Response to Reviewer # 2
Review of hess-2013-294 "A Large-Scale, High-Resolution Hydrological Model Parameter Dataset for Climate Change Impact Assessment for the Conterminous United States"

General comments
This paper outlines the approach and methods used to develop a new set of historical hydrologic simulations for the conterminous U.S. at a high spatial resolution (HUC8 watersheds, using 1/24th degree data as a basis). This is a methodological paper, in which the data sources, calibration approach, and initial sensitivity testing are described – the stated intention being to provide a basis on which further improvements can be applied. The authors describe the considerable computational demand of such an undertaking, and note the substantial resources committed to the work at ORNL.

We would like to thank the anonymous reviewer for his/her comments and constructive criticism, which we believe have led to an improved manuscript. Below are specific answers to each review comment.

Overall, the effort has potential to be useful to the research community if the model, datasets, and future updates are made public.

As stated in the manuscript (P. 9595, lines 11-12), the meteorological and hydrological datasets that resulted from this work will be made available to the public. This statement is now emphasized in the abstract.

However, I have four chief concerns:
(1) Although potentially useful in vague terms, it is not clear exactly how this dataset (or the results) could be used to advance research, either in hydrology or climate impacts

It is well documented in previous studies 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). Additionally, high-resolution land hydrology is needed to address questions such as identifying climatic controls on the spatial variability of hydrologic parameters and the scale at which they are most dominant; examining the spatial scaling properties of the hydrologic parameters; performing statistical analysis between climatic variables and land surface processes and states, such as soil moisture and evapo-transpiration (ET); and developing subgrid parameterization approaches for the hydrologic parameters.

Another motivation of this study is, however, not simply to reconstruct the past observations at higher resolution, but rather to be able to evaluate the effects of long-term changes in extreme hydrologic events, for which an evaluation can only be conducted through refined spatial resolution. Moreover, it is necessary to accurately predict spatial variations in watershed hydrology at the subbasin scale in order to identify regionally specific management strategies to mitigate the potential impacts of climate changes on water resources at the regional scale.

(2) This is not the first effort of this kind. Indeed, the authors cite Maurer et al., but do not contrast it with their approach. A more thorough literature review should be included, along with a justification for this newer approach: how is this different from reinventing the wheel?

We agreed that this is not the first effort of this kind. Our approach is, however, different from previous studies such as Maurer et al., 2002 in several aspects, including:

1) Spatial scale:
   The dataset described in our study refines the spatial resolution from 1/8° to 1/24°. With the evolution of higher-resolution observation (e.g., radar precipitation) and increases in computing capabilities, other input datasets, such as remotely sensed land cover, continue to become available at finer spatial resolution. Providing the most up-to-date, high-resolution watershed soil, vegetation, elevation, and other hydrologic characteristics at 1/24° is a significant improvement over the previous studies. See Fig. R1 below for a comparison of simulated annual runoff with observed annual runoff by running the model with 1/24° resolution vs 1/8° resolution. The improvement of scale will lead to an improvement in the accuracy of hydrologic prediction.

