HESS Discussions, 10, 15219–15262, 2013

Challenges in conditioning a stochastic geological model of a heterogeneous glacial aquifer to a comprehensive soft dataset

Correspondence to: J. Koch (juko@geus.dk)

Author comment to the Editor: Prof. Mauro Giudici.

We would like to thank Mauro Giudici for managing the revision process of our manuscript. In total, three reviews were received and we strongly believe that revising the manuscript following the reviewers concerns will improve the scientific quality of our study. We are glad that Mauro Giudici acknowledges that most of the concerns from the first revision are answered by our first reply. In his comment, Mauro Giudici identifies a few main concerns from the first revision and beyond and we will gladly address them thoroughly and correct the manuscript accordingly. We will upload a final version of our manuscript and hope that it is then ready for publication.

> *The Authors' reply to the Reviewer's comments prove that the revised version of the paper could answer most of the complaints that were raised with the first review. I think that the paper could strongly improve if the Authors take in account in an accurate way all the questions posed by the reviewers.*
>
> *In particular, I stress that all the reviewers raised the problem of the resolution of the inverted geophysical data and this should be carefully addressed.*

We agree that the issue of the resolution of the airborne based TEM geophysical data was not sufficiently discussed in the first version of the manuscript and it is an important aspect to consider especially for choosing the grid size for the simulation and for decimating the soft conditioning dataset.

When addressing the resolution of SkyTEM data one has to consider the individual factors that influence the resolution. Initially the volume of sediment that gets "energized" strongly depends on the sediment; with a larger volume for high conductance materials (high clay content). That is the reason why the SkyTEM device tends to overlook small sand features. Additionally the resolution of the initial voltage data varies in lateral and vertical direction. The lateral resolution of a SkyTEM sounding increases with depth, as the penetration of the subsurface is shaped as a cone, with 15-20m on the surface to a larger support scale in larger penetration depths (at 30m depth the lateral support size will be in the range of 50m). The glacial sequence which defines the model domain is between 10m and 40m thick. The vertical support scale of the SkyTEM measurement device will be rather constant over depth and will thus not vary as much with penetration depth as the lateral support scale. The second step of the data processing is the inversion of voltage data to resistivity. This step is supposed to inflate the resolution by smoothing the values. Last the inverted data are interpolated using kriging, which is known to be a smooth interpolator.

Taking these three points into consideration one can summarize that due to the physics of the measurement the resolution increases with depth and with conductance of the sediment. Disregarding depth, the entire domain is affected by the inversion and the kriging interpolation, both smoothing the data and inflating the resolution.

However, one can assume that the chosen grid size of 20m x 20m x 2m is suitable for near surface resistivity values. For assessing deep buried geological heterogeneity (e.g. around 100m and more) the grid size must be increased. Nevertheless, one has to consider that with increasing depth the support size may grow larger than the defined grid size, where the lateral direction is more affected than the vertical direction. One can conclude that the resolution of the geophysical data is constantly smaller than the correlation length, which is approximately 500m in vertical direction and 5m in lateral direction. This again supports the chosen grid size.

Please find the relevant sections in the revised manuscript: 3.3 lines 213-218 and 6.3 lines 570-581.

> *Also the discussion on the method use for decimating the soft conditioning data set should be considered with great care.*

We decided to add additional detail on describing the way the decimation is conducted and also on the discussion on the implications of decimating the conditioning dataset.

The decimating is conducted by sampling the original soft conditioning dataset at different distances, thus out-thinning. Moving and static sampling are applied at different distances. The static sampling generates one sampling grid, where the distances between the sampling locations are constant. The moving sampling approach does not use a moving window thus there is no resampling, averaging or interpolation included. Instead, it generates different location grids for the samples. The n different location grids have the same distance between the samples for each chosen distance (100m, 200m, etc.), but each has an accumulated shift of the origin (+ sampling distance/n in X and Y direction). For the TProGS application five location grids are generated, which yields five independent soft conditioning datasets. For the 100m moving sampling approach the first sampling grid has the origin (0,0) the second (20,20), the third (40,40), etc. (Figure 1). Five realizations are computed for each soft dataset; giving a total of 25 realizations.

