Interactive comment on “A comparison of three simple approaches to identify critical areas for runoff and dissolved reactive phosphorus losses” by C. Hahn et al.

Anonymous Referee #2

Received and published: 21 January 2014

The authors compare the critical areas for runoff and DRP losses derived for 2 small Swiss catchments from 3 different types of models: RRP is a semi-distributed conceptual model at hourly resolution, DoRP is a largely static runoff processes classification scheme with hydrological dynamics added on through a simple bucket model, SCIMAP is a static risk mapping approach. The conclusions of the paper (section 5) unfortunately do not go much beyond what was already known prior to model application (the same is true for section 4.2 on model limitations).

The study is further limited methodologically by using the RRP model as a benchmark to compare the other models against. The authors justify this based on their calibration/validated of the model as reported in Hahn et al., 2013). However, the calibration timeseries was extremely short (7-17 July 2000, P14504, L5) and I doubt that all important modes of behaviour of the system are reflected in the calibration information and hence the model parameters. The validation periods used in the present paper are equally short (Mar-Nov 1999 and May-Aug 2010 for the 2 catchments, respectively). So RRP is, in my opinion, not a robust benchmark. The authors state regarding the validation of the DoRP model that “no reliable statement for the Stägbach catchment is possible due to the limited number of observations” (P14505, L17-18), and I believe the same is true for the other catchment and for RRP.

In addition, I had the following comments:

P14498, L28-29: Is this not pre-empting the results? What about DoRP?
P14500, L7: Please explain uniform MC method.
P14500, L9-12: Is this not the classic GLUE method? What is the justification of the choice of performance measure (NSC) and behavioural threshold (0.6), particularly in relation to more sophisticated methods such as formal Bayesian methods and extended GLUE (e.g. Romanowicz & Beven, 2003; Rankinen et al., 2006; Winsemius et al., 2009; Krueger et al., 2012)? This is not discussed in in the original modelling study (Hahn et al, 2013).
P14506, L23: Homogeneous rather that heterogeneous?
P14507, L12-14: I wonder whether the lambda/NI comparison can be given more prominence, perhaps as a new focus of the paper?
P14507, L21; P14515, L16: Re tile drains, how significant are they in the 2 catchments? If important then topography might not be a good predictor of runoff generation. The same would apply for NI in terms of pollution risk.
P14509, L14-17: Here I’m missing a formal spatial comparison, e.g. via Cohen’s kappa.
I do not think this is a problem since these are the risky times, no?

It is not very clear what was done here – please try and revise.

How was the spatial information aggregated over the event timesteps?

Here, too, I'm missing a formal significance test, e.g. via ANOVA.

References


Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 10, 14495, 2013.