Hydrology and Earth System Sciences
Open Access
Discussions

Large scale 3-D modeling by integration of resistivity models and borehole data through inversion

N. Foged[1], P. A. Marker[2], A. V. Christansen[1], P. Bauer-Gottwein[2], F. Jørgensen[3], A.-S. Høyer[3], and E. Auken[1]

[1]HydroGeophysics Group, Department of Geoscience, Aarhus University, Aarhus, Denmark
[2]Department of Environmental Engineering, Technical University of Denmark, Copenhagen, Denmark
[3]Geological Survey of Denmark and Greenland, Aarhus, Denmark

HESSD

11, 1461–1492, 2014

Large scale 3-D modeling by integration of resistivity models

N. Foged et al.

Title Page

Abstract	Introduction
Conclusions	References
Tables	Figures

◄◄ | ►►
◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Abstract

We present an automatic method for parameterization of a 3-D model of the subsurface, integrating lithological information from boreholes with resistivity models through an inverse optimization, with the objective of further detailing for geological models or as direct input to groundwater models. The parameter of interest is the clay fraction, expressed as the relative length of clay-units in a depth interval. The clay fraction is obtained from lithological logs and the clay fraction from the resistivity is obtained by establishing a simple petrophysical relationship, a translator function, between resistivity and the clay fraction. Through inversion we use the lithological data and the resistivity data to determine the optimum spatially distributed translator function. Applying the translator function we get a 3-D clay fraction model, which holds information from the resistivity dataset and the borehole dataset in one variable. Finally, we use k means clustering to generate a 3-D model of the subsurface structures. We apply the concept to the Norsminde survey in Denmark integrating approximately 700 boreholes and more than 100 000 resistivity models from an airborne survey in the parameterization of the 3-D model covering 156 km^2. The final five-cluster 3-D model differentiates between clay materials and different high resistive materials from information held in resistivity model and borehole observations respectively.

1 Introduction

In a large-scale geological and hydrogeological modeling context, borehole data seldom provide an adequate data base due to low spatial density in relation to the complexity of the subsurface to be mapped. Contrary, dense areal coverage can be obtained from geophysical measurements, and particularly airborne EM methods are suitable for 3-D mapping, as they cover large areas in a short period of time. However, the geological and hydrogeological parameters are only mapped indirectly, based on site-specific relationships. Therefore e.g. the link between hydrological properties

and electrical properties has been an area of increased interest as reviewed by Slater (2007).

Integrating geophysical models and borehole information has proved to be a powerful combination for 3-D geological mapping (Jørgensen et al., 2012; Sandersen et al., 2009) and several modeling approaches have been reported. One way of building 3-D models is through a knowledge-driven (cognitive), manual approach (Jørgensen et al., 2013a). This can be carried out by making layer-cake models composed of stacked layers separated by surfaces or by making models composed of structured or unstructured 3-D meshes where each voxel is assigned a geological/hydrogeological property. The latter allows for a higher degree of model complexity to be incorporated (Turner, 2006; Jørgensen et al., 2013a). The cognitive approach enables various types of background knowledge such as the sedimentary processes, sequence stratigraphy, etc. to be utilized. However, the cognitive modeling approach is difficult to document and reproduce due to its subjective nature. Moreover, any cognitive approach will be quite time-consuming, especially when incorporating large airborne electromagnetic (AEM) surveys, easily exceeding 100 000 resistivity models.

Geostatistical modeling approaches such as multiple-point geostatistical methods (Daly and Caers, 2010; Strebelle, 2002), transition probability indicator simulation (Fogg, 1996) or sequential indicator simulation (Deutsch and Journel, 1998), provide models with a higher degree of objectivity in shorter time compared to the cognitive, manual modeling approaches. An example of combining AEM and borehole information in a transition probability indicator simulation approach is given by He et al. (2013). The use of only borehole data in geostatistical modeling problems often faces the problem that the data are too sparse to represent the lateral heterogeneity. Including geophysical data enables a more accurate estimation of geostatistical properties, especially laterally, but opens the question of what to use as *hard* and *soft* data in the model simulations and estimations. Combined use of geostatistical and cognitive approaches can be a suitable solution in some cases (Jørgensen et al., 2013b; Raiber et al., 2012; Stafleu et al., 2011). Integration of borehole information and geological knowledge as

prior information directly in the inversion of the geophysical data is another technique to combine the two types of information and thereby achieve better geophysical models and subsequently better geological and hydrological models (Høyer et al., 2014; Wisén et al., 2005).

Geological models are commonly used as the basis for hydrostratigraphical input to groundwater models. While model predictions are sensitive to variations in hydrostratigraphy, non-uniqueness with respect to hydrostratigraphy is inherent to groundwater models (Seifert et al., 2012).

Sequential, joint and coupled hydrogeophysical inversion techniques (Hinnell et al., 2010) have been used to inform groundwater models with both geophysical and traditional hydrogeological observations. Such techniques use petrophysical relationships to translate between geophysical and hydrogeological parameter spaces. For applications in groundwater modeling using electromagnetic data see e.g. Dam and Christensen (2003) and Herckenrath et al. (2013). Also clustering analyses can be used to delineate subsurface hydrogeological properties. Fuzzy c means clustering has been used to delineate geological features from measured EM34 signals with varying penetration depths (Triantafilis and Buchanan, 2009) and to delineate the porosity field from tomography inverted radar attenuation and velocities and seismic velocities (Paasche et al., 2006).

We present an automatic method for parameterization of a 3-D model of the subsurface. The geological parameter we map is the clay fraction (CF), expressed as the cumulated thickness of *clay* in a depth interval relative to the interval length. The method integrates lithological information from boreholes with resistivity information, typically from large-scale geophysical AEM surveys. We obtain the CF from the resistivity data by establishing a petrophysical relationship, a translator function, between resistivity and the CF. Through an inverse mathematical formulation we use the lithological borehole data to determine the optimum parameters of the translator function. Hence, the 3-D CF-model holds information from the resistivity dataset and the borehole dataset in one variable. As a last step we cluster our model space represented by the CF-model

and geophysical resistivity model using ⬜means clustering to form a structural 3-D cluster model with the objective of further detailing for geological models or as direct input to groundwater models.

First, we give an overall introduction to the CF-concept, and then we move to a more detailed description of the different parts: observed data and uncertainty, forward modeling, inversion and minimization, and clustering. Last we demonstrate the method in a field example with resistivity data from an airborne SkyTEM survey combined with quality-rated borehole information.

2 Methodology

Conceptually, our approach sets up a function that best describes the petrophysical relationship between clay fraction and resistivity. Through inversion we determine the optimum parameters of this translator function, by minimizing the difference between the clay fraction calculated from the resistivity models (Ψ_{res}) and the observed clay fraction in the lithological well logs (Ψ_{log}). We do this for a number of elevation intervals (calculation intervals) that are mutually constrained to cover an entire 3-D model space. Having the optimized and spatially distributed translator function we can transform the resistivity models to form a 3-D clay fraction model, incorporating the key information from both the resistivity models and the lithological logs into one parameter. The CF-concept is a further development to three dimensions of the accumulated clay thickness concept by Christiansen et al. (2013), which is formulated in 2-D.

