



A large-scale, high-resolution hydrological model parameter data set for climate change impact assessment for the conterminous US

A. A. Oubeidillah¹, S.-C. Kao¹, M. Ashfaq¹, B. S. Naz¹, and G. Tootle²

¹Oak Ridge National Laboratory, Oak Ridge, TN, USA

²University of Alabama, Tuscaloosa, AL, USA

Correspondence to: S.-C. Kao (kaos@ornl.gov)

Received: 2 July 2013 – Published in Hydrol. Earth Syst. Sci. Discuss.: 22 July 2013

Revised: 28 October 2013 – Accepted: 28 November 2013 – Published: 7 January 2014

Abstract. To extend geographical coverage, refine spatial resolution, and improve modeling efficiency, a computation- and data-intensive effort was conducted to organize a comprehensive hydrologic data set with post-calibrated model parameters for hydro-climate impact assessment. Several key inputs for hydrologic simulation – including meteorologic forcings, soil, land class, vegetation, and elevation – were collected from multiple best-available data sources and organized for 2107 hydrologic subbasins (8-digit hydrologic units, HUC8s) in the conterminous US at refined 1/24° (~4 km) spatial resolution. Using high-performance computing for intensive model calibration, a high-resolution parameter data set was prepared for the macro-scale variable infiltration capacity (VIC) hydrologic model. The VIC simulation was driven by Daymet daily meteorological forcing and was calibrated against US Geological Survey (USGS) WaterWatch monthly runoff observations for each HUC8. The results showed that this new parameter data set may help reasonably simulate runoff at most US HUC8 subbasins. Based on this exhaustive calibration effort, it is now possible to accurately estimate the resources required for further model improvement across the entire conterminous US. We anticipate that through this hydrologic parameter data set, the repeated effort of fundamental data processing can be lessened, so that research efforts can emphasize the more challenging task of assessing climate change impacts. The pre-organized model parameter data set will be provided to interested parties to support further hydro-climate impact assessment.

1 Introduction

With the advance of high-performance computing and more abundant historical observation data, hydrologists and water resource engineers are better equipped to improve the scale, resolution, and accuracy of hydrologic simulation. Depending on the need, a suitable hydrologic model may range from a statistical model (e.g. artificial neural network, ANN) or a conceptual model (e.g. bucket model) to a more sophisticated model that can simulate a series of hydrological processes based on physical mechanisms. However, although a statistical model can generally simulate hydrologic variables well with fewer predictors, the assumption of stationarity may be questionable in a changing environment in which many hydrologic processes are expected to be disrupted (Milly et al., 2008). Under such conditions, historical relationships may not provide fully accurate information about future streamflow and water availability. One example is ANN (and the various related machine learning algorithms). Although these types of advanced statistical methods are extremely powerful in forecasting reservoir outflows with minimal observation, the physical relationships among various predictors cannot be interpreted (Govindaraju and Rao, 2000); this hinders their direct extension across different locations and climate patterns. Therefore, these methods may not be suitable choices for climate-related research.

Unlike statistical models, process-based models are theoretically justifiable for climate research. Since most hydrologic mechanisms are simulated through deterministic laws, the assumption of stationarity is a lesser issue. However, a high number of observations and parameters is required to drive process-based models. For instance, a distributed

rainfall–runoff model may require fine-resolution inputs of vegetation, precipitation, temperature, solar radiation, topography, and many other soil properties to simulate various hydrologic processes such as evapotranspiration, infiltration, vegetation root absorption, snowmelt, and runoff generation. With increasing complexity in model resolution, scale, and processes, the required computational resources also increase exponentially. As a result, it is generally more challenging to conduct a process-based hydrologic simulation for a large study area with fine spatial resolution. A tradeoff usually must be made between scale and resolution because of resource limitations.

Numerous studies have investigated the hydrological impacts of climate change in the US using process-based models (Mote et al., 2005; Christensen et al., 2004; Payne et al., 2004; Maurer et al., 2002; McCabe and Hay, 1995; Hamlet and Lettenmaier, 1999; Wolock and McCabe, 1999; Ashfaq et al., 2010). Output from global climate models (GCMs) is usually downscaled, bias corrected, and used in conjunction with hydrologic models to assess future water availability. However, owing to resource limitations, many hydro-climate impact assessments either are focused on smaller US regions or provide lower spatial resolution (Christensen et al., 2004; Payne et al., 2004; Maurer et al., 2002; McCabe and Hay, 1995). Repeated efforts may be needed for fundamental data processing and model calibration, and these may unavoidably reduce the amount of computing time available for the more challenging task of assessing climate change impacts. To extend geographical coverage, refine spatial resolution, and make hydro-climate impact assessment more efficient, a comprehensive set of calibrated physical parameters is desired that can provide the most up-to-date, high-resolution watershed soil, vegetation, elevation, and other hydrologic characteristics. If a fine-resolution hydrological model parameter data set could be pre-organized, generally calibrated, and constantly updated, it would enable numerous researchers to easily extend hydro-climate impact assessment efforts to different watersheds.

One major structural difference between climate and hydrologic models is their respective requirements for vertical and horizontal resolution. Although spatial resolution is relatively less critical for current GCMs, they do need more vertical layers to better simulate the boundary layer and tropospheric processes that govern complex land–atmosphere–ocean interactions at varying timescales. However, horizontal resolution is the dominant factor for hydrologic models, since complicated topography and heterogeneous land surface characteristics have the greatest impacts on model performance. Previous studies have shown that runoff is sensitive to spatial variations in soil properties, precipitation inputs, and topography (Haddelenad et al., 2002; Nijssen et al., 2001; Sharif et al., 2007; Dooge and Bruen, 1997; Merz and Plate, 1997; Shah et al., 1996; Wolock and Price, 1994) and finer resolution data input would result in better model performance in comparing simulated and observed discharge

(Shrestha et al., 2006). Additionally, fine-resolution hydrological model parameter data sets are needed to explore tasks such as identifying climatic controls on the spatial variability of hydrologic parameters; performing statistical analysis between climatic variables and land surface processes and states; and developing subgrid parameterization approaches for hydrologic parameters, particularly in wet and dry conditions. Therefore, to enhance the performance of hydrologic simulations for climate change impact assessment, one would need to simultaneously improve both spatial resolution (for hydrologic research needs) and geographical coverage (for climate change research needs). Maurer et al. (2002) has developed a long-term hydrologically consistent data set for the conterminous US at $1/8^\circ$ (~ 12 km) spatial resolution for water and energy budget studies. The data set was recently extended by Livneh and Lettenmaier (2013) for the time period of 1915–2011 with a refined $1/16^\circ$ (~ 6 km) resolution. With the continuous improvement of spatial resolution through regional climate models (e.g. North American Regional Climate Change Assessment Program and Coordinated Regional Climate Downscaling Experiment), refined climate projections will soon become available. A corresponding enhancement of hydrologic models is hence required for detailed hydro-climate impact assessment.

Given the motivation, a data- and computation-intensive effort was performed in this study. A widely used hydrologic model, variable infiltration capacity (VIC), was chosen as the baseline hydrologic model in this initial effort. Using multiple state-of-the-art geospatial data sets for the conterminous US, various key hydrologic model inputs – including topography, soil characteristics, vegetation, land surface classification, meteorological forcing, and runoff observation – were organized in a nationally consistent $1/24^\circ$ (~ 4 km) grid and grouped by the US Geological Survey (USGS) 8-digit hydrologic unit code (HUC8 or subbasin; Seaber et al., 1987). Since precipitation and other hydrologic variables are highly resolution dependent (Gao et al., 2006), the refinement from the commonly used $1/8^\circ$ (~ 12 km) resolution (e.g. Maurer et al., 2002; Mitchell et al., 2004) will allow more flexibility in characterizing detailed hydrologic responses, particularly for extreme hydrologic events. In addition, to spatially increase the model accuracy across different geographical locations, each HUC8 subbasin was calibrated independently through a computationally intensive exercise. By using a multi-site (watershed) calibration approach, we hope that the model performance can be consistently improved across various subbasins. This approach of refining model performance uniformly across all HUC8 subbasins is new and different from previous studies (e.g. Maurer et al., 2002). A parallelization scheme was developed for each of the 2107 HUC8 subbasins to greatly reduce the required computation time. Using high-performance supercomputing, an exhaustive model calibration exercise was then performed to improve the model accuracy across the conterminous US. The calibrated model was evaluated using USGS observed

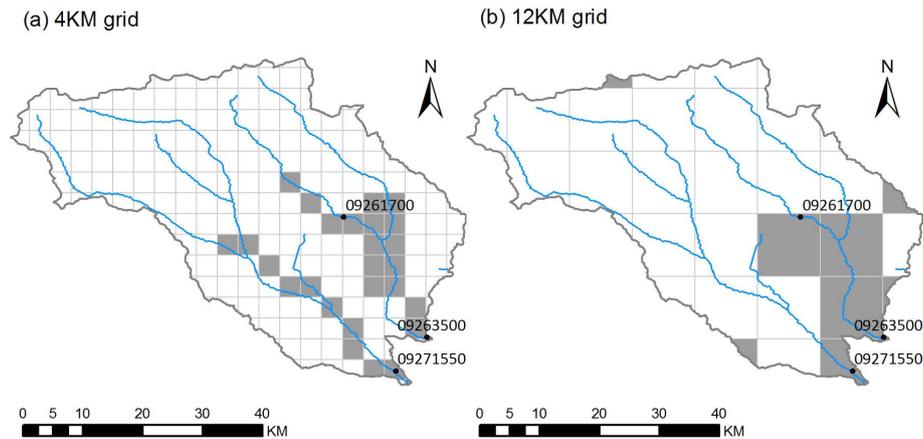


Fig. 1. Effect of spatial resolution in hydrologic modeling.

runoff, and the evaluation showed that the parameter data set results in satisfactory performance for most US HUC8 subbasins. In Sect. 2, the data source and methodology used to develop the parameter data sets are described. The current model performance is presented in Sect. 3, and a summary and conclusions are provided in Sect. 4.

2 Data and methodology

2.1 Study area and assessment grids

The effect of spatial resolution on hydrologic modeling is illustrated in Fig. 1 for the Ashley–Brush Subbasin in Utah (HUC8 ID 14060002). Based on the $1/3$ arcsec-resolution (~ 10 m) USGS national elevation dataset (NED; Gesch et al., 2002), the average elevation was computed for both $1/24^\circ$ (~ 4 km) and $1/8^\circ$ (~ 12 km) grids. Flow direction and flow accumulation were then derived using standard geographical information system (GIS) software. Grids with flow accumulation values of greater than 20 in Fig. 1a and greater than 2 in Fig. 1b are marked in grey, illustrating the locations of rivers seen by computer models. Three USGS gauge stations (09261700, 09263500, and 09271550) are also marked in Fig. 1. Whereas the $1/24^\circ$ grids can capture two streams, the $1/8^\circ$ grids fail to depict the system. Without additional information, the hydrological model might consider both outlets (09263500 and 09271550) as being located on the same river, which would directly affect the model calibration and validation. In addition, most of the $1/8^\circ$ grids flow to the neighboring subbasin instead of to the two outlets because of the insufficient spatial resolution for this watershed.

