostatistical Analyst*. ESRI: Redlands, CA, 300.

Willmott CJ, Matsuura K. 1995. Smart Interpolation of annually averaged air temperature in the United States. *J. Appl. Meteorol.* **34**: 2577–2586, doi: 10.1175/1520-0450(1995)034<2577:SIOAAA>2.0.CO;2.

Willmott CJ, Robeson SM. 1995. Climatologically aided interpolation (CAI) of terrestrial air temperature. *Int. J. Climatol.* **15**: 221–229, doi: 10.1002/joc.3370150207.

Willmott CJ, Ackleson SG, Davis RE, Feddema JJ, Klink KM, Legates DR, O′ Donnell L, Rowe CM. 1985. Statistics for the evaluation and comparison of models. *J. Geophys. Res. Oceans* **90**: 8995–9005, doi: 10.1029/JC090iC05p08995.