Response to interactive comment on "The effect of input data complexity on the uncertainty in simulated streamflow in a humid, mountainous watershed"

Linh Hoang, Rajith Mukundan, Karen E. B. Moore[2], Emmet M. Owens and Tammo S. Steenhuis

Point by point response (indicated by →) to reviewers' comments

The italic text in quotation is quoted from the original manuscript.

W. Vervoort (Referee)

willem.vervoort@sydney.edu.au

Comment

This manuscript describes how an earlier developed version of SWAT with a wetness index spatial layer as an additional input, is tested for different input layer uncertainties. In particular, the paper focuses on the scale of the DEM and scale of the land use and soil layers. In principle, this is an interesting concept, however, I believe that, at the moment, the authors don't do a good job identifying and describing what the important (of interest to a global audience of HESS) findings are. As a result, I struggled to understand why this paper should be published in its current form.

Response

→ While earlier studies have looked at the effect of differences in scale of DEMs, land use, and soil layers on SWAT model prediction (streamflow, nutrients etc.), the focus of our study is on the uncertainty in parameter values and output uncertainty (streamflow and saturated areas in this study) due to differences in input layers. We found that increasing input data complexity/resolution does not help to reduce parameter uncertainty and the uncertainty in model predictions. However, using multiple types of observed datasets such as spatial observations in addition to the conventional temporal observations (flow at the outlet) can eliminate a high number of unsuitable parameter sets and guide selection of the appropriate parameter sets that give good temporal and spatial predictions for streamflow and saturated areas. These results are important and has wider applicability when identifying critical runoff generating areas and locations within the watershed where management interventions for water quality improvement (e.g.

Phosphorus loading) are most effective. Our results are applicable to regions with similar land use, topography, and climate that are dominated by saturation-excess runoff. To apply to other region, similar/independent work is needed using the methodology described in this paper.

Although we think that our study provides a good scientific contribution, we agree with the reviewer that we did not give a proper explanation in the two major issues that the reviewer mentioned. We tried our best to respond to all comments you have as below, and will base on those responses to revise our manuscript.

Comment

There are to me two major issues that need to be expanded on in more detail and can actually be the bit that makes this paper acceptable:

1. What is the influence of adding the wetness index layer to the input mix, and how does this interact with the other layers? I checked Hoang et al. (2017), which is actually twice in your references, and I did not see any analysis of this either. It mainly covers the improved predictions, and this is an important addition. However, thinking about the input layers, there has to be some sort of interaction between the slope, soil and wetness index, and this is not really explored. Note that I am not saying the algorithm is not valuable, it clearly is, and however, what remains unexplored is how this addition interacts with the existing components. For example you would expect that at some of the DEM and soil map resolution, the soil slope interaction would be similar to your wetness index layer. Of course without the underlying lateral flow and surface aquifer algorithm, this would be useless information. What is valuable in your current research is that you seem to discover this in your results. All your results, I think, point to the fact that the wetness index layer dominates the actual flow behavior, but I am not sure if this is specific to your water shed. For example, the latb parameter is a sensitive parameter. Is this because the watershed is dominated by lateral flow, or is it because you have introduced the wetness index layer? As you only calibrate on streamflow with some comparison of the saturated areas, we don't actually know. The dominance of the wetness index layer also explains why the uncertainty in the soil and land use layers is minimal, specifically since you a-priori decided on a 10m DEM for that test. In summary, I believe you need to investigate this further and figure out how this exactly works in relation to the algorithm that you have introduced.

Response

➔ We agree with the reviewer that we did not make it clear the importance of the wetness class map and its interaction with others layers. We hope our response to your comment below will clarify it.

1. The role of wetness class map

The SWAT-HS model uses topographic index (TI) as the basis for hydrological modeling, like some other variable source area models: TOPMODEL (Quinn and Beven, 1993;Beven and Kirkby, 1979), SWAT-VSA (Easton et al., 2008), SWAT-WB (White et al., 2011). To keep the model semi-distributed, we divide the watershed into a limited number of wetness classes (maximum 10 classes in the current SWAT-HS). Each wetness class is assigned a soil water storage capacity (or called as saturation deficit in the TOPMODEL). This is different from the TOPMODEL in which saturation deficit is calculated from the value of topographic index, soil water storage capacities in wetness classes in SWAT-HS is assumed to follow Pareto contribution as:

$$edc_i = S_{\max}\left[1-(1-A_i)^{1/b}\right] \quad (1)$$

where edc_i is the soil water storage capacity in wetness class *i*, *Smax* is the maximum soil water storage capacity of the watershed, A_i is the fraction of the watershed for which the storage capacity is less than edc_i, and *b* is the shape parameter.

The lower *edc* values are assigned to the wetness classes having high TI values, locating in downslope areas ("wetter" wetness classes) while higher *edc* values are assigned to wetness classes with low TI values in upslope areas ("drier" wetness classes). *Smax* and *b* are two parameters controlling the Pareto distribution that we use in calibration. Note that this Pareto distribution is already used in other models to simulate saturation-excess runoff like: the Xinanjiang model (Zhao et al., 1995;Zhao et al., 1980), the VIC model (Wood et al., 1992;Liang and Lettenmaier, 1994) and the PDM model (Moore, 2007). These three models simulate saturation-excess runoff and estimate the saturated fraction of the watershed, however, they are not able to identify the specific spatial locations of saturated areas.

In conclusion, the role of wetness class map is to divide the watershed into areas with different saturation deficits which are lower in areas with high TI values and higher in areas with low TI values.

2. Interaction of wetness class layer with other input layers

To create HRUs in SWAT-HS, first, the soil map was overlaid with the wetness class map to create a new soil map in which the same soil types in different wetness classes have different soil names but retain the same soil characteristics (Hoang et al., 2017). The new soil name reflects both wetness class and soil type. Subsequently this new soil map is overlaid with land use map to create HRUs using the regular procedure in SWAT for

HRU definition. As we mentioned in the paper, we assumed that slope is not a part of HRU discretization for simplification purpose although slope is used in topographic index calculation and thus incorporated into wetness classes.

Once HRUs are created, each of them has details of its land use type, its soil type with soil name reflecting the wetness class number in which it is located. Each HRU has the initial soil water storage capacity (saturation deficit) depending on which wetness class it is located. This storage capacity will change over time depending on climate inputs and other processes occurring in the soil profile (percolation, uptake, evaporation, generation of different types of flow) which are affected by soil and land use information.

In conclusion, the wetness class map defines the initial saturation deficit in HRUs while the land use and soil type gives information to calculate hydrological processes that changes the saturation deficit and subsequently update the saturation deficit values for HRUs.

We agree with the reviewer that all our results point to the fact that the wetness index layer dominates the actual flow behavior. It is reasonable because SWAT-HS uses topographic index as the basis for hydrological modeling. And we think this is applicable to all watersheds that SWAT-HS is suitable to be used, i.e. watersheds dominated by saturation-excess runoff. We also agree with you that the dominance of the wetness class layer also explains why the uncertainty in the soil and land use layers is minimal. When the appropriate DEM resolution is used, soil and land use information become less sensitive to hydrological predictions.

