Review 1

General comments

The effect of non-water saturated sediments and that of groundwater quality needs to be stated more explicit. I would think that there is data available from watersamples, well-logging or any other information, that provides information about the height of the watertable and confirms that the groundwater is fresh, and as such not a major factor in the resistivity.

Actually, the effect of saturation is quite substantial and we have added a detailed comment on this in the discussion. As long as the water saturated sand formation resistivity is higher than that of a “clay”-formation our basic assumption is not violated and the translator function will, ideally, adjust accordingly. In the specific case the pore water resistivity is sufficiently high that the clay layers are still the most conductive.

Specific comments

There are some issues in the paper I do not understand / are not clarified satisfactorily. One of the main issues is scale. The translator function is defined on a 1km grid and then applied to boreholes in order to obtain consistency between clay fraction from the lithology log and clay fraction from the resistivity models, Fig. 1 and 2.

On page 1468, the authors mention the procedure to define the translator function at the resistivity models, but the effects of the large distance between grid-node of the translator model and the resistivity models is not discussed.

The final model has a grid size of 100m x 100m, which is considerable more detailed than the translator model. The consequences of this difference in scale should be discussed.

We have added a detailed discussion on these relevant issues in the ´discussion’ section. The scale of the translator function is defined by the ‘scale’ of changes in the resistivity-clay translation, and these are generally thought to be slow. The resistivity data are used to described the actual positioning of clay and sand units in the entire volume regardless of the translation, and it therefore make sense to have a much denser grid here.

Besides that, the consistency comparison between clay fraction from the resistivity models and from the lithology logs (Fig. 1) involve some decisions about which borehole to use for the comparison. For example, is there a distance constraint used for comparing boreholes with nearest resistivity model?

I believe this is a misunderstanding of the concept. There is no such thing as ‘a closest resistivity model’. The comparison is done in the borehole positions based on values kriged from the resistivity positions. This is done exactly to avoid having to discuss direction, search radius etc. There is of course an effective search radius, but it is chosen so big (500 m) that several geophysical models contribute for most boreholes.

On page 1468, lines 14-18, the migration of the translator function to areas with few / no boreholes needs justification. The decision to do this is rather crucial for the resulting model and at least an attempt should be made to estimate the effects.

We agree that the choice of constraint strengths is important for the outcome. Setting the constraints very loose we would be able to (over-)fit most boreholes, but it would at the price of an unrealistic looking model. As we have no 'true model' to compare against the evaluation is done based on the classical balance between fitting the data while having a reasonable model. These evaluations are primarily based on visual evaluations comparing the results against key boreholes. A clarifying sentence have been added in the 'Methodology' section and detailed paragraphs are also added to the 'Discussion'.

Page 1468, lines 19-23, the procedure is explained for obtaining the clay-fraction from the resistivity model at the location of the borehole. Point kriging is used, and I would recommend that the authors make clear that this is carried out with keeping in mind the maximum correlation distance. Beyond that distance, the interpolation is merely a local averaging.

The is absolutely correct, but we find that this is going into too much detail, as we have references to the kriging method itself. If the reader is unfamiliar with kriging many other aspects would require a deeper discussion to be fulfilling.

The results, as displayed in Fig 6 and 7 are promising. It seems to confirm the general geology of the area, but there is no rigorous validation of the procedure, e.g. performing cross-validation (leaving boreholes out of the dataset, one by one, and comparing the estimate with the borehole data) to judge the performance. Another option would be to split the dataset (e.g. 20%-80%) and estimate the quality of the procedure on using 80% of the data on the remaining 20%. This would give the reader a better "feel" of the quality of the results.

I see the point, but I think that the reader would only be more confused. Though, we did not even report the data fit of the inversion result, which should have been there and we have added it now. The data fit is a significant number saying if the data (boreholes) can be fitted by the model suggested by the inversion process. It seems to me that the suggested approach requires that the boreholes are looked at as "hard information", which is contrary to the approach here assigning actual noise to the borehole descriptions. Also, given that the optimization is handled as an inversion problem removing parts of the data set does not make much sense in my opinion. We would get data fits at the removed data points a little worse than what we report here (1.26 – just outside the assigned noise), but how should that then be interpreted? It is similar to taking a schlumberger sounding (VES) and removing some data points and see if you can back-fit them with the remaining data. You can do that, but the fit would be a little poorer than having all the data. If you remove the insignificant data the effect would be small; if you remove crucial data the result would be worse. I am confident the result would be the same here – removing one borehole at a time we would see that the remaining boreholes would produce an almost equally good fit at the position of the missing borehole. A little bit worse as suggested by inversion theory, and we would not have learned much.

The results are defined in terms of clay-fraction: the fraction of the length of an interval that is clay. How would this convert to hydrological parameters?

