

Constructing Retrospective Gridded Daily Precipitation and Temperature Datasets for the Conterminous United States

MAURO DI LUZIO

Blackland Research Center, Texas Agricultural Experiment Station, Texas A&M University System, Temple, Texas

GREGORY L. JOHNSON

West National Technology Support Center, U.S. Department of Agriculture Natural Resources Conservation Service, Portland, Oregon

CHRISTOPHER DALY

PRISM Group, Department of Geosciences, and College of Oceanographic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon

JON K. EISCHEID

Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado

JEFFREY G. ARNOLD

Grassland Research Laboratory, U.S. Department of Agriculture Agricultural Research Service, Temple, Texas

(Manuscript received 22 November 2005, in final form 27 March 2007)

ABSTRACT

This paper presents and evaluates a method for the construction of long-range and wide-area temporal spatial datasets of daily precipitation and temperature (maximum and minimum). This method combines the interpolation of daily ratios/fractions derived from ground-based meteorological station records and respective fields of monthly estimates. Data sources for the described implementation over the conterminous United States (CONUS) are two independent and quality-controlled inputs: 1) an enhanced compilation of daily observations derived from the National Climatic Data Center digital archives and 2) the Parameter–Elevation Regressions on Independent Slopes Model (PRISM) maps. The results of this study show that this nonconventional interpolation preserves the spatial and temporal distribution of both the PRISM maps (monthly, topography-sensitive patterns) and the original daily observations. Statistics of a preliminary point comparison with the observed values at high-quality and independent reference sites show a reasonable agreement and a noticeable improvement over the nearest station method in orographically sensitive areas. The implemented datasets provide daily precipitation and temperature values at 2.5-min (around 4 km) resolution for 1960–2001. Combining seamless spatial and temporal coverage and topographic sensitivity characteristics, the datasets offer the potential for supporting current and future regional and historical hydrologic assessments over the CONUS.

1. Introduction

Precipitation and temperature are driving variables in the simulation of physical and biological processes occurring in the landscape. Consequently, modeling

evaluations of a growing number of hydrological and environmental issues are increasingly requiring reliable areal (area average) meteorological datasets as well as practical methods for their generation. The availability of these datasets with adequate spatial and temporal resolution is particularly critical when modeling applications are designed over large spatial and temporal domains.

This is the case of the Conservation Effects Assessment Project (CEAP; Mausbach and Dedrick 2004), a

Corresponding author address: Mauro Di Luzio, TAES–Blackland Research Center, 720 E. Blackland Rd., Temple, TX 76502.

E-mail: mdiluzio@brc.tamus.edu

program initiated by the U.S. Department of Agriculture (USDA), which includes outlooks for water quality at the watershed as well as the national scale. Topics of CEAP are agricultural conservation practices, which are recognized as reducing local losses of soil, nutrients, pesticides, pathogens, and other biological and chemical materials from agricultural lands, as well conserving natural resources, enhancing the quality of the agroecosystem and wildlife habitat. The core of the CEAP's assessment approach binds hydrology-based software models simulating on a daily time step and a number of supporting digital databases to derive annual estimates of the agricultural conservation benefits obtained by the program for the nation. The spatial domain is the conterminous United States (CONUS), which is going to be segmented, at the first level, in a set of composing hydrologic areas, such as the 18 major water-resources regions out of the overall 21 as defined by the U.S. Geological Survey (USGS; Seaber et al. 1987). The currently established modeling spatial framework is supported by the Hydrologic Unit Modeling for the United States (HUMUS) databases. HUMUS (Arnold et al. 1998) includes databases on land use and nonpoint and point sources of pollution populated in correspondence to each elementary unit of the current watershed framework, the hydrologic unit codes (HUC) polygons, also named 8-digit cataloging units, developed by the USGS (Seaber et al. 1987).

The large spatial domain of the application project prevents using detailed onsite collection of near-surface data, which are only available for traditional large station networks and only recently from remote sensors (radars and satellites). Although a source of valuable information, with extended geographic distribution, duration of record, frequency of observations, and standardized measurement techniques, station network data still requires methods for the spatial generalization of point observations across application domains and remote sites. Much of the literature concerning the distribution of meteorological fields over large areas focuses on long-time-step average (e.g., seasonal, annual, monthly), satellite-based, global spectrum and low-resolution estimates (Krajewski et al. 2000; Meneghini et al. 2001), whereas short-time-step (e.g., hour, minute) radar precipitation and high-resolution estimates are only recently available (Fulton et al. 1998). Our objectives appertain to the spatial interpolation of daily meteorological fields (precipitation and temperature) from historical station networks data, which also has mainly dealt with long-time-step average estimates. A wide range of suitable interpolation methods, from simple (e.g., nearest neighbors, inverse distance) to sophisticated statistical (e.g., spline, kriging), have been

developed and tested (Tabios and Salas 1985; Seo et al. 1990; Kurtzman and Kadmon 1999; Xia et al. 2001), mainly challenging the more severe variability of the precipitation fields (Hubbard 1994). These methods, which are mainly based on a weighting and/or geostatistical combination of point values from station records, when applied to existing networks do not provide explicit evaluation of areal estimates, in particular over a range of elevation zones. Issues arise because the majority of measurements are traditionally taken from stations located at low elevations and with a very sparse coverage of the territory (Daly 2002; Groisman and Legates 1994). Precipitation assessment improvements have been obtained using supplementary information, mainly introducing topographic data into the basic interpolation or in more elaborate methods (Garen et al. 1994; Piper and Stewart 1996; Pardo-Iguzquiza 1998; Prudhomme and Reed 1999; New et al. 1999; Daly et al. 1994, 2002; Hijmans et al. 2005). In general, the improvements were achieved in zones where the spatial density of stations allowed establishing satisfactory relationships with the topography and/or for the estimation of annual and monthly amounts. Challenges are presented in particular by the variability of precipitation fields on a short time step, which have stimulated the development of methods combining radar and rain gauge measurements (Steiner et al. 1999), but which are not suitable for retrospective analysis because of the recent availability of the remote sensing data. To overcome these various issues (e.g., historical and regional coverage, topographic adjustment) and limitations (e.g., station location and density, variability for daily time step) daily gridded estimates data have been obtained using techniques based on the additional identification of meteorological patterns before the application of a spatialization procedure. To this end Thornton et al. (1997) introduced a daily precipitation occurrence algorithm, as a precursor to the prediction of daily precipitation amount, which was obtained using the spatial convolution of a truncated Gaussian weighting filter with the set of station locations. Courault and Monestiez (1999) proposed a conditioning approach of air temperature data taking into account the atmospheric circulation patterns identified from outputs of large-scale forecast models, before interpolating using an ordinary kriging. Hewitson and Crane (2005) conditioned a gridding inverse distance interpolation of daily precipitation to the procedural identification of the daily synoptic forcing state of the recording stations based on the identification of a spatial pattern of wet/dry conditions.

This paper describes the details and the implementation of a pattern-based method for the development

of daily total precipitation and temperature (daily maximum and minimum) gridded datasets, which are ultimately needed to meet the following modeling requirements: 1) seamless spatial coverage of the entire project application area (the CONUS), 2) representation of sequential daily values, 3) serially complete over an extended historical period, 4) an adequate resolution to support the applied hydrologic models at the current (and most probably becoming finer in the near future) spatial hydrologic segmentation, and 5) provision of orographic adjustment. The first requirement is dictated by the geographic scope of the model simulations along with the necessity, also related to the rest of the requirements, of their calibration using observed data (e.g., stream flows, sediment loads), which are expected to be correlated to the implementation of the conservation practices, thereby excluding the usage of generated weather records. In addition, as previously described, the model simulations require daily input time series. Their historical extent and spatial and temporal variability are fundamental for the achievement of the project goals and any water resource management plan. The inclusion of the strong variation of climate with elevation is obviously important and in addition provides background data for the concurrent estimation of the atmospheric deposition loads.

The remainder of this paper is subdivided into four sections. The first section describes the conceptual approach and the data sources input to the method implemented on the CONUS. The second section describes the method with specific references to the implementation data. The third section reports the results of basic verifications of the output data. The last section discusses and summarizes the main results.

2. Gridding approach and data sources

In developing our datasets, we embraced the concept that the station observations are the true values on a surface component dominated by the prevailing weather systems determined by large-scale synoptic forcing (atmospheric pattern) acting at the monthly base, mixed with a modulating daily component determined by local forcing. For the precipitation field $P(x, y, t)$, the magnitude of the first component conceptually was considered to be equivalent to the cumulative topography-based sum $P_c(x, y, t)$ modulated by a second component, such as pointwise time-varying ratios $P_r(x, y, t)$ defined by the daily pattern, such as

$$P(x, y, t) = P_c(x, y, t) \times P_r(x, y, t); \quad (1)$$

$P_r(x, y, t)$ was derived from the interpolation of ratios calculated using station records and the procedure described in the next section. In place of $P_c(x, y, t)$, for

each month we used the distinct Parameter–Elevation Regressions on Independent Slopes Model (PRISM; Daly et al. 1994, 1997, 2002) cumulative precipitation fields at the full resolution (4 km). This decomposition–interpolation–composition approach was introduced by Widmann and Bretherton (2000), and used in other linked research efforts (Maurer et al. 2002; Hamlet and Lettenmaier 2005) only for coarse griddings ($1/8^\circ$ spatial resolution and larger) on which the PRISM monthly fields were averaged and $P_r(x, y, t)$ was normalized before the interpolation. In our implementation we use the full-resolution (4 km) and also introduce a similar approach to obtain daily temperature fields (maximum and minimum), decomposed in its monthly mean $\bar{T}(x, y, t)$ and additive daily anomaly $T_a(x, y, t)$, such as

$$T(x, y, t) = \bar{T}(x, y, t) + T_a(x, y, t). \quad (2)$$

The anomaly-average ratio $T_r(x, y, t)$ was defined as

$$T_r(x, y, t) = T_a(x, y, t) / \bar{T}(x, y, t); \quad (3)$$

$T_r(x, y, t)$ was derived from the interpolation of fractions calculated using station records and the procedure described in the next section. In place of $\bar{T}(x, y, t)$, for each month we used the distinct PRISM average temperature (maximum and minimum) fields at the full resolution (4 km) that combined with $T_r(x, y, t)$ allowed the spatial distribution of the daily values representing the estimated daily fields (see implementation details below).

A deterministic interpolation method [the inverse-distance-weighted (IDW) method; Watson and Philip 1985] was implemented to specifically assign fraction values to missing locations based on the surrounding measured values. IDW, although lacking in optimality criteria, is in general recognized as more appropriate than the classic nearest-neighbor method (Thiessen 1911), which in turn introduces discontinuous surfaces and is traditionally used for large-area hydrological assessments. For a given estimation point, the IDW technique provides a set of weights that sum to unity and that are inversely related to the distances to the data points. The IDW(x, y) estimation at (x, y) is a linear combination of the observed values, such as

$$\text{IDW}(x, y) = \sum_j w_j f(x_j, y_j), \quad (4)$$

with the weights w_j defined as follows:

$$w_j = \frac{d_j^{-p}}{\sum_k d_k^{-p}}, \quad (5)$$

where d_k is the distance from (x, y) to (x_k, y_k) and $f(x_k, y_k)$ is the observed value at (x_k, y_k) ; p is a positive real number that influences the character of the inter-

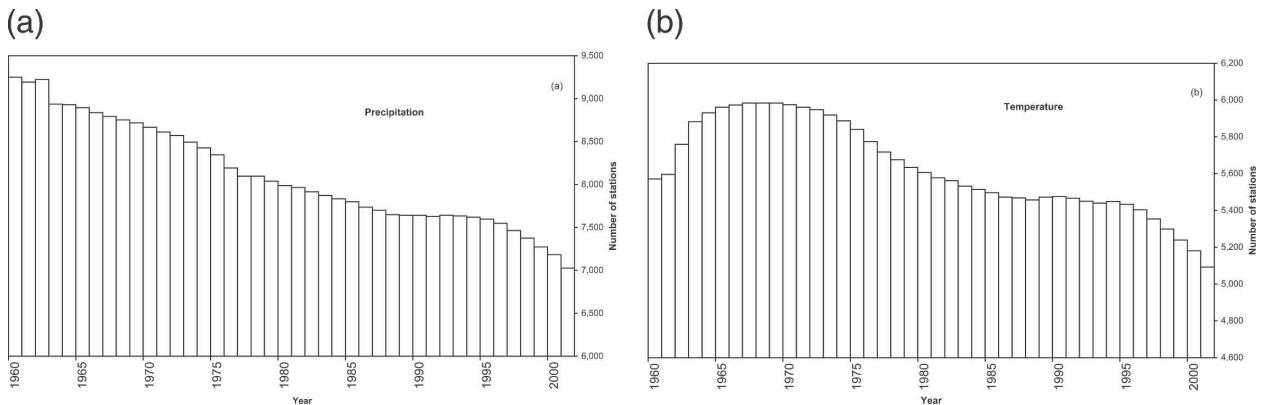


FIG. 1. Variation of the number of COOP stations measuring (a) precipitation and (b) temperature in the period 1960–2001.

polation, from local to global: the higher the value is, the stronger is the influence of the closer sample points. Here $p = 2$ has been used (inverse square interpolation), which still determines a local dominating weight to a particular measurement when it is located near the estimation point, and in addition returns a smooth transition of the interpolated surface (the first derivative is zero at the data point). The input set of data points (stations) for calculating each interpolated point have been limited to 12. Using IDW, the range of interpolated values is limited to the range of the measured variable. In general this is considered a major disadvantage because the interpolation is not fully responsive to local trends (e.g., for unsampled hill tops and valley bottoms). The importance of this issue is minimized here since the fractional values are interpolated and the local trends are accounted for at the monthly level.

