Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-216-AC1, 2020

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Implications of Model Selection: A Comparison of Publicly Available, CONUS-Extent Hydrologic Component Estimates" *by* Samuel Saxe et al.

Samuel Saxe et al.

ssaxe@usgs.gov

Received and published: 14 September 2020

Note: The author responses were originally going to be provided as PDFs but upload failed due to technical errors on the Copernicus website. If any reviewers or editors would prefer author responses in PDF or Word format, please contact the lead author via email.

The authors are grateful for the valuable input from the referee on our paper "Implications of Model Selection: A Comparison of Publicly Available, CONUS-Extent Hydrologic Component Estimates". The major comments, especially, highlight suggestions that will enhance the relevance and clarity of our manuscript.

Major comments

Regarding CV (I): With the coefficient of variation (CV) metric, our primary goal is to quantify inter-model variability. Our secondary goal is to quantify temporal variability (inter-annual), though this is performed in only a rudimentary fashion by dividing datasets into "Early" and "Late" time periods. The use of CV is best exemplified in Figure 5. Confusion is likely arising from the early use of CV in Figure 2 where only a single value is presented for each component (e.g. precipitation, evapotranspiration, etc.). In the case of Figure 5, CV was calculated between all models for a single water year for each component for each ecoregion. Using Figure 5/component AET/ecoregion Medit. CA/period Early as an example, the data points used to generate the boxplot are the CV between all models for each water year from 1985-1999. We agree with the referee's suggestion to use the term "uncertainty" in place of CV to avoid confusion with temporal variability, this will be corrected in the revised manuscript. To answer the referee's question "Can they clarify if/how CV values for individual water years were aggregated over 15 years?": Data presented in Figure 5 show the spread of individual years. In the text, CV values are summarized over 15 year periods using the median CV measure.

Regarding CV (II): The referee makes a very valid point regarding CVs inflating the importance of relatively unimportant hydrologic components. We agree that that relating uncertainty relative to precipitation, rather than the underlying mean, would relate the importance of model uncertainty by hydrologic component. However, this is probably only be effective for flux components (precipitation, ET, runoff) and not storage components (SWE, soil moisture).

Regarding sample size: Agreed that uncertainty will increase as sample size increases. Bootstrapping using a constant sample size would provide an additional measure of uncertainty and would help to better inform readers.

Regarding soil moisture: The usage of the term "rootzone" was carried over from the first few models we included in this study (NLDAS and GLDAS land surface models). However, it is at best misleading and at worst incorrect to call the included soil moisture datasets "rootzone" products. We will amend the manuscript to use only "soil moisture". Remotely sensed products (ESA-CCI, SMOS-L4) cover different soil depth ranges, computed in terms of volumetric soil water content. ESA-CCI is an assimilation of data from various remote sensing instruments, while SMOS-L4 uses a double bucket water budget model to extrapolate surface soil moisture (0-5cm) to the "rootzone" domain (5-200cm). Again, though, the use here of the term "rootzone" is inaccurate and we will amend the manuscript accordingly.

Regarding water balance residuals: We used the term "imbalances" rather than "residuals" specifically to inform the reader that excess water, or lack thereof, in our water budget analyses is not a measure of accuracy. Rather, "imbalances" are used to show the reader that the presence and magnitude of gains or losses in a water budget equation are highly uncertain due to the uncertainty in terrestrial hydrology estimates, thereby showing the possible ranges in terrestrial water storage and groundwater storage. We excluded changes in groundwater storage from this study for a few reasons.

(1) Availability of modeled estimates: At the CONUS scale, there are few publicly available datasets to compare and use for estimates of uncertainty, relative to other components (e.g. precipitation, SWE). While we were offered results of a few models, those data were not available to the public, and thus not used in the context we discuss in the introduction. Furthermore, groundwater datasets covered a shorter time range than other datasets in this study and did not cover the conterminous U.S. This spatiotemporal heterogeneity would conflict with the homogeneity of the surface water flux and storage datasets used in this study, reducing the effectiveness of our uncertainty analysis.

(2) GRACE: Groundwater estimates are commonly derived from GRACE terrestrial water storage (TWS) data and provide an effective means of measuring changes in groundwater storage. However, to calculate changes in groundwater from GRACE

TWS requires the subtraction of modeled data estimating soil moisture, snow, and surface water storage. We believe that this is outside the scope of this study as it would constitute a new analysis focusing on the implied uncertainty in groundwater derived from GRACE TWS. As a case in point, we are submitting a manuscript to another journal this week where we do just that: we provide a comprehensive analysis of uncertainty in GRACE-derived groundwater by quantifying how trend magnitude and direction are affected by model selection using ten surface water storage estimates (soil moisture and snow) and five GRACE TWS solutions (3 spherical harmonic and 2 mascon) over a seven year period.

Regarding ecoregions: The EPA ecoregions were used for classification for two reasons:

(1) Common use: the EPA ecoregions are commonly applied across a range of subdisciplines within the hydrological, ecological, and geological scientific communities. They are also "understandable" in the sense that readers intuitively understand the differences between a region called "Eastern Temperate Forests" and "Northwestern Forested Mountains". With this paper, our target audience is not only the modeling community (e.g. land surface modelers, catchment modelers, remote sensing modelers) but also the users of model outputs who apply estimates to local and regional scale analyses who would otherwise be constrained in terms of technical abilities and computational resources from understanding and incorporating measures of uncertainty in their projects. We believe that using EPA ecoregions provides a useful, albeit highly simplified, classification system for intercomparing modeled hydrologic estimates.