The flowchart in Fig. ⬜provides an overview of the CF-concept. The observed clay fraction (Ψ_{log}) is calculated from the lithological logs (box 1) in the calculation intervals. The translator function (box 2) and the resistivity models (box 3) form the forward response which produces a resistivity-based clay fraction (box 4) in the different calculation intervals. The parameters of the translator function are updated during the inversion to obtain the best consistency between Ψ_{res} and Ψ_{log}. The output is the optimum resistivity-to-clay fraction translator function (box 5) and when applying this to the

resistivity models (the forward response of the final iteration), we obtain the optimum Ψ_{res} and block kriging is used to generate a regular 3-D CF model (box 6).

The final step is a *k* means clustering analysis (box 7). With the clustering we achieve a 3-D model of the subsurface delineating a predefined number of clusters that represent zones of similar physical properties, which can be used as input in, for example, a detailed geological model or as structural delineation for a groundwater model.

The subsequent paragraphs detail the description of the individual parts of the concept.

2.1 Observed data – lithological logs and clay fraction

The common parameter derived from the lithological logs and resistivity datasets is the clay fraction (Fig. 1, box 1–4). It is a common assumption that a petrophysical relationship between resistivity and clay content can be established shown for instance by Waxman and Smits (1968) and Shevnin et al. (2007). From the lithological logs we only have a lithological description, and in many cases only a very simple one; *sand*, *clay*, *gravel*, *chalk*, etc. Even in cases where more detailed descriptions with for instance sedimentary facies (e.g. *clay till*) or age (e.g. *Palaeogene clay*) are available, is it not possible to obtain the actual clay content from the descriptions. This is only possible if detailed lab-analyses have been carried out, which are extremely rare on a larger scale. In this paper we therefore refer to *clay* as material described as clay in a lithological well log regardless the type of clay; *clay till*, *mica clay*, *Palaeogene clay*, etc. This term is robust in the sense that most geologists and drillers have a common conception on the description of *clay*. The *clay fraction*, Ψ_{log}, of a given depth interval in a borehole is therefore calculated as the cumulative thickness of layers described as *clay* divided by the length of the interval. By using this definition of clay and clay fraction we can easily calculate the clay fraction in depth intervals for any lithological well log. Having retrieved the Ψ_{log} values we then need to estimate their uncertainties since a variance estimate, σ_{log}, needed in the evaluation of the misfit to Ψ_{res}. The

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

drilling method is one of the key parameters affecting the uncertainty of the well log data.

The drillings are conducted with a range of different methods. This has a large impact on the uncertainties of the lithological well log data. The drilling methods span from core drilling resulting in a very good base for the lithology classification, to direct circulation drillings (cuttings are flushed to the surface between the drill rod and the formation resulting in poorly determined layer boundaries and a very high risk of getting samples contaminated due to the travel time from the bottom to the surface. Other parameters affecting the uncertainty of the Ψ_{log} are, parameters like sample interval and density, accuracy of the geographical positioning and elevation, and the credibility of the contractor.

2.2 Forward data – the translator function

For calculating the clay fraction for a resistivity model, Ψ_{res}, we use a simple two-parameter translator function as shown in Fig. 2. The translator function is described fully by a scaled complementary error function:

$$W(\rho) = 0.5 \cdot \text{erfc} \left(\frac{K \cdot (2\rho - m_{\text{up}} - m_{\text{low}})}{(m_{\text{up}} - m_{\text{low}})} \right), \, K = \text{erfc}^{-1}(0.0025 \cdot 2) \tag{1}$$

where m_{low} and m_{up} are defined as the resistivity at which the translator function, $W(\rho)$, returns a weight of 0.975 and 0.025 respectively. For a layered resistivity model, the Ψ_{res} value in an interval, is then calculated as:

$$\Psi_{\text{res}} = \frac{1}{\sum t_i} \cdot \sum_{i=1}^{N} W(\rho_i) \cdot t_i \tag{2}$$

where N is the number of resistivity layers in the calculation interval, $W(\rho_i)$ is the clay weight for the resistivity in layer i, t_i is the thickness of the resistivity layer, and Σt_i is the

length of the calculation interval. In other words, W weights the thickness a resistivity layer, so for a resistivity below m_{low} the layer thickness is counted as clay ($W \approx 1$) while for a resistivity above m_{up} the layer is counted as non-clay ($W \approx 0$).

The resistivity models are also associated with an uncertainty and if the variance estimates of the resistivities and thicknesses for the geophysical models are available we take these into account. The propagation of the uncertainty from the resistivity models to the Ψ_{res} values is described in detail in Christiansen et al. (2013).

To allow for variation, laterally and vertically, in the resistivity to Ψ_{res} translation, a regular 3-D grid is defined for the survey block (Fig. 2). Each grid node holds one set of m_{up} and m_{low} parameters. The vertical discretization follows the clay fraction calculation intervals, typically 4–10 m intervals. 2-D bilinear horizontal interpolation of the m_{up} and m_{low} is applied to define the translator function uniquely at the positions of the resistivity models.

To migrate information of the translator function from regions with many boreholes to regions with few boreholes or with no boreholes, horizontal and vertical smoothness constraints are applied between the translator functions at each node point as shown in Fig. 2. The smoothness constraints furthermore act as regularization and stabilize the inversion scheme.

Finally, we estimate Ψ_{res} values at the Ψ_{log} positions (named Ψ^*_{res}) using point kriging interpolation. The experimental semi-variogram is calculated from the Ψ_{res} values for the given calculation interval and can normally be approximated well with an exponential function, which then enters the kriging interpolation. The code Gstat (Pebesma and Wesseling, 1998) is used for kriging, variogram calculation, and variogram fitting. By using kriging for interpolation the spatial variance of Ψ_{res} is taken into account and even more important, it provides uncertainty estimates (σ^*_{res}) of the Ψ^*_{res} values, which include the original uncertainty of Ψ_{res} and the interpolation uncertainty. These uncertainty estimates are needed for a meaningful evaluation of the data misfit at the borehole positions.

2.3 Inversion – objective function and minimization

The inversion algorithm in its basic form consists of a nonlinear forward mapping of the model to the data space:

$$\delta\Psi_{obs} = \mathbf{G}\,\delta m_{true} + e_{log} \tag{3}$$

where $\delta\Psi_{obs}$ denotes the difference between the observed data (Ψ_{log}) and the non-linear mapping of the model to the data space (Ψ_{res}). δm_{true} represents the difference between the true translator function and an arbitrary reference model. e_{log} is the observational error, and $\mathbf{G}$ denotes the Jacobian matrix that contains the partial derivatives of the mapping. The general solution to the non-linear inversion problem of Eq. (3) is described by Christiansen et al. (2013) and is based on Auken and Christiansen (2004) and Auken et al. (2005).

