Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2018-284-AC2, 2018

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Statistical approaches for assessment of climate change impacts on low flows: temporal aspects" *by* Anne Fangmann and Uwe Haberlandt

Anne Fangmann and Uwe Haberlandt

fangmann@iww.uni-hannover.de

Received and published: 1 October 2018

Dear Dr. Hanel,

thank you very much for reviewing our manuscript. We appreciate your suggestions, which definitely help improve the paper. Please find below our responses to your comments:

1. Motivation of the study. In introduction the authors write: "Even though the application of process-based hydrological models is supposedly the most accurate means for analysis of climatic impacts, model set up and application may be difïňĄcult and time

consuming, especially for detailed regional analyses and if large numbers of climate change scenarios and climate models are to be considered. Also, data scarcity may pose an issue for model calibration." While I agree that statistical models are valuable for drought assessment I ïňĄnd the motivation for their use in this study somewhat unconvincing. The authors use many meteorological indices requiring at-site/catchment data as well as indices calculated from runoff time series. I.e. all information needed for calibration of hydrological model is in principle available. Then, in my opinion, the process-based hydrological models should be always preferred, especially for climate change studies, since it might be not clear what happens if the statistical model (calibrated on current conditions) is applied for climate that is warmer e.g. by 4âŮęC. For instance the snow water storage regime may change dramatically etc., introducing non-linearity and making the extrapolation uncertain. In addition, the statistical framework presented by authors is rather complex. One may ask, whether the application of simple process-based hydrological model is not easier even for large number of catchments, especially when lot of hydrological models are available for easy scripting e.g. in R. Therefore I see the potential in the presented framework rather in assessment of drought drivers or understanding the key processes responsible for drought generation in current climate (through regression analysis, PCA, ïňĄnding the most informative variables etc.) than in drought prediction. In this perspective, I think the paper would beneïňĄt from focusing more on the information we can get from the statistical models.

- We agree with the issues concerning the motivation for use of statistical approaches. The initial idea behind applying a few meteorological indicators as predictors was the provision of a different impact model approach that – in practice – would be straightforward in its formulation and application. With all the adaptations and model variants, the final model became more complex.

- Still, even with a number of steps involving index computation, data transformation, dimensionality reduction, variable selection etc. the methods are extremely fast and straightforward, especially during application, given the simplicity of the model and the annual basis for calculation. We understand that there is a number of deterministic models that require similarly little input, calibration effort and are easily applied. Nonetheless, we still believe that the ease of application, especially for settings with large numbers of catchments, is an important asset of the statistical approach.

- We completely agree to stress the importance of the approach for identification of drought drivers. We changed the title, the introduction and the motivation accordingly and try to focus more on the information we get from the statistical models during the analysis. A section has been added that specifically deals with analysis of drivers for the individual low flow indices.

2. Study design Authors calibrate the statistical models using observed data for present climate, ïňĄnd the best model and apply it on bias corrected RCMs for present and future climate. Then the bias for present conditions and the changes between control and future periods are assessed. I ïňĄnd this rather unfortunate. There are many reasons why the observation-based statistical model cannot work with the RCM simulations. First of all, as stated by authors, the bias from the climate models is removed by monthly linear scaling. It is well known, that the bias in climate models has multi-scale nature and correction at one scale does not warrant correction at different scales (see e.g. Hanel et al., 2017). Therefore the daily meteorological indices are likely biased impacting the results of the whole analysis. Second, there is spatial scale mismatch between the climate model and observed data, as also noted by authors. Especially the indices related to precipitation extremes are then smoothed and also characteristics of temperature related to altitude are likely biased. More natural approach would be to ïňĄt the statistical model considering the climate model derived meteorological indices as input and the observed drought indices as output. This would then include the correction into the statistical model.

- The calibration has been carried out as suggested by the referee. In order to address the mentioned scale issues, the models have been re-calibrated using the climate model data instead of the observed data. Since the temporal reference between

climate model data and observed drought indices is not given, a re-calibration strategy (regression coefficient values are re-estimated) is used that aims at reproducing the overall distribution of the drought indices, rather than the annual values, by using ordered series of observed low flow and climatic indices/principal components thereof obtained from the climate models. Instead of trying to reproduce the observed low flow series, the initial model fitted to the observation has been used as target variable for calibration, which already contains a certain error. Matching the climate model data directly to the observed time series would assumingly lead to wrong parameter estimates.