Therefore, to refine hydro-climate assessment from the regional to the watershed scale, spatial resolution is key. After evaluating the available resources, it was decided to select 4 km as the targeted resolution, which would require 9 times the computational resources needed for the commonly used 12 km grids, plus some additional effort for large data set

management and quality control. The 4 km grids selected herein in fact follow the same configuration as the Parameter-elevation Regressions on Independent Slopes Model meteorological data sets (PRISM; Daly et al., 2002), with the north boundary extending to 53° N to cover the entire Columbia River basin on the Canadian side. Since PRISM is recognized as the most accurate grid-based monthly meteorological observation of precipitation and temperature based on ground-based meteorological stations, it is convenient to use the same grid configuration for future model evaluation and comparison. Each 4 km grid was given a unique identifier and further labeled with USGS HUC8 IDs. While our focus in this paper is on the conterminous US, some headwater basins in Canada (e.g. British Columbia) and Mexico also contribute to the downstream hydrology and should eventually be considered. As a placeholder, the watershed boundary outside the conterminous US was obtained from the USGS National Hydrography Dataset Plus, version 1 (EPA/USGS, 2010). When the forcing and parameter data are available, they are also organized in a consistent format. Further simulation and calibration for the non-US region will be conducted in future studies. Overall, there are $\sim 480\,000$ grid points and 2107 HUC8s in 18 hydrologic regions (HUC2) in the conterminous US. HUC2 and HUC8 are illustrated in Fig. 2.

2.2 Meteorological forcing

“Meteorological forcing” refers to the required meteorological inputs for hydrologic modeling, including precipitation, temperature, wind speed, and others. Conventionally, these values are looked up from surface weather observation stations (e.g. National Weather Service Cooperative Observer Program) and then spatially interpolated for further hydrologic application. However, given the heavy data processing requirements, such an approach is not applicable for large-scale hydrologic simulation; therefore, pre-processed grid-based observations are primarily used.

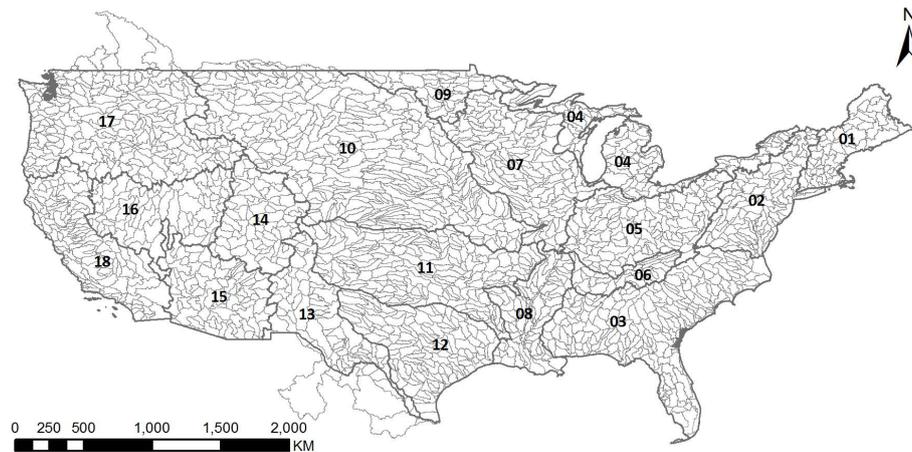


Fig. 2. Study area in this study. The smaller polygons represent the hydrologic subbasins (HUC8) and the larger polygons with thicker grey boundaries represent the hydrologic regions (HUC2). The 2-digit numbers represent the HUC2 ID.

Several meteorological forcing data sets are commonly used in hydrologic studies for the conterminous US. These data sets are either fully based on observation or partially assimilated through weather forecasting models. Four were selected for consideration in this study: PRISM, Maurer, Daymet, and North American Regional Reanalysis (NARR). PRISM provides accurate grid-based observations of precipitation and temperature by considering topographical effects and some other adjustment factors (Daly et al., 2002, 2008). It is available on a monthly timescale from 1895 to the present and provides $1/24^\circ$ (~ 4 km) spatial resolution (<http://www.prism.oregonstate.edu/>). Maurer et al. (2002) is a widely used forcing data set for hydrologic studies (Livneh and Lettenmaier, 2013). It is based on observation and is available on a daily timescale from 1950 to 2010 in $1/8^\circ$ (~ 12 km) spatial resolution. Targeted for fine-scale ecological studies, Daymet (Thornton et al., 1997) is another commonly used meteorological data set based on ground-based meteorological stations (White et al., 2006; Keane et al., 2008; Manter et al., 2005). Daymet is available on a daily timescale from 1980 to the present at a projected 1 km spatial resolution (<http://daymet.ornl.gov/dataaccess>). NARR (Mesinger et al., 2006) is an assimilated meteorological reanalysis data set that provides a complete set of meteorological variables (e.g. pressure, wind). It is available on a 3 h timescale from 1979 to the present at a 36 km horizontal grid spacing. Some other new meteorological forcing data sets are also available (e.g. Abatzoglou, 2013).

To compare differences, the four selected meteorological forcing data sets were processed to a consistent format on the 4 km grids described in Sect. 2.1. Spatial interpolation was performed for both Maurer and NARR from $1/8^\circ$ (~ 12 km) and 36 km to $1/24^\circ$ (~ 4 km) for direct comparison. The 3 h wind speed from the lowest layer in NARR was used to calculate mean daily wind speed for comparison with the wind speed provided by Maurer et al. (2002).

Spatial aggregation was performed for Daymet to gather information from 1 km to $1/24^\circ$ (~ 4 km) horizontal grid spacing. Overall, daily precipitation and minimum and maximum temperatures are available from 1980 to 2010 for Maurer, Daymet, and NARR; but daily wind speed is available only for Maurer and NARR. An overall comparison is presented in Sect. 3.1. Methodologically speaking, aggregation is considered to be more justifiable than interpolation since it does not result in information loss. The direct interpolation of Maurer and NARR to $1/24^\circ$ (~ 4 km) resolution, while sufficient for the purpose of comparison at the HUC8 scale in Sect. 3.1, is questionable for $1/24^\circ$ (~ 4 km) simulation unless other factors such as topography and elevation are also considered during interpolation. Since the main purpose of this study is not to create a new forcing data set, Daymet was considered to be the most appropriate choice for $1/24^\circ$ (~ 4 km) simulation given its finer 1 km spatial resolution. Daymet may be further improved through monthly scale bias adjustment (e.g. Ashfaq et al., 2010), but that was not done in the current study. For non-US regions in which Daymet is unavailable, NARR information was used.

2.3 Soil parameters

Soil parameters are mainly used to describe the process of infiltration and base flow generation in hydrologic modeling. Given their heterogeneous nature and the lack of an effective remote sensing method, soil parameters remain the most uncertain of all parameters. Intensive hydrological model calibration is usually performed on soil parameters to improve overall model performance. In this initial effort, the conterminous US soil data set (CONUS-SOIL; Miller and White, 1998) was used to provide the required soil information for hydrologic modeling. CONUS-SOIL was derived from the State Soil Geographic data set (Schwarz and Alexander, 1995). It provides commonly used soil characteristics, such

as texture, bulk density, and porosity, arranged in 11 standard layers ranging from 0 to 2.5 m in depth and is specifically aimed at hydro-climate applications. The CONUS-SOIL data set is available in 1 km spatial resolution and is provided in common GIS formats (e.g. raster or polygon). Each CONUS-SOIL grid is spatially joined to the 4 km grids (described in Sect. 2.1) so that the required soil characteristics can be summarized efficiently for further hydrologic application. Future effort will be invested in collecting soil characteristics for non-US regions that are not covered by CONUS-SOIL.

2.4 Vegetation parameters

In considering the water budget for a large study domain, uptake and evapotranspiration from vegetation is a critical factor, since it has a significant influence on the seasonality of the simulated hydrology. Because of the rapid improvement in remote sensing data over the past decade, historical surface vegetation can now be more effectively captured to support large-scale hydro-climate simulation.

In this study, both the University of Maryland (UMD) land cover classification (Hansen et al., 2000) and the National Aeronautics and Space Administration Moderate Resolution Imaging Spectroradiometer (MODIS) model 15A2 leaf area index (LAI) information were imported. The UMD land cover classification consists of 14 categories of vegetation (e.g. evergreen needleleaf forest, mixed forest) and is available at 1 km spatial resolution in GeoTIFF format. To link the 1 km land cover classification to the 4 km grids used in this study, a conversion table for these two grid systems was developed. Both grids were first converted to polygons and then spatially intersected to form a massive table that contains the overlapping surface area of each spatial unit and the unique identifiers from both data sets. This massive conversion table was then used to summarize the portion of the UMD land cover classification in each 4 km grid efficiently.

To capture the seasonal pattern of surface vegetation, the MODIS15A2 LAI was included. LAI, defined as the green leaf area per unit of ground area (leaf area/ground area), is a widely used dimensionless canopy index. Depending on the type of vegetation, LAI may show significant seasonal trends. The historical time series of LAI were constructed using MODIS remote sensing data. The MODIS15A2 information is available every 8 days and is stored in HDF format at approximately 1 km spatial resolution in sinusoidal projection. For consistency with the UMD land classification, the MODIS LAI values were spatially interpolated to the UMD 1 km grids. The interpolated 8 day LAI values were then aggregated for each month and linked to the 4 km grids via the same conversion table developed for the UMD grids. Overall, the monthly LAI time series are organized from 2003 to the present and can be used to support various hydrologic applications. The overall statistics are summarized in Sect. 3.2.

2.5 Elevation and topography

Elevation and topography have a significant influence on surface hydrology. For instance, slope directly affects flow velocity, and local topographical depressions may create impoundments and delay surface runoff. Snow accumulation and snowmelt are also closely related to elevation. To refine the spatial resolution of a hydro-climate assessment, a fine-resolution elevation data set is needed. In this study, the 1/3 arcsec-resolution (~ 10 m) USGS NED (Gesch et al., 2002) was used for the conterminous US. NED is a seamless data set with the best available raster elevation data in the US. Similar to the treatment of UMD grids, each NED grid was labeled with a unique 4 km grid identifier. Average elevation, average slope, and the histogram of the elevation at each 4 km grid were then computed. For regions outside the US, the 90 m Shuttle Radar Topography Mission elevation was used instead (Farr et al., 2007). The pre-organized information can be processed efficiently for further applications.

2.6 Observed streamflow and runoff

Historical hydrologic observations are required for model calibration and validation. Two types of observations, streamflow and runoff, can be used to support hydrologic model calibration. Gauge-based streamflow observations directly measure flow discharge at a specific river section. Comprehensive daily flow observations can be obtained from the USGS National Water Information System (NWIS) for more than 22 000 current and retired gauge stations throughout the US.