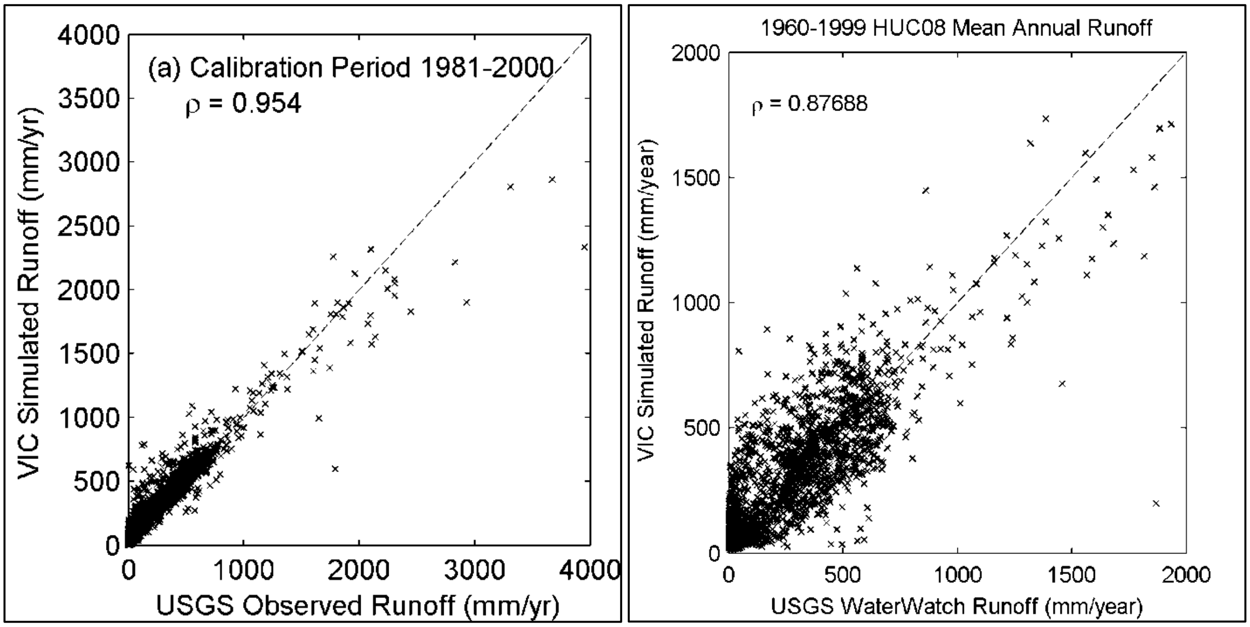


Figure R1: The USGS WaterWatch observed runoff versus the VIC simulated annual total runoff (surface runoff+baseflow) using spatial resolution of (a) 1/24 degree and (b) 1/8 degree.

2) Multi-site (watershed) calibration approach:
   Most of the previous studies focused on the change in the overall water budget rather than the spatial and temporal variations in streamflow variability. These studies calibrated the VIC model at fewer selected locations using the streamflow measurements at a single watershed outlet. To reduce the possibility of apparently accurate simulation at the watershed outlet resulting from the combination of locally inaccurate simulations, we used multi-site (watershed) calibration at the subbasin scale. The last point, which separates our approach from nearly all the previous work, is crucial in the context of water resource management because accurate information about spatial variations in streamflow is critical to identifying regionally specific management strategies in order to mitigate the potential impacts of climate changes on water resources at the regional scale.

Citations to multiple papers have been included in the revised manuscript to address this concern.

(3) Quality of the input datasets is key to obtaining reliable estimates of hydrology – the authors appear to have paid insufficient attention to the quality of the input datasets (e.g., meteorology, soil parameters) – e.g., alternative options, differences in quality/methodology, past validation studies, etc.

We do not agree with this comment. Please note that most of our parameters were assimilated from high-resolution and long-term coverage datasets. These datasets were stored in various different formats, projections, and required a huge effort for data processing, organization, and quality control. Such a data-intensive effort is unprecedented in the previous studies, and hence the new resource will be valuable because it will prompt future hydro-climate studies. Several of our major activities included the following:

1) For meteorological forcings, we made an effort to collect, organize, and compare various daily-resolution datasets. Particularly, DAYMET is available at a very high (1km) spatial resolution. The organization of a complete 1980-2008 DAYMET daily record for the entire conterminous US is not a trivial task. The applicability of DAYMET for a hydrologic simulation of the conterminous US has never been shown.

2) We also made an effort to collect the wind speed data from NARR, a dynamically downscale meteorological dataset from the coarser-resolution reanalysis data (over 1 degree to 36km). As shown in the manuscript, the more detailed local wind speed pattern can be better simulated.

3) For vegetation, we made a huge effort to process and organize 2003-2008 8 day spacing MODIS LAI at 1km resolution for the entire conterminous US. We are not aware of such application in hydrologic simulation for the entire US.

4) For elevation, we collected the information from both 10m resolution NED and 90m resolution SRTM (for non-US regions). It was not a simple spatial aggregation. We calculated a spatial histogram and used it to develop elevation bands.

5) While we invested relatively less time on collecting new soil parameters, recognizing the high measuring uncertainty for soil parameters, we utilized high-performance computing to calibrate main soil parameters to improve the overall modeling performance. To the best of our knowledge, such an intensive data calibration effort is novel in hydro-climate studies.

6) For runoff, we calibrated each HUC8 independently through the use of the WaterWatch dataset. This is a new strategy for large-scale hydrologic model applications.

