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

Authors: Nikolaj Foged[1], Pernille Aabye Marker[2], Anders Vest Christansen[1], Peter Bauer-Gottwein[2], Flemming Jørgensen[3], Anne-Sophie Høyer[3], Esben Auken[1]

[1]HydroGeophysics Group, Department of Geoscience, Aarhus University, Denmark.

[2]Department of Environmental Engineering, Technical University of Denmark

[3]Geological Survey of Denmark and Greenland

ABSTRACT

We present an automatic method for parameterization of a 3D model of the subsurface, integrating lithological information from boreholes with resistivity models through an inverse optimization, with the objective of further detailing of 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 3D 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 3D model of the subsurface structures. We apply the procedure 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 3D model covering 156 km2. The final five-cluster 3D model differentiates between clay materials and different high resistive materials from information held in the 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 3D mapping, as they cover large areas in a short period of time. However, the geological and hydrogeological parameters are only mapped indirectly, and an interpretation of the airborne results is needed, often based on site-specific relationships. Linking electrical resistivity to hydrological properties is thus an area of increased interest as reviewed by Slater (2007).

Integrating geophysical models and borehole information has proved to be a powerful combination for 3D geological mapping (Jørgensen et al., 2012; Sandersen et al., 2009) and several modeling approaches have been reported. One way of building 3D-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 or by making models composed of structured or unstructured 3D 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 to 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 (Carle and 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. (2014). Geostatistical modeling approaches based primarily on borehole data often face the problem that the data are too sparse to represent the lateral heterogeneity at the desired spatial scale. Including geophysical data enables a more accurate estimation of the geostatistical properties especially laterally. This could be determination of the transition probabilities and the mean lengths of the different units. Though, the geophysical data also opens the question of to what degree the different data types should be honored 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. However, even though groundwater model predictions are sensitive to variations in the hydrostratigraphy, the groundwater model calibration is non-unique, and different hydrostratigraphic models may produce similar results (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 procedure for parameterization of a 3D model of the subsurface. The geological parameter we map is the clay fraction (CF). In this paper we 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* and it can easily be derived from the lithological logs. The clay fraction is then the cumulated thickness of *clay* layers in a depth interval divided with the length of the depth interval. The CF-procedure 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 3D 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 k-means clustering to form a structural 3D cluster model with the objective of further detailing for geological models or as direct input to groundwater models.

Lithological interpretation of a resistivity model is not trivial since the resistivity of a geological media is controlled by: porosity, pore water conductivity, degree of saturation, amount of clay minerals, etc. Different, primarily empirical, models try to explain the different phenomena, where Archie's law (Archie, 1942) is the most fundamental empirical model taking the porosity, pore water conductivity and, the degree of saturation into account, but does not account for electrical conduction of currents taking place on the surface of the clay minerals. The Waxman and Smits model (Waxman and Smits, 1968) together with the Dual Water model of Clavier et al. (1984) provides a fundamental basis for widely and repeatedly used empirical rules for shaly sands and material containing clay (e.g. Bussian, 1983; Sen, 1987; Revil and Glover, 1998). However, in a sedimentary depositional environment it can be assumed in general that clay or clay rich sediments will exhibit lower resistivities than the non-clay sediments, silt, sand, gravel, and chalk. As such, discrimination between clay and non-clay sediments based on resistivity models is feasible and the CF-value is a suitable parameter to work with in the integration of resistivity models and lithological logs. A 3D CF-model or clay/sand model will also contain key structural information for a groundwater model, since it delineates the impermeable clay units and the permeable sand/gravel units.

With the CF- procedure we use a two-parameter resistivity to CF translator function, which relies on the lithological logs providing the local information for the optimum resistivity to CF translation. Hence, we avoid describing the physical relationships underlying the resistivity images explicitly.

First, we give an overall introduction to the CF- procedure, 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}).

A key aspect in the CF-procedure is that the translator function can change horizontally and vertically adapting to the local conditions and borehole data. The calculation is carried out in a number of elevation intervals (calculation intervals) to cover an entire 3D model space. Having obtained the optimum and spatially distributed translator function we can transform the resistivity models to form a 3D clay fraction model, incorporating the key information from both the resistivity models and the lithological logs into one parameter. The CF-procedure is a further development to three dimensions of the accumulated clay thickness procedure by Christiansen et al., 2014, which is formulated in 2D.

The flowchart in Figure 1 provides an overview of the CF-procedure. 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 3D CF model (box 6).

The final step is a k-means clustering analysis (box 7). With the clustering we achieve a 3D 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 CF-procedure.

2.1 Observed data - lithological logs and clay fraction

The common parameter derived from the lithological logs and resistivity datasets is the clay fraction (Figure 1, boxes 1-4). The *clay fraction*, of a given depth interval in a borehole (named Ψ_{log}) is 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 Ψ_{log} in depth intervals for any lithological well log as the example in Figure 2a shows. Having retrieved the Ψ_{log} values we then need to estimate their uncertainties since a variance estimate, σ^2_{log} is needed in the evaluation of the misfit to Ψ_{res}.

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 the samples contaminated due to the travel time from the bottom to the surface. Other parameters affecting the uncertainty of the Ψ_{log} are sample intervals and sample density, accuracy of the geographical positioning and elevation, and the credibility of the driller to mention a few important ones..

2.2 Forward data – the translator function

For calculating the clay fraction for a resistivity model, Ψ_{res}, we use the translator function as shown in Figure 2b, which is defined by a m_{low} and a m_{up} parameter. With the CF-procedure we primarily want to determine resistivity threshold values for a clay-sand interpretation of the resistivity models. Thin geological layers are often not directly visible in the resistivity models, whereas they will most often appear in carefully described boreholes. The length of the calculation intervals reflects the resolution capability of the geophysical method of choice, which means that in some cases the calculation intervals contain both sand and clay layers when imposed on the lithological logs. The translator function must therefore be able to translate resistivity values as partly clay and partly sand to obtain consistency with the lithological logs. This is possible with the translator function in Figure 2b, where m_{low} and m_{up} represent the clay and sand cut-off values. So for resistivity values below m_{low} the layer is entirely clay (weight ≈ 1) and for resistivity values above m_{up} the layer is entirely sand or non-clay (weight ≈ 0).

