Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2016-307-RC1, 2016

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Picturing and modelling catchments by representative hillslopes" *by* Ralf Loritz et al.

Anonymous Referee #1

Received and published: 19 July 2016

Summary and Recommendation:

In this paper the authors address two basic questions:

1) If you have a lot of spatially-distributed information about the geology and soil-hydraulic properties in a catchment, can you parameterize a high-dimensional, spatially-distributed model (without any calibration or inverse optimization) to accurately represent water flow within a single 2-d hillslope, based on that existing knowledge?

2) If your knowledge-based (not optimized) model domain and parameterization prove reasonably representative, can you then extrapolate this representative 2-d hillslope across the 3-d volume of the entire catchment, to simulate hydrograph dynamics and the annual water balance for the entire catchment?

To address these questions the authors employ a Richards-equation-based model with evapotranspiration module and overland flow routing modules. They apply the model to simulate hillslope-scale soil moisture dynamics and water-balance partitioningâĂŤand after extrapolation, whole-catchment streamflow dynamicsâĂŤfrom two catchments in Luxembourg with varying geology, topography, soil, and vegetation. In addition to the modeling, their analysis includes extensive, and impressive data sets representing spatially distributed soil-hydraulic properties, geologic features, plant transpiration, and topography. These questions, the observations, and the methodological approach adopted here are of interest in scientific hydrology and would be received with interest by readers of HESS. The writing is mostly clear in a grammatical sense, though somewhat desultory and technically ambiguous in many areas. The model setup is technically sophisticated in many ways, though not so in others. The graphics are of very high quality. The organization of the paper is logical, though as I suggest below, there are significant portions that might be omitted to maintain a consistent focus of the paper throughout. Overall there were too many competing objectives in the paper. As such, any salient result or conclusion is hard to discern. The results and discussion could be greatly improved.

I recommend that this paper could be accepted for publication in HESS, but I believe some major revisions are needed beforehand, possibly including revised simulations and results. Those suggested revisions are outlined in the following General Comments section, with references to more specific Technical Comments.

General Comments:

The manuscript can be improved upon significantly by reducing total length, omitting or clarifying the use of excessive jargon, and possibly removing sections of the paper to minimize superfluous and unfocused commentary. Some specific instances are noted in Technical Comments 5-10, 24, 27, and others.

The manuscript may be much improved by omitting much of the commentary about modeling evapotranspiration, along with the virtual experiment 3 (VE3) and associated results, and rather focusing on the importance of explicitly representing (or not) landscape heterogeneity for the purpose of simulating hydrographs. The authors talk quite a lot about all the uncertainties associated with ET modeling in hillslope/catchment hydrology, but their modeling approach does not reflect the state of the science (e.g. as presented in disciplines such as hydrometeorology and plant biophysics), so this discussion does not seem warranted. See Technical Comments 12, 29, and 56.

The methodological approach is inadequately described in many instances, with some revision being needed. See Technical Comments 16-26, 36, 39, and others.

There are some aspects of the model domain and parameters that seem very unrealistic. For example, the bedrock for both modeled catchments is parameterized to have porosity of 40-45% (Table 1). That's comparable to, or greater than, the porosity of many soils. I can't imagine how SchistâĂŤa metamorphic crystalline rockâĂŤcan be composed of 40% air space. This is certainly not consistent with most reported porosities for Schist, which are typically 10% or less. This will have a large impact on the flow simulations, since about half of the hillslope domain is bedrock. If there is some justification for this, and some other aspects of the model, then perhaps the only revision that is needed is to provide that justification. Otherwise, many of the simulations may need to be repeated with more appropriate parameterization. See Technical Comments 19-26, and others.