We have included multiple new references and some new analysis in the revised manuscript (as suggested by the reviewers); however, we would like to emphasize that our intention is to organize spatially consistent forcing and parameter datasets for the conterminous US. Therefore, the spatial coverage is one major consideration. Although there are some alternative, higher-resolution and higher-quality data (e.g., lidar elevation, SURRGO dataset), they are unavailable uniformly across the entire US and hence were not selected in this study.

(4) Calibration on such small scales may be problematic, since it may result in simulations that "get the right answer for the wrong reason" and therefore have incorrect sensitivities. I'm not convinced that calibration at the HUC8 level – in particular given the approximations of WaterWatch – is the right approach.

We believe that calibrating the model at multiple locations at a small scale may improve the streamflow prediction as recommended by other studies (e.g., Chien et al., 2013; Zhang et al., 2008; Wang et al., 2012; Jetten et al., 2003). For example, Jetten et al. (2003) concluded that predicting the hydrologic response at the 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 the single watershed outlet from the combination of locally inaccurate simulations.

In addition, it should be noted that many of the unregulated streamflow gauges that are used for hydrological model calibration actually have drainage areas less than HUC8 Subbasins. Therefore, the HUC8-based WaterWatch observation should not be considered small scale. We have shown that the results from WaterWatch-based calibration are not very different from the gauge-based calibration. The WaterWatch approach actually provides a new opportunity because the data are available for each HUC8 in a consistent format so that some gauge data processing time can be saved.

Specific comments:

p. 9578, line 14: it is false to say that spatial resolution is not important for climate modeling – it may be less critical than for hydrology, but typical GCM resolutions are far from adequate at the scale of HUC8 watersheds.

The sentence has been revised for clarity.

p. 9579, line 15: this is a very vague statement. Be more concrete – e.g., median and range of NSE, etc.

The sentence has been revised for clarity.

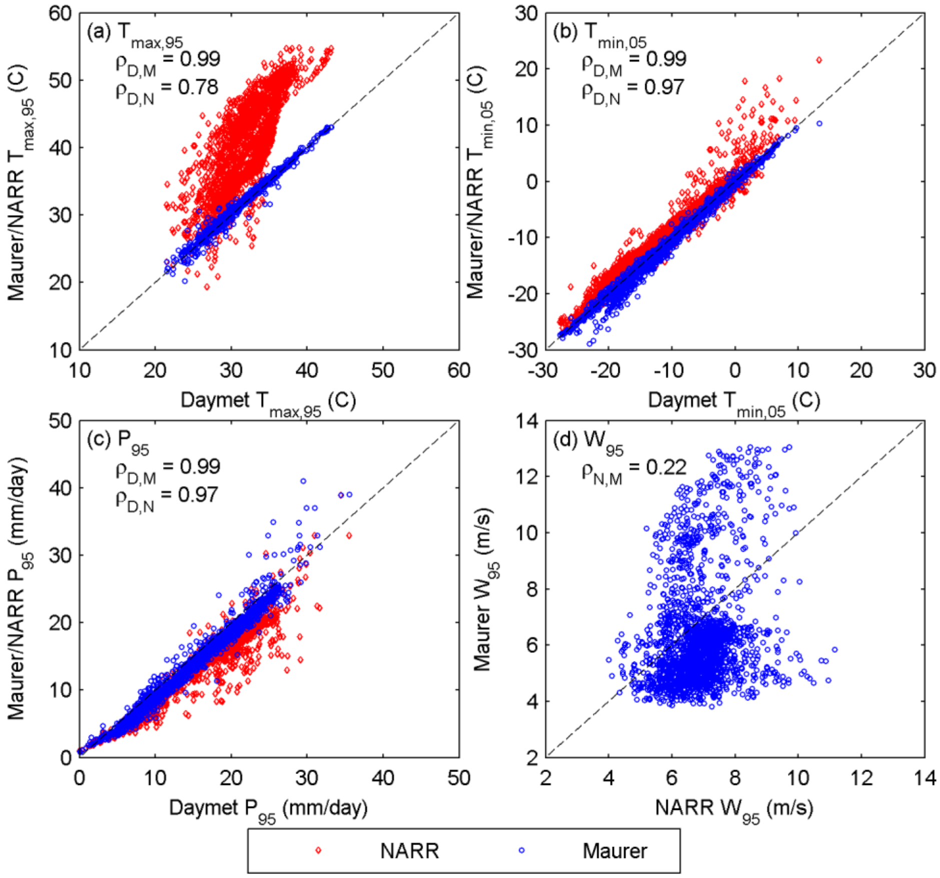
p. 9580, line 11: the 1/24th degree grids should require exactly 9 times the computational resources needed for the 1/8th degree grid, not "more than 10 times".