Many functions fulfilling the above criteria could have been chosen, but we use the one shown because it is differentiable throughout while being flat at both ends and fully described by just two parameters. The translator function (*W(ρ)*) is mathematically a scaled complementary error function, defined as:

$$W(\rho) = 0.5 \cdot \mathrm{erfc}\left(\frac{K \cdot (2\rho - m_{up} - m_{low})}{(m_{up} - m_{low})}\right)$$

$$K = \mathrm{erfc}^{-1}(0.05) \qquad (1)$$

*w*here m_{low} and m_{up} are defined as the resistivity (ρ) at which the translator function, $W(\rho)$, returns a weight of 0.975 and 0.025 respectively (the *K*-value scales the erfc function accordingly). For a layered resistivity model the Ψ_{res} for a single resistivity model value in one calculation interval is then calculated as:

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

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 ($W \approx 0$). Figure 2a shows how a single resistivity model is translated into Ψ_{res} in numbers of calculation intervals.

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. (2014).

To allow for variation, laterally and vertically, in the resistivity to Ψ_{res} translation, a regular 3D grid is defined for the survey block (Figure 3). Each grid node holds one set of m_{up} and m_{low} parameters. The vertical discretization follows the clay fraction calculation intervals, varying between 4-20 m increasing with depth. The horizontal discretization is typically 0.5-2 km and a 2D 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 or no boreholes, horizontal and vertical smoothness constraints are applied between the translator functions at each node point as shown in Figure 3. Choosing appropriate constraints is based on the balance between fitting the data while having a reasonable model. The balance is site and data specific, but would typically be based on visual evaluations comparing the results against key boreholes. The smoothness constraints furthermore act as regularization and stabilize the inversion scheme.

Finally, we need to estimate Ψ_{res} values at the Ψ_{log} positions (named Ψ^{*}_{res}) for evaluation. We estimate the Ψ^{*}_{res} values by making a point kriging interpolation of the Ψ_{res} values and associated uncertainties within a search radius of typically 500 m. 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. Hence, for the output estimates of the Ψ^{*}_{res} both the original variance of Ψ_{res} and the variance on the kriging interpolation itself is included to provide total variance estimates of the Ψ^{*}_{res} values (σ^{2*}_{res}), which 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:

$$\delta\Psi_{obs} = \mathbf{G}\delta\mathbf{m}_{true} + \mathbf{e}_{log} \tag{3}$$

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 model parameters (m_{up}, m_{low}) of the true, but unknown, translator function and an arbitrary reference model (the initial starting model for the first iteration, then at later iterations the model from the previous iteration). e_{log} is the observational error, and G denotes the Jacobian matrix that contains the partial derivatives of the mapping. The general solution to the non-linear inversion problem of equation (3) is described by Christiansen et al. (2014) 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:

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

where N_{dat} is the number of Ψ_{log} values and σ^2_i is the combined variance of the *i*'th Ψ_{log} (σ^2_{log}) and Ψ_{res} (σ^{2*}_{res}) given as:

$$\sigma_i^2 = \sigma_{log,i}^2 + \sigma_{res,i}^{2*} \qquad (5)$$

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

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

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 equation (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 equation (7) similar to R_{con} in equation (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. (2014).

2.4 Cluster analysis

The delineation of the 3D 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 of minimizing 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 (Figure 4) around the town of Odder (Figure 5) 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 with 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 Neogene 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 Neogene 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. Today, the Jupiter database holds information about more than 240,000 boreholes. All borehole layers in the database are assigned a lithology code, which makes it easy to extract the different types of clay layers for the calculation of the Ψ_{log} values in the different calculation intervals.

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 system, where each borehole is rated from 1-4 (He et al., 2014). The ratings are used to assign different uncertainty (weights) to the lithological logs/the Ψ_{log} values in the CF-procedure.

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

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

- Vertical 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 assigned 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/or drilled as geophysical shot holes for seismic exploration.

The locations, quality ratings and drill depths of the boreholes are shown in Figure 5b. The drill depths and quality ratings are summarized in Figure 6. As the top bar in Figure 6 shows that 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 Figure 6. 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 (Figure 5a).

The SkyTEM data were collected with the newly developed SkyTEM^{101} system (Schamper et al., 2014b). 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., 2014a). The recorded times span the interval from ~3 µs 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 (Figure 5a). 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 1D resistivity models. The inversion was carried out in a spatially constrained inversion setup (Viezzoli et al., 2008) with a smooth 1D-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 individually at their 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 soundings 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.1D layered resistivity models with 3 to 5 layers were used in the interpretation of the TEM sounding data.

3.3 Model setup

The 3D 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 above sea level (asl) (highest surface elevation) to 120 m below sea level (bsl) l. The vertical discretization is 4 m for layers asl and 8 m for layers bsl, which results in 40 calculation intervals. Hence, in total, the model grid holds 16x18x40=11,520 translator functions each holding two parameters. Translator functions in the 3D grid situated above terrain, below DOI of the resistivity models, and outside geophysical coverage does not contribute at all, and are only included to make the translator function grid regular for easier computation/bookkeeping. The effective number of translator functions, is therefore close to 5,200.

The regularization constraints between neighboring translator functions nodes are set relatively loose to promote a predominantly data driven inversion problem. In this case we use horizontal constraint factors of 2 and vertical constraint factors of 3. This roughly allows the two parameters of the translator function to vary with a factor of 2 (horizontal) and a factor of 3 (vertical) relative to adjacent translator function parameters. The resulting variations in the translator model grid are a trade-off between data, data uncertainties and the constraints (equation (7)). A spatially uniform initial translator function was used with m_{low} = 35 Ωm and m_{up} = 55 Ωm.

To create the final regular 3D CF-model the Ψ_{res} values from the geophysical models, the Ψ_{log} values from the boreholes, and associated variances are used in a 2D-kriging interpolation for each calculation interval. The 2D-grids are then stacked to form the 3D-CF-model. The Ψ_{log} values are primarily used to close gaps in the resistivity dataset where boreholes are present, as seen for the large central hole in the resistivity survey (Figure 8b), which is partly closed in the CF-model domain (Figure 8d) by borehole information. In order to match the computational grid setup of a subsequent groundwater model, a horizontal discretization of 100 m is used for the 3D-CF-model grid. In this case the dense EM-airborne survey data could actually support a finer horizontal discretization (25-50 m) in the CF-model.

The k-means clustering is performed on two variables, the CT-model and resistivity model, in a 3D grid with regular horizontal discretization of 100 m and vertical discretization of 4 m between 96 and 0 m asl and 8 m between 0 and 120 m bsl. 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

CF-modeling results from the Norsminde area are presented in cross sections in Figure 7 and as horizontal slices in Figure 8. The total misfit of equation (7) is 0.37, but probably more interesting the isolated data fit (equation (3)) is 1.26 meaning that we fit the data almost to the level of the assigned noise. Figure 7a and b show the inversion results of the m_{low} and m_{up} parameters in section view. The vertical variation in the translator is pronounced in the resistivity transition zones, because sharp layer boundaries have a smoother representation in the resistivity domain.