Another observational product, the USGS WaterWatch runoff (Brakebill et al., 2011), was found to be more useful in this study. Derived from the comprehensive NWIS gauge observation, WaterWatch runoff is the assimilated time series of flow per unit of area calculated for each conterminous HUC8 subbasin. For each HUC8 subbasin, multiple NWIS gauge stations located within or downstream of the HUC8 were used to estimate the runoff generated locally at each HUC8. The contributing drainage areas (both gauge-to-HUC8 and HUC8-to-gauge) were converted as weighting factors to merge runoff time series from all stations. As a result, gauges with drainage coverage most similar to that of the particular HUC8 received the highest weights. Therefore, the influence of highly regulated gauge stations (usually with large drainage coverage across multiple HUC8s) could be reduced. This approach may effectively assimilate streamflow observations from multiple gauge stations as a consistent areal HUC8 runoff measurement that has a unit similar to that for precipitation (depth/time). WaterWatch runoff is available monthly from 1901 to the present. Note that WaterWatch runoff is based on an earlier version of watershed boundaries and was found to be slightly different from the new watershed boundaries adopted in Sect. 2.1. Using the polygon shapefiles from both versions of the watershed

boundaries, a conversion table based on overlapping drainage areas was developed to adjust the WaterWatch runoff to a consistent watershed boundary for further comparison.

Although WaterWatch runoff is a promising new data set for hydro-climate studies and has been found useful in Ashfaq et al. (2013) and Sale et al. (2012), there are some questions and limitations that deserve further exploration. For instance, for HUC8 subbasins with very few gauge stations (e.g. arid regions in the central US), data accuracy and quality would be an issue. In addition, since the exact weighting factors for each station are not published by USGS, it would be difficult to examine and further quantify the influence of highly regulated stations (e.g. reservoir or diversion) in WaterWatch runoff. Nevertheless, for the purposes of initial model development and calibration, WaterWatch should be an ideal choice given its consistent format across various subbasins. Future detailed studies on the accuracy of WaterWatch runoff would be highly useful.

2.7 Baseline hydrologic model application

In this study, the widely used VIC model (Nijssen et al., 1997; Liang et al., 1994, 1996; Cherkauer et al., 2002) was selected as the baseline hydrologic model for the conterminous US. VIC has been successfully tested in a many hydrologic studies and large river networks (Gao et al., 2010; Ashfaq et al., 2010; Su et al., 2005; Nijssen et al., 1997, 2001; Bowling et al., 2004; Lohmann et al., 1998). The current VIC model studies were mostly conducted at $1/8^\circ$ spatial resolution (~ 12 km). Given its wide acceptance and the fact that it can be directly implemented for parallel computing, the VIC model was considered the most suitable baseline hydrologic model for this initial effort. VIC is a process-based hydrological model that simulates evapotranspiration, snow pack, surface runoff, base flow, and other hydrologic mechanisms at daily or subdaily time steps. The water and energy balances are solved for multiple elevation bands and vegetation types, which allows the model to capture the subgrid-scale variability of these land surface features. The model simulates all processes in each grid cell independently in an equally spaced grid. Infiltration and runoff are estimated using the VIC model curve, which uses the soil moisture content of the upper two soil layers to approximate the spatial variability of surface saturation. The empirical Arno curve is used to generate base flow based on the soil moisture content in the bottom layer (Cherkauer and Lettenmaier, 2003). A routing algorithm external to the VIC model can then be used to simulate the streamflow at a specified location by routing runoff and base flow from each grid cell (Lohmann et al., 1998).

VIC requires a large number of parameters, including soil, vegetation, elevation, and daily meteorological forcings, at each grid cell. By taking daily precipitation, maximum/minimum temperature, and wind speed as inputs, VIC computes potential evapotranspiration using the Penman–Monteith equation (see Maidment, 1993). Other forcings,

including shortwave and long-wave radiation, relative humidity, and vapor pressure, are estimated using algorithms from MTCLIM (Kimball et al., 1997; Thornton and Running, 1999) at subdaily time steps (3 h). Additionally, subdaily temperatures are estimated within the model as a parameterization of maximum/minimum temperature (Bohn et al., 2013).

For soil physical properties, the CONUS-SOIL information was divided into three layers covering the total depth from 0 to 2.5 m (CONUS-SOIL layers 1 and 2 to VIC model layer 1, CONUS-SOIL layers 3–7 to VIC model layer 2, and CONUS-SOIL layers 8 and 9 to VIC model layer 3). The 1 km CONUS-SOIL information was then aggregated to 4 km grids. When a three-layer configuration was used, a total of 53 soil parameters was required, including saturated hydrologic conductivity, initial soil moisture, bulk density, layer thickness, fraction of soil moisture at wilting point, and some other conceptualized parameters such as the variable infiltration curve parameter ($b_{infiltr}$), which required further calibration. Although CONUS-SOIL contains a number of different soil characteristics, only a few are directly called out in the VIC model soil parameter file (e.g. bulk density). Most VIC soil parameters are derived from porosity and soil texture class according to a standard index table provided by the VIC modeling group. A few other non-soil parameters requested in the soil parameter file – such as the average annual air temperature, average annual precipitation, average elevation, and slope (used to derive the maximum velocity of the base flow) – were derived from Daymet and NED. If the CONUS-SOIL information was totally unavailable for a specific grid point, the information from the nearest grid point was used instead.

The VIC model vegetation parameter file describes the number and percentage of vegetation types in each grid cell. The conversion table described in Sect. 2.4 was used to efficiently summarize the UMD vegetation classification to the format requested by VIC. To improve the characterization of surface vegetation, a data-intensive enhancement was applied to import the monthly LAI observation from MODIS at each 4 km grid. The monthly LAI was first computed from the 2003–2008 MODIS at each 1 km UMD grid and then converted to the subgrid vegetation information in the VIC model format. Although the time frame between UMD (before 2000) and MODIS LAI (2003–2008) is somewhat inconsistent, given that there is no suitable alternative, both data sets are considered to be the best proximity for the actual vegetation class and LAI.

To better represent snow accumulation and snowmelt, subgrid elevation band information can be set up (i.e. fractions of the grid area with their corresponding mean elevations). Since NED provides a much finer spatial resolution (10 m) than the 4 km grid, the subgrid elevation band can be described in considerable detail. However, the elevation band should also be considered in terms of the required computation resources. It was found that the required

Table 1. Selected VIC parameters for calibration.

Parameter	Range	Units	Description
b_{infiltr}	0.001 ~ 0.8	N/A	Variable infiltration curve parameter
exp	8 ~ 30	N/A	Exponent of the Brooks–Corey drainage equation
thick ₂	0.1 ~ 2	m	Thickness of soil layer 2
Ds	0 ~ 1	N/A	Fraction of the maximum velocity of base flow where nonlinear base flow begins
Ws	0.5 ~ 1.0	N/A	Fraction of maximum soil moisture where nonlinear base flow occurs

computational time is roughly proportional to the number of elevation bands; therefore, in some flat regions, it might not be worthwhile to use multiple bands. For flexibility, instead of creating elevation bands at fixed intervals, the histogram of elevation at each 4 km was summarized. Depending on the required research question and the available resources, a suitable number of elevation bands can be generated in the VIC model format efficiently.

2.8 Calibration through high-performance computing

Although the process-based models incorporated various explicit physical mechanisms, given the complexity of hydrologic phenomena, parts of the processes still relied on conceptual statistical parameterization. As a result, several non-physically based parameters required further calibration before a model could be put to use. Calibration was also required for those parameters with high measurement uncertainty (e.g. most of the soil parameters), since they may affect the performance of hydrologic modeling significantly. Nevertheless, a full hydrologic calibration is extremely resource intensive, especially when the model application requires fine resolution and sophisticated mechanisms. As recommended by other studies, calibrating the model at multiple locations at a small scale may improve streamflow prediction (Chien et al., 2013; Zhang et al., 2008; Wang et al., 2012; Jetten et al., 2003). For example, Jetten et al. (2003) conclude that predicting the hydrologic response at a single watershed outlet may result in the phenomenon of “predicting the correct result for the wrong reasons” and suggested using multi-site calibration within a watershed to reduce the possibility of accurate simulation at a single watershed outlet from a combination of locally inaccurate simulations. For this reason, we performed a first-order modeling calibration for each HUC8 consistently through a computationally exhaustive algorithm to improve the overall model performance. This large-scale calibration was mainly targeted at narrowing the possible range of suitable parameter values in each HUC8. Depending on the needs of future research, further fine calibration can be performed efficiently.

Following the sensitivity analysis by Demaria et al. (2007), five sensitive VIC parameters were selected for calibration, including the variable infiltration curve parameter (b_{infiltr}), exponent of the Brooks–Corey drainage equation (exp), thickness of soil layer 2 (thick₂), fraction of the maximum velocity

of base flow at which nonlinear base flow begins (Ds), and fraction of maximum soil moisture at which nonlinear base flow occurs (Ws). Although other VIC parameters could also be important (e.g. thickness of soil layer 3, thick₃), they were not considered in the current effort given the computational resource limitations. Although the soil parameters were obtained with a pre-specified soil depth, the thickness of soil layer 2 (root layer) was treated as a parameter and can be changed during calibration. Given that the USGS WaterWatch runoff can provide an estimate of local runoff at each HUC8, our calibration was performed by matching the simulated total monthly runoff (base flow + surface runoff) to the observed WaterWatch monthly runoff. In other words, the offline routing model was not used, a decision similar to that of Demaria et al. (2007). For each parameter, three combinations, including the upper and lower bounds and the suggested default values, were chosen for calibration (Table 1). A total of 243 (3^5) parameter scenarios were then prepared. The VIC model simulation was driven by Daymet daily meteorological forcing (precipitation and minimum/maximum temperature) and NARR daily wind speed from 1980 to 2008. Year 1980 was treated as the model startup period, 1981–2000 as the calibration period, and 2001–2008 as the validation period. The VIC model simulation was performed in 3 h time steps using the energy and water balance mode. VIC version 4.1.1 was used in the current study, and the recently released VIC 4.1.2 will be incorporated after the model development work has been stabilized.

To effectively manage the data flow, all forcing, soil, vegetation, global parameter, and output flux files were organized in separate HUC8 folders. Depending on the total watershed area, all grid points within a HUC8 were subdivided into 16, 32, or 48 computation units. Each computational unit had separate global, soil, vegetation, and elevation parameter files and could be executed independently. Computation was performed using Oak Ridge National Laboratory’s Titan supercomputer, a Cray XK7 system with 18 688 computational nodes, each equipped with four quad-core CPUs and two GPU cards. The extensive simulation exhausted ~ 1.5 million CPU-hours (i.e. the number of CPUs multiplied by the average hours used by each CPU), approximately 171 calendar years if done by a single-core desktop machine. Note that we had originally planned to use a 10-layer elevation band during calibration, but it would have

Table 2. Matrices used for model calibration and evaluation.

Matrix	Name	Equation
R^2	Coefficient of determination	$\frac{\left[\sum_{t=1}^n (O_t - \bar{O})(Y_t - \bar{Y}) \right]^2}{\left[\sum_{t=1}^n (O_t - \bar{O})^2 \right] \left[\sum_{t=1}^n (Y_t - \bar{Y})^2 \right]}$
Nash	Nash–Sutcliffe model efficiency coefficient	$1 - \frac{\sum_{t=1}^n (O_t - Y_t)^2}{\sum_{t=1}^n (O_t - \bar{O})^2}$
MAE	Mean absolute error	$\frac{1}{n} \sum_{t=1}^n O_t - Y_t $
RMSE	Root mean square error	$\sqrt{\frac{1}{n} \sum_{t=1}^n (O_t - Y_t)^2}$

resulted in a huge increase in the required computational time (an estimated 15 million CPU hours, approximately) over our allowable resource. Given that the current calibration was targeted for total monthly runoff and was less affected by the elevation band, a one-layer elevation band was used. When hydro-climate projections are produced for future research, multiple elevation bands will be implemented.