More comments on this issue have been added to this, but is also on purpose not to dive too deep into this discussion as we are really trying to be general about the conceptual idea and not link it too tightly to a specific use (even though the hydrological modelling is obvious…)

The authors mention that, after clustering, the Norsminde are can be divided into sub-areas, with different hydrological parameters. Is there a way to use the results of the clay-fraction model directly into groundwater models?

See above

1 Detailed comments (annotations in PDF-document)

Page, Line	***Review remarks***	***Authors response***
1462,26	What does this mean in the context of 3D mapping?	Rephrased
1463,8	layers = surfaces; so what do you mean?	Corrected
1463,24	not proper English	Rephrased
1463,25	what is meant by geostatistical properties?	Rephrased
1463,26	explain what is meant by hard and soft data. Does not occur in the manuscript after this	Rephrased
1464,5-8	What do you want to say? It is not clear what this sentence means.	We have rephrased this sentence
1465,1	c or k?	K-mean (type-setting error "K" should not be italic)
1465,16	Add "Established"	Sentence rephrased.
1465,21	In Fig. 1, the resistivity models are not listed as data, but it is data, isn't it?	From and inversion point of view the resistivity models are not "data" in the concept. The data (observed data) that are fitted during the inversion are CF-data of the boreholes. The resistivity models is a part of the forward response (forward data) as described in section 2.2. The labels in fig. 1 is therefore correct.
1466, 27	statistical variance is denoted as sigma^2, sigma = standard deviation	Agree. Corrected throughout the paper incl. in formulas.
1467, 7	sediment?	No change, we believe it is clear as it is.
1467, 20	Not all parameters are described / explained: K, rho	K is defined in equation 1., but we have clarified the text.
1468, 7	reference not very satisfacory: in review	Agree, but … The referenced paper is in print (proofread recently), and there is no good alternative reference.

1468, 11	horizontal discretization? 1km?	Rephrased: "The horizontal discretization is typically 500-1000 m and a 2D bilinear horizontal interpolation of ...."
1468, 18	This is a rather tricky business, migrating to areas without supporting data. You need to justify this!	It is true that it is tricky business to setup constraints that migrate information to less data dense areas. Here, it is merely a statement on how the inversion works, but we added a short extra sentence and addressed the question in more general terms in the discussion section.
	Kriging is not taken the spatial variance into account but uses the spatial correlation (as captured in the variogram) to estimate spatial interpolation variance. Except, when you mean that you are using "kriging with uncertain data", in that case it should be stated explicitly. You probably mean the spatial variation	"kriging with uncertain data" is used in this case. Paragraph is rephrased to make it clear.
1468, 18	How? See previous remark!	We believe this is covered by the stated reference for the used kriging code (Pebesma and Wesseling, 1998)
1469, 14-15	is this the standard deviation of the variance?. this means the standard deviation!	Corrected, see also authors response 1466, 27
1472, 14	superfluous remark	Removed
1472, 18	length?	Corrected to "calculation intervals" to be consistent with the concept explanation in section 2
1472, 28	you mean the vertical density?	Yes. Corrected to "Vertical sample density"
1473, 7	are often drilled for the purpose of	Rephrased
1474, 23	Are this factors that are in line with other studies / experiences? They do not mean anything to me	Paragraph extended and rephrased to add some qualitative statements Comment: Since the constrains are specified directly on the translator model parameters there is nothing to compare with as this is the first time the concept is presented.
1474, 27	sentence is not correct: "through subsequent test-inversions"	Corrected
1475, 7	I do not understand this! what does "are included" mean?	The sentence have been rephrased and extended substantially.
1475, 9	How do you obtain this 100m model? Your input is the resistivity models, converetd to clay fraction, I assume. Some kind of interpolation? Which technique?	Valid point - the explanation is heavily extended on this part.
1475, 21	Of course these are smooth, because they originate from a 1km grid	Rephrased for clarity

