

Editors remarks in black

Authors reply in red

Editors - reviewer

Comments to the Author:

I expect that the Authors can answer most of the comments raised by the two reviewers. They already included a first revised draft, but I strongly recommend the Authors to provide a very accurate revision, in order to properly and completely take into account the reviewers' complaints. Moreover, I attach an annotated copy of their draft revised version, where I inserted several additional comments of mine.

We thank you for this review of the updated version of the paper. We have done our best to address your additional remarks and annotations in the pdf-version.

I think that in this work and in related papers, the effect of pore water saturation and conductivity is often neglected or underestimated. This question was raised by both reviewers and it was only partially answered.

We have added some facts regarding the pore water saturation and conductivity for the Norsminde case as requested, together with a few more discussion comments: "In the Norsminde area used in the case history the groundwater table is generally located a few meters below the surface. This means that the water saturation is not a major factor for the resistivity values and thus the translator function. Though, even if the water table were deeper the effect might not be important as long as we have boreholes with the variations in them. In this case – if we have boreholes containing variable saturation levels (even if it is not measured), and we have geophysical data for the same area, the translator function will automatically adapt to the saturation effect on the resistivity images over the same formation"

Also the topic of SKYTEM sensitivity and its relationship with scaling should be clearly addressed.

This is a comprehensive but a relevant topic. We have added a paragraph in the discussion section with some general remarks regarding sensitivity/resolution for EM-surveys and some more specific remarks regarding the Norsminde SkyTEM- survey and stated a very relevant reference for the Norsminde SkyTEM survey on this topic: Schamper, C., F. Jørgensen, E. Auken, and F. Effersø, 2014, **Assessment of near-surface mapping capabilities by airborne transient electromagnetic data - An extensive comparison to conventional borehole data**: Geophysics, 79, B187-B199.

A careful revision of the reference list is necessary. Several papers are still missing details and some papers are cited in the text but not listed in the bibliography.

The reference list is now fully updated and the various annotations in the reference are also fixed. We apologize for not detecting these in the first version or first revision.

Non-public comments to the Author:

I strongly invite you to prepare a very careful revision of the paper, in order to improve its quality not only from the scientific point of view, but also for language and clarity.

In the answer to reviewer #2, a couple of sentences started with the singular pronoun "I", instead of "We". I wonder whether all the Authors gave their contribution to the paper and to its revision! This is a serious ethical concern.

The authors are of course involved in the revision of the manuscript (to a varying degree depending on the issues raised) and it was simply the writer putting an "I".

A final minor comment: I do not like the widespread use of the word "concept" in the paper; this word could be often substituted with "procedure".

Concept replaced with procedure throughout.

Annotations in Document (The line number refer to the pdf-version with the Editor's annotations)

Editor remarks	Authors response
<p>Line 168. Please, modify. "The geological parameter we map is the clay fraction (CF), expressed as the cumulated thickness of <i>clay</i> in a depth interval relative to the interval length."</p>	<p>Rephrased: The geological parameter we map is the clay fraction (CF). In this paper we refer to <i>clay</i> as material described as clay in a lithological well log regardless the type of clay; <i>clay till, mica clay, Palaeogene clay</i>, etc. This term is robust in the sense that most geologists and drillers have a common conception on the description of <i>clay</i> and it can easily be derived from the lithological log. The clay fraction will then be the cumulated thickness of <i>clay</i> layers in a depth interval deviated with the length of the depth interval.</p>
<p>Line 240. Yellow marking of deleted text.</p>	<p>Note: The deleted part was Rephrased and moved to the introduction.</p>
<p>Line 288: Please, rephrase this sentence. "parameters affecting the uncertainty of the ρ_{log} are parameters like sample interval and density, accuracy of the geographical positioning and elevation, and the credibility of the contractor."</p>	<p>Sentence has been rephrased for clarity</p>
<p>Line 289 Please, rephrase. Possibly substitute "contractor" with "driller"</p>	<p>Corrected</p>
<p>Why this choice and not a different function?</p> $W(\rho) = 0.5 \cdot \operatorname{erfc}\left(\frac{K \cdot (2\rho - m_{up} - m_{low})}{(m_{up} - m_{low})}\right)$	<p>Text updated with: "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 commonly 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"</p>
<p>Substitute 0.0025 x 2 with 0.005</p> $K = \operatorname{erfc}^{-1}(0.0025 \cdot 2)$	<p>Done. (an uncorrected zero has also been removed) Value=0.05</p>
<p>Line 240:</p>	<p>The deleted part was split up, rephrased and placed in the introduction and the discussion sections to keep the focus in the methodology section on the CF-procedure.</p>
<p>Line 321: varying between 4 m and 20 m</p>	<p>Updated</p>
<p>Line 433+440+714: missing in the reference list</p>	<p>The reference list is now fully updated and the various annotations in the reference are also fixed.</p>
<p>Line442: please, explain. "For the lithological logs a fixed lithology code list"</p>	<p>Rephrased to:" All borehole layers in the database are assigned a lithology code, which makes it easy to</p>

