

Associate Editor (Dr Kerstin Stahl)

The authors thank Dr Kerstin Stahl for her constructive comments on the manuscript. We agree with the comments and we explain below how we have modified the text to account for her comments.

Dear Mehmet and co-authors

Thanks for submitting your revised manuscript, which I think is now almost ready for publication.

In my assessment of your revisions, I found that some of the justified criticisms of Referee 2 that you rebutted should at least be discussed in the discussion section to give a fair assessment of the found implications to the reader. Please revise these aspects as detailed below, along with some editorial issues in order to bring this manuscript into its final, publishable, shape.

Best regards,
Kerstin

Comments:

Comment 1) Suggested change about unregulated flow in the Moselle:

You are modelling the Moselle and not the Rhine. The Moselle originates from the Vosges and not from the Alps. Please delete the first two sentences added on the Rhine. They are irrelevant to this study.

It is not relevant here what you assume, but what is the case and how it may influence the modelling. If you are confident that the Moselle flow you model has no flow alterations that would not be captured by the model's dynamics, then just say so – but I suggest to say that the “alteration of flow magnitude and dynamics are minimal”, rather than to say “human influence on the river is minimal”, which is a different thing (incl. anything from urbanisation in the basin to bank stabilization etc. which is high everywhere in Europe).

However, your models are calibrated. So for the model experiment, what may matter more than general human influence is perhaps that there were no (low)flow-altering changes to regulation during the time of calibration and application. Is this the case? Tests performed on data homogeneity could prove that. Have you done any?

Reply from authors: We agree with the comment. We removed the two sentences about the River Rhine and applied the Mann Kendall trend test to the daily P , PET and Q series. We found significant trends in all three time series at different alpha significance levels. To understand the direction of the trend we plotted them below.

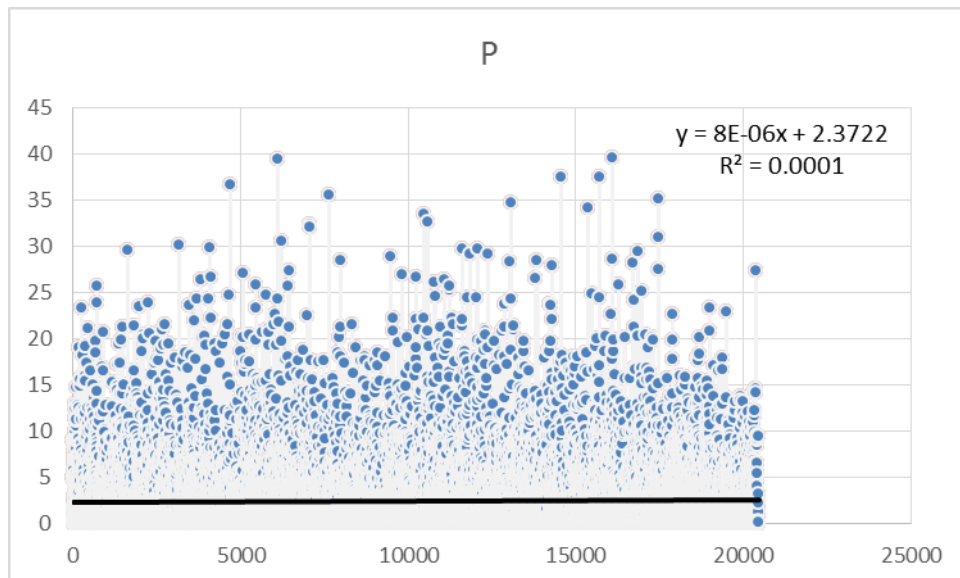
Significant trends in P , PET and Q do not show always the human influences in the river basin. For example, if there is a positive trend in P and a negative trend in PET resulting in a positive trend in Q this might be caused by climatic changes instead of human influences. If trends in P and PET are not consistent with trends in Q one might conclude that human interventions also play a role. We will report the results in the revised version of the manuscript.