I strongly suggest that the authors consider refocusing this paper on two subject areas. First, concentrate on the modeling of spatially-distributed soil moisture dynamics and the temporal dynamics in the hydrograph. You use a spatially distributed model but nowhere do you assess the models ability to accurately represent any spatial pattern. You have an enormous amount of interesting spatially distributed data, so this could become one of the most rigorous tests of the Richards equation model at the hillslope scale ever published. See Technical Comments 33-34, 47-48, and 54. I recommend the second focal point to be the argument that extrapolating the parameterization of

a single 2-d hillslope to an entire 3-d catchment may, or may not, be defensible. At present this is stated as an objective, but the results and discussion are ambiguous, or possibly in disagreement, about this point. Showing these outcomes would be challenging enough, and a good contribution to contemporary scientific hydrology. Toward that aim, I suggest omitting the virtual experiments and associated discussion. I found the virtual experiments and the associated discussion around them to be desultory and vastly oversimplified. It was not clear to me how they relate to any previously stated objective. Those virtual experiments mainly consisted of changing a single variable (e.g. total relief, timing of bud break for plants, or hydraulic conductivity), and speculating broadly about the implications of the resulting simulations. See Technical Comments 50, 52-53, and 55-56.

Last, the authors seem to be advocating that high-dimensional, spatially-distributed models will always be wrong for a variety of reasons, but that they're still important for helping us learn about which state-variables and flow processes dominantly control the emergent streamflow dynamics at the catchment outlet. I think this is a relevant and worthwhile argument, but I encourage the authors to better focus their writing on this topic. This effort may be aided by omitting some other sections of the paper, as noted above. At present the authors present some seemingly conflicting (at least to me) statements about what exactly is the merit of taking this approach, and just how well it did or did not work out for them. See Technical Comments 34, 46, 47-48, and 54.

Technical Comments/Corrections:

1) Line 54: Change "reflect" to "reflects".

2) Line 57: I suggest using "e.g." throughout the manuscript when the cited work is only one example of all the works that could be cited to support a particular statement, as you did here for the Brontstert and Plate reference, but not the others. Certainly there must be innumerable works on hillslope energy and sediment fluxes.

3) Line 58: I would suggest using either "perceptual" or "perceptional" as an adjective to describe "model", but not both.

4) Lines 58-61: Citations not needed for this subjective statement.

5) Line 67: Not immediately clear to the reader what distinguishes a "conceptual model" from a "perceptual model".

6) Lines 85-88: Consider rephrasing or deleting. Not clear what is the message of the sentence. The result of any mathematical model must be compared to observations in the modeled system. This goes without saying, and certainly doesn't require a supporting citation. Perhaps I am just not clear what you mean by "benchmarking".

7) Lines 49-97: This is very clearly written, but for sake of making your paper as concise as possibleâĂŤand therefore more likely to be read in fullâĂŤyou might consider abbreviating the section, or deleting. It doesn't assert much, or highlight some problem with the status quo in catchment modeling. It mainly states that model-based analysis are useful for learning about hydrological systems, which I think is already acknowledged ubiquitously in the community of hydrological scientists. It's your call, but there is no shortage of papers on hillslope modeling, and I always prefer to read a more concise one than a longer one.

8) Line 101: Use of "behavioral" is ambiguous, maybe omit or clarify.

9) Line 113: "functional behavior of catchments of organized complexity" is hard to interpret. Caution in using too much ambiguous jargon.

10) Line 128-130: Consider rephrasing using the plainest language possible, since this is seemingly an important part of your rationale statement for the study.

11) Line 137-140: Rephrase "behavioral physically model structures". Also, comparing model outputs to observed data sets (like tracer time series) doesn't inherently reduce the number of degrees of freedom in the modeling procedure. It might help constrain parameter values. If that observed data set somehow informs the modeler that a par-

ticular parameter is unnecessary, or that a spatially-distributed domain can be adequately represented in a lumped way, then the degrees of freedom might be reduced. Are the works you cite here examples of the latter? For example, in the application of Richards equation with the van Genuchten-Mualem soil-hydraulic model, you can't just decide based on some observation that you no longer need the shape parameter in the hydraulic model – it still has to be there. If the observational data set leads you to a coarser grid resolution for the domain, then that would be a reduction in degrees of freedom, since the number of spatially distributed elements where the equation is solved/averaged is reduced. But really, for multi-parameter models that are spatially distributed, the degrees of freedom are always grossly high, not even considering the fact that we most often ignore anisotropy and hysteresis in soil hydraulics in hillslope- to catchment-scale applications. That's the whole motivation for lumped models, right?