	extract the different types of clay layers for the calculation of the Ψ_{\log} values in the different calculation intervals."
Line 462: is this right? "The boreholes are awarded points in"	"Awarded" replaced with "assigned "
Line: 495: please, rephrase. "Data were inverted single site using a 1D layered"	Rephrased to: " 1D layered resistivity models with 3 to 5 layers were used in the interpretation of the TEM sounding data
Line 500 please, use a unique notation and define "m asl" and "m bsl".	Defined and corrected throughout the paper.
Line 565: please, clarify. -10 m asl or 10 m below the ground surface?	Corrected to "(deeper than 10 m bsl)"
Line 878 Figure 1 captions. Please, avoid repetition.	Rephrased
Line 899 Figure 3 This figure is not very informative.	Yes that is actually correct. BUT it is our experience that giving a visual idea of the spatial distribution of, in this case the translator functions and how they are constrained laterally and vertically increases the immediate understanding of the procedure. If need be, we will remove it and replace it by a few lines of text.
Line 928: Figure 6 captions. Please, modify.	Rephrased
Line 950 Figure 7 captions. Please, modify. Possibly "Black and yellow vertical bars show the positions of boreholes: black blocks mark the clay layers, yellow blocks mark sand and gravel layers"	Rephrased
Line 960 Figure 8. What is this white hole? it is partly visible also in the the d) map.	<p>"To create the final regular 3D CF-model the Ψ_{\log} 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.</p> <p>"</p> <p>To avoid confusion we have made a note of this in the figure caption as well.</p>

Review 1 (Jan Gunnik)

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 describe the actual positioning of clay and sand units in the entire volume regardless of the translation, and it therefore makes 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 be 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 has 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

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 σ^2 , σ = 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 satisfactory: 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, convertd 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 now 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 there is no 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 towards 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 do not end up in a cluster!.

Reviewer 2

We have tried our best to follow the comments given by reviewer 2. It was quite difficult as the comments were not referred to specific places in the manuscript, which means that we had to guess in some cases what the particular comments were referring to.

Scientific questions and issues.

In general I identify three general comments/questions on the scientific contents of the manuscript.

1) First question: Is the proposed method an inversion approach or an integrated interpretation approach based on a "geo-statistical approach" through an optimization approach? The proposed method in my opinion regard the integrated interpretation and structural calibration of the geophysical 3-D model obtained with the airborne EM method with borehole data, even though an inverse problem approach is used.

Well, we see this as an approach as the title says: "integrating lithological information from boreholes with resistivity models through an inverse optimization". It is therefore an integrated approach using an inverse formulation and in that sense it is not either or. Though, the inversion part is, in our opinion, what makes it unique.

2) Second question: The authors should specify, for a better reader comprehension, the reason to select the "Clay-Fraction" (CF) as characteristic descriptive hydro-geophysical parameter of the model. As the authors claim (2.1 paragraph; with references Waxman and Smits 1968 and Shevnin et al 2007) that, "It is a common assumption that a petrophysical relationship between resistivity and clay content can be establish...". Detail about this relationship should be given.

The Waxman and smiths relation is quite trivial (basically: $\sigma = \sigma_{\text{water}}/F + \sigma_{\text{clay}}$) stating archies law plus a clay term. For the purpose here we find that it would make the introduction overly long if this (and similar) should be added in detail. Though, we have added more paragraphs on this both in the introduction and in the discussion part.

In my opinion the selection of this parameter should be deeply explained. Is the CF a sensitive parameter for hydro-geological process description (in spatial and time scales)? In particular how the "CF", which is an integral descriptor parameter (as a consequence of its definition), could be used in the hydrogeological modelling, in which probably the fine-distribution of CF parameter of the stratigraphic units are requested for an accurate predictive modelling.

We have added an extended paragraph in the introduction explaining and giving reasons for our choice of "clay-fraction" as the key parameter, and for the concept in general. The discussion part have also been extended with sections discussing related issues

Nevertheless I agree with authors that an approach in which we model the hydrogeophysical model is parameterised in terms of a set of parameters that characterize the "homogeneous hydro-geological units" but this should explained in detail, (ie: why only one parameter?), also in terms of the errors that this choice induces in the predictions in the hydro-geological modelling, when a such approach is used.

We have tried to include an elaborated discussion on this as well in the introduction

Really we are dealing with an hydro-geological conceptual model and, in this context, I advice to use the term "CF-conceptual model", more than "CF-concept". They should explain the basic "adopted conceptual model". Conversely if you want perform a "calibrated" structural interpretation of the EM data with a geo-statistical approach considering one parameter we should say it explicitly, clarifying the proper use of the obtained model, as, it seems, the authors definitively claims in the conclusion "With the CF-concept and clustering we aim at building 3-D models suitable as structural input for groundwater models".

We apologize, but we are not quite sure what is meant here. We are definitely NOT dealing with model building using geo-statistical approaches (whatever that precisely is). The model in the end could be called “CF conceptual model”, but we would like to refer to the overall procedure as the “CF concept”. Though, “Conceptual model” indicates in our opinion that it is a rough model based on limited background information. Here, we actually present an approach taking in all the borehole information AND all the structural information in the resistivity model to produce a clay-fraction model.

Finally connected to this point the CF vs Resistivity relationship is not a single values relationship, as pointed out also by the authors to justify the results. So, again, why they chose an integration procedure with a single parameter?

Again, we are unfortunately not sure what is meant here. The CF is a single parameter being an output of a relationship that is a spatially distributed two-parameter function. We choose the single parameter in lack of good options. We hope that the extended paragraphs in the discussion section adds to the confusion addressed by the reviewer.

3 – The differences in spatial sampling between boreholes and airborne EM resistivity. The authors should spend more effort in describing the spatial parameter setup (spatial analysis of the data, mesh-grid selection, smoothing and interpolation parameters).

Yes, this is pointed out also by the other reviewer and we have added an extended paragraph in the discussion.

