

Answer to Editor – A. Butturini

Editor comments were copied in italics, answers of authors are in normal font.

I would thank the two referees for their valuable contributions. Overall, both coincided that the data presented in this paper is an “interesting” contribution that clearly fit the HESS aims and scope. However an evident revision is necessary before to resubmit the manuscript to HESS. Therefore I strongly suggests to analyse in depth all suggestions and critiques provided by reviewers. It is crucial that authors make an effort to remove all reviewers doubts.

What my concern, I agree totally with the referee #1 with respect to the need to focus on concentrations rather than to the nitrogen loads.

Reply: The reason why we are not focusing on concentrations is simply, that the existing model structures we are using have been developed for N load predictions, and not for concentrations. The models are based on mass-balances calculations and were designed to provide water and N loads as primary outcome. Most of the applications of the models we used, and that we are aware of, predicted long term N loads or balances and source apportionments.

*The lack of information about the modelled dissolved nitrogen fractions concentrations strongly contrast with the exhaustive description of the observed dissolved nitrogen fractions in the study site description section (see page. 5306 and lines 15-17, and figure 2**). It is essential to fill this gap.*

Reply: We do not really agree with the criticism the description of observed dissolved N fraction is too long. With overall 35 pages, less than one page of the paper is dedicated to the basis of all modelling approaches, i.e. the field experimental work. In addition, each description of the model that we included in the ensemble summarizes the different nitrogen fractions that can be simulated with the respective model (Table 1 and P5306-5311).

Under this perspective, figure 5 suggests that most of the models tend to subestimate the measured nitrogen loads (i.e. most of dots are below the 1:1 line. The expectation is the NO₃ and total-N fluxes at Savja, see below). If I couple this observation with your sentence that reveals “high concentrations and high fl se during winter” (see pag. 5306 and lines 15-17), a reader might suspect that models tend to fail in the activation of the nitrogen mobilization from hillslope to the streams during winter storm evens (or the spring snow melt episodes). Then, leaching processes might a feeble node in simulations modules. I believe that a detailed exploration of simulated dissolved nitrogen fractions concentrations might strongly help to obtain a more depth understanding of models strength and weakness. As mentioned before, and outlined by the authors as well, (see pag 5320, lines 1-3) the most noticeable exception is that for total-N and NO₃ at Savja. In this case the best MME have a good prediction along the entire range of observed total- N (and nitrate) fl In .

Reply: Strengths and weaknesses of models exist as illustrated by the heterogeneity of our results. However, the aim of this study was neither to try to understand the reason of these discrepancies nor the reason why some models would perform better than others (P5304 L3-5 ‘We focused our evaluation on the effect of merging results rather than on the results themselves’). In-depth descriptions of the internal mechanisms of these models were already published (Global overview:

Breuer et al., 2008; SWAT: Arnold et al., 1998; HBV-N-D:Lindgren et al., 2007; LASCAM: Sivapalan et al., 1996a, 1996b; Viney et al., 2000; CHIMP/INCA: Wade et al., 2002; Whitehead et al., 1998) and this was definitely not the frame of our study.

Nevertheless, an investigation of why models fail is needed in the future to improve our general understanding of hydro-biogeochemical fluxes. But to do so, we need spatially distributed data sets with high temporal resolution. This is not available in the current study area.

We will point out these two aspects in detail in the conclusion of the revised paper.

Nevertheless a question arise at this point: is it the ensemble modelling approach really necessary? At the end of the discussions authors are rather ambiguous: “the improvements were not very high compared to those of the best SMEs” (pag 5320, line 5). Effectively, from figures 5 and table 4 it appears that SWAT model is typically better than other models, and its output quality is similar to that obtained with the Multi-Model Ensembles (MME). I believe that a critical discussion of this apparent incongruence might strongly enhance the entire discussion section.