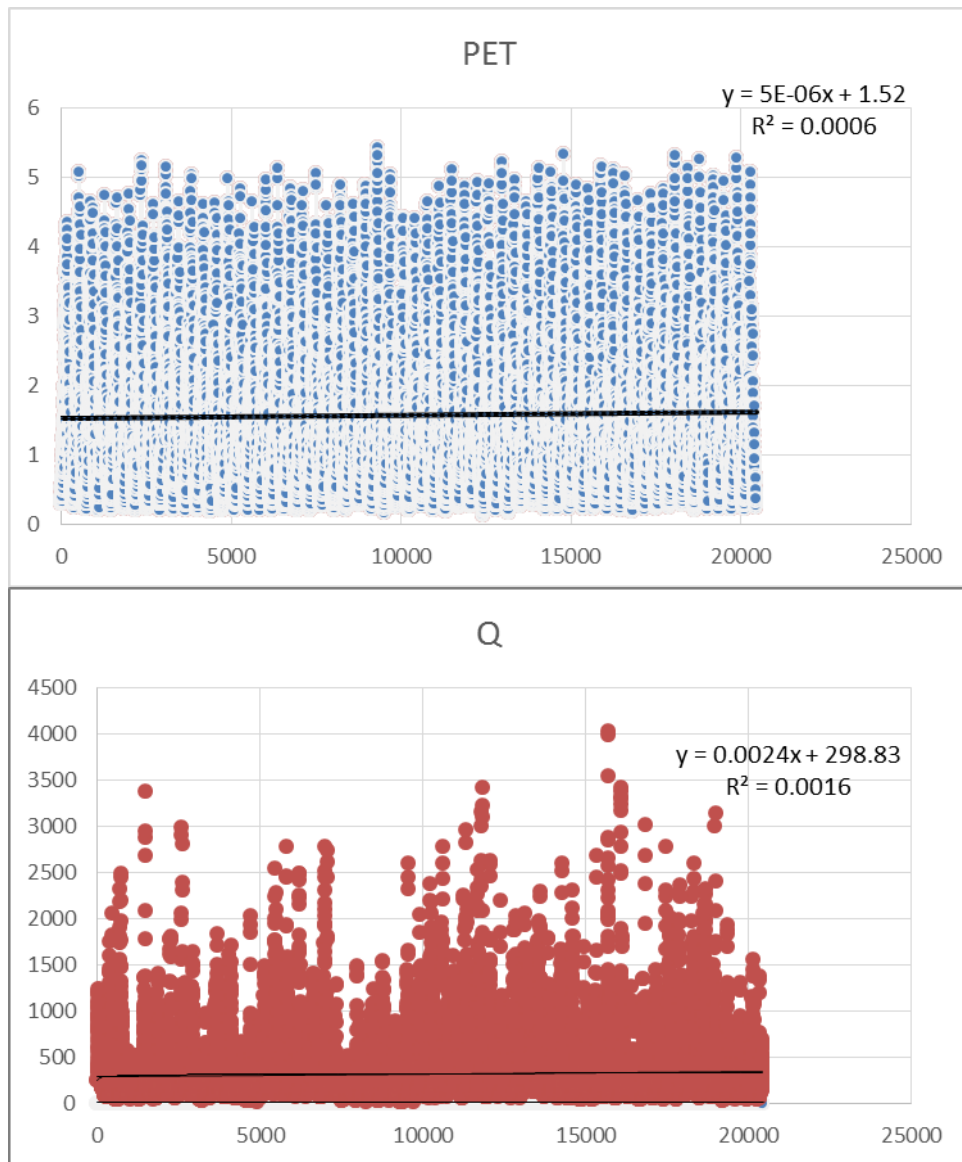
manuscript, lines 400-405:

It should be noted that the effect of anthropogenic activities (e.g. flood preventive regulations and urbanisation) on the alteration of flow magnitude and dynamics is not obvious as we found weak positive trends in all P, PET and Q series ($p < 0.025$ for the three variables using Man Kendall method) which might be caused by climatic changes. However, the effect of uncertainties due to the anthropogenic activities on the three low flow models is minimal as the models are successfully calibrated for the study area. Other studies reported that the trends in flood stages in Moselle River were not significant (Bormann et al., 2011).

<i>Variable</i>	<i>H value</i>	<i>Alpha</i>
P	<i>1</i>	5.9e-05
PET	<i>1</i>	0.024
Q	<i>1</i>	7.7e-07

Note: The code performs original two tailed Mann-Kendall test. It tests the null hypothesis of trend absence in the vector V, against the alternative of trend. The result of the test is returned in H = 1 indicates a rejection of the null hypothesis at the alpha significance level. H = 0 indicates a failure to reject the null hypothesis at the alpha significance level.





Comment 2) The referee's valid concerns with the objective function are not properly addressed. The choice of an objective function is a crucial step in modelling and you have to remember that literature (also your paper in the future) will be used by others to decide and justify their decision. Therefore, you need to be fair and at least discuss transparently a) your decision path (based on literature and your aims) and b) all caveats encountered with your choice. This needs to be addressed with a proper discussion section to make problems and limitations very clear to the reader. Please add a fair assessment of the issues (incl the unit discrepancy) raised by the referee and based on these issues, consider toning down the conclusion (where only the further implications should be raised and not the study and results be repeated). Please revise Discussion and Conclusion accordingly.

Reply from authors: We agree with the comment. We added the green shaded text to the Discussion section.