- As seen in Fig 1., the climate change signals are comparable to the previous results in direction, their median values are similar but the differences between the individual stations, i.e. the spread seen in the boxplots becomes larger. The former results have been replaced by the newly calibrated analyses, the analysis of the 20th century reference period has been discarded, as the bias becomes negligible.

3. Clarity Although the paper is in general well written, some methods are not presented clearly. See speciïňĄc comments. SpeciïňĄc and minor comments: 1. p. 3, l.10 "A hydrological model could only be set up for 7 of these stations" – why?

- The hydrological model had already been set up for previous analyses within the scope of the project this article originates from. The amount of work required for setting up the model for the remaining stations for mere comparison of model performance would be too high. Thus, the 7 stations are used. An explanation has been added in the manuscript: "In previous work (NLWKN, 2017), a hydrological model has been set up for 7 of these stations, as indicated in Fig. 1. This selection of stations will be used for comparing statistical and hydrological model performance."

2. p. 6, l.5 "The climate model data is available for a smaller domain than the observed climate data" - actually the RCM data are available for the whole Europe . . .

- Same holds for the regional climate model data, which has been pre-processed within the scope of the project.

3. p.7, l.9 change "to be e relevant" to "to be a relevant"

- “To be a relevant” has been corrected.

4. p.7, l.18 "X" should be changed to Chi

- “Chi^2” has been corrected.

5. p.9, l.25-33 - The procedure for detection of non-stationarity - not very clear to me Authors use bootstrap to examine the nature of the relation between variables. For a candidate model, they resample with replacement the time series. Then, it is not clear whether they reïňĄt the model or just check if the regression ïňĄtted for original series is still valid for the resampled series (they say "the regression is signiïňĄcant"). Any case, they further write that "if the regression is signiïňĄcant for all subsamples, the relationship is considered continuously stationary. If, on the other hand, the slope of the regression line is time dependent, random samples taken from different periods will not yield a signiïňĄcant slope coefïňĄcient." I have several concerns regarding the procedure: - It would be more natural to allow for some fraction of insigniïňĄcant samples. The requirement that ALL subsamples yield signiïňĄcant regression seems to be quite strong - If the models were reïňĄtted for each subsample then in principle even if the regression is signiïňĄcant, the coefïňĄcient can be oposite, or very different... The regression may be also insigniïňĄcant due to natural variability not as a result of non-stationarity. - If the models were not reïňĄtted for each subsample then the procedure is kind of cross-validation - In summary, the procedure needs to be described in more detail, consider the concerns above and be preferably based on published literature

- We agree that the requirement of all subsamples having to yield a significant regression is too strong. Actually the test has been set up to at least require 90% of the subsample regressions to be significant. This has been communicated wrongly in the original manuscript and has been corrected. “If the regression is significant for more

than 90% of the subsamples, the relationship is considered continuously stationary.”

- The models are indeed refitted for each subsample. The idea behind testing the significance of slope parameters for different subsamples is the following: if the slope of the regression line varies between different sections of the time series (which cannot be tested due to limited time series length), then a model fitted to randomly selected data from different sections of the series would not be significant. Figure 2 shows that this procedure automatically selects those explanatory variables that simultaneously show least variation in the slope and mostly little variation in the intercept parameter (example for random station, tested for the entire 60-year period).

- Unfortunately, there is no available literature for the procedure.

6. p.10, l.5-10 - Calibration/validation procedure "... the calibration consistently precedes the validation period. This set-up is chosen to test the ability of models ïňĄtted to a past period of time to predict ’future’". I agree that it is useful to assess the performance of model ïňĄtted in calibration period in the validation period. I would, however, be careful with presenting this as a test for the ability of models to predict future. Especially in the climate change context, the differences between the calibration and projection period are much larger, especially for temperature.