Technical issues

Review remarks	Authors response
In the equation 1 should be used a notation 3D using the discrete indices,	Equation 1 is not intended as describing the 3D distribution of the translator function. Here it is merely a general description of the translator function without any dimensionality (i.e referring only to the inset of figure 2). For clarification we have introduced the translator function in its own figure without the 3D grid.
explicating also the rho meaning (even though is trivial).	Done
Also the figure 2 should be modified inserting the grid notation.	Fig. 2 is a <u>principle</u> sketch for the translator function grid and constraints. Adding i,j,k indexes for the three directions (if that what meant by “grid notation”) will just add unneeded complexity to the figure in our opinion.
The constraints in m_up and m_low should be explicated: what do you intend: a smoothing, limits and why you need to constrain these values, How do you set these constraints?	This is also partly addressed by the other reviewer and we have added more text to clarify these choices. Though, the purpose of the constraints is already explained: <i>“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... The smoothness constraints furthermore act as regularization and stabilize the inversion scheme.”</i> The paragraph explaining the constraint setup for the case has been rephrased and extended: <i>“The regularization constraints between neighboring translator model nodes are set relatively loose to</i>

	<i>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."</i>
From eq.1 I think that the translator function (probably better "CF profiler function") is isotropic but really when I read the entire procedure, due to lateral smoothing operation seems to be anisotropic; please explain this aspect.	The translator function is NOT isotropic, and it will vary vertically and horizontally as dictated by the data. We have emphasized this in the beginning of the Methodology section.
Probably it will be interesting to show to the reader, integrating the figure 2, for a vertical profile: geo-stratigraphy with description about "clay contents" and the corresponding Psilog, and show the corresponding electro-stratigraphy, the corresponding 'Translator function' and the derived Psires. This also to demonstrate the basic assumption of the approach(eq. 1 and 2).	This is a good idea, and we have added this information to a new figure 2.
In the equation 3 it should be explain the meaning of m (the parameter of the translator function)	Minor update/explanation of m added.
What are the resistivity errors.	The paper already hold this paragraph: <i>"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 ρ_{res} values is described in detail in Christiansen et al., 2013."</i>
What is the sensitivity in depth of resistivity inversion?	Sensitivity/resolution for the Airborne EM results is a topic worth several papers in itself and hence out of the scope of this paper in our opinion. Though, it is an important issue when building models based on airborne EM resistivity models and we mention this now in the rewritten discussion section
A central technical issue of the application of the procedure is the spatial sampling and in particular the difference between the sampling in resistivity and boreholes. The first issue regards the vertical (z) sampling and resolution. The good geo-stratigraphic data (quality 1 and 2 following the author classification) probably are characterize by an oversampling with spatial wavelength of 1-2 m despite the inverted resistivity that has a higher sampling 4 or 8/10 m.	The lithological logs are generally not oversampled (on the contrary). Some boreholes have lithological samples for each 1 m some for each 5 m, while others only define when the lithology changes. The vertical discretization of the CF-models (thickness of the calculation layers) has been selected to some degree to reflect the vertical resolution in the resistivity models since it is the translation of the resistive model that gives the structure in the CF-model.
In the horizontal direction we have the inverse situation with resistivity data sampled at about 15 m along the line and 50-100m between the line and the distance between borehole surely greater of 100 m but, I think , and comparable with about 1000 m which is the lateral grid used in the geo-statistical optimization with the proposed procedure. About this the authors should analyze and report	In our opinion it is probably a bit too detailed to describe in detail the distribution of boreholes as it would be difficult to make any real use of this information. However, we have updated two key figures to provide overview "statistical" information on the borehole distribution: Figure 4a has been updated to also show the drill

<p>some statistic parameter of spatial distribution of the borehole. Probably a statistical evaluation of the Voronoi area could be suitable to fix the minimum horizontal area including the min depth information in the whole data space (geo-stratigraphy+resistivity). Probably an areal pre-selection based on the areal distribution of the data could drive the optimization. If the studied area is about 156 km², and we suppose an homogeneous spatial distribution of the deeper boreholes up to 90m which are 125 (100 of which up to 60m), we have about 0.8 boreholes/km². This is quite in agreement with the horizontal grid used in the CF procedure (1 km), but if we see the figure 4b, 6c and 6d, the spatial boreholes distribution is highly variable.</p>	<p>depths, so the borehole distribution vs. depth can be examine and fig. 5 has been updated with boreholes/km² information for the different depths intervals.</p> <p>Also, a paragraph elaborating the choice of horizontal discretization for the translator model vs. borehole and resistivity model density has been added to the Discussion section.</p>
<p>Furthermore the depth sampling of boreholes seems poor with respect the resistivity one. So the obtained results, as claimed by the authors, is mainly driven by the starting model for 6320 CF points over a total of 11520! This aspect should be emphasized.</p>	<p>We agree that this was not clear in the original manuscript. Paragraph updated:</p> <p><i>“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.”</i></p> <p>A discussion on selecting starting model has also been added to the discussion.</p>
<p>How is obtained the starting/reference model for m. “starting model and constrains setup are based on experience and the expected geologic variability and fine-tuned through a subsequent of test inversion”(3.3 paragraphs). It should spend some explanation about this; what is the type of information you intend as experience and degree of geological expectation?</p>	<p>Paragraph about this added to the Discussion section.</p>
<p>It is possible to perform trials or numerical experiment and test to study the robustness of the procedure respect the starting model, procedure parameters and constrains? The strength of an automatic procedure of data interpretation is connected to her sensitivity to the initial setup (a priori information, starting model, procedural parameter setup). The author in 3.3 paragraph claim that to setup the inputs of procedure “fine tuned test inversion” were performed. Which tests was performed, which are the results of these test in order to drive the setup of the inputs of the procedure?</p>	<p>Any inversion problem will be sensitive to the starting model and the setup of constraints. Normally, very little is written in papers about the fine-tuning of the inversions, but in this case the data density becomes very low for the deeper layers with few boreholes and obviously this increases the effect of the starting model.</p> <p>An extended discussion on starting model and constraints has been included in a discussion section.</p>
<p>-Results. Can you give a measure of errors in the optimized parameters (even though difficult for non-linear inversion) or the reduction in residuals?</p>	<p>We could give these estimates as we have everything ready at hand. However, the uncertainties are fairly difficult to use by themselves as we are really interested in the uncertainties of the resulting clay fraction model and not the translator model parameters. For the clay fraction model we have chosen not to show the uncertainties to avoid this complexity level as the uncertainties are by far dominated by the uncertainty in the kriging interpolation, which means that the uncertainty map is, to the first degree, a visualization of the data density, without much unique information.</p>