Four statistical matrices, the coefficient of determination (R^2), Nash–Sutcliffe model efficiency coefficient (Nash), mean absolute error (MAE), and root mean square error (RMSE), were used to evaluate model performance. The matrices are summarized in Table 2, in which O_t and Y_t represent the observed and modeled monthly total runoff from month 1 to n , and \bar{O} and \bar{Y} represent the mean of O_t and Y_t . For each HUC8, daily total runoff was computed by summing base flow and surface runoff at each 4 km grid and then aggregated up to calculate the HUC8 monthly runoff (Y_t). The observed monthly runoff (O_t) from the USGS WaterWatch was then used for model evaluation. In addition to runoff analysis, the simulated 1 April snow water equivalent (SWE) was compared with the snow course observations. The results are reported in Sect. 3.4.

3 Results and discussions

3.1 Differences among forcing data sets

To understand the differences among four selected forcing data sets (Maurer, PRISM, Daymet and NARR), an overall comparison was performed (illustrated in Fig. 3). Average daily maximum temperature (T_{\max}), daily minimum temperature (T_{\min}), annual total precipitation (P), and average wind speed (W) from 1980 to 2008 for each of the 2107 HUC8s were computed for comparison. The correlation coefficients among the HUC8 average values were also computed. Given PRISM's high accuracy (Daly et al., 2008), it is placed on the x axis as the target for comparison. Figure 3a

illustrates the difference for T_{\max} . Although both Maurer and Daymet are close to the PRISM observation, NARR seems to be much warmer in most of the HUC8s. The difference is not as significant for T_{\min} in Fig. 3b, where most of the data sets are similar to each other; NARR is slightly warmer than the Daymet and Maurer data sets. The warm bias of NARR was also reported by Royer and Poirier (2010). Additionally, Daly et al. (2008) showed a cold bias in the Daymet data set compared with the PRISM January minimum temperature, whereas differences between PRISM and Daymet July maximum temperatures are relatively small. For the difference of P , a consistent observation can be made in Fig. 3c. Both Daymet and Maurer are closer to PRISM, but NARR is more divergent than the other data sets. Daly et al. (2008) showed that the differences of P , T_{\min} and T_{\max} between Daymet and PRISM were more prominent in the western US than in the east. They also indicated that Daymet tended to be drier than PRISM in some locations and wetter in others; this was because Daymet did not resolve rain shadows well owing to an inability to recognize topographic facets. Given that wind speed is available only for Maurer and NARR, only one set of points is plotted in Fig. 3d. A significant difference can be seen in the two data sets, with a correlation coefficient of around 0.28.

A further comparison of high/low daily quantiles is shown in Fig. 4. The mean annual 95 % quantile of daily maximum temperature ($T_{\max,95}$, Fig. 4a), 5 % quantile of daily minimum temperature ($T_{\min,05}$, Fig. 4b), 95% quantile of daily precipitation (P_{95} , Fig. 4c), and 95% quantile of daily wind speed (W_{95} , Fig. 4d) from 1980–2008 are plotted for each HUC8. Since PRISM is available only in a monthly scale, it is not included in the comparison. For $T_{\max,95}$, Daymet remains similar to Maurer, but the difference between NARR and others becomes more significant. The comparison of $T_{\min,05}$ is similar to that of T_{\min} (Fig. 3b), but some HUC8s seem to be much cooler in Maurer than in Daymet. For high precipitation P_{95} , although Daymet, Maurer, and NARR are generally linearly correlated, Daymet was

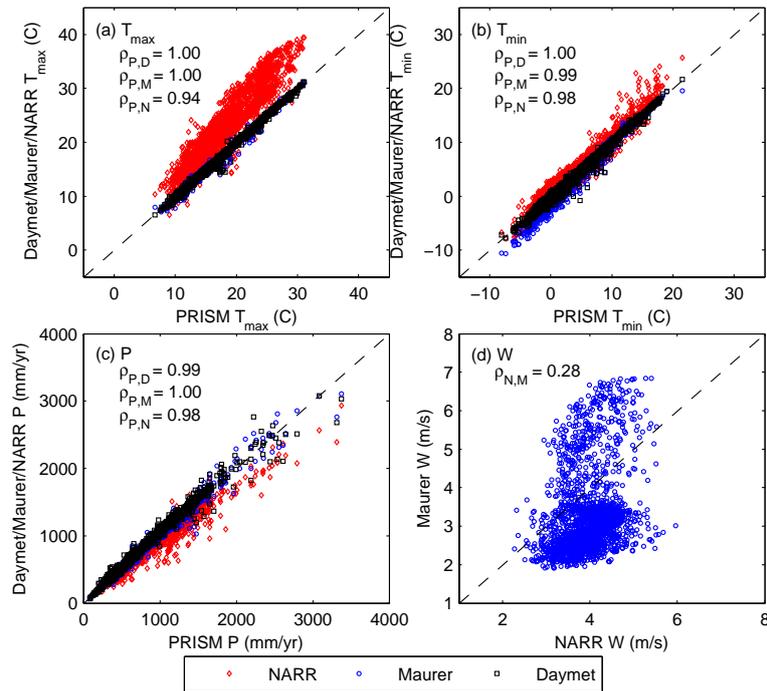


Fig. 3. Comparison among the 1980–2008 PRISM, Daymet, Maurer, and NARR (a) mean daily maximum temperature T_{max} , (b) mean daily minimum temperature T_{min} , (c) annual total precipitation P , and (d) mean daily wind speed W . Each point represents the HUC8 annual average over the entire period.

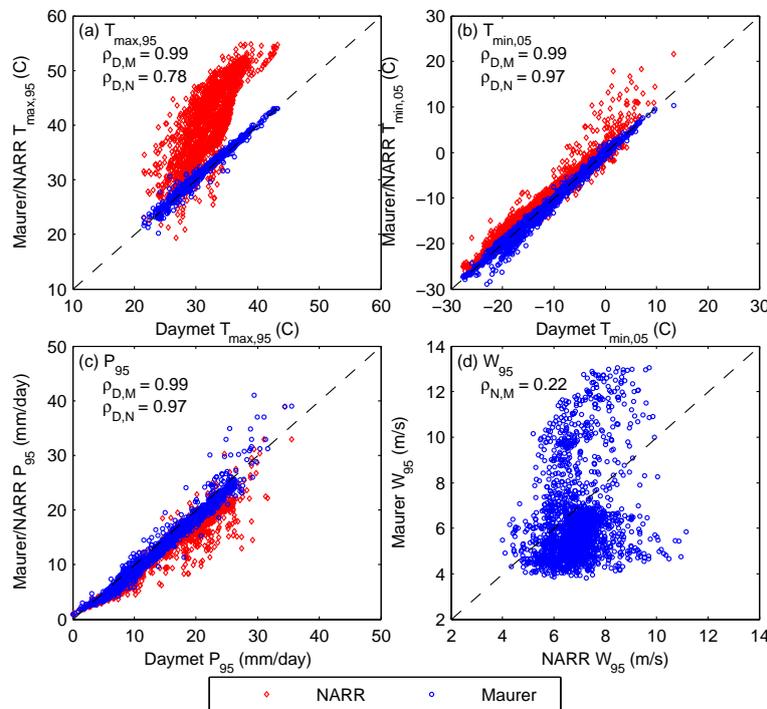


Fig. 4. Comparison among the 1980–2008 Daymet, Maurer, and NARR (a) mean annual 95 % quantile daily maximum temperature $T_{max,95}$, (b) mean annual 5 % quantile daily minimum temperature $T_{min,05}$, (c) mean annual 95 % quantile daily precipitation P_{95} , and (d) mean annual 95 % quantile daily wind speed W_{95} . Each point represents the HUC8 annual average over the entire period.

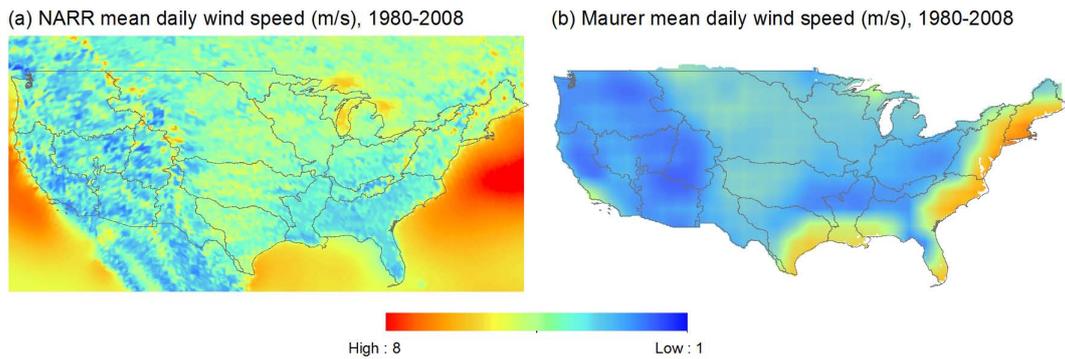


Fig. 5. Maps of mean daily wind speed.

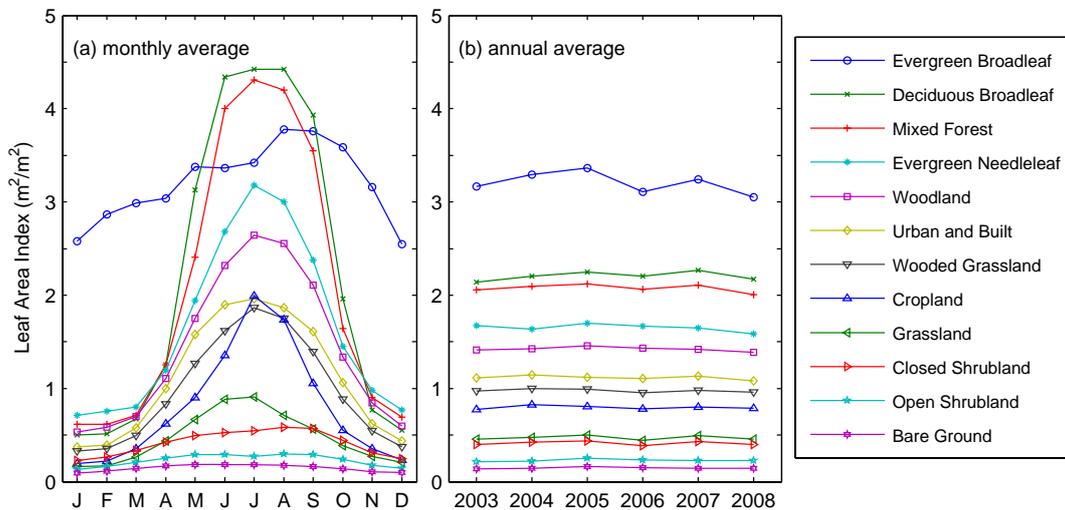


Fig. 6. The MODIS leaf area index summarized by the UMD land cover classification.

found to consistently provide a higher value than Maurer and NARR. That is not a surprise, since precipitation is highly resolution dependent (Gao et al., 2006); therefore, the finer resolution Daymet should report precipitation extremes more accurately at the 4 km scale. In other words, even using the same ground-based observations, the precipitation extremes calculated at 12 km resolution will be smoothed further than those calculated at 4 km; and it is unlikely that such spatial variability can be faithfully reconstructed from coarser resolution to finer resolution. For W_{95} , the comparison is very similar to that for W (Fig. 3d).

