Synopsis:

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? If so, this supports the notion that a "representative hillslope" is a sufficiently good representation of an *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 partitioning from 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 authors did a nice job of revising this paper from its original version. The organization of the paper now seems much more coherent, and it was a pleasure to read.

I recommend that this paper could be accepted for publication in *HESS* with only minor remaining revisions. I urge the authors to reconsider the phrasing of their hypothesis, and some instances where they discuss the hypothesis later in the paper. They claim to test whether a representative hillslope is "the most parsimonious" representation of a catchment. I think they have gone on to show that a representative hillslope can be used to reasonably represent streamflow generation processes for a whole catchment. But, to say that the hillslope is the most parsimonious representation would require testing the representative hillslope against some other alternatives. That was not done here. I think it is important to clarify this point, but I don't think it will require major revision—just some changes in phrasing (see comments 3 and 13). My other comments are mostly technical details. I urge the authors to consider them all carefully in their revision.

Specific Comments:

1. Page 4; line 118: Can you indicate what land areas are associated with the "lower mesoscale" designation?
2. Page 4; lines 124-128: Maybe consider rephrasing this. It seems the main point you're trying to make is that it's not really possible to develop a model that accurately represents a hillslope, or an average of hillslopes, since we'll always lack sufficient knowledge of the subsurface. As such, we have to develop a simplified conceptual model. As it reads

now, you say "cannot be a simple copy of a real hillslope", which is a little off, because a copy of a real hillslope wouldn't be simple at all.

3. Page 5; line 142: Do you really test this hypothesis in this paper? If so, please indicate to the reader here how you will determine if the hillslope is the *most* parsimonious representation of the catchment, rather than any other conceptual representation, or analog measurement, such as bedrock topography or porosity, soil permeability, catchment area, net radiation (i.e. Budyko model), etc. Please state clearly to the reader what will be the basis for accepting or rejecting the hypothesis. I would argue that the stated objectives that follow—while worthwhile and interesting—do not provide a rigorous test of this stated hypothesis. It seems the work is more objective driven than hypothesis driven. I don't think this is an impediment to publication of your work (most hydrological science is objective driven, rather than by hypothesis), but whether or not you're testing a hypothesis, and on what basis, should be crystal clear up front.
4. Page 9; line 285-287: I think you should put appendix A2 in the main text. From Figure 7 I notice that many of your soil samples have porosity > 0.6. That's quite high porosity. Are the soils high in clay content, and very low in bulk density?
5. Page 10; lines 314-317: Why do you use this proxy instead of actual sapflow [L^3 T^{-1}]? You have the radially-distributed measurements of heat dissipation across the xylem tissue, so you apparently have what you need to integrate and get reasonable volumetric flows. So why do you use this proxy? Also, you say you use a 12-h daily mean. Mean of what exactly, hourly-maximum velocities, or something else? It's not clear what you're averaging. I encourage you to put these calculations in equation form and report the dimensions, so this is all completely clear to the reader.
6. Page 11; lines 345-351: Excellent, thank you! This is clearly explained and much easier to understand than the previous version.
7. Page12; lines 374-375: Usually a seepage-face boundary is a no-flow boundary except at nodes where pressure head is greater than 0 (i.e. greater than local atmospheric pressure). Is that what you mean here?
8. Page 15; line 476: Can you explain further what exactly was done here? It seems one of the key points of your paper is that you don't use inverse optimization procedures to defined best parameters for your model. Rather, you develop all your parameter sets based on measurement or perception. When you say that both hillslopes were setup in a few "test simulations", does this mean that you actually did optimize the parameters based on iterative simulation results and your error metric?
9. Page 16; lines 499-500: Maybe another sentence or two here to explain a little further what this means?
10. Page 17; lines 513-514: "12-hour-rolling median of *daily* sap flow velocities" is not clear. Can you put this into an equation so it is explicit? I'm still not entirely clear why you don't use your field measurements to quantify *sap flow* [L^3 T^{-1}] across time, rather than looking at maximum values of *sap flux* [L T^{-1}]. If you quantify *sap flow* using the field measurements then you can compare it more directly to simulated transpiration. I use italics for emphasis here. You may want to adopt the sap flow and sap flux phrases for volumetric flows and fluxes, respectively. For example, you refer to "oberved sap flow", but from the observations you're only using fluxes, no? And comparing those to simulated transpiration?