	<p>However, the text has been slightly updated to emphasize key issues about uncertainty and we have added key misfit numbers on the data side as these were missing in the first version.</p>
<p>Conclusions: the binomial behaviour should be expected due to the math feature of the “translator function”, i.e. an on-off or low-pass filter. What you think about?</p>	<p>Not sure what is meant here? The statement as written here is correct and is also what we conclude in the paper looking at it from the lithology side: <i>“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 calculation interval”.</i></p>
<p>Could you present, if is available for the same area, examples of the application of other approaches as it is reported in the introduction paragraph and references?</p>	<p>Very good idea, but no, not at present. For the Norsminde area a model comparison paper is under preparation comparing: 1) A traditional “cognitive” geological model 2) stochastic generated models using transition probability geostatistics and conditional Sequential Indicator Simulation”. 3) The CF model of this paper.</p>
<p>- The figure 3 could be reported as an inset panel in figure 4.</p>	<p>Fig. 4 is already relatively compact. Not really room for fig. 3. as a panel.</p>
<p>- In figure 4b beside the quality could be interesting to insert a color or symbol size to represent the maximum boreholes depths.</p>	<p>Yes – good point. Figure updated with drill depth information.</p>
<p>Figure 6 what the colour palette of borehole indicate the clay layer. Further in figure 6d it could be better to represent the CF obtained from boreholes using the palette of CF or representing the values in a CF vs Z profile, probably better in a inset zoom panel or another figure.</p>	<p>The figure text clearly explains the color code of the boreholes: <i>“Black borehole colors mark the clay layers, while yellow colors mark sand and gravel layers”.</i></p> <p>If we use the same color scale for the boreholes as for the CF-model (Red- brown) will it be very difficult to see the boreholes in figure 6d!.</p>
<p>What is the reason for CF mapping of the use of a colour palette with a different colour tunes? Why the authors didn’t use a standard 5 colour palette like those used in m values or resistivity?</p>	<p>The color scale for the CF-model (red- brown) has been selected to have it stand out from the resistivity scales indicating that this is a totally different regime.</p>

1 Large scale 3D-modeling by integration of 2 resistivity models and borehole data 3 through inversion

4

5 Authors: Nikolaj Foged¹, Pernille Aabye Marker², Anders Vest Christansen¹, Peter Bauer-Gottwein²,
6 Flemming Jørgensen³, Anne-Sophie Høyer³, Esben Auken¹

7 ¹HydroGeophysics Group, Department of Geoscience, Aarhus University, Denmark.

8 ²Department of Environmental Engineering, Technical University of Denmark

9 ³Geological Survey of Denmark and Greenland

10 **ABSTRACT**

11 We present an automatic method for parameterization of a 3D model of the subsurface, integrating
12 lithological information from boreholes with resistivity models through an inverse optimization, with the
13 objective of further detailing [of geological models](#) or as direct input to groundwater models. The
14 parameter of interest is the clay fraction, expressed as the relative length of clay-units in a depth interval.
15 The clay fraction is obtained from lithological logs and the clay fraction from the resistivity is obtained by
16 establishing a simple petrophysical relationship, a translator function, between resistivity and the clay
17 fraction. Through inversion we use the lithological data and the resistivity data to determine the optimum
18 spatially distributed translator function. Applying the translator function we get a 3D clay fraction model,
19 which holds information from the resistivity dataset and the borehole dataset in one variable. Finally, we
20 use k-means clustering to generate a 3D model of the subsurface structures. We apply the [procedure](#) to
21 the Norsminde survey in Denmark integrating approximately 700 boreholes and more than 100,000
22 resistivity models from an airborne survey in the parameterization of the 3D model covering 156 km².
23 The final five-cluster 3D model differentiates between clay materials and different high resistive materials
24 from information held in [the](#) resistivity model and borehole observations respectively.

Deleted: for

Deleted: concept

27 1 INTRODUCTION

28 In a large-scale geological and hydrogeological modeling context, borehole data seldom provide an
29 adequate data base due to low spatial density in relation to the complexity of the subsurface to be mapped.
30 Contrary, dense areal coverage can be obtained from geophysical measurements, and particularly airborne
31 EM methods are suitable for 3D mapping, as they cover large areas in a short period of time. However,
32 the geological and hydrogeological parameters are only mapped indirectly, and an interpretation of the
33 airborne results is needed, often based on site-specific relationships. Linking electrical resistivity to
34 hydrological properties is thus an area of increased interest as reviewed by Slater (2007).