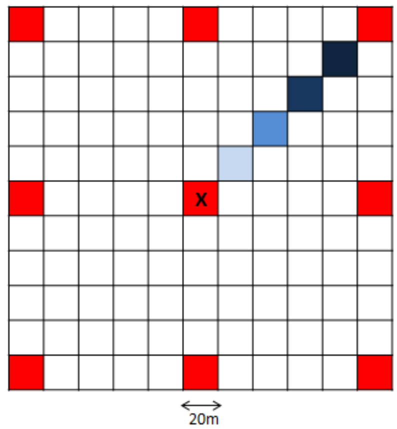

Figure 1. The systematical pattern of the moving sampling of soft data. The red cells exemplify the 100m static sampling grid, with the **X** as the origin (0/0). The blue cells symbolize the sampling grid origins of the additional sampling grids for the 100m moving sampling scenario in the moving sampling conditioning (20/20), (40,40), etc. This generates 5 conditioning datasets sampled from the original soft conditioning dataset with equal spacing but different origins.

Decimating the soft conditioning dataset may seem as an overly simplistic and very crude approach, but we aim at finding the balance between too few data and too many data. The first case is presented in section 5.2 (onlyBH) and 5.4 (BH-Sky500static); one can say that the risk to miss important features is high when conditioning to too few data. Our study mainly deals with the latter case, where too many data lead to an underestimation of the simulation uncertainty. We agree, and have acknowledged in our manuscript, that valuable local information might be sacrificed in the process of out-thinning the conditioning dataset. However the entire SkyTEM dataset is considered during the validation process, which ensures that the TProGS results, which are conditioned to a decimated dataset, are validated against the entire geophysical information available. In our study we tried different sampling techniques; static and moving sampling at various distances. We analyzed the tradeoff between an increase of sampling distance and a decrease in accuracy of a set of realizations, where 200m moving sampling performed best. This was essential a trial and error approach with no profound scientific background. However the matrix of performance criteria (Table 4 in manuscript) allowed us to validate which sampling distance performed best. We believe that a 200m sampling distance is still sufficient to adequately capture all relevant geological features proxied by the entire dataset; this can be argued by the fraction between the observed mean length and the conditioning spacing. The mean length of a sand lens is found to be 500m and can proxy the correlation length. With a horizontal length scale of 500m and sampling at 200m we still condition the simulation with two to three soft data points in each horizontal direction for each mean sized sand feature.

Please find the relevant sections in the revised manuscript: 4.3 lines 278-290 and 6.5 lines 599-621.

With regard to the remark by Reviewer #1 that the title is very general, I suggest to discuss whether the *conclusions of this study could be generalized also for the use of other methods of stochastic simulation and other techniques of geophysical prospecting.*

This a very interesting point that was not included in the initial discussion section. Firstly we cannot directly conclude that overconditioning is a general problem in stochastic simulations where a vast conditioning dataset is applied. However we can presume that heavily spatially correlated data will affect also other stochastic simulation algorithms. He et al. (2013) incorporated a very similar SkyTEM dataset in a stochastic simulation using multi point statistics; SNESIM algorithm (Liu, 2006). The vast and dense geophysical dataset was indirectly used to derive the training image and not directly for conditioning the simulation. The conditioning included simply lithological borehole data.

TProGS was clearly not developed to run with such comprehensive conditioning. To our knowledge, the problem of overconditioning has not yet been reported and with our study we would like to create awareness that might help other researches when designing their stochastic simulation. We believe that possible implications of spatially correlated conditioning data ought to be considered when performing stochastic simulations.

In regard to the technique of geophysical prospecting it can be concluded that the problem of overconditioning is clearly not limited to airborne based TEM data. Other geophysical methods that generate dense data, e.g. seismic methods can show similar complications.

Please find the relevant sections in the revised manuscript: 6.5 lines 622-628.

The Authors state that "no general empirical relationship between resistivity measured by an airborne based TEM method and hydrological attributes has been studied". However, there is a great literature about laboratory, field and theoretical studies about the relationship between electrical and electromagnetic parameters (no matter which prospecting method is used) and properties of the sediments and of the pore fluid.

We agree with the editor in this point. We make a very general comment which only applies to airborne based TEM methods. At the same time we forget about other studies that address the relationship between electrical resistivity and hydrological properties. We acknowledge the point and want to bring forward three interesting studies. (Infante et al., 2010) compare vertical electrical sounding data (resistivity) with seismic data and two boreholes containing lithological descriptions at a field site. The three data types correlate well and it is concluded that both geophysical methods are influenced by common geological attributes such as water content, lithological description, and geological structure. Another study by (Bowling et al., 2007)integrate direct current resistivity (DC) data with ground penetrating radar (GPR) and lithological data (grainsize analysis) at a field site. Resistivity is linked to grainsize distribution and is used to delineate mayor geological structures. A strong positive correlation between gravel content in the lithology and resistivity was observed. (Bowling et al., 2005) investigate a sand and gravle outcrop to establish an empirical relationship between geophysical attributes (DC and GPR data) and lithostratigraphic properties. However we believe that our study focuses rather on the application of airborne based TEM data and a thorough discussion on different approaches goes beyond the scope of our study.

