Hydrol. Earth Syst. Sci. Discuss., 6, C2598–C2604, 2009
www.hydrol-earth-syst-sci-discuss.net/6/C2598/2009/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Iterative approach to modeling subsurface stormflow based on nonlinear, hillslope-scale physics" *by* J. H. Spaaks et al.

J. H. Spaaks et al.

jspaaks@uva.nl

Received and published: 13 November 2009

We would like to thank Anonymous Referee #1 for taking the effort of writing a clear and concise Interactive Comment. We feel that the manuscript has benefited greatly from the constructive feedback we received. In the sections below, we cover specific issues in detail. The Referee's comments are verbatim in bold face; our comments are in normal face.

1 Major points

1) The authors use a Darcy-law approach to compute lateral flow (Eq. 3), doesn't this contradict to previous findings (including papers by McDonnell!), which highlighted the importance for macro-pore flow at this hillslope?

The paper is intended to start out from a very simple formulation of subsurface stormflow, which is then improved in subsequent steps. After the first step, it became clear that $Model_1$ was not able to reproduce the observed discharge curve, so we needed to introduce some other process. Since macropore flow —or preferential flow in general— has been shown to be important at Panola, this was a likely candidate. Doing this, however, was easier said than done, since we do not have much information on how far the soil pipes at the Panola trench face extend uphill, whether or not such features occur elsewhere in the hillslope, what their diameter, length, and tortuosity is, and at what depth they occur in the soil profile. In addition to preferential flow at the plot scale, we also had to incorporate preferential flow at the hillslope scale: patches of transient saturation have been shown to connect to the trench face, effectively forming a hillslope-scale preferential flow path. To account for these preferential flow processes, we used travel time distributions in $Model_2$ and the subsequent $Model_3$.

2) Even if we assume that flow can be simulated using Darcy law, I find it surprising that Ksat is assumed to be constant with depth. Many papers have shown that Ksat often decreases with depth, and I am a bit surprised that the authors do not test this. Looking at the hydrograph, I would guess that a different formulation for the Ksat-profile would have improved model 1 considerable. By not testing this, the question remains whether the more complex models are needed because of spatial heterogeneity (as discussed by the authors) or a wrong assumption on Ksat. (a similar argument can be made for the choice of g(t) in model 2))

The Referee suggests that the discharge curve from $Model_1$ could have been improved by allowing K_{sat} to change with depth. Such a model formulation could indeed have been an intermediate step in the Iterative Research Cycle, and would undoubtedly have improved the performance. The main reason why we did not include a depth-variable K_{sat}, is that we wanted to test a model concept that allowed for the existence of a hillslope scale preferential flow path of transient saturation (see also our comments on the previous issue) within one spatial element. The existence of such a flowpath at Panola had become apparent from the research of Ilja Tromp-van Meerveld and her co-workers.

3) I am not sure I understand the physical reasoning for the travel times in model 2. As I understand the authors, the motivation for g(t) is the delay of the water entering the soil storage (i.e., vertical flow). I would argue that for shallow gw-systems, such as in this study, delays are rather caused by lateral flow than by vertical flows.

As the Referee points out, the motivation for $g(t)$ is the delay of water before reaching the saturated storage. This is indeed a *lateral* delay, because we assume the saturated zone to be spatially discontinuous, as depicted in Figure 3B on page 5238.

4) Why do the authors limit themselves to simulating only one event? This is a severe limitation and limits the value of the results considerable. With only one event I don't think we can draw any general conclusions. Previous work at the Panola site present an impressive data set of many events (e.g. Freer or Tromp-van Meerveld), and I can't see any valid argument why the analysis presented in the present paper should be restricted to using only one event!

Generally speaking, calibration using multiple events essentially amounts to compensating for errors in the initial and boundary conditions (as well as allowing the parameters to compensate for errors in the model structure). This is why, for prediction

purposes, it is important to use as many events as possible. However, when trying to 'finger down' into how a model result came about, it is not necessary to do this. In fact it may even be better to have a short period of time. For instance, none of our models would perform well when the summer season is included, simply because we have not yet incorporated the processes relevant during summer. Since it is impossible to manipulate the hillslope system as we would do in a laboratory experiment, limiting the simulation to a short time period is the next best thing. While this implies that one cannot extend the conclusions to the summer season, it is helpful in separating the effects of different processes.