1476, 8	What is the height of the water-table in the area?	This is relevant, but the whole idea is to address many different issues with one parameter. More justification has been added in the introduction and in particular in the discussion.
1476, 9	This is quite troublesome, since this would also have an effect on all previous calculations! What is known about the groundwater quality / salinity?	This issue is know elaborated in the Discussion section. Regarding salinity: Saltwater intrusion is not a of major concern for the Norsminde area since the clay sequence extends almost to the surface in the coastal area.
1476, 18	What do you mean by "correct" ?	Rephrased
1476, 22	With	Corrected
1476, 27	Although you will have layers that cross the discretisation interval, with part in one interval and part in the lower lying interval. This also causes non-binary intervals.	Rephrased for clarity
1477, 9	Well., this is only one section and a visual inspection of the results. I would like to see a more rigorous comparison, e.g. cross-validation, see general comments.	See comment under general comments above
1477, 27	insert: "are able to"	Corrected
1481, 21	Replication	Corrected
Fig.1	resistivity model is also data?	See 1465,21
Fig.2	What is the spatial lay-out of the resistivity models, compared to this layout of the translator function? Gives an idea of the scale differences	Fig. 2 is a principle sketch for the translator function grid and constraints. For the Case story the layout of the EM-survey/resistivity model is described in section 3.2. The setup of the translator function grid for the case story is specified in section 3.3 and the horizontal node discretization can be see e.g. in fig. 7a.
Fig.7	How come there is a CF model while ther ei sno resistivity?	See 1475, 7
Fig. 7	First time you mention that resistivity is interpolated	It is only for presenting a resistivity slice that the resistivity value has been interpolated. The CF-concept do not use interpolated resistivities as input as described in section 2.2. Fig. label updated to: "Resistivity slice (interpolated)"

Fig.9	Why not use relative frequency on the y-axis? No. of voxels is not very informative. And how does this compare to the borehole data? Is the frequency similar?	Y-axis: Agree. Figure axis change to "percent of voxels". 1) The distribution of the borehole CF values are not really comparable with the resistivity CF-values distribution, since the sampling of the model space is heavily biased to wards the near surface and non clay areas for the Boreholes. The drills are also typically ending when reaching the pre-quaternary low resistivity "bottom" clay layer. 2) The borehole CF-values dose not end up in a cluster!.

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 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 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 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 3D model covering 156 km2. The final five-cluster 3D model differentiates between clay materials and different high resistive materials from information held in resistivity model and borehole observations respectively.

2 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 interpretation of the airborne results is needed, which is often based on site-specific relationships. Linking electrical resisvivity to hydrological properties is therefore 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 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. (2014). Geostatistical modeling approaches based primarily on borehole data often faces the problem that the data

Deleted: Therefore e.g. the link
Deleted: between
Deleted: and electrical properties has
Deleted: been
Deleted: separated by surfaces
Deleted: He et al.
Deleted: 2013
Deleted: The use of only borehole data in g
Deleted: problems

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 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 method for parameterization of a 3D 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. 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 log. 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 3D CF-model holds information from the resistivity dataset and the borehole dataset in one variable. As a last step we cluster

Deleted: especially laterally

Deleted: ,

Deleted: but

Deleted: of what to use as *hard* and *soft* data

Deleted: Jørgensen et al., 2013b

Deleted: Høyer et al., 2014

Deleted: While m

Deleted: y, non-uniqueness with respect to hydrostratigraphy is inherent to groundwater models

Deleted: ¶

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-concept we uses 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 parameters explaining the resistivity images explicitly.

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.

3 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 concept 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-concept is a further development to three dimensions of the accumulated clay thickness concept by Christiansen et al., 2014, which is formulated in 2D.

The flowchart in Figure 1 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 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 concept.

3.1 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

Deleted: We do this for

Deleted: that are mutually constrained

Deleted:

Deleted: mized

Deleted: Christiansen et al., 2013

Deleted: Christiansen et al., 2014

Deleted: 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*.

Deleted: Ψ_{log},

Deleted: therefore

Deleted: the

Deleted: clay fraction

Deleted: The drilling method is one of the key parameters affecting the uncertainty of the well log data. ¶

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.

3.2 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. m_{low} and m_{up} represents the clay and sand cut-off values. So for resistivity values below m_{low} the layer is counted as clay ($W \approx 1$) and for resistivity values above m_{up} the layer is counted as sand ($W \approx 0$). The translator function ($W(\rho)$) is ~~mathematically~~ a scaled complementary error function, defined as:

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

$$K = \operatorname{erfc}^{-1}(0.0025 \cdot 2)$$

(1)

where 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:

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

(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 a 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

Deleted: ,

Deleted: a simple two-parameter

Deleted: The translator function is described fully by

Deleted: ,

Deleted: an

Deleted:

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, typically 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 boreholes or with 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.

3.3 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

Deleted: Christiansen et al., 2013

Deleted: Christiansen et al.

Deleted: 2014

Deleted:

Deleted:

Deleted: 1

Deleted: intervals. A

Deleted: .

Deleted: using

Deleted: By using kriging for interpolation the spatial variance of Ψ_{res} is taken into account

Deleted: , and even more important, it provides uncertainty estimates (σ^{*}_{res}) of the Ψ^{*}_{res} values, which include

Deleted: uncertainty

Deleted: . These uncertainty estimates

Deleted:

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).

Deleted: Christiansen et al. (2013)

Deleted: Christiansen et al.