The objective function, Q, to be minimized includes a data term, R_{dat}, and a regularization term from the horizontal and vertical constraints, R_{con}. R_{dat} is given as:

$$R_{dat} = \sqrt{\frac{1}{N_{dat}} \cdot \sum_{i=1}^{N_{dat}} \frac{\left(\Psi_{log,i} - \Psi_{res,i}^{*}\right)^2}{(\sigma_i)^2}} \tag{4}$$

where N_{dat} is the number of Ψ_{log} values and σ_i is the combined variance of the ith Ψ_{log} (σ_{log}) and Ψ_{res} (σ_{res}^{*}) given as:

$$\sigma_i = \sqrt{\sigma_{log,i}^2 + \sigma_{res,i}^{*2}}. \tag{5}$$

The inversion is performed in logarithmic model space to prevent negative parameters, and R_{con} is therefore defined as:

$$R_{con} = \sqrt{\frac{1}{N_{con}} \cdot \sum_{i=1}^{N_{con}} \frac{\left(\ln\left(m_j\right) - \ln(m_k)\right)^2}{\left(\ln\left(e_{r,i}\right)\right)^2}} \tag{6}$$

where e_r is the regularizing constraint between the two constrained parameters m_j and m_k of the translator function and N_{con} is the number constraint pairs. The e_r values in Eq. (6) are stated as constraint factors, meaning that an e_i factor of 1.2 corresponds approximately to a model change of ±20 %.

In total the objective function Q becomes:

$$Q = \sqrt{\frac{N_{dat} \cdot R_{dat}^2 + N_{con} \cdot R_{con}^2}{(N_{dat} + N_{con})}}. \tag{7}$$

Furthermore, is it possible to add prior information as a prior constraint on the parameters of the translator function, which just adds a third component to Q in Eq. (7) similar to R_{con} in Eq. (6).

The minimization of the non-linear problem is performed in a least squares sense by using an iterative Gauss–Newton minimization scheme with a Marquardt modification. The full set of inversion equations and solutions are presented in Christiansen et al. (2013).

2.4 Cluster analysis

The delineation of the 3-D model is obtained through a k means clustering analysis which distinguishes groups of common properties within multivariate data. We have based the clustering analysis on the CF-model and the resistivity model. Other data, which are informative for structural delineation of geological or hydrological properties, can also be included in the cluster analysis. For example this could be geological a priori information or groundwater quality data. The resistivity model is part of the CF-model, but is reused for the clustering analysis because the representation of lithology used in the CF-model inversion has simplified the geological heterogeneity captured in the resistivity model.

K means clustering is a hard clustering algorithm used to group multivariate data. A k means cluster analysis is iterative optimization with the objective to minimize a

distance function between data points and a predefined number of clusters (Wu, 2012). We have used Euclidean length as a measure of distance. We use the k means algorithm in MATLAB R2013a, which has implemented a two-phase search, batch and sequential, to minimize the risk of reaching a local minimum (Wu, 2012). K means clustering can be performed on several variables, but for variables to impact the clustering equally, data must be standardized and uncorrelated. The CF-model and resistivity model are by definition correlated. We use Principal Component Analysis (PCA) to obtain uncorrelated variables.

Principal component analysis is a statistical analysis based on data variance formulated by Hotelling (1933). The aim of a PCA is to find linear combinations of original data while obtaining maximum variance of the linear combinations (Härdle and Simar, 2012). This results in an orthogonal transformation of the original multi-dimensional variables into a space where dimension one has largest variance, dimension two has second largest variance, etc. In this case the PCA is not used to reduce variable space, but only to obtain an orthogonal representation of the original variable space to use in the clustering analysis. Principal components are orthogonal and thus uncorrelated, which makes the principal components useful in the subsequent clustering analysis. The PCA is scale sensitive and the original variables must therefore be standardized prior to the analysis. Because the principal components have no physical meaning a weighting of the CF-model and the resistivity model cannot be included in the k means clustering. Instead the variables are weighed prior to the PCA.

3 Norsminde case

The Norsminde case model area is located in eastern Jutland, Denmark (Fig. 3) around the town of Odder (Fig. 4) and covers 156 km^2, representing the Norsminde Fjord catchment. The catchment area has been mapped and studied intensely in the NiCA research project in connection to nitrate reduction in geologically heterogeneous catchments (Refsgaard et al., 2014). The modeling area has a high degree of geological

complexity in the upper part of the section. The area is characterized by Palaeogene and Neogene sediments covered by glacial Pleistocene deposits. The Palaeogene is composed of fine-grained marl and clay and the Neogene layers consist of marine Miocene clay interbedded with deltaic sand layers (Rasmussen et al., 2010). The Neo-
gene is not present in the southern and eastern part of the area where the glacial sediments therefore directly overlie the Palaeogene clay. The Palaeogene and Neo-gene layers in the region are frequently incised by Pleistocene buried tunnel valleys and one of these is present in the southern part, where it crosses the model area to great depths with an overall E-W orientation (Jørgensen and Sandersen, 2006). The
Pleistocene deposits generally appear very heterogeneous and according to boreholes they are composed of glacial meltwater sediments and till.

3.1 Borehole data

In Denmark, the borehole data are stored in the national database Jupiter (Møller et al., 2009) dating back to 1926 as an archive for all data and information obtained by drilling.
Similar databases are maintained in other countries. Today, the Jupiter database holds information about more than 240 000 boreholes. For the lithological logs a fixed lithol-ogy code list is available and all different types of clay layers are easily identified, and the Ψ_{log} values for the desired elevation intervals can be calculated.

For the model area, approximately 700 boreholes are stored in the database. Based
on borehole meta-data found in the database we use an automatic quality rating sys-tem, where each borehole is rated from 1–4 (He et al., 2013). The ratings are used to apply the lithological logs with uncertainty (weights) used in the inversion.

The meta-data used for the quality-rating are:

– drill method: auger, direct circulation, air-lift drilling, etc.

– sample density

– accuracy of the geographical position: GPS or manual map location

- accuracy of the elevation: differential GPS or other

- drilling purpose: scientific, water abstraction, geophysical shot holes, etc.

- credibility of drilling contractor.

The boreholes are awarded points in the different categories and finally grouped into four quality groups according to their total score. Boreholes in the lowest quality group (4) are primarily boreholes with low sample frequencies (less than 1 sample per 10 m), low accuracy in geographical position, and geophysical shot holes for seismic exploration.

The locations and quality ratings of the boreholes are shown in Fig. 4b, while the drill depths and quality ratings are summarized in Fig. 5. As the top bar in Fig. 5 shows, 4 % of the boreholes are categorized as quality 1, 46 % as quality 2, 32 % as quality 3, and 18 % as quality 4. The uncertainties of the Ψ_{log} values for the quality groups 1–4 are based on a subjective evaluation and are defined as 10, 20, 30, and 50 %, respectively. The number of boreholes drastically decreases with depth as shown in Fig. 5. Thus, while about 100 boreholes are present in a depth of 60 m, only 25 boreholes reach a depth greater than 90 m.

3.2 EM data

The major part of the model area is covered by SkyTEM data and adjoining ground based TEM soundings are included in the resistivity dataset (Fig. 4a).

