Responses Referee #1

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 difficult 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 find 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 4∘C. 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, finding the most informative variables etc.) than in drought prediction. In this perspective, I think the paper would benefit 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, find 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 find 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 fit 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 specific comments. Specific 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…. 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

- *"Chi²" 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 refit the model or just check if the regression fitted for original series is still valid for the resampled series (they say "the regression is significant"). Any case, they further write that "if the regression is significant 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 significant slope coefficient." I have several concerns regarding the procedure: - It would be more natural to allow for some fraction of insignificant samples. The requirement that ALL subsamples yield significant regression seems to be quite strong - If the models were refitted for each subsample then in principle even if the regression is significant, the coefficient can be oposite, or very different... The regression may be also insignificant due to natural variability not as a result of non-stationarity. - If the models were not refitted 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 fitted to a past period of time to predict 'future'". I agree that it is useful to

assess the performance of model fitted 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 fitting 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 defined

- *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 configurations...". It seems that in the boxplots also the models that are not appropriate are presented. I find 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 fitting of models with linear time dependence of the coefficients (validated via likelihood ratio tests) was tried ... but could not yield the expected results for the small calibration period." - could you be more specific 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 figure 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.*

Responses Referee #2

General comments

p.9, L.25-33: Restricted models: "For the OLS, GLS and principal component model, a second variant with restricted variable selection is applied, via resampling 30 yrs. of data by bootstrapping." I wonder what criterion is exactly used to separate significant time-dependence from randomness? (the criterion "all subsamples" would include randomness, which is normally restricted to some confidence level alpha of e.g. 5%). It is also said that "The random sampling disrupts the chronological order of the time series." And "If the regression is significant for all subsamples, the relationship is considered continuously stationary." I guess not the chronological order is of matter, but the change of coefficients for different time slices?

- *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."*
- *Testing the significance of slopes fitted to individual subsamples implies analysis of the variability in the regression parameters, as shown in Fig. 2 (test carried out for random station). The advantage of using the significance test is that specific thresholds can be deduced (5% significance level). The description has been adjusted, "disrupting the chronological order" has been removed.*

p.14, L.10– p.15, L.19: Modeling of other low flow indices: This is a very useful discussion but should be extended to include the process-based model as well. The latter is expected to better cover the dynamics of low flow events and this should be an advantage of the more dynamic indices,such as onset, duration, peak. It would be interesting to learn more about this point. I think that such an evaluation might be not too difficult as predictions of the process-based model are already available. A discussion could be added under the assumption that NM7Q is a relevant fitting criterion for the low flow parts of the hydrograph and therefore useful for all mentioned low flow statistics.

- *The time series simulated by the hydrological model have been used to compute all considered indices, which are directly compared to the indices simulated by the statistical approach. The NM7Q appeared to be a good fitting criterion for the hydrological model to reproduce the NM7Q, NM30Q, Q95 and Q80. In order to better reproduce volume and duration based indices, a second calibration has been applied, using the Vmean as fitting criterion, which showed one of the lowest validation performances for the statistical approach. The section and belonging Figure have been changed accordingly.*

p.21, L.18 ff: Analysis of temporal shifts: I agree that the analysis of winter low flow indices provide some evidence for seasonal shifts (although more specific analyses should be more informative). However, I find the analysis of annual NM7Q not conclusive, as most catchments may be expected to have a clear summer regime and this is simply reflected in the annual values in the same way as in the summer low flow indices. Suggest to leave it out.

- *We agree that the analysis of the annual values is redundant and removed it from the manuscript. Numbers that state the percentages of summer and winter low flows in past and future periods have been added instead.*

p.23, L.10: "Non-stationarity within the relationship ... appear to be an issue, ... Ideally the models should be revised through inclusion of methods to map potential non-stationary processes and interrelationships." I guess this conclusion is not comprehensive enough, and should be formulated more strongly. Stationarity in the sense of parameter-stability is a pre-requisite for future predictions (e.g. Parajkaetal. (2016)). In your MS you have proposed a model selection method that allows only for stable parameters during model selection. The question remains if this is sufficient to guarantee parameter stability in the future... Priority could have been given to the restricted PC model for the sake of more robust future predictions (cp. To p.16, L.4).

- *The discussion with regard to non-stationarity has been intensified (section 4.1).*
- *A discussion about the limitation of the restricted models for future application has been added (section 4.3).*
- *The restricted PC model is now used for the assessment of future changes.*

The paper performs a spatio-temporal assessment on a station-by station basis. A reference to space-time methods could be given in the outlook, together with some expectations based on the assessments of Fangmann (2017).

- *Outlook was added (p.33, l.12-17).*

Minor comments

Sect. 4.1 Model performance: I think the term G.O.F measures is not properly used here: All of the measures are "performance measures", only those referring to the calibration (=fitting) period are usually termed "G.O.F" measures and the ones referring to an independent validation period "predictive performance". Please clarify and make consistent use of terms to avoid confusion (text and figure captions).

- *GOF has been changed to "predictive performance" where appropriate.*

p. 17 L.1: "NSE of the ranked simulated and observed index time series" - Does this mean that you are assessing rank statistics (ranks) instead of original data? Pls. Clarify.

- *The assessment is not based on rank statistics but simply on the deviance in the original data of the two ordered samples (due to the temporal dissociation between climate model data and observed data). Clarification is added in the text.*

p.21, L.3: Can the effect of area be made more explicit, e.g. by adding some statistical figures in the text, or by co-plotting the area in Fig. 12?

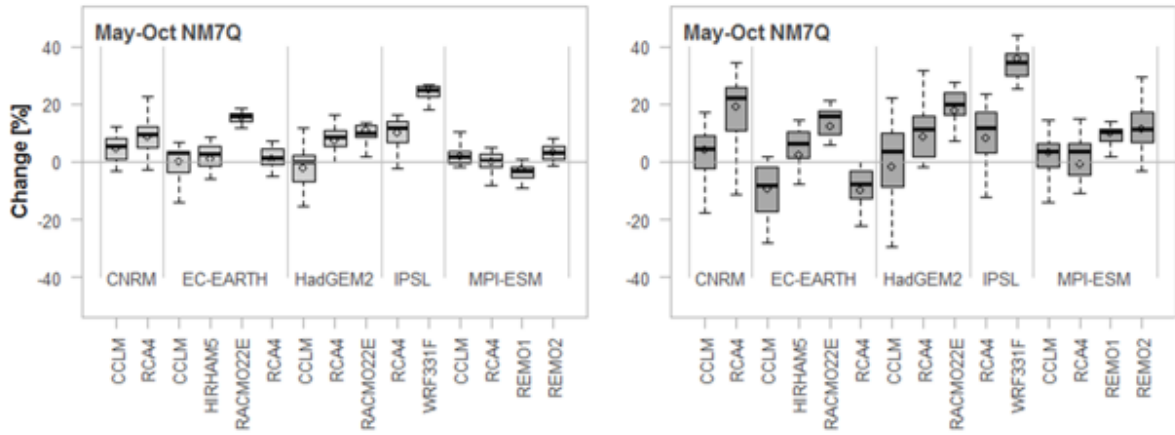
- *A Figure showing the distribution of area has been added.*

p.23, L. 6: Modelling of annual low flow values as a function of annual meteorological indices - I think annual and seasonal would be correct (cp. to temporal index scheme in Fig. 2, and p.11, L.22-23) – pls clarify.

- *By "annual" we wanted to state that one value per year was considered, independent of its base period used for computation. In order to avoid misunderstandings "annual" was removed.*

Technical comments See the annotated MS for further, mainly technical comments on the MS.

- *The technical comments have been acknowledged and changes have been made to the affected sections. Please see the attached supplement.*

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

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

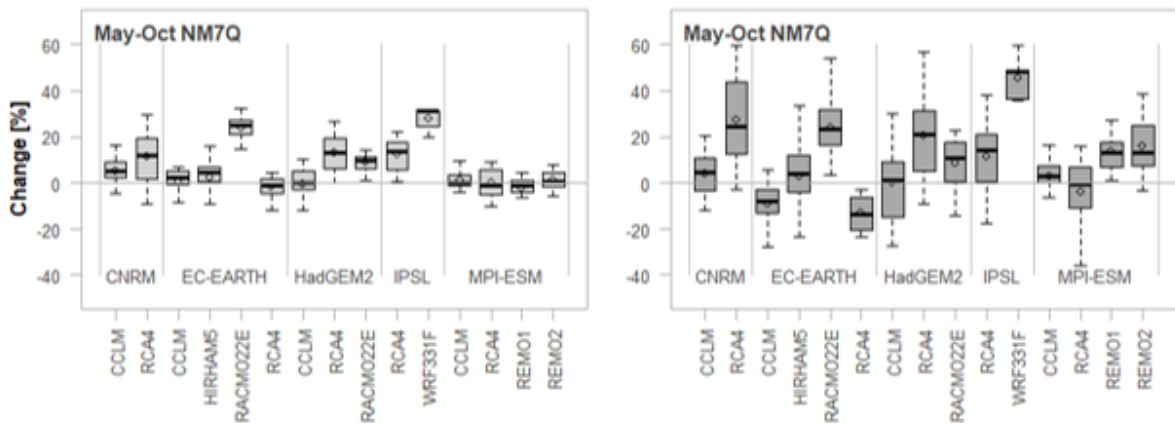



Figure 1: Prognoses after re-calibration.

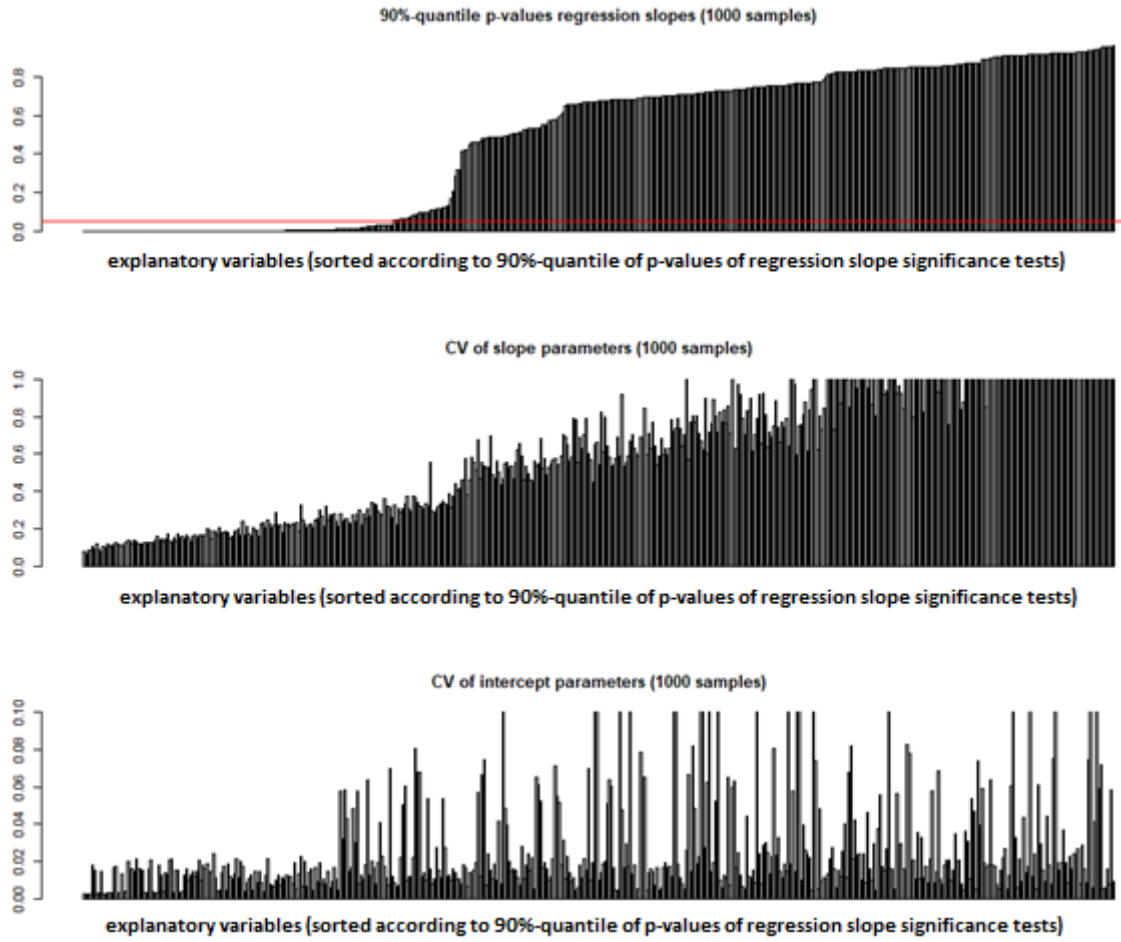



Figure 2: Selection of potentially stationary variables according to significance of regression slope (random station 1000 subsamples).