Deleted: 2014

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:

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

Deleted: $(\sigma_i)^2$

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:

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

Deleted: σ_i

Deleted: $\sqrt{\sigma\ {}^{2}_{log,i}\ \sigma_{log,i}{}^2 + \sigma^{*}_{res,i}{}^2}$

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

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

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 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:

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

(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).

Deleted: Christiansen et al.(2013)
Deleted: Christiansen et al.
Deleted: 2014
Deleted: .

3.4 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 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.

4 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 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 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.

4.1 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. For the lithological logs a fixed lithology code list is available and the different types of clay layers are easily identified, and the Ψ_{log} values for the calculation intervals can be calculated.

Deleted: Similar databases are maintained in other countries.

Deleted: all

Deleted: desired elevation 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 apply the lithological logs with uncertainty (weights) used in the inversion.

Deleted: He et al., 2013

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

Deleted: S

- 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/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, 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.

Deleted: and

Deleted: , while the

Deleted: t

4.2 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., 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 ~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 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 1D layered resistivity model with 3 to 5 layers depending on the number of layers needed to fit the data.

4.3 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 masl (highest surface elevation) to 120 mbsl. The vertical discretization is 4 m above sea level and 8 m below sea level, 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 model nodes are set relatively loose to promote a predominantly data driven inversion problem. In this case we uses horizontal constraint factors of 2 and vertical constraint factors of 3. This roughly corresponds to allowed translator parameter variations of a factor of 2 (horizontal) and a factor of 3 (vertical) relative to adjacent translator parameters. The resulting variations in the translator models grid is 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 masl and 8 m between 0 and 120 mbsl. 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

Deleted: Node points in the t

Deleted: regularization constraints between neighboring nodes are set to a factor of

Deleted: , while the horizontal

Deleted: regularization

Deleted: is set to a

Deleted: The Sstarting model and the constraints setup are based on a visual comparison of the resistivity models compared to key lithological logs combined with experience and the expected geological variability and fine-tuned through a subsequent of test-inversions.

Deleted: Node points in the translator function grid situated in major data gaps (above terrain, below DOI of the resistivity models, and outside geophysical coverage) dose not come in to play at all, and are only included to made the translator function grid regular for easier computation/bookkeeping. 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 are therefore approximately 5,200

Deleted: In the interpolation to make the regular 3D CF-model, Ψ_{log} values are included together with the Ψ_{res} values 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.

4.4 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 (below elevation -10 m) 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 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.

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

Deleted: and discussion

Deleted: The variations in the translator function are relatively smooth, especially in the

Deleted:

Deleted: . The smooth

Deleted: for the deeper part

Deleted: due to

Deleted: ¶
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 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.

Deleted: to

Deleted: if more

Deleted: cal

Deleted: layers

Deleted:

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 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 spanning 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 overlapping 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~~)~~.

5 DISCUSSION

Deleted: ¶
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 (Figure 9Figure 9a).¶

5.1 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. Though, in the case of saline pore water, the pore water resistivity needs to be taken into account in the interpretation. This is particularly important in the (rare) case where the presence of saline pore water might violate the basic assumption that clay rich formations are more conductive than coarse-grained sediments.

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 fration of this concept, 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 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 becomes 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-concept 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).

5.2 Clustering and validation

For the clustered 3D-model each cluster represent some unit with fairly uniform characteristics. It could be hydrostratigraphic units where the hydraulic conductivity of the cluster units are 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.

6 CONCLUSION

Deleted: AND OUTLOOK

We have presented a concept to produce 3D 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 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 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 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-concept 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~~en~~ 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).

Deleted: 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 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.

Deleted: (Jørgensen et al., 2013c)

7 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, 2004, Layered and laterally constrained 2D inversion of resistivity data: Geophysics, 69, 3, 752-761.

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

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, 2014, An overview of a highly versatile forward and stable inverse algorithm for airborne, ground-based and borehole electric and electromagnetic data: Exploration Geophysics, Submitted.

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

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

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

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

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

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

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

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

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

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

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

He, X., J. Koch, T. O. Sonnenborg, F. Jørgensen, C. Schamper, and J. C. Refsgaard, 2014, 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.

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

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

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

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

Jørgensen, F., R. R. Møller, L. Nebel, N. Jensen, A. V. Christiansen, and P. Sandersen, 2013a, 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.

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

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

Jørgensen, F., W. Scheer, S. Thomsen, T. O. Sonnenborg, K. Hinsby, H. Wiederhold, C. Schamper, Roth.B., R. Kirsch, and E. Auken, 2012, 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.

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

Pebesma, E. J. and C. G. Wesseling, 1998, Gstat: A Program for geostatistical Modelling, Prediction and Simultation: Computers & Geosciences, 24, 1, 17-31.

Raiber, M., P. A. White, C. J. Daughney, C. Tschritter, P. Davidson, and S. E. Bainbridge, 2012, 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.

