Machine learning-based downscaling of modelled climate change impacts on groundwater table depth

Raphael Schneider[1], Julian Koch[1], Lars Troldborg[1], Hans Jørgen Henriksen[1], Simon Stisen[1]

[1]Department of Hydrology, Geological Survey of Denmark and Greenland (GEUS), 1350 Copenhagen K, Denmark

5 *Correspondence to*: Raphael Schneider (rs@geus.dk)

Abstract. There is an urgent demand for assessments of climate change impact on the hydrological cycle at high spatial resolution. In particular, the impacts on shallow groundwater levels, which can lead to both flooding and drought, have major implications for agriculture, adaptation and urban planning. Predicting such hydrological impacts is typically performed using physically-based hydrological models. However, such models are computationally expensive, especially at

10 high spatial resolutions.

This study is based on the Danish national groundwater model, setup as a distributed, integrated surface-subsurface model at 500 m horizontal resolution. Recently, a version at a higher resolution of 100 m was created; amongst others, to better represent the uppermost groundwater table and to meet end user demands for water management and climate adaptation. The increase in resolution of the hydrological model, however, also increases computational bottleneck. To evaluate climate

15 change impact, a large ensemble of climate models was run with the 500 m hydrological model, while performing the same ensemble run with the 100 m resolution nation-wide model was deemed infeasible. The desired outputs at 100 m resolution were produced by developing a novel, hybrid downscaling method based on machine learning.

Hydrological models for five subcatchments, covering around 9% of Denmark and selected to represent a range of hydrogeological settings, were run at 100 m resolution with forcings from a reduced ensemble of climate models. Random

20 Forest algorithms were established using the simulated climate change impacts (future – present) on water table depth at 100 m resolution from those submodels as training data.

The trained downscaling algorithms then were applied to create nation-wide maps of climate change-induced impacts on the shallow groundwater table at 100 m resolution. These downscaled maps were successfully validated against results from a validation submodel at 100 m resolution excluded from training the algorithms, and compared to the impact signals from the

25 500 m hydrological model across Denmark.

The suggested downscaling algorithm also opens for the spatial downscaling of other model outputs. It has the potential for further applications where, for example, computational limitations inhibit running distributed hydrological models at fine resolutions.

1 Introduction

30   Groundwater accounts for a substantial part of the active hydrological cycle, and exhibits a wide range of interactions with ecosystems, food and energy security, as well as climate and flood regulation (Gleeson et al., 2016). Interactions with the land surface and other components of the hydrological cycle are particularly pronounced where the uppermost groundwater table is found within few metres of the surface. Such high groundwater tables affect up to a third of the earth's land surface (Fan et al., 2013; Soylu and Bras, 2022). Moreover, across many parts of the world, groundwater resources are not only

35   affected by human interactions, but also changing due to climate change (Rodell et al., 2018).

Denmark, with its gentle topography and temperate climate, has particularly high groundwater tables: The uppermost groundwater table (which we hereafter refer to as shallow groundwater) is located only few metres to decimetres below the surface across large parts of Denmark (Henriksen et al., 2020a; Koch et al., 2021); see Figure 1. This is also reflected in roughly 50% of agricultural land in Denmark being artificially drained (Møller et al., 2018; Olesen, 2009). Along with rising

40   precipitation in future climate (Pasten-Zapata et al., 2019), the shallow groundwater levels are generally expected to rise (Henriksen et al., 2020b; Refsgaard et al., 2016; van Roosmalen et al., 2007). The projected rise in groundwater level is most pronounced in winter, where groundwater levels are highest to start with, due to increased precipitation. During summer, the picture is more complex for shallow groundwater levels, with some areas showing falling groundwater tables due to increased temperature and evapotranspiration. More extreme and higher groundwater levels in the future pose significant

45   challenges for infrastructure, agriculture and ecosystems (Halsnæs et al., 2022). Due to the considerable small-scale variability of shallow groundwater levels (Koch et al., 2021), which are mainly controlled by topographic variability and hydrogeology, high-resolution information is required for purposeful groundwater management and climate adaptation.

This requirement for high resolution data is particularly relevant when evaluating climate change-induced impacts on the shallow groundwater table. The national water resource model of Denmark (the so-called DK-model) is based on a

50   distributed, coupled surface-subsurface model at 500 m horizontal resolution, and recently at a higher-resolution 100 m version with specific focus on surface-near processes. As forcing for climate change impact studies, a large ensemble of locally bias-corrected climate models is available (Pasten-Zapata et al., 2019). Ideally, national hydrological impact assessment would be based on the high-resolution version of the hydrological model. However, the 25-fold increase in computational nodes for the 100 m model compared to the 500 m model makes such a modelling task infeasible due to the

55   computational burden.

In order to obtain national impact projections at high resolution based on a large climate ensemble, but with a minimized computational cost, it is proposed to develop a machine learning (ML) based downscaling method to refine impact simulations from the computationally feasible 500 m hydrological model (HM) to a resolution of 100 m.

Within other fields such as remote sensing, ML-based spatial downscaling algorithms have been explored and used for

60   several years, for example in the DisALEXI modelling system (Anderson et al., 2004, 2021). The background here often is to bridge gaps between coarse-resolution imagery from some satellites that typically have a frequent revisit time, and high-

resolution imagery from other satellites with lower temporal resolution (Im et al., 2016; Yang et al., 2017). Moreover, image sharpening techniques that build scale-independent models are utilized to increase the spatial resolution of satellite data (Guzinski and Nieto, 2019). Similarly, spatial downscaling is widely applied for weather predictions and climate models
65   (e.g. Cheng et al., 2020; Sun and Tang, 2020). We hypothesize that applying similar techniques on outputs from complex HMs, combining the information of physically-based models with various high-resolution spatial data and ML methods will increase relevance of HM outputs. This is a relatively new field with few applications so far: For example, Koch et al., 2021 used HM outputs as covariates in ML algorithms predicting groundwater table depth at high spatial resolution, or Zhang et al., 2021 downscaled GRACE total water storage data with the help of HM output from the GLDAS Global Land Data
70   Assimilation System. Here, the presented research is contributing to further integrating physically-based HMs with ML techniques. This can also contribute to the discussion of further development of knowledge-guided or theory-informed ML techniques in hydrological science (Nearing et al., 2021).

Hence, the objective of this work is to develop a ML-based algorithm for spatial downscaling of physically-based HM outputs. This is applied to the downscaling of climate change impact predictions on the simulated depth to the shallow
75   groundwater table for Denmark, downscaling from a resolution of 500 m to 100 m. We favour a ML-based approach over simple interpolation methods or topography-driven downscaling, because ML has the capability to effectively learn multivariate relationships which we expect to be highly relevant for a complex variable like the shallow groundwater table. Even though running the DK-model nationwide at the fine 100 m resolution with a full ensemble of climate models is computationally too expensive, it is feasible to run some selected submodels at 100 m resolution and utilize their high
80   resolution outputs as training data for the ML-based downscaling. In addition, a single national scale deterministic run with historic climate is possible at 100 m, which can serve as valuable information to the downscaling algorithm. Also, all the relevant model input data is available at a resolution of at least 100 m.

2. Data and methods

2.1 DK-model HIP

85   The national water resource model for Denmark, the DK-model, covers most of the Danish land surface area of approximately 43,000 km². It has been continuously developed over several decades (Henriksen et al., 2003; Højberg et al., 2013; Stisen et al., 2019), and been used in various research projects (recent examples: Koch and Schneider, 2022; Noorduijn et al., 2021; Schneider et al., 2022; Soltani et al., 2021), as well as public consultancy and in relation to the EU Water Framework Directive. It targets, for example, questions of water resource availability, water quality, and future
90   impacts on the hydrological cycle due to climate and land use change. Most versions of the DK-model have a horizontal resolution of 500 m, while a version at 100 m horizontal resolution was created as part of the Danish Hydrologic Information and Prognosis System (HIP) (Henriksen et al., 2020b). This most recent version, hereafter referred to as DK-model HIP, is the basis of the presented work. Figure 1 displays mean historic depths to the shallow groundwater table as simulated by the

DK-model HIP: Shallow groundwater levels are within the first one to two metres from the surface for large parts of the
95   country, with variations mostly controlled by topography and geology.

2.1.1 General hydrological model setup in MIKE SHE

The DK-model HIP is an integrated, distributed surface water-groundwater model, setup in the MIKE SHE model code
(Abbott et al., 1986; DHI, 2020). MIKE SHE is used to couple 3D subsurface flow, 2D overland flow, a simple 2-layer
description of the unsaturated zone, and 1D kinematic routing of flow in the stream network. It is run as a transient model
100   with daily climate forcing and a maximum timestep of 24 hours.

The description of the DK-model is kept short here, as the model itself and its development is not the focus of this paper. For
details, the readers are referred to the provided literature.

2.1.2 Input data and conceptualization

The DK-model HIP exists in two distinct horizontal resolutions of 500 m and 100 m, with all relevant input data available at
105   100 m resolution. The saturated zone is divided into 9 to 11 computational layers of varying thickness, depending on the
region. The 3D unit-based parameterization of the subsurface is based on a nationwide hydrogeological model, with the
exception of the uppermost computational layer with a constant thickness of 2 m which is parameterized based on the Danish
soil map (Jakobsen et al., 2015).

Human water use is included to the extent that groundwater extractions for drinking water and irrigation are included as well
110   as sewage plant outflows are added to the model. Moreover, the model also includes a representation of (subsurface)
drainage, which is widespread in Denmark (as described in Schneider et al., 2022).

As historic climate forcing for precipitation, temperature and potential evapotranspiration, the gridded, daily data from the
Danish Meteorological Institute is used (Scharling, 1999a, 1999b). Temperature and potential evapotranspiration are
provided at 20 km resolution, and precipitation at 10 km. Daily precipitation is corrected as described by Stisen et al., 2011.

115   ### 2.1.3 Model calibration

The national DK-model HIP was calibrated in its 500 m version, against observations of groundwater heads and streamflow
for the period 2000 to 2010, and validated from 1990 to 1999. For calibration, the model optimization tool PEST (Doherty,
2015) was used with its implementation of the Gauss-Marquardt-Levenberg algorithm. Again, due to excessive
computational effort, it was not possible to conduct such a calibration, requiring a large number of model runs, with the full
120   100 m national model. However, through a validation comparison, it was deemed feasible to directly transfer the calibrated
parameters from the 500 m version of the model to its 100 m version without loss of performance. A re-calibration of a
limited set of parameters for ten subcatchments in 100 m yielded no substantial additional improvement in model
performance. Details regarding the model calibration are beyond the scope of this paper and can be found in Henriksen et al.,
2020a.

2.2 Input data for future climate scenarios

2.2.1 Climate models

An ensemble of regional climate models (RCM) from the Euro-CORDEX initiative (Jacob et al., 2014) was used in this work. Those RCMs were locally bias corrected for Denmark and remapped to the same 10 km and 20 km resolution of the historical climate data by (Pasten-Zapata et al., 2019). All RCMs cover the years 1971 to 2100 with daily timesteps. For the work presented here, we used the 17 RCMs representing the RCP8.5 greenhouse gas concentration scenario (van Vuuren et al., 2011). Out of that ensemble of 17 RCMs, we selected a subset of five RCMs. Those five RCMs were selected based on their ranges of projected precipitation, to cover a wide range of future climates, and ensuring that the median precipitation predictions across the five selected RCMs are close to the median precipitation predictions of the entire ensemble of 17. This subset of five RCMs was used with the 100 m submodels (Figure 1), producing the training data for the downscaling.