Nevertheless, it has become apparent from the reviewer's comments that some clarification was needed in the description of our experimental design's rationale. In the revised version of the paper, we therefore included a section similar to our explanation above.

5) I must admit, I am a bit confused about the model setup. At first I thought that the simulations were done using a grid (e.g., P5213, 14: drainage direction in steepest gradient direction), but then I missed any information on, for instance, grid resolution or numerical schemes. Later I also got the impression that the first three models were lumped, since the authors describe the difference of model 4 as the hillslope being represented by 8 elements. Clarification is needed!

We noticed that Referee #1 is understandably "confused" by the text on page 5213. In the revised version of the manuscript, we have added some text to clarify our concept. Also, we deleted the reference to the flow direction algorithm on page 5213, line 14–15. While the models can be set up in lumped, transect, or 2-D spatially distributed mode, it is not necessary to include this reference here, since we only use lumped and transect setups.

6) Basic information about the model application seems to lack. What is the time step for the simulations? How have initial conditions been specified?

We added the model time step as used by each model formulation to Tables 1 and 3. The initial conditions have been specified there as well.

2 Minor points

I would recommend not using terms like field capacity, which have a clear meaning in soil hydrology but are used in a different way in this study. The authors use, for instance, the term field capacity for a moisture content at which no further drainage occurs, but their value is independent of depth and height above groundwater level. In reality, the water content for the no-drainage condition obviously varies with depth and is related to the position of the groundwater table. Please note: my point is not that the approach used by the authors is wrong (we often have to use 'wrong' formulations as useful approximations) but that the terminology is misleading.

We changed this in the revised version of the manuscript to avoid confusion.

The structure of the paper with its jumping between methods, results and discussions is a bit awkward.

The 'jumping structure' as used in the paper is perhaps a bit unconventional, but we believe that it follows naturally from adhering to the Iterative Research Cycle. We believe that, from a reader's perspective, it would be confusing to lump together all the model descriptions, discussion parts, etc, in their separate sections. At the very least it would take away from the emphasis on learning from exactly in what way an experiment (model formulation) went wrong, and how it is (perhaps) improved for the

next cycle.

I think we should be a bit more careful with the term 'virtual experiment'. There are certainly cases where a modelling study can be called virtual experiment for good reasons, here I don't see why we need this fancy term. Different model structures are tested, but I don't see any 'virtual experiments'.

The term 'virtual experiments' was coined by Markus Weiler and Jeff McDonnell in their 2004 paper. The authors present an approach with which first-order controls on flow and transport at the hillslope scale can be identified. They compare results from 2 model setups that use a different drainable porosity. This is somewhat similar to a sensitivity analysis, in that the effect of changing the value of one parameter is isolated. It also goes beyond a simple sensitivity analysis in that it evaluates in what way specifically the model results differ. Their paper aims to better understand why the model is doing what it is doing and how that relates to observable patterns in the real world. This is exactly what we were aiming for with our virtual experiments, in particular for the transect case (virtual experiment 4). From a philosophical perspective, the only difference is that we are comparing model structures, rather than parameterizations.

P 5218, 23-24: an efficiency of 0.85 is rather high, BUT of course here this value is computed for only one event. One might question whether the efficiency which is usually used for continuous simulations is appropriate for evaluating one event.

We agree with the Referee that the Nash-Sutcliffe efficiency is more commonly used for describing the performance of longer time series. However, we use the Nash-Sutcliffe efficiency as an objective measure to intercompare the model formulations, and to assess whether model changes resulted in actual improvements. We do not make any explicit comparison between the performance of our model formulations and those reported in the literature. Also, we do not focus solely on the event-aggregated Nash-Sutcliffe criterion, but we split the hydrograph into various parts. The behavior in

these parts is then scrutinized separately, albeit in a qualitative sense.

In the revised version of the paper, we will make sure to emphasize that our efficiencies should be interpreted as a measure of how well each model could be calibrated to the event, and that care should be taken to interpret it in terms of the models' predictive power for longer time series.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 6, 5205, 2009.