3. *The latb parameter is a sensitive parameter. Is this because the watershed is dominated by lateral flow, or is it because you have introduced the wetness index layer?*

Yes, *latb* parameter is a sensitive parameter because our watershed is dominated by lateral flow. The below figure, which is in our previous paper (Hoang et al., 2017) with DEM 10m and the most complex soil and land use maps, shows the comparison of different types of flow in the Town Brook watershed. The figure shows that lateral flow is the dominant type of flow. Our result is compatible with the finding of Harpold et al. (2010) in an intensive field survey in a 2.5 km^2 headwater watershed of Town Brook watershed where hillside lateral preferential flow paths rapidly transported water to near - stream saturated areas during runoff events under relatively dry antecedent conditions. Harpold et al. (2010) also suggested that the lateral redistribution of water from hillside areas reduces the influence of surface topography and channel topology on the sources of stream runoff.

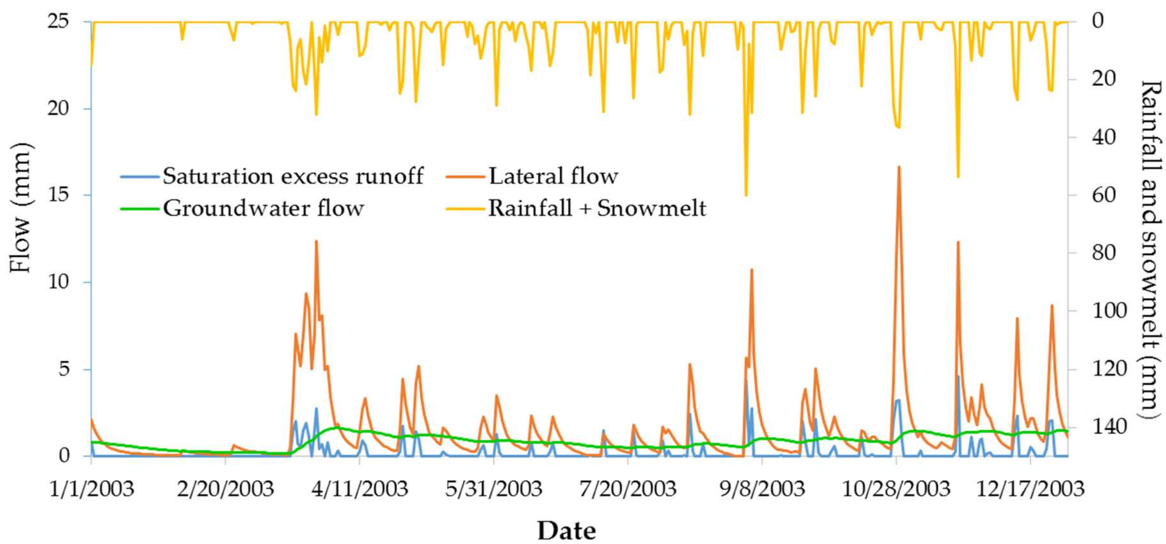

Figure 2: Time series of flow components simulated by SWAT-HS in the Town Brook watershed (Hoang et al., 2017)

Comment

The second major issue is that it is unclear from your research whether the results are more generally applicable. What is the global significance of your research? I am asking this as there is now real way of telling whether your results are water shed specific. You even seem to write to a local audience in the paper, often referring to your results as being specific and decisions being specific for this watershed. In a way, this is fine, but for HESS, the real value is in research that is of interest to a global public. This means that I believe that you need to define this better or test this better.

Response

➔ Thank you very much for your constructive comment. We are confident that our results are not only important in the New York City watershed region, but also have wider applicability.

For the New York City watershed region, our study will be a guidance for choosing input data (DEM resolution and the degree of complexity for soil and land use map) to apply SWAT-HS for a larger scale watershed which requires division into multiple subbasins and a certain degree of complexity for soil and land use information. Our results are important when we use SWAT-HS to identify critical runoff generating areas and locations within the watershed where management interventions for water quality improvement (e.g. Phosphorus loading) are most effective. In our follow-up work which uses the results of this study to scale up the application of SWAT-HS to a larger scale watershed to simulate streamflow and phosphorus (paper under review), we got good results for phosphorus calibration due to the correct simulation of hydrology (surface runoff in particular). Identifying the correct/optimum input layers is thus an important step. Our experience also suggests that model run time can also be reduced when using the optimum input layers which is particularly important during the calibration process.

Our results are applicable to watersheds with similar land use, topography, and climate but similar/independent work is needed in other regions using the methodology described in this paper. In our case DEM 10m gave the best results due to better physical representation of the landscape and is a compromise between the high resolution DEMs 1m and 3m that provide too much spatial details that interfere with calculation of upslope contributing areas and topographic index, and coarse resolution DEM 30m that average out the necessary fine details. DEM10m was also chosen as optimal in several studies in other regions by Kuo et al. (1999) and Zhang and Montgomery (1994), but was evaluated worse than other finer resolution like in Buchanan et al. (2014). Therefore, we think that the sensitivity of DEM resolution depends on the scale and characteristics of the watershed. Therefore, to choose the appropriate DEM resolution for other region, we recommend carrying out similar work with the methodology used in our study.

What we learned from this study is that, hydrological prediction is very sensitive to choice of DEM (with higher effect on prediction of saturated areas than streamflow), when using a hydrologic model that uses topographic index as the basis for hydrological modeling, to simulate hydrology in a watershed that is dominated by saturation-excess runoff. Besides SWAT-HS, some other watershed models using topographic index are: TOPMODEL (Quinn and Beven, 1993;Beven and Kirkby, 1979), SWAT-VSA (Easton et al., 2008), SWAT-WB (White et al., 2011). With SWAT-HS and models that are based on topographic index in general, DEM resolution is more sensitive than the complexity of soil/land use information. When the appropriate DEM resolution is used, soil and land use information become less sensitive to hydrological predictions.

We are sorry that the manuscript in current form does not transfer sufficient knowledge and information to the general audience. Based on our response to reviewer's comment, we will improve this in the revised manuscript.

Comment

An example is your DEM result, you point out that the 10m DEM seems best comparing the NSE and looking at saturated, but you don't seem to be able to explain why (which is the real wider interest). Is this because of the specific physiography of your watershed, or is this due to your specific model algorithms? Citing other literature that found similar things does not really help unless this helps you explain your result. So in summary, your results need to be explained better and it should be clearer what the value of your results to the wider research community. I have added many more comments on the attached pdf that are probably useful to address these issues.

Response

➔ We totally agree with the reviewer for lack of proper explanation on why DEM10m is the optimal resolution in our study. We tried our best to find the answer as below.

The prediction of saturated areas are based on wetness classes which are classified based on values of topographic index. Therefore, the sensitivity of DEMs on saturated areas predictions can be explained by the effect of DEM resolution on topographic index (TI).

Note that the basic equation for topographic index is TI = ln (contributing area/slope angle)

The below figure shows the relationships of TI with slope angle, upslope contributing area and elevation using 2 representative DEM resolutions: 1m and 10m. It is clearly observed that DEM 1m can capture a significantly wider range of slope than DEM 10m because of its finer resolution. Also, the percentage of grids that has low values of TI is significantly higher in DEM 1m than in DEM 10m (in figure below use red lines for reference), which also can be seen in figure 3d in the main manuscript. Low TI values are usually found in grids with steep slope or with low upslope contributing areas. Because DEM 1m captures steep slope at local scale and has a high number of grids with low upslope contributing area (figure 3c in the main manuscript), the percentage of low TI values in DEM 1m is much higher. If we look at the relationship between TI and elevation, we can see that the distribution of TI values in DEM 1m spread out wider than in DEM10m at all elevations. This explains why the distribution of wetness classes in DEM1m has a more complex pattern with every wetness class spread-out while DEM10m has a more coherent pattern with high TI wetness class well compatible with the stream network (Figure 4 in the main manuscript).