Please find the relevant sections in the revised manuscript: 3.2 lines 196-201.

> *Therefore, I would appreciate an improvement of the description of the method applied to transform electrical conductivity in lithological properties, by taking into account also physical aspects, such as variability of pore water conductivity and interactions between pore water and solid grains.*

Please find a more detailed description of translating electrical conductivity into lithological properties:

The SkyTEM data, measured in resistivity, needs to be linked to a facies type in order to be used in a stochastic geological simulation. The procedure to integrate SkyTEM data with facies information obtained from borehole data used in this study is the histogram probability matching method (HPMM) (He et al., 2014). The idea behind this method is the assumption that the probability of facies occurrence is positively correlated to the occurrence of resistivity over a certain range. Therefore the continuous SkyTEM data are classified into bins with a defined range (10Ωm). A fixed vertical discretization is defined representing the scale of the assessed heterogeneity, 2m in this case. The geophysical data is then compared with the categorical borehole data at collocated cells and the data pairs are grouped after the chosen bin width (10Ωm). Thus each bin contains a number of data pairs and a facies fraction of the categorized borehole data can be calculated respectively. The fraction can be plotted as bars in a histogram and polynomial curve fitting allows translating any resistivity observation into a probability of facies occurrence (Figure 3, section 3.2 in manuscript).The HPM-method used in this study is purely based on spatial correlations and is not build up on physical relationships. The main limitation is that it is site specific and cannot be applied to other catchments. On the other hand it was never the intention to create a general histogram with universal applicability.

As mentioned correctly by the editor, there are many sources of uncertainty that will affect the relationship between electrical conductivity and facies information. The HPM-method lumps various sources and the shape of the fitted curve reflects those, especially the width of the transition zone. The deviation between the fitted curve and the "ideal" curve (where only zeros exist below the cut-off value and ones above the cut-off value) indicates the extent of the combined uncertainty. Thus, the larger the combined uncertainty is, the lower the slope of the fitted curve will be. He et al. (2014) discussed the prevalent uncertainties: first, borehole descriptions are not accurate, and classification of borehole lithology is subjective. Second, there are uncertainties on the resistivity data due to the resolution of the physics itself, the geophysics instruments, field measurements and signal processing (inversion). Third, there is no unique relationship between resistivity and lithology, and the curve can therefore be fitted in various ways. Last, there are uncertainties related to the scale of aggregation, since the borehole data and geophysical data have different resolutions and hence different supporting scales.

Please find the relevant sections in the revised manuscript: 3.3 lines 219-243.

References

Bowling, J. C., D. L. Harry, A. B. Rodriguez and C. M. Zheng, Integrated geophysical and geological investigation of a heterogeneous fluvial aquifer in Columbus Mississippi, Journal of Applied Geophysics, 62(1), 58-73, 2007.

Bowling, J. C., A. B. Rodriguez, D. L. Harry and C. M. Zheng, Delineating alluvial aquifer heterogeneity using resistivity and GPR data, Ground Water, 43(6), 890-903, 2005.

He, X., T. O. Sonnenborg, F. Jorgensen, A. S. Hoyer, R. R. Moller and K. H. Jensen, Analyzing the effects of geological and parameter uncertainty on prediction of groundwater head and travel time, Hydrology and earth system sciences, 17(8), 3245-3260, 2013.

He, X., J. Koch, T. O. Sonnenborg, F. Jorgensen, C. Schamper and J. C. Refsgaard, Uncertainties in constructing stochastic geological models using transition probability geostatistics and transient AEM data., Water Resources Research, 2014.

Infante, V., L. A. Gallardo, J. C. Montalvo-Arrieta and I. N. de Leon, Lithological classification assisted by the joint inversion of electrical and seismic data at a control site in northeast Mexico, Journal of Applied Geophysics, 70(2), 93-102, 2010.

Liu, Y. H., Using the Snesim program for multiple-point statistical simulation, Computers & geosciences, 32(10), 1544-1563, 2006.