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

46 Geostatistical modeling approaches such as multiple-point geostatistical methods (Daly and Caers, 2010;
47 Strebelle, 2002), transition probability indicator simulation (Carle and Fogg, 1996) or sequential indicator
48 simulation (Deutsch and Journel, 1998), provide models with a higher degree of objectivity in shorter
49 time compared to the cognitive, manual modeling approaches. An example of combining AEM and
50 borehole information in a transition probability indicator simulation approach is given by He et al., (2014).

51 Geostatistical modeling approaches based primarily on borehole data often face the problem that the data
52 are too sparse to represent the lateral heterogeneity at the desired spatial scale. Including geophysical data
53 enables a more accurate estimation of the geostatistical properties especially laterally. This could be
54 determination of the transition probabilities and the mean lengths of the different units. Though, the
55 geophysical data also opens the question of to what degree the different data types should be honored in
56 the model simulations and estimations. Combined use of geostatistical and cognitive approaches can be a
57 suitable solution in some cases (Jørgensen et al., 2013b; Raiber et al., 2012; Stafleu et al., 2011).

58 Integration of borehole information and geological knowledge as prior information directly in the
59 inversion of the geophysical data is another technique to combine the two types of information and

- Deleted: which
- Deleted: an interpretation that
- Deleted: Therefore e.g. the link
- Deleted: v
- Deleted: between
- Deleted: and electrical properties has
- Deleted: therefore
- Deleted: been
- Deleted: separated by surfaces

Deleted: Fogg, 1996

- Deleted: He et al.
- Deleted: 2013
- Deleted: The use of only borehole data in g
- Deleted: problems
- Deleted: s

- Deleted: especially laterally
- Deleted: ,
- Deleted: ,
- Deleted: but
- Deleted: of what to use as *hard* and *soft* data
- Deleted: Jørgensen et al., 2013b

83 thereby achieve better geophysical models and subsequently better geological and hydrological models
84 ([Høyer et al., 2014](#); Wisén et al., 2005).

Deleted: Høyer et al., 2014

85 Geological models are commonly used as the basis for hydrostratigraphical input to groundwater models.
86 [However, even though groundwater model predictions are sensitive to variations in the hydrostratigraphy,](#)
87 [the groundwater model calibration is non-unique, and different hydrostratigraphic models may produce](#)
88 [similar results](#) (Seifert et al., 2012).

Deleted: While m

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

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

98 We present an automatic [procedure](#) for parameterization of a 3D model of the subsurface. The geological
99 parameter we map is the clay fraction (CF). [In this paper we refer to clay as material described as clay in](#)
100 [a lithological well log regardless the type of clay; clay till, mica clay, Palaeogene clay, etc. This term is](#)
101 [robust in the sense that most geologists and drillers have a common conception on the description of clay](#)
102 [and it can easily be derived from the lithological logs. The clay fraction is then the cumulated thickness of](#)
103 [clay layers in a depth interval divided with the length of the depth interval.](#) The CF-procedure integrates
104 lithological information from boreholes with resistivity information, typically from large-scale
105 geophysical AEM surveys. We obtain the CF from the resistivity data by establishing a petrophysical
106 relationship, a translator function, between resistivity and the CF. Through an inverse mathematical
107 formulation we use the lithological borehole data to determine the optimum parameters of the translator
108 function. Hence, the 3D CF-model holds information from the resistivity dataset and the borehole dataset
109 in one variable. As a last step we cluster our model space represented by the CF-model and geophysical
110 resistivity model using k-means clustering to form a structural 3D cluster model with the objective of
111 further detailing for geological models or as direct input to groundwater models.

Deleted: ¶

Deleted: method

Deleted: , expressed as the cumulated thickness of clay in a depth interval relative to the interval length.

Deleted: will

Deleted: be

Deleted: e

Deleted: ated

Deleted: method

112 [Lithological interpretation of a resistivity model is not trivial since the resistivity of a geological media is](#)
113 [controlled by: porosity, pore water conductivity, degree of saturation, amount of clay minerals, etc.](#)
114 [Different, primarily empirical, models try to explain the different phenomena, where Archie's law](#)
115 [\(Archie, 1942\) is the most fundamental empirical model taking the porosity, pore water conductivity and,](#)

Deleted: ,

132 [the degree of saturation into account, but does not account for electrical conduction of currents taking](#)
133 [place on the surface of the clay minerals. The Waxman and Smits model \(Waxman and Smits, 1968\)](#)
134 [together with the Dual Water model of Clavier et al. \(1984\) provides a fundamental basis for widely and](#)
135 [repeatedly used empirical rules for shaly sands and material containing clay \(e.g. Bussian, 1983; Sen,](#)
136 [1987; Revil and Glover, 1998\).](#) [However, in a sedimentary depositional environment it can be assumed](#)
137 [in general that clay or clay rich sediments will exhibit lower resistivities than the non-clay sediments, silt,](#)
138 [sand, gravel, and chalk. As such, discrimination between clay and non-clay sediments based on resistivity](#)
139 [models is feasible and the CF-value is a suitable parameter to work with in the integration of resistivity](#)
140 [models and lithological logs. A 3D CF-model or clay/sand model will also contain key structural](#)
141 [information for a groundwater model, since it delineates the impermeable clay units and the permeable](#)
142 [sand/gravel units.](#)

Deleted: ¶

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

Deleted: s

Deleted:

Deleted: -

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

Deleted: concept

150 2 METHODOLOGY

151 Conceptually, our approach sets up a function that best describes the petrophysical relationship between
152 clay fraction and resistivity. Through inversion we determine the optimum parameters of this translator
153 function, by minimizing the difference between the clay fraction calculated from the resistivity models
154 (Ψ_{res}) and the observed clay fraction in the lithological well logs (Ψ_{log}).