To understand the differences, the geographical wind speed patterns are plotted in Fig. 5. Clearly, the inconsistency should be from the differences in the original data sources. The wind speed provided in Maurer's data set was calculated from the coarser-resolution reanalysis data set and hence shows a smoother pattern in Fig. 5b. Given that NARR can provide a more delicate local wind speed pattern, the NARR wind speed was chosen as the default wind speed in this study.

3.2 Monthly and annual statistics of LAI

To examine the variability of the MODIS LAI, the monthly and annual average LAI are plotted in Fig. 6. For each UMD class, the mean monthly MODIS LAI values were calculated for the entire conterminous US from January 2003 to December 2008. The monthly average LAI values are shown in Fig. 6a, in which the highest LAI values are found for evergreen broadleaf, deciduous broadleaf and mixed forest, and the lowest LAI values for bare ground, open shrubland, and closed shrubland. In terms of seasonal pattern, the LAI values for evergreen broadleaf are consistently high across all seasons. Both deciduous broadleaf and mixed forest have the strongest seasonal variation and can be larger than evergreen broadleaf during summer. The annual averages are plotted in Fig. 6b and further broken down to 18 conterminous US hydrologic regions in Fig. 7. Given that the annual variability is not significant in both figures, it should be justifiable to calculate the required VIC monthly vegetation parameters by averaging the monthly MODIS LAI values from 2003 to 2008. This simplification is needed because the current VIC

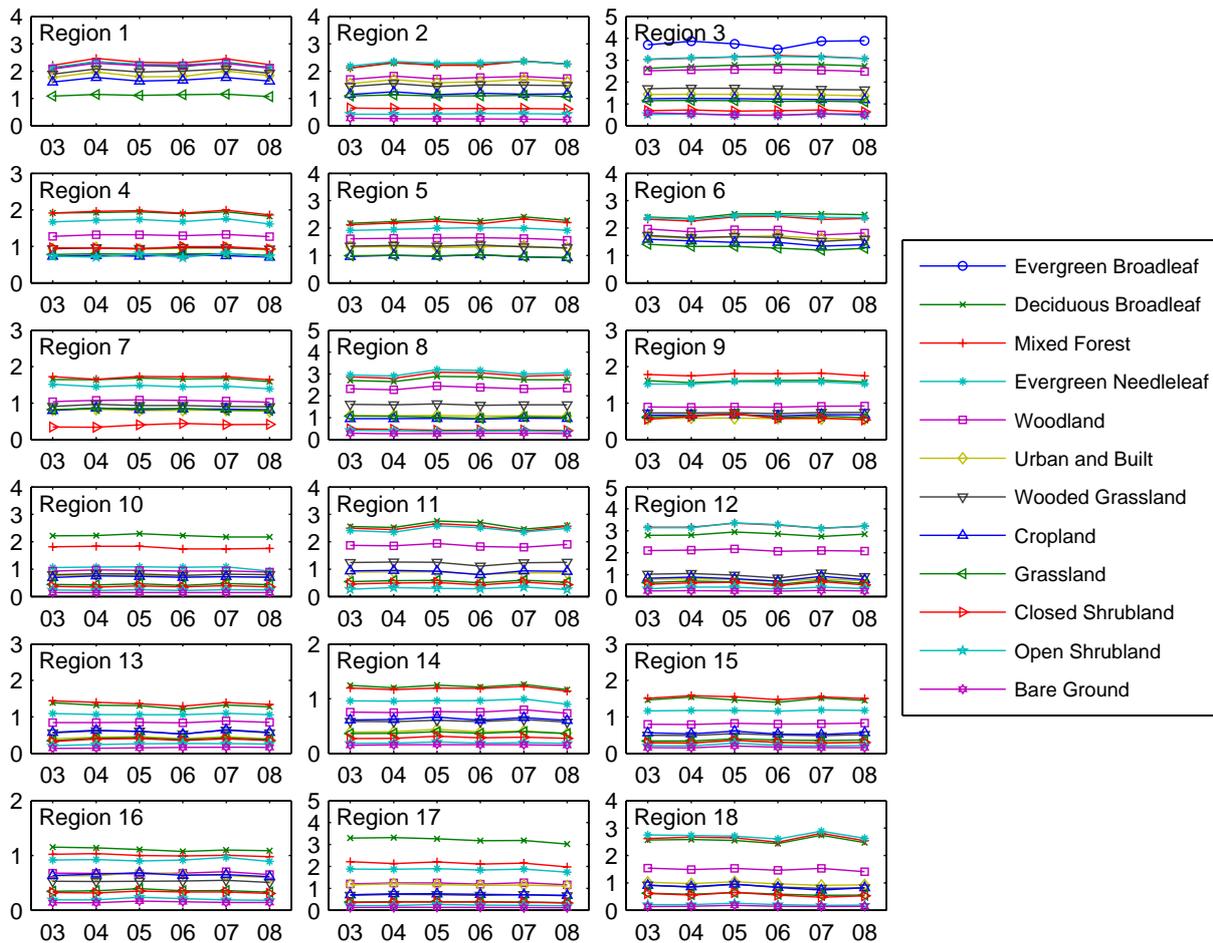


Fig. 7. Summary of MODIS leaf area index ($\text{m}^2 \text{m}^{-2}$, y axis) from 2003 to 2008 (x axis) by the UMD land cover classification for each US hydrologic region.

model does not allow dynamic vegetation simulation. It is interesting to note that the LAI values reported in the urban and built land class are not among the smallest. Those results may be explainable by considering that the findings are based on 1 km grid resolution, and at that resolution, many suburban areas are still covered by vegetation.

3.3 Difference between runoff aggregation and routing

To simulate streamflow in the VIC model, a separate routing model developed by Lohmann et al. (1998) was required. The routing model simulated a channel network with a number of nodes, each of which represented information from a grid cell. A unit hydrograph was then used to route the simulated surface runoff and base flow through a channel network using a linearized St. Venant’s equation (Lohmann et al., 1998). The routing model required five types of input: flow direction, grid area fraction, flow velocity, watershed boundary mask, and gauge locations. Whereas flow direction, grid area fraction, and watershed boundary can be derived from digital elevation models, flow velocity and unit hydrograph

involve larger uncertainty and cannot be easily estimated. As discussed in Sect. 2.1, the resolution of digital elevation models also has a significant influence on the accuracy of river networks.

Although the major purpose of a routing model is to account for the travel time of river flow, this step may reasonably be skipped in smaller watersheds (i.e. when travel time is short) by using a simpler runoff aggregation method (Demaria et al., 2007). To evaluate the difference, a comparative analysis was performed for two randomly selected USGS gauge stations. Both gauge 01047000 in HUC01030003 and 02342500 in HUC03130003 have experienced little or no anthropogenic disruption and have complete records from 1980 to 2008. Routing models were set up to calculate streamflow at these two gauge stations. The simulated total monthly runoff (surface runoff + base flow) and monthly average streamflow are illustrated in Fig. 8; runoff observations were taken from WaterWatch and gauge observations from NWIS. The correlation coefficient (ρ) between the observed and simulated runoff/streamflow was also calculated.

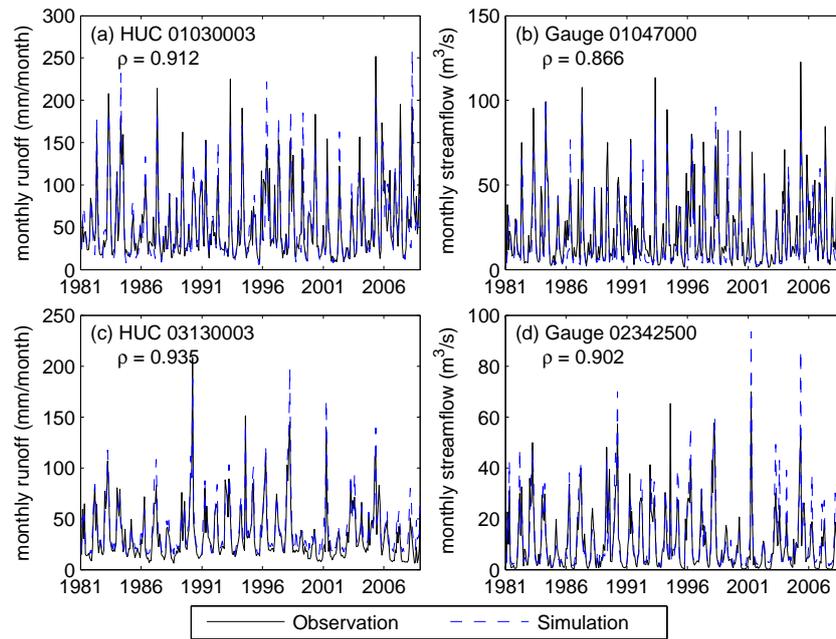


Fig. 8. The comparison between observed and simulated runoff (left panels) and observed and simulated streamflow (right panels). Gauge 01047000 is located at HUC01030003 and gauge 02342500 is located at HUC03130003.

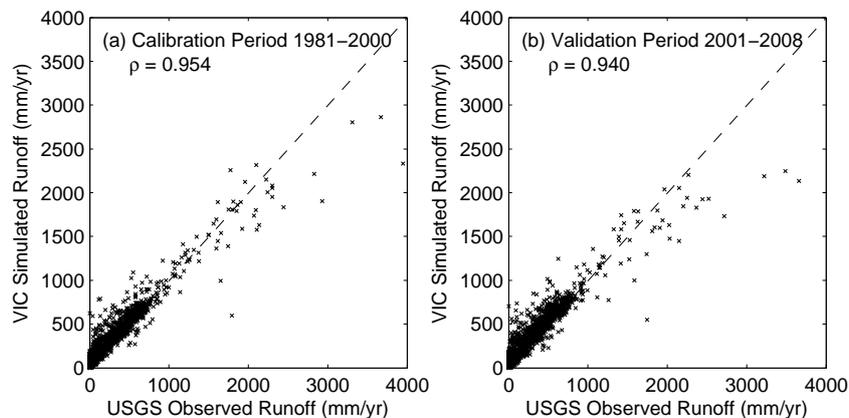


Fig. 9. The USGS WaterWatch observed runoff versus the VIC simulated annual total runoff (surface runoff + base flow) for both calibration and validation periods for each HUC8 subbasin.

Generally speaking, the model performance showed large similarities for both approaches, with correlation coefficients between simulation and observation varying from 0.8 to 0.9. Although the travel time was not modeled in the runoff aggregation approach, the extra uncertainty induced by the routing model was also avoided (e.g. flow speed, routing resolution), so pros and cons exist for both methods. Given that our main objective was to provide a first-order calibrated hydrological parameter data set to expedite further efforts at fine calibration, it is more efficient and consistent to calibrate VIC for each HUC8 through the runoff aggregation approach. Also, since it is extremely time-consuming to develop reasonable routing models for all HUC8s in the US, the

runoff aggregation approach is an easier alternative than spatially examining the model performance for a great number of watersheds in the US.