Our findings are in agreement with Lane et al. (2004) who used high resolution LiDAR 2m DEM in the TOPMODEL which simulates hydrology based on topographic index. The TOPMODEL predicted the widespread existence of disconnected saturated zones that expand within an individual storm event but which do not necessarily connect with the drainage network. They found that using the LiDAR 2m DEM, the topographic index has a complex pattern, associated with small areas of both low and high values of the topographic index, leading to the appearance of disconnected saturated areas. After remapping the topographic data are remapped at progressively coarser resolutions by spatial averaging of elevations within each cell, they found that as the topographic resolution is coarsened, the number and extent of unconnected saturated areas are reduced: the catchments display more coherent patterns, with saturated areas more effectively connected to the channel network. Moreover, in another study, Quinn et al. (1995) showed how progressively fining model resolution from 50 m to 5 m reduces the kurtosis in the distribution of topographic index values and increases quite substantially the number of very low index values.

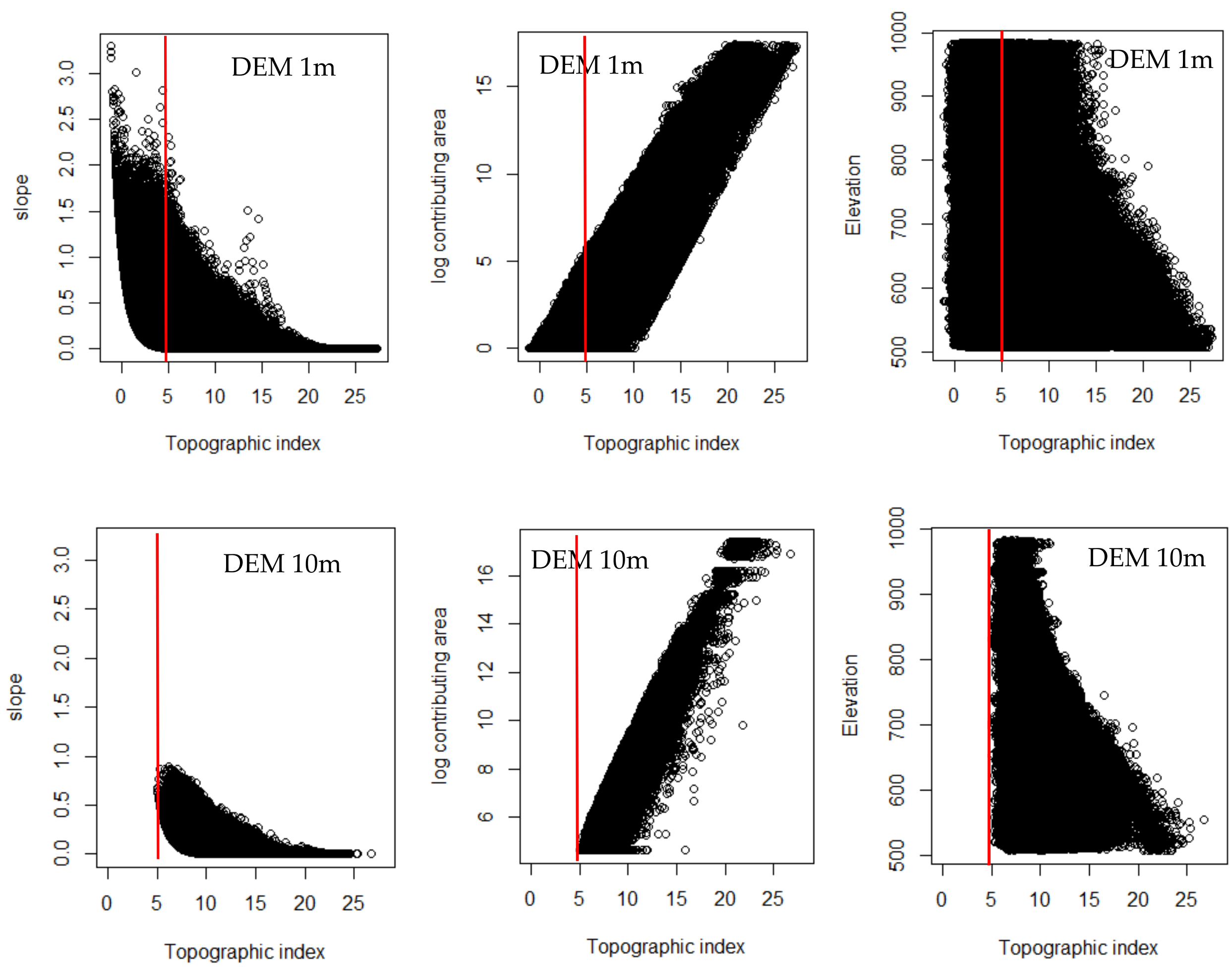

Figure 2: Relationship of topographic index with slope, upslope contributing area and elevation with two resolution of DEM: 1m and 10m

Comment

I have added many more comments on the attached pdf that are probably useful to address these issues.

Response

➔ We copied all the comments in the attached pdf excluding the comments on grammar and English as below and responded to them one by one.

Other comments in the supplement

Comment

P3 L1: *"This so called equifinality is very common in hydrological models and 1 is the cause for uncertainties in model predictions"* The main cause? The only cause? I think you just argued that there are three sources of uncertainty, but here you reduce this back to parameter uncertainty?

Response

➔ We are sorry for creating confusion here. We edited the text in the revised version as:

"This so called equifinality is very common in hydrological models and is one of the main causes for uncertainties in model predictions"

Comment

P5, L20: The research questions "*what is the suitable DEM resolution in order to get good model performance*" and ", and "*what is the appropriate complexity of the distributed input data*".

How dependent is this on the location where you are? How can you test that your results are generally applicable. I am worried that given the number of studies in this area that your results are in fact watershed dependent. This is also true for your second objective.

Response

➔ We agree with your comment. As stated earlier, our results are applicable to watersheds with similar land use, topography, and climate, and dominated by saturation-excess runoff. However, similar/independent work is needed in other regions using the methodology described in this paper.

The novelty of our study is that the focus is on the uncertainty in parameter values and output uncertainty (streamflow and saturated areas in this study) due to differences in input layers, while earlier SWAT studies have looked at the effect of differences in scale of DEMs, land use, and soil layers only on model predictions (streamflow, nutrients etc.).

Based on our response to this comment and other similar comments, we will provide a discussion about the general applicability of this study in the revised version.

Comment

P8, L3-5: "*Subsequently, we divided the remaining areas into 8 wetness classes (wetness class 2 – 9) with approximately equal areas (~ 6% each) based on TI values.*"

So is there any chance that this fairly arbitrary, but practical, division is suited only to your water shed? Or is this a general rule that should be applied if including the wetness index in SWAT. What is the best way to decide how the wetness classes should be scaled?

Response

➔ SWAT-HS gives flexibility for users to divide the watershed into wetness classes with the maximum number of wetness classes being 10. The division of wetness classes is arbitrary and requires general knowledge of the studied watershed.