This confusion was brought up by multiple reviewers, and it has been clarified in the revised manuscript. We agreed that the 1/24° grids would result in exactly 9 times of the computational resources when compared with the 1/8° grids. However, given the expanded data flow, there will be some additional computational demand for data management and quality control, which were not required for the simpler 1/8° grids.

Section 2.2: All of the met datasets listed have major limitations, and ALL (including DAYMET) involve some sort of interpolation from sparse station observations. It is certainly not sufficient to simply interpolate Maurer and NARR, since at minimum some crude correction should be made for differences in elevation. In addition, some degree of validation is certainly possible and necessary to include in a study such as this one – at minimum a discussion of the validation presented in the main reference for each dataset.

We appreciate this comment. We have included several references to describe the limitations of the selected forcings datasets. For instance, at the regional scale, Daly et al. (2008) showed a cold bias in the DAYMET dataset when compared to the PRISM January minimum temperature, while differences between the PRISM and DAYMET July maximum temperatures are relatively small. For precipitation, DAYMET tended to be drier than PRISM in some locations and wetter in others because DAYMET did not resolve rain shadows well owing to an inability to recognized topographic facets (Daly et al., 2008). This study, however, showed that differences in precipitation, t_{min} and t_{max}, are more prominent in the western US than in the eastern regions. This discussion has been added in the revised manuscript.

To our knowledge, elevation and topography have been considered for the creation of DAYMET, PRISM, and Maurer when interpolating meteorological observations from surface weather stations. Since we do not intend to develop our own weather interpolation, we were seeking a most justifiable way to come up with the meteorological forcing at 4km resolution using preexisting datasets. For this purpose, DAYMET is the most appropriate choice since it was already in finer (1km) resolution. As evidenced in the new analysis (Fig. 4 in the revised manuscript), the 95% quantile of precipitation P_{95} is consistently higher in DAYMET than in Maurer and NARR, suggesting that higher spatial resolution is necessary to capture the precipitation extremes. The direct interpolation of Maurer and NARR, while sufficient for the purpose of comparison, may not provide the best results at a resolution of 4km.



New Figure 4 in the revised manuscript. 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.

I'm concerned by the fact that the authors seem to have applied very little rigor to the choice of meteorological dataset – this is likely a dominant source of error in hydrologic simulations – a problem that calibration rarely addresses, instead ensuring that you get the "right answer for the wrong reason".

As stated earlier, DAYMET was chosen as the default meteorological dataset since it was already in 1km resolution and did not require interpolation (neither direct nor topography-based interpolation). In the revised manuscript, Daly et al. (2008) has been included to describe the limitations of DAYMET. The new result (Fig. 4) on high/low quantiles suggested that DAYMET may better capture the precipitation extremes, which is an important contribution that cannot be shown from coarser-resolution datasets.

*In addition to the monthly analysis, some analysis should be included regarding differences among *daily data from DAYMET, NARR, and Maurer – these comparisons would likely show much larger discrepancies. Furthermore, a review of previous validation efforts should be included – surely all of these datasets have been subject to scrutiny elsewhere in the literature.*

We appreciate this suggestion. To address this concern, a new comparison on high/low daily quantiles (Fig. 4) is now included. Particularly, the new results show that the precipitation extremes are consistently higher in DAYMET than in Maurer and NARR. These results are consistent with our understanding that precipitation is highly resolution-dependent (Gao et al., 2006), and hence a finer-resolution dataset is required to better capture and simulate the extreme events.

Section 2.3: Why not use the new dataset of Livneh et al., 2013?

We agree that the Livneh et al. (2013) dataset is also available at fine spatial resolution (1/16 degree, ~ 6km). However, this recent dataset was not available publicly when we started our study. Since we have utilized high-performance computing to calibrate the key soil parameters for better model performance, we think our soil parameter set may serve a good modeling purpose as well.