The SkyTEM data were collected with the newly developed SkyTEM[101] system (Schamper et al., 2013). The SkyTEM[101] system has the ability to measure very early times, which improves the resolution of the near surface geological layers when careful system calibration and advanced processing and inversion methodologies are applied (Schamper et al., 2014). The recorded times span the interval from $\sim 3\,\mu s$ to 1–2 ms after end of the turn-off ramp, which gives a depth of investigation (Christiansen and Auken, 2012) of approximately 100 m for an average ground resistivity of 50 Ωm. The

SkyTEM survey was performed with a dense line spacing of 50 m for the western part and 100 m line spacing for eastern part (Fig. 4a). Additional cross lines were made in a smaller area, which brings the total up to approximately 2000 line km. The sounding spacing along the lines is approximately 15 m resulting in a total of 106 770 1-D resis-
tivity models. The inversion was carried out in a spatially constrained inversion setup (Viezzoli et al., 2008) with a smooth 1-D model formulation (29 layers, with fixed layer boundaries), using the AarhusInv inversion code (Auken et al., 2014) and the Aarhus Workbench software package (Auken et al., 2009) . The resistivity models have been terminated at the estimated depth of investigation (DOI) calculated as described by
Christiansen and Auken (2012).

The ground based TEM soundings originate from mapping campaigns in the mid-1990s. The TEM sounding were all acquired with the Geonics TEM47/PROTEM system (Geonics Limited) in a central loop configuration with a 40 by 40 m^2 transmitter loop. Data were inverted single site using a 1-D layered resistivity model with 3 to 5 layers
depending on the number of layers needed to fit the data.

3.3 Model setup

The 3-D translator function grid has a horizontal discretization of 1 km, with 16 nodes in the x direction and 18 nodes in the y direction. Vertically the model spans from 100 m a.s.l. (highest surface elevation) to 120 m b.s.l. The vertical discretization is
4 m a.s.l. and 8 m b.s.l., which results in 40 calculation intervals. Hence, in total the model grid holds $16 \times 18 \times 40 = 11\,520$ translator functions each holding two parameters.

The horizontal regularization constraints between neighboring nodes are set to a factor of 2, while the horizontal regularization constraint is set to a factor of 3. A uniform
initial translator function was used with $m_{low} = 35\,\Omega m$ and $m_{up} = 55\,\Omega m$. Starting model and constrains setup are based on experience and the expected geological variability and fine-tuned through subsequent of test-inversions.

Node points in the translator function grid situated in major data gaps (above terrain, below DOI, outside geophysical coverage) are purely driven by the model constraints and the starting model. The effective number of translator functions, that are situated in the vicinity of resistivity models and borehole data is approximately 5200.

In the interpolation to make the regular 3-D CF-model, Ψ_{log} values are included together with the Ψ_{res} values to close gaps in the resistivity dataset where boreholes are present.

The k means clustering is performed on two variables, the CT-model and resistivity model, in a 3-D grid with regular horizontal discretization of 100 m and vertical discretization of 4 m between 96 and 0 m a.s.l. and 8 m between 0 and 120 m b.s.l. CF-model values range between 0 and 1 and have therefore not been standardized. The resistivity values have been log transformed and standardized by first subtracting the mean and then dividing by four times the standard deviation. The standardization of the resistivity was performed in this way to balance the weight between the two variables in the clustering. A five cluster delineation is presented for the Norsminde case in the result section.

3.4 Results and discussion

CF-modeling results from the Norsminde area are presented in cross sections in Fig. 6 and as horizontal slices in Fig. 7.

Figure 6a and b show the inversion results of the m_{low} and m_{up} parameters. The variations in the translator function are relatively smooth especially in the deeper part of the model (below -10 m). The smooth translator function for the deeper part corresponds well to the general geological setting of the area with relatively homogenous clay sequences in the deeper part, but is also due to very limited borehole information for the deeper part. The general geological setting of the area is also clearly reflected in the translator function in the horizontal slices in Fig. 7a and b. The eastern part of the area with lowest m_{low} values (dark blue in Fig. 7a) and lowest m_{up} values (light blue/green in Fig. 7b) corresponds to the area where the Palaeogene highly conductive

clays are present. In the western part of the area the cross section intersect the glacial complex, where the clays are mostly tills, and higher m_{low} and m_{up} values are needed to get the optimum translation.

Beside the regularization and initial starting model two main parts control the resulting m_{low} og m_{up}. The first part concerns the fact that both units described as clay and non-clay in the lithological logs can exhibit a relatively wide range of resistivities. For example, heavy clays may have resistivities of 2–3 Ωm and firm and dry clay tills can have relatively high resistivities in the range of 80 Ωm. Furthermore, changes in resistivity occur within the same geological unit due to changes in the pore water resistivity as described by Archie's law. The second issue concerns the resolution of the true formation resistivity in the resistivity models. Lithological logs contain point information with a good and uniform vertical resolution. Contrary, AEM data provide a good spatial coverage, but the vertical resolution for the EM resistivity models is relatively poor and not necessarily returning the true resistivity of the formation. Especially thin high-resistivity layers (sand layers) at great depth are poorly resolved by the EM-methods making geological interpretation difficult. By allowing spatial variation in the translator function we can, to some degree, resolve weak layer indications in the resistivity models lithologically correct while also accounting for variations in the pore water resistivity and other resistivity changes within the same lithological description.

The resistivity cross section in Fig. 6c and the slice section in Fig. 7c reveal a detailed picture of the effect of the geological structures seen in the resistivity data. Generally, a good correlation to the boreholes is observed. Translating the resistivities we obtain the CF-model presented in Figs. 6d and 7d. The majority of the voxels in the CF-model have values close to 0 or 1. This is expected since the lithological logs are described binary clay/non clay, and Ψ_{log} values not equal to 0 or 1 can only occur if more lithological layers are present in the calculation interval.

Evaluating the result in Figs. 6d and 7d, it is obvious that the very resistive zones are translated to a CF-value close to 0 and the very conductive zones are translated to CF-value close to 1. Focusing on the intermediate resistivities (20–60 Ωm) it is clear

HESSD

11, 1461–1492, 2014

Large scale 3-D modeling by integration of resistivity models

N. Foged et al.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

that the translation of resistivity to CF is not one-to-one. For example, the buried valley structure (profile coordinate 6500–8500 m, Fig. 6d) has mostly high-resistive fill with some intermediate resistivity zones. In the CF-section these intermediate resistivity zones are translated to zones of high clay content, consistent with the lithological log at profile coordinate 7000 m that contains a 25 m thick clay layer. The CF-section sharpens the layer boundaries compared to the smooth layer transitions in the resistivity section. The integration of the resistivity data and lithological logs in the CF-concepts results in a high degree of consistency between the CF-results and the lithological logs, as seen in the CF-section in Fig. 6d.

Horizontal slices of the 3-D cluster model are shown in Fig. 8. The near-surface part of the model (Fig. 8a–b) are dominated by clusters 2 and 4, while the deeper parts of the model (Fig. 8c–d) are dominated by clusters 3 and 5, with the east-west striking buried valley to the south, (Fig. 8c), is primarily represented by clusters 1 and 2.