For the deeper part of the model (deeper than 10 m bsl) the translator functions are less varying. This corresponds well to the general geological setting of the area with relatively homogenous clay sequences in the deeper part, but it is also a result of very limited borehole information for the deeper model parts. The general geological setting of the area is also clearly reflected in the translator function in the horizontal slices in Figure 8a and b. The eastern part of the area with lowest m_{low} values (dark blue in Figure 8a) and lowest m_{up} values (light blue/green in Figure 8b) corresponds to the area where the Palaeogene highly conductive clays are present. In the western part of the area the cross section intersects the glacial complex, where the clays are mostly tills, and higher m_{low} and m_{up} values are needed to get the optimum translation.

The resistivity cross section in Figure 7c and the slice section in Figure 8c reveal a detailed picture of the effect of the geological structures seen in the resistivity data. Generally, a good correlation with the boreholes is observed. Translating the resistivities we obtain the CF-model presented in Figure 7d and Figure 8d. 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 both clay and non-clay lithologies are present in the same calculation interval in a particular borehole.

Evaluating the result in Figure 7d and Figure 8d, 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 that the translation of resistivity to CF is not one-to-one. For example, the buried valley structure (profile coordinate 6500-8500m, Figure 7d) 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 7,000 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-procedure results in a high degree of consistency between the CF-results and the lithological logs, as seen in the CF-section in Figure 7d.

Horizontal slices of the 3D cluster model are shown in Figure 9. The near-surface part of the model (Figure 9a-b) are dominated by clusters 2 and 4, while the deeper parts of the model (Figure 9c-d) are dominated by clusters 3 and 5, with the east-west striking buried valley to the south, (Figure 9c), is primarily represented by clusters 1 and 2.

The histograms in Figure 10 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 10 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 (Figure 10a), while they are clearly separated in the CF-model space and span the entire interval (Figure 10b). The opposite is observed for clusters 3, 4, and 5, which are clearly separated in the resistivity space (Figure 10a), but strongly overlap in the CF-model space (Figure 10b).

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, are able to distinguish between clay types. This results in the two-part clustering of the low resistivity (>20 Ωm) values as seen in Figure 10a.

4 DISCUSSION

4.1 Translator function, grid and discretization

The spatially varying resistivity to CF translator function is the key to achieve consistency between the borehole information and the resistivity models, and the spatial variations of the translator model accounts for, at least, two main phenomena: 1) Changes in the resistivity-lithology petrophysical relationship, 2) The resolution capability in the geophysical results.

The first issue includes spatial changes in the pore water resistivity, the degree of water saturation, and/or contents of clay minerals for the sediments described lithologically as clay. The spatial variation in the pore water resistivity on this modeling scale is probably relatively smooth and small and will therefore only have a minor impact on the resistivity to lithology/clay fraction translation. Even in the case with larger fluctuations in the pore water resistivity (e.g. present of saline pore water) the translator function will automatic adapt to this as long as we have borehole information available that resembles the changes and the basic assumption that the clay rich formations are more conductive than coarse-grained sediments is fulfilled.

In the Norsminde area used in the case history the groundwater table is generally located a few meters below the surface and the groundwater is fresh. This means that the neither pore water resistivity nor the water saturation plays a major role for the resistivity-clay fraction relationship and thus the translator function. Though, in the case with a thicker unsaturated zone like for the pore water resistivity, the translator function will automatically adapt to this situation as long as borehole information is available.

The varying content of clay minerals in the lithologies described as *clay* will effect the translator model. The correlation between the clay mineral content and resistivity is quite strong and could be the key parameter instead of the simple clay fraction of this procedure, but it would require clay mineral content values available in boreholes on a large modeling scale, which is why we disregard this approach and use the intentionally simple definition of clay and clay fraction.

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 is relatively poor and decreasing with depth. Detailed geological layer sequences might only be represented by an average conductivity or only have a weak signature in the resistivity models. By allowing spatial variation in the translation we can, to some degree, resolve weak layer indications in the resistivity models by utilizing the vertically detailed structural information from the lithological logs via the translator function.

The resolution in the final CF-model is strongly correlated to the resolution in the resistivity model, since the resistivity dataset contribute with the majority of the information. In general EM-methods are sensitive to absolute changes in the electric conductivity, which makes the resolution in the low resistivity end superior to the resolution of high-resistivity contrasts. The diffusive behavior of EM-methods results in a decreasing horizontal and vertical resolution capability with depth, and the vertical resolution capability furthermore strongly depends on the layer sequence. A sequence of thin lithological layering may therefore be represented as a single resistivity layer with an average conductivity, which is obviously challenging for the geological interpretation. The horizontal resolution strongly depends on the sample/line density of the geophysical measurements, but the footprint of a single measurement sets the lower limit for the horizontal resolution. The Norsminde airborne SkyTEM survey is conducted with a very dense line spacing giving a very high lateral resolution, which could actually support a finer horizontal discretization (25-50 m) in the CF-model. The 100 m horizontal discretization of the CF-model and cluster-model was selected to match the computational grid setup of a subsequent groundwater model. A detailed overview of resolution capabilities of the Norsminde SkyTEM survey is given by Schamper et al. (2014b) including an extensive comparison to borehole data

The horizontal sampling of the translator function should in principle be able to reproduce the true (but unknown) variations in the resistivity to CF translation. Though it is primarily the borehole density and secondarily the complexity of the petrophysical relationship between clay and resistivity that dictates the needed horizontal sampling of the translator function. To our experience a horizontal discretization of the translator function grid of 1-2 km (linearly interpolated between nodes) is sufficient to obtain an acceptable consistency between the lithological logs and the translated resistivities. For the deeper part of the model domain where the borehole information is sparse, a coarser translator function grid would be sufficient.

Starting model values for the translator function in the inversion scheme become important in areas with very low borehole density, primarily the deeper part of the model domain. The starting model values are selected based on experience and by a visual comparison of the resistivity models to key lithological logs. The horizontal and vertical constraints to migrate some information from regions with many boreholes to regions with few boreholes or with no boreholes. As in most inversion tasks a few initial inversions are performed to fine-tune and to evaluate the effect of different starting models and constraints setup.

The CF-procedure supports both uncertainty estimates on the input data, on the output translator functions, and on the final CF-model. Generally, the uncertainties in the CF-model are closely related to the borehole density and quality, as well as resolution and density of the resistivity models. The calculation and estimation of input and output uncertainties is described en detail in Christiansen et al. (2014).