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

Deleted: We do this for

Deleted: that are mutually constrained

Deleted:

Deleted: mized

Deleted: concept

Deleted: concept

Deleted: Christiansen et al., 2013

Deleted: Christiansen et al., 2014

175 The flowchart in Figure 1 provides an overview of the CF-procedure. The observed clay fraction (Ψ_{\log}) is
176 calculated from the lithological logs (box 1) in the calculation intervals. The translator function (box 2)
177 and the resistivity models (box 3) form the forward response, which produces a resistivity-based clay
178 fraction (box 4) in the different calculation intervals. The parameters of the translator function are
179 updated during the inversion to obtain the best consistency between Ψ_{res} and Ψ_{\log} . The output is the
180 optimum resistivity-to-clay fraction translator function (box 5), and when applying this to the resistivity
181 models (the forward response of the final iteration), we obtain the optimum Ψ_{res} and block kriging is used
182 to generate a regular 3D CF model (box 6).

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

187 The subsequent paragraphs detail the description of the individual parts of the CF-procedure.

188 2.1 Observed data - lithological logs and clay fraction

189 The common parameter derived from the lithological logs and resistivity datasets is the clay fraction
190 (Figure 1, boxes 1-4). The clay fraction, of a given depth interval in a borehole (named Ψ_{\log}) is calculated
191 as the cumulative thickness of layers described as clay divided by the length of the interval. By using this
192 definition of clay and clay fraction we can easily calculate Ψ_{\log} in depth intervals for any lithological well
193 log as the example in Figure 2a shows. Having retrieved the Ψ_{\log} values we then need to estimate their
194 uncertainties since a variance estimate, σ_{\log}^2 is needed in the evaluation of the misfit to Ψ_{res} .

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

202 2.2 Forward data – the translator function

203 For calculating the clay fraction for a resistivity model, Ψ_{res} , we use the translator function as shown in
204 Figure 2b, which is defined by a m_{low} and a m_{up} parameter. With the CF-procedure we primarily want to
205 determine resistivity threshold values for a clay-sand interpretation of the resistivity models. This

Deleted: concept

Deleted: concept

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: 2

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

Deleted: ,

Deleted: parameters like

Deleted: contractor

Deleted: a simple two-parameter

Deleted: 2

Deleted:

Deleted:

246 geological layers are often not directly visible in the resistivity models, whereas they will most often
 247 appear in carefully described boreholes. The length of the calculation intervals reflects the resolution
 248 capability of the geophysical method of choice, which means that in some cases the calculation intervals
 249 contain both sand and clay layers when imposed on the lithological logs. The translator function must
 250 therefore be able to translate resistivity values as partly clay and partly sand to obtain consistency with the
 251 lithological logs. This is possible with the translator function in Figure 2b, where m_{low} and m_{up} represent
 252 the clay and sand cut-off values. So for resistivity values below m_{low} the layer is entirely clay (weight ≈ 1)
 253 and for resistivity values above m_{up} the layer is entirely sand or non-clay (weight ≈ 0).
 254 Many functions fulfilling the above criteria could have been chosen, but we use the one shown because it
 255 is differentiable throughout while being flat at both ends and fully described by just two parameters. The
 256 translator function ($W(\rho)$) is mathematically a scaled complementary error function, defined as:

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

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

257 (1)

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

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

262 (2)

263 where N is the number of resistivity layers in the calculation interval, $W(\rho_i)$ is the clay weight for the
 264 resistivity in layer i , t_i is the thickness of the resistivity layer, and $\sum t_i$ is the length of the calculation
 265 interval. In other words, W weights the thickness a resistivity layer, so for a resistivity below m_{low} the
 266 layer thickness is counted as clay ($W \approx 1$) while for a resistivity above m_{up} the layer is counted as non-
 267 clay ($W \approx 0$). Figure 2a shows how a single resistivity model is translated into Ψ_{res} in numbers of
 268 calculation intervals.

Deleted: and

Deleted: commonly

Deleted: e

Deleted:

Deleted:

Deleted: 2

Deleted: of the translator function

Deleted: counted as

Deleted: counted as

Deleted: The translator function is described fully by

Deleted: ,

Deleted: 0

Deleted: 2

Deleted: · 2

Deleted: ,

Deleted: an

Deleted: ,

Deleted: 2

Deleted:

Deleted: a

290 The resistivity models are also associated with an uncertainty, and if the variance estimates of the
291 resistivities and thicknesses for the geophysical models are available we take these into account. The
292 propagation of the uncertainty from the resistivity models to the Ψ_{res} values is described in detail in
293 [Christiansen et al. \(2014\)](#).

Deleted: Christiansen et al., 2013

Deleted: Christiansen et al.

Deleted: 2014

294 To allow for variation, laterally and vertically, in the resistivity to Ψ_{res} translation, a regular 3D grid is
295 defined for the survey block (Figure 3). Each grid node holds one set of m_{up} and m_{low} parameters. The
296 vertical discretization follows the clay fraction calculation intervals, [varying between 4-20 m increasing](#)
297 [with depth. The horizontal discretization is typically 0.5-2 km and a](#) 2D bilinear horizontal interpolation
298 of the m_{up} and m_{low} is applied to define the translator function uniquely at the positions of the resistivity
299 models.