- The selection of the calibration and validation period was intended to assess problems (like non-stationarity) that can arise when applying the models in climate change impact analysis. It was in no way seen as a guarantee that the models will also work for climate model data, but simply as the best available option for model set up based on the limited observation period. We added the following sentence in the discussion to put the concerned statement into perspective: “It should be noted that the set up for model calibration and validation allows for assessment of changes between directly adjacent periods only. Application of the models to predict low flows in a more distant future under more severe climatic change may significantly enhance the error due to non-stationarity.”

7. Model ïňĄtting and evaluation section - it is not clear which combinations of input variables were tested (where they all possible combinations considering the restrictions?). How many models it was?

- The variable selection was carried out in a step-wise procedure rather than in an exhaustive search, starting with a random variable, adding the next BIC-minimizing variable and (if applicable) removing any variables if their removal minimizes the BIC further. The paragraph has been altered for clarification.

8. p.10, l.5-10 It is not clear, what you mean by "time series are reversed" and "the inversion is applied to preserve the continuity of the time series". Are the calibration and validation periods just swapped?

- The time series have been inversed, i.e. calibration and validation period are exchanged. The aim of this is to assess differences in selected parameters and model performance, as discussed in several places. In order to preserve the structure of potential non-stationarity, all values are inversed, i.e. year 60 becomes year 1, year 59 year 2 and so on.

9. p.10, l.23 "MLR" not deïňĄned

- MLR definition has been added in section 3.1.

10. Model performance section (p. 10 and further) In Figure 4, the authors present "the performance of all tested model conïňĄgurations...". It seems that in the boxplots also the models that are not appropriate are presented. I ïňĄnd this strange, especially for situations when the same predictors are used e.g. in one model assuming independence of residuals and the other model allowing for auto correlation. Then only one of the two models should be considered (based on the likelihood ratio test), since the other is not appropriate. In general, it would be more interesting to focus on the information we can gain from such regression exercise.

- We agree that the benefits of the individual modifications in the regression methods

could be highlighted in more detail. Still, the aim of the comparisons was to consider the entire set of stations in the study area and assess the maximum performance we could obtain with adjustment of the regression approach but without violating regression prerequisites. For better assessment of the information gain of the individual approaches, numbers have been added on how many models have been successfully fitted.

11. p.13, l.19-21 "Maximum likelihood ïňĄtting of models with linear time dependence of the coefïňĄcients (validated via likelihood ratio tests) was tried ... but could not yield the expected results for the small calibration period." - could you be more speciïňĄc on this, please? - what you mean by expected results?

- The procedure will be explained in more detail. Regression models were fitted where slope and intercept parameters were estimated as linear functions of time. Whether the models were appropriate was then tested using the maximum likelihood test. In case of linear time dependence of the regression, the models would yield better results than the stationary models. The results indicated, however, that such a linear time dependence was not given.

12. Figure 6 and related discussion It seems that the mean error increases with the temporal distance from the calibration period. Does this also apply when you swap the calibration and validation periods? Could this be a sign of trend in temperature? Would the ïňĄgure look same if you omit models relying on temperature (and ETP)?

- The mean error increases independent of the direction the model has been fitted in (i.e. swapping of calibration and validation period). As has been discussed previously (p.10, l.10), the positive bias occurs in both directions. Also, the positive bias occurs in models exclusively made up of precipitation based explanatory variables.

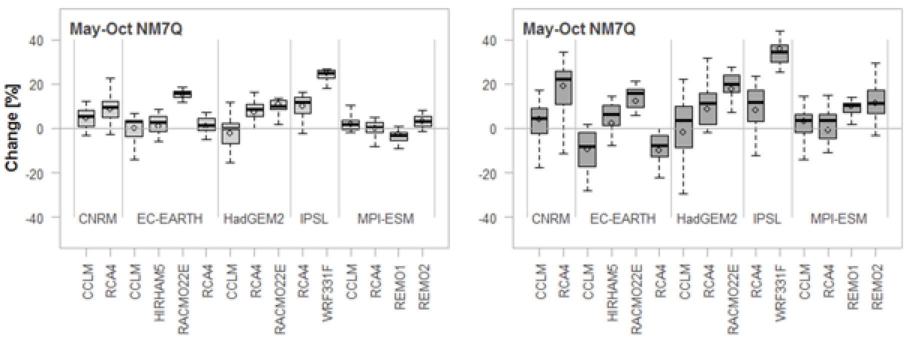
13. Prognosis section - see general comments 14. How was the ETP calculated for climate model data?