12) Lines 166-184: Evapotranspiration is represented in a rudimentary way in many hydrological models precisely because those models have the primary aim of predicting hydrographs. For modelers with this primary interest, there will inevitably be a greater effort spent on representing such processes as non-equilibrium flow than on ET, because the former is of more interest. Models aimed at predicting streamflow use long-standing, and possibly antiquated ET models, because they're convenient, not because enhanced knowledge of stomatal dynamics and plant phenology is absent. Tree physiologists and hydrometeorologists have highly advanced understanding of these phenomena, and their discipline-specific models reflect that. These arguments are worth keeping in mind when you're noting the "uncertainty in the community on how to represent plant physiological controls on transpiration in hydrological and land surface models." Is it uncertainty, or just lack of interest/effort to study and implement models that reflect contemporary knowledge in plant physiology and boundary-layer biophysics? As an example, you go to great lengths here to incorporate small-scale non-uniformities into the subsurface flow domain, but you use a pretty standard version of the PM approach for ET that's been around for over 30 years now. You assume homogenous land cover in Colpach catchment, and you use vegetation parameters

from a non-local catchment in Germany, with phenology assumed invariant from year to year (I assume that's what you mean by "fixed"). With that model setup, it's really not appropriate for you to be talking about how uncertain the community is in how to represent the complexity of these processes in models. It wouldn't be very complex for you to considerably improve this standard version of the PM model just by using site-specific data, a dynamic representation of phenology, and to accurately reflect the relative abundance of forest versus other vegetation cover in the catchment. I don't want to be overly critical, because some assumptions and simplifications are always made in modelling. My main point is that you don't want to go on philosophizing about how we learn from still-uncertain models, if the model you're employing is not nearly state-of-the-art, or not parameterized nearly as carefully as it could be.

13) Line 198-201: Or stated more directly, "To assess the ability of a spatially-distributed, physically-based model to accurately simulate multiple state and flux variables (not just streamflow), when the parameterization of the model is based on observed catchment properties, not on an optimization algorithm."

14) Line 240-242: This concept certainly precedes the work of Zehe 2014. For example, consider these important papers, which are notably absent from works cited in your introduction:

[Berne et al., 2005; Harman and Sivapalan, 2009]

Berne, A., R. Uijlenhoet, and P. A. Troch (2005), Similarity analysis of subsurface flow response of hillslopes with complex geometry, Water Resources Research, 41(9) Harman, C., and M. Sivapalan (2009), A similarity framework to assess controls on shallow subsurface flow dynamics in hillslopes, Water Resources Research, 45 15) Lines 323-324: Please provide some explanation of what you mean by the onset of vegetation period? Presumably that would be timing of leaf development for crops and deciduous plants, but evergreen plants will be physiologically active even through winter and spring, albeit at lower rates than in summer. Then there is of course an extended pe-

riod when crops and deciduous plants transition from no foliage to the maximum leaf area they will obtain that year. Than transition can span weeks to more than a month. 16) Lines 355-368, and Figure 7: The variability among measured moisture retention curves for your soil samples is remarkable. Can they be logically grouped in any way, for example, by landscape position or soil depth? If so, a color scheme to illustrate that would be very interesting. Some additional detail about where, and at what depths, the soil samples were taken is needed.

17) Lines 385-388: Are the sapflow sensors collocated with rainfall and soil moisture measurements? Please elaborate on the exact type of sensors, depth of installation into trees, and other important details about their operation.

18) Lines 424-426: Any consideration of the soil-moisture dependence of vapor diffusivity in soil? It varies as a power-law function of air-filled porosity, over about 4 orders of magnitude.