The histograms in Fig. 9 show how the original variables, the CF-model and the resistivity model, are represented in the five clusters. Clusters 3 and 5 have resistivity values almost exclusively below 10 Ωm and CF values above 0.7, but mostly close to 1. In the resistivity model space clusters 2 and 4 represent high and intermediate resistivity values respectively with some overlap, while cluster 1 overlap both clusters 2 and 4. Figure 9 also clearly shows that both the resistivity values and the CF-values contribute to the final clusters. The clusters 1, 2, and 4 span only part of the resistivity space with significant overlaps (Fig. 9a), while they are clearly separated in the CF-model space and spanning the entire interval (Fig. 9b). The opposite is observed for clusters 3, 4, and 5, which are clearly separated in the resistivity space (Fig. 9a), but strongly overlapping in the CF-model space (Fig. 9b).

The CF-model does not differentiate between clay types, contrary the EM-resistivity data that have a good resolution in the low resistivity range and therefore, to some degree distinguish between clay types. This results in the two-part clustering of the low resistivity (> 20 Ωm) values as seen in Fig. 9a.

4   Conclusion and outlook

We have presented a concept to produce 3-D clay-fraction models, integrating the key sources of information in a well-documented and objective way.

The concept combines lithological borehole information with geophysical resistivity models in producing large scale 3-D clay fraction models. The integration of the lithological borehole data and the resistivity models is accomplished through inversion, where the optimum resistivity to clay fraction function minimizes the difference between the observed clay fraction from boreholes and the clay fraction found through the geophysical resistivity models. The inversion concept allows for horizontal and lateral variation in the resistivity to clay fraction translation, with smoothness constraints as regularization. The spatially varying translator function is the key to achieve consistency between the borehole information and the resistivity models. The concept furthermore handles uncertainties on both input and output data.

The concept was applied to a 156 km^2 survey with more than 700 boreholes and 100 000 resistivity models from an airborne survey. The output was a detailed 3-D clay fraction model combining resistivity models and lithological borehole information into one parameter.

Finally a cluster analysis was applied to achieve a predefined number of geological/hydrostratigraphic clusters in the 3-D model and enabled us to integrate various sources of information, geological as well as geophysical. The final five-cluster model differentiates between clay materials and different high resistive materials from information held in resistivity model and borehole observations respectively.

With the CF-concept and clustering we aim at building 3-D models suitable as structural input for groundwater models. Each cluster will then represent a hydrostratigraphic unit and the hydraulic conductivity of the units will be determined through the groundwater model calibration constrained by hydrological head and discharge data. For the case study, we have not evaluated cluster validity, i.e. how many clusters the data can support. Cluster validity can be assessed with various statistical measures (Halkidi et

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

al., 2002). If the cluster model is used as structural input to a groundwater model the number of clusters resulting in the best hydrological performance (keeping in mind the principle of parsimony) might also be used as a measure of cluster validity.

The 3-D clay fraction model can also been seen as a binomial geological sand-clay model by interpreting the high and low CF-values as clay and sand respectively, as the color scale for the CF-model example in Figs. 6 and 7 indicated. Integration and further development of the CF-model into more complex geological models have been carried out with success (Jørgensen et al., 2013c).

References

Auken, E. and Christiansen, A. V.: Layered and laterally constrained 2D inversion of resistivity data, Geophysics, 69, 752–761, 2004.

Auken, E., Christiansen, A. V., Jacobsen, B. H., Foged, N., and Sørensen, K. I.: Piecewise 1D Laterally Constrained Inversion of resistivity data, Geophys. Prospect., 53, 497–506, 2005.

Auken, E., Christiansen, A. V., Westergaard, J. A., Kirkegaard, C., Foged, N., and Viezzoli, A.: An integrated processing scheme for high-resolution airborne electromagnetic surveys, the SkyTEM system, Explor. Geophys., 40, 184–192, 2009.

Auken, E., Christiansen, A. V., Kirkegaard, C., Fiandaca, G., Schamper, C., Behroozmand, A. A., Binley, A., Nielsen, E., Effersø, F., Christensen, N. B., Sørensen, K. I., Foged, N., and Vignoli, G.: An overview of a highly versatile forward and stable inverse algorithm for airborne, ground-based and borehole electric and electromagnetic data, Expl. Geophys., in review, 2014.

Christiansen, A. V. and Auken, E.: A global measure for depth of investigation, Geophysics, 77, WB171–WB177, 2012.

Christiansen, A. V., Foged, N., and Auken, E.: A concept for calculating accumulated clay thickness from borehole lithological logs and resistivity models for nitrate vulnerability assessment, J. Appl. Geophys., in review, 2013.

Daly, C. and Caers, J. K.: Multi-point geostatistics – an introductory overview, First Break, 28, 39–47, 2010.

Dam, D. and Christensen, S.: Including geophysical data in ground water model inverse calibration, Ground Water, 41, 178–189, 2003.

Deutsch, C. V. and Journel, A. G. GSLIB: geostatistical software library and user's guide, 2nd Edn., Oxford University Press, 1998.

Fogg, G. E.: Transition Probability-Based Indicator Geostatistics, Math. Geol., 28, 453–476, 1996.

Geonics Limited: http://www.geonics.com/index.html (last access: 3 February 2014), 2012.

Härdle, K. W. and Simar, L.: Applied Multivariate Statistical Analysis, Springer, Berlin, Heidelberg, New York, doi:10.1007/978-3-642-17229-8, 2012.

He, X., Koch, J., Sonnenborg, T. O., Jørgensen, F., Schamper, C., and Refsgaard, J. C.: Uncertainties in constructing stochastic geological models using transition probability geostatistics and transient AEM data, Water Resour. Res., in review, 2013.

Herckenrath, D., Fiandaca, G., Auken, E., and Bauer-Gottwein, P.: Sequential and joint hydrogeophysical inversion using a field-scale groundwater model with ERT and TDEM data, Hydrol. Earth Syst. Sci., 17, 4043–4060, doi:10.5194/hess-17-4043-2013, 2013.

Hinnell, A. C., Ferr, T. P. A., Vrugt, J. A., Huisman, J. A., Moysey, S., Rings, J., and Kowalsky, M. B.: Improved extraction of hydrologic information from geophysical data through coupled hydrogeophysical inversion, Water Resour. Res., 46, W00D40, doi:10.1029/2008WR007060, 2010.

Hotelling, H.: Analysis of a complex of statistical variables into principal components, J. Educ. Psychol., 24, 417–441, 1933.

Høyer, A.-S., Jørgensen, F., Lykke-Andersen, H., and Christiansen, A. V.: Iterative modelling of AEM data based on geological a priori information from seismic and borehole data, Near Surf. Geophys., in press, 2014.

Jørgensen, F. and Sandersen, P. B. E.: Buried and open tunnel valleys in Denmark-erosion beneath multiple ice sheets, Quaternary Sci. Revi., 25, 1339–1363, 2006.

