Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/hess-2022-281-RC1, 2022

Comment on hess-2022-281

Keith Beven (Referee)

Referee comment on "Why do our rainfall–runoff models keep underestimating the peak flows?" by András Bárdossy and Faizan Anwar, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2022-281-RC1, 2022

I am always a bit wary of papers that compare models with models in order to constrain the uncertainties involved in the modelling process. I fully understand why they do so – it is so difficult to get any grip on the real uncertainties in model inputs and "observed" stream discharges and other sources of epistemic uncertainty. But the question is whether what is learned is more than what is lost by not addressing the real problems of model uncertainty directly.

Here the reference model is the SHETRAN model driven by a relatively dense network of raingauges. How snow accumulation and melt is handled in the observations is not made clear. SHETRAN is described as a physically-based model, but its physical basis is wrong, particularly at the 1km grid scale used (Richards equation with no account of sub-grid variability, effects of assuming effective parameters at the grid scale). It was already described as a lumped conceptual model at the grid scale nearly 35 years ago (Beven, JH1989). It is not made clear how the daily time step is implemented for the reference model. It would have to have a smaller internal time step, so are the inputs simply averaged over the day? And any surface runoff production surely occurs at much smaller scales than 1km? Does that not already suggest and expectation that the reference will already underestimate peaks in unrealistic ways (unless compensations exist, for example in the routing)? As a virtual reality this does not affect the current study (there was no need to see whether this was a valid model of the catchments) – but as an example of the hydrological expectations we might have BEFORE making such a study, it is surely relevant.

And talking of expectations, it is stated that the subcatchments of the Neckar used were large enough to allow for daily time step simulations. The hydrographs shown in Figures 9, 10 and 13 really do not support this. Again, I understand the use of a daily time step so as to maximise the precipitation stations available but the discretisation effects would suggest that this is surely not properly reflecting the hydrology of these catchments. Equally HBV is run at the 1km scale, but it appears as if the outputs from each grid square are simply added, with no account taken of distance from the outlet (and the carry-over from day to day this might produce).

So we are left with the conclusion that a model might get better (in representing an incomplete virtual reality) as the number of gauges to define the inputs increases. A model using spatially distributed inputs will generally do better than one using an average input (all other things being spatially equal, which of course, in general they are not). But we knew that already, so have we really learned anything new about WHY our models "keep underestimating peak flows"? Except that, as shown in Figure 13, they do not – the global recalibration of HBV, given a set of inputs, can overcompensate for some storms.

We know there are subtle interactions between different types of model calibration, time steps, routing methods, parameter sets and their complex interactions, and sources of epistemic uncertainty, including both inputs and the rating curve (this study eliminates the discharge uncertainties by design but these can be important in practice). So if we already have an expectation that, in general, catchment averaged or poorly sampled inputs will lead to a tendency to underprediction of peaks, then the question that the paper should be addressing is what to do about that in real applications (where the discharge uncertainties also come into play). We cannot invent input data, so we will normally use as much as is available (which would be even more than the 150 gauges in this study). We know that the calibration will depend on, and compensate for, the limitations of the observed data that are available – but the smaller the sample then the greater the resulting uncertainties in the predicted discharges might be (see figure 13 again). This paper does not address those uncertainties (except in comparing the 5 samples at each density). It does not even make use of the uncertainties associated with the kriging interpolation (which would seem to be a good reason for using the kriging interpolator, despite the necessary assumptions and requirement to already have many gauges to estimate variograms).

So I do not think the paper can be published in this form. As far as I can see we cannot use any of the results to improve practical applications (other than a general exhortation

to use as many input gauges as possible and try to take account of spatial patterns – but even then the common problem of not having gauges in higher elevations has not been mentioned and there is a lack of information about how snow is handled). I would suggest it needs to be more complete and more ambitious and address the question of IF we only have a certain density of gauges available (remembering that in practice we cannot resample from a larger set), then how should the modelling workflow compensate to get better estimates of the (uncertain) peak flows. Does calibration provide sufficient compensation? Do the acceptable parameter sets change with the input scenarios? Can the authors suggest "uncertainty multipliers" as the density of inputs decreases (but that requires estimation of addressing the simulation uncertainties more directly)? Does this vary between models (though I understand the problem of calibrating SHETRAN if it was used in the comparison)?

There are many other comments on the manuscript.

Keith Beven

Please also note the supplement to this comment:
https://hess.copernicus.org/preprints/hess-2022-281/hess-2022-281-RC1-supplement.pdf