The implementation for the CONUS used the following precompiled datasets: 1) corrected and quality-controlled National Weather Service Cooperative Observer (COOP) daily observations from the National Climatic Data Center (NCDC) and 2) the PRISM monthly grid estimates, both of which are here briefly described along with the national segmentation currently adopted in the project.

a. Daily precipitation and temperature data and the hydrologic units map

The dataset used in this project is a serially complete (no missing values) daily total precipitation and maximum–minimum temperature time series developed initially for the western United States (Eischeid et al. 2000) and extended to the entire United States. It was compiled, purging and/or correcting extreme errors and/or missing values traditionally included in observation records, creating quality-controlled, serially complete data in support of natural resource modeling. The

source records were from the COOP stations: namely the NCDC Summary of Day (TD3200) data. In the creation of the final serially complete datasets the following refinement steps were applied: 1) quality control that identifies unreliable reporting stations and records that had been flagged as missing values; 2) replacement of missing daily values based on the use of simultaneous values at nearby stations along with six different, seasonally and geographically dependent, methods of spatial interpolation to calculate estimated values for those specific days (Eischeid et al. 1995); and 3) a final data-consistency check and eventual correction.

The total numbers of distinct stations in the serially complete dataset, operative in the period 1895–2001, are 12 540 and 7998, respectively, for precipitation and maximum–minimum temperature. The number of operative stations and their distribution results change over the years. Our target period (1960–2001) avoids large temporal data inhomogeneities in the COOP data prior to 1950 that were noted by Hamlet and Lettenmaier (2005). These persistent temporal inhomogeneities are due to undocumented changes in stations and station locations. To limit this problem, whereas Eischeid et al. (2000) retained the stations with at least 10 yr of data and with no more than 48 missing months, a procedure has been applied to remove a few spatially redundant stations (with the same coordinates), retaining only the longest recording station. The numbers of distinct stations recording at any time within our target period (1960–2001) are 11 680 and 7565 for precipitation and temperature, respectively. Figures 1a,b show the temporal variation of the number of the measuring COOP stations. The spatial distribution of the stations is shown in Figs. 2a,b (maximum and minimum temperature stations share identical locations) with reference to the USGS water-resources regions shown in Fig. 2c.

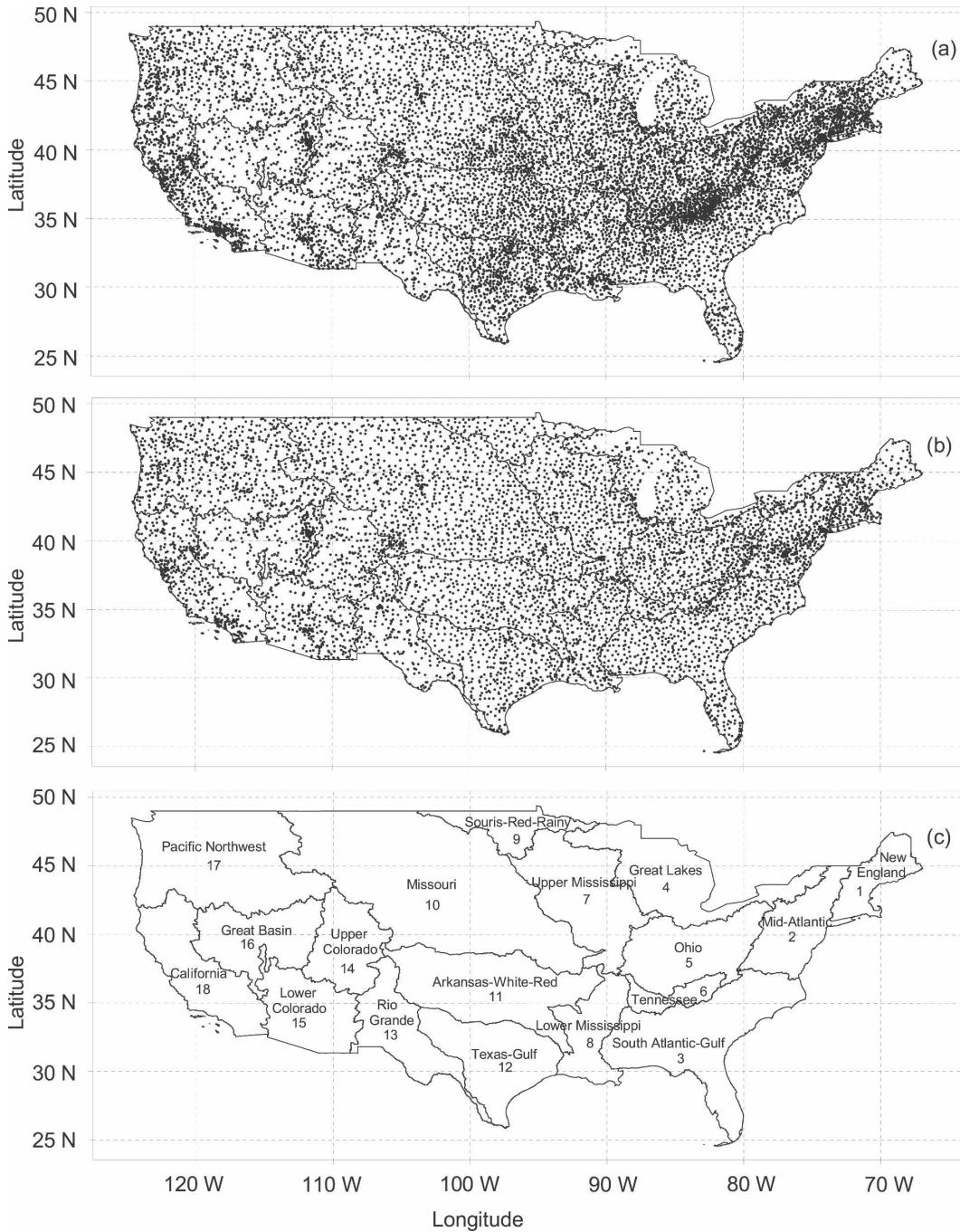


FIG. 2. Spatial distribution of the (a) 11 680 COOP stations measuring precipitation and (b) 7565 COOP stations measuring temperature in the period 1960–2001. (c) USGS water-resources regions in the CONUS.

The supporting digital spatial dataset used in this project is the 1:250 000-scale HUC (Steeves and Nebert 1994), which counts 2150 HUCs for the entire United States. After some minor simplifications and aggregations, the revised dataset contains 2108 units. Figures 3a,b show for precipitation and temperature, respectively, the distribution (average and standard deviation)

of the area associated with each station and of the number of stations per HUC in each of the 18 water-resources regions in the CONUS. There are 82 HUCs without precipitation stations. These HUC polygons are shaded in Fig. 4a. The most precipitation stations (43) are located within HUC number 6010105, located in region 6, across Tennessee and North Carolina. HUC

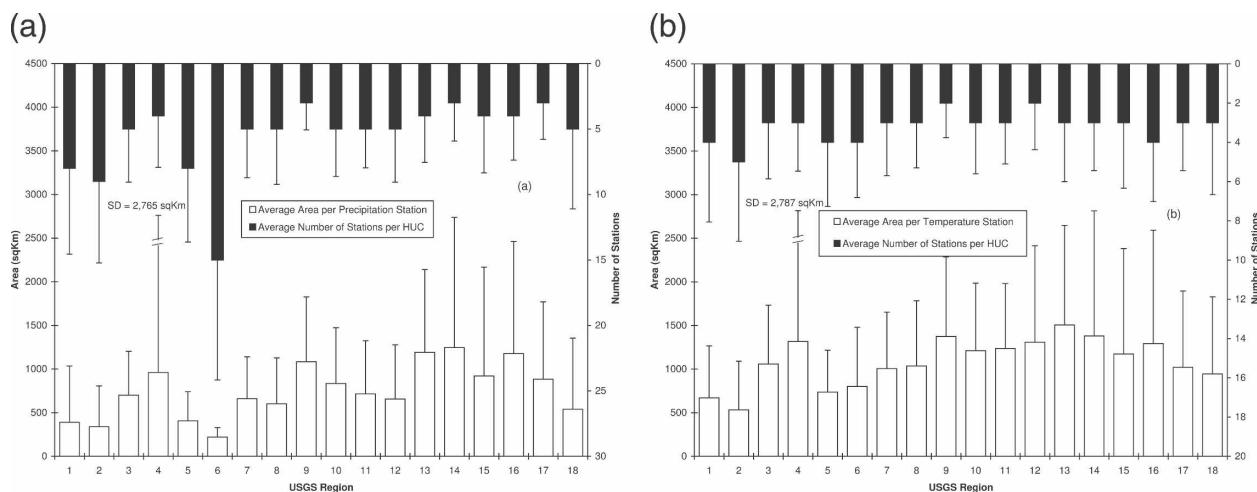


FIG. 3. Average and standard deviation of the area of each station and of the number of stations per HUC measuring (a) precipitation and (b) temperature in the period 1960–2001 in each of the 18 regions of the CONUS.

number 18070106, located in region 18, in southern California, has the smallest mean area associated with each station (59.75 km^2). These HUC polygons are also highlighted in Fig. 4a. There are 141 HUCs without temperature stations. These HUC polygons are shaded in Fig. 4b. The most temperature stations (29) are located within HUC number 3090202, located in region 3, in southern Florida. HUC number 18070106, located in region 18, in northern California, has the smallest mean area associated with each station (96.22 km^2). These HUC polygons are also highlighted in Fig. 4b.

b. PRISM grids

The PRISM climate mapping system was used to create the gridded climate datasets described in this study. PRISM is a knowledge-based system that uses point data, a digital elevation model (DEM), and many other geographic datasets to generate gridded estimates of monthly and event-based climatic parameters (Daly et al. 1994, 2001, 2002, 2003; Daly 2006). PRISM has been used extensively to map precipitation, temperature, dewpoint, weather-generator parameters, and other climate elements over the United States, Canada, China, and other countries (USDA-NRCS 1998; Daly and Johnson 1999; Johnson et al. 2000; Plantico et al. 2000; Daly et al. 2001, 2003; Gibson et al. 2002; NOAA-NCDC 2002; Zhu et al. 2003; Daly and Hannaway 2005; Hannaway et al. 2005; Simpson et al. 2005).

PRISM adopts the assumption that for a localized region, elevation is the most important factor in the distribution of temperature and precipitation (Daly et al. 2002). PRISM calculates a linear climate–elevation relationship for each DEM grid cell, but the slope of

this line changes locally with elevation as dictated by the data points. Beyond the lowest or highest station, the function can be extrapolated linearly as far as needed. A simple, rather than multiple, regression model was chosen because controlling and interpreting the complex relationships between multiple independent variables and climate is difficult. Instead, weighting the data points (discussed later) controls the effects of variables other than elevation. The climate–elevation regression is developed from x, y pairs of elevation and climate observations supplied by station data. A moving-window procedure is used to calculate a unique climate–elevation regression function for each grid cell. The simple linear regression has the form

$$Y = \beta_1 X + \beta_0, \quad (6)$$

where Y is the predicted climate element; β_1 and β_0 are the regression slope and intercept, respectively; and X is the DEM elevation at the target grid cell.

