Hydrology and Earth System Sciences Discussions
Hydrol. Earth Syst. Sci. Discuss., author comment AC2
https://doi.org/10.5194/hess-2022-122-AC2, 2022

Reply on RC2

Raphael Schneider et al.

Author comment on "Machine-learning-based downscaling of modelled climate change impacts on groundwater table depth" by Raphael Schneider et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2022-122-AC2, 2022

Reply to Review by Anonymous Referee #2

[Reviewer comments in normal font; *Author replies in itialic*]

[Additional figures in attached pdf]

Schneider et al proposed a RF-based downscaling method to downscale changes in the simulated water table depth over Denmark from 500 m resolution to 100 m resolution under different future climate scenarios. The method was trained on data from five submodels that cover a wide range of geologic, topographic, and hydrologic variability occurring across Denmark, and validated on data from another submodel (VI). The results obtained by the proposed method outperformed 500m-resolution water table depth and its bilinear interpolation in showing the climate change-induced changes to the shallow groundwater table. The paper would be of interest to the hydrological community. Overall, it is well-written and the related questions are discussed thoroughly. However, I have the following concerns regarding the paper.

Reply: *We thank the reviewer for their positive and constructive feedback to improve the manuscript. Below, we outline how we consider responding to the concerns raised by the reviewer in a revision of the manuscript.*

General comments:
1. Traditional downscaling techniques downscale a product at a coarse resolution to the same product at a finer resolution. Here the authors used different statistics calculated from the coarse-resolution product (TBDV). Why didn't the authors directly use the 500m water table depth as a covariate here?

Reply: *Actually, we use the same statistics (mean, Q01, Q99, 1mex of changes to groundwater table) in 500m resolution (resampled/interpolated to 100m) as a covariate in downscaling to the respective output in 100m (i.e. again mean, Q01, Q99, or 1mex, respectively). Hence, if we understand the reviewer's comment correctly, our method is in that respect in line with "traditional downscaling techniques".*
On a side note: We also expect the proposed method to work with time-varying groundwater depth maps (as mentioned in the 2nd paragraph of the Conclusions).

2. Which criteria did the authors use to select their validation submodel (VI. Aarhus

Å/Aarhus)? From Fig.1, the submodel has a very shallow mean water table depth (0.5-2.5m). There are many areas in Demark having water table depth > 10 m, like the submodel V. I wonder if the selection of VI would give a biased conclusion for the RF validation.

Reply: *The submodels were picked from a set of 10 submodels in total, which were part of the model calibration (see section 2.1.3 and Henriksen et al., 2020a). These ten submodels already were chosen to be representative of hydrologic variations across Denmark. The submodels used in the downscaling training then were further selected from these ten based on their representativeness of relevant covariates (see Figure 2).*
Maybe this cannot be seen very well in Figure 1, but as can be seen in the first plot in the attached pdf file (Figure A1), submodel VI actually represents the Danish conditions quite well. The plot shows a histogram of the mean historic depth to the groundwater table, separately for Denmark, the five training submodels I to V and the validation submodel VI; with linear scale y-axis in the top plot, and log-scale in the bottom plot. Hence, while the reviewer correctly noted that there are significant areas with deeper groundwater tables in Denmark, we remain confident that those conditions are well covered by our training models (in particular submodel V) and also our validation submodel VI. (Furthermore, when it comes to vulnerability to climate change impacts, areas with currently shallow groundwater levels (within the first few metres of the surface) are of greatest interest.)
Plan for revision: Extend the section 2.3.3 concerning the submodel choice, and potentially include the validation submodel in the histograms in Figure 2 as shown in the attached Figure A1.

3. I really like the idea to study the importance of each covariate (feature) used in RF. I also think that determining the feature importance based on ML model performance is a feasible method. However, the authors may need to check the independence of their covariates before implementing such an approach. If two or more covariates are strongly correlated, perturbing one of them may not impact the ML performance, which leads to wrong results. I would like to know how the authors dealt with this issue.