Deleted:

Deleted: typically

Deleted: 1

Deleted: intervals, A

Deleted: .

300 To migrate information of the translator function from regions with many boreholes to regions with few
301 [or no](#) boreholes, horizontal and vertical smoothness constraints are applied between the translator
302 functions at each node point as shown in Figure 3. [Choosing appropriate constraints is based on the](#)
303 [balance between fitting the data while having a reasonable model. The balance is site and data specific,](#)
304 [but would typically be based on visual evaluations comparing the results against key boreholes.](#) The
305 smoothness constraints furthermore act as regularization and stabilize the inversion scheme.

Deleted: or with no boreholes

306 Finally, we [need to](#) estimate Ψ_{res} values at the Ψ_{log} positions (named Ψ_{res}^*) [for evaluation. We estimate the](#)
307 [\$\Psi_{res}^*\$ values by making a point kriging interpolation of the \$\Psi_{res}\$ values and associated uncertainties within](#)
308 [a search radius of typically 500 m.](#) The experimental semi-variogram is calculated from the Ψ_{res} values
309 for the given calculation interval and can normally be approximated well with an exponential function,
310 which then enters the kriging interpolation. The code Gstat (Pebesma and Wesseling, 1998) is used for
311 kriging, variogram calculation, and variogram fitting. [Hence, for the output estimates of the \$\Psi_{res}^*\$ both the](#)
312 original [variance](#) of Ψ_{res} and [the variance on the kriging interpolation itself is included to provide total](#)
313 [variance estimates of the \$\Psi_{res}^*\$ values \(\$\sigma_{res}^{2*}\$ \), which](#) are needed for a meaningful evaluation of the data
314 misfit at the borehole positions.

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

315 2.3 Inversion - objective function and minimization

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

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

318 (3)

337 where $\delta \Psi_{obs}$ denotes the difference between the observed data (Ψ_{log}) and the non-linear mapping of the
 338 model to the data space (Ψ_{res}). δm_{true} represents the difference between the [model parameters \(\$m_{up}\$, \$m_{low}\$ \)](#)
 339 [of the true, but unknown, translator function and an arbitrary reference model \(the initial starting model](#)
 340 [for the first iteration, then at later iterations the model from the previous iteration\)](#). e_{log} is the
 341 observational error, and G denotes the Jacobian matrix that contains the partial derivatives of the
 342 mapping. The general solution to the non-linear inversion problem of equation (3) is described by
 343 [Christiansen et al. \(2014\)](#) and is based on Auken and Christiansen (2004) and Auken et al. (2005).

Deleted:

Deleted: (3)

Deleted: Christiansen et al. (2013)

Deleted: Christiansen et al.

Deleted: 2014

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

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

Deleted: $(\sigma_i)^2$

346 (4)

347 where N_{dat} is the number of Ψ_{log} values and σ_i^2 is the combined variance of the i 'th Ψ_{log} ($\sigma_{log,i}^2$) and Ψ_{res}
 348 ($\sigma_{res,i}^2$) given as:

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

Deleted: σ_i

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

349 (5)

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

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

352 (6)

353 Where e_r is the regularizing constraint between the two constrained parameters m_j and m_k of the translator
 354 function and N_{con} is the number constraint pairs. The e_r values in equation (6) are stated as constraint
 355 factors, meaning that an e_r factor of 1.2 corresponds approximately to a model change of +/- 20%.

356 In total the objective function Q becomes:

357

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

366 (7)

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

369 The minimization of the non-linear problem is performed in a least squares sense by using an iterative
 370 Gauss-Newton minimization scheme with a Marquardt modification. The full set of inversion equations
 371 and solutions are presented in [Christiansen et al., \(2014\)](#).

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

372 2.4 Cluster analysis

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

380 K-means clustering is a hard clustering algorithm used to group multivariate data. A k-means cluster
 381 analysis is iterative optimization with the objective of minimizing a distance function between data points
 382 and a predefined number of clusters (Wu, 2012). We have used Euclidean length as a measure of
 383 distance. We use the k-means algorithm in MATLAB R2013a, which has implemented a two-phase
 384 search, batch and sequential, to minimize the risk of reaching a local minimum (Wu, 2012). K-means
 385 clustering can be performed on several variables, but for variables to impact the clustering equally, data
 386 must be standardized and uncorrelated. The CF-model and resistivity model are by definition correlated.
 387 We use Principal Component Analysis (PCA) to obtain uncorrelated variables.

- Deleted: to
- Deleted: e

388 Principal component analysis is a statistical analysis based on data variance formulated by Hotelling
 389 (1933). The aim of a PCA is to find linear combinations of original data while obtaining maximum
 390 variance of the linear combinations (Härdle and Simar, 2012). This results in an orthogonal
 391 transformation of the original multi-dimensional variables into a space where dimension one has largest
 392 variance, dimension two has second largest variance, etc. In this case the PCA is not used to reduce
 393 variable space, but only to obtain an orthogonal representation of the original variable space to use in the
 394 clustering analysis. Principal components are orthogonal and thus uncorrelated, which makes the

402 principal components useful in the subsequent clustering analysis. The PCA is scale sensitive and the
403 original variables must therefore be standardized prior to the analysis. Because the principal components
404 have no physical meaning, a weighting of the CF-model and the resistivity model cannot be included in
405 the k-means clustering. Instead the variables are weighed prior to the PCA.