Stations surrounding the target grid cell provide the X and Y pairs for the regression function. Upon entering the function, each station is assigned a weight that is based on several factors. In the general PRISM formulation, the combined weight of a station is a function of distance, elevation, cluster, vertical layer, topographic facet, coastal proximity, and effective terrain weights, respectively. Thus, the combined weight W of a station is a function of the following:

$$W = f(W_d, W_z, W_c, W_p, W_b, W_f, W_e), \quad (7)$$

where W_d , W_z , W_c , W_p , and W_e are the distance, elevation, cluster, coastal proximity, vertical layer, topographic facet, and effective terrain weights, respec-

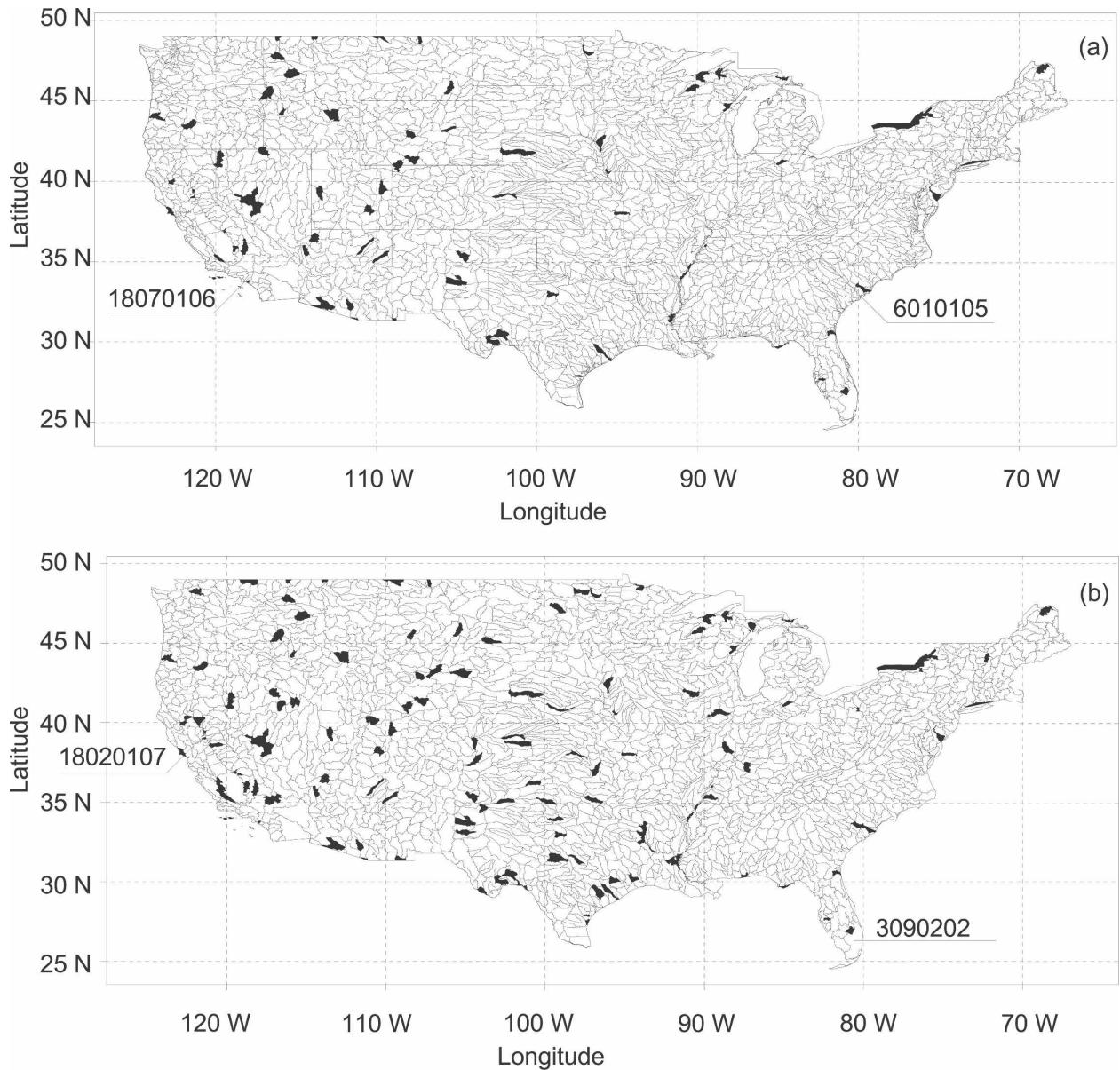


FIG. 4. Hydrologic units map of the United States: Highlighted are the HUCs without a station measuring (a) precipitation and (b) temperature in the period 1960–2001. In (a), HUC 6010105 has the highest number and HUC 18070106 has the smallest mean associated area; in (b), HUC 3090202 has the highest number and HUC 18070106 has the smallest mean associated area.

tively. Distance, elevation, and cluster weighting are relatively straightforward in concept. A station is down-weighted when it is relatively distant or at a much different elevation than the target grid cell, or when it is clustered with other stations (which leads to overrepresentation). Coastal proximity weighting is used to define gradients in precipitation or temperature that may occur because of proximity to large water bodies (Daly et al. 2002, 2003; Simpson et al. 2005). A description of a trajectory model developed to estimate coastal proximity for Puerto Rico is provided in Daly et al. (2002);

this model is routinely used in the United States. Vertical layer weighting is used to simulate situations where rapid changes, or even reversals, in the relationship between climate and elevation are possible (i.e., temperature inversions). When the potential for such situations exists, the climate stations entering the regression are divided into two vertical layers, and regressions run on each separately. Layer 1 represents the boundary layer, and layer 2 represents the free atmosphere above it (Daly et al. 2002). Topographic facet weighting accounts for discontinuities in the climate

field caused by terrain breaks (e.g., rain shadows). A topographic facet is a contiguous terrain slope with a common orientation, delineated at a variety of scales, from the major leeward and windward sides of large mountain ranges, to north- and south-facing hill slopes. At each grid cell, the model chooses the topographic facet scale that best matches the data density and terrain complexity, and assigns the highest weights to stations on the same topographic facet (Daly et al. 2002). Effective terrain weighting accounts for differences in the ability of terrain features to enhance precipitation through mechanical uplift of moisture-bearing air. Features having relatively steep, bulky profiles typically produce strong precipitation–elevation relationships; while low, gently rolling features have weaker relationships (Daly 2002). Detailed discussions of these weighting functions, and typical PRISM parameter values used in the CONUS, are available from Daly et al. (2002) and Daly (2002). [Many of the papers cited here, as well as a downloadable presentation on PRISM, were available from the PRISM Internet site at <http://prismclimate.org>.]

The development of the monthly time series grids used in this study was performed in two phases. The first phase involved the period 1895–1997 and was a collaborative effort between Oregon State University (OSU), the National Center for Atmospheric Research (NCAR), and NCDC. NCDC provided all available COOP station data, including those from the Historical Climate Network (HCN), and pre-1948 data digitized by the Midwestern Climate Center. These sources were supplemented by data from the USDA–Natural Resources Conservation Service Snowpack Telemetry (SNOTEL) network, as well as those from various local networks. A two-tiered spatial quality-control (QC) system developed by OSU was applied to the station data. The QC system identified metadata errors, including station location and elevation problems, as well as spatial inconsistencies in the monthly data (Gibson et al. 2002). NCAR used spatial statistical tools to create serially complete monthly data for all stations for the entire 1895–1997 period (Johns et al. 2003). At OSU, PRISM was parameterized for optimal performance by applying a statistical version of the model that determines optimal model performance from a jackknife cross-validation error analysis. Run for a subset of years and months in an iterative fashion, and a robust, universal parameterization sufficient for all months and years was obtained. PRISM was then applied to the CONUS in three overlapping regions: western, central, and eastern; these regions were then merged together into a single grid with a resolution of 2.5 arc min (~4 km). The resulting grids were examined for accuracy

and reasonableness by OSU, NCAR, and NCDC. The grids served as the basis for the Vegetation–Ecosystem Modeling and Analysis Program bioclimatic dataset (Kittel et al. 2004), used in U.S. Global Change Research Program assessments.

Immediately after the release of this dataset, there were calls for the dataset to be updated to a more current year. The second phase of development involved developing an operational system for producing monthly temperature and precipitation grids from 1998 to the present, and extended on a near–real time basis (Daly et al. 2004). OSU worked in partnership with the Western Regional Climate Center, which provided near–real time COOP and National Weather Service Automated Service Observing System station data. These were supplemented with data from SNOTEL, other national networks, and regional mesonets such as those operating in Oklahoma, Minnesota, and Texas.

PRISM mapping of near–real time data required an approach slightly different than the standard modeling procedure. The amount of station data available for recent months is always much less than that available for the high-quality climatological PRISM precipitation and temperature maps, such as the expert-reviewed PRISM 1961–90 mean precipitation maps (USDA–NRCS 1998). This meant that mapping climate in recent months using available data only would sacrifice a significant amount of the spatial detail present in the long-term mean climate maps. Results from a pilot study showed that the PRISM 1961–90 mean monthly climate maps were excellent predictors of a given month’s climate, much better than elevation, which is typically the predictor variable in PRISM applications. The relationships between 1961–90 mean monthly and individual monthly climate values were strong because much of the incorporation of the effects of various physiographic features on climate patterns had already been accomplished through the careful creation of the 1961–90 grids. This method is in common use, and was termed “climatologically-aided interpolation” (CAI) by Willmott and Robeson (1995).

PRISM was parameterized to use existing 1961–90 mean monthly grids of minimum and maximum temperature and precipitation as the predictor grids in the interpolation of these climate elements for recent months. The PRISM weighting functions for distance, elevation, topographic facet, atmospheric layer, orographic effectiveness, and coastal proximity were all retained. This allowed the moving-window regression function of observed monthly climate with the 1961–90 mean to be interpolated with sensitivity to physiographic factors, such as might occur when a month has below-normal temperatures along the coast and above-

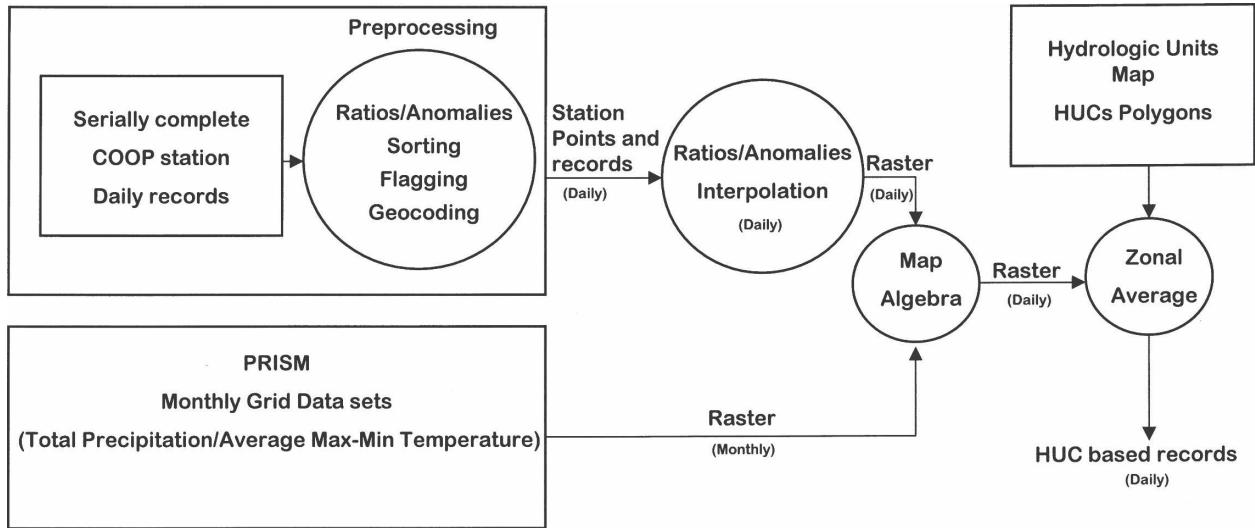


FIG. 5. Scheme of the input–output data elaboration.

normal temperatures inland. PRISM was run in this fashion for the period January 1998–December 2002 to complete a seamless time series from 1895 to 2002. Starting in 2002, PRISM has been run operationally to produce grids of the most recent month, as well as previous months for which there are new station data available. When a new version of a monthly grid is produced, it replaces the old one. The resulting process is akin to a several-month temporal moving window of modeling and remodeling of climate elements. Each month typically goes through about 6–9 iterations, depending on the length of time for all of the COOP data to become available in final form. Occasionally it is necessary to update a month in the more distant past if new data or a new version of the present data becomes available. A major update of the entire gridded time series is expected in 2008. The update will use a consistent CAI interpolation method for all years, using new, 1971–2000 PRISM mean climate grids as the predictors. When completed, the resulting dataset will be available for download from the PRISM Internet site (<http://prismclimate.org>).

3. Implementation

Daily precipitation and maximum–minimum temperature spatial datasets were created by linking and combining the data sources outlined above (daily station records and PRISM monthly grids). The implementation, illustrated in Fig. 5 and conceptually described in section 2, relies on the input PRISM grids to reproduce the climate patterns as well as to fasten the accumulated values on the monthly base (total precipitation and average daily temperatures). The linkage is

established by defining, for each station in the database and for each day of the analyses, the daily fractional contribution to the total monthly precipitation and the fractional daily anomaly with respect to the average monthly values for daily maximum and minimum temperature (see first and second procedures below). An IDW interpolation function is applied to expand these point-sample ratios (fractional anomalies) over the spatial scope, such as the CONUS territory (see third procedure below). The spatial combination (see fourth procedure below) of the resulting grids with the PRISM grids returns the new spatial dataset at the daily time step, which is expected to be consistent with the monthly precipitation totals (monthly average daily temperatures) provided by the PRISM grids. An additional spatial function was applied to obtain records at the HUC level (see fifth procedure below).

The method was applied to the project target period (1960–2001) by incorporating five specific procedures. The first procedure is processing of the serially complete station records. This procedure included the following tasks: 1) identification and extraction of the stations operating any time within the target period (see Figs. 2a,b), 2) flagging of the station days on which the respective station had not been operative (this is for the days outside the gapless station-specific serially complete period), and 3) for each selected precipitation station computation of the monthly precipitation total (called monthly P_c) and ultimately the daily ratio (called daily P_r) of the monthly total precipitation. This computation is done as follows:

$$\text{daily } P_r = \text{daily } P / \text{monthly } P_c, \quad (8)$$

where “daily P ” is daily recorded precipitation. For each selected temperature station, monthly average “monthly \bar{T} ” is calculated (independently for maximum and minimum temperature) as

$$\text{monthly } \bar{T} = \sum_{i=1}^N \text{daily } T_i / N, \quad (9)$$

where “daily T_i ” is daily recorded temperature at the day i and N is the number of days in the respective month. The daily anomaly-average ratio “daily T_r ” from the average monthly temperature is calculated as follows:

$$\text{daily } T_r = \frac{\text{daily } T - \text{monthly } \bar{T}}{\text{monthly } \bar{T}}. \quad (10)$$

To avoid problematic zeros, all temperature computations were operated on a shifted dominion in which a value equal to 100 was added to all of the variables.

