

Interactive comment on “Modelling of the shallow water table at high spatial resolution using Random Forests” by Julian Koch et al.

Anders Bjørn Møller (Referee)

anbm@agro.au.dk

Received and published: 11 June 2019

General comments

I have read the manuscript “Modelling of the shallow water table at high spatial resolution using Random Forests” submitted to HESS by Koch et al., in order to provide a referee comment.

The manuscript is well structured, clear, concise and well written. It addresses the depth to the shallow water table, which is a highly relevant issue, and introduces a number of novel methods in doing so. Some parts of the introduced methods have great potential, not only for hydrological applications but for spatial predictions with machine learning in general.

C1

My main concerns with the manuscript lie with some of the specific choices that the authors make in implementing the methods, especially related to the assessment of the accuracy and uncertainties of the predictions. I will elaborate on these concerns in the following section. However, given that the authors address them, the manuscript is highly suitable for publication in HESS.

Specific comments Firstly, I am wondering why the authors choose to map the depth to the shallow water table rather than the elevation of the shallow water table. I would expect the elevation of the shallow water table to show less spatial variation than the depth from the surface. It should therefore be easier to predict, all other things equal. I am sure the authors have good reasons for this choice, but I would like to see them stated explicitly.

Secondly, I would like to comment on the use of a sine function to model an annual minimum event. I think it is a useful and generally robust way to address the issue of working with limited data. However, the method could be improved upon in a number of ways. Firstly, the maximum of the curve does not match the maximal observed water levels. The authors could therefore have calculated the uncertainty related to the sine model and, ideally, used these uncertainties in the Random Forest model. The authors already state this in the manuscript, but my second comment is related to the same issue. For training locations with sparse data, the authors used the maximum of the sine curve, but for training locations with more observations, the authors used observed maximum water levels. This choice muddles the results, both in terms of the predicted values and their accuracies. Is it a map of the expected minimum depth to the shallow water table, averaged over a number of years? Or is it a map of an extreme event, observed only in some years? The mixture of training data makes this question difficult to answer.

Thirdly, I have concerns about the way that the authors assess the accuracy of the predictions. The training dataset shows a high degree of clustering. Therefore, when the authors use the out-of-bag predictions for assessing the accuracy, the points used for

C2

assessing the accuracy will be located close to the training points used for making the predictions. It is very likely that the values are spatially autocorrelated, and the stated accuracy is therefore probably not representative for the study area as a whole. I would expect the accuracy to be lower for the parts of the study area that do not have a high density of observations. A spatially structured accuracy assessment, as proposed for example by Muscarella et al. (2014), would most likely provide a more representative accuracy assessment. Furthermore, I am wondering if the authors used all the training points for the predictions. The training dataset contained both groundwater and surface water observations. However, the aim is not to predict surface water levels, and I would therefore say that one could justify removing the surface water points from the out-of-bag predictions when assessing the accuracy.

Fourthly, I very much like the way that the authors handle covariate importance. Being able to assess covariate importance in geographic space is potentially extremely useful, for both researchers and end users. However, I do not think that decrease in R^2 is the best measure of covariate importance. One can potentially obtain a high R^2 , even if the absolute values are inaccurate. A better choice would therefore be to assess the relative change in a measure that accounts for absolute values, such as RMSE, Lin's concordance or the Nash-Sutcliffe efficiency.

Fifthly, while I appreciate that the authors assessed the uncertainties of the predictions in two different ways, I do not think that combining them is justified. The theoretical basis for the approach seems scarce. Both the RF uncertainties and the residuals used for kriging relate to the same model, and it is therefore a stretch to say that they are independent. Furthermore, quantile regression forest should be able to assess uncertainties quite accurately without any further elaboration, as shown for example by Rudiyanto et al. (2018). I think a large part of the spatial autocorrelation in the residuals would disappear, if one takes into account the uncertainties related to the RF predictions. The uncertainties in the predictions make the residuals uncertain as well, which complicates regression-kriging. When experimenting with techniques, as

C3

the authors do, it is important to set aside an independent part of the dataset to be able to assess the accuracy of the estimated uncertainties. However, the authors do not do this, and it is therefore impossible to assess if the error propagation actually leads to a better estimate of the uncertainties. Unless the authors can adequately address these shortcomings, the section on error propagation should be removed. I am also wondering why the authors used the out-of-bag residuals and not the residuals from the actual RF predictions. I have not seen any other studies using out-of-bag residuals for regression-kriging, and the authors should elaborate on their reasons for this choice.

Sixthly, the authors use the hydrological DK-model as a covariate in the random forest model. I am wondering if the training points used in the RF model were also used for calibrating the DK-model. If this is the case, it creates a risk of circular logic, as the covariate contains information on the target variable at the location of the training points.

Seventh, the authors state that the sine model used to estimate extreme events could be replaced by an updated version of the DK-model. While I agree that this would improve the estimate of extreme events, it would also introduce another potential source of circular logic, if the DK-model was still used as a covariate. The approach would therefore need to be implemented with great care in order to avoid this.

