Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/hess-2022-258-RC1, 2022

Comment on hess-2022-258

Scott Wilson (Referee)

Referee comment on "Estimation of groundwater age distributions from hydrochemistry: comparison of two metamodelling algorithms in the Heretaunga Plains aquifer system, New Zealand" by Conny Tschritter et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2022-258-RC1, 2022

This paper makes an excellent and novel contribution to predicting transit times using a wider dataset than isotopic age tracers. I found the text enjoyable and easy to read, and the perspective is fairly balanced. I have two main comments on the approach taken, which have some bearing on the conclusions that can be derived from this work.

The main drawback of the approach taken is that the chemically based models use the lumped model age estimates as a response variable, ie training a model on another model which is acknowledged as having shortcomings, although this is the motivation for the paper. The difficulty is that this creates an ambiguity as to whether mismatches in the trained models are due to poor lumped model estimates, or poor local performance of the trained models, or both, or something else (eg parameter or model selection). As a suggestion, an alternative or complementary approach would be to firstly train models to predict the isotopic tracer con\centration. This would provide some prior information on the mismatch between the lumped model predictions, and ensemble predictions. This approach could perhaps inform how the lumped parameter estimates could be improved, which was suggested in 553.This is not a step necessary for this paper, but perhaps something that could be carried out in future work.

The modelling approach applied here is to generate a global model from a subset of individual models, and it is assumed that the input data are spatially and temporally independent. However, the hydrochemical clustering results do indicate that the chemistry data have a predictable spatial and temporal variability. Some evidence of this infleunce is apparent in the results (eg Fig 6, T2_34), and hence in the applications section some spatial and temporal discrepancies are acknowledged. To overcome this, it may have been beneficial to train models on each hydrochemical cluster, although it has to be acknowledged that there is little data available for clusters 4 and 5. An alternative approach would be to introduce some additional predictive parameters in the model to account for spatial and temporal variability, eg elevation, depth, position, hydrochemical cluster. Some of these parameters have been used for validation, but they could also have been training parameters, or tested to see if they do inform model predictions. In doing

so, one could have more confidence in the application of the models to areas with no age data.

The paper would also benefit from some corections and the clarification of some points listed below.

Title and line 60: SR and GBR methods are not metamodels per se. They are applied in this paper as metamodels because they are trained on the LPMs rather than raw observation data

97: Should be Heretaunga Plain not Plains (also elsewhere)

125: There are red lines on Fig 1 which are unreferenced. Are these flow barriers? It seems odd that there is a flow path towards a flow barrier (centre top)

154: The clustering detailed in the hydrochemistry section provides some background context, but is not used in the modelling or subsequent analysis.

219: It's good practice to state that this is the response variable for the statistical modelling, and the hydrochemistry data are the predictor variables

247-249: How much error is the distance to these input signal datasets likely to introduce to the age estimates, and how would that compare to the error introduced by the EPM?

269: Were not was

273: The primary aim of tuning is to improve model performance, not assist convergence

278: The terms 'chained' and 'unchained' models is unorthodox, and perhaps not an apt description of what the models represent. Perhaps these would be better referred to as 'independent' (see line 276) or 'individual models', and the chained models as 'ensemble models'

286: Why do the train/test splits differ for the two models? This approach doesn't enable a

clear comparison of modelling performance between the two models to be made

286: As a comment, a 10/90 split is quite heavy-handed and could lead to overfitting. The unchained GBR R2 values are very high, although this is also true for the SR R2 values

290: There seems to be an error in the Pearson formulas

375: Last Glacial (is a noun)

399: The third value is 1.7 (ie >1)

405: Perhaps the models could achieve good age distributions with substantially less parameters?

410: It might have been more informative to plot the cluster results here rather than the ensemble weights, since the most informative parameters are already described in the text. As a reader, I'm intrigued by the relationship between the model performance and the clusters.

434: Perhaps water chemistry has some influence of the source rock, which wouldn't necessarily be reflected in the age estimates

517: It's ambiguous how these parameters were treated. Were their values set to the detection limit?

522: I think this claim is a bit of a stretch since there are no spatial aspects to this study. The model is aspatial, and global, and appears to generalise well to most, but not all the data. The model has the potential to be applied to other areas with confidence if the successful or unsuccessful predictions could be identified as having an association with something eg a particular cluster. NB this comment also applies to the last sentence of the abstract.

543-547: I don't think these statements are valid, particularly in light of the preceding sentences. There is no spatial aspect to the modelling to this modelling approach, it only uses age and chemistry data.

578: Which of these models would you have the most confidence to apply elsewhere?