2.2.2 Further input data for future scenarios

Beyond climate forcing, some further input data were adapted to account for future conditions: Sea level is included as fixed head boundaries along the HM's sea boundaries. For the historic model runs, the sea level is kept constant at elevation 0 m. For the future scenarios, sea level rise was accounted for as projected by the Danish Meteorological Institute (Thejll et al., 2021).

Changes in groundwater abstraction rates, sewage plant outflows and land use were not projected, but current historic values and maps were assumed for the future period. Irrigation is simulated demand driven in the model; hence, it automatically accounts for climatic changes.

2.3 Climate change impact runs with the hydrological models

An overview over the different climate change impact runs performed with the HMs is given in Table 1: Due to computational limitations, we ran the full, 17 RCM ensemble only with the 500 m national HM (C; B is a subset of C). Besides that, we selected five training subcatchments, covering in total about 9% of Denmark's surface area, and ran those with five selected RCMs at 100 m resolution (D). To allow for a fully independent validation, a sixth validation submodel was run at 100 m resolution, with the full RCM ensemble (E). Finally, the DK-model HIP was also run for all of Denmark with observed climate forcing, for the period 1990 to 2019. These historic runs were performed both at 500 m and 100 m resolution (A and *). Relevant HM outputs were initially stored at daily timestep and then aggregated as described below.

For the climate change impact evaluation, the HMs were run using each individual member of the ensemble of RCMs as forcing, for three different periods: The reference period, serving as baseline, as well as the near future period (nf), and the far future period (ff). For the reference period, model outputs were extracted from 01-01-1991 to 31-12-2020, for the near future from 01-01-2041 to 31-12-2070, and for the far future from 01-01-2071 to 31-12-2100. The HMs were started from

155 initial conditions taken from a continuous run of the corresponding HM across the entire period of 1991 to 2100. A warmup
period of four years prior to the start dates for result extraction was allowed for each of the HM runs.

2.3.1 Climate change-induced changes to the shallow groundwater table

Finally, the climate change impacts of interest were the average conditions across each of the future 30-year periods,
subtracted from the reference 30-year period. These climate change impacts on the shallow groundwater table are hereafter
160 also referred to as to-be-downscaled-variables (TBDV).

That means we focus on *changes* to the groundwater table caused by future climate change, as opposed to absolute values, as
changes can be more useful than absolute values when comparing present-day to future conditions. This is mostly because
changes to the groundwater table are typically small compared to the absolute depth to the groundwater table and the
uncertainties in the modelled absolute depth. Hence, when evaluating future conditions, an absolute value predicted by the
165 HM can only be used in direct comparison to present-day output of the same HM. If being used in different contexts, for
example in comparison with more detailed local HM results or observations of the groundwater table, the discrepancies
between these absolute values will dominate and potentially mask the projected changes. When using predicted changes
between reference and future period HM outputs with the same HM setup, those projected changes can be attributed with
high confidence to the changes in climate.

170 Hence, we downscale projected climate change-induced changes to the shallow groundwater table. Different aggregated
statistics were chosen, and changes were calculated for both the near and far future periods. The chosen aggregated statistics
are the changes to the mean, 1st percentile and 99th percentile of the simulated groundwater table (high and low groundwater
tables, respectively), as well as changes to the 1 m exceedance probability. The latter represents the fraction of time the
groundwater table is closer than 1 m to terrain during the given period. This results in a total of eight statistics to be
175 downscaled (TBDV, see also Table 2). For example, the change to 1st percentile depth to the groundwater table for the near
future period $Q01_{nf}$ is calculated as

$$Q01_{nf} = median(Q01(nf_i) - Q01(ref_i)) \tag{1}$$

where $Q01(nf_i)$ is the 1st percentile of depth to the groundwater table simulated in each of the HM's grid cells, across all
daily values of the 30-year near future period, and $Q01(ref_i)$ is the respective value for the 30-year reference period. i refers
180 to each of the individual members of the RCM ensemble used, and the actual TBDV then is the median change across the
RCM ensemble. In equivalent manner, all the eight TBDV $mean_{nf}$, $mean_{ff}$, $Q01_{nf}$, $Q01_{ff}$, $Q99_{nf}$, $Q99_{ff}$, $ex1m_{nf}$, and $ex1m_{ff}$ are
calculated. Moreover, in the same manner as the TBDV were calculated based on the 500 m national DK-model HIP outputs,
the respective training data was calculated based on the 100 m submodels.

2.3.2 National climate change impact runs in 500 m

185 The full national DK-model HIP was run at 500 m resolution with the full ensemble of 17 RCMs (which includes the five selected training RCMs). This 500 m model output represents the data that is to be downscaled to finer 100 m resolution with the presented method.

2.3.3 Submodel runs in 100 m

Six submodels distributed across Denmark (outlined and marked in Figure 1), each representing a hydrologic subcatchment
190 and setup at 100 m resolution, were chosen to produce training and validation data for the downscaling algorithm:

I.   Suså
II. Odense Å
III. Kongeå/Kolding Å
IV. Storå
195 V. Simested Å/Mariager Fjord
VI. Aarhus Å/Aarhus

The five submodels I to V were used as training data, whereas submodel VI was used as a fully independent validation dataset. To reduce computational effort spent on HMs, the five training models were run with the five selected RCMs, i.e. the training data is based on the reduced RCM ensemble. For the validation, to address the extrapolation capability of the
200 downscaling algorithm, the validation model was also run with the full ensemble of 17 RCMs. The submodels were run with the same general model setup, periods and warmup as for the national model described at the start of section 2.3. Along the submodels' land boundaries, dynamic boundary conditions were applied based on the simulated groundwater heads in the corresponding 500 m national HM.

The five training submodels, along with the sixth validation model, were selected to cover a wide range of geologic,
205 topographic and hydrologic variability occurring across Denmark. ML methods are known to perform poorly when used to extrapolate data beyond the range of data they are trained against (Meyer and Pebesma, 2021). Figure 2 displays the histograms of one of the variables to be downscaled (TBDV), together with some important covariates. The values covered by the five training submodels closely resemble the distribution and ranges seen across all of Denmark; i.e. the submodels are considered representative of all of Denmark.

210 ## 2.4 ML-based downscaling

2.4.1 Covariates

Table 2 presents an overview over the covariates used in the downscaling algorithm. The first type of covariates are topography-related, and contain six different covariates. Secondly, 14 covariates derived from the DK-model HIP parameters were used: This group contains the horizontal transmissivities for the uppermost seven computational layers of the model's
215 saturated zone, and the mean horizontal hydraulic conductivities across five depth intervals, as well as the spatially

distributed drain depths and drain time constants used in the model. The last group contains the actual 500 m change to the groundwater table that is to be downscaled (the TBDV). Besides that, the respective *absolute* value corresponding to each TBDV from the 100 m historic HM is included, as well as the differences in TBDV between a historic dry and wet period, which can be seen as a proxy for expected changes with future climate.

220 All covariates are available at the target resolution of 100 m for all of Denmark, to allow the application of the downscaling algorithm to the entire country. The only exception obviously is the TBDV itself, which only is available in 500 m nationwide. It was resampled from its native resolution to 100 m using bilinear interpolation, as initial tests showed performance improvements of the downscaling algorithm if interpolated TBDV were used.

More covariates have been used in preliminary tests. Analyses of covariate importance and collinearity lead to the final set of

225 23 covariates.

2.4.2 Random Forest regressor

For the downscaling task at hand, we decided to use a Random Forest (RF) regressor. RF goes back to Breiman, 2001 and has over time proven to be a powerful and versatile data-driven modelling tool for a range of applications in environmental sciences. For example, RF was used in the context of groundwater pollution (Rodriguez-Galiano et al., 2014; Tesoriero et al.,

230 2017), data analysis and predictions in large-sample hydrology (Addor et al., 2018; Ghiggi et al., 2019), as well as some of the remote sensing downscaling models mentioned in section 1. Examples from related Danish contexts include the prediction of groundwater level changes based on well observations (Gonzalez and Arsanjani, 2021), modelling of the historic depth to shallow groundwater and redox boundary (Koch et al., 2019b, 2019a), or the modelling of artificial drainage properties (Møller et al., 2018; Motarjemi et al., 2021).

235 RF is a supervised ML learning method; that means it requires training data. Based on the training dataset, a RF model learns about relationships between a set of covariates and the target (training) data. In the next step, predictions at unsampled locations beyond the training data can be made. Tyralis et al., 2019 provide a concise overview over the theory behind and the use of RF in hydrological contexts.

We used the implementation of RF regressors in Python 3.8 from the scikit-learn package, version 1.0.2 (https://scikit-

240 learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html).

In Figure 3, a workflow diagram of the downscaling algorithm is shown: All covariates exist at 100 m resolution for all of Denmark, with the exception of the actual TBDV, which originates from the 500 m HM. The TBDV exists at the finer target resolution of 100 m for the five training submodels – these data are used as training data for the RF downscaling regressors. After training, the RF regressors are used to produce Denmark-wide maps of the TBDV in 100 m. The downscaled outputs

245 are validated against both national HM output, the training data itself, and a sixth, independent validation submodel in 100 m. Details are presented in the following sections.

2.4.3 Dummy points for RF training

Initial tests of the downscaling algorithm showed issues with the robustness of predictions outside the submodels, i.e. outside the regions with training data. To increase robustness, it was decided to include additional training data sampled from the
250 entirety of Denmark outside the training submodels, which enhances the spatially limited training dataset. These points were sampled randomly in space. The actual training target – the TBDV in 100 m – is lacking for those points, and instead the TBDV in 500 m was used as training target. Hence, we refer to these points as dummy points. The inclusion of such dummy points increases the robustness of the RF downscaling outside the training data areas; however, the more dummy points are included, the closer the downscaled output resembles the original TBDV in 500 m, which is undesired. Considering this
255 trade-off, we tested different amounts of dummy points, and settled on an optimal number of 20,000 dummy points covering all of Denmark to be included in the training data. This adds 5% to the total number of data points in the training dataset comprised of the five training submodels, where all 100 m HM grids were included in the training data (approximately 400,000 points).

2.4.4 Training the RF regressors

260 The four different climate change impact statistics regarded in this work – that is changes to the depth to the groundwater table in average (*mean*), high (*Q01*), and low (*Q99*) conditions, as well as the probability of the groundwater table exceeding 1 m below terrain (*ex1m*) – are expected to be controlled by different properties. Hence, it was decided to train a separate RF downscaler for each of the four TBDV. However, to ensure the consistency across near and far future and to increase the robustness of the downscaler by increasing the amount of training data, each of the downscalers was trained against data
265 from both near future and far future periods simultaneously. That means that four RF downscaler models were trained; one for each *mean*, *Q01*, *Q99*, and *ex1m*.

The training was performed using the 100 m target values from the five fine-scale submodels (D in Table 1), based on the reduced RCM ensemble. The respective TBDV in 500 m was taken from the national HM (B in Table 1). Beyond the training data in 100 m and the TBDV in 500 m, the other covariates presented in Table 2 were used.