The second procedure is to arrange the data by date. Derived data were reorganized into a comprehensive sequence of daily records, such as a database table. Each record (line of the database) contained all station data for a single day. Data included the daily ratios, the anomalies, and flags calculated for all available stations (see the second task of procedure 1). This time-indexing procedure was made and applied to facilitate the computations that follow.

The third procedure is a sequence of interpolations for the time series of daily ratios. Following geocoding of the station location points, this procedure provided a day-by-day sequential interpolation of the daily ratios (see procedure 2), dynamically associated only with the stations operative on the currently analyzed day. The resulting IDW continuous surfaces are the daily grids (raster data) that cover the target period and represent the spatial extension from the sampled locations to all the locations where measures are not available of the daily ratios on the monthly bases. The spatial analysis environment used for this procedure was adopted from the PRISM datasets: namely, a geographic, World Geodetic Spheroid 1972 (WGS72) coordinate system, and the 2.5-min (~ 4 km) cell resolution.

The fourth procedure involves daily spatial combination. This procedure included the application of map algebra functions for combining the surface interpolation maps (ratios; see procedure 3) and the respective PRISM monthly grids. The map algebra functions combine data on a cell-by-cell basis to derive the final target information grid dataset. In this way, operating on each cell, the target daily precipitation grid was obtained as the result of the following combination:

$$\text{daily } P(i) = \text{daily } I_r(i) \times \text{monthly } P_c(i), \quad (11)$$

where daily $P(i)$ is precipitation grid at day i , daily $I_r(i)$ is grid of IDW interpolated station ratios [see Eq. (8)] at day i , and monthly $P_c(i)$ is PRISM total precipitation grid for the respective month. The daily temperature grid (maximum and minimum) was obtained using the following combination:

$$\text{daily } T(i) = \text{monthly } \bar{T}(i) \times \text{daily } I_r(i) + \text{monthly } \bar{T}(i), \quad (12)$$

where daily $T(i)$ is temperature grid at day i , daily $I_r(i)$ is grid of IDW interpolated station anomaly-average ratios [see Eq. (10) deviation from the monthly average] at day i , and monthly $\bar{T}(i)$ is PRISM average temperature grid for the respective month.

The fifth and last procedure is to create a hydrologic-unit average. A further step was required to provide the CEAP models with daily time series over each computation unit (HUC watersheds). For this aim, each of the previous grids was spatially averaged (simple average of all HUC-contained grid cells) within each HUC.

The daily grids (precipitation and maximum–minimum temperature in the period 1960–2001 at the 2.5-min resolution) and, secondary, the daily records stored for each HUC polygon in the national framework are the results of the implementation. These results were then verified as described in the next section.

4. Verification

Two kinds of tests were performed on the created datasets: the first one verifies the overall areal consistency with the reference PRISM grids, and the second one includes point-based comparisons between predicted and observed values at selected COOP and reference station locations.

a. Areal consistency

The areal (average area) consistency was analyzed on the 2108 HUC polygons of the simulation framework used in the CEAP project as well as on each elementary 4-km cell. In each of these geographic features (HUC and cell), the newly generated datasets and the monthly PRISM grids were accumulated in our target period, equivalent to 42 yr, 504 months, or 15 341 days.

For precipitation, the difference in percentage (PRISM subtracted from predicted) is summarized for the 18 large basins (shown in Fig. 2c) in Fig. 6a. In general, the resulting gridded values have a tendency to underestimate slightly the PRISM accumulated values (mean difference less than 1%), with higher gaps (mean negative

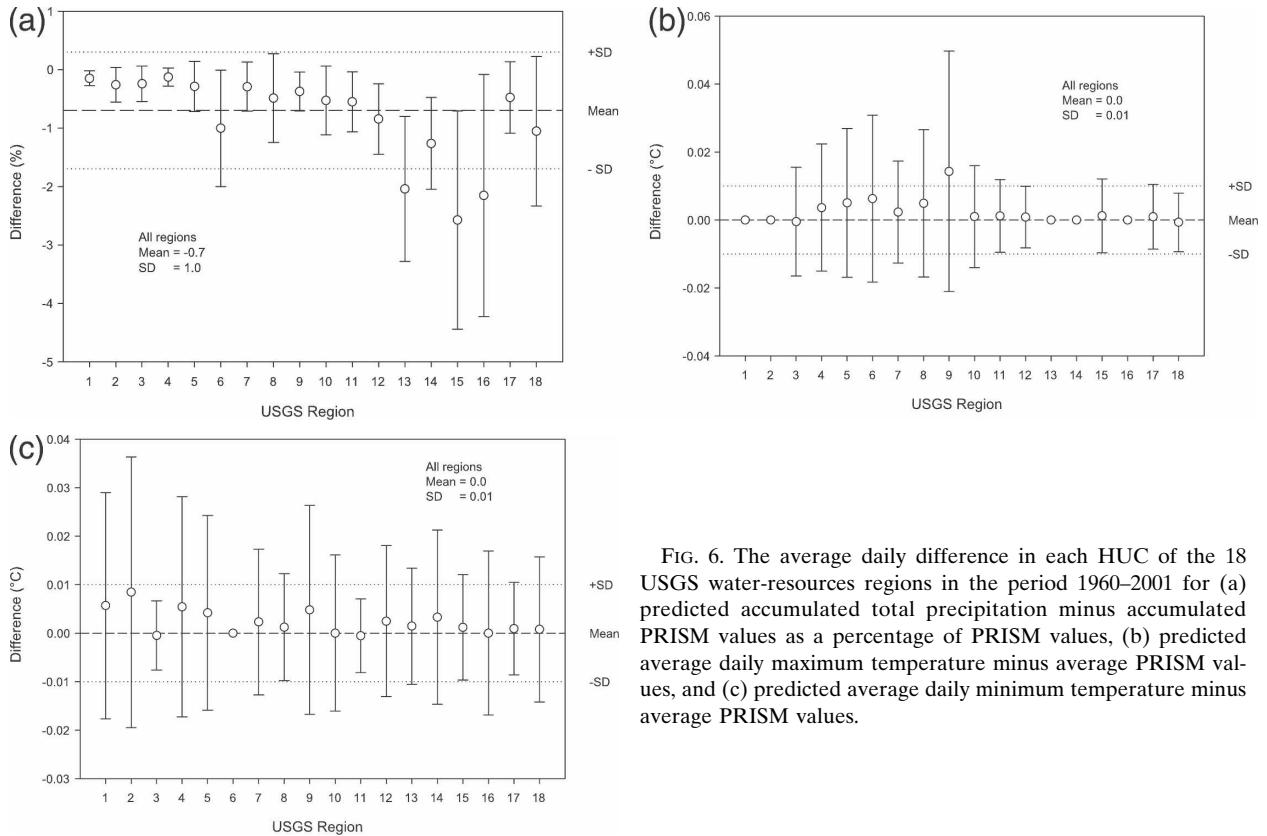


FIG. 6. The average daily difference in each HUC of the 18 USGS water-resources regions in the period 1960–2001 for (a) predicted accumulated total precipitation minus accumulated PRISM values as a percentage of PRISM values, (b) predicted average daily maximum temperature minus average PRISM values, and (c) predicted average daily minimum temperature minus average PRISM values.

difference around 2%–3%) seen in regions 13, 15, and 16. A linear regression (the plot is not shown) between the accumulated monthly PRISM records (independent variable) and the accumulated daily records predicted (dependent variable) in all the HUCs returned a correlation coefficient R equal to 0.99 and a slope very close to 1. Maps of the annual average PRISM and of the predicted precipitation grids are shown side by side in Figs. 7a and 7b. The two maps show a general agreement in terms of spatial distribution, patterns, and range of values. The cell-by-cell difference map (PRISM subtracted from predicted, as percentage of PRISM) is depicted in Fig. 7c. This figure confirms a general agreement between the predicted and the PRISM datasets. A number of spots where the gaps appear more pronounced are visually noticeable and more frequent the arid regions 13, 15, and 16, as previously noted. A closer comparison of the monthly records relative to the COOP stations and the PRISM fields for these and other sensitive zones (i.e., those located in region 6), revealed local divergences, with PRISM values in general greater than the respective station ones.

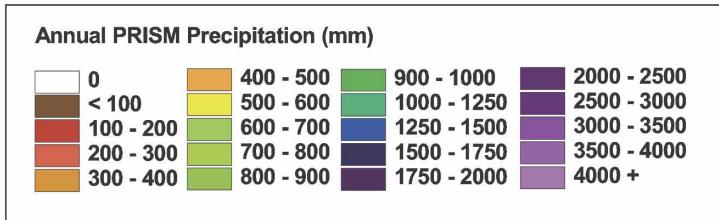
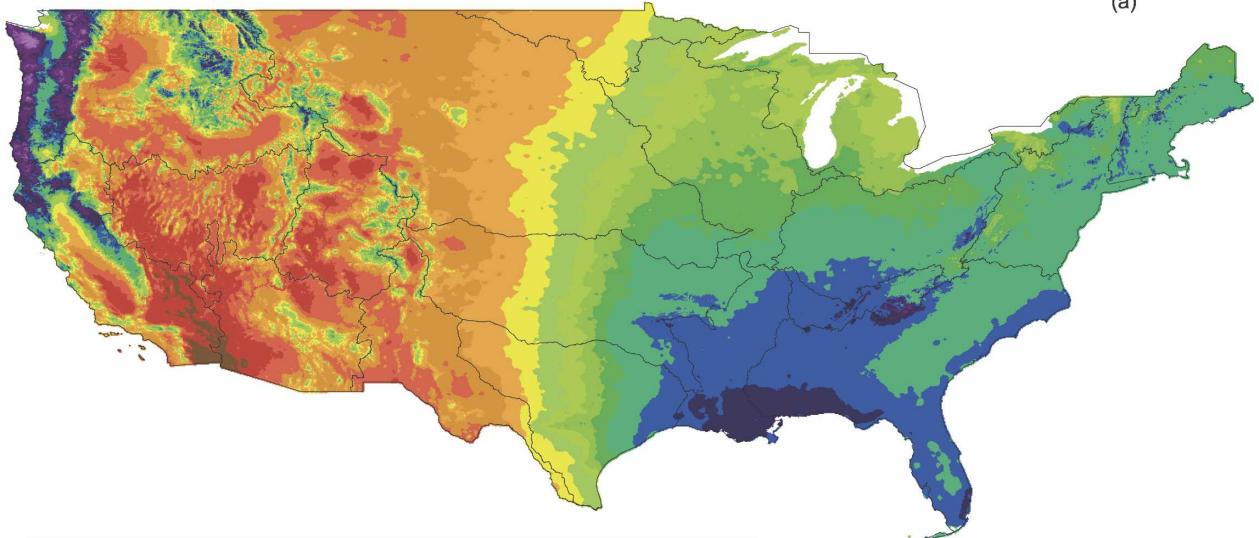
For temperature, the average daily differences (PRISM subtracted from the predicted) are summarized for the

18 large basins (shown in Fig. 2c) in Fig. 6b (maximum temperature) and Fig. 6c (minimum temperature). In general the resulting gridded values resemble the PRISM accumulated values on the HUC areas within each region. This is also confirmed by the linear regressions (the plots are not shown) between daily average temperatures (maximum and minimum) from the PRISM records (independent variable) and the predicted daily records predicted (dependent variable) in all the HUCs, which returned in both cases a correlation coefficient (R) very close to 1 and a slope equal to 1. Similar to the precipitation, a cell-by-cell analysis was obtained subtracting the daily average grid obtained from the PRISM maps from the daily average obtained using the predicted grids. The results (the maps are not shown) show less than $(1/10)^\circ\text{C}$ of gap across the country, both for the maximum and minimum daily average temperature.

b. Point-based comparison

The series of predicted grid values were compared with gauge records collected at COOP and reference stations. Because of the large amount of data, the findings are summarized to illustrate the general results. For precipitation, percent bias is defined as

(a)



(b)

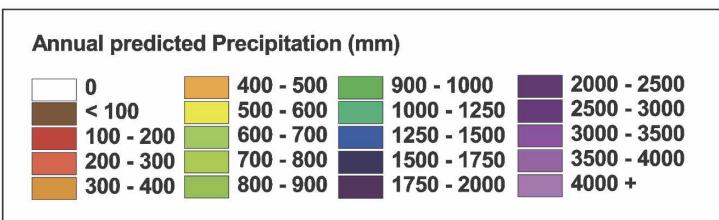
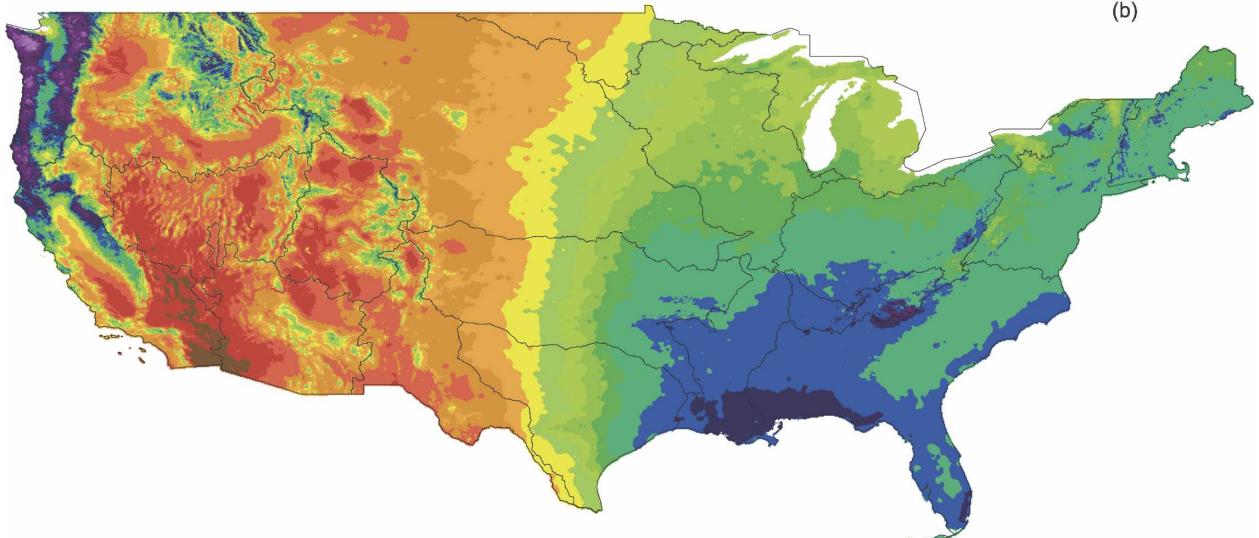


FIG. 7. Annual average (a) PRISM and (b) predicted precipitation accumulated in the period 1960–2001. (c) Cell-by-cell difference between predicted and PRISM average annual precipitation (PRISM subtracted from the predicted) as a percentage of PRISM in the period 1960–2001.