List of main changes made to the manuscript

- The objective of the manuscript was restated, highlighting more the identification of relevant processes that influence low flow formation. This lead to changes in title, abstract and introduction.
- An additional section was introduced dedicated to a more detailed analysis of meteorological low flow drivers
- The performance of the hydrological models in simulating different low flow index types was additionally assessed.
- The models used for low flow prediction were changed from the non-restricted to the restricted principal component model, assuming these models are more robust.
- The calibration strategy for future low flow prediction was altered: instead of using the observed data, the models were calibrated using the individual climate model data. This procedure aims at minimizing potential bias.
- Clarifications of several methodological aspects were added.

Statistical approaches for ~~assessment~~ identification of ~~climate change impacts on~~ low flow~~s~~ drivers: temporal aspects

Anne Fangmann[1], Uwe Haberlandt[1]

[1]Institute of Hydrology and Water Resources Management, Leibniz University of Hannover, Germany

Correspondence to: Anne Fangmann (fangmann@iww.uni-hannover.de)

Abstract. The characteristics of low flow periods, especially regarding their low temporal dynamics, suggest that ~~estimation of~~ the dimensions of the metrics related to these periods may be ~~carried out~~easily related to their meteorological drivers using simplified statistical model approaches~~that base on a rudimentary input of aggregated local meteorological information. Compared to physically-based or even strongly conceptualized hydrological models, such approaches may have the advantage of being easily set up, applicable over large study areas in a fraction of the time, and easily transferrable between regions, given that predictions are made with the accuracy required for a given purpose.~~ In this study, linear ~~simplified~~ statistical models based on multiple linear regressions (MLR) ~~for the use in regional climate change impact analysis~~ are proposed. Study area is the German Federal State of Lower Saxony with 28 available gauges for analysis. A number of regression approaches are evaluated. An approach using principal components of local meteorological indices as input appeared to show the best performance. In a second analysis it was assessed whether the formulated models may be eligible for application in climate change impact analysis. The~~is~~ model~~s~~ ~~type was eventually~~were therefore applied to a climate model ensemble based on the RCP8.5 scenario. Analyses in the baseline period revealed that some of the meteorological indices needed for model input could not be fully reproduced by the climate models. The predictions for the future show an overall increase in the lowest average 7-day flow (NM7Q), projected by the majority of ensemble members and for the majority of stations.

1 Introduction

~~The assessment of climate change impacts on streamflow has been a long-established field in hydrology. Future hydrological change has been projected all over the world using various types of emission scenarios, climate model ensembles, impact models and other sources of information.~~

~~Analysis of the climate impact on streamflow usually requires a model that allows making predictions of a desired flow metric in time based on expected changes in the climate. The probably most common approach in hydrological impact analysis is to drive hydrological models, usually rainfall-runoff models of variable complexity, with regional climate model data input, obtained for various emission scenarios. Also for low flows this model chain for impact assessment has been applied in several studies, e.g. by de Wit et al. (2007), Schneider et al. (2013), Forzieri et al. (2014), Wanders and Van Lanen (2015), van Vliet et al.(2015), Roudier et al. (2016) or Gosling et al. (2017).~~

~~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 difficult 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. Alternative, simpler impact model approaches that are easily set up and applied to large numbers of catchments and extensive climate model ensembles - while predicting hydrological quantities with sufficient accuracy - may thus offer a different, yet convenient tool for comprehensive regional climate change impact analyses.~~

Low flows appear ideal for the ~~development of alternative model approaches~~analysis of statistical dependencies with their meteorological drivers, due to their specific characteristics. As opposed to high flows, low streamflow is usually much less dynamic and its meteorological drivers, i.e. primarily a lack of rain and increased evaporation, are observable over much larger scales in both time and space. Accordingly,

~~Se~~several studies have aimed at explaining observed peculiarities in low flows by climatic phenomena, including Mosley (2000), who analyzed the influence of El Niño and La Niña effects on monthly lowest 7-day flow in New Zealand. He determined a major deviation of low flows from the normal in La Niña years. Haslinger et al. (2014) found that correlations between streamflow anomalies and meteorological drought indices are high, especially within the low flow period. Van Loon and Laaha (2015) analyzed the dependence of streamflow drought duration and deficit volume on climatic indicators and catchment descriptors, while Liu et al. (2015) estimated the parameters of a GEV distribution fitted to the low-flow time series at their gauge in China as functions of a set of climatic indices.

~~Based on the special characteristics of low flow periods~~Based on these findings it is hypothesized that streamflow metrics describing ~~these~~ low flow events may be ~~predicted~~ efficiently assessed based on their relationship to metrics of meteorological states. This assumption entails that simple statistical models may be formulated that allow ~~prediction of future~~a reproduction of low flow~~s~~ metrics based on ~~any~~ meteorological input~~, including climate model data ensembles and extrapolated climate observations~~. In this study the relationship between annual low flow and a variety of meteorological indices will be analyzed with the aim of formulating simple statistical models. Such~~These~~ models will be set up individually for each catchment in the study area and are supposed to be capable of predicting low flow as a simple function of several relevant meteorological indices in time. Various temporal scales will be considered for the meteorological data and the most adequate ones for each catchment will be identified. The model fitting itself will be carried out using various statistical approaches in order to account for different effects, like non-stationarity or serial dependence in the annual low flow indices. In order to assess the performance of t~~T~~he strictly linear statistical approaches, they will ~~will~~ be compared to~~with~~ a hydrological model, which is assumed to accurately reproduce rainfall-runoff relationships by involving more complex processes, relying on continuous meteorological input. This analysis will reveal whether the statistical models fail to acknowledge important non-linear relationships.

Apart from identifying relevant low flow drivers, it is tested whether the use of the formulated statistical models may be extended to prediction of future low flows, using meteorological input from climate models. ~~Analysis of the climate impact on streamflow usually requires a model that allows making predictions of a desired flow metric in time based on expected changes in the climate.~~ The probably most common approach in hydrological climate change impact analysis is to drive hydrological models, usually rainfall-runoff models of variable complexity, with regional climate model data input, obtained for various emission scenarios. Also for low flows this model chain ~~for impact assessment~~ has been applied for impact assessment in several studies, e.g. by de Wit et al. (2007), Schneider et al. (2013), Forzieri et al. (2014), Wanders and Van Lanen (2015), van Vliet et al.(2015), Roudier et al. (2016) or Gosling et al. (2017).

The statistical models may pose an alternative simpler impact model approach that is easily set up and applied to large numbers of catchments and extensive climate model ensembles - while predicting hydrological quantities with sufficient accuracy. They may thus offer a different, yet convenient tool for comprehensive regional climate change impact analyses. According to this hypothesis, the formulated models will

5   ~~and after validation be eventually~~be applied to an ensemble of regional climate model data in order to assess the plausibility of their low flow projections in the future.

~~The assessment of climate change impacts on streamflow has been a long-established field in hydrology. Future hydrological change has been projected all over the world using various types of emission scenarios, climate model ensembles, impact models and other sources of information.~~

2 Study area and data

2.1 Study area

The area under investigation is the federal state of Lower Saxony situated in northwestern Germany. Its total area of 47,634 km^2 extends from the North Sea in the northwest to the Harz mountains in the south-east. The largest portion

15   of Lower Saxony is part of the Northern German Plain, where the terrain is generally shallow. Highest elevations are found in the secondary mountains of the Harz with up to 971 m a.s.l.. The topography is depicted in Fig. 1.

2.2 Base data

The meteorological data used in this work comprises of daily time series of precipitation, temperature and global radiation. The data is made available continuously in space via interpolation onto a 1x1-km raster for the total area of

20   Lower Saxony. Regionalization is based on 771 precipitation gauges and 123 climatic stations and carried out for a period ~~of~~from 1951 to 2010. The regionalization process is described in detail in NLWKN (2012).

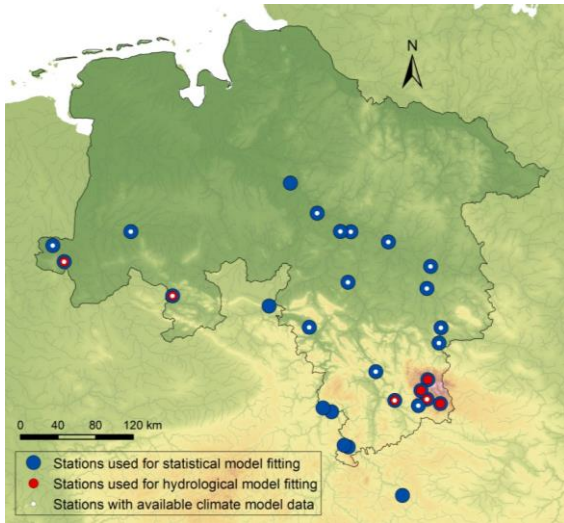



Figure 1: Study area with stations available for different analyses.

Also for streamflow daily time series are used. In total, there are 353 flow records of average daily discharge available in and around the study area. Record lengths range from 7 to 193 years. For analysis in this work, gauges have been selected that have a maximum overlap with the available climate data, i.e. available records from 1951 to 2010. After removal of heavily influenced gauges, 28 stations remain for analysis, as depicted in Fig. 1. The belonging catchment sizes range from 24 to 37,720 km². In previous work (NLWKN, 2017), Aa hydrological model could only behas been set up for 7 of these stations, as indicated in Fig. 1Fig. 1. This selection of stations will be used for comparing statistical and hydrological model performance.

2.3 Indices

The daily time series of meteorological and streamflow data are used to compute indices, which pose the basis for all models. Target variables are a variety of annual low flow indices. They are selected in a way to represent different quantities of interest for low flow analysis (e.g. intensity, duration and timing). The low flow indices used in this study are shown in Table 1Table 1. In order to exclude winter low flows, which may be subject to different underlying processes, the indices are computed for the summer half year (May to October) only. Within the region most low flow events occur between between June and October.

Table 1: List of low flow indices.

Index	Unit	Description
Q_{95} Q_{80}	m³/s	5- and 20-percent non-exceedance quantile of the daily average discharge
NM7Q NM30Q	m³/s	Lowest 7- and 30-day average flow
V_{max} V_{mean}	m³	Maximum and mean deficit volume: sum of daily discharge below the long term 20-percent non-exceedance quantile
D_{max} D_{mean}	d	Maximum and mean low flow duration: number of days with discharge below the long term 20-percent non-exceedance quantile
timing	-	Day of the year at which smallest daily flow occurs

The selected low flow indices represent a small fraction of possible indices but cover quite a range of low flow quantities relevant for water resources management and planning. Non-exceedence quantiles represent the overall low flow situation in a year without specific relation to a certain event. The same holds for average volume and duration.

5   NMxQ, maximum volume and duration, as well as timing, on the other hand, focus on the greatest events per year in their respective terms and characterize them accordingly. The consideration of variations within the individual index types (e.g. Q_{95} and Q_{80}) is not solely owing to different management requirements but used to investigate the various models' capability to reproduce both more and less extreme low flows.

10   Meteorological indices are used as input for the models. Like for their low flow counterparts, annual values are calculated from daily time series of precipitation sums, mean and maximum temperature and global radiation, averaged over the respective basins of the considered discharge gauges. Table 2Table 2 summarizes the applied indices and gives a short description.

Table 2: Meteorological indices based on precipitation, temperature and potential evapotranspiration.

Index	Unit	Description
P_{mean}	mm/d	Average daily precipitation
P_x	mm/d	Non-exceedance quantiles of daily precipitation sums
SPI	-	Standardized Precipitation Index: standardized deviation of accumulated precipitation sums from the long term normal (Mckee et al., 1993)
DSD_{mean} DSD_{max}	d	Mean and maximum dry spell duration: number of days with preciptation <1mm/d
WSD_{mean} WSD_{max}	d	Mean and maximum wet spell duration: number of days with preciptation >= 1mm/d
T_{min} T_{mean} T_{max}	°C	Minimum, mean and maximum daily average temperature
HWD	d	Heat wave duration: Number of days above 90-percent non-exceedance quantile of maximum temperature calculated for each specific day of the year
ETP_{mean}	mm/d	Average daily potential evaporation calculated according to Turc-Wendling
$P\text{-}ETP_{mean}$	mm/d	Average climatic water balance: precipitation minus potential evapotranspiration
P/ETP_{mean}	-	Aridity index: Ratio of average precipitation and potential evapotranspiration.
SPEI	-	Standardized Precipitation Evaporation Index: standardized deviation of accumulated climatic water balance from long term normal (Vicente-Serrano et al., 2010)

While the low flow indices are calculated for a fixed period within the year (the summer half year), the period for computation of the meteorological quantities is varied in length and position. Figure 2Figure 2 shows the scheme according to which every meteorological index has been computed. I_{NQ} represents the low flow index, calculated for its fixed period within a given year. I_M denotes the meteorological index. The first number in the index notation represents the length of the base period for calculation (eg. 3, 6 or 12 months), while the second one indicates the lead time, i.e. the number of months the period is shifted back in time relative to the low flow index. A 0-shift indicates that the calculation period ends simultaneously with the low flow period.