19) Lines 432-446: How are the probabilities of the Poisson process determined? Is this based on some knowledge that informs your perceptual model, or are though chosen arbitrarily? Or, do you determine some best-performing parameters based on a sensitivity/optimization process? Please describe in a little more detail, and consider providing an image of what these structures look like in your final model domain. Is this what we're seeing in Figure 3C,D? If so, please just allude to that figure here in the text.

20) Lines 453-454: Considering the horizontal resolution of model elements is 1-m, I'm wondering how realistically the vertically-oriented, preferential-flow zones can be represented? Certainly the macropores in your photographs are not 1-m wide. Representing their tortuosity by vertically-offset grid cells of 1-m width seems like a gross distortion as well. Can you discuss how you rationalize this model domain and horizontal grid resolution, especially with regard to those grid cells that are imposed to represent preferential flow structures?

Also, can you provide detail about the dye-irrigation studies from which the photograph were derived? The pictures are very insightful. However, it is well known that irrigation studies often impose exceptionally high input fluxes, and with sprinkler systems that exhibit enormous spatial variability. Can you comment specifically on the irrigation rates and the spatial uniformity of the irrigation system, and how those irrigation rates compare to the frequency distribution of rainfall intensities that occur at these field sites?

21) Lines 465-466: I'm personally not familiar with the phrasing "free outflow boundary" and "gravitational flow boundary". Please clarify exactly what these mean, and maybe represent in an equation. Does "free outflow boundary" imply a seepage-face, where there is no flow until water-pressure head exceeds atmospheric pressure? And does gravitational flow boundary imply a zero gradient in soil-water pressure head (i.e. flow is governed by the elevation gradient and saturated hydraulic conductivity)?

22) Line 477: 40 – 45% porosity seems exceptionally high for a fractured bedrock. Any evidence to support that number? That's equally porous as most soils. It's hard for me to visualize how a cubic meter of schist could be 40% air space. Porosity for metamorphic crystalline rocks is typically reported to be less than 10%. I think this parameterization is patently wrong, and will have a significant effect on your simulation results since a majority of the domain is defined as bedrock.

23) Line 479: Here again, it would be good to know about the depth distribution of the soil samples collected for hydraulic characterization. Did any actually come from that depth?

24) Lines 486-490: This sounds wonderfully sophisticated. I have no idea what it means. I'm probably not alone in that regard. It seems like important information about how spatial heterogeneity of hydraulic properties are generated in the model domain, so maybe a sentence or twoâĂŤin plain languageâĂŤto build the intuition of the reader/s that are not intimately familiar with this jargon.

25) Line 528: Do you mean left boundary? The right boundary is no flow.

26) Line 527-530: You're simulating flow through a single 2-dimensional hillslope profile, so how can you compare the composite discharge (overland flow, subsurface flow, and deep drainage) to the measured streamflow for the whole 3-d catchment? Are you integrating the hillslope response over the entire 3rd dimension of the catchment? Please explain this. Also, do you think it is appropriate to include the deep drainage flux in this composite outflow when comparing to the stream hydrograph? It could conceivably travel through an aquifer system that discharges outside the boundaries of your catchment, no?

27) Lines 531-533: What exactly do you mean by "validated"? What's the difference between validating and benchmarking? The phrase "tuning against" is not readily interpretable.

28) Line 555: Left boundary?

29) Line 566-568: The important trend for the catchment water balance is the timing of the leaf area expansion, maximum, and decline (in fall). The leaf area is the dominant control on transpiration and net radiation. So, I don't understand how or why you change the timing of phenology without changing the temporal dynamics of leaf area. Please explain the rationale for this.

30) Line 594: Figure 9B, rather than 10B?

31) Lines 614-615: Are you talking about Weierbach catchment in Germany? It's irrelevant. I suggest you delete that and stay focused on your catchments.

32) Line 619-622: Run-on sentence that is very hard to interpret. Please rephrase. Also, please explain what inference you think is made possible by comparing NSE with log(NSE).