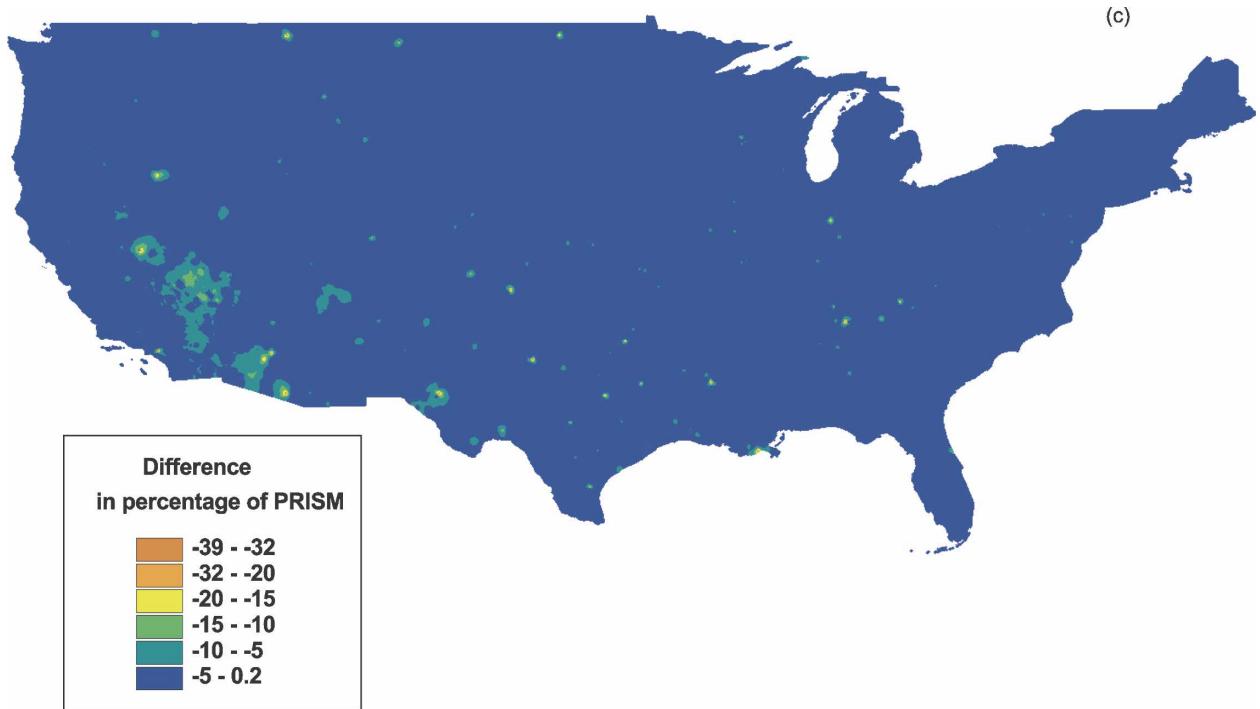


FIG. 7. (Continued)

$$\text{bias_}\% = \frac{\sum_{i=1}^N (P_i - O_i)}{\sum_{i=1}^N O_i} \times 100$$

and percent mean absolute error is defined as

$$\text{MAE_}\% = \frac{\sum_{i=1}^N |P_i - O_i|}{\sum_{i=1}^N O_i} \times 100.$$

Bias, in degrees Celsius for temperature and in millimeters for precipitation, is defined as

$$\text{bias} = \frac{1}{N} \sum_{i=1}^N (T_i - O_i),$$

and the mean absolute error, in degrees Celsius for temperature and in millimeters for precipitation, is defined as

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |T_i - O_i|,$$

where P_i and T_i are the predicted precipitation and temperature values (generated gridded values at the station location) and O_i are observed values (observations from the station), for the i th time step of the total N , respectively. In addition, root-mean-square error (RMSE) and correlation coefficient R , calculated for each reference station, are reported using daily values.

1) COOP STATIONS

A subset of 1800 stations, 100 stations for each of the 18 USGS regions, was randomly selected from the entire database (see section 2a). Results of the comparison between the predicted and observed daily records are reported in Table 1 for precipitation and Table 2 for maximum and minimum temperature. Table 1 also reports differences calculated for the total accumulated volumes for the period 1960–2001.

Table 1 shows that, in general, the newly estimated precipitation values slightly change and overpredict the recorded values. The overall estimated precipitation volumes deviate less than 2% from the observed records. Only for regions 6, 14, and 15 was this limit slightly exceeded (2.39%, 2.11%, and 2.13%). Average R values were all above 0.95 and bias_% was all positive and within 1.8, except for region 6 (6.2%). Bias values were all positive and less than 0.1 mm, except for region 6 (0.21 mm).

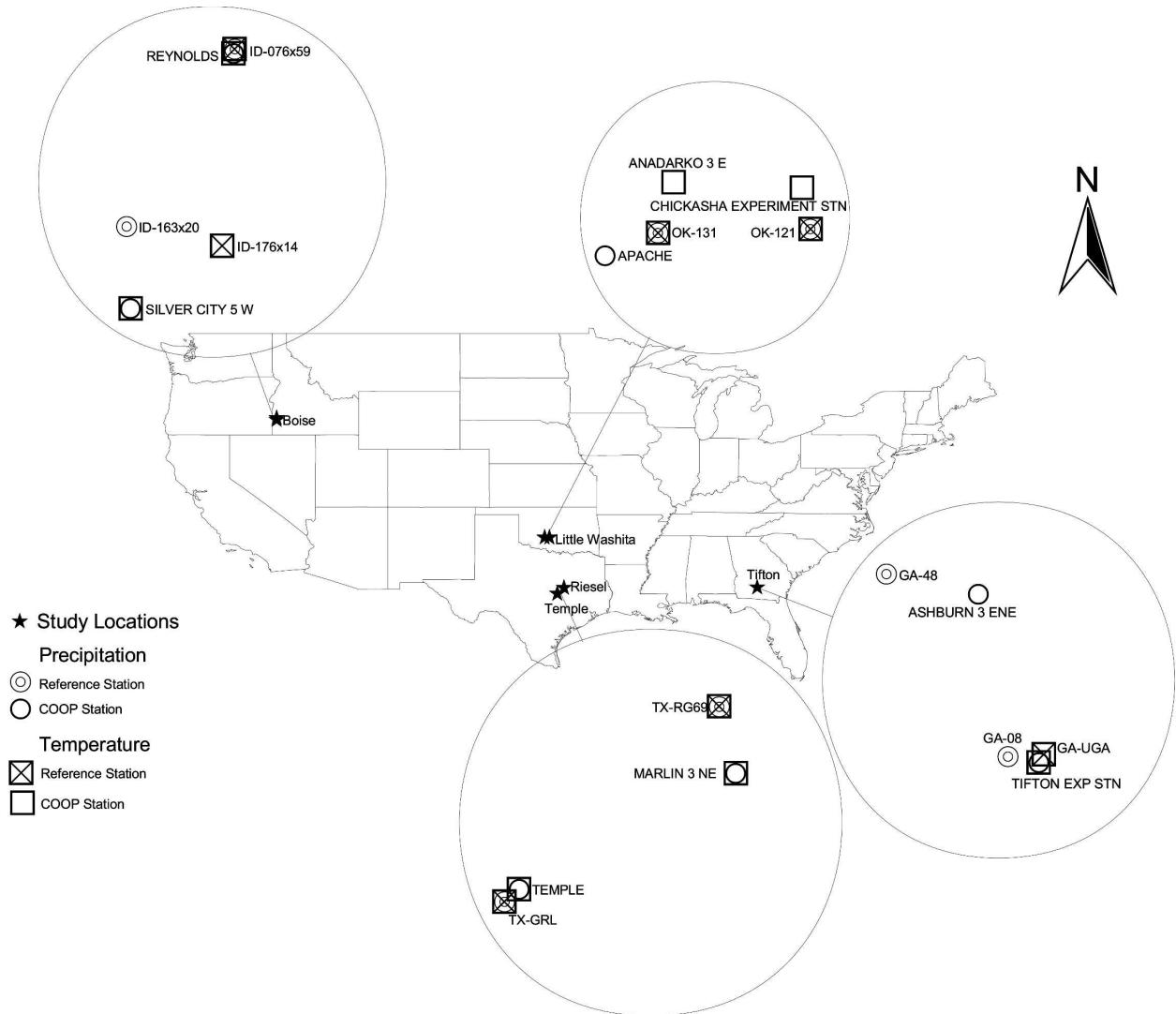


FIG. 8. Location map of the study areas, showing the stations used in the point comparison.

Noticeable are the MAE_% and MAE values, which are higher than the respective bias_% and bias values. The latest note indicates that the predicted and observed values are similar in intensity but with gaps of storm timing occurrence (rain/no-rain phase) introduced by the contribution of the surrounding stations. Thus, RMSEs parallel the bias_% values.

Table 2 shows that in general the newly estimated daily maximum temperature values resemble the observed ones. The *R* values are all above 0.99, and the bias absolute values are within 0.15°C, except for regions 1, 14, 15, 17, and 18 (biases are positive and within 0.5°C). MAE values are all within 1°C. Similar results are obtained for the minimum temperature, with absolute values of bias exceeding 0.15°C in region 17 (0.27°C) and region 18 (0.44°C).

2) REFERENCE STATIONS

The selected stations are not part of the COOP database. For precipitation, the data from eight watershed research stations managed by the U.S. Department of Agriculture Agriculture Research Service (USDA-ARS) were used (Slaughter and Richardson 2000). For maximum and minimum temperature, data from seven stations, including six ARS stations and one managed by the University of Georgia, were used. As shown in Fig. 8, the stations are located in Texas, Idaho, Oklahoma, and Georgia. Daily records from all of these stations were readily available over the Internet (except for the precipitation stations located in Georgia). Among the multiple stations available for each watershed, stations located at different elevations were se-

TABLE 3. Names, locations, and other information for the reference stations used in the point-based comparison; *P* = precipitation, and *T* = temperature.

Location	Name	Length of record/data type	Lat	Lon	Elev (m)
Riesel, TX	TX-RG69	1960–2001/ <i>P, T</i>	31°28'40"N	96°53'07"W	174
Temple, TX	TX-GRL	1960–2001/ <i>P, T</i>	31°03'23"N	97°20'55"W	195
Boise, ID, Reynolds Creek	ID-076x59	1962–96/ <i>P, 1964–96/T</i>	43°12'11"N	116°44'57"W	1207
Boise, ID, Reynolds Creek	ID-163x20	1962–96/ <i>P</i>	43°04'38"N	116°49'34"W	2170
Boise, ID, Reynolds Creek	ID-176x14	1967–96/ <i>T</i>	43°03'49"N	116°45'29"W	2097
River Watershed, OK, Little Washita	OK-121	1995–2001/ <i>P, T</i>	34°57'31"N	97°53'55"W	342
River Watershed, OK, Little Washita	OK-131	1995–2001/ <i>P, T</i>	34°57'01"N	98°14'01"W	458
Tifton, GA	GA-48	1968–2001/ <i>P</i>	31°44'42"N	83°44'27"W	145
Tifton, GA	GA-08	1968–2001/ <i>P</i>	31°29'12"N	83°34'25"W	89
Tifton, GA	GA-UGA	1996–2001/ <i>T</i>	31°29'39"N	83°31'35"W	120

lected. Table 3 shows the main characteristics of the stations along with the length of the records used in this study.

We compared the following time series: (a) the values recorded at the reference stations listed in Table 3, (b) the records extracted from the derived grids at the respective locations, and (c) the values recorded at the respective nearest COOP stations. Table 4 lists the information of the COOP stations selected for this elaboration, including the calculated distances from the reference stations. The locations of these stations are also mapped in Fig. 8. Properties comparisons were calculated in the respective analyses period indicated in Table 3. Table 5 and Table 6 report statistics for precipitation and temperature, respectively, calculated for each reference station using daily values. Both tables report two sets of statistics calculated against the reference data, the first set using the predicted time series and the second set using nearest COOP station time series. Their comparison is used in the remainder of the paper to indicate improvement and/or worsening of the predicted results. These tables show several elements.