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, 2014, Nitrate reduction in geologically heterogeneous catchments - A framework for assessing the scale of predictive capability of hydrological models: ScienceDirect, 468-469, 1278-1288.

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

Sandersen, P., F. Jørgensen, N. K. Larsen, J. H. Westergaard, and E. Auken, 2009, 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.

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

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

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

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

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

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

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

Triantafilis, J. and S. M. Buchanan, 2009, 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.

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

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

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

Wisén, R., E. Auken, and T. Dahlin, 2005, 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.

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

FIGURES AND FIGURE CAPTIONS

Figure 1

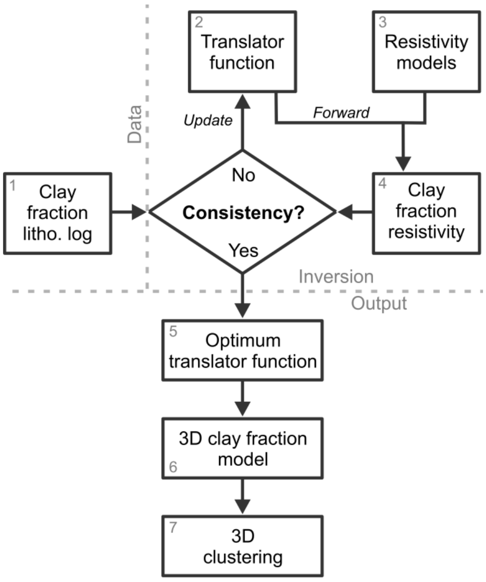

Figure 1. Conceptual flowchart for the clay fraction concept and clustering.

Figure 2

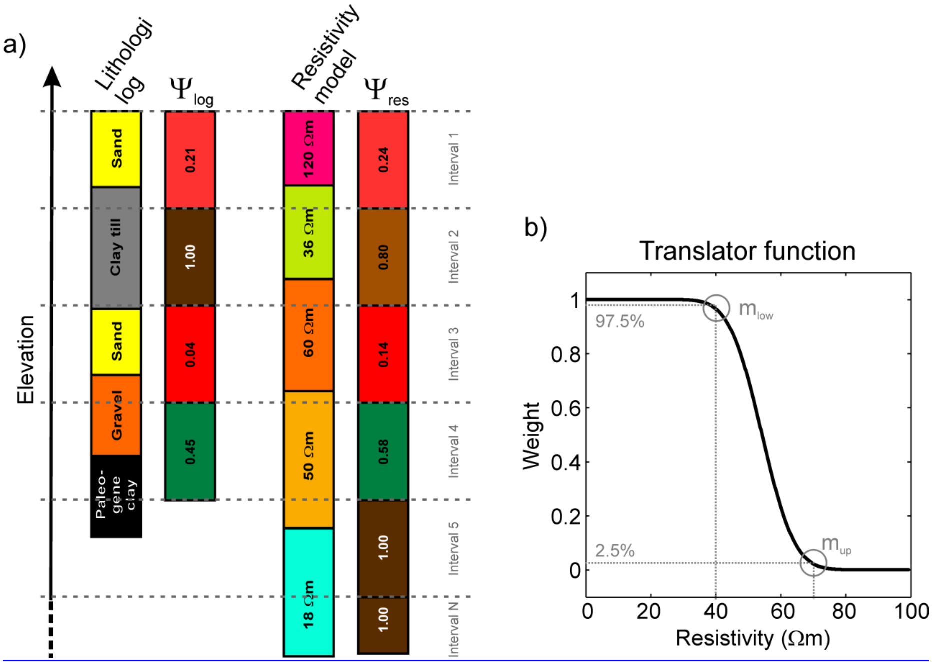

Figure 2. a) Example on how a lithological log is translates into Ψ_{log} and how a resistivity model translates into Ψ_{res}, for a numbers of calculation intervals. The resistivity values and the resulting clay fraction values are stated on the bars, but also indicated with colors with reference to 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.

Moved (insertion) [1]

Deleted: 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 m_{low}, and m_{up} parameters correspond to 40 Ωm and 70 Ωm respectively