33) Lines 622-627: These statements are questionable. Are you claiming that infiltration-excess overland flow is occurring, or overland flow due to saturation ex-

cess? You say that the model erroneously generates overland flow in the summer in Wollefsbach due to convective storms. The saturated-hydraulic-conductivity parameter you report in Table 1 is 2.9 x 10-4 m/s, or about 1.8 cm/minute. This is quite a high hydraulic conductivityâĂŤwhat one might expect for coarse sand. Are your surficial soils sandy? Are those convective storms producing rainfall flux greater than 1.8 cm/minute? Seems doubtful storms like that would occur frequently. How is the model generating so much overland flow if the rainfall rates are (I assume) always considerably lower than the saturated conductivity? In Figure 9D it looks like your simulated-average-soil moisture is in excess of almost all the measured time series. Maybe you're way overestimating soil-water storage and generating saturation-excess overland flow, rather than infiltration excess overland flow. If that's the case, and if it's saturation-excess overland flow, then you can't immediately assume that enhancing Ksat to represent soil cracks is the next, necessary step to improving the model. You need to get the soil moisture dynamics correct before you can go off exploring that speculation. You're Ksat value is already pretty high, is it based on measurements?

Also, it's somewhat a shame that you have all those soil moisture observations, and a spatially-distributed model, but you only compare the mean-simulated soil moisture (I assume it's the mean) to the observations. The spatially-distributed model gives you lots of spatially-distributed results to compare to spatially-distributed observations. If you're just going to look at the average-simulated soil moisture, you're giving up all that detail that is provided by the model, and one has to ask why not just use a lumped model? You should compare simulated soil moisture at specific points in the landscape to the observed soil moisture at those same points.

34) Lines 631-635: I am quite confused by this statement. You are using a spatially-distributed model, so why are you claiming that it's unrealistic to expect the model to accurately represent the spatially-distributed nature of soil-moisture dynamics? It should be able to represent at least coarsely the spatial distribution of soil moisture, for example, differences in upslope versus downslope positions, or differences in ar-

eas overlying saturated bedrock depressions versus those areas where the bedrock roughly parallels the land surface. Again, if you don't expect your spatially-distributed model to accurately represent any of these spatial patterns (which are important for runoff), then why are you using a spatially-distributed model to begin with?

35) Lines 641-642: This statement is inevitably true for every catchment in the world, and hence does not rely on any measured or simulated soil moisture dynamics. Maybe just delete.

36) Lines 671-684, and Figure 10A: This material needs much improvement. First, measuring sapflow in trees is a delicate business, with major discrepancies existing between methodologies, and significant errors arising from inexact application of methods (e.g. due to radially-varying flux rates within the sapwood, a well-documented phenomenon in the tree physiology literature). There are several reviews of this topic in the plant physiology literature (e.g. Steppe et al. 2010, Agricultural and Forest Meteorology, 150). You have provided essentially no detail about the nature of your field-based sapflow measurements. Also, how are you normalizing the measurements?

Second, you use a version of the PM model to simulate water vapor flux from the plant canopy, not sapflow (L3 T-1). The two cannot be assumed to be equal. If you consider the tree as a system spanning the point of your sapflux measurement (breast height on the stem) to the canopy, then the sapflow (input to the system) only equals the volumetric flow out of the leaves (outflow from the system) if the system is in steady state (i.e. inflow = outflow and storage is constant). Water storage in the tree stems and canopy foliage is dynamic. I am not sure how good or bad is the assumption of steady state in your system, but it is certainly an assumption you should carefully consider and provide some justification for why this comparison (between measured sapflow and modeled canopy vapor flux) is valid. Without that, you should probably omit this text and figure 10A from the manuscript.

37) Lines 691-692: You might be careful in projecting your own expectations, surprises,

and uncertainties onto your readers. The result you describe here is not counterintuitive to me; it's exactly what I would expect, for the exactly the reason you state. Higher gradient = more rapid drainage = less persistent storage (assuming all else is equal, which is what you've assumed for this virtual experiment).