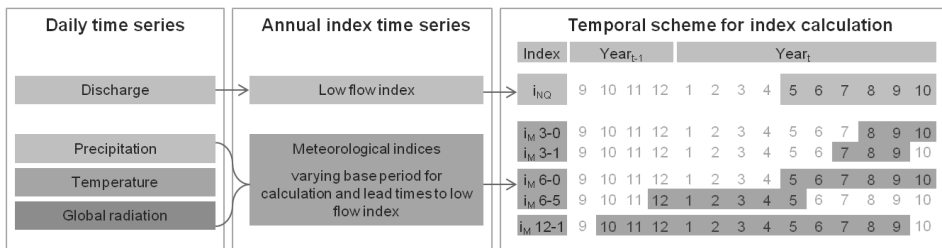



Figure 2: Calculation scheme for low flow indices with fixed base period and meteorological indices with varying base period and lead times relative to the low flow calculation period.

2.4 Climate model data

An ensemble of 15 coupled global and regional climate models is applied for climate change impact analysis. An overview is shown in Table 3~~Table 3~~. The data sets base on simulations of global climate models of CMIP5 run on an RCP8.5 scenario. The data is available on a 10x10-km grid.

The climate model data has been pre-processed in previous work (NLWKN, 2017) and is available for a smaller domain than the observed climate data. Therefore, basin averages can only be computed for 17 of the 28 stations, as shown in Fig. 1~~Fig. 1~~. ~~Temperature and precipitation have been bias corrected for all model chains, using a monthly linear scaling. The procedure is described in detail in NLWKN (2017).~~

Table 3: List of coupled global and regional climate models used for analysis.

Global model	Regional model	Short name
CNRM-CERFACS-CNRM-CM5 run1	CLMcom-CCLM4-8-17	CNRM-CCLM
	SMHI-RCA4	CNRM-RCA4
ICHEC-EC-EARTH run12	CLMcom-CCLM4-8-17	EC-EARTH-CCLM
ICHEC-EC-EARTH run3	DMI-HIRHAM5	EC-EARTH-HIRHAM5
ICHEC-EC-EARTH run1	KNMI-RACMO22E	EC-EARTH-RACMO22E
ICHEC-EC-EARTH run12	SMHI-RCA4	EC-EARTH-RCA4
MOHC-HadGEM2-ES run1	CLMcom-CCLM4-8-17	HadGEM2-CCLM
	RACMO22E	HadGEM2-RACMO22E
	SMHI-RCA4	HadGEM2-RCA4
IPSL-INERIS-CM5A-MR run1	SMHI-RCA4	IPSL-RCA4
IPSL-CM5A-MR	WRF331F	IPSL-WRF331F
MPI-M-MPI-ESM-LR run1	CLMcom-CCLM4-8-17	MPI-ESM-CCLM
	SMHI-RCA4	MPI-ESM-RCA4
	REMO2009 run1	MPI-ESM-REMO1
	REMO2009 run2	MPI-ESM-REMO2

3 Methods

The aim is to model a desired low flow index on an annual basis as a function of a combination of meteorological indices observed in time. Several methods will be analyzed for this purpose in order to identify the one most suitable to model said relationship under consideration of later application for far future predictions. These methods are described in the following.

3.1 Multiple linear Regression

Multiple linear regression (MLR) is probably the most common method to model relationships between variables and will pose the basis for several methods in this work. It aims at reproducing a target variable y as linear combination of k explanatory variables $x_1, \ldots x_k$. The general shape of a multiple linear regression model is the following:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon \tag{1}$$

The regression coefficients $\beta_1, \ldots \beta_k$ are estimated in two different ways, i.e. a) an ordinary least squares (OLS) procedure and b) a generalized least squares (GLS) fitting. The latter method involves the inclusion of the covariance matrix of the residuals, which allows for correction for both heteroscedasticity and dependence of the residuals. Since heteroscedasticity did not appear to be e a relevant issue in this study, while serial correlation in the data did, autoregressive (AR) covariance structures of various orders are tested to describe the residual covariance. The elements of an AR(1) covariance structure, for example, can be described via

$$\sigma_{ij} = \sigma^2 \rho^{|i-j|}, \tag{2}$$

where ρ denotes the autocorrelation between terms with lag 1.

The need for and order of the considered AR process is case wise determined using the likelihood ratio test. This parametric test is developed to test the superiority of nested more complex models over their simpler forms, i.e. whether a model with a higher number of parameters performs significantly better than the same model with fewer parameters. Performance is thereby measured in terms of maximized log-likelihood function values and the $(1 - \alpha)$ quantile of the ChiX^2 distribution, whose degrees of freedom are chosen as the difference in number of parameters between the compared models (Coles, 2001).

3.2 Principal component analysis

The fitting of multiple linear regression models is restricted by sample size. Fitting a model with a large number of regressors to a small data set will result in over-fitting. Additionally, the problem of multicollinearity between the regressors in a model arises. Consequently, for a small data set, only a small number of uncorrelated regressors should be selected. At the same time, leaving out important explanatory variables may drastically lower the predictive power of the model. In order to overcome the restrictions given through the limited period of observation, a principal component analysis (PCA) is applied, merging many explanatory variables into a few uncorrelated components that pose the ideal basis for model fitting on limited data with maximum exploitation of information. PCA is carried out by firstly centering the set of p explanatory variables X^* via subtraction of their respective means. Then, the (p x p) covariance matrix Σ is computed and its eigenvalues $\lambda_1, \lambda_2, \ldots \lambda_p$ and eigenvectors $\gamma_1, \gamma_2, \ldots, \gamma_p$ are determined. Multiplication of the eigenvectors with X^* yields the following system of equations

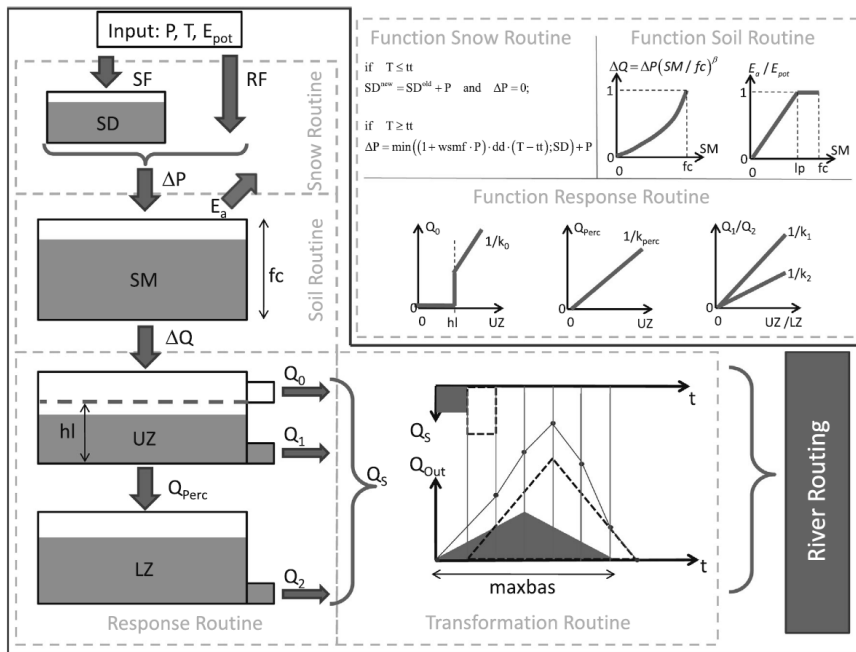
$$y_1 = \gamma_{11}x_1 + \gamma_{21}x_2 + \ldots + \gamma_{p1}x_p$$
$$y_2 = \gamma_{12}x_1 + \gamma_{22}x_2 + \ldots + \gamma_{p2}x_p \quad ,$$
$$\vdots$$
$$y_p = \gamma_{1p}x_1 + \gamma_{2p}x_2 + \ldots + \gamma_{pp}x_p$$

(3)

where y_1, y_2, …, y_p denote the principal components, sorted according to their contribution to the total variance of data at hand. The eigenvectors represent the respective loading of a variable x on a principal component y.

3.3 Hydrological modeling

5   In order to evaluate the performance of the statistical approaches, hydrological modeling is applied as the benchmark for prognosis of future flow. The rainfall-runoff model used here is an adaptation of the Swedish HBV by Lindstrom et al. (1997), denoted HBV-IWW. The peculiarities of the model are described in detail e.g. in Wallner et al. (2013). A schematic overview over the structure of the model is presented in Fig. 3Fig. 3. The model is semi distributed and can be applied on the sub-catchment scale. Input is daily precipitation, temperature and potential evapotranspiration.

10 The model comprises of 5 routines and a variety of parameters controlling the translation between the various model components. The parameters are automatically optimized using the evolutionary multimethod algorithm AMALGAM (Vrugt and Robinson, 2007).

15   **Figure 3: Vertical structure of the HBW-IWW model according to Wallner et al. (2013)**

3.4 Model fitting and evaluation

Given the vast number of meteorological indices, variable selection for the statistical models becomes paramount. In order to select appropriate indices and restrict the numbers of selected regressors, a two-way stepwise approach for variable selection is chosen that aims at minimizing the Bayesian information criterion (BIC; Schwarz,1978) of the final model. The BIC is closely related to the Aikaike information criterion (AIC; Aikaike, 1974) but penalizes variable inclusion stricter for larger sample sizes, which is the reason for its selection in this study. The smaller the BIC, the better the predictive power of a model. The algorithm for variable selection randomly selects a variable and adds variables to the model that decrease the overall BIC. After each addition it is tested whether the removal of any of the existing variables in the model leads to further decrease in the BIC.

Since the number of potential regressors in this study is high and variables may be strongly related, a second criterion is included into the BIC-minimization procedure. In order to prevent multicollinearity, the variance inflation factor (VIF) of each variable added to the model is computed. The VIF is a measure of how much of the variance of a predictor in a model can be explained by the other predictors in the same model. Computation is done via

$$\text{VIF} = \frac{1}{1 - R^2} \, , \tag{4}$$

where R² denotes the coefficient of determination of a regression model that describes the variable to be added as a function of the existing variables in the model. The recommended limits for the VIF vary throughout literature. Since multicollinearity is a major issue for the analyses of this study, a low value of 5 has been chosen. Variables that show higher VIFs are not included into the existing model.

For the GLS models with AR-correlation structure, variable selection was carried out according to OLS model fitting, evaluating the necessity of the AR-structure for each added variable using the likelihood ratio test. For variable selection using principal components, the BIC-minimization strategy is extended, not adding each additional variable directly to the model but computing principal components first and testing them as potential regressors of the model.

For the OLS, GLS and principal component model, a second variant with restricted variable selection is applied. This approach aims at filtering out variables that have a non-stationary relationship to the target variable. This filtering is achieved via resampling. Randomly, 30 years of data are sampled with replacement from the calibration data set 1000 times. A simple linear regression model is fitted between target variable and all potential explanatory variables for each subsample. ~~The random sampling disrupts the chronological order of the time series.~~ Testing the significance of regression slopes fitted to random samples gives an estimate of the variability of the slope coefficient over segments of the time series. If the regression is significant for ~~all~~ more than 95% of the 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 significant slope coefficient. Only the former variables are included in further variable selection. The models with restricted variable selection will be denoted OLS-R, GLS-R and PC-R.

In contrast to the statistical models the hydrological model bases on daily input and output and is therefore capable of modeling an entire years' hydrograph rather than a single low flow metric. Calibrating the hydrological model on the complete hydrograph may reduce the predictability of the single low flow event and favor the statistical models (directly calibrated on the low flow value) in a comparison of model performances. Therefore, the HBV-IWW is

calibrated in a strategy comparable to the statistical approaches, namely by including nothing but the desired annual low flow metric in the objective function.

In order to allow for a proper assessment of the models' capabilities of predicting future low flows, the available record at each station is split equally into a calibration and a validation period. Both periods are continuous in time and the calibration consistently precedes the validation period. This set-up is chosen to test the ability of models fitted to a past period of time to predict "future", i.e. the validation period's characteristics. In order to double the available periods for validation, the time series are ~~reversed~~ inversed and models are fitted to the former validation period and evaluated in the calibration period. The inversion is applied in order to preserve the continuity of the time series.

In order to compare the quality of the different model approaches, a selection of performance ~~goodness of fit~~ measures is computed. These measures include the coefficient of determination (R^2), the normalized root mean square error (NRMSE) and the percent bias (pbias). The criteria are selected to show various aspects of model quality, i.e. information about the similarity of the course of simulated and observed time series, the average fit, and systematic errors.

3.5 Climate change impact analysis

The ~~best performing~~most suitable statistical model is eventually applied using the ensemble of climate model data. The first step is the assessment of the performance of the model chains in reproducing the observed low flow. This analysis is done in a reference period from 1971-2000. The projections and changes in the low flow will be analyzed in a near (2021-2050) and far future (2071-2100) period. It should be noted that the aim of the prognosis using actual climate model data is primarily the validation of the statistical model approaches rather than a detailed regional assessment of expected changes.