270 Based on the training data, the importances of each of the covariates was evaluated: The information content (importance) that each covariate delivers to the RF regressors is determined by randomly perturbing each of the covariates one by one, retraining the model, and observing the decrease in model fit.

2.4.5 RF hyperparameter search

RF regressors have built-in parameters controlling model behaviour and complexity, the so-called hyperparameters. To fine-
275 tune and optimize the RF regressor performance, a hyperparameter search was performed using the standard grid-search procedure. The tested parameter values, resulting in 144 unique combinations, are listed in Table 2. For each of the parameter combinations, and each of the TBDV, the following was performed:

(i) An evaluation of the ability of the RF regressor to reproduce the 100 m HM results by determining Pearson's R. This was performed in a spatial cross-validation test, where the RF regressor was trained against four of the five training submodels,

280   and its results validated against the hold-out submodel, looping through all five permutations of train/hold-out submodels.

(ii) An evaluation of the Denmark-wide bias introduced by the downscaling. Here, the RF regressor was trained against all five training submodels, and then used for predictions for all of Denmark. Those predictions were compared against the national 500 m HM results.

In this setup, (i) presents an indication of the RF regressor's ability to reproduce the fine-scale 100 m results, whereas (ii)

285   represents an indication of the transferability and robustness of the RF regressor, with decreasing performance in case of overfitting. Based on averaging these results across all TBDV, a hyperparameter set representing an optimal trade-off was determined. These hyperparameters, marked bold in Table 3 alongside the other reported parameters, then were used in all subsequent tasks.

2.4.6 Validation of the RF regressors' downscaled outputs

290   Subsequent to training the RF regressors with the optimal hyperparameters, they were applied to produce predictions for the entirety of Denmark. That means that the four RF downscaler models were used to downscale the national 500 m HM outputs to 100 m, separately for both the near and far future period, and for both the ensemble of five RCMs used in the training as well as for the full ensemble of 17 RCMs. The only difference in downscaling for the different periods and RCM ensembles is in the used TBDV in 500 m, which always represents the corresponding period and RCM ensemble.

295   Validation of the downscaled results in 100 m then was performed in two distinct ways: First, against the national 500 m HM output, verifying that the downscaling does not introduce an overall bias. Second, against the 100 m HM output from the validation submodel VI, which was not used in the training process.

3. Results and Discussion

3.1 Covariate importance

300   Figure 4 summarizes the results of the covariance importance analysis, for each of the final trained RF downscalers. The covariates are sorted after their average importance across the four downscalers. Based on the average importance scores, it is noteworthy that the TBDV turns out as the most important covariate, followed by the respective historic absolute value of the TBDV and the difference between historic dry and wet periods. That means that the downscaling is largely guided by information on the TBDV itself, which is desired as our aim was to develop a *downscaling* algorithm. Turning to the

305   topography and model parameter-related covariates, and comparing the different downscalers, it can be seen that *Q01* (high groundwater tables) is strongly guided by the hydraulic conductivities in the uppermost layers (the four most important covariates from that group being trh_lay0, trh_lay1, trh_lay3, and kh_mean_2-5m). *Q99* (low groundwater tables), on the

other hand, is more guided by topography and hydraulic conductivity from lower layers (the four most important covariates from that group being topo, kh_mean_2-5m, kh_mean_5-10m, and trh_lay5).

3.2 Downscaled output

Figure 5 presents different input data and results for the same area located at the north eastern edge of submodel III Kongeå/Kolding Å (extent indicated in Figure 1), for the example TBDV $mean_{ff}$. The top row shows examples for two of the used covariates, the relative topography and the horizontal hydraulic transmissivity in the second layer of the HM. The centre right map shows the simulated mean groundwater heads from the historic 100 m HM, which is used as a covariate. In the centre left map, the used training data from the 100 m hydrological submodel can be seen, whereas the bottom left map shows the corresponding downscaled result at 100 m resolution. Comparing those two maps, together with the actual TBDV in the bottom right map, i.e. the results from the HM at 500 m resolution, gives a good visual example of the capability of the downscaling algorithm to reproduce the details at 100 m resolution from the 500 m coarse resolution input. It also emphasizes the value of the fine-scale information: As can be seen, the 100 m HM output resolves significant variations in both the absolute values of the depth to the groundwater table as well as the climate change-induced changes to the groundwater table that are completely lost at 500 m resolution. The complexity of the fine-scale variations also goes beyond what can be achieved by merely interpolating the coarser 500 m to the finer 100 m resolution. Rather, it becomes apparent that local variations in groundwater levels and level changes are controlled by an interplay of topography and geology.

Figure 6 illustrates downscaling results for the fully independent validation submodel VI Aarhus Å/Aarhus (extent indicated in Figure 1), for the example of TBDV $ex1m_{ff}$. The left column shows the median change of the five RCMs used in the training exercise, where the top displays the TBDV from the 500 m HM (B in Table 1). The middle shows the respective TBDV as simulated by the 100 m HM submodel (E in Table 1). The bottom map then displays the respective downscaling results in 100 m. The right column shows the corresponding results for the full RCM ensemble of 17 models. These 17 RCMs were not used in training, but only in validation. Even if confronted with a different climate signal (compare the differences between the 500 m HM outputs in the top row Figure 6), and used outside the area of training data, the downscaling algorithm still works well. That means it is robust enough to be trained based on a (small) RCM ensemble, and then be applied to HM output based on a different (larger) RCM ensemble, which also is in line with the findings from the covariate importance analysis, where the corresponding TBDV showed to be the most important covariate.

These visual evaluations of the downscaling outputs can be confirmed quantitatively – Figure 7 presents an overview. The top row presents a Denmark-wide evaluation of the climate change-induced changes to the shallow groundwater table. The bars present the mean changes across all of Denmark for each of the eight TBDV, as predicted by the 500 m national HM runs, and the downscaled outputs in 100 m, respectively. It can be observed that generally, the downscaling does not introduce overall biases to the predicted changes, but remains true to the general picture predicted by the coarser national HM. This applies to both the ensemble of five RCMs, as well as the full RCM ensemble with 17 models. The bottom two rows of Figure 7 display validation results of the downscaling output against 100 m HM outputs for submodel VI. I.e. this is

a fully independent validation of the 100 m RF downscaling output versus 100 m HM outputs. Spatial transferability to independent areas without training data is a well-known weakness of the used ML techniques. However, our RF downscaling model proves to be robust also in this context: Pearson correlation coefficients between the 100 m RF output and the 100 m HM output are consistently higher than correlation coefficients between 500 m HM output and 100 m HM

345   output. It can also be seen that the RF downscaling adds more information than a simple interpolation, as results are given separately for the 500 m HM outputs and a bilinear interpolation of the 500m HM outputs. The latter is considered the benchmark to be beaten to determine whether a downscaling algorithm actually delivers additional information over the coarse-scale HM outputs. The same applies to the mean absolute error (MAE) between the 100 m RF downscaling output and the 100 HM, which is consistently smaller than the MAE between the 500 m HM and 100 m HM.

350   For the example of the TBDV $mean_{ff}$, per-pixel data is presented in scatterplots in Figure 8. The left panel shows a comparison of the 100 m training data from the HM of the five training submodels against the corresponding TBDV from the 500 m HM. Most values range between roughly -10 cm and 30 cm, which means that the mean groundwater table for the far future period is predicted to mostly be between 10 cm lower to 30 cm higher than in the reference period. There is good general agreement between the 500 m and 100 m HMs and no general bias, but a considerable scatter caused by the

355   differences in resolution. This is in line with the previously discussed lack of bias in the predicted changes from the 500 m and 100 m HMs as apparent from the top row of Figure 7. The right panel shows a comparison of the 100 m training data and the RF downscaling predictions in 100 m. Here it can be seen that the downscaled results follow the 100 m HM outputs more closely than the 500 m HM outputs, with a much narrower scatter.

Finally, Figure 9 displays Denmark-wide results of the downscaling algorithm, for all four TBDV of the far future period

360    $mean_{ff}$, $Q01_{ff}$, $Q99_{ff}$, and $ex1m_{ff}$. Each of the detail maps shows the same extent: They serve as a good illustration that groundwater tables (and changes to them) are controlled by small-scale landscape features, such as topography and geology. If only 500 m information was available, many details would be lost that are apparent in 100 m resolution. For many purposes, e.g. agriculture, information on the groundwater table at the coarse scale of 500 m is inappropriate – a grid size of 25 ha is larger than typical field size. However, with a resolution of 100 m, i.e. 1 ha, we move into more relevant, near field-

365   size-resolutions. Also, as the downscaled outputs are still based on physically-based HMs, fine-scale physical information is added on the interplay of controls such as topography and geology on the groundwater table.

Moreover, Figure 9 also displays the initially mentioned tendency to more extreme groundwater levels: The low groundwater tables ($Q99$), typically occuring during the summer months, are projected to fall across many regions of the country, especially across Zealand and southern Jutland where groundwater levels are very shallow and affected by

370   evapotranspiration. The high groundwater tables ($Q01$), typcially occuring during the wetter winter months, are simulated to rise further for most of the country. Some of the highest rises are simulated for areas in central Jutland where the shallow groundwater levels are deep below surface (compare Figure 1). Moderate rises are also simulated in regions where the groundwater tables are very shallow (within the first 1 m or 2 m below surface), such as parts of Zealand and Funen.

3.3 Overarching discussion

375   Despite the discussed capabilities of ML techniques in exploiting large datasets at high spatial resolution, physically-based HMs are still needed to predict climate change-induced changes to the hydrological cycle. Purely data-driven methods (often based on ML algorithms) struggle with predicting previously unobserved states, which holds true for both, spatial and temporal extrapolation (Koch and Schneider, 2022; Mai et al., 2021; Meyer and Pebesma, 2021). As an attempt to overcome the computational burden of high resolution, complex HM, Tran et al., 2021 have developed a ML-based emulator of a HM.

380   Read et al., 2019 and Koch and Schneider, 2022 showed that guiding ML models with simulation data from a physically-based model enhanced the capability of the ML model to extrapolate. The downscaling algorithm presented in this work increases the resolution of HM outputs, guided by fine-resolution HM outputs.

As our method is based on output from distributed, physically based HMs, we (i) are able to obtain fully distributed climate change impact evaluations, and (ii) have higher confidence in extrapolating the models to values outside observed ranges,

385   which is unique and, for example, goes beyond recent efforts with purely data-driven extrapolations of climate change impacts on groundwater levels for selected wells across Germany (Wunsch et al., 2022). Our developed downscaling algorithms were shown to be robust to be transferrable in space, and also transferrable to a different climate signal, i.e. transferable to a different RCM ensemble.

Acknowledging computational limitations, we based our 100 m training data on five HM subcatchments, representing

390   around 9% of Denmark, which were run with forcings from a reduced ensemble of five RCM. This meant that computational time for hydrological model runs was reduced to less than 3% of what would be required for a nation-wide full ensemble run at 100 .

The downscaling was facilitated by the consistency of parameter sets between the 500 m and 100 m HM setups: We do not intend to compensate for potential discrepancies in simulated groundwater heads due to differences in parameter values, but

395   merely want to downscale the 500 m national HM results to a finer scale as simulated by the corresponding model in higher resolution. The same applies for all the other input data: For both 100 m and 500 m HMs, the same input data was being used, with the only exception of scale.