4.2 Clustering and validation

For the clustered 3D-model each cluster represents some unit with fairly uniform characteristics. It could be hydrostratigraphic units where the hydraulic conductivity of the cluster units is determined through a subsequent groundwater model calibration, typically constrained by hydrological head and discharge data. Groundwater model calibration of the Norsminde 3D-cluster model has been performed with a preliminary positive outcome, but more experiments are needed before drawing final conclusions. In this process one needs to evaluate the cluster validity, i.e. how many clusters the data can support. Cluster validity can be assessed with various statistical measures (e.g. Halkidi et al., 2002). The number of clusters resulting in the best hydrological performance might also be used as a measure of cluster validity. The validity of the clusters and the resulting groundwater model is still to be explored in more detail.

5 CONCLUSION

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

The CF-procedure combines lithological borehole information with geophysical resistivity models in producing large scale 3D 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 CF-procedure 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 CF-procedure furthermore handles uncertainties on both input and output data.

The CF-procedure 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 3D 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 3D-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-procedure and clustering we aim at building 3D 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.

The 3D clay fraction model can also be 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 Figure 7 and Figure 8 indicated. Integration and further development of the CF-model into more complex geological models have been carried out with success (Jørgensen et al., 2013b).

6 ACKNOWLEDGEMENTS

The research for this paper was carried out within the STAIR3D-project (funded by Geo-Center Danmark) and the HyGEM-project (funded by the Danish Council for Strategic Research under contract no. DSF **11-116763**). We also wish to thanks the NiCA research project (funded by the Danish Council for Strategic Research under contract no. DSF 09-067260) for granting access to the SkyTEM data for the Norsminde case and senior advisor at the Geological Survey of Denmark and Greenland (GEUS), Claus Ditlefsen, for his work and help with quality rating of the borehole data. Finally a great thanks to our colleagues, Ph.D. Casper Kirkegaard for help with optimization of the numerical code and Professor emeritus Niels Bøie Christensen for insightful comments on the uncertainty migration.

REFERENCES

Auken, E. and A. V. Christiansen, Layered and laterally constrained 2D inversion of resistivity data: Geophysics, 69, 3, 752-761, 2004.

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

Auken, E., A. V. Christiansen, C. Kirkegaard, G. Fiandaca, C. Schamper, A. A. Behroozmand, A. Binley, E. Nielsen, F. Effersø, N. B. Christensen, K. I. Sørensen, N. Foged, and G. Vignoli, An overview of a highly versatile forward and stable inverse algorithm for airborne, ground-based and borehole electromagnetic and electric data: Exploration Geophysics, 1-13, DOI: 10.1071/EG13097, 2014.

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

Bussian, A. E., Electrical conductance in a porous medium: Geophysics, 48, 9, 1258-1268, 1983.

Carle, S. F. and G. E. Fogg, Transition Probability-Based Indicator Geostatistics: Mathematical Geology, 28, 4, 453-476, 1996.

Christiansen, A. V. and E. Auken, A global measure for depth of investigation: Geophysics, 77, 4, WB171-WB177, 2012.

Christiansen, A. V., N. Foged, and E. Auken, A concept for calculating accumulated clay thickness from borehole lithological logs and resistivity models for nitrate vulnerability assessment: Journal of Applied Geophysics, 108, 69-77, 2014.

Clavier, C., G. Coates, and J. Dumanoir, Theoretical and experimental bases for the dual-water model for interpretation of shaly sands: Society of Petroleum Engineers Journal, 24, 2, 153-168, 1984.

Daly, C. and J. K. Caers, Multi-point geostatistics - an introductory overview: First Break, 28, 9, 39-47, 2010.

Dam, D. and S. Christensen, Including geophysical data in ground water model inverse calibration: Ground Water, 41, 2, 178-189, 2003.

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

Geonics Limited, http://www.geonics.com/index.html: 2012.

Halkidi, M., Y. Batistakis, and M. Vazirgiannis, Clustering validity checking methods: Part II: Sigmod Record, 31, 19-27, 2002.

Härdle, K. W. and L. Simar, Applied Multivariate Statistical Analysis: Springer,2012.

He, X., J. Koch, T. O. Sonnenborg, F. Jørgensen, C. Schamper, and J. C. Refsgaard, Transition probability based stochastic geological modeling using airborne geophysical data and borehole data: Water Resources Research, Special Issue on Patterns in Soil-Vegetation-Atmosphere Systems: Monitoring, Modeling and Data Assimilation, 1-23, 2014.

Herckenrath, D., G. Fiandaca, E. Auken, and P. Bauer-Gottwein, Sequential and joint hydrogeophysical inversion using a field-scale groundwater model with ERT and TDEM data: Hydrology and Earth System Sciences, 17, 4043-4060, 2013.

Hinnell, A. C., T. P. A. Ferre, J. A. Vrugt, J. A. Huisman, S. Moysey, J. Rings, and M. B. Kowalsky, Improved extraction of hydrologic information from geophysical data through coupled hydrogeophysical inversion: Water Resources Research, 46, 4, DOI: 10.1029/2008WR007060, 2010.

Hotelling, H., Analysis of a complex of statistical variables into principal components: Journal of Educational Psychology, 24, 417-441, 1933.

Høyer, A.-S., F. Jørgensen, H. Lykke-Andersen, and A. V. Christiansen, Iterative modelling of AEM data based on geological a priori information from seismic and borehole data: Near Surface Geophysics, 12, -DOI: 10.3997/1873-0604.2014024, 2014.

Jørgensen, F., R. R. Møller, L. Nebel, N. Jensen, A. V. Christiansen, and P. Sandersen, A method for cognitive 3D geological voxel modelling of AEM data: Bulletin of Engineering Geology and the Environment, 72, 3-4, 421-432, DOI: 10.1007/s10064-013-0487-2, 2013a.

Jørgensen, F. and P. B. E. Sandersen, Buried and open tunnel valleys in Denmark-erosion beneath multiple ice sheets: Quaternary Science Reviews, 25, 11-12, 1339-1363, 2006.

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

Jørgensen, F., W. Scheer, S. Thomsen, T. O. Sonnenborg, K. Hinsby, H. Wiederhold, C. Schamper, Roth.B., R. Kirsch, and E. Auken, Transboundary geophysical mapping of geological elements andsalinity distribution critical for the assessment of future sea waterintrusion in response to sea level rise: Hydrology andEarth SystemSciences, 16, 1845-1962, 2012.

Møller, I., H. Verner, V. H. Søndergaard, J. Flemming, E. Auken, and A. V. Christiansen, Integrated management and utilization of hydrogeophysical data on a national scale: Near Surface Geophysics, 7, 5-6, 647-659, 2009.

Paasche, H., J. Tronicke, K. Holliger, A. G. Green, and H. Maurer, 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 C. G. Wesseling, Gstat: A Program for geostatistical Modelling, Prediction and Simultation: Computers & Geosciences, 24, 1, 17-31, 1998.