11. Figure 11: It seems the model systematically overestimates soil moisture—being above the range of all observations in some cases. The measured moisture-retention curves you present in a figure are notable in that the porosities seem very high (>0.6). I'm wondering if, under the wettest observed conditions, you ever see volumetric water content measurements as high as that in the field? Figure 11 suggests your field observations of volumetric water content almost never exceed 0.4. Any chance there is some bias in those lab-measured moisture retention curves toward excessive porosity? If so, and if you reduced the porosity parameter in your model (which is based on those lab measurements), your simulated time series of soil moisture should come down closer into the range of what you observe in the field.
12. Page 20; Section 4.4 and Figure 12: I would recommend that you just omit this section and Figure 12. From the double mass curves it is clear that your model is reasonably approximating the catchment water balance at daily to seasonal time scales. Implicitly this must mean that your transpiration routine is approximately correct. Looking at these time series doesn't add anything for me. For one, even though you have normalized both variables, the comparison is still possibly confounded. The total water storage within tree stems can change (tree physiologists regularly measure this). You could have similar measured flux on two days, but slightly different volumetric flow out of the leaves (transpiration), if on one of those days part of the flux contributed to storage change wtihin the stem, rather than direct discharge from leaf to atmosphere. Also, it remains unclear exactly how you derive those daily flux estimates. Providing equations might help clarify, as noted in previous comments. Omitting this section and figure would also shorten the paper a bit, which I suspect will increase your readership. You have good depth of content even without this section and figure—to me it's just a distraction. Your decision though.
13. Page 21; lines 638-640: Here, as well as where you discuss your hypothesis in the introduction, you're referring to a "representative hillslope" as the "most parsimonious representation" of a catchment. I think you should reconsider the phrasing. Your model is not too parsimonious—it's fairly complex. A transfer function in an instantaneous-unit hydrograph approach would be much more parsimonious, and might predict time series of streamflow, and the double-mass curve, equally well. As I noted regarding your stated hypothesis, you haven't really examined other null or alternative hypotheses. I think it is more appropriate to say that you "tested the hypothesis that a representative, 2-d hillslope—conceived and parameterized based solely on observations—*can reasonably embody* all the pedological and geomorphic complexities that control streamflow generation at the whole catchment scale." To support your claim that the hillslope is "the *most parsimonious* representation of a catchment", you would have needed to compare this representation to some others, but that wasn't done here.
14. Page 22; lines 700-702: I agree, and your sensitivity analysis seems to support this.
15. Page 23; lines 730-733: Maybe clarify what you mean by "groundwater". You do have unconfined aquifers in the catchment, no, with full saturation and positive pore-water pressures, even if they are transient? I presume these exist in the riparian areas and extend upslope during large storms? It will help to more specifically distinguish "groundwater" and "subsurface storm flow".
16. Page 25; lines 767-770: Again, regarding your moisture retention curves, in the model I see you parameterized them with n values of near 1 within the Mualem-vanGenuchten

hydraulic model. That is an exceptionally low n value, and will result in very strong suction pressures even under relatively small declines in volumetric-water content. Possibly worth looking back at that parameter selection (also see comment 11 above).

17. Page 25; line 771: It's hard to image this being a problem consistently for so many probes. If you wet the soil and compress it around the sensor before installation, you normally have sufficiently good contact to avoid major errors due to air space. Also, the soil always shrinks and swells somewhat during wetting and drying, so voids that might exist due to disturbance during installation should become compressed in fairly short time.