

Hydrol. Earth Syst. Sci. Discuss., author comment AC1  
<https://doi.org/10.5194/hess-2022-122-AC1>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Reply on RC1

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-AC1>, 2022

---

## Reply to Review by Anonymous Referee #1

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

The manuscript "Machine learning-based downscaling of modelled climate change impacts on groundwater table depth" by Schneider et al. presents a novel downscaling method which uses hydrological model simulation data at a coarse scale (500 meters) together with ancillary data (e.g. topography and hydrogeologic information) to derive indicators for groundwater changes for future climate scenarios at higher spatial resolution (100 meters). Model simulations at a scale of 100 meters for five selected catchments and five input data sets from different regional climate model simulations are used as training data for the downscaling algorithm which is based on the Random Forest method. Estimates of groundwater changes at high resolution are made by using hydrologic simulations at coarse scale (500 meters) with input from 18 regional climate model simulations. The downscaling method is verified with data from a high resolution (100 meters) simulation for one additional catchment.

The topic of the paper is relevant to the hydrologic community as it describes an interesting possibility to provide stakeholders with high resolution information on potential changes in groundwater resources with an affordable computational cost. Generally, the paper is well written but there are a few issues that need to be clarified in my opinion.

**Reply:** *We thank the reviewer for their positive and constructive feedback to improve the manuscript. Below, we outline how we consider to respond to the issues pointed out by the reviewer in the revision.*

General comments:

- The proposed downscaling method can be seen as a data-driven surrogate model for generating high resolution data out of the simulation results of the 500 meter model. This allows to avoid the computationally expensive direct simulation at this higher resolution, but adds some additional uncertainties and errors. In order to judge the quality and usefulness of these high resolution data, the user would still require some information on how the predictions improve when going from 500 to 100 meter resolution. Currently, the manuscript only provides information on how well the downscaling algorithm works but it does not describe the practical benefits and improvements of the higher resolution. Hence I would suggest to add a paragraph (e.g. around line 123) that summarizes the main advantages of the high resolution model as inferred from previous comparisons of the low

and high resolution model with observation data.

**Reply:** *The reviewer raises a valid point. When originally developing/calibrating the two versions (100m and 500m) of the model, the 100m resolution performed slightly better in terms of groundwater head performance (especially for shallow wells). However, we expect the 100m model to generally be better able to reproduce fine-scale variations of the uppermost groundwater level, as these are controlled majorly by topography. And many of the relevant topographic variations will be smoothed out at 500m resolution, but remain visible in 100m resolution. These variations are hard to show with conventional groundwater observations, for example because some of the relevant areas such as river valleys are under-represented. However, we managed to show some of this benefit of the 100m by comparing satellite land surface temperature products (as a proxy for the shallow groundwater table) with modelled results across river valleys.*

*Plan for revision: Extend the section mentioned by the reviewer by more clearly pointing out some of the benefits of the 100m vs. 500m model wrt. the representation of the uppermost groundwater level*

- In section 2.4.3 (lines 248-258) it is mentioned that additional points outside the five 'calibration' catchments were used in the calibration procedure of the algorithm to improve the robustness of the method. Can you explain in more detail what kind of robustness issues you detected? Do you have any explanation why these additional points were necessary although the five chosen 'calibration' catchments closely resembled the statistical properties for whole Denmark (Figure 2)? Which additional information did these 'dummy points' provide?

**Reply:** *With "robustness", we mean spatial transferability/performance on the spatial hold-out. While it is true that the covariate space seems to be adequately covered by the training catchments, a random sampling of all of Denmark still seems to be adding some covariate values/covariate combinations that inform the Random Forest regressor. (furthermore, performance of a Random Forest algorithm or similar is not only determined by covering the covariate space, but also by covering the relevant combinations of the different covariates – a thought that was behind the development of the dissimilarity index by Meyer and Pebesma, 2021)*

*Plan for revision: Clarify the idea behind the dummy points with the above mentioned*

- Additionally, the selection of additional calibration data through the 'dummy points' is not really in line with the argumentation in the rest of the paper which only refers to a calibration procedure with data from the five subcatchments. I would suggest to clearly state in all relevant parts how the calibration dataset was chosen (i.e. also mentioning the 'dummy points').