3.4 Overall model performance

Overall model performance is illustrated in Fig. 9. For each HUC8, the WaterWatch-observed and VIC-simulated total annual runoff (base flow+surface runoff) were computed for both calibration and validation periods. The correlation coefficients ρ between observed and simulated HUC8 annual runoff are 0.954 in the calibration period and 0.940 in the validation period, which is satisfactory overall. The results

Table 3. Summary of median HUC8 matrices for each hydrologic region.

	Median statistics							
	Calibration period (1981–2000)				Validation period (2001–2008)			
	R^2	Nash	MAE (mm yr ⁻¹)	RMSE (mm yr ⁻¹)	R^2	Nash	MAE (mm yr ⁻¹)	RMSE (mm yr ⁻¹)
01 New England	0.843	0.826	12.2	17.8	0.829	0.810	14.4	21.2
02 Mid-Atlantic	0.825	0.798	9.8	14.3	0.824	0.746	11.1	15.4
03 South Atlantic–Gulf	0.830	0.807	10.0	14.2	0.813	0.705	10.9	14.8
04 Great Lakes	0.757	0.703	9.2	13.1	0.748	0.633	9.7	14.0
05 Ohio	0.849	0.824	10.4	14.9	0.865	0.823	10.8	14.6
06 Tennessee	0.809	0.776	12.5	16.9	0.815	0.766	12.7	17.8
07 Upper Mississippi	0.713	0.668	8.5	12.6	0.764	0.692	9.0	12.8
08 Lower Mississippi	0.774	0.704	16.7	23.4	0.744	0.596	17.8	24.9
09 Souris–Red–Rainy	0.488	0.416	3.6	6.3	0.421	0.269	4.9	8.1
10 Missouri	0.368	0.156	2.8	4.0	0.433	−0.296	2.7	3.5
11 Arkansas White–Red	0.562	−0.006	8.9	12.2	0.596	−0.921	9.0	11.7
12 Texas Gulf	0.519	−0.442	11.1	14.8	0.573	−1.929	14.1	17.8
13 Rio Grande	0.078	−63.106	5.4	6.8	0.060	−56.580	5.2	6.4
14 Upper Colorado	0.339	0.113	5.0	6.6	0.409	−0.767	4.3	5.7
15 Lower Colorado	0.160	−12.028	7.2	9.0	0.130	−34.077	4.7	5.3
16 Great Basin	0.394	−0.183	6.1	8.0	0.415	−2.020	5.6	7.0
17 Pacific Northwest	0.669	0.527	15.2	21.3	0.652	0.478	13.3	18.2
18 California	0.702	0.590	9.6	17.1	0.689	0.255	8.4	14.4

represent an improvement from 0.906 using an uncalibrated 4 km data set (i.e. with default parameters) and from 0.877 (Fig. 10) using the Ashfaq et al. (2010) 12 km data set, both with a much larger spread.

To examine the model performance in different seasons, Fig. 9 is further broken down into winter (December–January–February, DJF), spring (March–April–May, MAM), summer (June–July–August, JJA), and fall (September–October–November, SON) in Fig. 11. It can be observed that the current parameter set works best in winter, with $\rho = 0.966$ in the calibration period and $\rho = 0.954$ in the validation period. A strong dry winter bias is observed in some HUC8s, which require further calibration for future application. Following winter, the model performance in spring and fall remains good, with ρ greater than 0.929 in the calibration period and 0.908 in the validation period. The worst model performance is observed in summer, with $\rho = 0.866$ in the calibration period and 0.830 in the validation period. The worse performance in summer is not surprising. Since we are calibrating monthly time series across different seasons, the Nash–Sutcliffe coefficient will be controlled more by wetter months (winter) than drier months (summer). If the future focus is on summer months, different objective functions should be designed and used for further model adjustment.

In Fig. 9, it can be seen that the current parameter sets overestimate runoff in multiple drier HUC8s. To spatially examine the model performance for the entire conterminous US, the observed and simulated annual runoff, R^2 ,

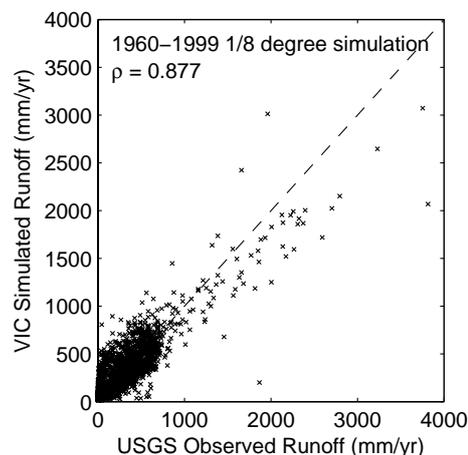


Fig. 10. The 1960–1999 USGS WaterWatch observed runoff versus the VIC simulated annual total runoff (surface runoff + base flow) for each HUC8 subbasin using the Ashfaq et al. (2010) 1/8° (~ 12 km) resolution data set.

Nash, MAE, and RMSE for each HUC8 are illustrated in Fig. 12. The median values of the HUC8 evaluation matrices in each hydrologic region are also summarized in Table 3. From Fig. 12a and b, it can be seen that VIC generally captures the spatial patterns of WaterWatch runoff. However, the simulated runoff is higher in many HUC8s, especially in very dry regions such as the Rio Grande (HUC 13), Lower Colorado (HUC 15), Texas (HUC 12), Great Basin (HUC 16),

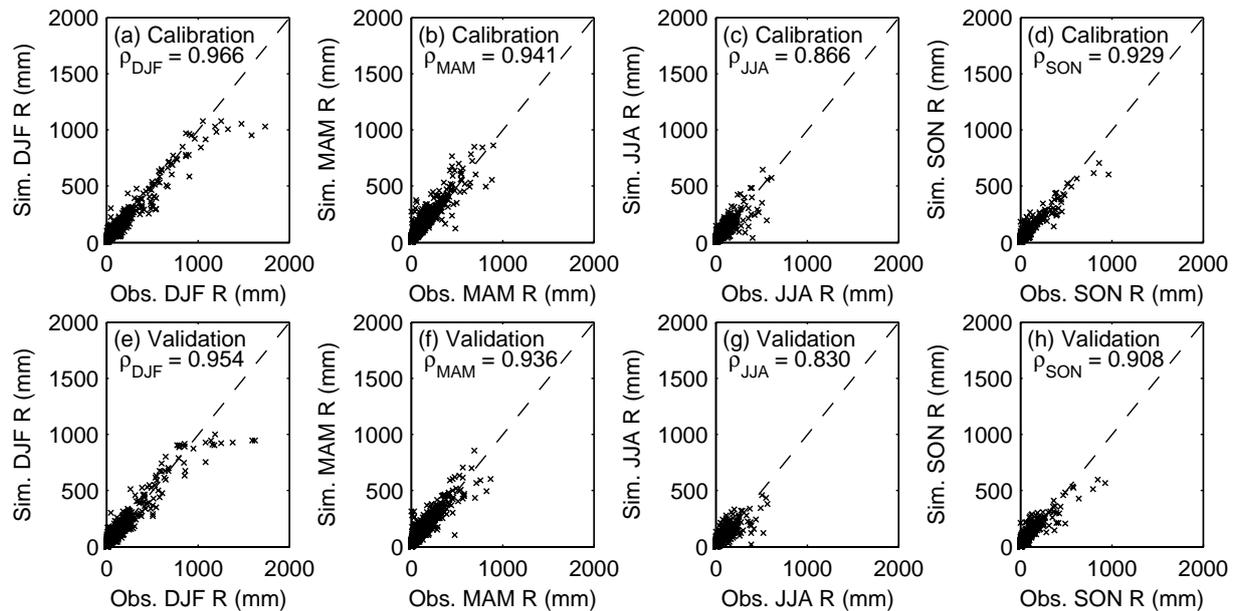


Fig. 11. The USGS WaterWatch observed runoff (x axis) versus the VIC simulated seasonal total runoff (surface runoff+base flow, y axis) for both calibration periods (a)–(d) and validation periods (e)–(f) for each HUC8 subbasin. Winter runoff from December to February is plotted in (a) and (e), while spring runoff from March to May in (b) and (f), summer runoff from June to August in (c) and (g), and fall runoff from September to November in (d) and (h).

and Arkansas White–Red (HUC 11). To enable a closer look, both R^2 (Fig. 12c) and Nash (Fig. 12d) between the observed and simulated monthly runoff are illustrated. Clearly, while the current parameter sets may provide satisfactory results for wetter regions, it is challenging to capture the monthly runoff time series in drier regions. Further studies in dry regions are required, since the suitable parameter values and model setup may have exceeded the currently suggested ones. This issue may be also related to meteorological forcing. It was noticed that Daymet generally provided higher precipitation ($\sim 25\%$ greater than PRISM) in dry HUC8s; and under such a situation, the VIC model cannot be further improved unless precipitation is reduced in proportion to the PRISM values. Other factors such as groundwater that could not be simulated by VIC, and/or managed flow that was not captured by WaterWatch runoff, could also affect this erroneous runoff simulation. Nevertheless, since these regions are fairly dry, they in fact have smaller MAE (Fig. 12e) and RMSE (Fig. 12f) compared with wet regions. Therefore, the overall impact of wet bias in drier regions may not be significant since it is on a small scale.

To evaluate the model performance for other variables, the simulated 1 April SWE was compared with the observed snow course data used by Mote et al. (2005). Focusing on the 1981–2000 period, 784 snow stations with complete annual 1 April SWE observations were selected. For each station, the simulated 1 April SWE at the nearest grid was looked up. Since the point observation may be on a different scale from the grid-based SWE, the correlation coefficient

between observation and simulation is computed for evaluation. The results are summarized in Fig. 13. Figure 13a plots the histogram of ρ and shows that high correlation coefficients can be seen in most of the stations. To examine the statistical significance, the histogram of the p value is plotted in Fig. 13b. The p value is less than 0.05 for nearly 700 stations (i.e. correlation is statistically significant under 5% significance level), which suggests that the simulation may capture the annual trend for most of the stations. To check the spatial pattern, ρ values are further plotted in Fig. 13c. Generally speaking, except for eastern Wyoming, northern Colorado, and northern California, the simulated SWE showed good correlation with observations. As mentioned, the current simulation was conducted using one elevation band (for efficiency of model calibration). With an increase in elevation bands in future simulations, the performance of snow simulation could be further improved.