Raiber, M., P. A. White, C. J. Daughney, C. Tschritter, P. Davidson, and S. E. Bainbridge, 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: Journal of Hydrology, 436-437, 13-34, 2012.

Rasmussen, E. S., K. Dybkjær, and S. Piasecki, Lithostratigraphy of the upper Oligocene - Miocene succession of Denmark: Geological Survey of Denmark and Greenland Bulletin, 22, 1-92, 2010.

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

Revil, A. and P. W. J. Glover, Nature of surface electrical conductivity in natural sands, sandstones, and clays: Geophysical Research Letters, 25, 5, 691-694, 1998.

Sandersen, P., F. Jørgensen, N. K. Larsen, J. H. Westergaard, and E. Auken, Rapid tunnel-valley formation beneath the receding Late Weichselian ice sheet in Vendsyssel, Denmark: BOREAS, 38, 4, 834-851, DOI: 10.1111/j.1502-3885.2009.00105.x, 2009.

Schamper, C., E. Auken, and K. I. Sørensen, Coil response inversion for very early time modelling of helicopter-borne time-domain electromagnetic data and mapping of near-surface geological layers: Geophysical Prospecting, 62, 3, 658-674, DOI: 10.1111/1365-2478.12104, 2014a.

Schamper, C., F. Jørgensen, E. Auken, and F. Effersø, Assessment of near-surface mapping capabilities by airborne transient electromagnetic data - An extensive comparison to conventional borehole data: Geophysics, 79, 4, doi: 10.1190/geo2013-0256.1, B187-B199, 2014b.

Seifert, D., T. O. Sonnenborg, J. C. Refsgaard, A. L. Højberg, and L. Troldborg, Assessment of hydrological model predictive ability given multiple conceptual geological models: Water Resources Research, 48, 6, DOI: 10.1029/2011WR011149, 2012.

Sen, P. N., Electrochemical origin of conduction in shaly formations: Society of Petroleum Engineers: Presented at 62nd Annual Technical Conference and Exhibition, 1987.

Slater, L., Near surface electrical characterization of hydraulic conductivity: From petrophysical properties to aquifer geometries - A review: Surveys in Geophysics, 28, 2-3, 169-197, 2007.

Stafleu, J., D. Maljers, J. L. Gunnink, A. Menkovic, and F. S. Busschers, 3D modelling of the shallow subsurface of Zeeland, the Netherlands: Geologie en Mijnbouw/Netherlands Journal of Geosciences, 90, 4, 293-310, 2011.

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

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

Turner, A., Challenges and trends for geological modelling and visualisation: Bulletin of Engineering Geology and the Environment, 65, 2, 109-127, 1-5-2006.

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

Waxman, M. H. and L. J. M. Smits, Electrical Conductivities in Oil-Bearing Shaly Sands: Society of Petroleum Engineers Journal, 8, 107-122, 1968.

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

Wu, j., Advances in K-means Clustering: A Data Mining Thinking: Springer,2012.