Figure 3

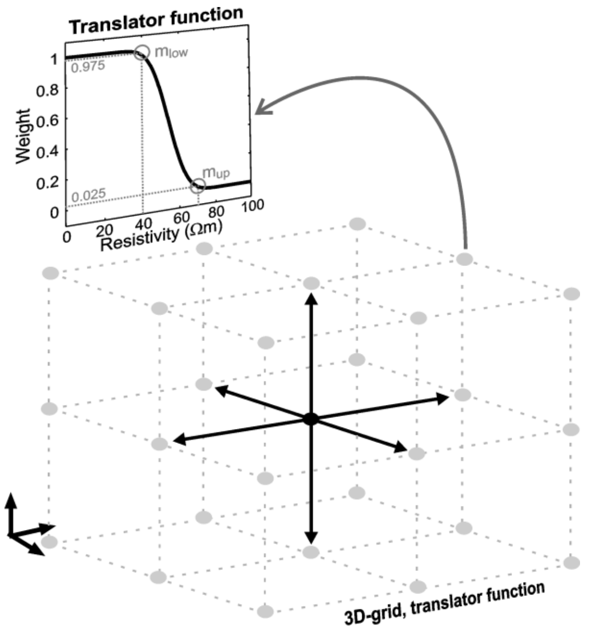

Figure 3. The translator function and 3D translator function grid. Each node in the 3D translator function grid holds a set of m_{up} and m_{low}. The m_{up} and m_{low} parameters are constrained to all neighboring parameters as indicated with the black arrows for the black center node.

Deleted: 2

Deleted: 32

Moved up [1]: 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 m_{low}, and m_{up} parameters correspond to 40 Ωm and 70 Ωm respectively

Deleted: .

Figure 4

Deleted: 3

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

Deleted: 43

Figure 5

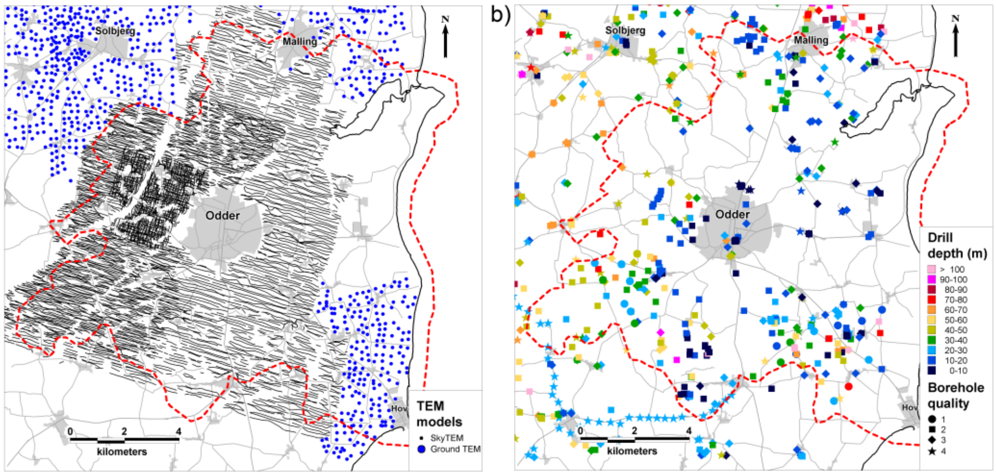

Figure 5, a) Resistivity model posistions for the SkyTEM survey and the ground-based TEM soundings, b) Borehole locations, quality (shape), and drill depth (color). Quality 1 corresponds to the highest quality and 4 to the lowest quality. The red dashed line outlines the catchment area (156 km^2).

Figure 6

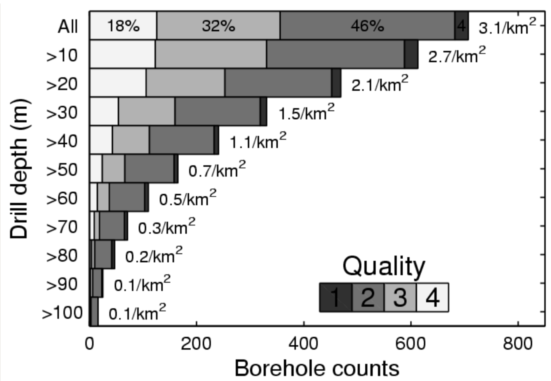

Figure 6, Number of boreholes vs. drill depth. The bars show how many boreholes reach a certain depth. The value to the right of the bars state the number of boreholes per km^2 for minimum depth of the interval. The color coding of the bars marks the quality grouping.