In order to account for potential bias in the climate model data used for prediction of future low flow, the individual station models are recalibrated for each climate model separately. The problem that needs to be overcome for this process is the matching of the temporarily dissociated series of meteorological indices generated by the climate models with the observed low flow time series, which is handled as follows. At first, the explanatory variables identified in the observed model fitting procedure are carried over, i.e. no new variable selections are carried out. This is meaningful, since the selected variables represent observed low flow drivers. The principal components are then re-estimated for the selected variables as projected by the individual climate models. In order to relate the climate model data to the observed time series, the regression parameters are optimized in such a way, that the empirical distribution of the predicted low flow matches that of the observed sample. In order for the recalibrated model to be no more precise than the originally fitted model, which may lead to false estimates of the regression coefficients, the model fitted to the observed data is used as target during re-calibration rather than the observed low flow series itself.

4 Results and discussion

4.1 Model performance

Performance of the individual MLR model variants are compared using the aforementioned performance goodness of fit measures. Figure 4Figure 4 shows the model performance of all tested model configurations exemplary for NM7Q prediction in the calibration (top) and validation period (bottom). For the validation period one can clearly observe an increase in model performance from left to right, i.e. an increasing predictive power with increasing model complexity. The non-restricted OLS-fitted model shows poorest performance with a median R^2 of only 0.43. The removal of variables with a potential non-stationary relationship to the target variable increases model performance significantly, as seen for the OLS-R model. The R^2 increase in median to 0.48 and NRMSE drops from 78.9 to 72.7%. The restriction in the variable selection thereby leads to an alteration of 58% of the originally fitted OLS models.

Highest goodness of fitpredictive performance has been achieved with the principal component model without restriction during variable selection (PC). The coefficient of determination for the MLR model over all stations has a median value of 0.58. The PC-model also shows smallest percent bias with a median of 1.30 %. The PC-R model, where potential non-stationary variables have been removed, shows slightly lower performance. On the other hand, the restriction significantly reduces the number of variables used for formation of the principal components and hence potentially yields more robust parameter estimates for future predictions. Either way, Tthe inclusion of several variables in the form of principal components appears meaningful in order to avoid omittance of important regressors and potential multicollinearity, granted by the orthogonality of the components. Additionally, the principal component models areis supposedly more robust against outliers in single variables

Second best performance shows the GLS-fitted model with an AR-correlation structure and restriction during variable selection via exclusion of variables with non-stationary relationship to the target variable (GLS-R). One should note that not for every station an AR-correlation structure has been used for model fitting. Only at those stations, where the likelihood ratio test favored an inclusion, a GLS-fitting was carried out. This was the case for 55% of the fitted models (35% AR(1) covariance structure, 20% AR(2) covariance structure). The bias in the calibration period arises through the approximation of the unknown true autocorrelation structure in the data. Since the values are negligibly small, this approximation is considered reasonable. As the number of variables included in the GLS model is significantly lower compared to the principal component model and the step of finding suitable components can be skipped, the performance based on the close validation period, as shown here, appears to be comparably good. However, inclusion of AR-processes via likelihood ratio testing results in some distortion of the residuals and violation of the prerequisites of linear regression fitting at several stations. Thus, the principal component model is considered favorable.

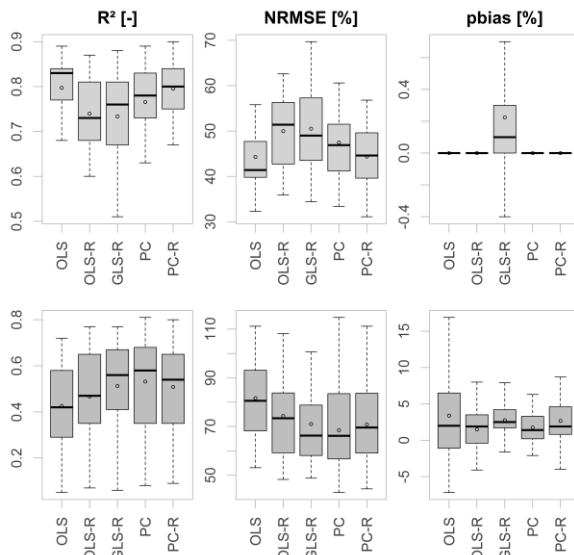



Figure 4: ~~Criteria for goodness of~~Performance criteria ~~fit~~ of different model approaches in the calibration period (top) and validation period (bottom) over 28 stations.

5   ~~The PC models for NM7Q estimation are used exemplary to evaluate the selected explanatory variables over all stations. The majority of models comprise of one or two principal components made up by 3-8 meteorological indices. Many of the selected variables appear to repeat for most of the models. The most common predictors of the NM7Q appear to be the SPI of various base periods and lead times (0 to 3 months). The second most frequent regressor is the aridity index P/ETP for 3- and 6-month base periods and for 1-4 months lead times, as well as the P-ETP for the same~~
10   ~~periods. Upper precipitation quantiles with various base periods and lead times follow in frequency. Several models also contain the ratio between average DSD and WSD as predictors. Pure temperature-based predictors are rare. Maximum HWD is considered in one model only.~~

Figure 5~~Fig. 5~~ shows the same criteria as Fig. 4~~Fig. 4~~ with inclusion of the hydrological model, and thus for a total of

7 stations only. The performance of the HBV model in both calibration and validation period is substantially higher in

15   all aspects with median validation R² of 0.50, an NRMSE of 76.1%, and a percent bias of 1.7% compared to the PC-model with median values of 0.42, 78.4%, and 5.8% respectively. This performance, however, could only be achieved using the calibration strategy of matching exclusively the annual NM7Q values at each station. Calibration using the entire hydrograph or parts of the flow duration curve did not explicitly outperform the statistical approaches. Nevertheless, HBV-IWW appears to be suitable for assessment of future low flow development when calibrated

20   specifically on ~~the desired~~ low flow ind~~iece~~ces~~x~~.

The statistical models appear to be positively biased. At almost all investigated stations a positive mean error could be observed. As models are fitted in both directions to the time series, this error appears to originate in the model itself, rather than in underlying processes of the time series that cannot be captured. Overall differences in model performance

between the stations could not be correlated to any specific regions or catchment features. The quality of the simulation did not depend on catchment size, position in the study area or any other distinguishable factor.

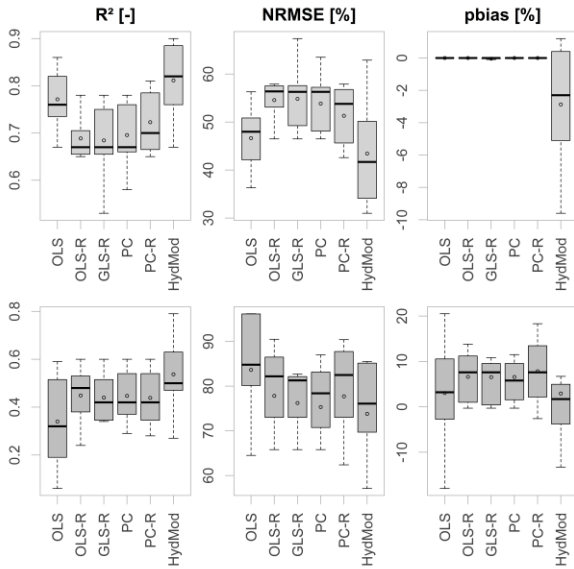



Figure 5: Criteria for goodness of fit of different model approaches in the calibration period (top) and validation period (bottom) over 7 stations.

An obvious problematic is the significant difference in model performance between the calibration and validation period. All quality criteria certify a much better performance during calibration than during validation. Table 4Table 4 and Table 5Table 5 show the median differences in goodness-of-fit measures for all model variants over all stations and over the stations at which hydrological modeling was possible. The differences are severe, even though quite a number of precautions have been taken during calibration. The R^2 for unrestricted PC, for example, differs by 0.22 between calibration and validation. For the 7 stations this difference increases to 0.23. Overfitting appears to be an issue, even though the number of regressors has been restricted. However, the differences are also observable for the hydrological model, in some cases even more drastic than for the statistical models. The median difference in R^2 over the 7 stations is 0.3 between calibration and validation period.

Table 4: Median absolute difference in quality criteria between calibration and validation period for 28 stations.

	OLS	OLS-R	GLS-R	PC	PC-R
NRMSE	37.8%	24.6%	20.3%	20.8%	26.2%
PBIAS	5.1%	3.2%	2.7%	2.7%	3.2%
R²	0.37	0.27	0.22	0.22	0.28

Formatiert: Schriftart: 10 Pt.

Formatiert: Schriftart: 10 Pt.

Through separation of the validation period into three parts it becomes obvious that performance decreases steadily with distance from the calibration period, as shown in Fig. 6~~Fig. 6~~. Compared are the estimated and the observed means for all statistical methods. The error increases also for methods that have been supposedly restricted to stationary relationships between target variable and regressors. The effect shows that non-stationary relationships between low flow and meteorological indices are of major relevance and need to be considered in statistical modeling. Logically, with a changing climate, interrelationships between individual meteorological variables and hence the ratio of the influence of individual drivers on low flow formation will change. At the same time, other forcings and feedbacks may become more or less relevant, which have not been included in the statistical model formulation in the first place. In an initial attempt at encompassing the problem by addressing a potential linear time dependence of the regression coefficients, the parameters have been re-estimated as linear functions of time using a ~~M~~maximum likelihood fitting. ~~of models with linear time dependence of the coefficients~~ (The required complexity of the models has been ~~assessed validated via~~ using likelihood ratio tests. ~~) was tried during this study to encompass the problem,~~ However, model fitting appeared problematic for the small calibration period and did not yield an improvement of the existing models.~~but could not yield the expected results for the small calibration period.~~

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.

Table 5: Median absolute difference in quality criteria between calibration and validation period for 7 stations.

	OLS	OLS-R	GLS-R	PC	PC-R	HydMod
NRMSE	34.4%	25.5%	21.4%	19.3%	25.5%	32.7%
PBIAS	6.9%	8.6%	8.4%	7.6%	8.6%	3.7%
R²	0.35	0.23	0.23	0.23	0.23	0.30

Apart from non-stationary relationship between target variable and regressors, another possible explanation – especially with respect to the observations regarding the hydrological model - could be a significant difference in flow behavior between the two periods selected for calibration and validation, potentially caused by other forcings than local climate. A previous study by Fangmann et al. (2013) has found that the time series at a majority of gauges in the study area are divided by significant break points (according to Pettitt, 1979) between the years 1987 and 1989. Since neither the hydrological nor the statistical models involve other input than local meteorology, this break cannot be accounted for by either approach. Also, anthropogenic interference poses a relevant factor. Even if indirectly considered in the statistical models, changes in the management patterns would significantly alter the prognosis; a factor that needs to be considered during model application.

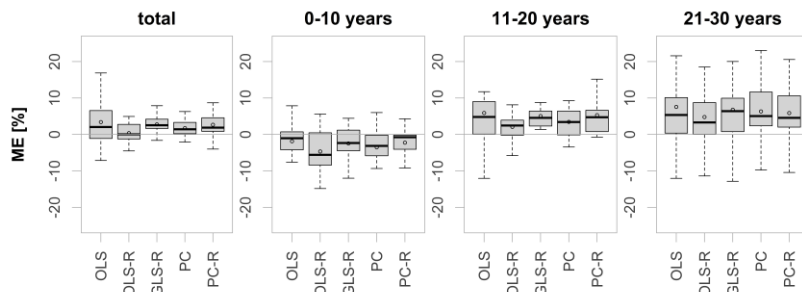



Figure 6: Mean deviation of estimated from observed means for the whole validation period (left), as well as the first, second and final 10 years of validation.

Modeling of the other tested indices showed quite a differentiated picture. As shown exemplary for 7 stations and the restricted GLS approach in Fig. 7~~Fig. 7~~, some indices were reproduced more (top), others less successfully (bottom) than the NM7Q. Estimation of the Q_{95} appeared slightly better in terms of all quality criteria. The D_{max} was modeled effectively via GLS in terms of R^2 and NRMSE (median values of 0.5~~4~~9 and ~~65~~80.~~0~~8 %) but showed quite a large ~~negative~~bias ~~bias (-~~2.1% median~~)~~. V_{max} and especially the annual low flow timing could not be modeled successfully by any of the statistical approaches (median R^2 of 0.~~28~~31 and 0.0~~3~~2, respectively).

It was noted that the overall model performance was slightly higher for the more average values, i.e. NM30Q and Q_{80} than for the more extreme indices NM7Q and Q_{95}. ~~The quality criteria of the more average values exceed the more extreme ones on average, while stations with extremely poor fit are no longer present.~~ The median R^2 values of the NM7Q and the NM30Q compare as 0.~~57~~47 and 0.5~~2~~61, the ones of the Q_{95} and Q_{80} as 0.~~58~~49 and 0.~~63~~62. Compared to the D_{max} and V_{max}, the annual average D_{mean} and V_{mean} cannot be predicted by the fitted models. The validation yields R^2 values below 0.1 for both cases.