FIGURES AND FIGURE CAPTIONS

Figure 1

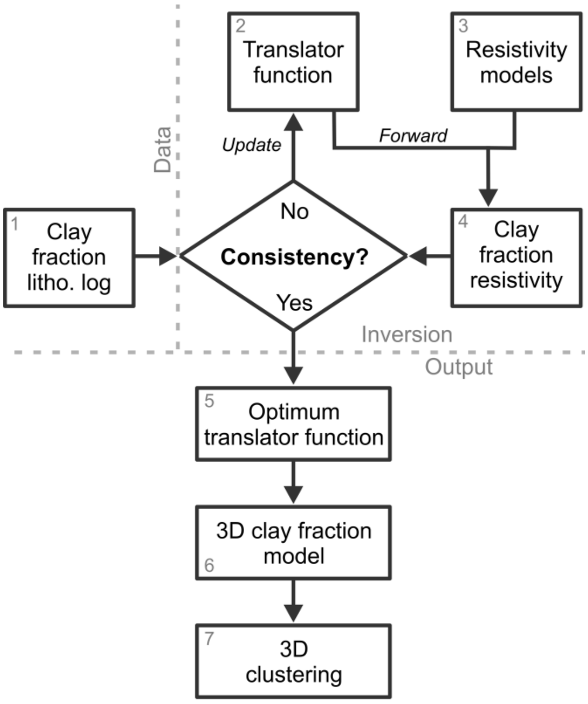

Figure 1. Conceptual flowchart for the CF-procedure and clustering

Figure 2

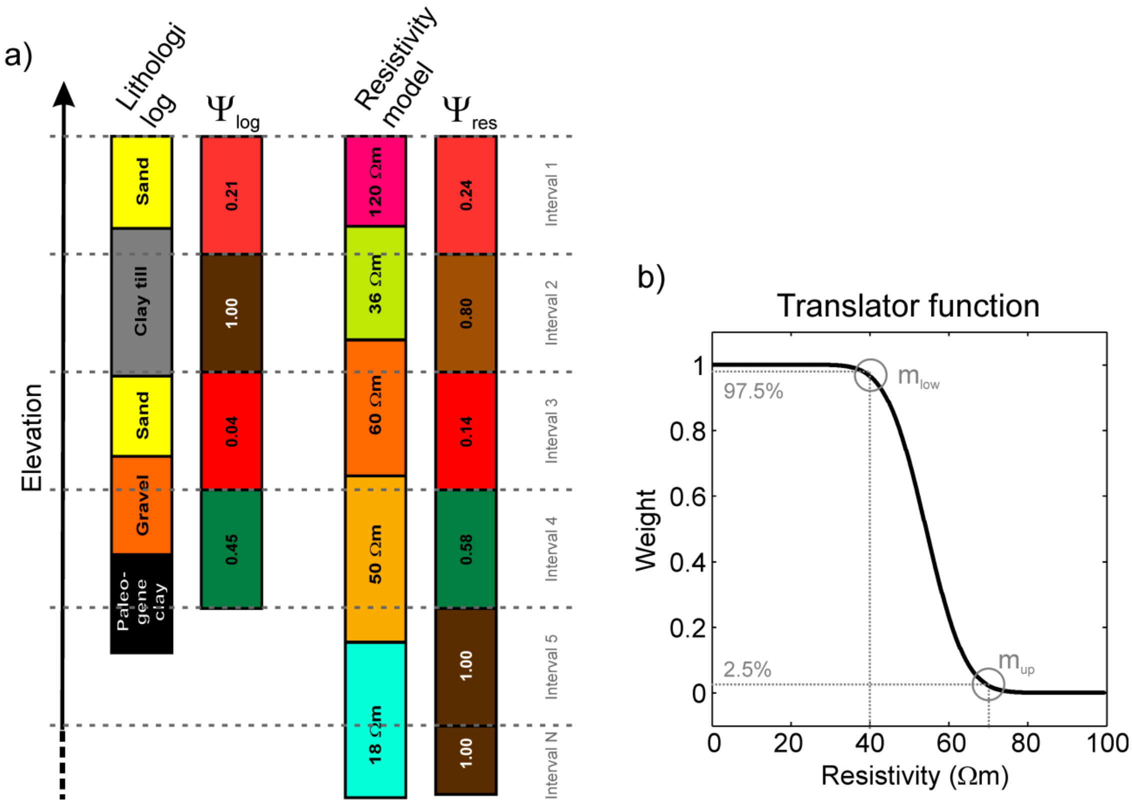

Figure 2. a) Example of how a lithological log translates into a Ψ_{log} and how a resistivity model translates into Ψ_{res}, for a numbers of calculation intervals. The resistivity values and the resulting clay fraction values are stated on the bars, but also indicated with colors with reference to the color scales of Figure 7. b) 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 m_{low}, and m_{up} parameters are 40 Ωm and 70 Ωm respectively.

Figure 3

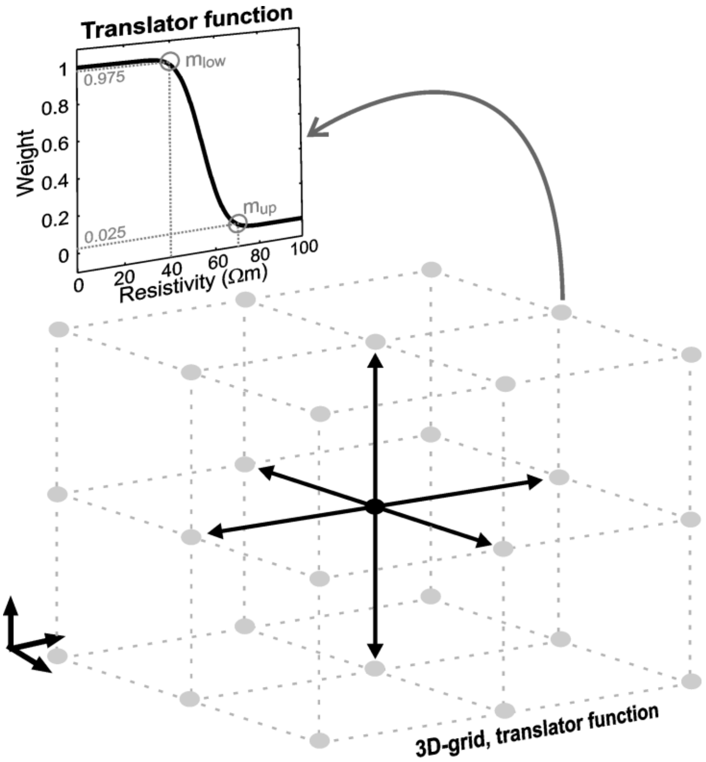

Figure 3. The translator function and 3D translator function grid. Each node in the 3D translator function grid holds a set of m_{up} and m_{low}. The m_{up} and m_{low} parameters are constrained to all neighboring parameters as indicated with the black arrows for the black center node.

Figure 4

Figure 4. The black square marks the Norsminde survey area.

Figure 5

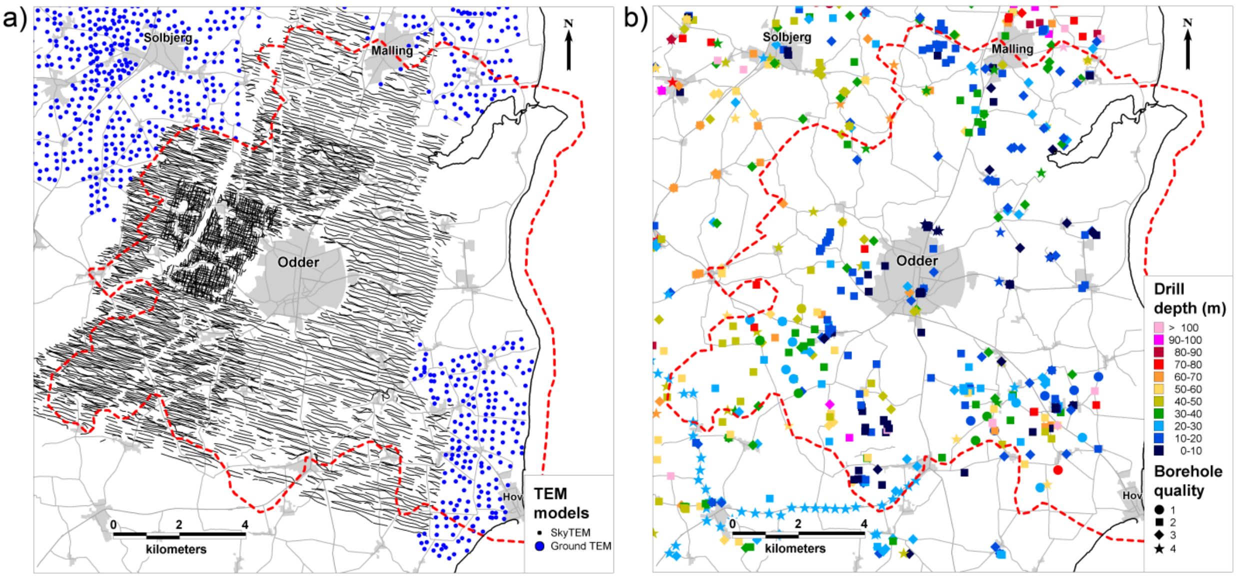

Figure 5. a) Resistivity model positions for the SkyTEM survey and the ground-based TEM soundings. b) Borehole locations, quality (shape), and drill depth (color). Quality 1 corresponds to the highest quality and 4 to the lowest quality. The red dashed line outlines the catchment area (156 km^2).

Figure 6

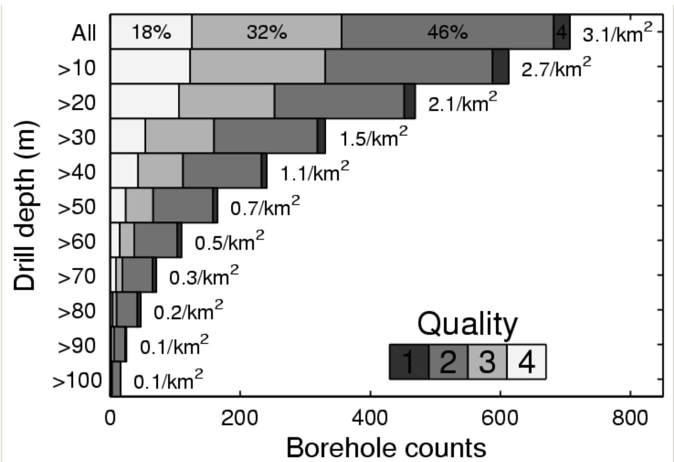

Figure 6. Number of boreholes vs. drill depth for the Norsminde survey area. The bars show how many boreholes reach a certain depth. The value to the right of the bars specifies the number of boreholes per km^2 at the different depths. The color coding of the bars marks the borehole quality grouping.

