Hydrology and
Earth System
Sciences

Open Access

Discussions

HESSD

Interactive
comment

Interactive comment on "SWAT Modeling of Water Quantity and Quality in the Tennessee River Basin: Spatiotemporal Calibration and Validation" *by* G. Wang et al.

Anonymous Referee #1

Received and published: 23 February 2016

Short Summary:

The manuscript outlines a framework for modeling water quantity and quality applying the Soil and Water Assessment tool (SWAT) to large river basins in the United States. The authors demonstrate their procedure with a case study in the Tennessee River Basin. To overcome shortcomings that are typical in such studies, for instance limitations in calibration data in terms of temporal and spatial resolution or the type of measured parameters, the authors state three innovations implemented in the study: 1) A parameter optimization framework was set up incorporating Shuffled Complex Evolution (SCE) as an optimization routine for SWAT that can provide intermediate response variables (variables that would not be recorded in the SWAT output files) for calibration.

2) For the calibration and validation as well as for model comparison, solely modeled data were incorporated, since no measured data were available within the considered time frame. 3) Correlating the results of several response variables to each other, and to the spatial properties of the modeled relationships among response variables as well between response variables and spatial attributes of SWAT sub basins.

General Comments:

With their work the authors raise important points eminent to almost every study analyzing in-stream water quality on large scale, such as limited data, particularly when it comes to nutrient fluxes, or identifying major contributors to nutrient loads. The study plots a workflow to incorporate the available datasets in the US in order to answer specific questions with process based modelling. When the approach is successful this might work as a blueprint for further studies on water quality in various US Catchments.

In general, I think the presented workflow of implementing available modelled data for analyzing US watersheds in a process based framework is a novel idea that could benefit further work on such topic. The outlined approach however, lacks some very critical considerations that must be covered before any publication of such framework. The in my opinion most critical points and my resulting suggestion to this manuscript are listed below:

1. My biggest concern is that the authors exclusively use modelled data derived from empirical models in their study! They state in the abstract, that they identified empirical data sets to compare model results to SWAT model responses. Further on in the paper however, the authors clearly state that the SWAT model was calibrated using computed runoff, as well as modelled LOADEST water quality data. These are clearly empirical datasets (which is mention in the manuscript further on and the authors are aware of). But most importantly, these datasets underlie large uncertainties in their responses. Therefore, implementing these

data, without thorough analysis of their quality is not an acceptable approach in my opinion. To prove the outlined approach as valid the calibrated (incorporating these empirical data) must be validated using observed data. The capabilities but also the shortcomings in comparison to observations require a detailed evaluation and discussion. The authors state that there is no observed data available for the model simulation period that can be used. This raises the question, which data was incorporated in the regression applying the LOADEST model in order to predict the nutrient loads for the respective time period. In a different paragraph the authors state that the major advantage of process based models over statistical models is the avoidance of extrapolation of statistical relationships. These statements contradict each other, when thinking of using extrapolated data from a statistical model to calibrate a process based model.

2. I do not agree with the authors statement, that the available automated calibration tools available for SWAT are incapable of including intermediate variables (as far as I understood this term right, these are variables comprised of other SWAT output variables, such as NO3 + NO2) in the calibration. When working with SWAT-CUP one can derive the product or the sum (etc.) of any variable that is evaluated in the subsequent post processing step. Therefore, less emphasis should be placed on this novelty in their study.

As the authors stated, these routines also offer SCE among other routines as sampling routine. Therefore, I do not agree with the authors that their routine is superior to others. I do however see the potential of this routine within the outlined workflow of specifically incorporating USGS data.

3. Apart from discharge, the performance of the SWAT model in its present state is low and the performance ratings are unsatisfactory for sediment, N, and P, especially when referring to Moriasi et al. (2007). As the authors used responses of a less complex model as calibration data for the more complex SWAT model, I question why SWAT is incapable of reproducing these results at all? The authors

state that SWAT is able to reconstruct the inter-annual repetitive patterns, but magnitudes of the peaks and periods with low nutrient loads are missed at all. Before publishing the model performance must improve. Any further model comparison of a poorly performing model to other models (in this case SPARROW) is in my opinion invalid.