The overall improvement from the computationally intensive calibration exercise is illustrated in Fig. 14. Focusing on Nash (Fig. 14a) and RMSE (Fig. 14b), the cumulative percentage of HUC8s is plotted. In terms of Nash, around 20% of HUC8s (~ 450 HUC8s) were improved from less than 0.5 to greater than 0.5. In terms of RMSE, the modeling errors were on average reduced by $\sim 5 \text{ mm yr}^{-1}$ for most of the HUC8s. While the best parameters at each HUC8 were identified, the performance of all examined combinations of parameters was also recorded to support further assessment. Calculating the sensitivity and trends for each parameter could make it possible to achieve further

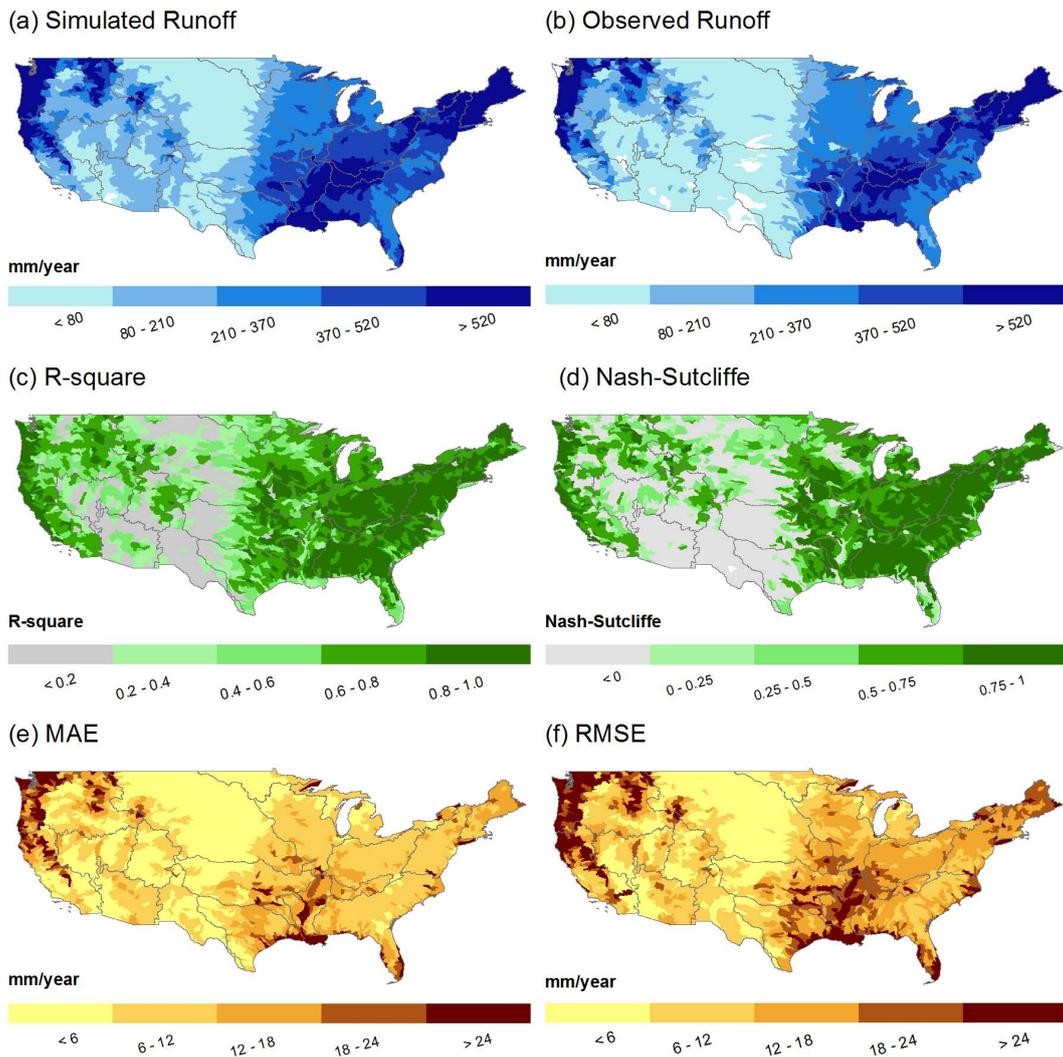


Fig. 12. Performance of calibrated VIC model at various HUC8s in the conterminous US.

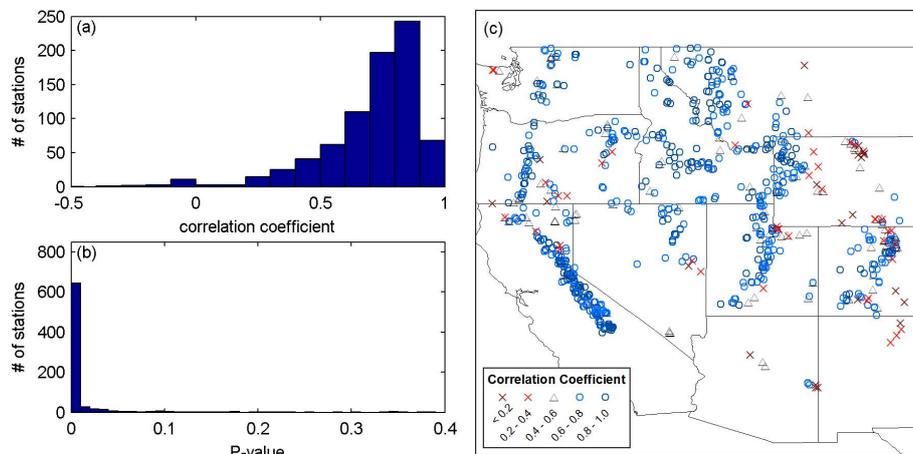


Fig. 13. Correlation coefficients between observed and simulated 1981–2000 1 April snow water equivalent: (a) histogram of correlation coefficients from 784 selected stations, (b) histogram of p value, and (c) spatial pattern of correlation coefficients for all selected stations.

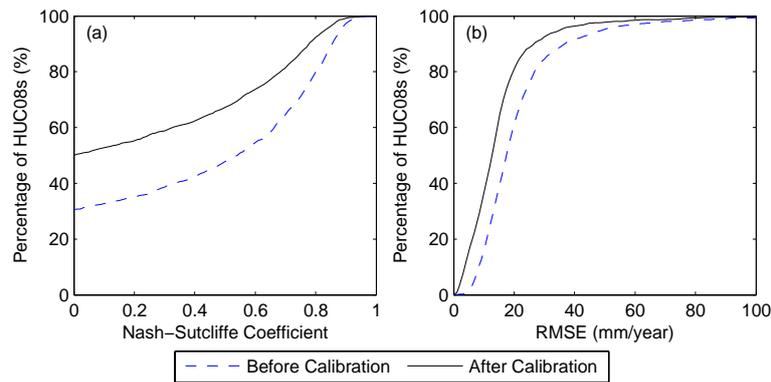


Fig. 14. Model performance before and after calibration, showing improvement.

model improvements with fewer iterations and computational hours. In other words, the proposed hydrological data set provides not only the currently best available parameters but also the tested parameter sensitivity to support further fine calibration.

4 Summary and conclusion

This study introduces an effort to prepare a comprehensive hydrological model parameter data set for large-scale, high-resolution climate change impact assessment. Several key inputs for hydrologic simulation – including meteorologic forcings, soil, land class, vegetation, and elevation – were collected and organized in refined 4 km grids. Using high-performance computing, a spatially consistent calibration was performed for the VIC model. The VIC simulation was driven by Daymet daily meteorological forcing and was evaluated by USGS WaterWatch runoff observations for 2107 HUC8 subbasins in the conterminous US. Overall, 1.5 million CPU-hours were used to develop a post-calibrated model parameter data set to support fine-scale future hydro-climate assessments. The pre-organized model parameter data set will be provided to interested parties to support further hydro-climate impact assessment. Although model calibration may yet be required for particular model applications, it is hoped the pre-organized data set will help reduce the amount of effort needed for basic data preparation and organization. Depending on the specific needs, the parameter data set can then be further calibrated effectively.

As a result of this exhaustive calibration exercise, it is now possible to more accurately estimate the resources required for further model improvement across the entire conterminous US. Calibrating a hydrologic model consistently for various watersheds can also increase understanding of the strengths and limitations of a particular hydrologic model across different climate regions. It can help in determining the best combination of hydrologic model and meteorological forcing data set for specific regions. Although the extensive model calibration was performed for the VIC model, the

computation and data framework were designed in a flexible manner so that other suitable hydrologic models could be incorporated in the future. By including multiple hydrologic model choices, we hope that it can provide flexibility for further application and increase understanding of the modeling uncertainty associated with different hydrologic models in hydro-climate impact assessment.

Acknowledgements. This research was funded by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC, for the US Department of Energy under contract DE-AC05-00OR22725. The US Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US Government purposes.

Edited by: J. Liu

References

- Abatzoglou, J. T.: Development of gridded surface meteorological data for ecological applications and modeling, *Int. J. Climatol.*, 33, 121–131, doi:10.1002/joc.3413, 2013.
- Ashfaq, M., Bowling, L. C., Cherkauer, K., Pal, J. S., and Diffenbaugh, N. S.: Influence of climate model biases and daily-scale temperature and precipitation events on hydrological impacts assessment: A case study of the United States, *J. Geophys. Res.*, 115, D14116, doi:10.1029/2009JD012965, 2010.
- Ashfaq, M., Ghosh, S., Kao, S.-C., Bowling, L., Mote, P., Touma, D., Rauscher, S., and Diffenbaugh, N.: Near-term acceleration of hydroclimatic change in the western U.S., *J. Geophys. Res.*, 118, 10676–10693, doi:10.1002/jgrd.50816, 2013.
- Bohn, T. J., Livneh, B., Oyster, J. W., Running, S. W., Nijssen, B., and Lettenmaier, D. P.: Global evaluation of MT-CLIM and related algorithms for forcing of ecological and hydrological models, *Agr. Forest. Meteorol.*, 176, 38–49, doi:10.1016/j.agrformet.2013.03.003, 2013.