Section 2.3: Given the approximations used to estimate HUC8 flows via WaterWatch, the discussion should include some information about how well this method performs and when/where it is most at risk of exhibiting errors. For instance, the authors hint at the problem of regulated gauge stations – this has potential to result in large errors in places with small reservoirs, diversions, etc. Similarly, there are other concerns – basins far from gauge observations, adjacent basins with very different exposures to precipitation, etc. Such questions should at least be acknowledged in the text – ideally, results of validation tests can be included.

We appreciate this comment. A new paragraph has been added in Section 2.6 to describe the limitation of WaterWatch. We note that future studies on the accuracy of WaterWatch runoff would be highly desired.

p. 9586, line 9: the 3 hourly time-step in VIC is a user-defined setting. I believe the default is for MTCLIM to run on a 1-hourly time step.

We appreciate this comment. It has been corrected in the revised manuscript.

Section 3.1, lines 3-7: Use of the term "daily" for each term is misleading since the comparison involves monthly means. This should be clarified to avoid confusion – both in the text and in the caption to Figure 3.

The use of "daily" (for temperature and wind speed) is accurate here. For temperature, while PRISM is available at a monthly resolution, the maximum and minimum temperatures actually refer to "Daily maximum/minimum temperature [averaged over all days in the month]" (http://www.prism.oregonstate.edu/ documents/PRISM_datasets_aug2013.pdf). For wind speed, we indeed computed and compared the mean daily wind speed from NARR and Maurer datasets. No change was made.

Figure 4: the mean daily wind speed from NARR appears to be quite noisy – e.g., no clear delineation of the Rockies, Cascades, or Sierras, where I would expect to see large increases in wind speeds. Could this be due to an artifact of the gridding process? Have NARR winds been validated elsewhere?

To address this concern, we used a web-based tool provided by the NOAA Earth System Research Laboratory (http://www.esrl.noaa.gov/psd/cgi-bin/data/narr/plotmonth.pl) to prepare independent wind speed plots for verification. The tool reads and plots NARR and NCEP reanalysis data on the server side, so the results are not affected by the gridding process as utilized in this study. Given the tool's limitations, we plotted 1980–2000 instead of 1980–2008 results in Fig. R2. However, the difference is really not differentiable. For NARR wind speed, the results from both approaches are similar, indicating that the spotty pattern was not a result of the gridding process. As expected, Maurer wind speed is also similar to the NCEP wind speed, with the spatial swift as noted on Maurer's website. Given that NARR is the downscaled product of NCEP, the fine-scale features should be better characterized than the NCEP features and would be a better choice for finer-scale hydrologic simulation.