In this study, we divided the Town Brook watershed into 10 wetness classes based on our expert knowledge of saturation, observations (Harpold et al., 2010) and predictions by other watershed models (SMR (Agnew et al., 2006), SWAT-VSA (Easton et al., 2008) and SWAT-WB (White et al., 2011). Agnew et al. (2006) developed a relationship between topographic index (TI) and probability of saturation P_{sat} for the Town Brook watershed using DEM 10m and suggested that the areas with TI >17.7 is always saturated. Based on this, we grouped the areas with $TI > 17.7$ as the "wettest" wetness class (wetness class 1) for DEM10m setup. For other DEMs resolution, we based on the distribution of wetness 1 in DEM10m to decide the TI threshold to use to create wetness class 1 for each DEM setup. Based on all mentioned information that we gathered, we know that saturated areas never exceeded 50% of watershed, we grouped 50% of the watershed with lowest TI values as the "driest" wetness class (wetness 10). Assigning the driest wetness class to half of the watershed allowed us to classify the remaining areas, which are more prone to saturation, into a greater number of wetness classes. We divided the remaining areas into eight wetness classes (wetness class 2–9) with equal areas based on TI values.

To apply SWAT-HS in a new watershed, we recommend users to initially adapt the procedure of wetness class classification used in this study. Wetness class 1 can be classified by choosing grid cells with upslope contributing areas higher than a reasonable threshold that makes wetness class 1 comparable to the stream network. Wetness 10 can be classified based on an expert knowledge on maximum estimate of saturation percentage, we believe that this information should not be difficult to find. Subsequently, wetness 2-9 can be created by dividing equally the remaining area.

Comment

P8, L13-14: "*…the model was calibrated and validated for the periods 2001-2007 and 2008-2012, respectively.*" What made you decide this order and this division? Why not the reverse and is there any difference in terms of climate between the periods?

Response

➔ We did not have particular reason to decide this order. We just wanted to have a reasonable number of years for the calibration and validation periods.

Comment

P8, L14-16: "*We excluded the year 2011 from the validation period because there were two extreme events (Hurricane Irene and Tropical Storm Lee) in August 2011 that the model could not capture well*." Isn't this a concern and actually worth investigating? Why does it not capture this well and how does it not capture this well? Is this related to the input uncertainty?

Response

➔ The below figure show the modeled streamflow versus observation in the year 2011 that we excluded from the validation period. Hurricane Irene and Storm Lee are the two extreme events indicated in the green box. Hurricane Irene and Storm Lee brought high rainfall amount for several continuous days with very high intensity for several hours of the day to the Catskill system of the New York City watershed while SWAT-HS predicts streamflow at a daily time step. Therefore, at a daily time step, SWAT-HS underestimated the magnitude of high streamflow caused by these two extreme events. However, the model captured the flow variation very well (see the below figure). We excluded the year 2011 from the validation period because we do not want the streamflow underestimation in these two events leads to an unfair evaluation of the model performance.

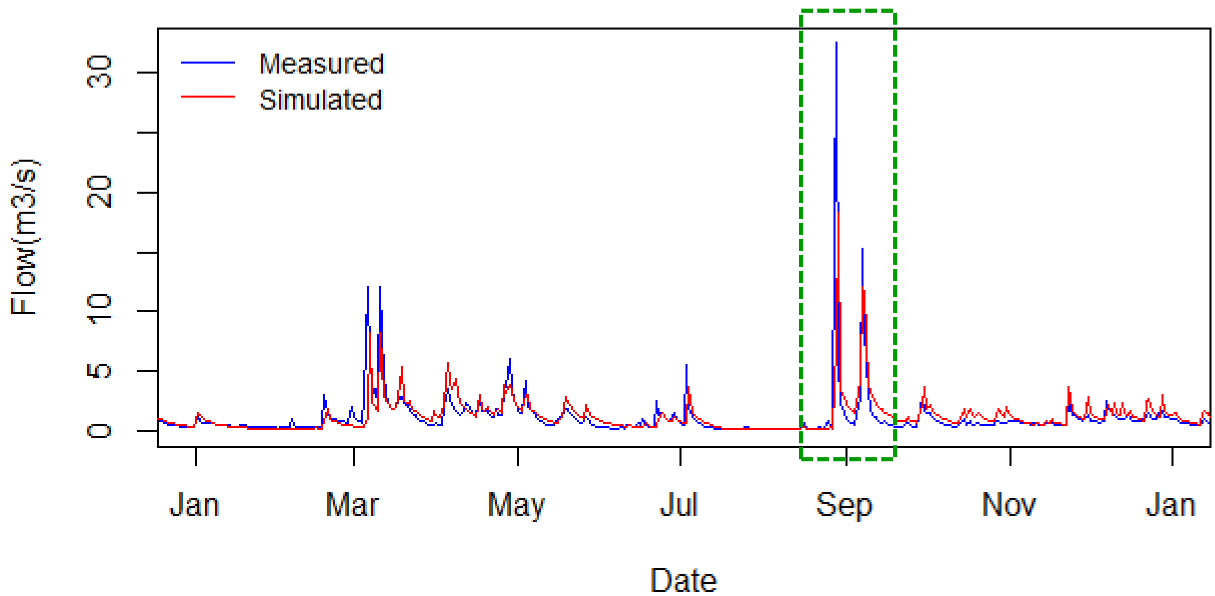

Figure 3: Comparison of simulated streamflow by SWAT-HS and observations in 2011 (peak flows caused by Hurricane Irene and Storm Lee are indicated in the green box)

Comment

P8, L20-22: "*The optimal parameter set from the snowmelt calibration was used in the final streamflow calibration*." How did you evaluate whether a model was "calibrated"? What did you use as objective function to assess your Monte Carlo output?

Response

➔ In the snowmelt calibration stage, we randomly generated 10,000 parameter sets for 5 snowmelt related parameters, ran those sets with SWAT-HS and compared streamflow predictions with observations The best parameter set giving the best fit to streamflow observations (highest NSE value) was chosen for the flow calibration stage. We had to use streamflow observations in the snowmelt calibration stage because observed snow data is not available.

We will clarify our calibration procedure by editing the text in the revised manuscript as:

"The calibration was carried out in 2 stages, i.e. snowmelt calibration and flow calibration, and by applying Monte Carlo sampling method. For snowmelt calibration, we calibrated 5 snowmelt related parameters in group (i) (Table 1) by generating randomly 10,000 parameter sets, running these sets using SWAT-HS, comparing the streamflow predictions with observations and choosing the best parameter set giving best fit to streamflow observations (highest value of daily Nash Sutcliffe Efficiency (NSE)) to use for the flow calibration stage. For flow calibration, 10,000 parameter sets of 9 flow parameters in group (ii) (Table 1) were generated which were then run with SWAT-HS. The simulations in the flow calibration stage were used for uncertainty analysis."

Comment

P9, L4-6: "*For each model setup, "good" simulations were identified as those with a Nash-Sutcliffe Efficiency (NSE) greater than 0.65 for use in uncertainty estimation of streamflow.*" How and why did you choose 0.65 as a threshold? How does this affect your results?