Reply: *Thanks. We agree with the reviewer, covariate correlation can affect feature importances. The covariates we used were already selected with covariate correlation in mind. Hence, covariate correlation is low for most of the covariates – see Figure A2 for a matrix of pairwise covariate correlations (Pearson's R) in the attached pdf.*
Plan for revision: Mention the issue of covariate correlation in the manuscript (likely in section 3.1). Potentially extend the feature importance analysis by a version where not only one covariate at a time is perturbed, but a whole group of (correlated) covariates – similar to Figure 4 in Koch et al., 2019a. In our case that for example would mean to perturb all the (moderately correlated) "kh_mean" covariates at a time for the feature importance analysis.

4. Please improve the quality of the figures.

Reply: *By that you mean the resolution and compression artefacts? In that case, we assume that the final article will be compiled in a different manner; the current quality issues are due to the pdf compiling.*

Specific comments:
1. Line 102, Page 4: "referred to the provided literature". Which literature? (Abbott et al., 1986; DHI, 2020)? Please specify there.

Reply: *Here we mean the various references provided throughout sections 2.1 covering different aspects of the DK-model (subsurface parameterization, climate input, MIKE SHE …).*

Plan for revision: Clarify this.

2. Line 112-114, Page 4: I am not an expert in hydrological model simulation, and I am a bit confused here. The authors mentioned that precipitation, temperature, and potential ET used for historic climate forcing to the DK-model HIP have various resolutions, 10 km or 20 km. However, in Line 105, they mentioned that all input data have a spatial resolution of 100 m. Therefore, did they downscale historical climate forcing data to 100 km or use them directly?

Reply: *Valid point. The climate forcing (at 10km/20km resolution) is interpolated to the model grid (500m or 100m).*
Plan for revision: Clarify this.

3. Climate models, Page 5: Can the authors clarify which 17 RCMs they chose and which 5 RCMs are used as a subset?

Reply: *Plan for revision: Add a table with the requested information.*

4. Line 173, Page 6: Why did the authors use changes to the 1m exceedance probability? Can the authors explain the practical meaning of this statistic?

Reply: *Good question, relevant to be clarified. The threshold of 1m was chosen in connection with stakeholders and users of the data. Water levels closer than a certain threshold to the surface can create various challenges in agriculture, infrastructure and flooding. In this context, a threshold of 1m was considered relevant (also, the widespread tile drains in Danish agriculture are located at around 1m depth). The exceedance probability then indicates how often (during an average year) the respective threshold of 1m is exceeded, and how that probability changes with climate change.*

5. Line 235, Page 8: "RF is a supervised ML learning method; that means it requires training data". This statement is wrong in my opinion. Unsupervised ML methods also require training data. I think here the authors meant supplementary teacher signals that are used to guide the training process. In addition, ML is the abbreviation for machine learning. Please delete the extra "learning" here.

Reply: *The reviewer is correct; we were not precise enough with the choice of our words here. We suggest reformulating to "RF is a supervised ML method, requiring labelled training data. Based on the training dataset, a RF regressor model learns about relationships between a set of covariates and the target (training) data values."*
Thanks for also noting the typo.

6. Please mark the locations of dummy points used in RF training in Fig.1 if possible.

Reply: *Due to the large number of dummy points (20,000), we think it is difficult to show them on the map. They are sampled randomly in space, from all of Denmark except for the areas covered by the training submodels.*

7. Line 276, Page 9: I believe there should be Table 3.

Reply: *Thanks for spotting that mistake; will be corrected.*

8. Line 335, Page 11: which statistic does "the climate change-induced changes to the shallow groundwater table" indicate?

Reply: *Figure 7 gives an overview over all the eight TBDV (i.e. mean, Q01, Q99, and 1mex for both near and far future). Plan for revision: Clarify this.*

9. Fig.7: Please explain the legends (e.g., 500m HM intp) in the caption.

Reply: *Valid point, will be added in the revision.*

Please also note the supplement to this comment:
https://hess.copernicus.org/preprints/hess-2022-122/hess-2022-122-AC2-supplement.pdf