Jørgensen, F., Scheer, W., Thomsen, S., Sonnenborg, T. O., Hinsby, K., Wiederhold, H., Schamper, C., Burschil, T., Roth, B., Kirsch, R., and Auken, E.: Transboundary geophysical mapping of geological elements and salinity distribution critical for the assessment of future sea water intrusion in response to sea level rise, Hydrol. Earth Syst. Sci., 16, 1845–1862, doi:10.5194/hess-16-1845-2012, 2012.

Jørgensen, F., Møller, R. R., Nebel, L., Jensen, N., Christiansen, A. V., and Sandersen, P.: A method for cognitive 3D geological voxel modelling of AEM data, Bull. Eng. Geol. Environ., 72, 421–432, doi:10.1007/s10064-013-0487-2, 2013a.

Jørgensen, F., Sandersen, P. B. E., Høyer, A.-S., Pallesen, T. M., Foged, N., He, X., and Sonnenborg, T. O.: A 3D geological model from Jutland, Denmark: Combining modeling techniques to address variations in data density, data type, and geology, 125th Anniversary Annual Meeting, Denver, Colorado, USA, 2013b.

Jørgensen, F., Sandersen, P. B. E., Høyer, A.-S., Pallesen, T. M., Foged, N., He, X., and Sonnenborg, T. O.: A 3D geological model from Jutland, Denmark: Combining modeling techniques to address variations in data density, data type, and geology, 125th Anniversary Annual Meeting, Denver, Colorado, USA, 2013c.

Paasche, H., Tronicke, J., Holliger, K., Green, A. G., and Maurer, H.: Integration of diverse physical-property models: Subsurface zonation and petrophysical parameter estimation based on fuzzy c-means cluster analyses, Geophysics, 71, H33–H44, 2006.

Pebesma, E. J. and Wesseling, C. G.: Gstat: A Program for geostatistical Modelling, Prediction and Simulation, Comput. Geosci., 24, 17–31, 1998.

Raiber, M., White, P. A., Daughney, C. J., Tschritter, C., Davidson, P., and Bainbridge, S. E.: Three-dimensional geological modelling and multivariate statistical analysis of water chemistry data to analyse and visualise aquifer structure and groundwater composition in the Wairau Plain, Marlborough District, New Zealand, J. Hydrol., 436–437, 13–34, 2012.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◀◀ | ▶▶

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Refsgaard, J. C., Auken, E., Bamberg, C. A., Christensen, B. S. B., Clausen, T., Dalgaard, E., Effersø, F., Ernstsen, V., Gertz, F., Hansen, A. L., He, X., Jacobsen, B. H., Jensen, K. H., Jørgensen, F., Jørgensen, L. F., Koch, J., Nilsson, B., Petersen, C., De Schepper, G., Schamper, C., Sørensen, K. I., Therrien, R., Thirup, C., and Viezzoli, A.: Nitrate reduction in geologically heterogeneous catchments – A framework for assessing the scale of predictive capability of hydrological models, Sci. Total Environ., 468–469, 1278–1288, 2014.

Sandersen, P., Jørgensen, F., Larsen, N. K., Westergaard, J. H., and Auken, E.: Rapid tunnel-valley formation beneath the receding Late Weichselian ice sheet in Vendsyssel, Denmark, Boreas, 38, 834–851, doi:10.1111/j.1502-3885.2009.00105.x, 2009.

Schamper, C., Jørgensen, F., Auken, E., and Effersø, F.: Resolution of thin and shallow geological layers using airborne transient electromagnetics, Geophysics, submitted, 2013.

Schamper, C., Auken, E., and Sørensen, K. I.: Coil response inversion for very early time modeling of helicopter-borne time-domain EM data and mapping of near-surface geological layers, Geophys. Prospect., in press, 2014.

Seifert, D., Sonnenborg, T. O., Refsgaard, J. C., Højberg, A. L., and Troldborg, L.: Assessment of hydrological model predictive ability given multiple conceptual geological models, Water Resour. Res., 48, WR011149, doi:10.1029/2011WR011149, 2012.

Shevnin, V., Mousatov, A., Ryjov, A., and Delgado-Rodriquez, O.: Estimation of clay content in soil based on resistivity modelling and laboratory measurements, Geophys. Prospect., 55, 265–275, 2007.

Slater, L.: Near surface electrical characterization of hydraulic conductivity: From petrophysical properties to aquifer geometries – A review, Surv. Geophys., 28, 169–197, 2007.

Stafleu, J., Maljers, D., Gunnink, J. L., Menkovic, A., and Busschers, F. S.: 3D modelling of the shallow subsurface of Zeeland, the Netherlands, Geologie en Mijnbouw/Netherlands Journal of Geosciences, 90, 293–310, 2011.

Strebelle, S.: Conditional simulation of complex geological structures using multiple-point statistics, Math. Geol., 34, 1–21, 2002.

Triantafilis, J. and Buchanan, S. M.: Identifying common near-surface and subsurface stratigraphic units using EM34 signal data and fuzzy k-means analysis in the Darling River valley, Aust. J. Earth Sci., 56, 535–558, 2009.

Turner, A.: Challenges and trends for geological modelling and visualisation, Bull. Eng. Geol. Environ., 65, 109–127, 2006.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

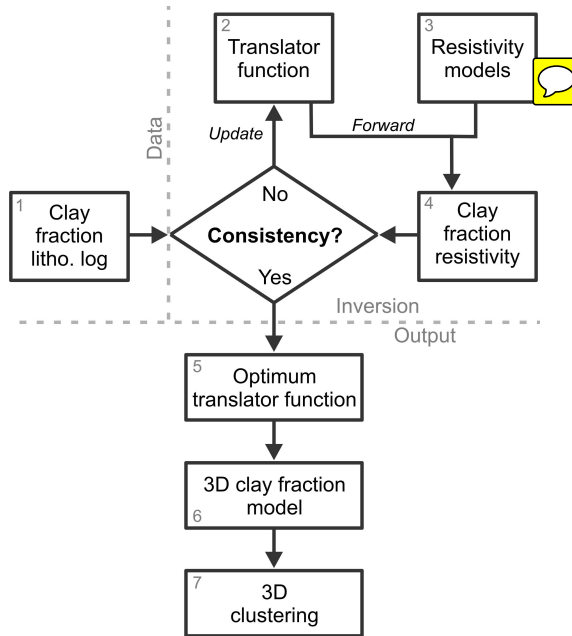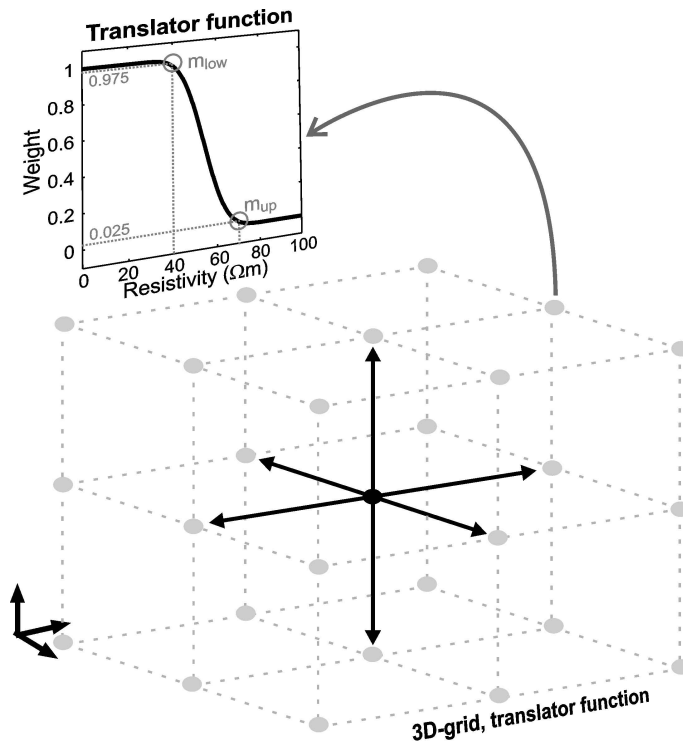Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Viezzoli, A., Christiansen, A. V., Auken, E., and Sørensen, K. I.: Quasi-3D modeling of airborne TEM data by Spatially Constrained Inversion, Geophysics, 73, F105–F113, 2008.