Response

➔ Figure 5a and figure 10a show the maximum daily NSE values in all our SWAT-HS setup which is 0.68-0.69. From our results, daily streamflow predictions with daily NSE higher than 0.65 results in monthly streamflow prediction with monthly NSE higher than 0.8. Based on guidelines for model performance evaluation by (Moriasi et al., 2007) that suggested "*good*" model performance for streamflow as corresponding to monthly NSE higher than 0.65, we are confident that our choice of NSE higher than 0.65 as good model performance at daily time-step is a reasonable choice.

Comment

P10, L4-5: "*In order to simplify the setup, we assumed that slope does not have an impact on HRU discretization.*"

Is there an interaction between slope and wetness index? I think this should be discussed. For that same matter, if a more detailed soil or land use map is available there could be an interaction between slope, wetness index, soil and land use. So in what way is the extra wetness index just a summary of other landscape elements? And how would this influence results. In a way you are creating a further split of the HRUs, making them more "specific", but a similar result could be achieved with a more detailed soil or land use map? So I think you need to be careful here to argue why the wetness index split is a better summary (thus can be coarser) then just using land use and soil and slope detail? As you have chosen to just use one DEM here, you cannot really investigate that detail.

Response

➔ The below figure shows the relationship between topographic index (TI) & the classified wetness classes and slope & upslope contributing area. This classification of wetness classes based on TI is used in our previous study (see Table 2 in Hoang et al. (2017)) using

DEM10m. Wetness class 1 with highest TI values include grids with gentle slope and high upslope contributing areas. Wetness class 10 which actually covers 50% of the watershed include grids with low contributing areas and wide range of slope angles.

As explained above, the role of wetness class map is to divide the watershed into areas with different saturation deficits. The most important processes of SWAT-HS: the simulation of saturation-excess runoff and the generation of lateral flow in saturated areas are based on the values of saturation deficits. Therefore, wetness class map is very important in HRU definition in SWAT-HS. We cannot replace it with more details of soil, land use or slope.

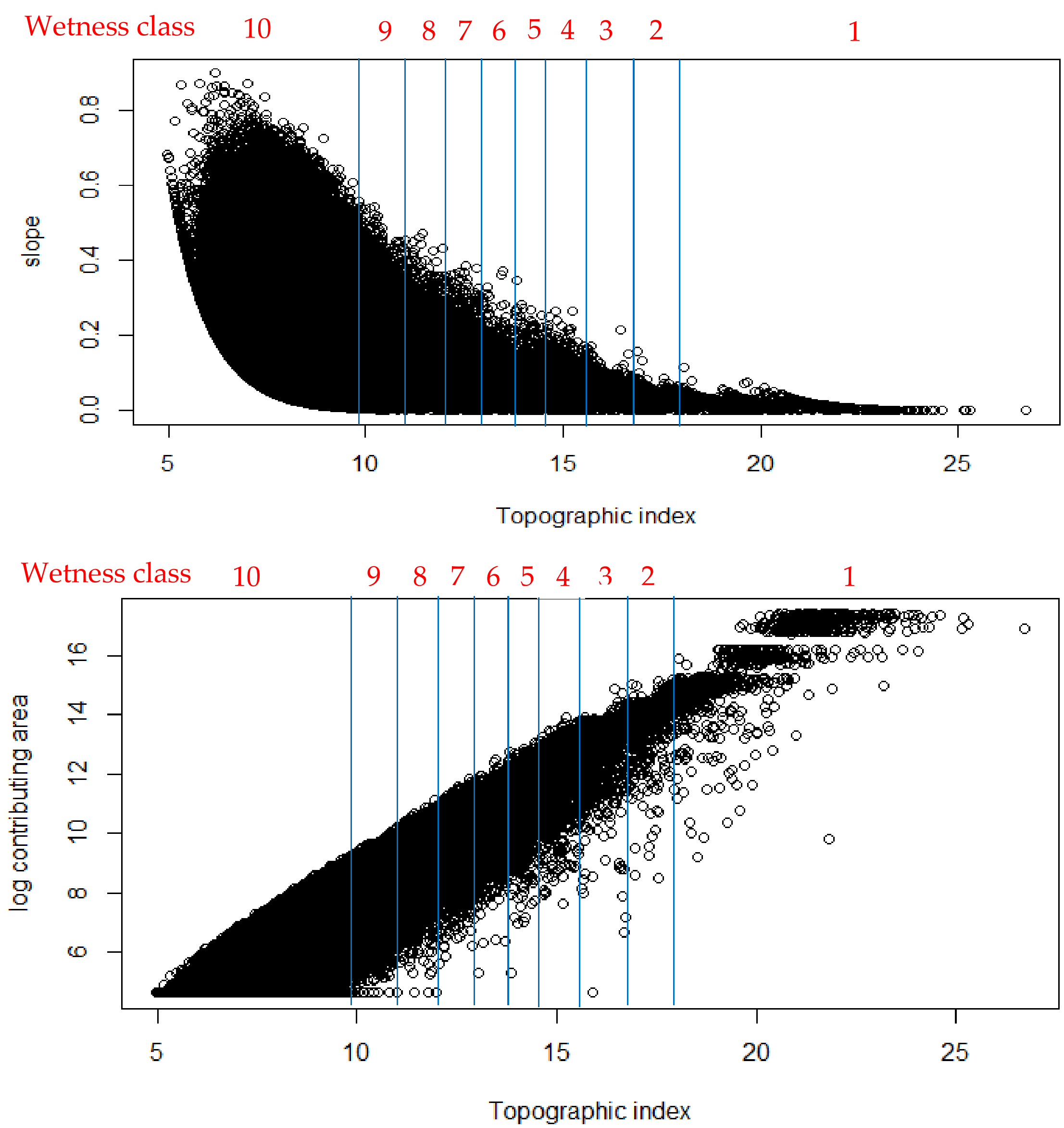

Figure 4: Relationship between slope, topographic index and wetness class using DEM10m

Comment

P12, L10-12: "*Generally, there are only slight differences in SWAT-HS performance on streamflow using different DEMs implying the insignificant effect of DEM resolution on streamflow simulation and the uncertainty of streamflow outputs.*"

I am not sure I would agree with you, you have just said that the finer DEM give fewer "good" performances. I would think that is quite an effect and requires discussion

Response

➔ We agree with the reviewer that we underestimate the effect of DEM resolution in this sentence. In the revised version, we deleted this sentence to prevent confusion because the following paragraph of this sentence already discussed about the effect of DEM resolution.

Comment

P12, L17-18: "*This implies better streamflow prediction by these two setups even without calibration*" Why?

Response

➔ In our response to your previous comment that required explanation on why DEM 10m is the best resolution to use, we already explained why the distribution of TI and wetness classes in fine resolution DEM (represented by DEM1m) has a more complex pattern low and high TI value grids spread out while DEM10m has a more coherent pattern with high TI grids well compatible with the stream network. Here we show the distribution of TI values using different DEMs in the below figure. It clearly shows that the coarser DEMs (10m and 30m) have grids with TI values compatible with the stream network, while in the finest DEM1m, it is hard to recognize where stream network is based on TI values. Realistically, highest TI value grids should locate in downslope, near-stream, low elevation areas while lowest TI value grid should be in upslope, high elevation areas. However, fine DEMs do not capture this physical representation of the landscape well while coarser DEMs do it better. We think that this is the reason why coarser DEMs give better predictions in both streamflow and saturated areas than the finer DEMs.