These models were calibrated using a hybrid low flow objective function. Although combining two metrics offered a selective evaluation of low flows, we have noted an

important caveat using the second component of the hybrid metric as it is less sensitive as compared to the first part of the hybrid metric resulting in higher (optimistic) values for most cases. The different units had no effect on our calibration results as the ultimate calibration target value is zero (i.e. unit independent). Other studies also combined different metrics with different units (Nash Sutcliffe, RMSE, R^2 and NumSC, i.e. the number of sign changes in the errors) into one objective function (Hamlet et al., 2013). However, the modellers should carefully use the hybrid function introduced in this study, in particular when comparing different model results. Plotting the two parts of this hybrid function as a Pareto front can lead to a more clear picture than simply summing the two metrics.

(Hamlet et al., 2013)

. Hamlet et al.

The objective function for the optimization process in this case was:

$$\begin{aligned} \text{Min } F = & -\text{NSE}(Q) - \text{NSE}(\log(Q + 1)) + \text{Vol_Err}(Q) \\ & - R^2(Q) + \text{Peak_Diff}(Q) + \text{RMSE}(Q) \\ & + \text{NumSC}(Q), \end{aligned} \quad (1)$$

where Q is the monthly streamflow; $\text{NSE}(Q)$ is the NSE (monthly flow), which varies between $[-\infty, 1]$ and typically between $[0, 1]$; $\text{NSE}(\log(Q + 1))$ is the NSE (monthly flow), which varies between $[-\infty, 1]$ (this metric places less emphasis on high flow errors in calculating NSE); $\text{Vol_Err}(Q)$ is the annual volume error (in 1000 acre feet); $R^2(Q) = R^2$ (squared correlation coefficient between simulated and observed Q), which varies between $[0, 1]$; $\text{Peak_Diff}(Q)$ is the mean hydrograph peak difference—the absolute value varies for different sites; $\text{RMSE}(Q)$ is the root mean square error, whose absolute value varies for different sites; and $\text{NumSC}(Q)$ is the number of sign changes in the errors (this metric penalizes simulations with too much month-to-month variability in comparison with observations).

Model calibration and validation used a split sample

Comment 3) Also, the (still very minimalistic) discussion needs to include the new finding with the apparent better forecast (higher hit rates) as a result of a more uncertain (larger range) longer-range forecast. I think this apparent better performance is a very important artefact, which actually requires to question (at least to discuss) the use of these metrics or at least what can be compared to what. And this really needs a proper discussion in order encourage follow up studies that use this paper to design their studies to address this issue better! There must be literature that has highlighted such an apparent improvement die to range previously, no? Please check and include a discussion accordingly.

Reply from authors: We added the green shaded text in the Discussion section in the revised version of the manuscript.

Moreover, our analyses show that the better forecast performance for longer lead times is an obvious artefact since the higher hit rates are the result of a more uncertain (larger range) forecasts. The probabilistic skill scores focuses on the forecasts, the uncertainty in the meteorological forcing data should be carefully scrutinized using different quantitative screening methods e.g. box plots.

This part has been added to the Conclusion part

The uncertainty increases in seasonal meteorological forecasts can lead to better skill scores as an artefact of large ranges in input. Therefore, the quality of the model inputs should be assessed in addition to the model outputs. The identified glitches in the second part of the hybrid objective function can be eliminated by plotting the two parts on two axes rather than simply summing the two metrics.

Comment 4) Table 6. I don't understand the answer. In addition: Table 6 needs to be changed with proper distinguishable variable names and all that text moved into table footnotes (or if the variables are explained before, then not needed) so that the reader can actually see immediately, which different counts comprise the contingency table. With all that text inside the table it is very confusing like it is. Explanations can be coordinated with the equation (maybe put this first and then use the table to explain the cases).

Reply from authors: We removed Table 6 to avoid ambiguity. Table 5 is a standard contingency table and can be used for interpreting low flow events too.

Comment 5) In all figures, the tick labels are too small and lines are too thin. In all figures there is space to make them larger. Axes labels are then sometimes excessively large, creating a very unprofessional look of the figures. Please check size requirements and conventions for sizes of labels and improve.

Reply from authors: We agree with the comment. We improved the readability of the figures.

Comment 6) Table 7 comes before Table 3 in the text. The order of mentioning tables and figures in the text MUST be consistent and Tables mentioned in the text must be placed as soon as possible after the mentioning, i.e. it is not possible to refer to a later table early in the text. Needs to be changed.

Reply from authors: We agree with the comment. We removed the Table 7 cross references in section 3.1.1 and 3.1.2 to keep the referencing consistent. We inserted the green shaded text in section 3.1. as shown below to guide the reader for the description of the parameters in section 4.1.

We provide a detailed description for each parameter of the three models in section **Error!**

Reference source not found..

References

Hamlet, A. F., Elsner, M. M., Mauger, G. S., Lee, S.-Y., Tohver, I., and Norheim, R. A.: An Overview of the Columbia Basin Climate Change Scenarios Project: Approach, Methods, and Summary of Key Results, *Atmosphere-Ocean*, 51, 392-415, 10.1080/07055900.2013.819555, 2013.