

Evaluation of the revised version

Dear Dr. C. Schürz

Thank you for the revised version of the manuscript and the effort you have put into it. In general, I consider the questions to be properly addressed. Nevertheless, some aspects call for clarification before I can accept the manuscript for publication.

Below I list a number of questions related to your response to the reviewers:

- Fig. A3: You mention in the response that the figure shall show any clustering of model parameters and possible parameter interactions. This figure is indeed very valuable for that purpose and illustrates a few interesting patterns. For example:
 - o There is one case of a strong correlation between two parameters (N concentration in rainfall, N percolation coefficient in the Raab catchment).
 - o some parameters were not relevant in one of the catchments (e.g., denitrification rate in the Raab catchment).
 - o Some parameters have strongly bi-modal distributions (e.g., available water capacity in soils in the Raab catchment).
 - o Some parameters have very different distributions for the two catchments (e.g., subsoil hillslope length).

However, you do not discuss any of these aspects (why to keep highly correlated parameters for example).

Can you please comment on

- o Whether these observations are plausible (why is denitrification important in the Schwechat but not in the Raab catchment)?
 - o How these patterns may have affected the outcome (how important for example are the three SLSSOIL parameter values for the resulting overall uncertainty)?
- Reviewer 3 commented on some apparently counter-intuitive results such as the limited effect of land use change on the N concentrations. You argue that this is supported by findings in the literature and you also provide evidence that it was not an artifact of the number of realizations of the respective input/setup. While I agree with your arguments I still think it was worth mentioning that land use change might cause much larger effects on nitrate levels for example in situations where larger fractions of the catchments were affected or changes were more dramatic (see e.g., Honti et al. (2017) as a quick example to illustrate the point (not because I think you have to cite it)). I suggest you clarify this aspect more clearly.
- Reviewer 4 asked for clarifications about the GSA sample size ($N = 7000$). Your explanation seems clear to me except for the determination of the base sample $N_{\text{base}} = 1000$. Was this an arbitrary decision or was there a further argument behind? Please clarify.

Below I list a number of questions related to revised manuscript:

p. 2, L. 1: I suggest to skip “proved to” (how did you proof it?). I think the statement “We present approaches for the visualization of the simulation uncertainties that support the diagnosis ... “ is clear enough.

p. 2, L. 7: What do you mean by “anomalies”? Please clarify.

p.7, L. 18: “where” should be corrected to “were”, I assume.

p. 8, L. 4: space missing in front of parenthesis.

p. 9, L. 21: Probably, one should insert “maximum” before “5%”.

p. 9, L. 22: How were the 42 parameters selected? If you argue that these parameters are frequently used for calibration, can you support this by a reference? Which fraction of the total number of parameters is covered by them?

p. 10, L. 5: It would be also informative to list those parameters that were not influential.

p. 10, L. 20: How can one calculate the target variables using the Nash Sutcliffe criterion? Please clarify.

p. 11, L. 19: What are these observable trends? Based on which data, references? Please clarify.

p. 12, L. 28: The N = 7000 represents almost half of all combinations (46%) for the Schwechat catchment (according to Tab. 3) and 13% for the Raab watershed, is this correct? Perhaps this information might be useful.

p. 12, L. 29: It is not clear what you mean “quasi random sampling”.

p. 13, L. 10: The target variable should be “y”, I assume.

p. 14, L: 9 – 10: What is meant by the “generic random sample”?

p. 15, L. 10: “Subsetting” isn’t a verb, I assume.

p. 15, L. 12: "temperature or precipitation anomalies": how did you define these terms?

p. 19, L. 11 – 13: "In comparison to the reference period (dashed line), wetter future climate scenarios (blue) simulated larger discharge and NO₃ -N loads, while dryer future conditions lead to a drastic reduction in discharge and NO₃ -N loads." With these changes, what are the implications for the N-balances: will N accumulate under drier conditions? Would one not expect feedback mechanisms to get activated? Can you comment on that?

Fig. A2: A log-scale for the y-axes would allow for a much better comparison of the observations and the simulations.

Please respond to these comments.

Sincerely

Dr. Christian Stamm
Editor HESS

References:

Honti, M., J. Rieckermann, N. Schuwirth and C. Stamm (2017). "Can integrative catchment management mitigate future water quality issues caused by climate change and socio-economic development?" Hydrological and Earth System Sciences **21**: 1593–1609.