4. Conclusions

We successfully designed and tested a RF-based spatial downscaling algorithm for outputs from distributed HMs. The usage

400   of ML techniques in spatial downscaling is widespread in fields such as remote sensing, however, still limited in hydrologic modelling. The developed downscaling algorithm is based on the existence of a coarse-scale HM for the full domain of interest, together with equivalent fine-scale HMs for limited parts of the domain. Furthermore, covariates in fine scale are required for the full domain. We trained the RF regressor based on the selected, representative fine-scale HMs covering only around 9% of the full domain, significantly reducing computational effort on the complex HMs. Furthermore, using the

405   downscaling algorithm in the context of an ensemble of RCMs, we could show that it is possible to train the downscaling

algorithm based on a smaller ensemble of RCMs, and apply it to the full ensemble, further reducing computational effort on HMs. The downscaling results were successfully validated against the output from a fully independent fine-scale validation HM, as well as the full-domain results of the coarse-scale HM.

The presented framework is envisioned to be transferrable to the downscaling of other spatial output from hydrological and
410   environmental models in general also beyond aggregated statistics (i.e. also in transient manner). It is generally acknowledged that there is a disparity between the – computationally possible – spatial resolution of HM output, and what is desired or required by users of the data, such as authorities, consultants or citizens (Samaniego et al., 2019). Also from a scientific point of view, we are commonly dealing with HM resolutions issues and spatial aggregation beyond what is appropriate, hindering process understanding (Nijzink et al., 2016; Wood et al., 2011).

415   This also applies to the greater context of this paper, the Danish HIP4Plus project (Henriksen et al., 2020b). Within this project, a larger dataset of climate change-induced changes to the shallow groundwater table was generated based on the presented framework. This includes predictions for different greenhouse gas concentration scenarios (RCP4.5 and RCP8.5), further statistics, and monthly as well as seasonally aggregated values; a total of 236 variables (as compared to the eight presented here). All downscaled maps and further data can be accessed via the HIP data portal
420   https://hip.dataforsyningen.dk/. Due to computational limits, the spatial resolution of HMs is, in many cases as in the presented example, below what is requested by end-users and decision makers. Downscaling can help bridge the gap to make HM output more relevant, in this case in the context of climate change impact evaluation and adaptation, to deliver valuable input to urban and infrastructure planning, or agriculture and future needs for artificial drainage or water pollution risks.

Data and code availability

425   Denmark-wide data produced with the presented downscaling algorithm, for a total of 236 different variables, is freely available from the HIP data portal: https://hip.dataforsyningen.dk/ Note, however, that this is not based on the exact same data used and presented in the paper. The exact data from this paper, as well as the used code, will be made available upon request by the main author without undue reservation.

Acknowledgements

430   The presented work was originally developed within the HIP4Plus project, headed by SDFE (Styrelsen for Dataforsyning og Effektivisering – the Danish Agency for Data Supply and Efficiency), and part of FODS6.1 (Fællesoffentlige Digitaliseringsstrategi Initiativ 6.1, Fælles data om terræn, klima, vand – the Danish public digitization initiative). Furthermore, we want to acknowledge the great team effort of the entire team of the HIP4Plus project at the Department of Hydrology, GEUS. The team encompassed, besides the authors of this paper: Jane Gotfredsen, Annesofie Jakobsen, Søren
435   Julsgaard Kragh, Maria Ondracek, Ernesto Pasten-Zapata, Per Rasmussen.

Author contribution

Raphael Schneider designed the downscaling algorithm, in close cooperation with Simon Stisen and Julian Koch, and further inputs from Hans Jørgen Henriksen and Lars Troldborg. The code for the downscaling algorithm, as well as validation and visualization were written by Raphael Schneider. Setting up the hydrological models was a joint effort of the entire project

440 team. Processing of the hydrological model output was handled by Lars Troldborg. Raphael Schneider prepared the manuscript with contributions from all co-authors.

Competing interests

The authors declare that they have no conflict of interest.

References

445 Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E. and Rasmussen, J.: An introduction to the European Hydrological System — Systeme Hydrologique Europeen, "SHE", 1: History and philosophy of a physically-based, distributed modelling system, J. Hydrol., 87(1–2), 45–59, doi:10.1016/0022-1694(86)90114-9, 1986.

Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N. and Clark, M. P.: A Ranking of Hydrological Signatures Based on Their Predictability in Space, Water Resour. Res., doi:10.1029/2018WR022606, 2018.

450 Anderson, M. C., Norman, J. M., Mecikalski, J. R., Torn, R. D., Kustas, W. P. and Basara, J. B.: A Multiscale Remote Sensing Model for Disaggregating Regional Fluxes to Micrometeorological Scales, J. Hydrometeorol., 5(2), 343–363, doi:10.1175/1525-7541(2004)005<0343:AMRSMF>2.0.CO;2, 2004.

Anderson, M. C., Yang, Y., Xue, J., Knipper, K. R., Yang, Y., Gao, F., Hain, C. R., Kustas, W. P., Cawse-Nicholson, K., Hulley, G., Fisher, J. B., Alfieri, J. G., Meyers, T. P., Prueger, J., Baldocchi, D. D. and Rey-Sanchez, C.: Interoperability of

455 ECOSTRESS and Landsat for mapping evapotranspiration time series at sub-field scales, Remote Sens. Environ., 252, 112189, doi:10.1016/j.rse.2020.112189, 2021.

Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology, Hydrol. Sci. Sci. Hydrol., 24(1), 43–69, 1979.

Breiman, L.: Random Forests, Mach. Learn., 45, 5–32, doi:10.1023/A:1010933404324, 2001.

460 Cheng, J., Kuang, Q., Shen, C., Liu, J., Tan, X. and Liu, W.: ResLap: Generating High-Resolution Climate Prediction through Image Super-Resolution, IEEE Access, 8, 39623–39634, doi:10.1109/ACCESS.2020.2974785, 2020.

Danish EPA: FOHM - Fælles offentlig hydrologisk model, [online] Available from: https://mst.dk/natur-vand/vand-i-hverdagen/grundvand/grundvandskortlaegning/kortlaegning-2016-2020/fohm-faelles-offentlig-hydrologisk-model/ (Accessed 11 March 2020), 2019.

465 DHI: MIKE SHE - User Guide and Reference Manual, , 822 p. [online] Available from:

https://manuals.mikepoweredbydhi.help/2020/Water_Resources/MIKE_SHE_Print.pdf, 2020.

Doherty, J.: Calibration and Uncertainty Analysis for Complex Environmental Models, Watermark Numerical Computing, Brisbane, Australia., 2015.

Fan, Y., Li, H. and Miguez-Macho, G.: Global Patterns of Groundwater Table Depth, Science (80-.)., 339(6122), 940–943,
470 doi:10.1126/science.1229881, 2013.

Ghiggi, G., Humphrey, V., Seneviratne, S. I. and Gudmundsson, L.: GRUN: an observation-based global gridded runoff dataset from 1902 to 2014, Earth Syst. Sci. Data, 11(4), 1655–1674, doi:10.5194/essd-11-1655-2019, 2019.

Gleeson, T., Befus, K. M., Jasechko, S., Luijendijk, E. and Cardenas, M. B.: The global volume and distribution of modern groundwater, Nat. Geosci., 9(2), 161–164, doi:10.1038/ngeo2590, 2016.

475 Gonzalez, R. Q. and Arsanjani, J. J.: Prediction of Groundwater Level Variations in a Changing Climate: A Danish Case Study, ISPRS Int. J. Geo-Information, 10(11), 792, doi:10.3390/ijgi10110792, 2021.

Guzinski, R. and Nieto, H.: Evaluating the feasibility of using Sentinel-2 and Sentinel-3 satellites for high-resolution evapotranspiration estimations, Remote Sens. Environ., 221, 157–172, doi:10.1016/j.rse.2018.11.019, 2019.

Halsnæs, K., Larsen, M. A. D. and Drenck, K. L.: Samfundsøkonomiske konsekvenser af oversvømmelser og investeringer i
480 klimatilpasning. [online] Available from: https://www.dnnk.dk/wp-content/uploads/2022/01/Samfundsoekonomiske-konsekvenser-af-oversvoemmelser-og-investeringer-i-klimatilpasning-_-final_reduced.pdf, 2022.

Henriksen, H. J., Troldborg, L., Nyegaard, P., Sonnenborg, T. O., Refsgaard, J. C. and Madsen, B.: Methodology for construction, calibration and validation of a national hydrological model for Denmark, J. Hydrol., 280(1–4), 52–71, doi:10.1016/S0022-1694(03)00186-0, 2003.

485 Henriksen, H. J., Kragh, S. J., Gotfredsen, J., Ondracek, M., van Til, M., Jakobsen, A., Schneider, R. J. M., Koch, J., Troldborg, L., Rasmussen, P., Pasten-Zapata, E. and Stisen, S.: Dokumentationsrapport vedr. modelleverancer til Hydrologisk Informations- og Prognosesystem. [online] Available from: https://sdfe.dk/media/2920327/hip4plus_dokumentationsrapport_v23feb.pdf, 2020a.

Henriksen, H. J., Kragh, S. J., Gotfredsen, J., Ondracek, M., van Til, M., Jakobsen, A., Schneider, R. J. M., Koch, J.,
490 Troldborg, L., Rasmussen, P., Pasten-Zapata, E. and Stisen, S.: Sammenfatningsrapport vedr. modelleverancer til Hydrologisk Informations- og Prognosesystem. [online] Available from: https://sdfe.dk/media/2920328/hip4plus_sammenfatningsrapport_v23feb.pdf, 2020b.

Højberg, A. L., Troldborg, L., Stisen, S., Christensen, B. B. S. and Henriksen, H. J.: Stakeholder driven update and improvement of a national water resources model, Environ. Model. Softw., 40, 202–213, doi:10.1016/j.envsoft.2012.09.010,
495 2013.

Im, J., Park, S., Rhee, J., Baik, J. and Choi, M.: Downscaling of AMSR-E soil moisture with MODIS products using machine learning approaches, Environ. Earth Sci., 75(15), 1120, doi:10.1007/s12665-016-5917-6, 2016.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K.,

500   Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J. F., Teichmann, C., Valentini, R., Vautard, R., Weber, B. and Yiou, P.: EURO-CORDEX: New high-resolution climate change projections for European impact research, Reg. Environ. Chang., 14(2), 563–578, doi:10.1007/s10113-013-0499-2, 2014.

Jakobsen, P. R., Hermansen, B. and Tougaard, L.: Danmarks digitale jordartskort 1:25000 - Version 4.0. [online] Available
505   from: http://ku-gis.dk/blog/wp-content/uploads/2018/01/jordart25000_beskriv.pdf, 2015.

Koch, J. and Schneider, R.: Long short-term memory networks enhance rainfall-runoff modelling at the national scale of Denmark, GEUS Bull., 49, 8292, doi:10.34194/geusb.v49.8292, 2022.

Koch, J., Stisen, S., Refsgaard, J. C., Ernstsen, V., Jakobsen, P. R. and Højberg, A. L.: Modeling Depth of the Redox Interface at High Resolution at National Scale Using Random Forest and Residual Gaussian Simulation, Water Resour. Res.,
510   55(2), 1451–1469, doi:10.1029/2018WR023939, 2019a.