First, both for precipitation and temperature, the predicted data series reasonably resembles the observed data. As expected, the temperature statistics are better than the respective precipitation given the known different degree of variability. Second, in general the predicted time series is an improvement over the nearest COOP station values. The *R* values are increased and/or other statistics decreased (consider absolute values of the calculated biases). The precipitation results for station ID-163x20 stand out: *R* value increases from 0.44 to 0.6 while the absolute value of bias_% decreases from 64% to 5%. The verification for this station represented the most stringent of our tests because of the significant gap in elevation and distance between the nearest COOP station and the reference one. Important results can also be noted for the similar ID-1763x14 station recording temperatures, which improved all the statistics. In general, minimum temperature statistics are slightly better than for maximum temperature statistics. Third, in areas of limited relief (all of the references stations in our study, except those located in Idaho), the overall statistics calculated on the daily val-

TABLE 4. Names, locations, and other information for the COOP stations (closest to the reference stations) used in the point-based comparison; *P* = precipitation, and *T* = temperature.

COOP station/data type	Lat	Lon	Elev (m)	Reference station name	Calculated distance (km)
Marlin 3 NE/ <i>P, T</i>	31°19'59"N	96°51'00"W	118	TX-RG69	16.4
Temple/ <i>P, T</i>	31°04'59"N	97°19'01"W	193	TX-GRL	4.2
Reynolds/ <i>P, T</i>	43°12'00"N	116°45'00"W	1198	ID-076x59	0.3
Silver City 5 W/ <i>P</i> (1 Jan 1978–30 Sep 1996)	43°01'08"N	116°49'23"W	1878	ID-163x20	6.5
Reynolds/ <i>P</i> (6 Jan 1962–31 Dec 1977)	43°12'00"N	116°45'00"W	1198	ID-163x20	15.0
Silver City 5 W/ <i>T</i> (1 Jan 1978–30 Sep 1996)	43°01'08"N	116°49'23"W	1878	ID-176x14	7.3
Reynolds/ <i>T</i> (6 Jan 1962–31 Dec 1977)	43°12'00"N	116°45'00"W	1198	ID-176x14	15.2
Chickasha Expt Station/ <i>P, T</i>	35°03'00"N	97°55'01"W	331	OK-121	10.3
Apache/ <i>P</i>	34°54'00"N	98°21'00"W	390	OK-131	12.0
Anadarko 3 E/ <i>T</i>	35°03'43"N	98°11'56"W	356	OK-131	12.7
Ashburn 3 Ene/ <i>P</i>	31°43'1"N	83°37'01"W	133	GA-48	12.6
Tifton Expt Station/ <i>P</i>	31°28'59"N	83°31'58"W	116	GA-08	4.2
Tifton Expt Station/ <i>T</i>	31°28'59"N	83°31'58"W	116	GA-UGA	1.4

TABLE 5. Statistics (computed from daily values) for the predicted daily total precipitation values and the closest COOP station records vs the observed records at each reference station.

Station statistic	TX-RG69		TX-GRL		ID-076x59		ID-163x20		OK-121		OK-131		GA-48		GA-08	
	Predicted	COOP	Predicted	COOP												
Bias_ %	2.37	4.68	-0.09	1.37	-3.33	-2.04	-5.3	-64.38	11.28	15.83	-4.98	4.59	-3.20	-1.03	-7.97	-0.34
Bias (mm)	0.06	0.12	0.0	0.03	-0.03	-0.02	-0.16	-1.97	0.23	0.32	-0.12	0.11	-0.11	-0.03	-0.27	-0.01
MAE_ %	99.09	105.66	99.44	105.55	103.84	104.00	87.21	87.65	116.33	125.94	114.08	129.26	107.56	122.45	117.54	125.9
MAE (mm)	2.49	2.66	2.42	2.57	0.79	0.80	2.67	2.68	2.34	2.53	2.64	3.00	3.57	4.06	3.91	4.19
RMSE (mm)	7.65	8.85	7.87	8.60	2.40	2.43	6.63	6.96	7.03	8.09	8.33	9.60	9.0	10.92	10.3	11.25
R	0.57	0.53	0.57	0.54	0.5	0.51	0.6	0.44	0.52	0.48	0.48	0.45	0.5	0.39	0.39	0.37

TABLE 6. Statistics (computed from daily values) for the predicted daily temperature (maximum and minimum) values and for the closest COOP station vs the observed records at each reference station.

Station statistic	TX-RG69		TX-GRL		ID-076x59		ID-1763x14		OK-121		OK-131		GA-UGA		
	Predicted	COOP	Predicted	COOP	Predicted	COOP	Predicted	COOP	Predicted	COOP	Predicted	COOP	Predicted	COOP	
Bias (°C)	Tmax	0.543	0.82	0.07	-0.19	0.04	0.55	4.14	5.52	0.49	1.14	0.24	-0.02	0.0	-0.017
	Tmin	-0.57	-0.10	-0.96	-0.68	-1.1	-0.72	0.05	0.59	-0.21	-0.01	-0.78	-1.38	-0.66	-0.66
MAE (°C)	Tmax	1.71	1.89	2.28	2.43	2.73	2.72	4.51	5.72	1.91	1.99	2.67	2.97	2.15	2.16
	Tmin	1.45	1.63	1.77	1.81	1.82	1.60	2.95	3.08	1.35	1.42	1.99	2.48	1.33	1.27
RMSE (°C)	Tmax	2.57	2.87	3.46	3.65	3.70	3.75	5.35	6.90	3.0	3.23	3.79	4.16	3.19	3.22
	Tmin	2.14	2.43	2.63	2.76	2.68	2.56	3.77	4.1	2.33	2.47	3.17	3.78	2.26	2.31
R	Tmax	0.96	0.95	0.92	0.91	0.94	0.94	0.94	0.92	0.96	0.96	0.93	0.92	0.90	0.90
	Tmin	0.96	0.97	0.96	0.95	0.95	0.89	0.88	0.88	0.97	0.97	0.95	0.94	0.96	0.96

TABLE 7. Annual total precipitation statistics for the predicted (second through seventh rows) and for the closest COOP station (last six rows) vs the observed records at each reference station (first row).

Station statistic	TX-RG69	TX-GRL	ID-076x59	ID-163x20	OK-121	OK-131	GA-48	GA-08
Avg reference (mm)	918.5	889.0	277.2	1109.7	650.3	752.5	1147.3	1185.4
Avg predicted (mm)	940.3	888.2	268.0	1050.8	723.7	715.0	1110.5	1091.0
Avg diff (mm)	21.8	-0.8	-9.2	-58.9	73.4	-37.4	-36.8	-94.4
Avg ratio	1.035	1.009	0.965	0.955	1.122	0.956	0.976	0.925
Std dev ratio	0.109	0.102	0.090	0.110	0.089	0.132	0.090	0.079
Bias_ %	2.37	-0.09	-3.33	-5.30	11.28	-4.98	-3.20	-7.97
MAE_ %	8.51	8.33	7.47	9.89	11.28	9.43	7.54	9.21
Avg nearest COOP station (mm)	961.5	901.2	271.5	395.2	753.2	787.0	1135.5	1181.4
Avg diff (mm)	43.0	12.2	-5.6	-714.4	102.9	34.5	-11.8	-4.0
Avg ratio	1.058	1.021	0.977	0.355	1.160	1.061	0.998	1.001
Std dev ratio	0.131	0.105	0.101	0.111	0.074	0.190	0.155	0.147
Bias_ %	4.68	1.37	-2.04	-64.38	15.83	4.59	-1.03	-0.34
MAE_ %	10.25	8.00	7.49	64.38	15.83	9.06	10.18	11.19

ues show parallel and close results between the predicted time series and the nearest COOP station ones. In particular, MAE_ % values are generally improved over the use of the nearest COOP station values but are still consistently higher than the respective bias_ % values. This is due to the local spatial setting of the closest COOP station and the reference station (the distance is reported Table 4), and also of the surrounding stations. The occurrence of the precipitation (rain/no-rain phase) at the involved locations might give different results. The relatively small bias_ % values show that the predicted and observed values are similar in intensity, but the relative high values of MAE_ % confirm gaps of timing occurrence. Given the sparse density of the COOP stations, primarily the closest station rain/no-rain phase and secondarily the phase of the surrounding local stations determine the MAE_ % calcu-

lated using daily values. This issue is not addressed by traditional interpolation techniques either.

Yearly and monthly accumulated precipitation and average temperature at each point were calculated by summing (for precipitation) or averaging (for temperature) the individual daily totals/averages. Tables 7 and 8, for precipitation and average temperature, report the average and other statistics. The prediction of annual precipitation shows the improvements (predicted vs closest COOP station) noted on the daily analysis, particularly for the most critical station (ID-163-x20). Exceptions are noted for the stations located in Georgia, which show a slight bias_ % increment. Improvements are generally noted for temperatures as well, with the exceptions of the slight bias of the minimum temperature for the stations located in Texas.

In general, monthly statistics follow the pattern of the

TABLE 8. Annual average daily temperature (maximum and minimum) statistics for the predicted (second through seventh rows) and for the closest COOP station (last six rows) vs the observed records at each reference station (first row).

Station statistic	TX-RG69		TX-GRL		ID-076x59		ID-176x14		OK-121		OK-131		GA-UGA	
	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min
Reference (°C)	25.0	13.2	25.4	13.5	15.6	2.1	8.9	0.8	22.8	9.8	22.2	10.0	25.0	13.6
Predicted (°C)	25.5	12.7	25.4	12.5	15.6	0.9	13.1	0.8	23.3	9.6	22.5	9.2	25.0	13.0
Diff (°C)	0.5	-0.5	0.0	-1.0	0.0	-1.2	4.2	0.1	0.5	-0.2	0.2	-0.8	0.0	-0.6
Avg ratio	1.022	0.958	1.004	0.929	1.003	0.365	1.482	0.573	1.021	0.98	1.010	0.922	1.0	0.951
Std dev ratio	0.025	0.029	0.026	0.031	0.016	0.405	0.139	1.956	0.017	0.020	0.012	0.019	0.007	0.029
Bias (°C)	0.54	-0.56	0.08	-0.96	0.04	-1.12	4.15	0.06	0.49	-0.20	0.23	-0.79	0.0	-0.67
MAE (°C)	0.70	0.59	0.49	0.98	0.22	1.12	4.15	0.60	0.57	0.23	0.23	0.79	0.13	0.67
Nearest COOP station (°C)	25.8	13.1	25.2	12.8	16.2	1.3	14.4	1.4	24.0	9.8	22.2	8.6	25.0	13.0
Diff (°C)	0.8	-0.1	-0.2	-0.7	0.6	-0.7	5.5	0.6	1.2	0.0	0.0	-1.4	0.0	-0.6
Avg ratio	1.033	0.992	0.993	0.951	1.036	0.623	1.651	0.998	1.050	0.999	0.999	0.863	1.0	0.951
Std dev ratio	0.030	0.036	0.031	0.052	0.018	0.314	0.292	2.925	0.025	0.025	0.026	0.043	0.010	0.033
Bias (°C)	0.81	-0.10	-0.19	-0.68	0.56	-0.73	5.52	0.60	1.13	-0.01	-0.03	-1.39	0.00	-0.65
MAE (°C)	0.91	0.36	0.61	0.75	0.56	0.74	5.52	0.77	1.13	0.19	0.46	1.39	0.20	0.65

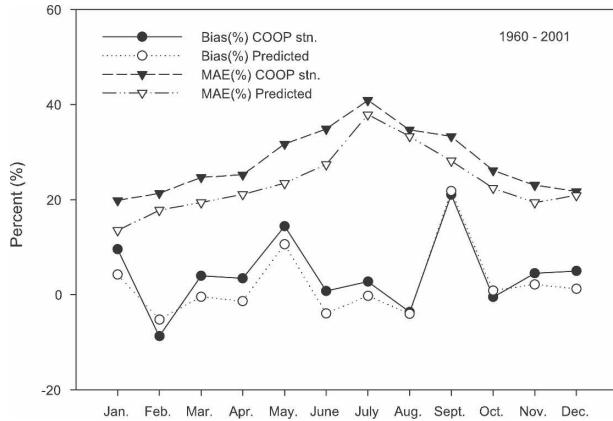


FIG. 9. Bias_% and MAE_% calculated on the predicted/closest COOP station monthly accumulated precipitation values for the TX-RG69 station.