(2) Dataset availability: The cited sources of classification schemes use somewhat limited sample sizes of 200-300 catchments (Sawicz et al., 2011; Berghuijs et al., 2014) that are disproportionately, or entirely, weighted towards the eastern CONUS. In all cases, the classification schemes are applied to cluster studied watersheds and not extrapolated over the CONUS, therefore not providing a continuous classification system that we could apply for our datasets.

Perhaps a useful future application of the datasets we collected for this study would be to extend the methods of Berghuijs et al., 2014 over the entirety of the CONUS to provide a multi-level classification system derived from hydrology rather than ecology.

Regarding Figure 7 and correlation of datasets against MOD16 and SSEBop: It is likely that poor correlation between MOD16-A2 and other datasets in the North American Deserts ecoregion is caused by poor model performance of the MOD16-A2 dataset. The MOD16-A2 dataset typically shows annual peak values occurring 1-3 months earlier than all other hydrologic and remote sensing models in the western area of the North American Deserts ecoregion.

In Mediterranean California, attributing poor correlation to either SSEBop or modeled datasets is more difficult. On one hand, SSEBop typically has positive biases in regions with high levels of bare ground fraction, such as sparse shrubland areas common to the Mediterranean California ecoregion, which may affect the timing of seasonal trends. On the other hand, bare ground fraction is even more common in the North American Desert ecoregion and yet correlation between SSEBop and other datasets is typically high. A more likely cause of the poor correlation is that high levels of irrigation in the Central Valley of California may not be properly accounted for in the ET models, whereas the SSEBop model correctly identifies them due to it's use of remotely sensed MODIS data. However, the MOD16-A2 dataset also uses MODIS data but shows better correlation with other datasets, further complicating the discussion. A deeper investigation into input parameters and calibration methods used in the hydrologic models would help to better understand these discrepancies.

Regarding recommendations/suggestions: We strongly agree with the comment that studies accounting for uncertainty in modeled estimates are rare despite the prevalence of literature noting and quantifying these values. Previous Model Intercomparison Projects (MIPs), such as those discussed in the introduction (e.g. CMIP, WaterMIP, AgMIP, etc.), provide interesting formats to provide the scientific community with evaluation and validation measures. However, those MIPs are often either short-lived or

limited in scope and don't provide measures of uncertainty that can be directly applied to individual studies. Our hope with this publication is that users of the study datasets can use our quantified uncertainty values within their own research to, at the very least, inform their subsequent readers of how uncertain their estimates of ET, P, etc. are and how that may impact results and conclusions.

A more systematic approach to providing measures of model uncertainty would be to compile a database in which a variety of modeled datasets (hydrologic, reanalysis, remote sensing models) are aggregated together in tandem with: (a) pre-computed intercomparison statistics, and/or (b) a GUI/app/R package/Python package to allow users to generate statistics based on their selected subsets of models. While collection of datasets is far easier today than 5-10 years ago, data are scattered across numerous databases that each require different access methods, permissions, output formats, coordinate reference systems, and spatiotemporal resolutions. To that end, we are tentatively planning a project to tackle this very issue, assuming that we can overcome the obstacles of data storage, GUI/app hosting costs, and dataset permissions.

In the context of this manuscript, we agree with the referee that the text should include examples of pre-existing sources of model subsets, such as the CAMELS database, to point readers in the direction of useful data. Additionally, all data created for this study (models aggregated by mean area weighting to 10 EPA Ecoregions at the monthly time step) will be made publicly available on a ScienceBase.gov page. Our revised manuscript will update the text to include these details.

Minor Comments

"L193: Currently there seems to be no section 2."

Correct, this is a typo that will be amended in the revision.

"L608-610: What does the range provided for each variable correspond to? I assume there are retrieved from Figure 5. Please also clarify this in the abstract (L23-25)"

The values provided for each variable are the ranges in coefficient of variation (CV) measurements for all ecoregions in either the western or eastern CONUS. For example, we state “Results show that flux and storage magnitudes disagree most greatly in the western CONUS, with CVs of precipitation (P) 11-22%...”. In this case, we are saying that the CV of precipitation datasets for ecoregions in western half of the CONUS range from as low as 11% (found in North American Deserts) and as high as 22% (found in Temperate Sierras). This will be clarified both in the lines referenced here as well as in the abstract.

“Figure 2a: Is the SD for RZSMV really close to 0 (perfect agreement between models)?”

This is the result of a difference in units between the datasets and an error in the vertical axis of the figure. Components P, AET, R, SWE, and RZSME are measured in equivalent water depth as millimeters. RZSMV is measured in volumetric water content fraction as m3/m3 (ranging from 0-1). So, the SD for RZSMV should be on a different vertical scale than the other components. This will be corrected in the manuscript revision.

“Figure 7: maybe clarify whether each circle/square corresponds to one dataset and whether these datasets are sorted based on rho (and hence their order changes from one panel to the next).”

Agreed that this should be clarified in the figure caption. Indeed, each point, whether it be a circle or square, corresponds to a single dataset and are sorted based on rho. We included a more detailed figure in the appendix (A2.10) for interested readers.

“Figure 8: “Histograms of eight ecoregion water budget relative residuals”, maybe clarify that it is relative to the total precipitation (Eq 7).”

This will be corrected in the revision. While this is stated in the manuscript methods, clarification in the figure caption will be useful for readers and skimmers.

“Figure A2.5 to A2.9: I suggest - removing caption mentions to colours not used in the plot (e.g. green in A2.5) - clarifying whether the boxplots are made using significant tau values only - clarifying whether the dots represents significant tau values - re-evaluating whether box plots are well-adapted to show 1 or 2 values”

We agree with all reviewer suggestions, and especially so on his last point. All suggestions will be addressed in the revisions.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-216, 2020.