Reply: The objective of this paper is first of all to show the value of a method (ensemble modelling and data fusion) that is standard in climatology and meteorology, becoming state-of-the-art in hydrology and is very new in hydro-biogeochemistry.

With regard to the comment that SWAT is already a good predictor, we see this a bit different. Table 4 shows that SWAT does a good job for NO₃ and total N, but not for NH₄. And, other models are also performing well for NO₃ and total-N such as LASCAM-S. It is true that MMEs have a low improvement rate compared to the other models, but still, the MMEs give the best results and this is in our sense the aim of any modelling effort. Nevertheless, we will extend and improve the discussion in this direction.

Additional comments: I have serious problem when I attempt to compare data in table 4 with those used to create figures 4. (See my attached file). Is it a misinterpretation?

Reply: The authors thank the editor for this nice editing comment. The actual values are those in the table and small error, leading to big discrepancies was to be found in the calculations. A corrected figure will be submitted along with the revised manuscript.

Pag. 5302, line 27: perhaps a reference might help to test your assertion.

Reply: We will add: Shamseldin, A. Y., O'Connor, K. M. and Liang, G.: Methods for combining the outputs of different rainfall–runoff models, J: Hydrol., 197(1-4), 203-229, doi:10.1016/S0022-1694(96)03259-3, 1997.

Pag 5318, line 16: “As illustrated on Fig. 3 the SWAT”. . .do you refer to figure 4?

Reply: Indeed! This will be corrected in the manuscript.

What does “Remaining N” means? Does the authors refer to the dissolved organic nitrogen fraction (DON)?

Reply: Remaining N is the difference between the Tot-N concentration and the sum of NH_4 and NO_2+NO_3 concentrations, DON concentrations being not directly available.

References

Arnold, J. G., Srinivasan, R., Muttiah, R. S. and Williams, J. R.: Large area hydrologic modeling and assessment part I: Model development, *J. Am. Water Resour. As.*, 34(1), 73–89, 1998.

Breuer, L., Vaché, K. B., Julich, S. and Frede, H.: Current concepts in nitrogen dynamics for mesoscale catchments / Concepts actuels relatifs à la dynamique de l'azote dans les bassins versants de méso-échelle, *Hydrolog. Sci. J.*, 53(5), 1059, doi:10.1623/hysj.53.5.1059, 2008.

Lindgren, G., Wrede, S., Seibert, J. and Wallin, M.: Nitrogen source apportionment modeling and the effect of land-use class related runoff contributions, *Nord. Hydrol.*, 38(4-5), 317, doi:10.2166/nh.2007.015, 2007.

Sivapalan, M., Ruprecht, J. K. and Viney, N. R.: Water and salt balance modelling to predict the effects of land-use changes in forested catchments. 1. Small catchment water balance model, *Hydrol. Process.*, 10(3), 393–411, 1996a.

Sivapalan, M., Viney, N. R. and Jeevaraj, C. G.: Water and salt balance modelling to predict the effects of land-use changes in forested catchments. 3. The large catchment model, *Hydrol. Process.*, 10(3), 429–446, 1996b.

Viney, N. R., Sivapalan, M. and Deeley, D.: A conceptual model of nutrient mobilisation and transport applicable at large catchment scales, *J. Hydrol.*, 240(1-2), 23–44, doi:10.1016/S0022-1694(00)00320-6, 2000.

Wade, A. J., Durand, P., Beaujouan, V., Wessel, W. W., Raat, K. J., Whitehead, P. G., Butterfield, D., Rankinen, K. and Lepisto, A.: A nitrogen model for European catchments: INCA, new model structure and equations, *Hydrol. Earth Syst. Sci.*, 6(3), 559–582, 2002.

Whitehead, P. G., Wilson, E. J. and Butterfield, D.: A semi-distributed integrated nitrogen model for multiple source assessment in catchments (INCA): Part I—model structure and process equations, *Sci. Total Environ.*, 210, 547–558, 1998.