38) Line 694: "is" rather than "and"

39) Table 2: Please provide some rationale for why you use multiple error metrics (e.g. NSE, KGE, logNSE) instead of just one. It's just confusing to the reader when you talk about quality of results in one case using KGE, and in another case using NSE. Also, the different metrics show different sensitivities to the domain-changes utilized in the virtual experiments. Why? Which one is most appropriate in light of those differences? You're using these various metrics to make inference about the relative importance of different model features, so you need to argue why one or the other metric is better. Or just use one metric for clarity.

40) Line 739: Sentence is unclear, please rephrase.

41) Figure 8D: Here, and in some other cases, the reader cannot see much of the observed and simulated dynamics because of the relatively long time scale of the x-axis, and the flash nature of the catchment. Consider using an axis break on the x-axis, or some other mechanism to expand the time series, so the reader can clearly see the dynamics. In Figure 8D, it's impossible for me to see what differences might exist between observations, and the standard and emergent-structure scenarios.

42) Line 756: Rephrase and omit use of "one to one". It implies exactness in the representation of scale, which is not the case. For example, your model grid cells are much larger than the preferential flow structures they are modified to represent.

43) Line 766: Again, careful with projecting your own reactions and perceptions onto your reader. The use of the word "astonishing" here seems hyperbolic to me.

44) Lines 780-781: Why is that remarkable? It's a predominantly upland catchment

with forested hillslopes. Was it your initial expectation (null hypothesis) that the model would be incapable of simulating streamflow?

45) Figure 1: Please add a color scale so we can see what are the associated elevations.

46) Lines 796-801: These statements are not very clear. You state, "We also found that benchmarking of the model against sapflow data provided additional information about the representation of vegetation controls, which cannot be extracted from the double mass curve or discharge data." Exactly what information are you talking about? Your Figure 10A basically just shows that your modeled water vapor flux from the canopy (based on PM model) trends in a similar way as field-based measurements of sapflow within tree stems. That's certainly to be expected, since both are driven by net radiation, but as noted in comment 36 above, it is not necessarily meaningful to compare magnitudes of those two fluxes, because they're not the same thing. So, the sapflow data don't tell you definitively that your simulations are right or wrong. Also, you say that this additional information could not be extracted from the double mass curves. Again, exactly what information are you talking about? Certainly the double mass curves in Figure 12A show a clear effect of changing the timing of bud break. The statement in lines 799-801 is presented as a conclusion, but it's not really a conclusion is it? You knew from the beginning that using a spatially-distributed, highly parameterized model would offer the opportunity to incorporate knowledge from soil-hydraulic measurements and geotechnical surveysâĂŤknowledge that would not necessarily be incorporated explicitly in a lumped model?

47) Lines 802-804: Are you so sure this can be concluded? It seems to me that if you want to advocate the use of highly parameterized, spatially-distributed models for the sake of learning about catchments, you need to illustrate that the model is accurately representing some of the spatial dynamics in the hillslope (or catchment that is a composite of your hillslopes). In those cases the matching between simulated and observed averages is not that great for soil moistureâĂŤthere are systematic errors

in all cases (Figure 9A-D). By comparing average-simulated soil moisture for the whole hillslope to the average-observed soil moisture for the whole catchment, you're failing to rigorously test the spatially-explicit predictions made possible by the model. You should try to show that the model actually properly represents spatial variability in soil moisture, saturated-zone expansion, hydraulic gradients, etc. If not, then it's hard to argue that the distributed model teaches us anything more than we would learn from a lumped model.

48) Lines 812-816: I fully agree with this statement. You don't need a high-dimensional, spatially-distributed model if all you want to do is predict runoff at an annual timescale, or even at shorter time scales. Use a transfer function, maybe even a time variable transfer functionâĂŤyou will still have vastly fewer degrees of freedom than in the spatially-distributed Richards equation model. But doesn't this statement contradict the overall message of your paper, that those more complex models are needed for learning about catchment functioning? The spatial variability of soil-hydraulic properties may be quite important, in fact, for properly simulating all those runoff peaks in the summer, where your model does quite poorly (Figure 8B,D, and Figure 12B).