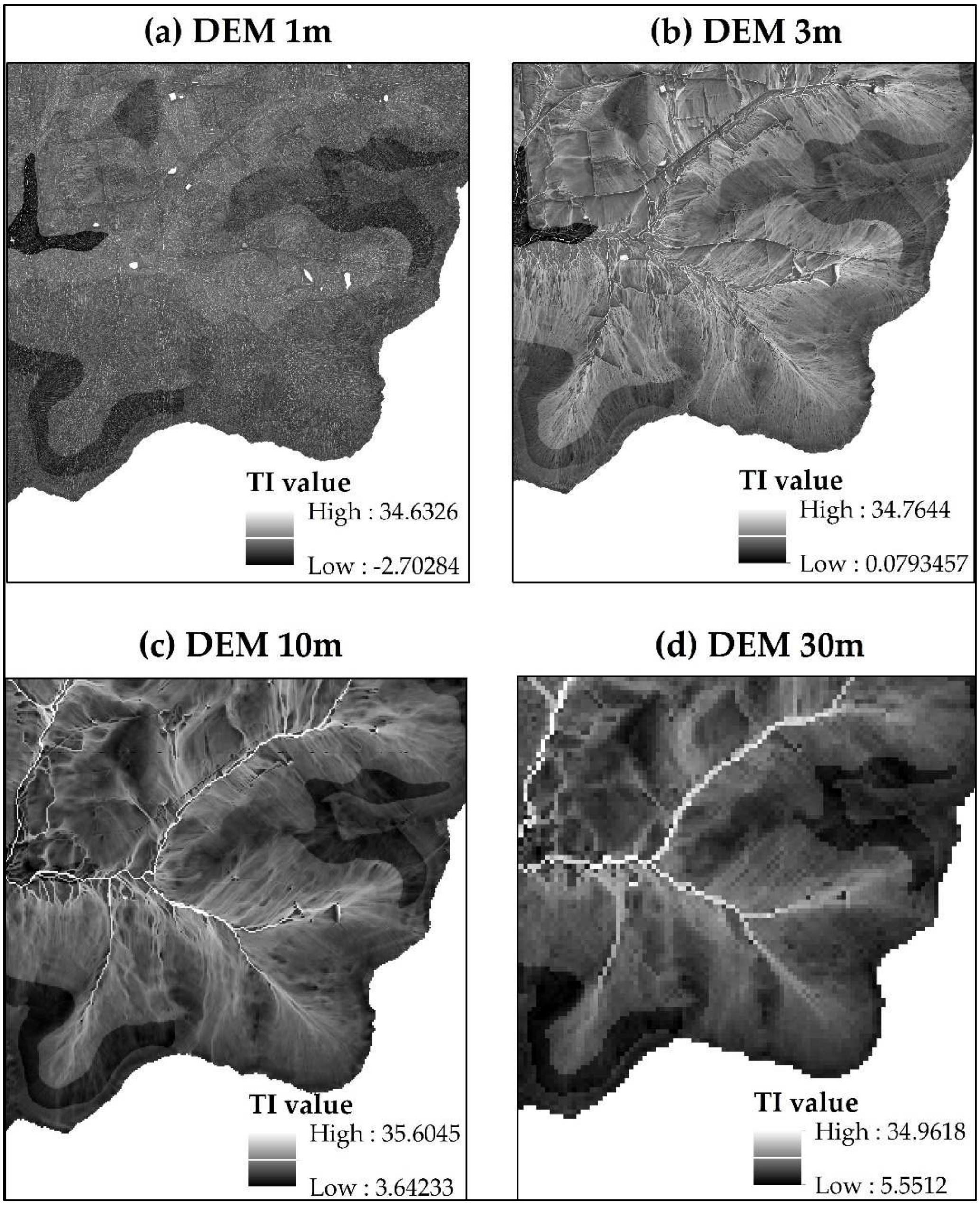

Figure 5: Distribution of topographic index values using different DEMs

Comment

P12, L24-25: *"The probability of saturation, which indicates the number of days in the calibration period when the wetness class is saturated, showed no significant difference among the four setups indicating that DEM resolution does not have an impact on the probability of saturation."*

Again, the real question here is why? If you look at your mapping of saturated areas against observed, it seems the coarse DEM does better (fig 8). Why is this? Is this related to your choice of dividing up the classes of the wetness index? I think this is a somewhat unexpected result

Response

➔ We tried to keep the areal percentage of each wetness class approximately similar in the four setups using different DEMs. The '*good*' parameter sets in 4 setups should give comparable predictions of streamflow, percentage of watershed area that is saturated, and the time that each wetness class saturated. That is the reason why Figure 6 shows that the probabilities of saturation for each wetness class are similar in 4 setups.

Comment

Section 3.2.2: I would just delete this section. You have just argued that the only difference in saturated areas is due to the DEM so using the same DEM would never give you different saturated areas? I guess this just points to the fact that your wetness index layer is overriding other effects (i.e. is much stronger than any land use and soil effects). But you can deal with that in one sentence, not a whole paragraph.

Response

➔ We agree with the reviewer. We replaced this section with one sentence in the revised manuscript as:

"Nine setups used the same DEM with 10m resolution and have the same distribution of wetness classes; therefore, the distributions of their predicted saturated areas, are similar and thus are not shown here."

Comment

P16, L2-4: "*Similar to the comparison of four setups using different DEMs, the nine setups with different degrees of complexity produced different numbers of good parameters for streamflow and saturated areas, but were similar in the shape of their distributions and value ranges.*"

Are these distributions very different from the DEM distributions? Not really, so why not? Why would two quite different input variations give you similar parameter distributions? You say *latb* is sensitive, but this again seems to point to an overriding effect of your wetness index distribution on the overall results. This is basically masking any other behavior. I think that is a worry. Or maybe a good thing and actually a reflection of your physiography and climate??

Response

➔ Yes, the parameter distributions in nine setups using different soil/land use complexity and four setups using different DEMs are similar. We think this is because we run sufficiently high number of Monte Carlo parameter sets that has good coverage of parameter space. Although different inputs result in varied number of good parameter sets, the numbers of 'good' parameter set in all setups are also sufficient to represent the distribution of 'good' parameter which reflects their sensitivities to hydrological prediction.

  As discussed previously, *Latb* parameter is a sensitive parameter because our watershed is dominated by lateral flow. This was supported not only by our results of the contribution of different flow components but also supported by field results (Harpold et al., 2010) in the watershed.

Comment

P16, L13-14: "*However, after calibration, the effect of DEM resolution on the uncertainty of streamflow prediction is very minor.*"

I agree, but....You found that in the Monte Carlo, the finer DEM had fewer "good" results. I still would like to know why! This is unexpected isn't it? Or does this have to do with how the wetness index is implemented?

Response:

Our response to this comment is same with the response to your previous comment. Finer DEMs had fewer good results because they do not capture the physical representation of the landscape realistically.

Comment:

P16, L18-19: "*These studies found that discharge was simulated equally well irrespective of DEM resolution as long as parameters are calibrated properly.*"

See earlier, I think this is not actually a good thing. This just means the model is flexible enough to "recover" from a bad input layer. But does this actually mean we get the "right answer for the right reasons"

Response:

➔ We agree with the reviewer. However, we do not think it is a concern. It means that we need to use more than one type of observations in evaluating the model performance. That is the reason why we also compare the four setups based on the streamflow results without calibration (considering all random parameter sets) and predictions of saturated areas.