Lastly, I would like to comment on the use of the term "validation" for accuracy assessment. This is a general concern with the literature as much as a comment on this manuscript in particular. Oreskes (1998) argues that a quantitative model of a complex natural system cannot be considered as truly "validated" until it is used. For example, a conceptually flawed model can still provide good accuracies. The issue becomes even more relevant for machine learning models, where the parameters represent only patterns in the data, not physical processes. Strictly speaking, a machine-learning model can therefore never be truly valid, although it may be accurate and useful. To emphasize this point, I will mention Fourcade et al. (2018), who accurately mapped species distributions with entirely nonsensical covariates. I will encourage the authors

C4

to consider these points when discussing the accuracy of the predictions.

Technical corrections and stylistic suggestions

Generally, the authors refer to “traditional physically-based modelling” several times in the manuscript. I think “conventional” would be more adequate than “traditional”, as science has conventions, not traditions. Tradition is a cultural phenomenon. Indeed, in most cases both “conventional” and “traditional” are redundant, as “physically-based modelling” accurately describes what the authors refer to, without any further need of clarification.

Page 2:

L5: “There exists a broad relevancy of the shallow groundwater” → “The shallow groundwater has a broad relevance”

L9 – L10: “energy partitioning” → “energy balance”

L13: “The shallow groundwater is also of importance in the urban context” → “The shallow groundwater is also important in urban contexts”

L19: “a 100 year event with respect to today’s average conditions” → “a 100-year event relative to present average conditions”

L21: “high permeable” → “highly permeable”

L28: “which hinders to conduct thorough calibration, sensitivity and uncertainty analysis at high resolution” → “which hinders thorough calibration, and sensitivity and uncertainty analyses at high resolution”

L29: “Further, there exists a general difficulty to parameterize subsurface processes regardless the scale” → “Furthermore, it is difficult to parameterize subsurface processes regardless of the scale”

Page 3: L3: “Hydrology” → “hydrology”

C5

L16: “mode” → “model”

L16: “Before machine learning techniques can build the toolbox of future’s environmental decision making” → “Before machine learning techniques can be considered as a toolbox for environmental decision making”

L25: “Opposed” → “However”

L29: “or” → “and”

Page 4:

L3: “The study area encompasses a large part of the Danish peninsular, which is located in Northern Europe (54.5–57.8°N and 8.0–10.9°E) and referred to as Jutland.” → “The study area encompasses a large part of the Jutland peninsula, located in Denmark in northern Europe (54.5–57.8°N; 8.0–10.9°E).”

L5: “the sequence” → “a sequence”

L6 – L8: The clay contents in eastern Denmark are not very high (10 – 20% for the topsoil). They are higher than the clay contents in western Denmark, but not relative to other areas in the world. It would be more accurate to say that the texture is loamy or that the clay contents are moderately high.

L8: “Weichselian sandy outwash plains” → “sandy Weichselian outwash plains”

Page 5:

L6: “well data [...] was” → “well data [...] were”

Page 6:

L6: “coast” → “the coastline”. This should be the case throughout the manuscript. Also “coastline” → “the coastline”.

Page 8:

C6

Table 1: Lowland classification and landscape typology should refer to Madsen et al. (1992).

Table 1: "Drain probability" → "Probability of artificial drainage"; "Drain class" → "Soil drainage class".

Page 9:

L13: Bootstrap samples on average contain 63.2% of the data, not 66%.

L25: "The concept of covariate permutation allows to assess the importance of each covariate" → "Covariate permutation allows an assessment of the importance of each covariate"

Page 12:

L20: "visual" → "visible"

Page 13:

L2 – L3: Delete "was evident".

Page 17:

L21: "clear a shortcoming" → "a clear shortcoming"

Page 19:

L3: "that region" → "the study area"

Page 20:

L14: "allows to model" → "enables"

References

Fourcade, Y., Besnard, A.G. and Secondi, J., 2018. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statis-

C7

tics. *Glob. Ecol. Biogeogr.* 27(2), 245-256. <http://dx.doi.org/10.1111/geb.12684>.

Madsen, H.B., Nørr, A.H. and Holst, K.A., 1992. The Danish soil classification. The Royal Danish Geographical Society, Copenhagen, Denmark.

Muscarella, R., Galante, P.J., Soley-Guardia, M., Boria, R.A., Kass, J.M., Uriarte, M., Anderson, R.P. and McPherson, J., 2014. ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods Ecol. Evol.* 5(11), 1198-1205. <http://dx.doi.org/10.1111/2041-210x.12261>.

Oreskes, N., 1998. Evaluation (not validation) of quantitative models. *Environ. Health Perspect.* 106(Suppl 6), 1453-1460. <http://dx.doi.org/10.1289/ehp.98106s61453>.

Rudiyanto, Minasny, B., Setiawan, B.I., Saptomo, S.K. and McBratney, A.B., 2018. Open digital mapping as a cost-effective method for mapping peat thickness and assessing the carbon stock of tropical peatlands. *Geoderma* 313, 25-40. <http://dx.doi.org/10.1016/j.geoderma.2017.10.018>.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2019-212>, 2019.

C8