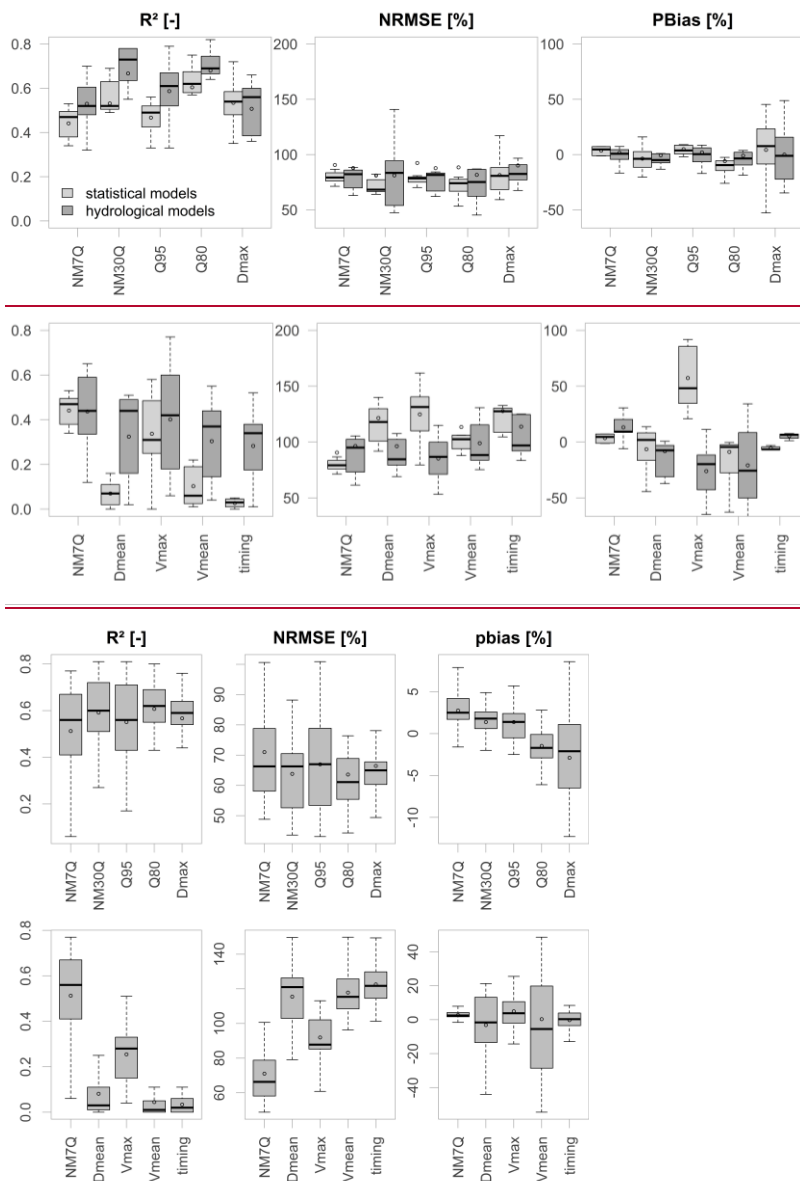



Figure 7: Validation results for the restricted GLS model and the hydrological model for various low flow indices at 7 stations performing better (top) and worse (bottom) than for the NM7Q.

The differences in reproducibility of the various metrics can be explained by the nature of the indices and the structure of the statistical models. The regression model uses aggregated meteorological features over a specific period of time for prediction of a desired low flow variable. The shorter the required period, the lower the degree of averaging and the greater the possibility to capture extremes that cause a subsequent low flow event. Therefore, indices related to a specific event, like the NM7Q or the D_{max}, can be modeled quite effectively based on previous meteorological states. The fact that more average indices like NM30Q and Q_{80} are reproduced better than more extreme ones can be explained by the same principle. Meteorological indices computed for longer base periods are required to model more average index values. Errors that occur if extremes cannot be explained by external variables are lower. Dmean and Vmean, however, do not represent averages of single but of multiple events. These features cannot be captured by small sets of regressors as used in this study.

Considering the models of the other indices, it can be seen that the selected regressors differ between the individual types. While Q95 and Q80 are primarily predicted by water balance parameters like the NM7Q and NM30Q, the models for Dmax, Vmax and timing count a high number of indices related to durations. Alongside the validation results of the statistical models those of the hydrological model are presented in Fig.7. The indices in the top panels are deduced from the simulation runs of the The hydrological mmodels that has been calibrated on the annual NM7Q. In order to better represent the indices related to duration and volume, the model used in the bottom panels has been calibrated using the annual Vmean as fitting criterion. has not been applied to other indices than the NM7Q but is expected to yield significantly better results, especially for deficit volume and low flow timing. One can see that the results for all indices improve for the hydrological model as expected, due to a more detailed process representation. This is especially the case for the average duration and volume based indices. Still, the differences in performance for the individual indices is visible also for the hydrological models.

It should be noted that the selected statistical methods exclusively take linear relationships between target variable and regressors into account. It appeared that these linear dependencies were strong and major portions of the variance in the target variables could be explained by the fitted models. At this point it cannot be precluded that non-linear relationships to some meteorological indices do exist but it is presumed that important influential quantities can be linearly related to the target variable.

4.2 Low flow drivers

The restricted PC models for NM7Q estimation are used exemplary to evaluate the selected explanatory variables over all stations. For the NM7Q Tthe majority of models comprise of one or two principal components made up by 32 88 meteorological indices. Many of the selected variables appear to repeat for most of the models. The most common predictors of the NM7Q appear to be the aridity index P/ETP for 3- and 6-month base periods and for 1-4 months lead times, as well as the P-ETP for the same periods. the SPI of various base periods and lead times (0 to 3 months). The second most frequent regressor is the SPI of various base periods and lead times (0 to 3 months) the aridity index P/ETP for 3- and 6-month base periods and for 1-4 months lead times, as well as the P-ETP for the same periods. Upper precipitation quantiles with various base periods and lead times follow in frequency. Several models

also contain DSD the ratio between average DSD and WSD as predictors. Pure temperature based predictors are ~~rare. Maximum HWD is considered in one model only~~rare, but do occur in some cases.~~.~~ The selected lead times of the temperature based indices are in these cases larger than for the precipitation or water balance indices. Lead times as long as 9 months, which would represent November-January temperature, are observed. Winter temperature hence appears to be an important predictor for summer low flow magnitude.

The inclusion of indices with differing base periods and lead times at the individual stations represents the heterogeneity of transformation processes within the study area. One of the major determinants of how meteorological processes translate into flow is catchment size. Larger, rather slowly responding catchments with significant storage capacity will less likely be affected by small meteorological events and low flows will occur moderated and with significant temporal delay. Therefore, statistical models fitted for these catchments will most likely include meteorological indices that have been computed for longer base periods and lead times. Within small, quickly responding systems, on the other hand, low flow magnitudes may be related to shorter dry periods that have occurred recently. Variable selection is thus carried out in favor of small base periods and lead times. Nevertheless, even for smaller catchments an inclusion of temporally leading meteorological indices may mimic initial storage conditions and determine the absolute magnitude of a summer low flow event.

The models fitted to the NM30Q are comparable to the ones of the NM7Q but contain on average fewer explanatory variables per model. As discussed before, the more average low flow values are more easily predicted, apparently requiring fewer external information. In contrast to the NM7Q, temperature based indices do not appear to be important predictors of the NM30Q. Considering the models of the other indices, it can be seen that the selected regressors differ between the individual types. While Q95 and Q80 are primarily predicted by water balance parameters, as are the NM7Q and NM30Q, the models for Dmax and Vmax count a high number of indices related to durations, especially average and maximum dry spell duration.

The regressors selected for the forward and inversed fitting carried out at each station are not always identical. Some of the independent variables are naturally very similar for various base periods and lead times that they are almost exchangeable, which in turn makes the fitted models very similar. Still, this observation hints again at the problem of non-stationarity and the observed breakpoint near the transition between calibration and validation period.

4.~~24~~.3   Prognosis

For prognosis, the ~~un~~restricted principal component models are applied using the ensemble of climate model data described above. Even though the non-restricted models showed overall higher performance in the observation period, the restriction via removal of potentially non-stationary variables is considered valuable for predicting future flows, especially with respect to parameter stability (Parajka et al, 2016). However, it needs to be noted that also variables, which have been determined as acceptably stationary in a small observed period may still exhibit non-stationary behavior in a more distant future, especially in their interactions.

For reasons of conciseness only the NM7Q is selected as target variable in this example and ~~assessed are~~ mere changes in the means for the near and far future period are assessed.

In order to firstly evaluate the climate models' capabilities of reproducing the required input variables for the impact models, the meteorological indices obtained from the models are compared to the observation in the reference period. At first the impact models are applied to the climate model data in the reference period from 1971 to 2000 in order to be able to compare estimated NM7Q values to the observation. Figure 8 shows the average deviation of the simulated from the observed values over all stations individually for each climate model (left) and the mean deviations for the individual stations over all model runs (right). The left panel shows that there is a slight systematic error within most climate model chains. Even though all medians are relatively close to 0 a number of station models appear to significantly underestimate the observed NM7Q.

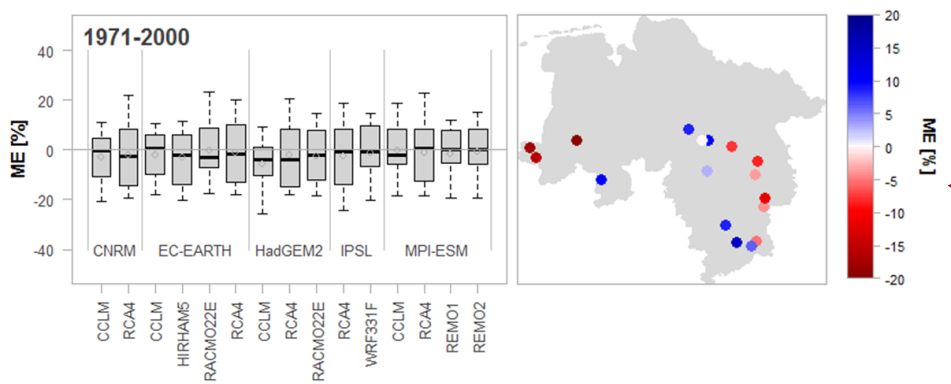



Figure 8: Mean error between simulated and observed NM7Q over 17 stations for the individual climate models (left) and over 15 climate models for the individual stations (right).

In the right panel one can see that severe underestimations appear to occur at the stations in the very west of the study area (up to -25.17 %). The eastern stations also show underestimations but slighter. Overestimations of up to 23.27 % can be found in the more central stations. This finding suggests that the quality of reproduction of the observed NM7Q is not as much determined by the selection of the climate model chain as by a station's location and its belonging impact model, i.e. the required external input variables.

The importance of input variable selection is underlined by Fig. 8 . It shows a comparison of meteorological indices derived for the climate model data in comparison to the observed values in the reference period summarizes the results. Three measures are used to assess the models' capability to reproduce the indices: a) the mean error, i.e. the difference in means between simulation and observation over all stations (computed for scaled data to make it comparable between indices) b) the number of stations where this difference is significant at a 5% -significance level, tested using a non-parametric Mann-Whitney test (Wilcoxon, 1945), and c) the NSE of the ranked simulated and observed index time series, both ordered by size, aas a means of assessing the similarity of the two distributions. The indices considered in the plot are calculated on a 6-months base period with 3 months shift and are representative for most base periods and lead times.

The left panel shows that deviations are observed for the climatic water balance and the aridity index, i.e. average P-ETP and P/ETP. The climate models consistently underestimate the observed values.

~~The stations in the eastern part of the study area require P-ETP or P/ETP input and consequently yield NM7Q values that are too low.~~

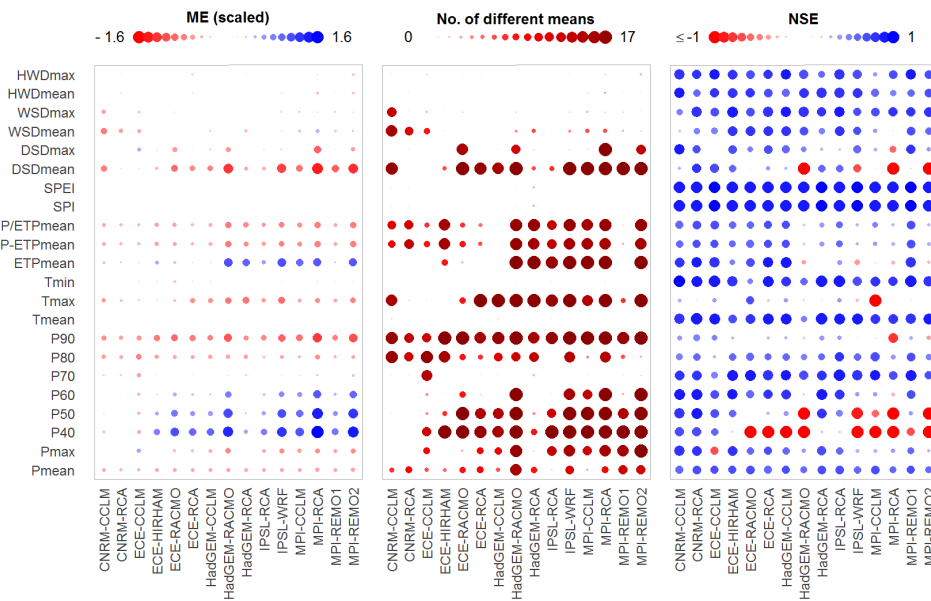



Figure 89: Comparison of simulated and observed meteorological indices for each model over 17 stations: mean error (left), number of significant shifts in location (center) and Nash-Suttcliffe efficiency of ordered series (right).

