

## ***Interactive comment on “An automated method to build groundwater model hydrostratigraphy from airborne electromagnetic data and lithological borehole logs” by P. A. Marker et al.***

**P. A. Marker et al.**

paam@env.dtu.dk

Received and published: 16 April 2015

First, the authors would like to thank Jan Gunnink for the comments on the paper. Referee comments will be presented in quotes.

“One issue that needs clarification is the fact that the airborne EM data is used in the inversion into resistivity, next the resistivity in an inversion together with borehole data in the Clay Fraction model and finally the clustering method uses both resistivity and Clay Fraction (which is again a derivative of the resistivity). This repetitive use of the airborne EM data is of course very clever, because the resistivity datasets provides a

C1076

dense 3D coverage, but it also needs some more explanation of the potential pitfalls. There is the danger of using the same data more than once that errors that are present in the data (as there always are) propagate along several ways into the end product.”

We basically start with two different and independent datasets: Resistivity from AEM and borehole lithology. Resistivity and borehole lithology are subsequently combined into the clay fraction model. However, resistivity variations are only partly due to clay content, so there is extra information in the resistivity which we want to retain for the clustering. Using the principal component analysis, we split the resistivity and clay fraction dataset into a correlated and an uncorrelated component and these two then go into the clustering algorithm. The clustering is not a data processing step but a spatial aggregation step. We agree that formal error propagation into the final hydrostratigraphic model is not performed in this paper, this is the subject of future work. The discussion will be amended in the revised manuscript.

“Also, by using the same dataset in different - but related - methods, one might end up in a “chicken-egg” situation. Is the result achieved by processing the same dataset sequentially using different methods really contributing to an improved end-product.”

This question is not easy to answer as we do not know the true hydrostratigraphy of the subsurface. The rationale in this paper is to let the hydrologic calibration decide whether or not the workflow results in a useful subsurface model in terms of being able to match hydrologic observations. We find that our approach is equally good or slightly better than the available benchmarks in that respect.

“Especially the fact that the combination of resistivity and borehole data is both used in the Clay Fraction model and in the clustering needs some better explanation.”

The clustering is performed in a 2-dimensional attribute space. Each subsurface voxel has two attributes, clay fraction and resistivity. Please remember that we start with two independent datasets (resistivity and borehole lithology), so the use of two attribute dimensions is entirely justified. Correlations between resistivity and clay fraction are

C1077

handled with the principal component analysis.

“Are we looking at real improvements or is it merely presenting the correlated datasets in different ways? For example, why not using a simple cut-off in the Clay Fraction model to derive clusters in the data? Because the resistivity is already used in making the Clay Fraction model, I would think that the result might not be very different. I understand that the Principal Component Analysis is crucial in the clustering method, but I am not sure that it is really contributing to a better end product. I am not saying that the cluster method is not correct to use here, but I would like to have a better understanding of the implications of the repetitive use of the same datasets.”

Yes, the CF-model alone could be used to do the clustering. However, the resistivity dataset contains signals that are not related to clay content (e.g. groundwater quality) and those should be retained in the clustering. The underlying assumption is that spatial patterns of resistivity/clay fraction are related to spatial patterns of hydraulic conductivity. We agree that empirical evidence for this hypothesis is required, which is the subject of future research.

“Another issue with the clustering method is the fact that k-means clustering aims at producing clusters of approx. the same size. This might have an unwanted influence on the results of the clustering, since there seems no reason to assume that the clusters in the hydrostratigraphic model needs to be of equal size.”

Yes. However the clusters found for our data are not of equal size. For example, for the 5-cluster model, clusters represent 9%, 26%, 9%, 12%, and 44% respectively of the total number of voxels. The behaviour of the k-means clustering algorithm is controlled by the chosen distance measure. Here, we used the standard Euclidean distance measure. There is scope to experiment with different distance measures in the clustering in the future.

“In the paper, another model is used as a “benchmark” or a so called reference model. This model I would not call a benchmark model, because it is also just a model! Al-

C1078

though there is some geological knowledge inserted into this model, this knowledge is later collapsed into only four hydrological units. Keep in mind the phrase: “all models are wrong, some are useful”. The results of the research in this paper indicates that for the purpose of hydrological modeling, the presented method shows better results in terms of deviation from measured groundwater heads, compared to another model.”

The reference model represent the hydrofacies as they were originally modelled in the integrated hydrological model. The reference model approach was chosen to have a baseline hydrological model performance for benchmarking. All four reviewers agree that the inclusion of the reference model has little added value and is partly confusing. We will therefore consider removing the discussion of the reference model from this paper.

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 12, 1555, 2015.

C1079