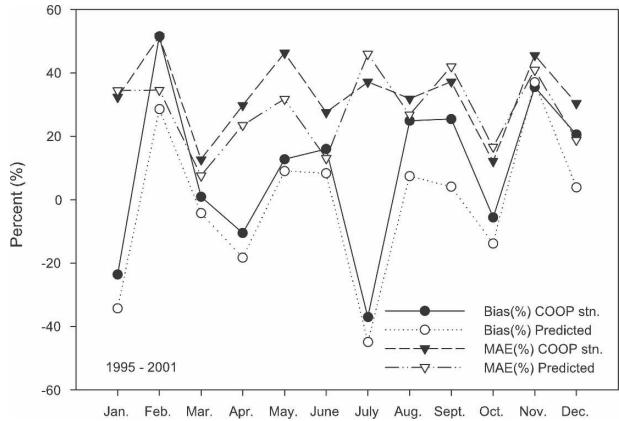


FIG. 11. As in Fig. 9, but for the OK-131 station.

daily and yearly ones. In particular, selected plots of monthly precipitation bias_% and MAE_% are shown in Figs. 9–12. The first of these figures, for the TX-RG69 station, shows the statistic parameters moving toward zero for each month. This pattern is magnified in Fig. 10, which refers to the significant reference station (ID-163x20) highlighted above. A clear improvement is not evident in Fig. 11, which refers to one of the reference stations located in Oklahoma (OK-131). However, statistics for this station were calculated using a time series with a limited number of monthly members (7) contained in the period 1995–2001. Figure 12, for a reference station located in Georgia (GA-08), highlights the expected improvement of the monthly MAE_% while the monthly bias_% appears to slightly and repeatedly worsen. For this and the other reference station located in Georgia (GA-48), this effect can be noted also on the bias_% calculated on the daily and yearly bases. These results can be expected in the

southeastern region of the United States, where rainfall patterns show great yearly and seasonal variability, with midsummer convective storms with high intensity and low spatial correlation. For this region, Bosch et al. (1999) highlighted low correlation of rain gauge records at distances over 2 km (e.g., the closest COOP station is 4.2 km away). An additional evaluation was performed on the accumulated monthly precipitation, arranging the various time series in deciles. Significant and representative results are reported in Figs. 13–15 for decile ranges 1–2, 3–6, and 7–10. Figure 13, for station TX-RG69, and Fig. 15, for station GA-08, show that the predicted time series are in good quantitative and qualitative agreement with the observed one. In particular, deciles 5–6 (representing precipitation amount not exceeded by the 50%–60% of occurrences/near normal) appear very well represented.

5. Discussion and conclusions

In this paper we have presented a method for constructing daily fields of precipitation and temperature

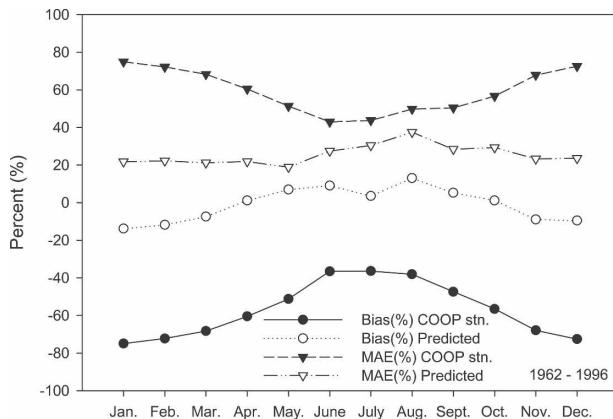


FIG. 10. As in Fig. 9, but for the ID-163x20 station.

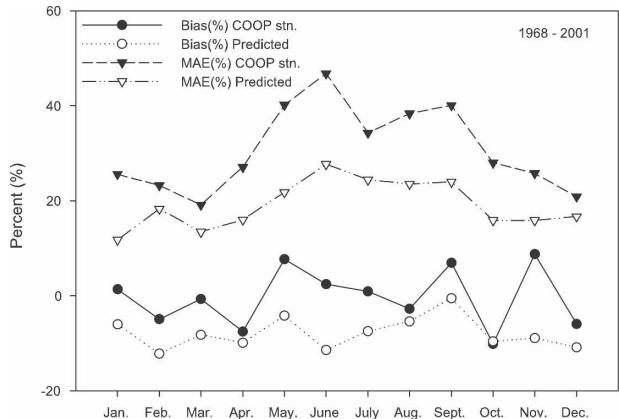


FIG. 12. As in Fig. 9, but for the GA-08 station.

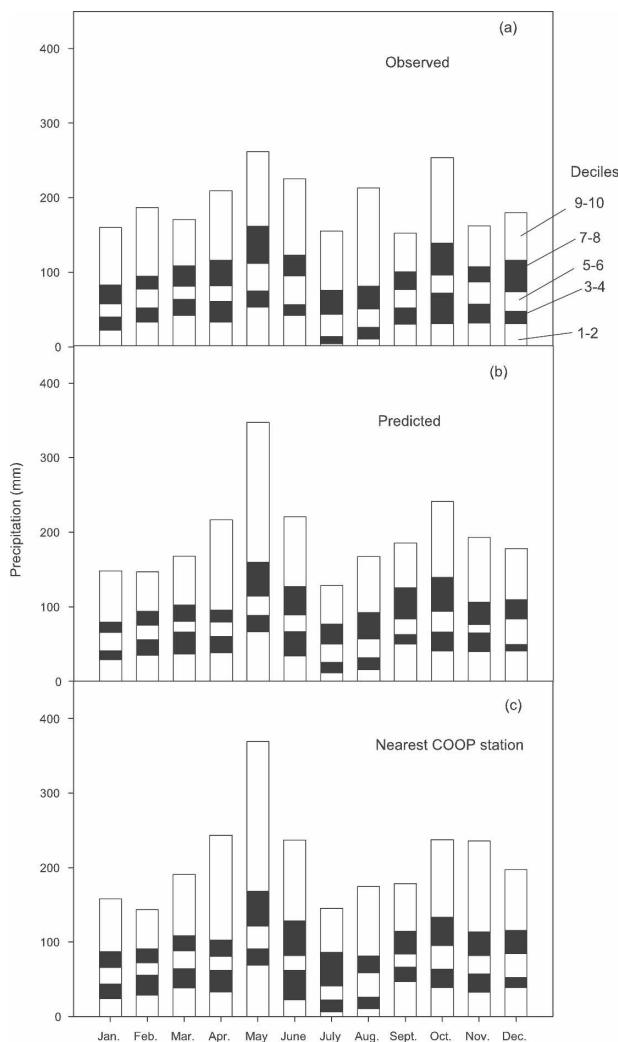


FIG. 13. Monthly accumulated precipitation deciles (a) observed at the TX-RG69 station, (b) predicted, and (c) observed at the nearest COOP station.

(maximum and minimum) by combining weather meteorological station records and monthly field estimates. The strength of the developed methodology is in accounting for climate patterns and for the variation of climate with elevation. In fact, key components of this nonconventional method are the identification of topographically adjusted meteorological patterns (monthly fields) and the application of a spatialization procedure applied to daily ratios versus whole recorded values. The method was implemented over the CONUS, using quality-controlled datasets, to overcome various needs, such as historical and regional coverage, topographic adjustment, and limitations such as station location and density, and variability for the daily time step. Previous similar attempts produced only datasets at coarser resolution and/or for a longer time step (monthly/yearly).

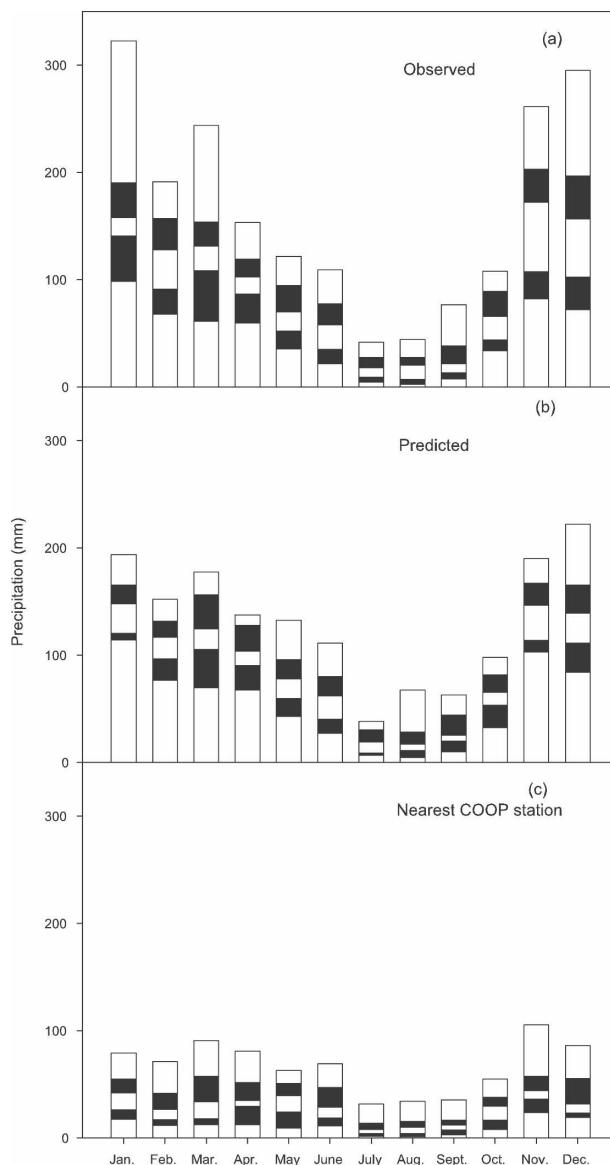


FIG. 14. As in Fig. 13, but for the ID-163x20 station.

The explicit purpose of this study was to estimate the daily regional response surfaces of precipitation and temperature (maximum and minimum) versus the interpolation to points. The lack of comparative “true” reference fields is a persistent drawback for this kind of assessment. Our strategy of assessment focused on verifying the monthly distribution of the variables derived from the daily surfaces with the original monthly surfaces, on observing the preservation of the source point records, and on evaluating the statistics for the available number of independent stations.

We have succeeded in our main objectives, producing precipitation and temperature fields seamlessly covering the entire CONUS for each daily time step over a

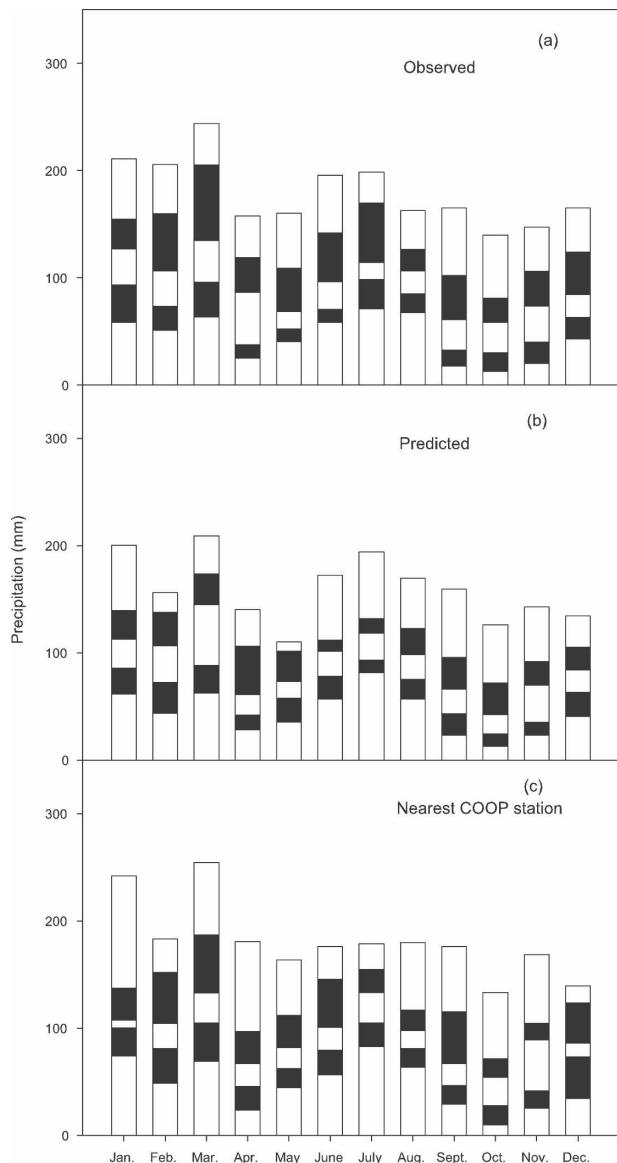


FIG. 15. As in Fig. 13, but for the GA-08 station.

temporal sequence of 15 341 days, from 1 January 1960 to 31 December 2001. This period covered an extended historical period needed to adequately describe meteorological fields (around 4-km resolution) to be used in our planned large-area hydrology modeling application. Our areal verification has shown that the new datasets are consistent with the governing data source monthly PRISM estimates. The method has shown to explicitly account for the critical topographic influence while maintaining a high degree of spatial correlation and integrity with the longer time step data. These combined achievements are particularly important in applied climatology as an application of spatial continuous input data to hydrologic simulation models at re-

gional scale, from which the assessments often look into broad and consistent evaluations of the water budget at the monthly time step. Although these results could have been expected, since monthly PRISM estimates are data source of the applied method, this could not have been guaranteed. In fact, PRISM was developed using monthly observation records, which are not necessarily consistent with the daily ones. In doing our verification, we actually highlighted a number of areas and associated stations that will need further records evaluation. This evaluation has the potential for future improvements of the PRISM grids and the station records both and also of our daily estimates. In fact, in some instances for precipitation and for the most arid zones, the local divergence of the monthly PRISM and the station records has shown the potential for affecting the final records estimated at the original COOP station locations.