Waxman, M. H. and Smits, L. J. M.: Electrical Conductivities in Oil-Bearing Shaly Sands, Soc. Petrol. Eng. J., 8, 107–122, doi:10.2118/1863-A, 1968.

5 Wisén, R., Auken, E., and Dahlin, T.: Combination of 1D laterally constrained inversion and 2D smooth inversion of resistivity data with a priori data from boreholes, Near Surf. Geophys., 3, 71–79, 2005.

Wu, J.: Advances in K-means Clustering: A Data Mining Thinking, Springer, Heidelbberg, New York, Dordrecht, London, doi:10.1007/978-3-642-29807-3, 2012.

Fig. 1. Conceptual flowchart for the clay fraction concept and clustering.

Fig. 2. The translator function and 3-D translator function grid. The translator function returns a weight, W, between 0 and 1 for a given resistivity value. The translator function is defined by the two parameters m_{low}, and m_{up}. In this example the the the m_{low}, and m_{up} parameters correspond to 40 and 70 Ωm respectively. Each node in the 3-D translator function grid holds a set of m_{up} and m_{low}. The m_{up} and m_{low} parameters are constrained to all neighboring parameters as indicated with the black arrows for the black center node.

Title Page

Abstract · Introduction

Conclusions · References

Tables · Figures

|◄ · ►|

◄ · ►

Back · Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Fig. 3. The black square marks the Norsminde survey area.
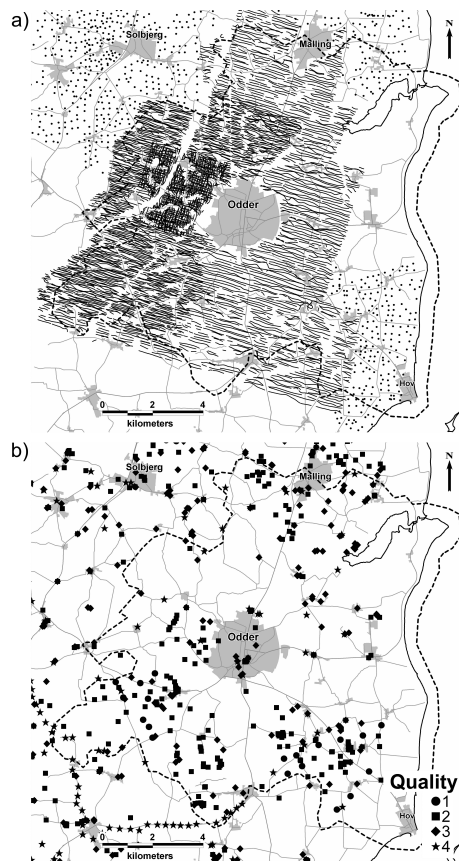
Fig. 4. (a) Location of the resistivity models. Small dots are SkyTEM models, larger scattered dots are ground-based models. **(b)** Borehole locations and borehole quality, where 1 corresponds to the highest quality and 4 to the lowest quality. The dashed line outlines the modeling area.
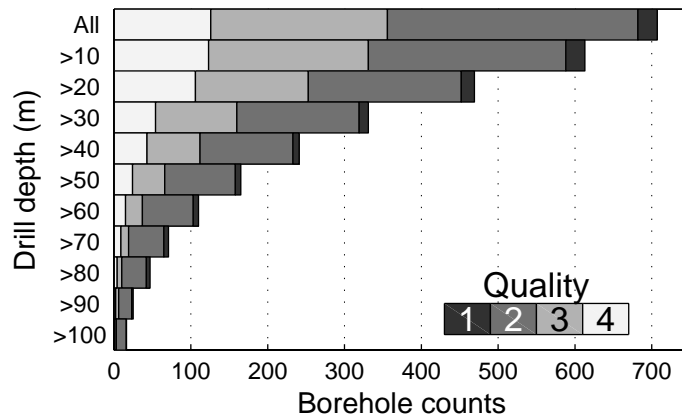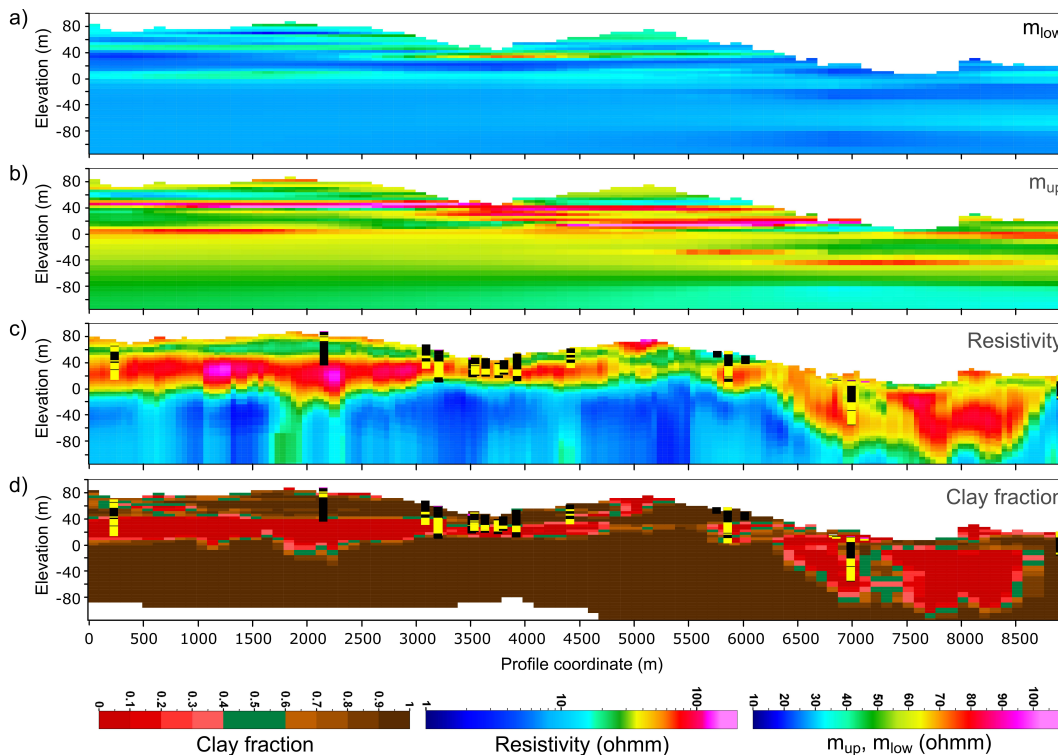
Fig. 5. Number of boreholes vs. drill depth. The bars show how many boreholes reach a certain depth. The color coding of the bars marks the quality grouping.