Comment

P17, L3-8: *"In our analysis of effect of DEM resolution on topographic characteristics, we observed that the statistical distribution of TI is very sensitive to DEM resolution (Fig. 3d), which results in considerable differences in spatial distribution of wetness classes (Fig. 4). This explains why the distribution of simulated saturated areas by SWAT-HS is also very sensitive to DEM resolution*."

So this is a concern isn't it? In the end the spatial distribution would affect water quality estimates and other things that you might want to simulate. Basically anything other than simply streamflow.

Response

➔ We do not think this is a concern, but tells us why it is important to identify the right DEM resolution for both streamflow and saturated area prediction.

SWAT-HS predicts saturation-excess runoff and saturated areas based on the classification of wetness classes. The classification of wetness classes is based on values of topographic index which is calculated from DEM. Therefore, it is logical that the distribution of simulated saturated areas by SWAT-HS is also very sensitive to DEM resolution. We are aware that the distribution of simulated saturated areas will control the prediction of water quality. This is actually our motivation to conduct this study to ensure that we use the appropriate DEM before applying SWAT-HS to predict water quality.

Comment

P17, L13-15: *"Therefore, DEM10m is the preferred choice to scale-up the application of SWAT-HS to larger watersheds in the New York City water supply system for future applications."*

OK, that is very specific. So what is the general knowledge that we can gain from your test, or are you just doing a sensitivity analysis of your specific model?

Response

➔ As stated earlier, our results are applicable to watersheds with similar land use, topography, and climate. However, similar/independent work is needed in other regions using the methodology described in this paper.

Comment

P18, L7-10: *"The difference in scale of case studies (field scale vs. watershed scale) and characteristics of case studies (agricultural fields vs. a mixture of forest and agriculture) between Buchanan et al. (2014) and our study may have resulted in different conclusions on choice of the appropriate DEM resolution."*

Ok, so this is the interesting stuff, and I think you need to spend more discussion on this point. It is not that interesting to argue why this is the "right" DEM, more interesting is how your "best" DEM relates to the local topography and therefore the "right" DEM can be chosen a-priori? You suggest some relationship between topography and variation of land use? How can we generalize this? What type of research would be needed to identify this? You just quote other case studies, are you saying there is no literature that looked at DEM resolution versus scale or spatial variation?

Response

➔ Based on our responses to your previous comments, we think that the sensitivity of DEM resolution depends on the scale and characteristics of the watershed. The dominant hydrological process in the watershed has a big impact on the sensitivity of DEM on hydrological prediction. For example, in our watershed, lateral flow is a dominant flow component and saturation-excess runoff is a dominant type of surface runoff, thus, topography is the most important factor. Therefore, the DEM that represents a realistic distribution of topographic index (TI) with high TI area compatible with the main stream network gave a better model performance. In a field scale watershed, finer DEM is probably better because it can capture a more detailed and realistic representation of TI distribution. In an agricultural area dominated by tile drainage, DEM resolution may not be sensitive.

Some of these questions about relationship between topography and variation of land use) are interesting but beyond the scope of this study. Topography is certainly related to land use type. For example, in this watershed, agricultural areas are located in the downslope, near stream areas which has high topographic index while forest is mostly distributed in the upland. We will consider researching on this subject in our future study.

Our literature review showed lack of SWAT studies on the relationship between DEM resolution and spatial predictions. There are studies looking at DEM resolution versus scale and spatial variation, for e.g. Wolock and Price (1994) using the TOPMODEL. In the revised manuscript, we will improve it by quoting more studies related to this topic.

Comment

P18 L24-27: "*However, with proper calibration, all nine models are able to provide good performances and their "good" parameter sets continue to perform equally well in the validation period. In addition to streamflow, all nine setups are able to capture saturated areas correctly on specific*"

Again this does not answer the question whether in your specific watershed the wetness index dominates over the other layers? I think this really needs to be discussed and I think that is what happens (based on your results). But that kind of makes your tests slightly irrelevant...

Response

➔ As our response to your previous comment, SWAT-HS is a topography dominated hydrological model. Therefore, the wetness class map dominates over the other layers. However, we do not think that this makes our test of comparing setups with different soil and land use complexity irrelevant. This test helps us to learn that DEM is the most important input for SWAT-HS and it is very important to choose the appropriate DEM resolution. The importance of soil and land use information is not as significant as DEM in hydrological modeling, but it will have a considerable impact when we use the model to simulate water quality.

Comment

P19, L1-3: "*We conclude that increasing spatial input details does not necessarily give better results for streamflow simulation as long as the model is properly calibrated*."

That is a bit of a cop out. That is saying: oh well I have enough parameters to fiddle with, so I can just feed the model rubbish. I don't think that is really the point. You have only compared to streamflow and saturated areas, but what about other variables (Actual ET, water quality?). So if you just are predicting water quantity, then yes maybe your statement is true given the large number of parameters that you can manipulate in SWAT, but it is not saying that the model is then a good representation of the catchment.

Response

➔ As mentioned above, SWAT-HS uses topographic index as the basis of hydrological modeling. Therefore, from our results, DEM resolution is more important input than soil and land use. This is the reason why we did not see significant differences from nine setups with different degrees of complexity when the appropriate DEM (DEM10m in this case study) was already used. We will have a more proper discussion on this issue in the revised manuscript.

Comment

P19, L7-8: *"It should be noted here that in this paper, hydrological response is the main focus of this study, and streamflow may not be very sensitive to the details of land use."*

Indeed, water quality would be much more sensitive.

Response

➔ Yes, we totally agree with the reviewer. Land use information controls the inputs of nutrients and information of other human activities that affect water quality, therefore, the water quality is expected to be very sensitive to the details of land use.

Comment

P19, L13-16: "*Petrucci and Bonhomme (2014) show that the inclusion of some basic geographical information, particularly on land use, improves the model performance, but further refinements are less effective.*"

This really depends on what other info is available as well. As I have indicated in your case I think wetness dominates and therefore land use is less important for water quantity.

Response

➔ We agree with the reviewer. Petrucci and Bonhomme (2014) conducted their study in a small residential area where the correct estimate or identification of impervious cover and path for surface water are very important for modeling. Therefore, including these information in the model setup helped to improve model performance.

We think that the importance of input data depend on the characteristics of the case study and the aspect that we want to model. Our studied watershed is rural and dominated by saturation-excess runoff. Therefore, topography and the wetness conditions of areas in the watershed are more important than land use in water quantity modeling. However, in water quality modeling, the importance of land use information which controls the input of nutrients is expected to be significant.

Comment

P19, L17-19: "*Finger et al. (2015) compared different setups with increasing detail in input information using the HBV model and three observational data sets. They found that enhanced model input complexity does not lead to significant increase in overall performance,…*"

In water quantity?

Response

➔ Yes, the study of Finger et al. (2015) focuses on water quantity evaluation. We clarified this by editing the text as:

"Finger et al. (2015) compared different setups with increasing detail in input information using the HBV model and three observational data sets. They found that enhanced model input complexity does not lead to significant increase in overall performance in water quantity,…"

Comment

P19, L21-23: "*Muleta et al. (2007) also showed that streamflow simulated by SWAT is relatively insensitive to spatial scale when comparing multiple watershed delineations from different soil and land use input data details.*"

Again, just water quantity??