Koch, J., Berger, H., Henriksen, H. J. and Sonnenborg, T. O.: Modelling of the shallow water table at high spatial resolution using random forests, Hydrol. Earth Syst. Sci., 23(11), 4603–4619, doi:10.5194/hess-23-4603-2019, 2019b.

Koch, J., Gotfredsen, J., Schneider, R., Troldborg, L., Stisen, S. and Henriksen, H. J.: High Resolution Water Table Modeling of the Shallow Groundwater Using a Knowledge-Guided Gradient Boosting Decision Tree Model, Front. Water,
515   3, 701726, doi:10.3389/frwa.2021.701726, 2021.

Mai, J., Tolson, B. A., Shen, H., Gaborit, É., Fortin, V., Gasset, N., Awoye, H., Stadnyk, T. A., Fry, L. M., Bradley, E. A., Seglenieks, F., Temgoua, A. G. T., Princz, D. G., Gharari, S., Haghnegahdar, A., Elshamy, M. E., Razavi, S., Gauch, M., Lin, J., Ni, X., Yuan, Y., McLeod, M., Basu, N. B., Kumar, R., Rakovec, O., Samaniego, L., Attinger, S., Shrestha, N. K., Daggupati, P., Roy, T., Wi, S., Hunter, T., Craig, J. R. and Pietroniro, A.: Great Lakes Runoff Intercomparison Project Phase
520   3: Lake Erie (GRIP-E), J. Hydrol. Eng., 26(9), 05021020, doi:10.1061/(ASCE)HE.1943-5584.0002097, 2021.

Meyer, H. and Pebesma, E.: Predicting into unknown space? Estimating the area of applicability of spatial prediction models, Methods Ecol. Evol., 12(9), 1620–1633, doi:10.1111/2041-210X.13650, 2021.

Møller, A. B., Beucher, A., Iversen, B. V. and Greve, M. H.: Predicting artificially drained areas by means of a selective model ensemble, Geoderma, 320, 30–42, doi:10.1016/j.geoderma.2018.01.018, 2018.

525   Motarjemi, S. K., Møller, A. B., Plauborg, F. and Iversen, B. V.: Predicting national-scale tile drainage discharge in Denmark using machine learning algorithms, J. Hydrol. Reg. Stud., 36, 100839, doi:10.1016/j.ejrh.2021.100839, 2021.

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C. and Gupta, H. V.: What Role Does Hydrological Science Play in the Age of Machine Learning?, Water Resour. Res., 57(3), doi:10.1029/2020WR028091, 2021.

530   Nijzink, R. C., Samaniego, L., Mai, J., Kumar, R., Thober, S., Zink, M., Schäfer, D., Savenije, H. H. G. and Hrachowitz, M.: The importance of topography-controlled sub-grid process heterogeneity and semi-quantitative prior constraints in distributed hydrological models, Hydrol. Earth Syst. Sci., 20(3), 1151–1176, doi:10.5194/hess-20-1151-2016, 2016.

Noorduijn, S. L., Refsgaard, J. C., Petersen, R. J. and Højberg, A. L.: Downscaling a national hydrological model to subgrid

scale, J. Hydrol., 603, doi:10.1016/j.jhydrol.2021.126796, 2021.

535   Olesen, S. E.: Kortlægning af potentielt dræningsbehov på landbrugsarealer opdelt efter landskabselement, geologi, jordklasse, geologisk region samt høj/lavbund. [online] Available from: http://web.agrsci.dk/djfpublikation/djfpdf/intrma21.pdf, 2009.

Pasten-Zapata, E., Sonnenborg, T. O. and Refsgaard, J. C.: Climate change: Sources of uncertainty in precipitation and temperature projections for Denmark, GEUS Bull., 43, e2019430102-01, doi:10.34194/GEUSB-201943-01-02, 2019.

540   Read, J. S., Jia, X., Willard, J., Appling, A. P., Zwart, J. A., Oliver, S. K., Karpatne, A., Hansen, G. J. A., Hanson, P. C., Watkins, W., Steinbach, M. and Kumar, V.: Process-Guided Deep Learning Predictions of Lake Water Temperature, Water Resour. Res., 55(11), 9173–9190, doi:10.1029/2019WR024922, 2019.

Refsgaard, J. C., Sonnenborg, T. O., Butts, M. B., Christensen, J. H., Christensen, S., Drews, M., Jensen, K. H., Jørgensen, F., Jørgensen, L. F., Larsen, M. A. D., Rasmussen, S. H., Seaby, L. P., Seifert, D. and Vilhelmsen, T. N.: Climate change

545   impacts on groundwater hydrology – where are the main uncertainties and can they be reduced?, Hydrol. Sci. J., 61(13), 2312–2324, doi:10.1080/02626667.2015.1131899, 2016.

Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J. T., Beaudoing, H. K., Landerer, F. W. and Lo, M. H.: Emerging trends in global freshwater availability, Nature, 557(7707), 651–659, doi:10.1038/s41586-018-0123-1, 2018.

Rodriguez-Galiano, V., Mendes, M. P., Garcia-Soldado, M. J., Chica-Olmo, M. and Ribeiro, L.: Predictive modeling of

550   groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (Southern Spain), Sci. Total Environ., 476–477, 189–206, doi:10.1016/j.scitotenv.2014.01.001, 2014.

van Roosmalen, L., Christensen, B. S. B. and Sonnenborg, T. O.: Regional Differences in Climate Change Impacts on Groundwater and Stream Discharge in Denmark, Vadose Zo. J., 6(3), 554–571, doi:10.2136/vzj2006.0093, 2007.

555   Samaniego, L., Thober, S., Wanders, N., Pan, M., Rakovec, O., Sheffield, J., Wood, E. F., Prudhomme, C., Rees, G., Houghton-Carr, H., Fry, M., Smith, K., Watts, G., Hisdal, H., Estrela, T., Buontempo, C., Marx, A. and Kumar, R.: Hydrological Forecasts and Projections for Improved Decision-Making in the Water Sector in Europe, Bull. Am. Meteorol. Soc., 100(12), 2451–2472, doi:10.1175/BAMS-D-17-0274.1, 2019.

Scharling, M.: Klimagrid Danmark - Nedbør, lufttemperatur og potentiel fordampning 20X20 & 40x40 km -

560   Metodebeskrivelse, Danish Meteorological Institute. [online] Available from: https://www.dmi.dk/fileadmin/user_upload/Rapporter/TR/1999/tr99-12.pdf, 1999a.

Scharling, M.: Klimagrid Danmark Nedbør 10x10 km (ver. 2) - Metodebeskrivelse. [online] Available from: https://www.dmi.dk/fileadmin/user_upload/Rapporter/TR/1999/tr99-15.pdf, 1999b.

Schneider, R., Stisen, S. and Højberg, A. L.: Hunting for Information in Streamflow Signatures to Improve Modelled

565   Drainage, Water, 14(1), doi:10.3390/w14010110, 2022.

Soltani, M., Bjerre, E., Koch, J. and Stisen, S.: Integrating remote sensing data in optimization of a national water resources model to improve the spatial pattern performance of evapotranspiration, J. Hydrol., 603, 127026,

doi:10.1016/j.jhydrol.2021.127026, 2021.

Soylu, M. E. and Bras, R. L.: Global Shallow Groundwater Patterns from Soil Moisture Satellite Retrievals, IEEE J. Sel.
570   Top. Appl. Earth Obs. Remote Sens., 15, 89–101, doi:10.1109/JSTARS.2021.3124892, 2022.

Stisen, S., Sonnenborg, T. O., Højberg, A. L., Troldborg, L. and Refsgaard, J. C.: Evaluation of Climate Input Biases and
Water Balance Issues Using a Coupled Surface-Subsurface Model, Vadose Zo. J., 10(1), 37–53, doi:10.2136/vzj2010.0001,
2011.

Stisen, S., Ondracek, M., Troldborg, L., Schneider, R. J. M. and van Til, M. J.: National Vandressource Model -
575   Modelopstilling og kalibrering af DK-model 2019., 2019.

Sun, A. Y. and Tang, G.: Downscaling Satellite and Reanalysis Precipitation Products Using Attention-Based Deep
Convolutional Neural Nets, Front. Water, 2, doi:10.3389/frwa.2020.536743, 2020.

Tesoriero, A. J., Gronberg, J. A., Juckem, P. F., Miller, M. P. and Austin, B. P.: Predicting redox-sensitive contaminant
concentrations in groundwater using random forest classification, Water Resour. Res., 53(8), 7316–7331,
580   doi:10.1002/2016WR020197, 2017.

Thejll, P., Boberg, F., Schmith, T., Christiansen, B., Christensen, O. B., Madsen, M. S., Su, J., Andree, E., Olsen, S., Langen,
P. L., Skovgaard, K. M., Olesen, M., Pedersen, R. A. and Payne, M. R.: Methods used in the Danish Climate Atlas., 2021.

Tran, H., Leonarduzzi, E., De la Fuente, L., Hull, R. B., Bansal, V., Chennault, C., Gentine, P., Melchior, P., Condon, L. E.
and Maxwell, R. M.: Development of a Deep Learning Emulator for a Distributed Groundwater–Surface Water Model:
585   ParFlow-ML, Water, 13(23), 3393, doi:10.3390/w13233393, 2021.

Tyralis, H., Papacharalampous, G. and Langousis, A.: A Brief Review of Random Forests for Water Scientists and
Practitioners and Their Recent History in Water Resources, Water, 11(5), 910, doi:10.3390/w11050910, 2019.

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V.,
Lamarque, J.-F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J. and Rose, S. K.: The representative
590   concentration pathways: an overview, Clim. Change, 109, 5–31, doi:10.1007/s10584-011-0148-z, 2011.

Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., de Roo, A., Döll, P., Ek, M.,
Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-
Lidard, C., Sivapalan, M., Sheffield, J., Wade, A. and Whitehead, P.: Hyperresolution global land surface modeling: Meeting
a grand challenge for monitoring Earth's terrestrial water, Water Resour. Res., 47(5), 1–10, doi:10.1029/2010WR010090,
595   2011.

Wunsch, A., Liesch, T. and Broda, S.: Deep learning shows declining groundwater levels in Germany until 2100 due to
climate change, Nat. Commun., 13(1), 1–13, doi:10.1038/s41467-022-28770-2, 2022.

Yang, Y., Cao, C., Pan, X., Li, X. and Zhu, X.: Downscaling Land Surface Temperature in an Arid Area by Using Multiple
Remote Sensing Indices with Random Forest Regression, Remote Sens., doi:10.3390/rs9080789, 2017.

600   Zhang, J., Liu, K. and Wang, M.: Downscaling Groundwater Storage Data in China to a 1-km Resolution Using Machine
Learning Methods, Remote Sens., 13(3), 523, doi:10.3390/rs13030523, 2021.