Figure 7

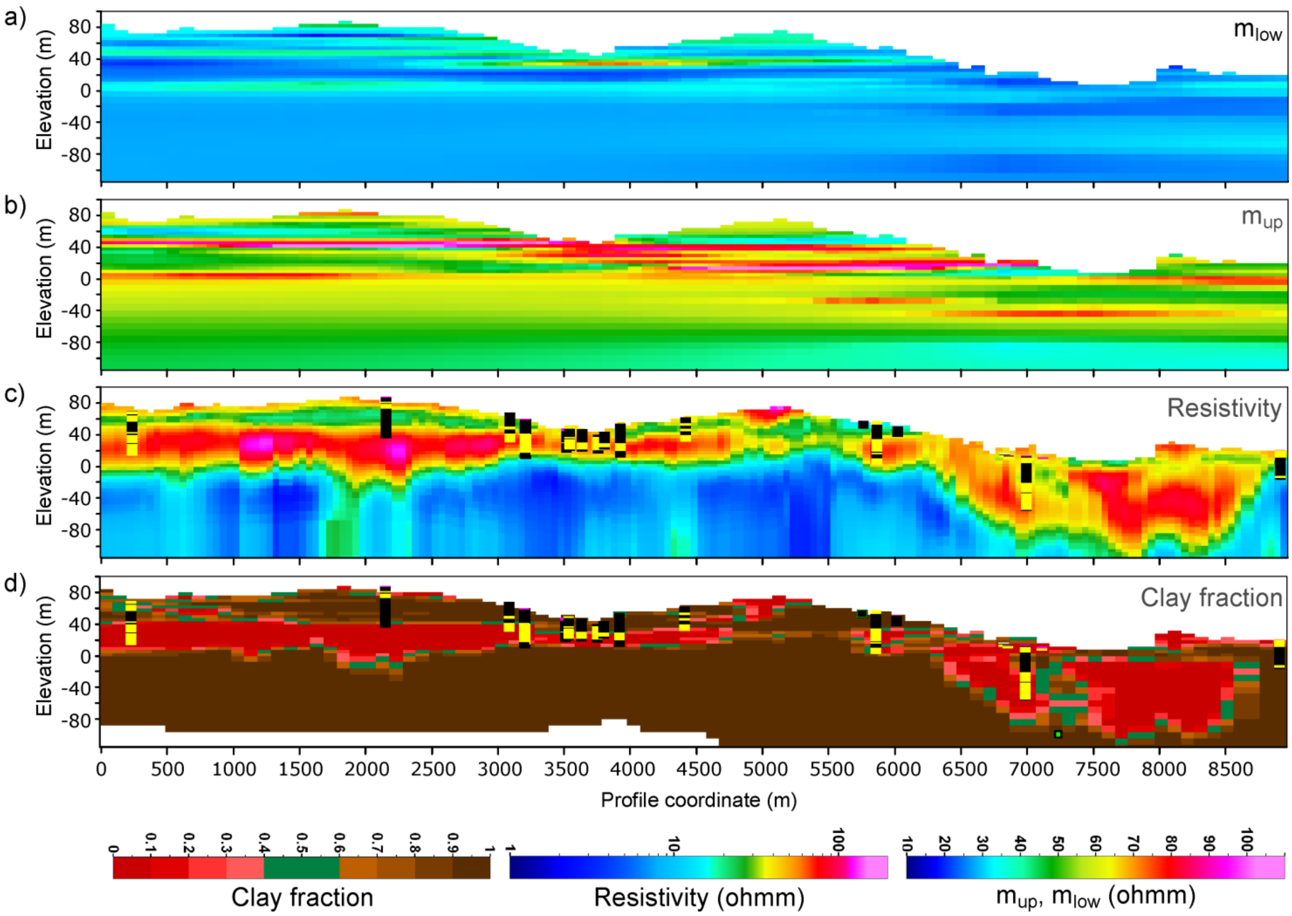

Figure 7. Northwest–southeast cross sections (vertical exaggeration x6). Location and orientation of the cross sections are marked in Figure 8. a) The m_{low} parameters of the translator function. b) The m_{up} parameters of the translator function. c) The resistivity section with boreholes within 200 m of the profile superimposed. Black and yellow vertical bars show the position of boreholes: Black blocks mark the clay layers, and yellow blocks mark sand and gravel layers. d) Clay fraction section and boreholes (same boreholes as plotted in the resistivity section).