Deleted: 4

Deleted: 54

Deleted: Location of the r

Deleted: s

Deleted: s

Deleted: . Small dots are

Deleted: models, larger scattered dots

Deleted: are

Deleted: models

Deleted: and borehole

Deleted: where

Deleted: modeling area

Deleted: 5

Deleted: 65

Figure 7

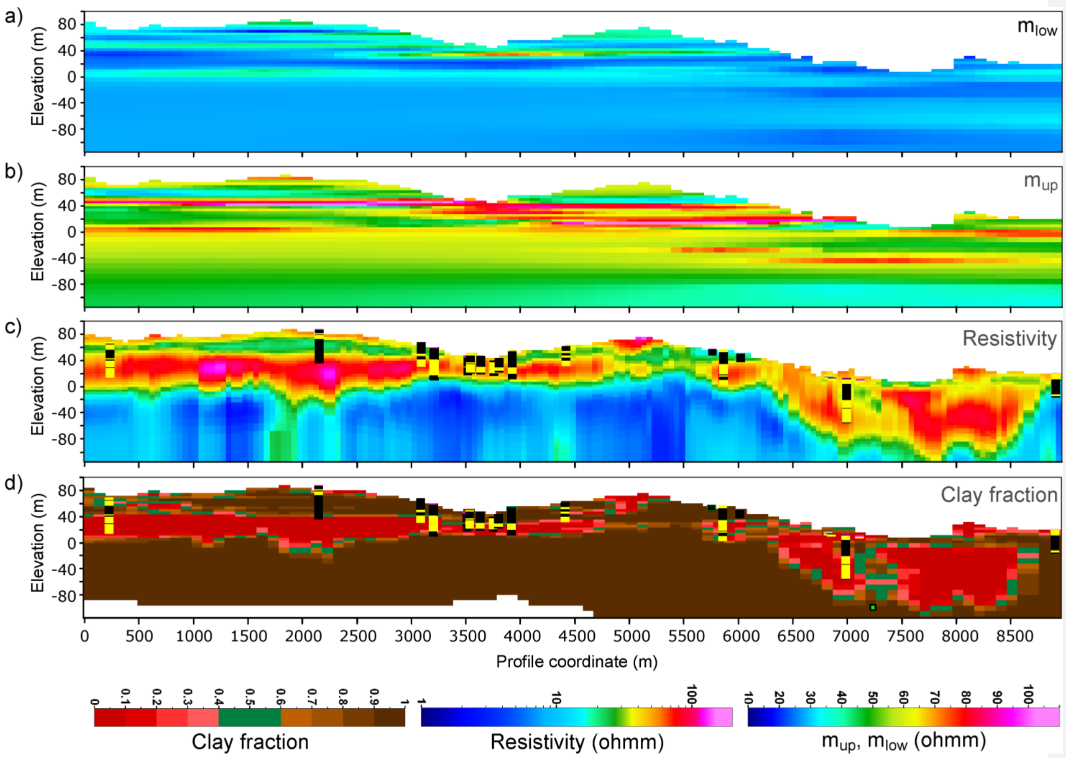

Figure 7. Northwest–southeast cross sections (vertical exaggeration x6). Location and orientation of cross sections are marked in Figure 8. a) The m_{low} parameters of the translator function. b) The m_{up} parameters of the translator function. c) The resistivity section with boreholes within 200 m of the profile superimposed. Black 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).