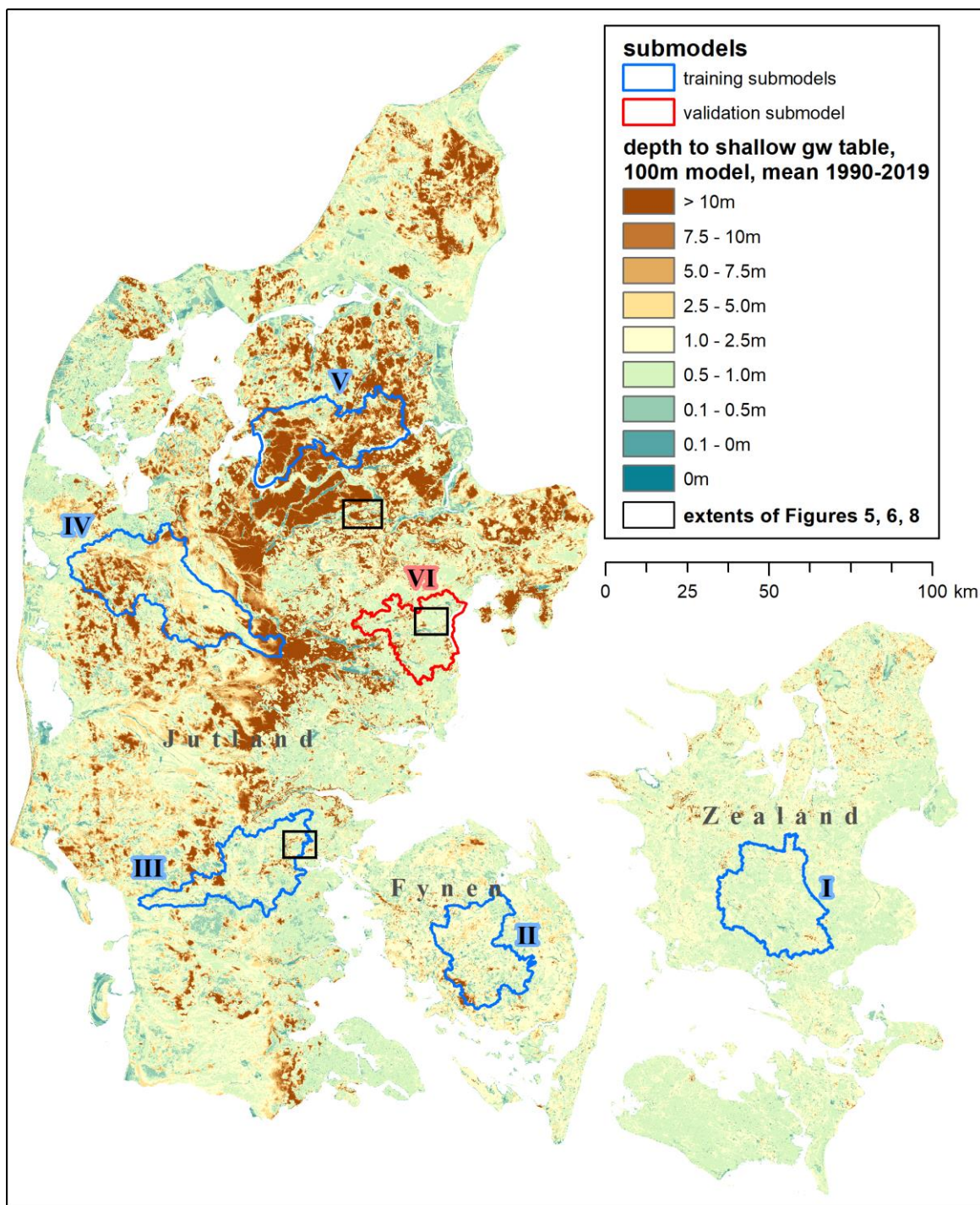



Figure 1: Denmark with the mean depth to the shallow groundwater table, as simulated by the DK-model HIP in 100 m for the period 1990 – 2019. The submodels used for training and validation of the downscaling algorithm are outlined in blue and red. The extents of the detail maps shown in Figures 5, 6, and 8 are indicated as black rectangles.

605

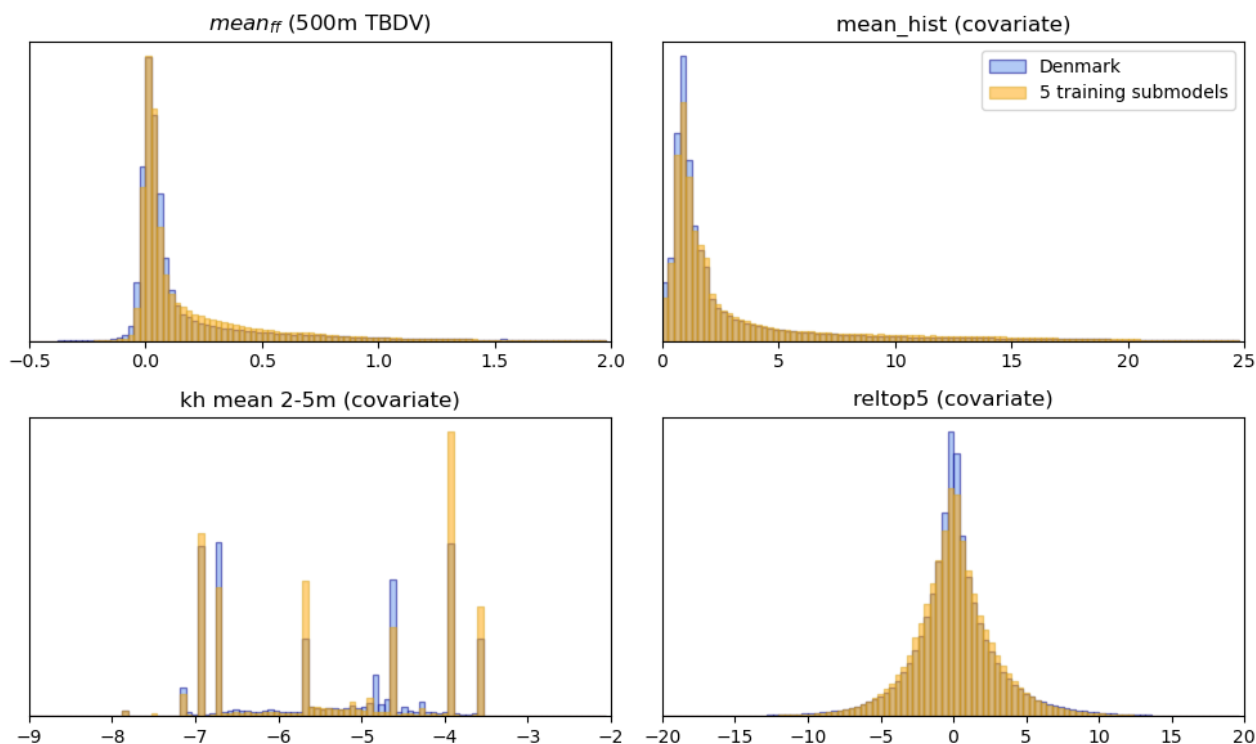



Figure 2: Histograms of covariate ranges covered by the five training submodels, compared to all of Denmark. Shown for the example of TBDV $mean_{ff}$. Top left: Change in mean depth to shallow groundwater from the 500 m model (i.e. the TBDV) [m]. Top right: Mean depth to shallow groundwater from the historic 100 m model (the respective most important covariate) [m]. Bottom left and right: kh mean 2-5m ([m/s], log-transformed) and reltop5 [m] as two more covariates.

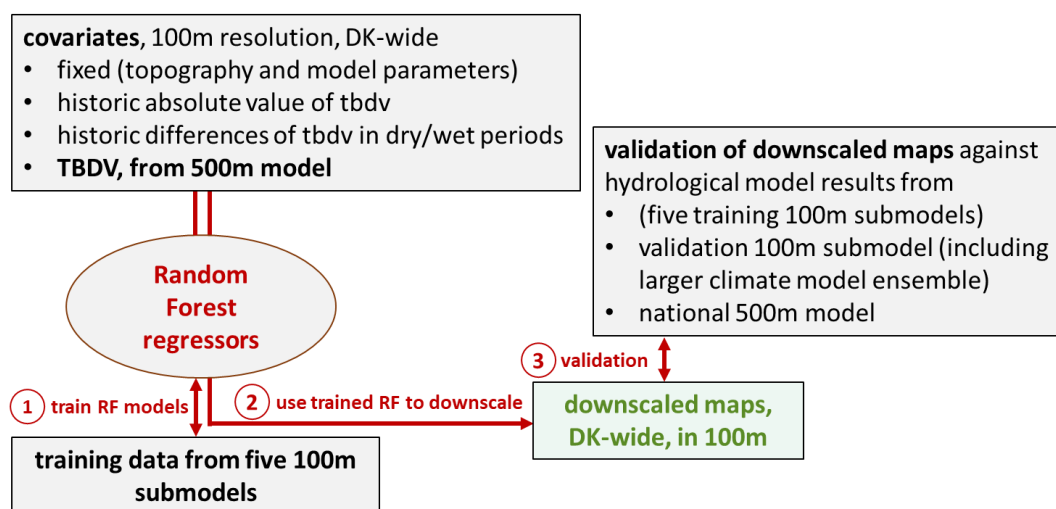



Figure 3: Diagram of the downscaling algorithm.

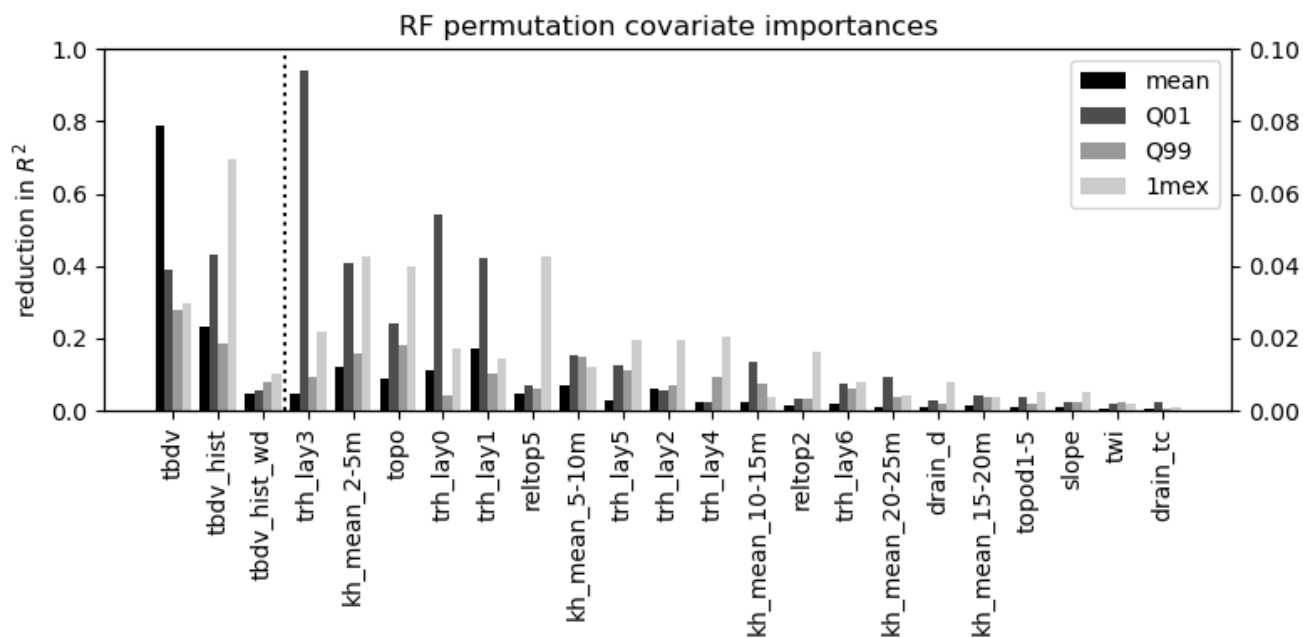



**Figure 4: Permutation covariate importance for the four RF downscaling regressors, for each covariate (see Table 2 for an
615 overview). The higher the reduction in R^2, the higher the importance of the covariate. Note that the three covariates to the left of
the dashed vertical line refer to the left y-axis, whilst the remainder of covariates refer to the right y-axis.**

