Hydrol. Earth Syst. Sci. Discuss., 10, C373–C377, 2013
www.hydrol-earth-syst-sci-discuss.net/10/C373/2013/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Resolving structural errors in a spatially distributed hydrologic model" *by* J. H. Spaaks and W. Bouten

Anonymous Referee #1

Received and published: 17 March 2013

The paper presents an innovation in model diagnostic approaches which makes use of the information contained in state updates of a Kalman filtering approach, analysing these updates both, as time series and depending on model state. The approach is demonstrated for a virtual catchment with the SWMS_3D soil model and the SODA framework using 105 error-free, virtual measurements of subsurface pressure head, an error-free, virtual observation of vertical water loss and an error-free, virtual observation of hillslope scale discharge.

The paper presents an interesting and novel idea with high relevance to the readers of HESS. I would like to congratulate on the idea to make use of the Kalman state updating information for model diagnostic and the way how the approach is presented in a clear, well structured way. The main concern arising when reading the manuscript

is whether the approach will work in a real world case. The design of the study makes a number of assumptions that are unnecessary optimistic. These are mainly the high number of measurements, measurements available exactly at the location of the hot spots and the error-free, virtual observations. The authors should present data documenting influence of these three factors on the performance of the approach. For example, what happens in the presence of systematic or random errors? What conclusions are drawn if only a limited number of for example 10 head measurements are available? What conclusions are drawn if the measurements do not coincide with the hot spots. These answers can all easily be answered within the experimental framework of this manuscript. If space limitations are of concern, the study would not loose much without the part on SCEM-UA. The core message does not depend on this part. Another question, which might be answered in a subsequent work is the influence of multiple, interacting model deficiencies on the interpretation drawn from this approach.

Specific comments:

- The title is too general and should be a bit more specific.

- Some references related to diagnostic approaches are missing (see at the end)

P 1826, L10: "forward model" and "inverse model" - the rational to choose these terms is not very clear. Why not using "reference model" and "test model" instead?

Fig 2 and Fig 4: as this is a virtual setup, presenting spatial distributions as smoothly distributed variables is somewhat misleading. I suggest to present rectangles for which a node is representative with the corresponding values.

P1827 L10-25: Do you use homogeneous soils and K values? Does this affect your main findings?

Fig 4: Please also refer to the symbol r_{sink}.

P 1828 L20: Explicitly state what the reference level for pressure head is. The method to obtain the initial state is not described sufficiently clear. Where do spatial hetero-

geneities in the initial state in Fig 5 come from? Is this due to spatially heterogeneous soil depths and r_{sink} values only?

P1829 L11: Related to the insufficient description of the initial state procedure: It is not sufficiently clear why 188 m^3 of water is present in the soil at the initial state.

P1829 L 20: This is a very rich set of observations that is hardly available in a real case catchment. The study will benefit much if a reduced set of pressure head observations is used in an alternative scenario.

P1830 L 9: Be explicit about how you treated the other parameters. Do you assume perfect knowledge about these? How does this influence results?

P1833 L14: Provide information about how to interface SCEM-UA and the model. Give exact information about implementation and version numbers for the model and SCEM-UA. Same for SODA further down.

P1835 L6: Based on the values for r_{sink} and the number of nodes, it should be possible to calculate a weighted average - this might be what we observe for r_{sink}. What else do you expect than a value in between the two extremes for r_{sink}?

Figure 5,7 and 9: Mark hot spots with an asterisk or similar. Use linear legend as the square representation suggests a bivariate color schema, which you are not using.

P1835 L11: How did you determine the spatial auto-correlation? Please report the measure for the spatial auto-correlation.

P1835 L20-21: Results supporting this argument are not clearly presented. I have a hard time to see this in Fig. 7.

Section 3.1: Your argumentation is not very convincing: If you want to improve understanding, you would make good effort to better understand what is going on in subsurface. A uniform leakage would not be a good assumption for this. Also $r_{sink} = r_{sink(low)}$ may be a good assumption for 93 out of 98 nodes, but not for the overall model. I would

suggest to leave out the entire part about SCEM-UA.

P1836 L1-5: Present the results of leave completely away

P1836 L 8: not very clear what you mean by "activation of response modes" – nowhere introduced.

P1836 L11-15: Not very clear how this derives from the results shown.

P1836 L 28-30: This thought, while interesting, could be made clearer by being clearer about some underlying conceptual ideas. For example introduce before, how and when patterns in residuals are related to physical processes.

P1837 L13-20 How were implicit sinks treated in the objective functions - where they neglected? What values do you get for OF1? Please present influence of the different OF on the selection of the parameters.

P1838 L7 Briefly state that you will explain nodes that need updating but are not hotspots a bit later.

Section 3.2.1 Title is not well chosen. The first part of the section is not related to "experimental design". In general, I find your suggestions for experimental design not very helpful - mostly what you are saying is: "make as much and as reliable measurements as possible". Could you try to make your recommendations in view of limited budgets? Is it better to use only few reliable (small errors) or a larger number, not so reliable sensors (larger errors)? If logger space is limited, is it better to make more frequent measurements of should longer periods be measured? (you are contradiction yourself within the manuscript. P1839 L 18: measure more frequently; P1836 L10: measure longer). Check your recommendations with experimentalists for the revised version or leave them out.

P1839 L30: if X15Y39 never shows the behaviour $B = r_{sink(high)} * h$, is it correct to speak of a hotspot then? In my view, in case of h < 0 for all times you can not distinguish the two kinds of nodes.

General: Provide virtual observations and results from the deficient model as supplementary material, for others to test their method on your example.

Additional references related to diagnostic approaches:

Bastidas, L., T. Hogue, S. Sorooshian, H. Gupta, and W. Shuttleworth (2006), Parameter sensitivity analysis for different complexity land surface models using multicriteria methods, Journal of Geophysical Research, 111: D20101

Reusser, D., and E. Zehe (2011), Inferring model structural deficits by analyzing temporal dynamics of model performance and parameter sensitivity, Water Resources Research, 47(7), W07,550.

Sieber, A., and S. Uhlenbrook (2005), Sensitivity analyses of a distributed catchment 738 model to verify the model structure, Journal of Hydrology, 310(1-4), 216-235.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 10, 1819, 2013.