~~The stations in the west, which show the largest negative bias in NM7Q do not include water balance parameters in their models. However, they base on~~What also becomes apparent is that upper precipitation quantiles (P70 – P90). ~~As seen in (left), these quantiles~~ are mostly underestimated. The amount of significant differences found for the means, especially in P90, suggest that there is a significant bias throughout the study area. Additionally, the NSE values are rather low, indicating a generally poor reproduction of these index values. Apart from the quantiles and the climatic water balance, errors ~~can be~~are found for maximum temperature and dry spell durations. ~~These variables, however, are hardly included in any station model.~~ The lower precipitation quantiles show quite high errors but are ~~also~~ rarely considered in the station models. For those variables that are considered, additional bias correction may be advisable. ~~The station models that overestimate the NM7Q in the reference periods also base on upper precipitation quantiles (P60 – P80). At these stations, however, the climate models overestimate the true values.~~

Apparently, ~~location is a relevant factor that determines~~ the climate ~~models'~~ models have limited ability to reproduce ~~the~~ some of the true meteorological indices. This effect may be due to the inability of the regional climate models to downscale the global climate data appropriately for the study area. One needs to also consider the difference in grid size of the observed, very detailed data of 1-km resolution and the simulation on a 10-km grid. Since mere averaging of the grid cells has been applied to obtain the basin averages, the smoothing effect may be recognizable, especially for small catchments and in heterogeneous terrain. The re-calibration technique described in section 3.5 was chosen to address these issues related to various forms of scaling.

~~Apart from the quantiles and the climatic water balance, errors can be found for maximum temperature and dry spell durations. These variables, however, are hardly included in any station model. The lower precipitation quantiles show quite high errors but are also rarely considered. For those variables that are considered, bias correction may be advisable.~~

Figure 9 shows the prognoses of the individual climate models over all stations. Along with the summer NM7Q (bottom), projections for the aridity index are shown, as they indicate the overall climatic development. Also, the index is included in numerous stations' impact models. The index's development is shown for a 6-month base period without lead time, which equals the computation basis for the NM7Q (May - October; center) and a 4-months shift (January - June), which equals the maximum lead-time considered in the station models (top).

The changes for ~~"winter"~~ January-June P/ETP are predominantly positive for the majority of models and stations. Only the EC-EARTH coupled with the CCLM and the RCA4 show negative trends. The total range of projections is -13.31 to +34.10% in the far future period. The positive change in the ratios is caused by a significant increase in January-June precipitation amounts, projected for all stations by all climate models, which exceeds the simultaneously projected increase in evapotranspiration. For summer, the projections for the P/ETP are more ambiguous. 9 ensemble members clearly predict a significant decline in the far future, while 5 models project an increase at all considered stations. The spread of the projection is accordingly large, ranging from -25.75 to +48.91%. The IPSL-WRF331F thereby shows by far the highest positive development and can potentially be considered an outlier. The rather negative trend in the summer P/ETP is related to both decreasing precipitation, as projected by most models, and increasing evapotranspiration.

For the NM7Q the spread of changes over the considered stations is significantly larger and increases further from the near to the far future period. This is a logical consequence, since flow behaves much more heterogeneously throughout space and direction and magnitude of change may greatly depend on local conditions.

Overall, t~~For the NM7Q the spread of changes over the considered stations is significantly larger and increases further from the near to the far future period. This is a logical consequence, since flow behaves much more heterogeneously throughout space and direction and magnitude of change may greatly depend on local conditions. In the statistical models this very behavior is accounted for by inclusion of several explanatory variables that differ from station to station. Larger, rather slowly responding catchments with significant storage capacity will less likely be affected by small meteorological events and low flows will occur moderated and with significant temporal delay. Therefore, statistical models fitted for these catchments will most likely include meteorological indices that have been computed for longer base periods and lead times. Within small, quickly responding systems, on the other hand, low flow~~

magnitudes may be related to shorter dry periods that have occurred recently. Variable selection is thus carried out in favor of small base periods and lead times. Nevertheless, even for smaller catchments an inclusion of temporally leading meteorological indices may mimic initial storage conditions and determine the absolute magnitude of a summer low flow event.

5   According to this simplified principle, the changes in the NM7Q appear to be a mixture of the patterns observed for P/ETP in both summer and winter. The highest increase in NM7Q is predicted by the IPSL-WRF331F model chain (median 3448.18.78%), which also shows highest increase in precipitation amounts and strongest decrease in evapotranspiration in both summer and winter. EC-EARTH-CCLM and –RCA4 predict the strongest decrease (median -77.9.65 and -5.2813.42%, respectively), which is also in accordance with declining P/ETP in both summer and winter.

10   Apart from these models, there appears to be a general increase projected for summer NM7Q over the majority of stations. It seems that on average the influence of spring and winter meteorological states weighs stronger on the formation of summer NM7Q than the actual summer conditions.

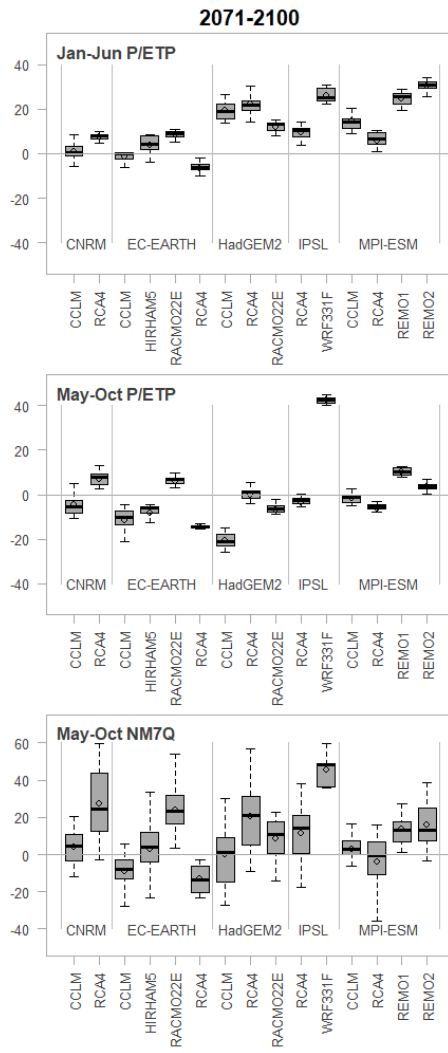
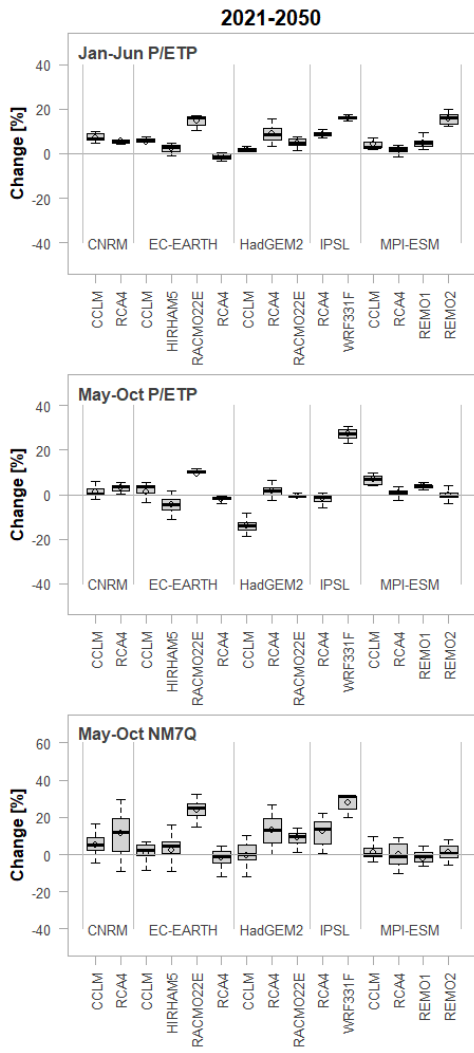
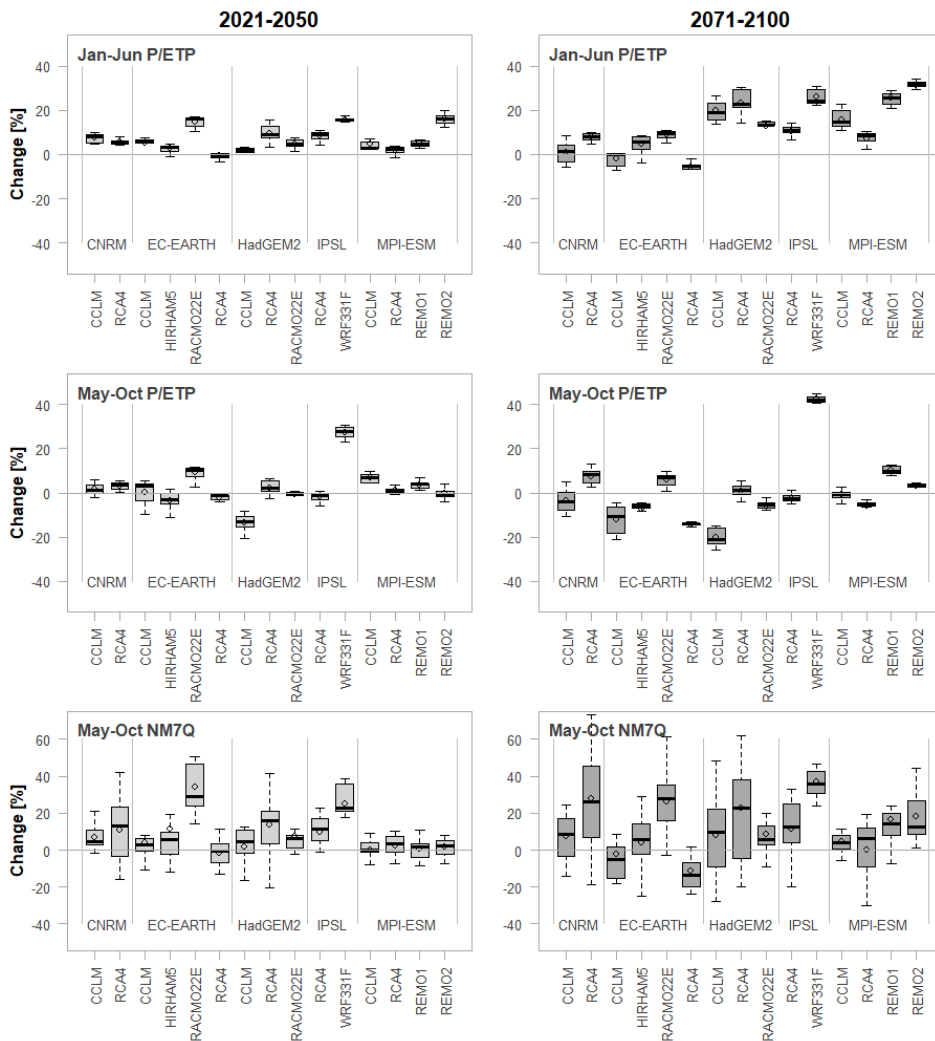



Figure 9: Projected changes in winter (top) and summer (center) aridity index and NM7Q (bottom) by all models over 17 stations and two future periods.

Figure 10 shows the spatial distribution of the projected changes for the two future periods and the variables analyzed before. Depicted are the median changes over all climate model ensemble members for the individual stations. All stations show an average increase in winter P/ETP ranging from +6.84 to +14.75% in the far future. For summer, at the majority of stations a decrease is projected, the largest changes occurring in the Harz mountains, i.e. the most southeastern stations in the study area, with an average decrease of -4.72%. Positive changes can be found in the northern stations with up to +3.12% in the near future but decreasing to +2.76% in the latter period. The NM7Q is on

average projected to increase by up to 2122.7978% at all but fourtwo stations, where a median n average decrease of up to -5.0212.01% is projected. Two of these stations are situated in or near the Harz Mountains, where flow forming processes and climate are quite different from the rest of the study area. Additionally, theirr areas are the smallest in the data set considered for impact analysis (44.5 and 68.7 to 363.0 km²). Apparently, the NM7Q in these areas is influenced more by summer meteorology than the rest of the study area.