49) Lines 823-825: Unclear, please rephrase.

50) Line 823-844: I would suggest you delete all of this. It seems wildly speculative and I have no idea how, based on the analyses performed in this paper, you conclude that, "equifinality and the concept of a representative hillslope is rather more a blessing than a curse since there is an infinite number of possible macropore setups which yield the same runoff characteristics. If this were not the case, we could not transfer macropore setups from the literature across system borders and successfully simulate two distinct runoff regimes which are strongly influenced by preferential flow." An infinite number of macropore setups that yield the same runoff characteristicsâĂŤwhat are you talking about here? You tested 2 such scenarios (Figure 3C,D). Your model does a fairly poor job at matching runoff peaks at many times of year. Those peaks are the hydrological attribute most likely to be influenced by the activation or latency of preferential flow

paths. When you say you that you "successfully simulate two runoff regimes" I presume you are talking about the double-mass curves, because your simulated hydrographs show significant errors at many times of the year.

51) Line 864-866: Do you mean, "below which", instead of "above which"? I wouldn't spend much time on that. By changing slope and nothing else, you're vastly oversimplifying how soils, geology, and geomorphology affect streamflow, and how all those variables are related to topography in naturally evolving landscapes.

52) Lines 888-889: How do you justify this statement? Did all of your virtual experiments where you manipulate the bedrock topography have an equal volume of depressions? And if so, how do you go about quantifying the volume of depressions in an undulating rock surface? What constitutes a depression or a high point, versus a portion of the rock surface that is part of a datum plane?

53) Lines 867 and 895: You're using the questions as section headers, but the subsequent content does not answer the questions. First of all, define what you mean by "first order control". Do you just mean that the response variable (annual runoff ratio, or other?) is a linear function of the independent variable (bedrock topography, or vegetation)? If so, do your data corroborate such a linear relationship? I'm not sure you can say, based on the limited scenarios of bedrock topography you tried. You would have to come up with some quantitative metric distinguishing one bedrock scenario from another. In terms of vegetation, all you did was try 2 different times for bud break. Can you discern a linear relationship between "vegetation" and some response variable based on these tests?

54) Lines 912-923: Of course you can! You can setup heterogeneous rainfall inputs at the soil surface in your model domain. You can setup different scenarios of incoming solar radiation along the hillslope domain to emulate aspect-related differences in the radiation balance, water budget, and possibly soil hydraulic/or geological characteristics. I'm really struggling to understand how you continually advocate for spatially

distributed models, but continually state that one can't expect them to accurately represent spatially-explicit hydrological processes. If you don't expect a spatially distributed model to accurately represent spatially-explicit hydrological processes, then why use it, instead of a lumped model?

55) Lines 937-969: I would suggest deleting this to shorten and focus your discussion. There is not much reference to your analysis in this section, it's a little bit of a ramble, and you don't say anything about the Richards equations that hasn't already been said many times before over the last 70-80 years.

56) Lines 971-993: I recommend you remove this from the manuscript. You're pontificating about all the ways the land surface models must be fundamentally improved for hydrological modeling, and in doing so you're demonstrating that you have no awareness of the related disciplines of hydrometeorology, plant physiology, and biophysics. All the phenomena that you imply are important, for example, "implies that phenology evolves in response to climate and hydrological controls, thereby creating feedbacks" are in fact known to be important by people in those fields, and others. The upgrades to our model representations that you suggest should happen, have in fact happened, and continue to be upgraded, for example, models that link plant metabolism and water use, or that utilize spatially- and temporally-dynamic root uptake schemes. You conclude by saying that the literature is full of more realistic models for parameterizing stomatal conductance, but you still use a fairly standard version of the PM model with non-local parameters. So I don't think your analyses are very relevant to the state of practice in evapotranspiration modeling. I think you should stay focused on the representation of runoff processes.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-307, 2016.