The analysis of the obtained results shows that the original daily values from the COOP stations are minimally altered by the applied method. Meanwhile, the results for a limited number of reference points scattered in the CONUS show that daily precipitation and temperature estimates relative to the nearest-neighbor method estimates are distinctively improved over an extended range of elevation and distance (e.g., station ID-163-x20 for precipitation and ID-163-x14 for temperature, which are located at a topographic setting very different from the nearest COOP recording station). The achieved improvements are either less evident or negligible for the other stations of our referencing set because of the quality and distance of the nearest station data, along with the local climatic patterns (e.g., southeastern region). More noticeable are the results at the longer time step (monthly/yearly) and the deciles distribution, which show a general improved agreement.

In addition, we have achieved two main goals and addressed associated needs. The first one is the demonstration that the method is suitable for being implemented in a GIS environment, thereby providing a way to temporally disaggregate the monthly PRISM maps to the daily time step. In fact, it was desirable to gain an alternative and simple method for the generalization of meteorological variables over large distances and elevations. We have demonstrated that this method, combining data and a simple IDW interpolator, provides reasonable results in critical situations, such as those with orographic influence and low spatial density of stations. The documented technique, components, and data sources can be systematically applied with a GIS using PRISM estimates, which are readily available on the Internet, and ground observation records. The second

one is the initialization of daily precipitation and temperature fields, which will be used for future comparisons/improvements of the approach, given the absence of “true” reference fields.

Based on these observations, we are planning a further verification of the data for an extended set of reference stations and a cross-validation error analysis to provide an uncertainty estimate. Future research plans include the areal comparisons in highly monitored watersheds, comparison with other similar datasets, investigation of the limitations of the IDW within the described methodology, inclusion of more sophisticated and/or adaptive (i.e., based upon station density and conditioned parameterization to address the underlined issue of the rain/no rain phase for the precipitation) interpolation methods, as well as taking into consideration the heterogeneous and biased distribution of the recording stations and quality of the data sources.

The datasets are serially and spatially complete (three sets of 15 341 grids with no gaps) for the period 1960–2001 and were created to fulfill the requirements of the hydrologic-based modeling framework being implemented in the CEAP project at the national scale, which plans to evaluate and track long-term effects on water quality from the implementation of conservation practices (CEAP currently implements the data summarized in the units composing the USGS HUC watersheds). The gridded data cell size (2.5 min, ~4 km) has the potential to provide adequate input to a wide spatial range of more detailed features replacing the HUCs in the future.

Acknowledgments. Funding for this study was provided by the USDA Natural Resources Conservation Service through a cooperative agreement with the Texas Agricultural Experiment Station in Temple, Texas, and in partnership with the USDA Agricultural Research Service. The study is part of the Conservation Effects Assessment Project.

REFERENCES

- Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams, 1998: Large area hydrologic modeling and assessment. Part I: Model development. *J. Amer. Water Resour. Assoc.*, **34**, 73–89.
- Bosch, D. D., J. M. Sheridan, and F. M. Davis, 1999: Rainfall characteristics and spatial correlation for the Georgia Coastal Plain. *Trans. ASAE*, **42**, 1637–1644.
- Courault, D., and P. Monestiez, 1999: Spatial interpolation of air temperature according to atmospheric circulation patterns in southeast France. *Int. J. Climatol.*, **19**, 365–378.
- Daly, C., cited 2002: Variable influence of terrain on precipitation patterns: Delineation and use of effective terrain height in PRISM. [Available online at <http://www.ocs.orst.edu/pub/prism/docs/effectiveterrain-daly.pdf>.]
- , 2006: Guidelines for assessing the suitability of spatial climate data sets. *Int. J. Climatol.*, **26**, 707–721.
- , and G. L. Johnson, 1999: PRISM spatial climate layers: Their development and use. *Short Course on Topics in Applied Climatology, Course Material, 79th Annual Meeting of the Amer. Meteor. Soc.*, Dallas, TX, Amer. Meteor. Soc., 49 pp. [Available online at <http://prism.oregonstate.edu/pub/prism/docs/prisguid.pdf>.]
- , and D. B. Hannaway, 2005: *Visualizing China's Future Agriculture Climate, Soils, and Suitability Maps for Improved Decision-Making*. Oregon State University, 296 pp.
- , R. P. Neilson, and D. L. Phillips, 1994: A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteor.*, **33**, 140–158.
- , G. H. Taylor, and W. P. Gibson, 1997: The PRISM approach to mapping precipitation and temperature. Preprints, *10th Conf. on Applied Climatology*, Reno, NV, Amer. Meteor. Soc., 10–12.
- , —, —, T. W. Parzybok, G. L. Johnson, and P. A. Pasteris, 2001: High-quality spatial climate data sets for the United States and beyond. *Trans. ASAE*, **43**, 1957–1962.
- , W. P. Gibson, G. H. Taylor, G. L. Johnson, and P. A. Pasteris, 2002: A knowledge-based approach to the statistical mapping of climate. *Climate Res.*, **22**, 99–113.
- , E. H. Helmer, and M. Quinones, 2003: Mapping the climate of Puerto Rico, Vieques, and Culebra. *Int. J. Climatol.*, **23**, 1359–1381.
- , W. P. Gibson, M. Doggett, J. Smith, and G. Taylor, 2004: Up-to-date monthly climate maps for the conterminous United States. Preprints, *14th Conf. on Applied Climatology*, Seattle, WA, Amer. Meteor. Soc., CD-ROM, P5.1.
- Eischeid, J. K., C. B. Baker, T. R. Karl, and H. F. Diaz, 1995: The quality control of long-term climatological data using objective data analysis. *J. Appl. Meteor.*, **34**, 2787–2795.
- , P. A. Pasteris, H. F. Diaz, M. S. Plantico, and N. J. Lott, 2000: Creating a serially complete, national daily time series of temperature and precipitation for the western United States. *J. Appl. Meteor.*, **39**, 1580–1591.
- Fulton, R. A., J. P. Breidenbach, D.-J. Seo, D. A. Miller, and T. O'Bannon, 1998: The WSR-88D rainfall algorithm. *Wea. Forecasting*, **13**, 377–395.
- Garen, D. C., G. L. Johnson, and C. L. Hanson, 1994: Mean areal precipitation for daily hydrologic modeling in mountainous regions. *Water Resour. Bull.*, **30**, 481–491.
- Gibson, W. P., C. Daly, T. Kittel, D. Nychka, C. Johns, N. Rosenbloom, A. McNab, and G. Taylor, 2002: Development of a 103-year high-resolution climate data set for the conterminous United States. *Proc. 13th Conf. on Applied Climatology*, Portland, OR, Amer. Meteor. Soc., 181–183.
- Groisman, P. Ya., and D. R. Legates, 1994: The accuracy of United States precipitation data. *Bull. Amer. Meteor. Soc.*, **75**, 215–227.
- Hamlet, A. F., and D. P. Lettenmaier, 2005: Production of temporally consistent gridded precipitation and temperature fields for the continental United States. *J. Hydrometeorol.*, **6**, 330–336.
- Hannaway, D. B., and Coauthors, 2005: Forage species suitability mapping for China using topographic, climatic and soils spatial data and quantitative plant tolerances. *Agric. Sci. China*, **4**, 660–667.
- Hewitson, B. C., and R. G. Crane, 2005: Gridded area-averaged daily precipitation via conditional interpolation. *J. Climate*, **18**, 41–57.

- Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis, 2005: Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.*, **25**, 1965–1978.
- Hubbard, K. G., 1994: Spatial variability of daily weather variables in the high plains of the USA. *Agric. For. Meteorol.*, **68**, 29–41.
- Johns, C. J., D. Nychka, T. G. F. Kittel, and C. Daly, 2003: Infilling sparse records of spatial fields. *J. Amer. Stat. Assoc.*, **98**, 796–806.
- Johnson, G. L., C. Daly, G. H. Taylor, and C. L. Hanson, 2000: Spatial variability and interpolation of stochastic weather simulation model parameters. *J. Appl. Meteorol.*, **39**, 778–796.
- Kittel, T. G. F., and Coauthors, 2004: VEMAP Phase 2 bioclimatic database. I. Gridded historical (20th century) climate for modeling ecosystem dynamics across the conterminous USA. *Climate Res.*, **27**, 151–170.
- Krajewski, W. F., G. J. Ciah, J. R. McCollum, and C. Bacotiu, 2000: Initial validation of the Global Precipitation Climatology Project monthly rainfall over the United States. *J. Appl. Meteorol.*, **39**, 1071–1086.
- Kurtzman, D., and R. Kadmon, 1999: Mapping of temperature variables in Israel: A comparison of different interpolation methods. *Climate Res.*, **13**, 33–43.
- Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen, 2002: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *J. Climate*, **15**, 3237–3251.
- Mausbach, M. J., and A. R. Dedrick, 2004: The length we go: Measuring environmental benefits of conservation practices. *J. Soil Water Conserv.*, **59**, 96–103.
- Meneghini, R., J. A. Jones, T. Iguchi, K. Okamoto, and J. Kwiatkowski, 2001: Statistical methods of estimating average rainfall over large space–timescales using data from the TRMM precipitation radar. *J. Appl. Meteorol.*, **40**, 568–585.
- New, M., M. Hulme, and P. Jones, 1999: Representing twentieth-century space–time climate variability. Part I: Development of a 1961–90 mean monthly terrestrial climatology. *J. Climate*, **12**, 829–856.
- NOAA-NCDC, 2002: The climate atlas of the United States, version 2.0. National Oceanic and Atmospheric Administration, 20023 pp., CD-ROM. [Available online at <http://www.ncdc.noaa.gov/oa/about/cdrom/climatls2/info/atlasad.html>.]
- Pardo-Iguzquiza, E., 1998: Comparison of geostatistical methods for estimating the areal average climatological rainfall mean using data on precipitation and topography. *Int. J. Climatol.*, **18**, 1031–1047.
- Piper, S. C., and E. F. Stewart, 1996: A gridded global data set of daily temperature and precipitation for terrestrial biosphere modeling. *Global Biogeochem. Cycles*, **10**, 757–782.
- Plantico, M. S., L. A. Goss, C. Daly, and G. Taylor, 2000: A new U.S. climate atlas. *Proc. 12th Conf. on Applied Climatology*, Asheville, NC, Amer. Meteor. Soc., 247–248.
- Prudhomme, C., and D. W. Reed, 1999: Mapping extreme rainfall in a mountainous region using geostatistical techniques: A case study in Scotland. *Int. J. Climatol.*, **19**, 1337–1356.
- Seaber, P., F. P. Kapinos, and G. L. Knapp, 1987: Hydrologic unit maps. U.S. Department of Interior U.S. Geological Survey, Water Supply Paper 2294, 63 pp.
- Seo, D.-J., W. F. Krajewski, and D. S. Bowles, 1990: Stochastic interpolation of rainfall data from raingages and radar using cokriging: 1. Design of experiments. *Water Resour. Res.*, **26**, 469–477.
- Simpson, J. J., G. L. Hufford, C. Daly, J. S. Berg, and M. D. Fleming, 2005: Comparing maps of mean monthly surface temperature and precipitation for Alaska and adjacent areas of Canada produced by two different methods. *Arctic*, **58**, 137–161.
- Slaughter, C. W., and C. W. Richardson, 2000: Long-term watershed research in USDA—Agricultural Research Service. *Water Resour. IMPACT*, **2**, 28–31.
- Steeves, P., and D. Nebert, 1994: 1:250,000-scale hydrologic units of the United States. U.S. Geological Service Open-File Rep. 94-0236. [Available online at <http://water.usgs.gov/GIS/metadata/usgswrd/XML/huc250k.xml>.]
- Steiner, M., J. A. Smith, S. Burges, C. V. Alonso, and R. W. Darden, 1999: Effect of bias adjustment and rain gauge data quality control on radar rainfall estimates. *Water Resour. Res.*, **35**, 2487–2503.
- Tabios, G. Q., and J. D. Salas, 1985: A comparative analysis of techniques for spatial interpolation of precipitation. *Water Resour. Bull.*, **21**, 365–380.
- Thiessen, A. H., 1911: Precipitation averages for large areas. *Mon. Wea. Rev.*, **39**, 1082–1084.
- Thornton, P. E., S. W. Running, and M. A. White, 1997: Generating surfaces of daily meteorological variables over large regions of complex terrain. *J. Hydrol.*, **190**, 214–251.
- USDA-NRCS, 1998: PRISM climate mapping project—Precipitation. Mean Monthly and Annual Precipitation Digital Files for the Continental U.S., USDA-NRCS, CD-ROM. [Available from the USDA-NRCS National Cartography and Geospatial Center, 501 West Felix St., Bldg. 23, Fort Worth TX 76115.]
- Watson, D. F., and G. M. Philip, 1985: A refinement of inverse distance weighted interpolation. *Geo-Processing*, **2**, 315–327.
- Widmann, M., and C. S. Bretherton, 2000: Validation of meso-scale precipitation in the NCEP reanalysis using a new grid-cell dataset for the northwestern United States. *J. Climate*, **13**, 1936–1950.
- Willmott, C. J., and S. M. Robeson, 1995: Climatologically aided interpolation (CAI) of terrestrial air temperature. *Int. J. Climatol.*, **15**, 221–229.
- Xia, Y., P. Fabian, M. Winterhalter, and M. Zhao, 2001: Forest climatology: Estimation and use of daily climatological data for Bavaria, Germany. *Agric. For. Meteorol.*, **106**, 87–103.
- Zhu, H., T. Luo, and C. Daly, 2003: Validation of simulated grid data sets of China's temperature and precipitation with high spatial resolution (in Chinese with English abstract). *Geogr. Res.*, **22**, 349–359.