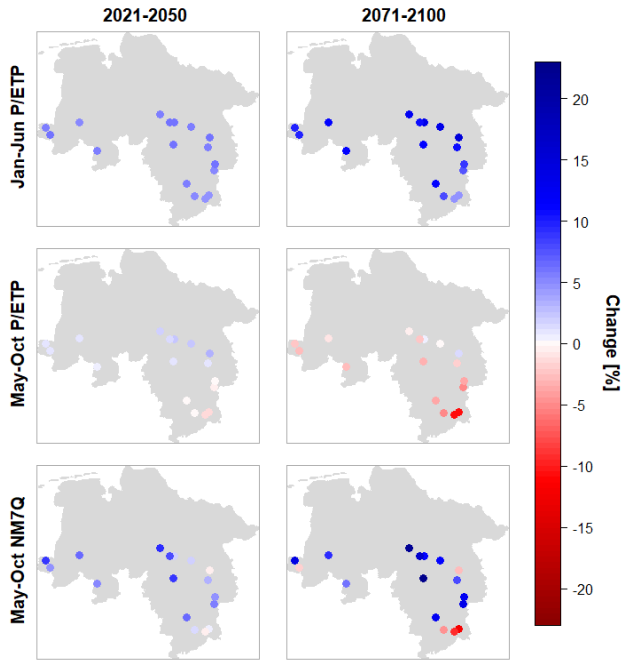
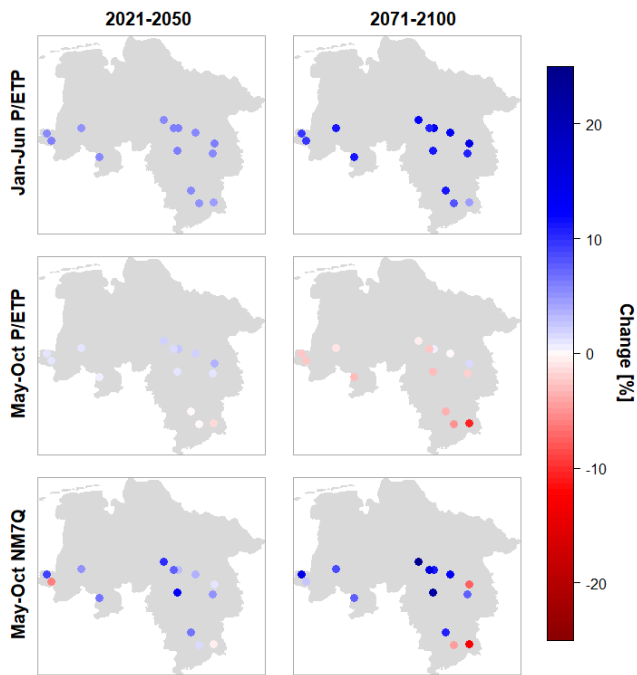



Figure 10: Medians of projected changes in winter (top) and summer (center) aridity index and NM7Q (bottom) by for individual stations over 15 climate models and two future periods.

Finally, ~~Fig. 11~~ shows boxplots for all stations and the projected changes in the NM7Q over all considered ensemble members. It can be seen that at most of the stations the predicted directions of change are quite unambiguous over the majority of climate models and projections are accordingly quite robust. Exceptions are stations 4819102 ~~and~~, 4882195 and 9286162, where the median change is close to 0. ~~What becomes also apparent in these plots is the spatial dependence between stations. Similar first digits of the station numbers indicate that stations belong to the same river basin, while increasing numbers indicate more downstream locations, i.e. larger size. The projected changes are obviously strongly related to catchment area, since continuous transitions are observable from most negative/smallest change in NM7Q at the smallest to most positive change at the largest catchments within each basin. It remains to be analyzed, e.g. via hydrological modeling, whether this is an artifact created by oversimplification of the statistical approaches (through neglect of other relevant, potentially counteracting factors) or a truly observable phenomenon.~~

What becomes also apparent in these plots is the spatial dependence between stations. Similar first digits of the station numbers indicate that stations belong to the same river basin, while increasing numbers indicate more downstream locations, i.e. larger size, as depicted in Fig. 12. The projected changes are obviously strongly related to catchment area, since continuous transitions are observable from most negative/smallest change in NM7Q at the smallest to most positive change at the largest catchments within each basin. Summer low flow in catchments with large storage capacity will more likely be influenced by the increasing annual and winter precipitation amounts, while the flow in

the smaller catchments is more drastically influenced by drier summer meteorological conditions. It remains to be analyzed, e.g. via hydrological modeling, whether this observation is an artifact created by oversimplification of the statistical approaches (through neglect of other relevant, potentially counteracting factors) or a truly observable phenomenon. Nonetheless, the coherence of the projected station means suggests that the statistical models – even though they have been fitted independently for every station - exhibit a meaningful spatial consistence within the region. Apparent outliers are not produced by any stations' statistical impact model.

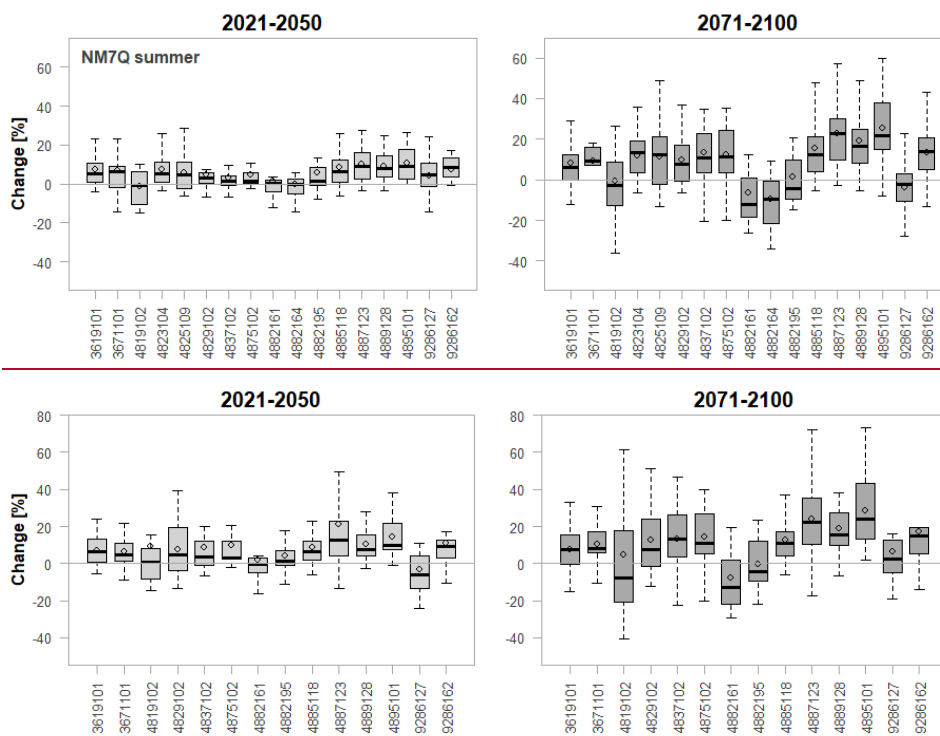



Figure 11: Projected changes in summer NM7Q for individual stations over 15 climate models and two future periods.

~~A comparison of (bottom) and shows that – other than for the error in the reference period – the projected changes appear to differ much more between the individual regional climate models than between the individual stations. The similarity of the projected station means suggests that the statistical models – even though they have been fitted independently for every station – exhibit a meaningful spatial consistence within the region. Apparent outliers are not produced by any stations' statistical impact model.~~

The significant increase in winter and partly projected decrease in summer P/ETP may suggest a temporal shift in minimum flows from summer/early fall to late fall/early winter in the study area. In order to test this hypothesis,

statical models are fitted to additionally model the winter ~~and annual~~ NM7Q. Modeling of ~~annual values appeared equally successful as the fitting to the summer values, modeling of~~ the winter values was more difficult than of the summer low flows since their causes may not be representable by the utilized indices (e.g. water retention by snow).~~.~~ The final projections ~~are depicted~~ in Fig. 13~~Fig. 13. The top panels suggest~~ suggest that the expected developments for winter are less conclusive than for summer. ~~.~~ At most stations mostly ~~negative~~ positive trends are projected for ~~the far~~both future periods with lacking robustness. A pronounced pattern of change contrasting the summer NM7Q cannot be observed. Accordingly, with overall increasing precipitation amounts the intensities of the severest events of the year are expected to decrease at most stations in the area, regardless of their time of occurrence. ~~For the entire year the trends are much more obvious with a pattern that seems to resemble the changes in the summer NM7Q. Still, the~~T hese findings indicate that there might be a slight delay in the onset of future low flow events. The ratio of winter to summer low flows appears to increase slightly at most stations. While for the observation 10.20% of the annual lowest low flow events occur in winter, the models predict on average 15.1% for the reference period, 20.00% for the near and 22.67% for the far future. ~~but that with overall increasing precipitation amounts the intensities of the severest events are expected to decrease.~~

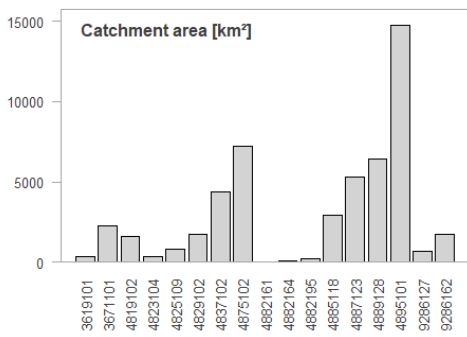



Figure 12: Catchment areas of the considered gauges.

A comparison of the signals projected in this work to other studies that analyze comparable regions and climate model ensembles but use hydrological models for impact assessment appears difficult, due to differences in scale, model ensemble and type of analysis (e.g. based on degree of heating rather than on fixed periods). A regional study that is comparable to this work has been carried out by Osuch et al. (2017) for Poland. They investigated climate change impacts on low flows, including the annual NM7Q, using the HBV model and, among others, an RCP 8.5 climate model ensemble. They found significant positive changes in the NM7Q for the RCP8.5 scenario with more than an average 40% increase in the study area in a period from 2071-2100. These changes are much larger than the ones found in this study but are obtained for a more continental climate. Other studies analyzing climate change effects on low flows over entire Europe (e.g. Marx et al., 2018) or global studies (e.g. Wanders and Yada, 2014) using RCP emission scenarios usually find inconclusive changes in low flow magnitude for the northwestern parts of Germany (between -

10 and +10%). The average changes in the annual NM7Q over all stations and ensemble members obtained in this study are +5.62~~4.18~~% for the near and +~~5.29~~8.96% for the far future period and thus within the range of these findings. Van Vliet et al. (2015) show in their Europe-wide study that especially in central western and northern Europe the projected changes in streamflow drought under an RCP8.5 scenario appear to differ between the two hydrological models they applied, both in direction and magnitude. In general, it appears difficult to assess future low flow development in the region, especially in comparison to the parts of Europe, where expected changes in precipitation are more pronounced and unequivocal between climate models.

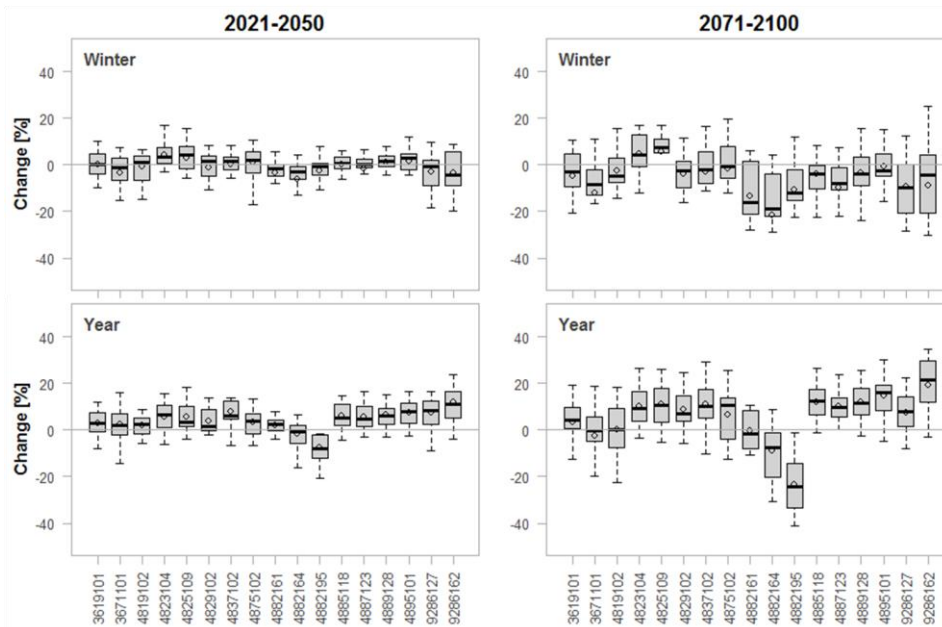
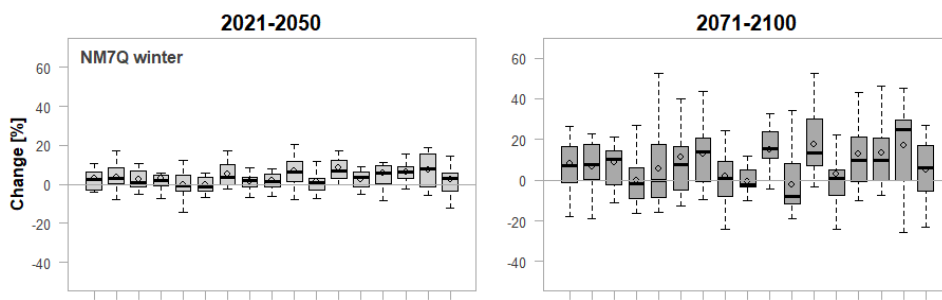



Figure 13: Projected changes in winter ~~(top) and annual (bottom)~~ NM7Q for individual stations over 15 climate models and two future periods.