Deleted: Page Break
¶
F

Deleted: 76

Deleted: Figure 8Figure 7

Figure 8

Deleted: 7

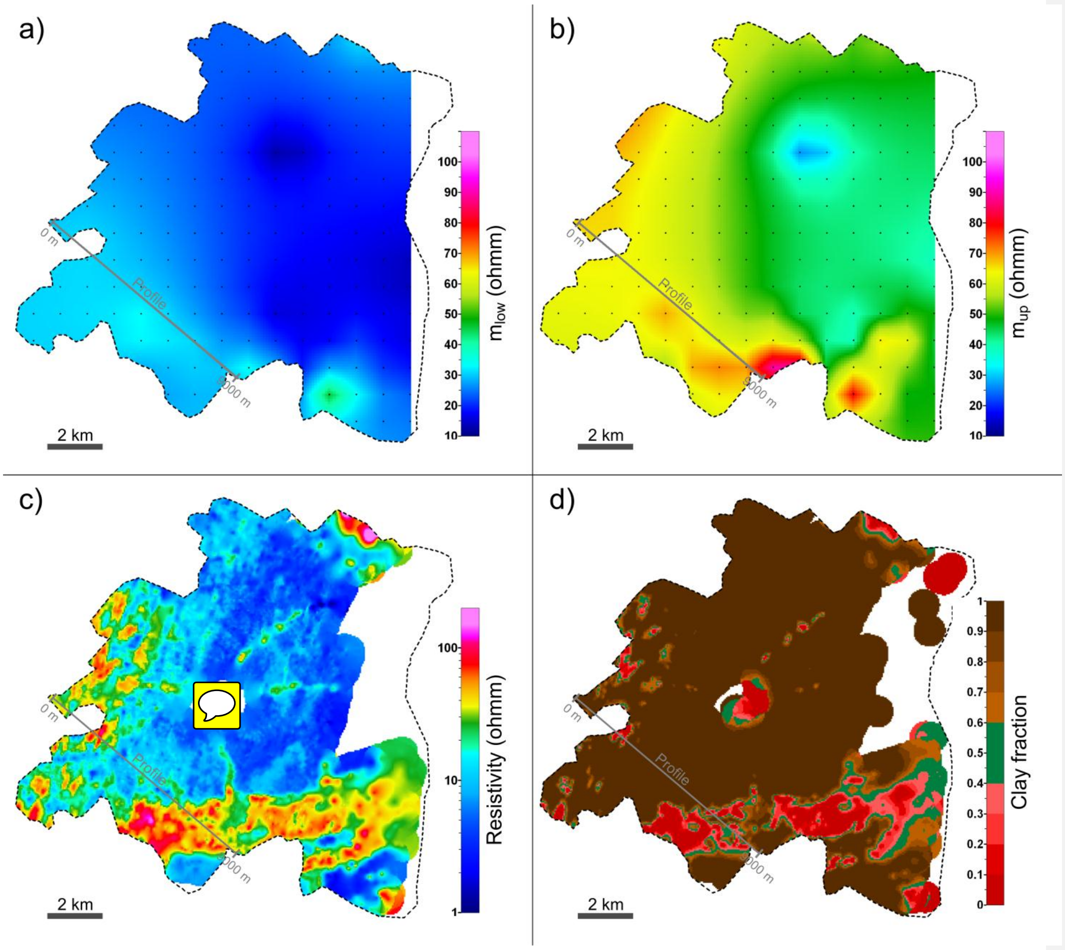

Figure 8. Horizontal slices at 2 mbsl cropped to the catchment area (dashed line). a) The m_{low} parameters of the translator function superimposed with the 1 km translator function grid (black dots). b) The m_{up} parameters of the translator function superimposed with the 1 km translator function grid (black dots). c) Resistivity slice (interpolated). d) Resulting CF- model.

Deleted: 87

Deleted: Interpolated r

Figure 9

Deleted: 8

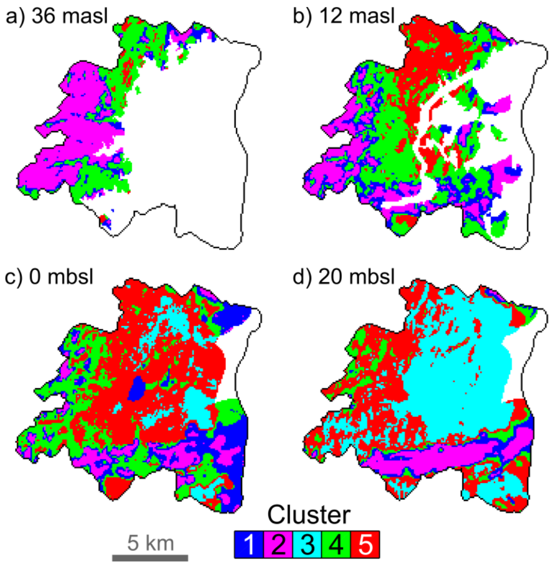

Figure 9. Horizontal slices in four depths of the 3D cluster model.

Deleted: 98

Figure 10

Deleted: 9

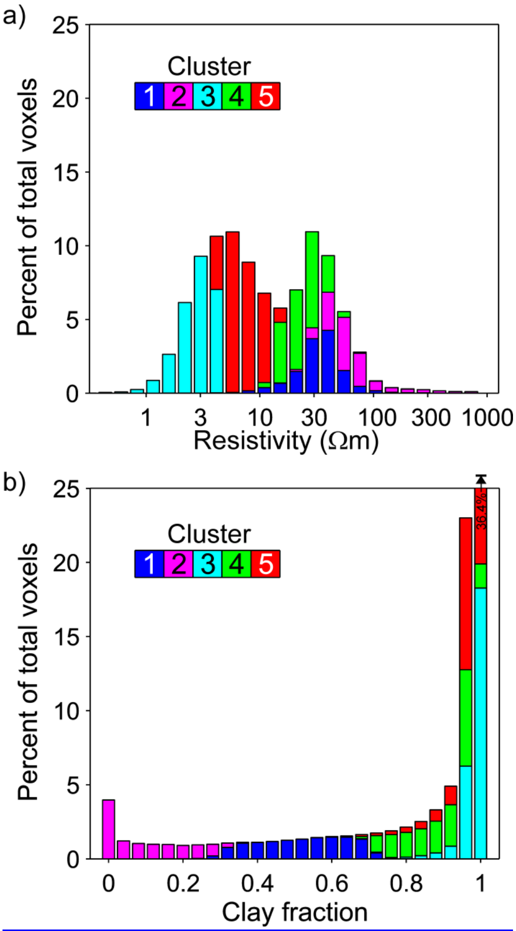

Figure 10. Cluster statistics. The histograms show which data from the original variables make up the five clusters. a) The distribution of the resistivity data in the five clusters. b) The distribution of the CF data in the five clusters.

Deleted: 109