Figure 8

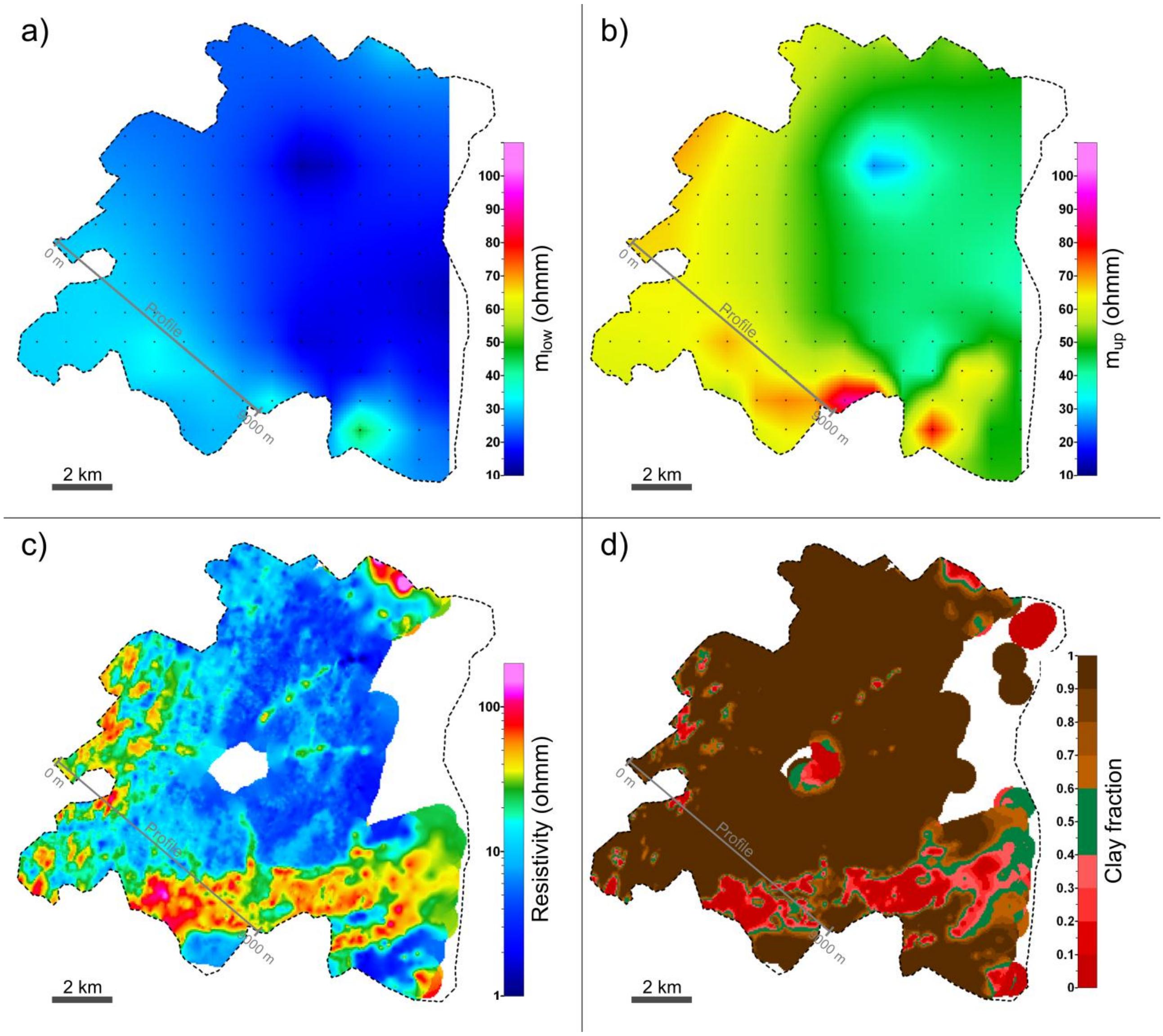

Figure 8. Horizontal slices at 2 m bsl cropped to the catchment area (dashed line). a) The m_{low} parameters of the translator function superimposed with the 1 km translator function grid (black dots). b) The m_{up} parameters of the translator function superimposed with the 1 km translator function grid (black dots). c) Resistivity slice (interpolated). Note that no EM-data is available around the town of Odder (see Figure 5a) resulting in a "hole" in the resistivity map. d) Resulting CF-model. The hole in the resistivity map is here partly closed because CF-values from boreholes are available in this area.

Figure 9

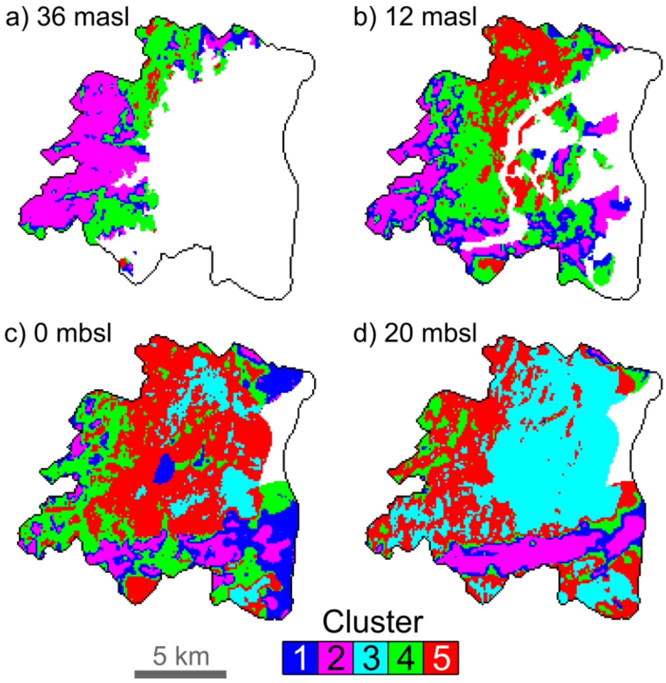

Figure 9. Horizontal slices in four depths of the 3D cluster model.

Figure 10

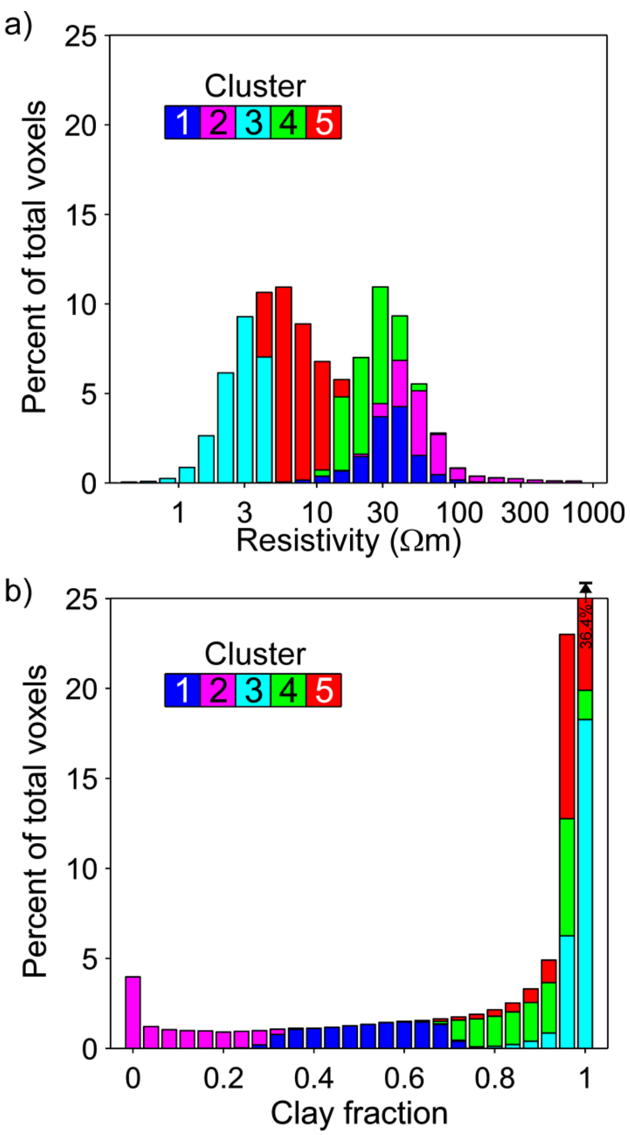

Figure 10. Cluster statistics. The histograms show which data from the original variables make up the five clusters. a) The distribution of the resistivity data in the five clusters. b) The distribution of the CF data in the five clusters.