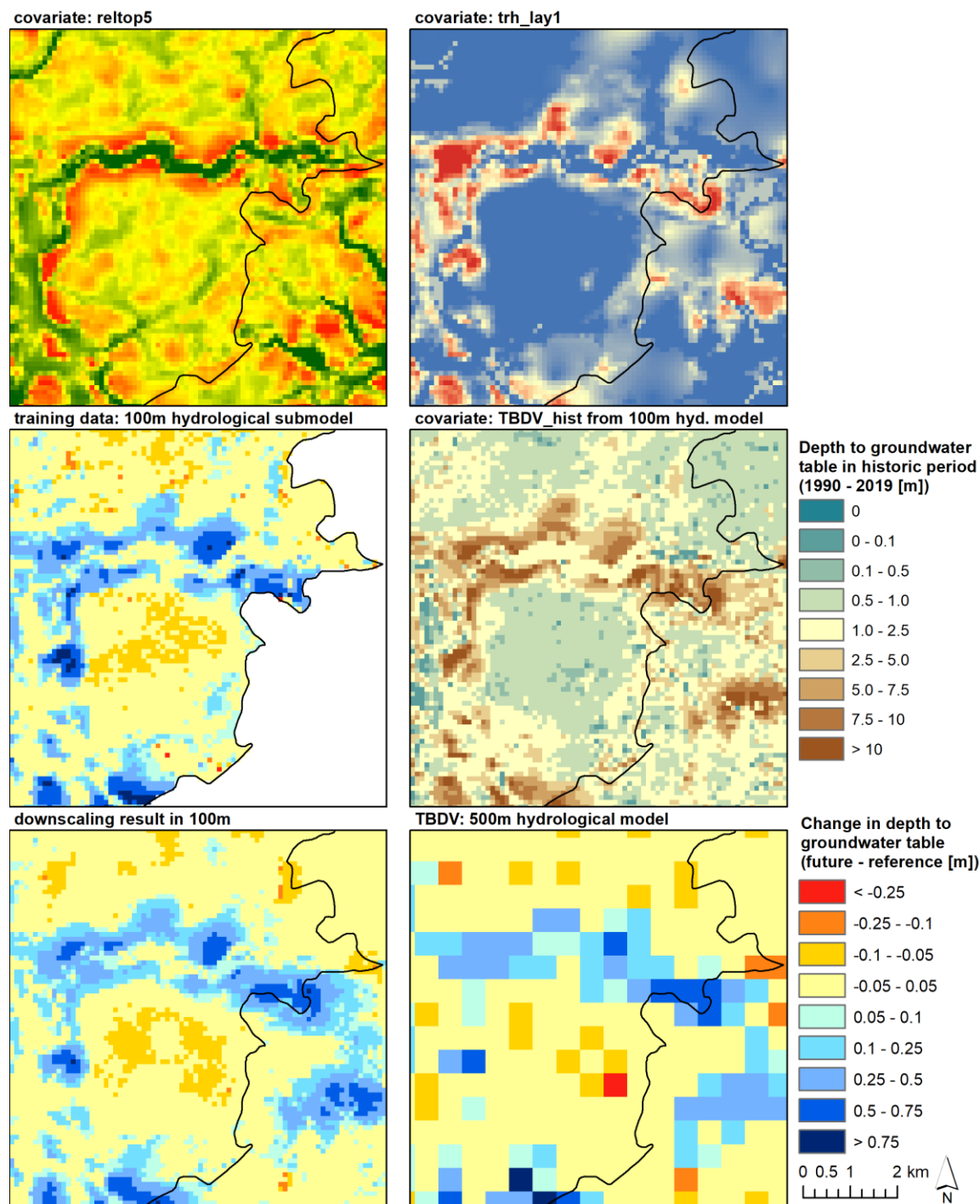



Figure 5: Example of downscaling results for changes to the mean depth to the groundwater table for the far future period (*mean_ff*). **All panels show the same extent. Top left and right: covariates reltop5 and trh_lay1. Centre left: 100 m HM submodel**
620   **results (training data). Centre right: 100 m historic absolute values (covariate). Bottom left: 100 m downscaling result. Bottom right: 500 m HM results (TBDV).**

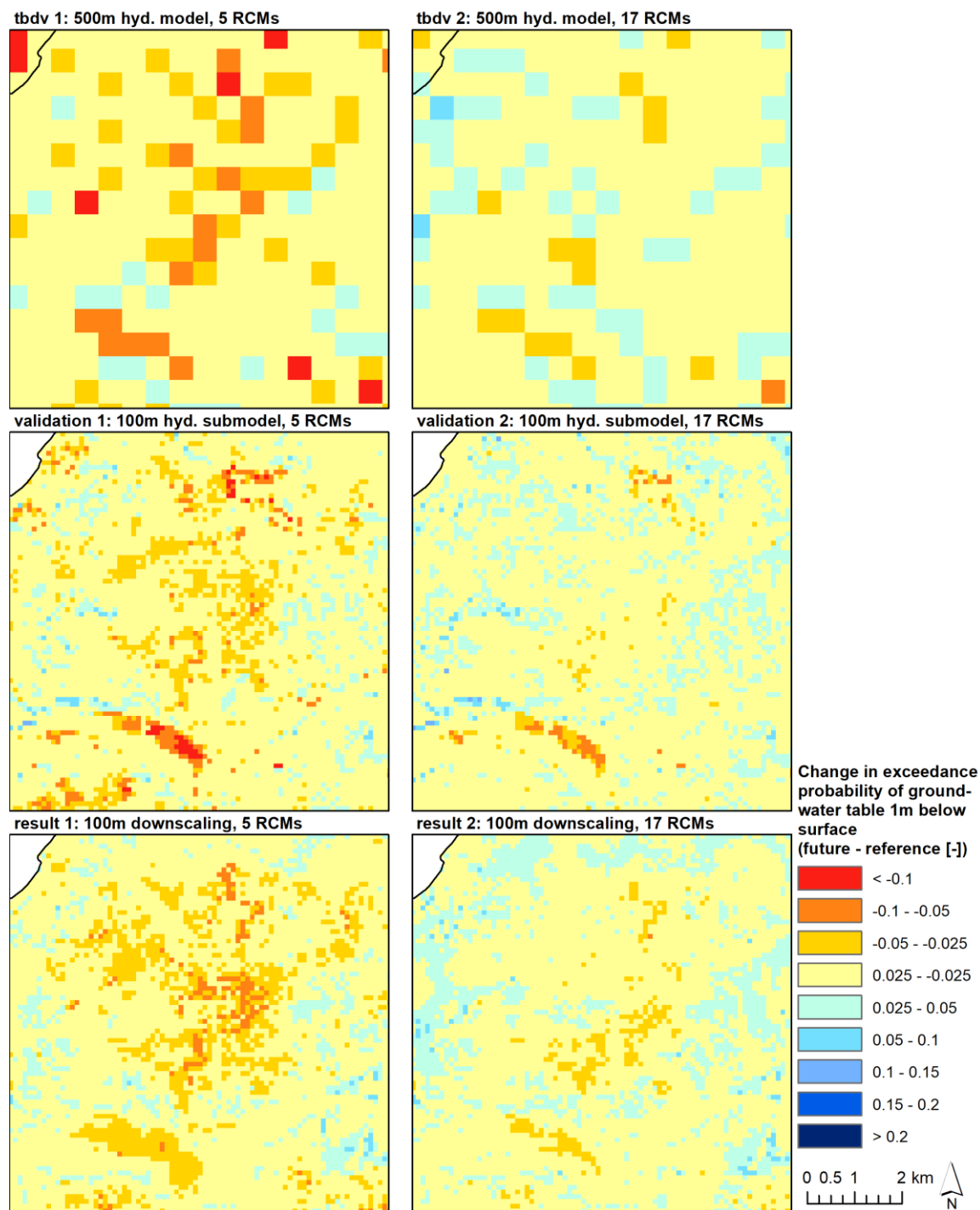
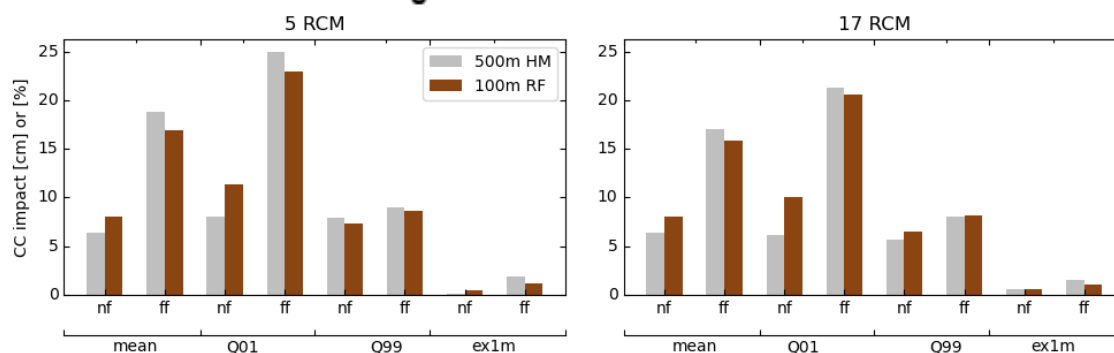



Figure 6: Downscaling outputs for validation submodel VI, for the example of *ex1m_{nf}*. The RF downscaling was trained to downscale the median change of the five RCM from the 500 m HM (top left) with the 100 m HM training data (centre left; however, submodel VI is not included in the training). Results are shown in the bottom left. If applied for a different RCM ensemble (top right), the results (bottom right) also align with the truth (centre right).