Response

➔ Yes, we quoted this reference for its result on the sensitivity of streamflow to spatial scale of input data. In our text, we wrote "*streamflow simulated by SWAT*", so we mean water quantity here.

Muleta et al. (2007) also studied the sensitivity of sediment yield to soil and land use input data details and found that sediment generated and sediment that leaves the watershed decreases as spatial scale gets coarser. However, in our study, we focus on only hydrology, therefore, we did not refer to the sediment result.

Comment

P20, L3-9: "*Therefore, we conclude that for this case study and the particular model SWAT-HS, using higher resolution DEM or adding complex information on soil or land use did not reduce parameter uncertainty or solve the equifinality problem. This statement may not be valid for other areas that are characterized by numerous land use and complex variations in topography and soil types. This is also not valid for physically based models which require detailed soil and land use information and a minimum number of parameters for calibration.*"

So what did we learn that is not "just for your case study"? I think this is important to highlight for HESS

Response

➔ What we learned from this study is that, hydrological prediction is very sensitive to choice of DEM (with higher effect on prediction of saturated areas than streamflow), when using a hydrologic model that uses topographic index as the basis for hydrological modeling, to simulate hydrology in a watershed that is dominated by saturation-excess runoff. Besides SWAT-HS, other examples of watershed models using topographic index are: TOPMODEL, SWAT-VSA, and SWAT-WB. With SWAT-HS and models that are based on topographic index in general, DEM resolution is more sensitive than the complexity of soil/land use information. When the appropriate DEM resolution is used, soil and land use information become less sensitive to hydrological predictions. Also our experience shows that with appropriate model input (using DEM10m and intermediate soil and land use complexity), model run-time can be reduced which makes calibration procedure faster.

We will add a paragraph in the revised manuscript to clarify this.

Comment

P20, L16-17: "Our study is not aimed at solving the equifinality problem, but rather reduces the number of solutions considered when using SWAT-HS to predict streamflow and water quality for decision-making."

I did not see any water quality results, so I think this over reaching.

Response

➔ We agree with the reviewer, we removed the "water quality" part in this sentence. We will revise the text in the revised manuscript as:

"Our study is not aimed at solving the equifinality problem, but rather reduces the number of solutions considered when using SWAT-HS to predict streamflow."

Comment

Reference Hoang et al. (2017) repeated twice

Response

➔ We removed the extra reference in the revised manuscript

REFERENCES

Agnew, L. J., Lyon, S., Gérard-Marchant, P., Collins, V. B., Lembo, A. J., Steenhuis, T. S., and Walter, M. T.: Identifying hydrologically sensitive areas: Bridging the gap between science and application, Journal of Environmental Management, 78, 63-76, http://dx.doi.org/10.1016/j.jenvman.2005.04.021, 2006.

Beven, K. J., and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant, Hydrological Sciences Bulletin, 24, 43-69, 10.1080/02626667909491834, 1979.

Buchanan, B. P., Fleming, M., Schneider, R. L., Richards, B. K., Archibald, J., Qiu, Z., and Walter, M. T.: Evaluation topograpic wetness indices across central New York agricultural landscapes, Hydrology and Earth System Science, 18, 3279-3299, 2014.

Easton, Z. M., Fuka, D. R., Walter, M. T., Cowan, D. M., Schneiderman, E. M., and Steenhuis, T. S.: Re-conceptualizing the soil and water assessment tool (SWAT) model to predict runoff from variable source areas, Journal of Hydrology, 348, 279-291, http://dx.doi.org/10.1016/j.jhydrol.2007.10.008, 2008.

Finger, D., Vis, M., Huss, M., and Seibert, J.: The value of multiple data set calibration versus model complexity for improving the performance of hydrological models in mountain catchments, Water Resources Research, 51, 1939-1958, 2015.

Harpold, A. A., Lyon, S. W., Troch, P. A., and Steenhuis, T. S.: The hydrological effects of lateral preferential flow paths in a glaciated watershed in the Northeastern USA, Vadose Zone Journal, 9, 397-414, 2010.

Hoang, L., Schneiderman, E. M., Moore, K. E. B., Mukundan, R., Owens, E. M., and Steenhuis, T. S.: Predicting saturation-excess runoff distribution with a lumped hillslope model: SWAT-HS, Hydrological Processes, 31, 2226-2243, 10.1002/hyp.11179, 2017.

Kuo, W. L., Steenhuis, T. S., McCulloch, C. E., Mohler, C. L., Weinstein, D. A., and DeGloria, S. D.: Effect of grid size on runoff and soil moisture for a variable-source-area hydrology model, Water Resources Research, 35, 3419-3428, 1999.

Lane, S. N., Brookes, C. J., Kirkby, M. J., and Holden, J.: A network-index-based version of TOPMODEL for use with high-resolution digital topographic data, Hydrological Processes, 18, 191-201, doi:10.1002/hyp.5208, 2004.

Liang, X., and Lettenmaier, D. P.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, 1994.

Moore, R.: The PDM rainfall-runoff model, Hydrology and Earth System Sciences Discussions, 11, 483-499, 2007.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, Transactions of the ASABE, 50, 885-900, 2007.

Muleta, M. K., Nicklow, J. W., and Bekele, E. G.: Sensitivity of a distributed watershed simulation model to spatial scale, Journal of Hydrologic Engineering, 12, 163-172, 2007.

Petrucci, G., and Bonhomme, C.: The dilemma of spatial representation for urban hydrology semi-distributed modelling: Trade-offs among complexity, calibration and geographical data, Journal of hydrology, 517, 997-1007, 2014.

Quinn, P. F., and Beven, K. J.: Spatial and temporal predictions of soil moisture dynamics, runoff, variable source areas and evapotranspiration for Plynlimon, mid Wales, Hydrological Processes, 7, 425-448, 10.1002/hyp.3360070407, 1993.

Quinn, P. F., Beven, K. J., and Lamb, R.: The in(a/tan/β) index: How to calculate it and how to use it within the topmodel framework, Hydrological Processes, 9, 161-182, doi:10.1002/hyp.3360090204, 1995.

White, E. D., Easton, Z. M., Fuka, D. R., Collick, A. S., Adgo, E., McCartney, M., Awulachew, S. B., Selassie, Y. G., and Steenhuis, T. S.: Development and application of a physically based landscape water balance in the SWAT model, Hydrological Processes, 25, 915-925, 10.1002/hyp.7876, 2011.

Wolock, D. M., and Price, C. V.: Effects of digital elevation model map scale and data resolution on a topography-based watershed model, Water Resources Research, 30, 3041-3052, doi:10.1029/94WR01971, 1994.

Wood, E. F., Lettenmaier, D. P., and Zartarian, V. G.: A land-surface hydrology parameterization with subgrid variability for general circulation models, Journal of Geophysical Research: Atmospheres, 97, 2717-2728, 10.1029/91JD01786, 1992.

Zhang, W., and Montgomery, D. R.: Digital elevation model grid size, landscape representation, and hydrologic simulations, Water resources research, 30, 1019-1028, 1994.

Zhao, R., Yilin, Z., Leren, F., Xinren, L., and Quansheng, Z.: The Xinanjiang Model, Hydrological Forecasting, 351-356, 1980.

Zhao, R., Liu, X., and Singh, V.: The Xinanjiang model, Computer models of watershed hydrology., 215-232, 1995.