**Reply:** *That is correct, we will clarify this, e.g. in the end of the introduction and in Figure 3. However, in case this was misunderstood, we want to point out that the dummy points originate from the coarse-scale resolution run of the hydrological model, so they did not require any additional runs of fine-scale hydrological models.*

- Is it possible to provide guidelines on the size of the training data set? This would be an important information when applying the proposed downscaling method to other regions.

**Reply:** *That is a relevant question, but also a difficult one. For a start, the necessary size of the training data depends a lot on the desired application. Are we only interested in (i) predictions within very limited areas/within the training catchments, or are we – as in the manuscript – interested in (ii) an algorithm that can be extrapolated beyond its training data?*

*In case of (i), much smaller datasets than the one used here might be sufficient. In case of (ii), any possible answer probably is less related to a size of a training dataset, but*

rather to how well the training dataset covers the covariates (and covariate combinations) of the area to be extrapolated to (as also mentioned in the comment above when discussing "robustness").

*Plan for revision: Add some of these thoughts to the Discussion, section 3.3*

- Some plots are difficult to understand and need to be revised (see specific comments below).

Specific comments:

- Line 150: "...aggregated as described below." Please add the section number you are referring to.

**Reply:** *This refers to section 2.3.1; will be added in revision*

- Line 154: It is not clear how the initial conditions were determined. Did you choose any random simulation time step between 1991 and 2100 as initial conditions or did you e.g. use the mean of this simulation period?

**Reply:** *The initial conditions were taken from a continuous run of the 500m national hydrological model with each of the respective climate models, where we used data from the simulation time step corresponding to the start time of each of the reference, near, and far future periods.*

*Plan for revision: Clarify.*

- Equation 1: Please make clear also through the notation that these statistics are calculated individually for each grid cell of the model.

**Reply:** *Valid point; will be made clear in revised version.*

- Line 218: Please provide details on the "...differences between a historic dry and wet period,..."

**Reply:** *For this, we took the difference between a relatively dry historic period (the 12 consecutive years between 1990 and 2001; average yearly precipitation 817mm) and a relatively wet historic period (2004 to 2015; average yearly precipitation 852mm)*

*Plan for revision: Clarify.*

- Line 222: Why is the 500 meter model output interpolated to 100 meters although this does not provide further information to the downscaling method? Is it a hard requirement of the algorithm to operate on equally sized vectors? Is there any explanation why the algorithm works better with interpolated TBDV data?

**Reply:** *Yes, the algorithm expects equally sized vectors, i.e. some kind of resampling from the coarse to the fine resolution has to be performed. Whether an interpolation (a simple bilinear interpolation in this case; not adding any data requirements or computational bottlenecks) is necessary or not remains unclear. However, in initial tests with non-interpolated data we experienced some artefacts from the edges of the 500m data in the 100m downscaled results.*

*Plan for revision: Discuss this.*

- Line 392: Unit missing.

**Reply:** *Thanks for noting; will be added (100 m)*

- Figure 4: The scale break in the figure is a bit counterintuitive and misleading. I would suggest to show the different factors on a plot with the same scale (0 to 1) and add an

additional plot (either separate or as an inset) with the second scale.

**Reply:** *For the revision, we suggest the following: A plot with a scale of 0 to 1 for all covariates, together with an inset with the less sensitive parameters and a scale of 0 to 0.1.*

- Figure 5: Legends for the plots in the uppermost row seem to be missing. Generally, it is not readily clear with legend applies to which subplot.

**Reply:** *Correct, legends for the uppermost row are missing. This is on purpose, as the absolute values have no importance in this context; the two maps of relative topography and transmissivity in layer 1 are mostly shown to get an idea of how patterns in covariates influence patterns in the climate change impact.*

*Plan for revision: We would suggest to keep this as it is; however with adding an explanation and making the relation between the existing legends and submaps more clear.*

- Figure 6: It is difficult to grasp what part of the verification data is shown in the different subplots (i.e. model input or output of the downscaling method). I would suggest to improve the figure headings and the caption text to guide the reader better through the figure.

**Reply:** *Good point, we will try to improve the figure; potentially by adding a "header" for each row*

- Figure 7: Please clarify the abbreviations in the figure, e.g. nf and ff. This might be guessed from the manuscript text but should also be made clear somewhere in the figure or the figure caption.

**Reply:** *The abbreviations are explained in section 2.3; however, this is long before Figure 7. We will repeat this in the caption text.*