Denmark-wide evaluation against 500 m HM results

Evaluation against 100 m HM for validation-subcatchment VI

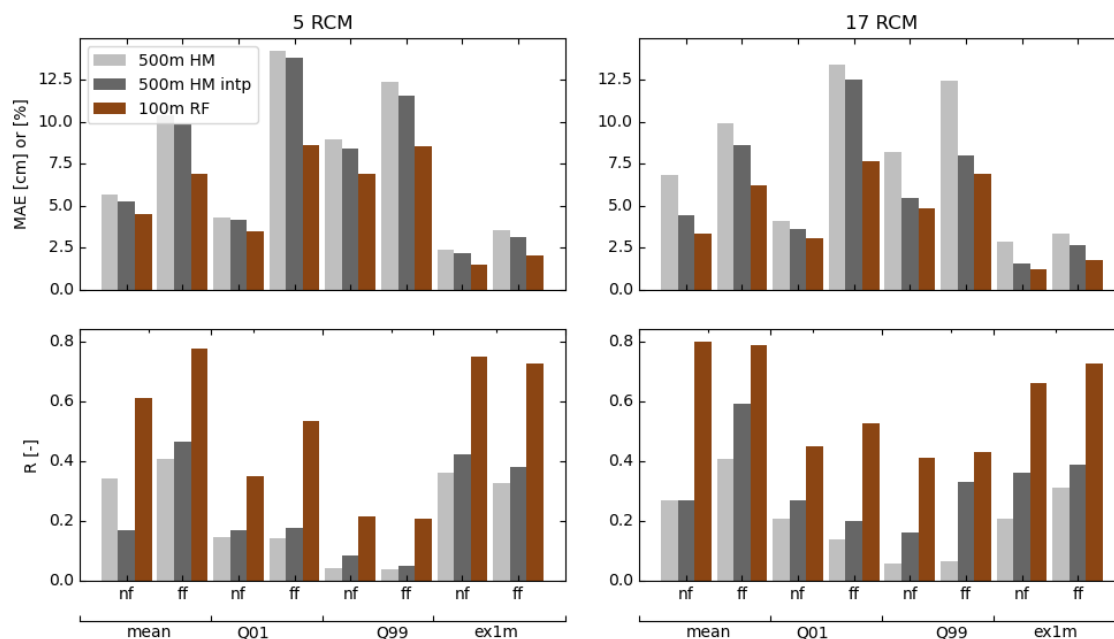



Figure 7: Overview over the validation results of the RF downscaling algorithm. 1st row: Climate change impact on the TBDV. 2nd and 3rd row: MAE and Pearson's R between the 100 m HM from the validation submodel and the corresponding 500 m HM and 100 m RF results. Values along the y-axis are given in cm for *mean*, *Q01* and *Q99*, and in % for *ex1m*.

630

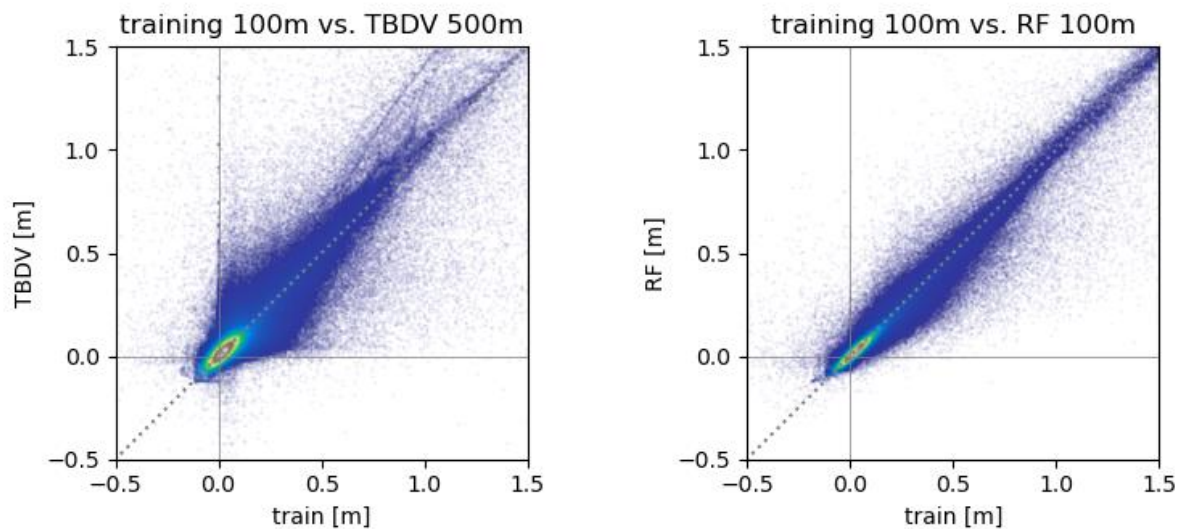



Figure 8: Scatterplots, for the example of TBDV $mean_{ff}$: Comparing the training data (i.e. the TBDV in 100 m) with the actual TBDV in 500 m (left panel), and with the RF downscaling results in 100 m (right panel) for the five training submodels.

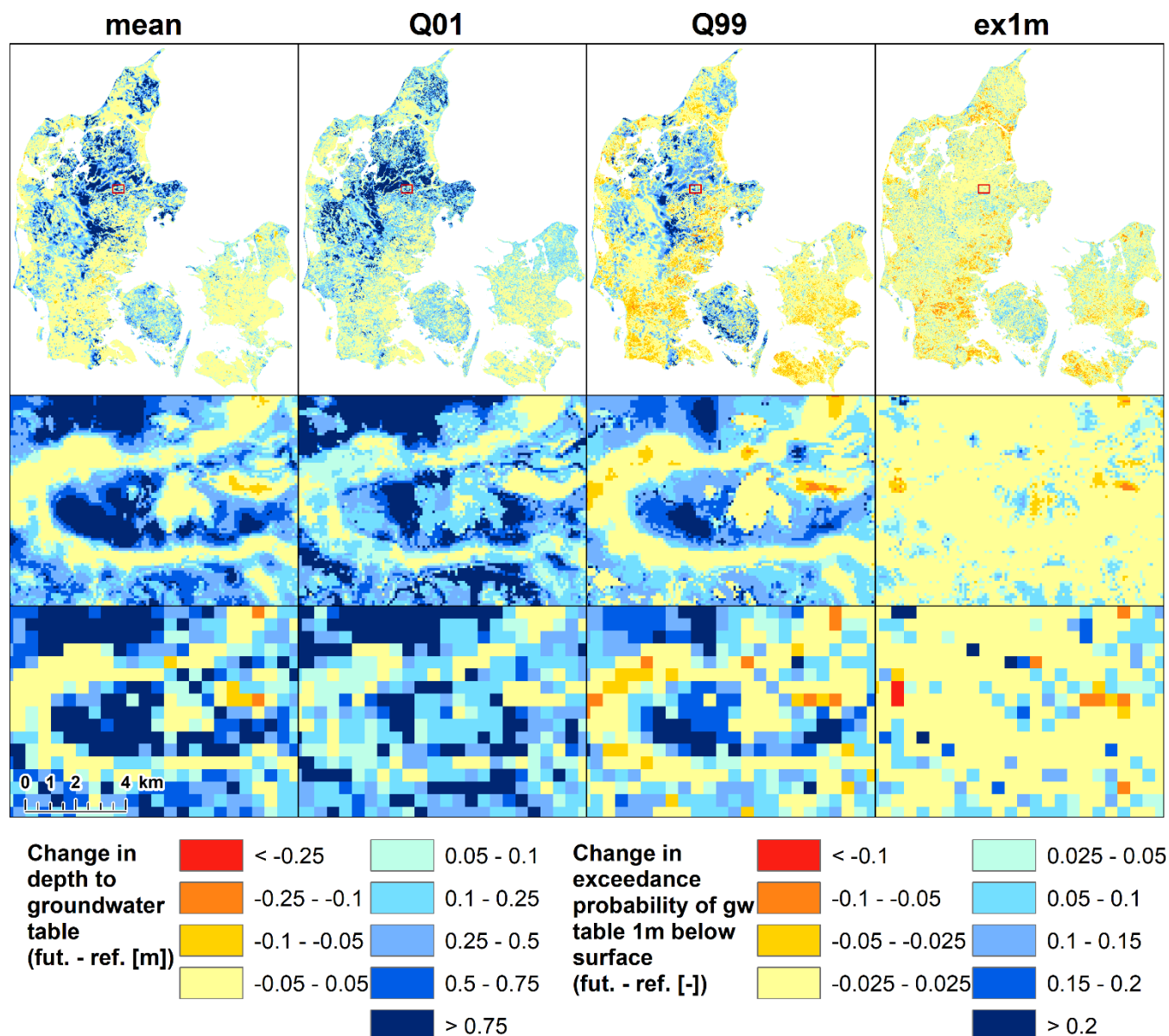



Figure 9: Results of the downscaling, for the far future of all four TBDV. Top row: Downscaling results in 100 m for all of Denmark. Middle row: Detail of the 100 m downscaling results (extent indicated in top row). Bottom row: Same detail for the corresponding TBDV in 500 m.

Table 1: Overview over the performed HM runs with different climate inputs for Denmark and different submodels. A: Used as covariate. B: To be downscaled, in training. C: To be downscaled, in validation. D: Training data. E: Validation data. *: Not directly used in this study.

640

		Denmark, 500 m	Denmark, 100 m	5 training submodels I – V, 100 m	validation submodel VI, 100 m
historic climate (1990 – 2019)		*	A		
17 RCMs (full ensemble)	reference period (1991 – 2020)	C			E
	near future (2041 – 2070)	C			E
	far future (2071 – 2100)	C			E
5 selected training RCMs	reference period (1991 – 2020)	B		D	E
	near future (2041 – 2070)	B		D	E
	far future (2071 – 2100)	B		D	E

Table 2: Overview over the covariates.

topography-related covariates

topo	Elevation	[mamsl]
topod1-5	Difference between elevation in 500 m and 100 m resolution	[m]
TWI	Topographic Wetness Index (Beven and Kirkby, 1979)	[-]
slope	Slope of terrain	[-]
reltop2	Relative topography within 200 m radius	[m]
reltop5	Relative topography within 500 m radius	[m]

model parameter-related covariates (from DK-model HIP in 100 m)

drain_d	Drain depth	[m]
drain_tc	Drain time constant	[s^{-1}]
trh_lay$<x>$	Horizonal transmissivity in computational layers 1 to 7	[$m^2\ s^{-1}$]
kh_mean_$<x>$	Geometric mean of horizontal hydraulic conductivity in five depth intervals: 2-5 m, 5-10 m, 10-15 m, 15-20 m, 20-25 m	[$m^2\ s^{-1}$]

Covariates related to TBDV

TBDV	"to-be-downscaled-variable", i.e. change in depth to shallow groundwater table. From 500 m DK-model HIP (bilinear interpolation to 100 m). B and C in Table 1. TBDV is one of	
	• $mean_{nf}$, $mean_{ff}$ (mean groundwater table)	[m]
	• $Q01_{nf}$, $Q01_{ff}$ (high groundwater table)	[m]
	• $Q99_{nf}$, $Q99_{ff}$ (low groundwater table)	[m]
	• $ex1m_{nf}$, $ex1m_{ff}$ (1 m exceedance probability)	[-]
TBDV_hist	Absolute value of TBDV for the historic period from DK-model HIP 100 m (e.g. the mean depth to the shallow groundwater table). A in Table 1.	
TBDV_hist_wd	Difference in TBDV between a wet and dry period (dry: 1990 – 2001, wet: 2004 – 2015) from the historic DK-model HIP 100 m (e.g. the mean depth to shallow groundwater table in the wet period minus the mean to shallow groundwater table in the dry period). A in Table 1.	

645

Table 3: The parameters used for the RF regressors. The first section presents the tested hyperparameters, where the selected values for each parameter are marked in bold. Further parameters are reported in the second section.

Name	Description	Values (**optimal**)
RF parameters tested in hyperparameter search		
max_depth	Maximum depth of tree. If "all", nodes are expanded until leaves are pure or all leaves contain less than min_samples_split samples	**all**, 30
max_features	Number of features (covariates) to consider when looking for the best split	n_features, **0.5*n_features**, sqrt(n_features)
min_samples_split	Minimum number of samples required to split an internal node	**2**, 5, 25
min_samples_leaf	Minimum number of samples required to be a leaf node	1, 3, 10, **50**
bootstrap	Use bootstrap samples to build trees	**True**, False
Further RF parameters		
n_estimators	Number of trees in the forest	200
criterion	Function to measure the quality of a split	mse

650