Fig. 6. Northwest–southeast cross sections (vertical exaggeration x6). Location and orientation of cross sections are marked in Fig. 7. **(a)** The m_{low} parameters of the translator function. **(b)** The m_{up} parameters of the translator function. **(c)** The resistivity section with boreholes within 200 m of the profile superimposed. Black borehole colors mark the clay layers, while yellow colors mark sand and gravel layers. **(d)** Clay fraction section and boreholes (same as plotted in the resistivity section).
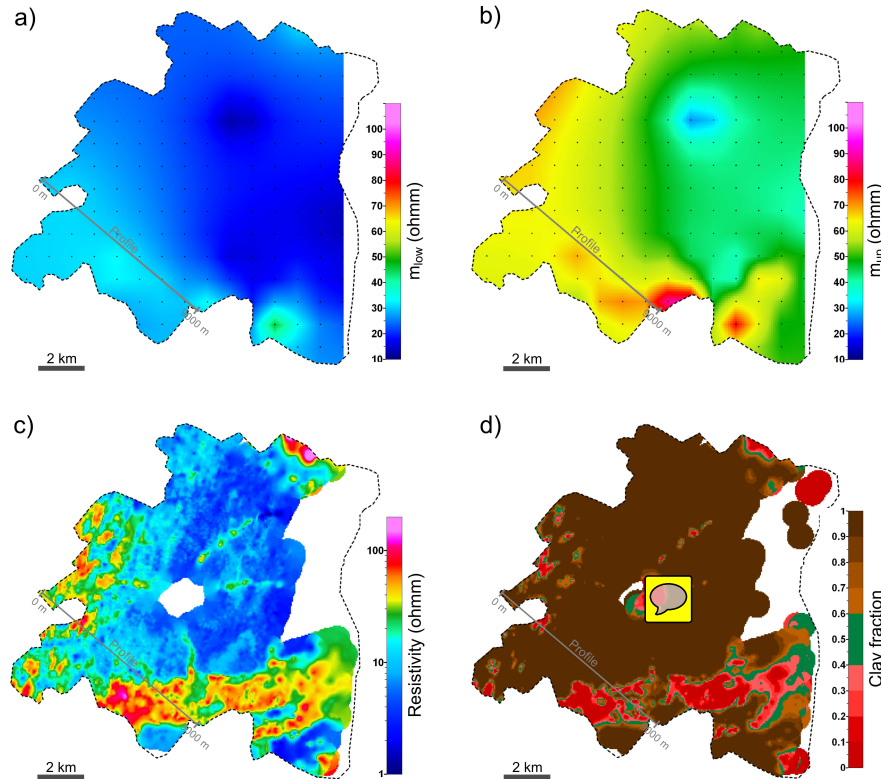
Fig. 7. Horizontal slices at 2 m b.s.l. cropped to the catchment area (dashed line). **(a)** The m_{low} parameters of the translator function superimposed with the 1 km translator function grid (black dots). **(b)** The m_{up} parameters of the translator function superimposed with the 1 km translator function grid (black dots). **(c)** Interpolated resistivity s[...] **(d)** Resulting CF- model.

Title Page

Abstract | Introduction
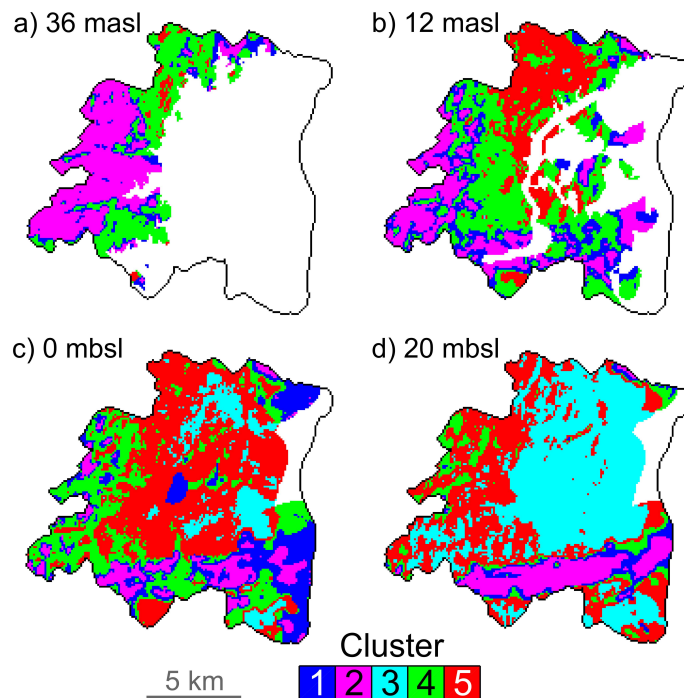
Conclusions | References
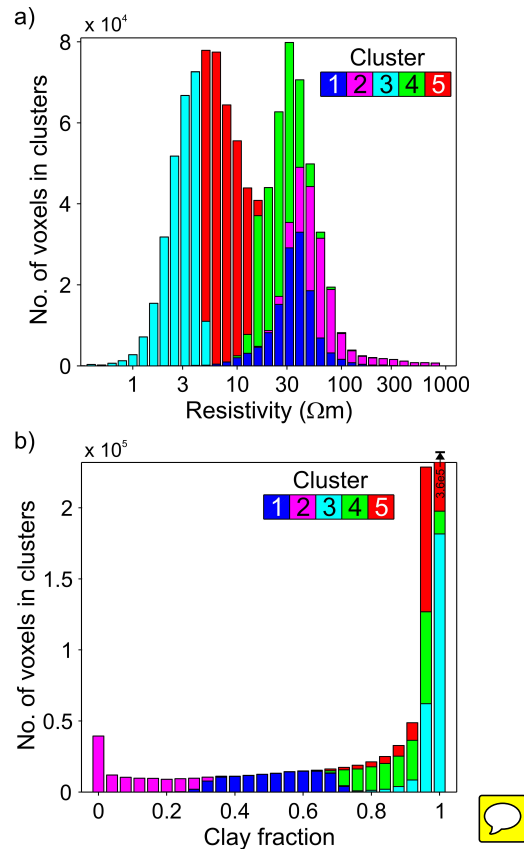
Tables | Figures

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Fig. 8. Horizontal slices in four depths of the 3-D cluster model.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Fig. 9. Cluster statistics. The histograms show which data from the original variables make up the five clusters. **(a)** The distribution of the resistivity data in the five clusters. **(b)** The distribution of the CF data in the five clusters.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion