Dear Dr. Bahremand,

let me first express that I am pretty excited that your manuscript excellently serves the purpose of an opinion paper to stimulate a vital debate. In line with the three reviewers I am very supportive concerning the tenor, title and abstract of the manuscript. In line with the reviewers I think that parts of the manuscript should be thoroughly reworked. This is not so much a point of “being wright or wrong” (as this is an opinion paper) but rather to write the revised manuscript as a synthesis by addressing the reviewer comments/ antitheses and your thesis in a dialectic sense.

As this opinion paper touches such an important topic you may forgive me that I add a couple of points, not mentioned by the reviewers, that might be helpful for sharpening the revised manuscript and particularly to make title, abstract and content more consistent. As I regard myself as process modeler and I am physicist by training these points have a certain bias, again I ask for your patience.

First of all I am astonished that, although your advocate we should focus on physics, nearly all your examples deal with bucket models such as Flex derivatives or the HBV, which I would not rank as physical (Hubert Savenije may forgive me). None of these models closes the energy balance and the momentum balance in the catchment, they all exclusively focus on the catchment mass balance. As all these balances are tightly linked in nature, a physical model implies a joint treatment.

I am a little astonished, that a consistency check in parameters, in a sense that the average travel time through the fast reservoir needs to be smaller than through the slow one, or that channel roughness is smaller than roughness of the flood plain, is reported as new insight. To my experience this part of good engineering practice and is reported the internal guidelines how to setup up a hydrological model I found in many engineering companies.

Personally I have the highest respect for Shervan Gharari and particularly Hubert Savenije. During the mentioned study (Gharari et al., 2014) they constrain the runoff coefficients in their model using annual observations for average and particularly dry conditions. This is a pretty strong constraint for the water balance, and does of course imply that the remaining model parameters require less calibration. There is nothing wrong with that, but one cannot claim this as independent, a priory assessment. In fact they use prior information on the runoff processes they tend to simulate. For me this is circular reasoning.

Is Grey Nearings comment “that we a priory know that all our model are imperfect” indeed so much surprising? Any theory and thus model is an empirical fit (or inference) on/from a class of phenomena we characterize with observations. A superior theory characterizes a wider class phenomena using the same amount of or even fewer “laws” (mostly in form of equations). This implies that the scope of any theory and thus also model is limited.

I would like to encourage you to be more concrete with your notion of the “new theory” and “new thinking”. To my notion a theory should draw from first principles/theorems and those observables one can currently assess or assess in the near future. Your study sets hydrological modelling pretty much equal to rainfall runoff modelling, in fact input output modelling. This implies rainfall and discharge (and maybe also radiation) are the prime observables. Do we indeed need so much of a new theory when focusing exclusively on rainfall runoff model – simple models explain 70% to 80% of the hydrograph (Although, in fact this is rather a manifestation of Weiherstrass’ approximation theorem than a matching for the wright reasons and the fact that we assume the subsurface watersheds to coincide perfectly with the surface watersheds, which is not so clear to me). Are the remaining 20 % of stream flow variance indeed a scientific challenge? Or do we waste or time with a problem with a very flat learning curve? Notwithstanding that errors in streamflow data are surely 10%.Consistent predictions of integral response behavior and distributed dynamics of storage an ET (patterns) in a terrestrial system (not necessarily a catchment) is already not so straight forward, dealing with non-stationarity, hydrological system adaption and feedbacks on process regimes is a cardinal challenge. The latter two challenges are where more physics and more ecology in our models can provide an added value for learning – but not for streamflow (unless we deal with water quality).

A last little thought on theory – many scientific disciplines have a commonly upon agreed set of equations for their models, which of course also gets closed by empirical parametrization (for instance on shallow turbulence or the stomata conductance). The perceptional model in this disciplines starts, as in hydrology, with discussing surface and subsurface structure, vegetation and how to best represent this in the common set of equation. I still think that the REW ideas, despite all is drawbacks and partly errors, was a good attempt because it tried to establish a common set of equations for the energy, mass, momentum and (entropy). We all know the challenge is in assessment of closure relations and that the proposed zero dimensional approach is too simple as it averages across different ensembles and across driving gradients. Nevertheless, I think the key is to agree on an improved set of common equations and join efforts to close them instead of dispersing our intellectual power within more than 100 different hydrological models (which often differ more with respect to their name that with respect to the implemented concepts).

I hope you find these points helpful to sharpen your manuscript and look forward to the revised version.

Erwin Zehe