- Bowling, L. C., Pomeroy, J. W., and Lettenmaier, D. P.: Parameterization of blowing-snow sublimation in a macroscale hydrology model, *J. Hydrometeorol.*, 5, 745–762, doi:10.1175/1525-7541(2004)005<0745:POBSIA>2.0.CO;2, 2004.
- Brakebill, J. W., Wolock, D. M., and Terziotti, S. E.: Digital hydrologic networks supporting applications related to spatially referenced regression modeling, *J. Am. Water Resour. Assoc.*, 47, 916–932, doi:10.1111/j.1752-1688.2011.00578.x, 2011.
- Cherkauer, K. A. and Lettenmaier, D. P.: Simulation of spatial variability in snow and frozen soil, *J. Geophys. Res.*, 108, 8858, doi:10.1029/2003JD003575, 2003.
- Cherkauer, K. A., Bowling, L. C., and Lettenmaier, D. P.: Variable Infiltration Capacity (VIC) cold land process model updates, *Global Planet. Change*, 38, 151–159, doi:10.1016/S0921-8181(03)00025-0, 2002.
- Chien, H., Yeh, P. J., and Knouft, J. H.: Modeling the potential impacts of climate change on streamflow in agricultural watersheds of the Midwestern United States, *J. Hydrol.*, 491, 73–88, 2013.
- Christensen, N. S., Wood, A. W., Voisin, N., Lettenmaier, D. P., and Palmer, R. N.: The effects of climate change on the hydrology and water resources of the Colorado river basin, *Climatic Change*, 62, 337–363, doi:10.1023/B:CLIM.0000013684.13621.1f, 2004.
- Daly, C., Gibson, W. P., Taylor, G. H., Johnson, G. L., and Pasteris, P.: A knowledge-based approach to the statistical mapping of climate, *Climate Res.*, 22, 99–113, doi:10.3354/cr022099, 2002.
- Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., and Pasteris, P. P.: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States, *Int. J. Climatol.*, 28, 2031–2064, 2008.
- Demaria, E. M., Nijssen, B., and Wagener, T.: Monte Carlo sensitivity analysis of land surface parameters using the Variable Infiltration Capacity model, *J. Geophys. Res.*, 112, D11113, doi:10.1029/2006JD007534, 2007.
- Dooge, J. C. I. and Bruen, M.: Scaling effects on moisture fluxes on unvegetated land surfaces, *Water Resour. Res.*, 33, 2923–2927, doi:10.1029/97WR01709, 1997.
- EPA – Environmental Protection Agency – and USGS – US Geological Survey: NHDPlus user guide, Environmental Protection Agency, available at: ftp://ftp.horizonsystems.com/NHDPlus/NHDPlusV1/documentation/NHDPLUSV1_UserGuide.pdf (last access: July 2013), 2010.
- Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., and Alsdorf, D.: The Shuttle Radar Topography mission, *Rev. Geophys.*, 45, RG2004, doi:10.1029/2005RG000183, 2007.
- Gao, H., Tang, Q., Ferguson, C. R., Wood, E. F., and Lettenmaier, D. P.: Estimating the water budget of major US river basins via remote sensing, *Int. J. Remote Sens.*, 31, 3955–3978, doi:10.1080/01431161.2010.483488, 2010.
- Gao, X., Xu, Y., Zhao, Z. Pal, J. S., and Giorgi, F.: On the role of resolution and topography in the simulation of East Asia precipitation, *Theor. Appl. Climatol.*, 86, 173–185, 2006.
- Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M., and Tyler, D.: The national elevation dataset, *Photogramm. Eng. Remote Sens.*, 68, 5–11, 2002.
- Govindaraju, R. S. and Rao, A. R.: Artificial neural networks: A passing fad in hydrology?, *J. Hydrol. Eng.*, 5, 225–226, doi:10.1061/(ASCE)1084-0699(2000)5:3(225), 2000.
- Haddelenad, I., Matheussen, B. V., and Lettenmaier, D. P.: Influence of spatial resolution on simulated streamflow in a macroscale hydrologic model, *Water Resour. Res.*, 38, 1124, doi:10.1029/2001WR000854, 2002.
- Hamlet, A. F. and Lettenmaier, D. P.: Effects of climate change on hydrology and water resources in the Columbia River basin, *J. Am. Water Resour. Assoc.*, 35, 1597–1623, 1999.
- Hansen, M., DeFries, R., Townshend, J. R. G., and Sohlberg, R.: Global land cover classification at 1 km resolution using a decision tree classifier, *Int. J. Remote Sens.*, 21, 1331–1365, doi:10.1080/014311698214235, 2000.
- Jetten, V., Govers, G., and Hessel, R.: Erosion models: quality of spatial predictions, *Hydrol. Process.*, 17, 887–900, 2003.
- Keane, R. E., Holsinger, L. M., Parsons, R. A., and Gray, K.: Climate change effects on historical range and variability of two large landscapes in western Montana, USA, *Forest Ecol. Manage.*, 254, 375–389, doi:10.1016/j.foreco.2007.08.013, 2008.
- Kimball, J. S., Running, S. W., and Nemani, R.: An improved method for estimating surface humidity from daily minimum temperature, *Agr. Forest Meteorol.*, 85, 87–98, doi:10.1016/S0168-1923(96)02366-0, 1997.
- Liang, X., Lettenmaier, D. P., Burges, S. J., and Wood, E. F.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, *J. Geophys. Res.*, 99, 14415–14428, doi:10.1029/94JD00483, 1994.
- Liang, X., Lettenmaier, D. P., and Wood, E. F.: Surface soil moisture parameterization of the VIC-2L model: evaluation and modification, *Global Planet. Change*, 13, 195–206, doi:10.1016/0921-8181(95)00046-1, 1996.
- Livneh, B., and Lettenmaier, D. P.: Regional parameter estimation for the unified land model, *Water Resour. Res.*, 49, 110–114, doi:10.1029/2012WR012220, 2013.
- Lohmann, D., Raschke, E., Nijssen, B., and Lettenmaier, D. P.: Regional scale hydrology, Part 1: Formulation of the VIC-2L model coupled to a routing model, *Hydrolog. Sci. J.*, 43, 131–141, doi:10.1080/02626669809492107, 1998.
- Maidment, D. R. (Ed.): *Handbook of Hydrology*, McGraw-Hill, New York, NY, 1993.
- Manter, D. K., Reeser, P. W., and Stone, J. K.: A climate-based model for predicting geographic variation in Swiss needle cast severity in the Oregon coast range, *Phytopathology*, 95, 1256–1265, doi:10.1094/PHYTO-95-1256, 2005.
- Maurer, E. P., Wood, A. W., Adam, J. C., and Lettenmaier, D. P.: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States, *J. Climate*, 15, 3237–3251, doi:10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2, 2002.
- McCabe, G. J. and Hay, L. E.: Hydrologic effects of hypothetical climate change in the East River basin, Colorado, *J. Hydrol. Sci.*, 40, 303–318, 1995.
- Merz, B. and Plate, E. J.: An analysis of the effects of spatial variability of soil and soil moisture on runoff, *Water Resour. Res.*, 33, 2909–2922, doi:10.1029/97WR02204, 1997.

- Mesinger, F., DiMegoand, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., Jovic, D., Woollen, J., Rogers, E., Berbery, E. H., Ek, M. B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., and Shi, W.: North American regional reanalysis, *B. Am. Meteorol. Soc.*, 87, 343–360, doi:10.1175/BAMS-87-3-343, 2006.
- Miller, D. A. and White, R. A.: A conterminous United States multilayer soil characteristics dataset for regional climate and hydrology modeling, *Earth Interact.*, 2, 1–26, doi:10.1175/1087-3562(1998)002<0001:ACUSMS>2.3.CO;2, 1998.
- Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., and Stouffer, R. J.: Climate change: stationarity is dead: Whither water management?, *Science*, 319, 573–574, doi:10.1126/science.1151915, 2008.
- Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Cosgrove, B. A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D., Lettenmaier, D. P., Marshall, C. H., Entin, J. K., Pan, M., Shi, W., Koren, V., Meng, J., Ramsay, B. H., and Bailey, A. A.: The multi-institution North American Land Data Assimilation System (NLDAS): utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, *J. Geophys. Res.*, 109, D07S90, doi:10.1029/2003JD003823, 2004.
- Mote, P. W., Hamlet, A. F., Clark, M. P., and Lettenmaier, D. P.: Declining mountain snowpack in western North America, *B. Am. Meteorol. Soc.*, 86, 39–49, doi:10.1175/BAMS-86-1-39, 2005.
- Nijssen, B., Lettenmaier, D. P., Liang, X., Wetzel, S. W., and Wood, E. F.: Streamflow simulation for continental-scale river basins, *Water Resour. Res.*, 33, 711–724, doi:10.1029/96WR03517, 1997.
- Nijssen, B., O'Donnell, G. M., Lettenmaier, D. P., Lohmann, D., and Wood, E. F.: Predicting the discharge of global rivers, *J. Climate*, 14, 3307–3323, doi:10.1175/1520-0442(2001)014<3307:PTDOGR>2.0.CO;2, 2001.
- Payne, J. T., Wood, A. W., Hamlet, A. F., Palmer, R. N., and Lettenmaier, D. P.: Mitigating the effects of climate change on the water resources of the Columbia River Basin, *Climatic Change*, 62, 233–256, 2004.
- Royer, A. and Poirier, S.: Surface temperature spatial and temporal variations in North America from homogenized satellite SMMR-SSM/I microwave measurements and reanalysis for 1979–2008, *J. Geophys. Res.*, 115, D08110, doi:10.1029/2009JD012760, 2010.
- Sale, M. J., Kao, S.-C., Ashfaq, M., Kaiser, D. P., Martinez, R., Webb, C., and Wei, Y.: Assessment of the Effects of Climate Change on Federal Hydropower, Technical Manual 2011/251, Oak Ridge National Laboratory, Oak Ridge, TN, 2012.
- Schwarz, G. E. and Alexander, R. B.: State Soil Geographic (STATSGO) Data Base for the Conterminous United States, US Geological Survey, Reston, VA, 1995.
- Seaber, P. R., Kapinos, F. P., and Knapp, G. L.: Hydrologic Unit Maps, US Geological Survey water supply paper 2294, US Geological Survey, Denver, CO, 1987.
- Shah, S., O'Connell, P., and Hosking, J.: Modelling the effects of spatial variability in rainfall on catchment response, 1. Formulation and calibration of a stochastic rainfall field model, *J. Hydrol.*, 175, 67–88, 1996.
- Sharif, H. O., Crow, W., Miller, N. L., and Wood, E. F.: Multidecadal High-Resolution Hydrologic Modeling of the Arkansas-Red River Basin, *J. Hydrometeorol.*, 8, 1111–1127, doi:10.1175/JHM622.1, 2007.
- Shrestha, R., Tachikawa, Y., and Takara, K.: Input data resolution analysis for distributed hydrological modeling, *J. Hydrol.*, 319, 36–50, doi:10.1016/j.jhydrol.2005.04.025, 2006.
- Su, F., Adam, J. C., Bowling, L. C., and Lettenmaier, D. P.: Streamflow simulations of the terrestrial Arctic domain, *J. Geophys. Res.*, 110, D08112, doi:10.1029/2004JD005518, 2005.
- Thornton, P. E. and Running, S. W.: An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation, *Agr. Forest Meteorol.*, 93, 211–228, doi:10.1016/S0168-1923(98)00126-9, 1999.
- Thornton, P. E., Running, S. W., and White, M. A.: Generating surfaces of daily meteorology variables over large regions of complex terrain, *J. Hydrol.*, 190, 214–251, 1997.
- Wang, S., Zhang, Z., Sun, G., Strauss, P., Guo, J., Tang, Y., and Yao, A.: Multi-site calibration, validation, and sensitivity analysis of the MIKE SHE Model for a large watershed in northern China, *Hydrol. Earth Syst. Sci.*, 16, 4621–4632, doi:10.5194/hess-16-4621-2012, 2012.
- White, M. A., Diffenbaugh, N. S., Jones, G. V., Pal, J. S., and Giorgi, F.: Extreme heat reduces and shifts United States premium wine production in the 21st century, *P. Natl. Acad. Sci. USA*, 103, 11217–11222, doi:10.1073/pnas.0603230103, 2006.
- Wolock, D. M. and McCabe, G. J.: Explaining spatial variability in mean annual runoff in the conterminous United States, *Climate Res.*, 11, 149–159, 1999.
- Wolock, D. M. and Price, C. V.: Effects of digital elevation model map scale and data resolution on a topography-based watershed model, *Water Resour. Res.*, 30, 3041–3052, 1994.
- Zhang, X., Srinivasan, R., and Van Liew, M.: Multi-site calibration of the SWAT model for hydrologic modeling, *T. ASABE*, 51, 2039–2049, 2008.