- The ETP for climate model data has been estimated according to Turc-Wendling, just as for the observation, stated in Table 2.

Please also note the supplement to this comment:
https://www.hydrol-earth-syst-sci-discuss.net/hess-2018-284/hess-2018-284-AC2-supplement.pdf

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-284, 2018.

Old climate change signals (model calibration using observed data only):



New climate change signals (after re-calibration for each climate model):

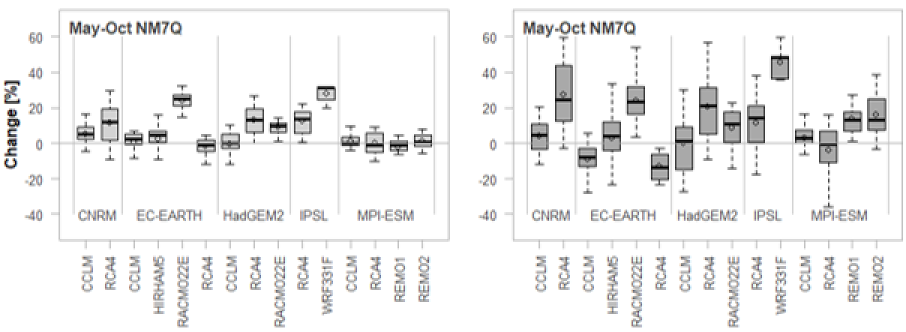


Fig. 1.

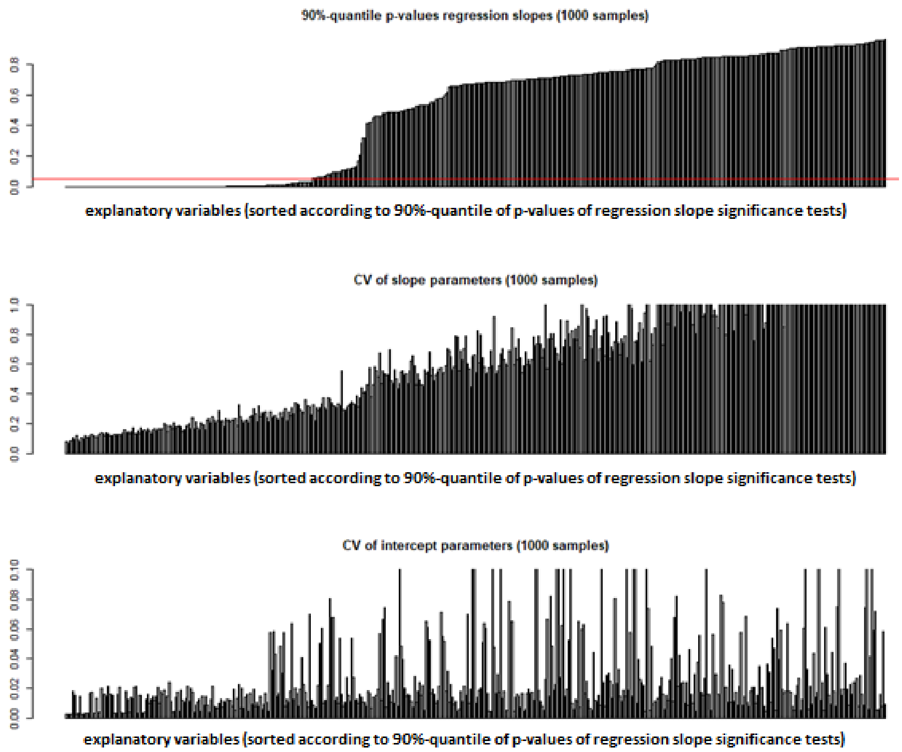

90%-quantile p-values regression slopes (1000 samples)
explanatory variables (sorted according to 90%-quantile of p-values of regression slope significance tests)
CV of slope parameters (1000 samples)
explanatory variables (sorted according to 90%-quantile of p-values of regression slope significance tests)
CV of intercept parameters (1000 samples)
explanatory variables (sorted according to 90%-quantile of p-values of regression slope significance tests)

Fig. 2.