5 Conclusions and outlook

The aim of this work was to assess key low flow drivers using , whether simplified statistical approaches would be a sufficient means for modeling climate change impacts on low flows. approaches. Several regression models with varying restrictions and assumptions have been set up and evaluated in a split-validation procedure and applied to an ensemble of climate model data. In order to assess potential future changes in the low flow, the models are applied to an ensemble of climate model data. The main findings are:

- Modeling of annual low flow values as a function of annual meteorological indices appears feasible. An approach using a few principal components of several variables as regressors yields the best results for the 60-year observation period. It showed to simultaneously countervail problems with multicollinearity and omittance of important explanatory variables.

- Non-stationarity within the relationship between meteorological predictors and low flow target variables appear to be an issue, resulting from the entire neglect of physical processes and changing interrelationships between explanatory metrics. This problematic needs to be considered, when applying the models to future data, especially in terms of parameter stability. Ideally the models should be revised through inclusion of methods to map potential non-stationary processes and interrelationships.

- More average low flow indices are reproduced better by the statistical models than their more extreme counterparts, e.g. the NM30Q compared to the NM7Q. Maximum duration and deficit volume can be adequately modeled, while the onset of an event can by no means be predicted by the suggested models.

- The main drivers for all low flow characteristics are metrics related to precipitation and evapotranspiration amounts. Temperature indices appear to be relevant at some stations for certain low flow metrics. Usually base periods for computation and lead times of the explanatory variables considered in the station models depend greatly on catchment size and response time of the catchments.

- The hydrological model outperforms the statistical approaches, when explicitly calibrated on the low flow events. However, the major discrepancy between its performances in the calibration and the validation period shows that either non-stationary processes may not be captured or that there is significant non-homogeneity between the calibration and validation period considered for analysis.

- Applying the non-restricted principal component model with an ensemble of coupled global and regional models based on an RCP 8.5 scenario yields on average an expected increase in summer NM7Q at most stations. Only two small headwater catchments show a decrease in NM7Q in the far future. These stations appear to not be influenced by preceding winter precipitation.

- A major weakness for the proposed model type is the reproduction of meteorological indices by the climate models. However, underestimation of the climatic water balance and upper precipitation quantiles may be a likewise limiting factor within process-based impact models. Bias correctionA direct calibration using climate model data instead of the observation can potentially help increase ing the quality of the projections made by the statistical models.

There are potential advantages to applying statistical approaches that have not been discussed before. One is that the required input consists of variables that are lumped over a significant amount of time (3 – 12 months). Such averages

can potentially be better reproduced by climate models than the daily variability required for other impact model types. Also, the statistical models may be easily extended to include other forcings like large-scale phenomena, which may be similarly described by indices like local meteorological conditions. Also land-use feedbacks might be includable to some extent. What remains difficult is to statistically model feedbacks between different external variables with the simplified approaches. This problem may, however, be moderated by including non-stationary regression coefficients for individual variables. Considering issues with the coupling of global and regional climate and impact models, as identified e.g. by Dai et al. (2013) or Wilby and Harris (2006), including issues with downscaling, a direct use of large-scale teleconnections and the possibility to apply the models directly to extrapolated climate data - rather than just climate model data – may suggest that statistical approaches can significantly increase the robustness of predictions, especially when considered in an ensemble of various model types and information sources. Such a framework has been suggested by Laaha et al. (2016) for proper assessment of climate change impacts on low flows.

The statistical approach can also make use and benefit from inclusion of spatial information, especially with respect to regional climate change impact analysis. According to the concept of "trading space for time", simultaneous analysis of all information in an area instead of a single-site approach will yield more robust parameter estimates and allow for estimation of climate change signals at unobserved sites. Extensions of the existing station models using different approaches of non-stationary regional frequency analysis have been proposed in Fangmann (2017). The simultaneous consideration of space and time within the simple statistical approaches appeared beneficial at the regional level.

All in all, the approach of modeling specific low flow indices as a function of meteorological indices appears promising. Model set up and computation are straightforward and quick, even over large study areas with high numbers of catchments. The method appears to be able to simulate flow for any catchment from small to large and flat to steep without any consideration of physiographical characteristics. Being able to consider larger periods for calibration would assumingly lead to an increase in robustness of the models and extend the potential horizon for prognosis. Even though the projected values should not be taken as basis for dimensioning, the statistical approaches can be of assistance for decision making. They can offer a good approximation of the future development, especially when considered in a framework of several climate-impact model ensembles and are worthy of consideration in the field of climate change impact analysis.

Parts of this study have been considered in greater detail in Fangmann (2017).

Acknowledgements

The results presented in this study are part of the KliBiW project (Wasserwirtschaftliche Folgenabschätzung des globalen Klimawandels für die Binnengewässer in Niedersachsen), funded by the Lower Saxon Ministry for Environment, Energy and Climate Protection (Niedersächsisches Ministerium für Umwelt, Energie und Klimaschutz), who are gratefully acknowledged. We would also like to thank the NLWKN (Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz) for provision and pre-processing of the data used in this study.

References

Akaike H.: New Look at Statistical-Model Identification. Ieee T Automat Contr, 19, 716-723, doi:10.1109/Tac.1974.1100705, 1974.

Coles S.: An introduction to statistical modeling of extreme values. Springer, London, 2001.

Dai A.G.: Increasing drought under global warming in observations and models. Nat Clim Change 3, 52-58, doi:10.1038/Nclimate1633, 2013.

de Wit M.J.M., van den Hurk B., Warmerdam P.M.M., Torfs P.J.J.F., Roulin E. and van Deursen W.P.A.: Impact of climate change on low-flows in the river Meuse. Climatic Change, 82, 351-372, doi:10.1007/s10584-006-9195-2, 2007.

Fangmann A., Belli A. and Haberlandt U.: Trends in beobachteten Abflusszeitreihen in Niedersachsen. Hydrol. Wasserbewirts., 57, 196-205, doi:10.5675/HyWa_2013,5_1, 2013, 2013.

Fangmann, A.: Low flow prediction in time and space: An adaptive statistical scheme for regional climate change impact assessment, Mitteilungen, Heft 106, Inst. of Hydrology, Leibniz University of Hannover, Hannover, 160 pp., 2017.

Forzieri G., Feyen L., Rojas R., Florke M., Wimmer F. and Bianchi A.: Ensemble projections of future streamflow droughts in Europe. Hydrol. Earth Syst. Sci., 18, 85-108, doi:10.5194/hess-18-85-2014, 2014

Gosling S.N., Zaherpour J., Mount N.J., Hattermann F.F., Dankers R., Arheimer B., Breuer L., Ding J., Haddeland I., Kumar R., Kundu D., Liu J.G., van Griensven A., Veldkamp T.I.E., Vetter T., Wang X.Y. and Zhang X.X.: A comparison of changes in river runoff from multiple global and catchment-scale hydrological models under global warming scenarios of 1 degrees C, 2 degrees C and 3 degrees C. Climatic Change, 141, 577-595, doi:10.1007/s10584-016-1773-3, 2017.

Haslinger K., Koffler D., ~~Schoner~~ Schöner W. and Laaha G.: Exploring the link between meteorological drought and streamflow, Effects of climate- catchment interaction. Water Resour. Res., 50, 2468-2487, doi:10.1002/2013WR015051, 2014.

Laaha G., Parajka J., Viglione A., Koffler D., Haslinger K., ~~Schoner~~ Schöner W., Zehetgruber J. and ~~Bloschl~~ Blöschl G.: A three-pillar approach to assessing climate impacts on low flows. Hydrol. Earth Syst. Sci., 20, 3967-3985, doi:10.5194/hess-20-3967-2016, 2016.

Lindstrom G., Johansson B., Persson M., Gardelin M. and Bergstrom S.: Development and test of the distributed HBV-96 hydrological model. J. Hydrol., 201, 272-288, doi:10.1016/S0022-1694(97)00041-3, 1997.

Liu D.D., Guo S.L., Lian Y.Q., Xiong L.H. and Chen X.H.: (2015) Climate-informed low-flow frequency analysis using nonstationary modelling. Hydrol. Process., 29, 2112-2124, doi:10.1002/hyp.10360.

Marx A., Kumar R., Thober S., Rakovec O., Wanders N., Zink M., Wood E.F., Pan M., Sheffield J. and Samaniego L.: Climate change alters low flows in Europe under global warming of 1.5, 2, and 3 degrees C. Hydrol. Earth Syst. Sci. 22, 1017-1032, doi:10.5194/hess-22-1017-2018, 2018.

Mosley M.P.: Regional differences in the effects of El Nino and La Nina on low flows and floods. Hydrolog. Sci. J., 45, 249-267, doi:10.1080/02626660009492323, 2000.

NLWKN: Globaler Klimawandel: Wasserwirtschaftliche Folgenabschätzung für das Binnenland, Oberirdische Gewässer Band 33, Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz, Betriebsstelle Hannover-Hildesheim, 2012.

NLWKN: Globaler Klimawandel: Wasserwirtschaftliche Folgenabschätzung für das Binnenland, Oberirdische Gewässer Band 41, Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz Betriebsstelle Hannover-Hildesheim, 2017.

Osuch M., Romanowicz R. and Wong W.K: Analysis of low flow indices under varying climatic conditions in Poland (in press). Hydrol. Res., doi:10.2166/nh.2017.021, 2017.

Pettitt A.N.: A non-parametric approach to the change-point problem. Appl. Stat.-J. Roy. St. C., 28, 126-135, doi:10.2307/2346729, 1979.

Roudier P., Andersson J.C.M., Donnelly C., Feyen L., Greuell W. and Ludwig F.: Projections of future floods and hydrological droughts in Europe under a +2 degrees C global warming. Climatic Change, 135, 341-355, doi:10.1007/s10584-015-1570-4, 2016.

Parajka J., Blaschke A.P., Blöschl G., Haslinger K., Hepp G., Laaha G., Schöner W., Trautvetter H., Viglione A. and Zessner M.: Uncertainty contributions to low-flow projections in Austria, Hydrology and Earth System Sciences, 20(5), 2085–2101, doi:10.5194/hess-20-2085-2016, 2016.

Schneider C., Laize C.L.R., Acreman M.C. and Florke M.: How will climate change modify river flow regimes in Europe? Hydrol. Earth Syst. Sci., 17, 325-339, doi:10.5194/hess-17-325-2013, 2013.

Schwarz G.: Estimating Dimension of a Model. Ann. Stat., 6, 461-464, doi:10.1214/aos/1176344136, 1978.

Van Loon A.F. and Laaha G.: Hydrological drought severity explained by climate and catchment characteristics. J. Hydrol. 526, 3-14, doi:10.1016/j.jhydrol.2014.10.059, 2015.

Vrugt J.A. and Robinson B.A.: Improved evolutionary optimization from genetically adaptive multimethod search. P. Matl. Acad. Sci. USA, 104, 708-711, doi:10.1073/pnas.0610471104, 2017.

Wallner M., Haberlandt U. and Dietrich J.: A one-step similarity approach for the regionalization of hydrological model parameters based on Self-Organizing Maps. J. Hydrol., 494, 59-71, doi:10.1016/j.jhydrol.2013.04.022, 2013.

Wanders N. and Van Lanen H.A.J.: Future discharge drought across climate regions around the world modelled with a synthetic hydrological modelling approach forced by three general circulation models. Nat. Hazard. Earth Sys., 15, 487-504, doi:10.5194/nhess-15-487-2015, 2015.

Wanders N. and Wada Y.: Human and climate impacts on the 21st century hydrological drought. J. Hydrol., 526, 208-220, doi:10.1016/j..hydrol.2014.10.047, 2015.

Wilby R.L. and Harris I.: A framework for assessing uncertainties in climate change impacts, Low-flow scenarios for the River Thames, UK. Water Resour. Res., 42, doi:10.1029/2005wr004065, 2006.

Wilcoxon F.: Individual Comparisons by Ranking Methods. Biometrics Bull., 1, 80-83, doi:10.2307/3001968, 1945.