4. Analyzing functional relationships of nutrient responses to any basin properties and their changes is in my opinion the actual strength of SWAT and therefore an interesting analysis to conduct. Many of the stated correlations however, are simply given by their functional relationship in the SWAT model and therefore obsolete to mention. The most surprising statement here was the very low correlation of sediment and P to runoff. These variables are usually highly correlated. Such findings require detailed analysis and discussion. The authors should not use SWAT as a black box, but should discuss the equations in the model in relationship to the simulated outputs. I advise the authors to restrict the thresholds of the correlation coefficient in their analysis. A correlation of 0.2 is not moderate in my opinion and gives the reader a wrong impression of the results. Taking the absolute value of the correlation coefficient leads to a loss of information.

Based on the given general statements I suggest **rejecting the paper** in its present form. I see however the potential of incorporating empirical data into a workflow of setting up SWAT models for US basins on HUC8 scale. Nevertheless, the above stated points must be conducted in order to prove this approach as valid.

Specific Comments:

Figures:

The figures do not fully express the results given in the manuscript, but rather provide a skewed view on them:

Figure 3 only gives the results for runoff, but omits demonstrating any results for sed-

iment and nutrients. This gives a wrong impression of good results in general. The loads of N and P estimated using SPARROW are given as well on HUC8 scale. The authors only plotted the SWAT results for mean annual loads with spatial reference in figure 4. The comparison of the SWAT outputs to SPARROW are done in lumped form using box plots. A spatial comparison would be of much more interest, revealing which areas contribute which amounts of nutrient loads when using the two different models.

The results in **figure 5** require statistical testing. To me, it seems that there is a possibility that these two distributions are significantly different. In the text the authors say that the results are comparable, statistical testing can support or contradict this statement.

The abbreviation *"Obs"* in **figures S2** and **S4** is misleading. It is explained in the text how the term observation is used here, but only looking at the graphs the impression that the authors have used observed data is falsely conveyed.

Paragraphs

L 31: The term "intermediate response variable" requires explanation much earlier in the manuscript as it is not fully self-explaining but is a key term in this work.

L 66 – 69: I do not fully understand this statement.

L78 – 80: This sentence gives the impression that smoothing out information on nutrient loads is a benefit of empirical models. When using them for calibration of a process based model this might be a serious issue.

L85 – 89: As stated above I do not fully agree to this statement.

L91: The authors mention here that one challenge is that the watershed is highly regulated due to many reservoirs. This issue however, is not covered in the paper at all.

L140 – 141: What are the elevation range and topography in the catchment? How do the authors argue these thresholds for the slope classification?

L141 – 144: How did the authors assign or distribute the specific land uses in the

catchment?

L144: Corn might be one of the main contributors of P and N loads in the catchment, even if their fraction in the catchment is that low. What was the fraction of corn (or soybean) after applying a threshold of 2

L147 – 148: I would not mix up the terms estimations and observations and use them for the very same data.

L151: Which input was used for the weather generator?

L156: If the SWAT outlets are set at the same position as the location of HUC8 runoff data, then runoff is no intermediate variable.

L200: When such long time series are available I recommend to use a larger fraction of it as warm up period. One year of warm up is very short.

L209 – 210: Why is it necessary to accumulate these runoffs? If the sub basins in the model are assigned properly then the runoff or water yield is explicitly given by the SWAT outputs.

L218 – 222: This statement still surprises me. There are data from 6000 Stations but none of these are adequate for use in validation?

L244 – 246: As stated above the defined thresholds are too low. Taking the absolute refuses one to identify positive or negative relationships.

L272 – 277: For calibration the total range of the NSE values was given. For validation only the CI is shown. It is logical that some sub basins perform much worse during validation. Refusing to show this information gives however a wrong impression.

L294: I would not call it a great improvement when a model performs just as good as the mean value.

L321 -325: Why are SPARROW TP estimates highly correlated to SWAT TN estimates?

These originate usually from different processes. This finding requires further analysis and discussion.

References:

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Binger, R. L., Harmel, R. D., Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE, 50(3), 885–900. http://doi.org/10.13031/2013.23153

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-34, 2016.

Printer-friendly version

Discussion paper