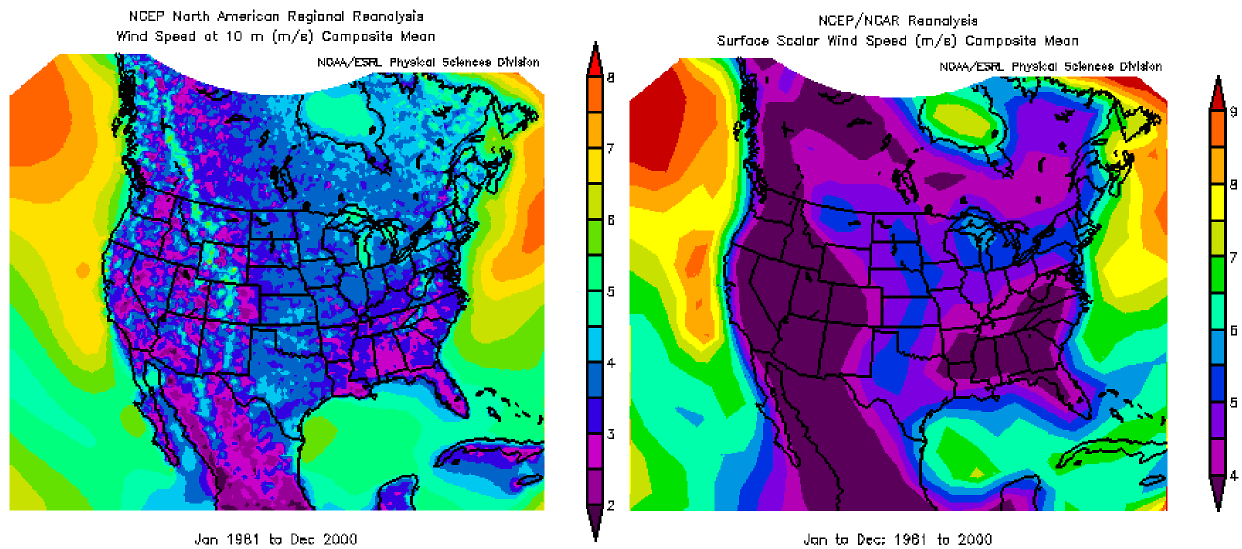
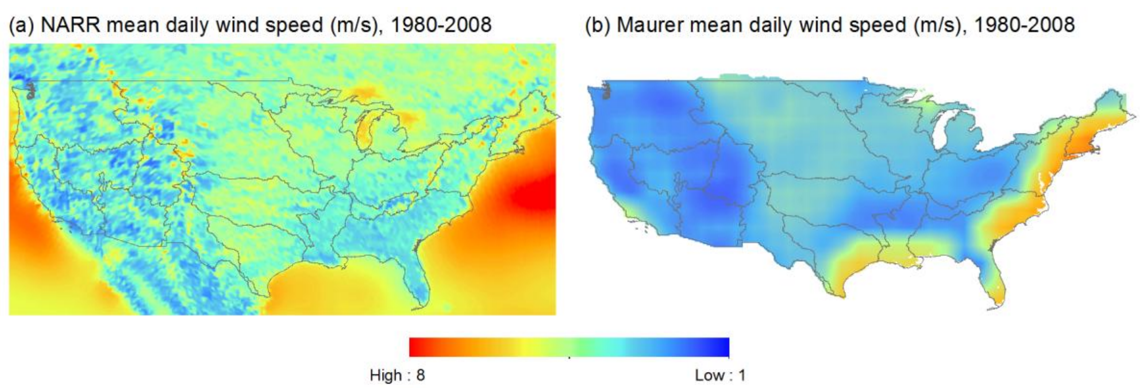


Figure R2: Maps of 1980–2000 mean daily wind speed (by http://www.esrl.noaa.gov/psd/cgi-bin/data/narr/plotmonth.pl)



New Figure 5 in the revised manuscript: Maps of mean daily wind speed.

p. 9593: What about groundwater? Managed flows? Could these be affecting the erroneous runoff simulations in Fig 8?

We appreciate this comment. These possible reasons have been added in the revised manuscript.

Technical corrections:

p. 9580, line 10: replace "a key" with "key" and "resource" with "resources"

We appreciate this comment. It has been corrected in the revised manuscript.

p. 9580, lines 11-12: replace "of computation resources comparing to" with "the computational resources required for" (There were other such grammatical errors and stylistically awkward phrasings throughout the text – please identify and correct these)

We appreciate this comment. It has been corrected in the revised manuscript.

p. 9580, line 27: weather station observations are not referred to as "gauge" observations – for one thing because the measurements are not made by gauges, but by instruments: temperature, precipitation, and wind sensors, for instance.

We appreciate this comment. It has been corrected in the revised manuscript.

Additional References

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., 2013.

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.

Dooge, J. C. I. and Bruen, M.: Scaling effects on moisture fluxes on unvegetated land surfaces, Water Resour. Res., 33(12), 2923–2927, doi:10.1029/97WR01709, 1997.

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(1–4), 173-185, 2006.

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.

Jetten, V., Govers, G. and Hessel, R.: Erosion models: quality of spatial predictions, Hydrol. Process, 17, 887–900, 2003.

Livneh, B., and Lettenmaier, D. P.: Regional parameter estimation for the unified land model, Water Resour. Res., 49, 110–114, 10.1029/2012WR012220, 2013.

Merz, B. and Plate, E. J.: An analysis of the effects of spatial variability of soil and soil moisture on runoff, Water Res. Resear., 33(12), 2909–2922, doi:10.1029/97WR02204, 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.

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. Hydrometeor, 8, 1111–1127, doi:http://dx.doi.org/10.1175/JHM622.1, 2007.

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, Hydrology Earth Syst. Sc., 16, 4621–4632, 2012.

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, Transactions of the ASABE, 51, 2039–2049, 2008.