406 3 NORSMINDE CASE

407 The Norsminde case model area is located in eastern Jutland, Denmark (Figure 4) around the town of
408 Odder (Figure 5) and covers 156 km², representing the Norsminde Fjord catchment. The catchment area
409 has been mapped and studied intensely in the NiCA research project in connection with nitrate reduction
410 in geologically heterogeneous catchments (Refsgaard et al., 2014). The modeling area has a high degree
411 of geological complexity in the upper part of the section. The area is characterized by Palaeogene and
412 Neogene sediments covered by glacial Pleistocene deposits. The Palaeogene is composed of fine-grained
413 marl and clay and the Neogene layers consist of marine Miocene clay interbedded with deltaic sand layers
414 (Rasmussen et al., 2010). The Neogene is not present in the southern and eastern part of the area where
415 the glacial sediments therefore directly overlie the Palaeogene clay. The Palaeogene and Neogene layers
416 in the region are frequently incised by Pleistocene buried tunnel valleys and one of these is present in the
417 southern part, where it crosses the model area to great depths with an overall E-W orientation (Jørgensen
418 and Sandersen, 2006). The Pleistocene deposits generally appear very heterogeneous and according to
419 boreholes they are composed of glacial meltwater sediments and till.

420 3.1 Borehole data

421 In Denmark, the borehole data are stored in the national database Jupiter (Møller et al., 2009) dating back
422 to 1926 as an archive for all data and information obtained by drilling. Today, the Jupiter database holds
423 information about more than 240,000 boreholes. All borehole layers in the database are assigned a
424 lithology code, which makes it easy to extract the different types of clay layers for the calculation of the
425 Ψ_{\log} values in the different calculation intervals.

426 For the model area, approximately 700 boreholes are stored in the database. Based on borehole meta-data
427 found in the database we use an automatic quality rating system, where each borehole is rated from 1-4
428 (He et al., 2014). The ratings are used to assign different uncertainty (weights) to the lithological logs/the
429 Ψ_{\log} values in the CF-procedure.

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

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

Deleted: to

Deleted: Similar databases are maintained in other countries.

Deleted: For the lithological logs a

Deleted:

Deleted: fixed

Deleted:

Deleted: list is available and all

Deleted: are easily identified, and

Deleted: for the

Deleted: desired elevation intervals can be calculated

Deleted: He et al., 2013

Deleted: apply the lithological logs with

Deleted: used

Deleted: inversion

- 448
- Vertical sample density
- 449
- Accuracy of the geographical position: GPS or manual map location
- 450
- Accuracy of the elevation: Differential GPS or other
- 451
- Drilling purpose: scientific, water abstraction, geophysical shot holes, etc.
- 452
- Credibility of drilling contractor

Deleted: s

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

Deleted: awarded

Deleted:

457 The locations, quality ratings and drill depths of the boreholes are shown in Figure 5b. The drill depths
458 and quality ratings are summarized in Figure 6. As the top bar in Figure 6 shows that, 4% of the boreholes
459 are categorized as quality 1, 46% as quality 2, 32% as quality 3, and 18% as quality 4. The uncertainties
460 of the Ψ_{\log} values for the quality groups 1-4 are based on a subjective evaluation and are defined as 10%,
461 20%, 30%, and 50%, respectively. The number of boreholes drastically decreases with depth as shown in
462 Figure 6. Thus, while about 100 boreholes are present in a depth of 60 m, only 25 boreholes reach a depth
463 greater than 90 m.

Deleted: and

Deleted: , while the

Deleted: t

Deleted: ,

Deleted:

Deleted:

Deleted:

Deleted:

464 3.2 EM data

465 The major part of the model area is covered by SkyTEM data and adjoining ground based TEM
466 soundings are included in the resistivity dataset (Figure 5a).

467 The SkyTEM data were collected with the newly developed SkyTEM¹⁰¹ system (Schamper et al., 2014b).
468 The SkyTEM¹⁰¹ system has the ability to measure very early times, which improves the resolution of the
469 near surface geological layers, when careful system calibration and advanced processing and inversion
470 methodologies are applied (Schamper et al., 2014a). The recorded times span the interval from ~3 μ s to 1-
471 2 ms after end of the turn-off ramp, which gives a depth of investigation (Christiansen and Auken, 2012)
472 of approximately 100 m for an average ground resistivity of 50 Ω m. The SkyTEM survey was performed
473 with a dense line spacing of 50 m for the western part and 100 m line spacing for eastern part (Figure 5a).
474 Additional cross lines were made in a smaller area, which brings the total up to approximately 2000 line
475 km. The sounding spacing along the lines is approximately 15 m resulting in a total of 106,770 1D
476 resistivity models. The inversion was carried out in a spatially constrained inversion setup (Viezzoli et al.,
477 2008) with a smooth 1D-model formulation (29 layers, with fixed layer boundaries), using the AarhusInv
478 inversion code (Auken et al., 2014) and the Aarhus Workbench software package (Auken et al., 2009) .

490 The resistivity models have been terminated individually at their estimated depth of investigation (DOI)
491 calculated as described by Christiansen and Auken (2012).

492 The ground based TEM soundings originate from mapping campaigns in the mid-1990s. The TEM
493 soundings were all acquired with the Geonics TEM47/PROTEM system (Geonics Limited) in a central
494 loop configuration with a 40 by 40 m² transmitter loop. 1D layered resistivity models with 3 to 5 layers
495 were used in the interpretation of the TEM sounding data.

496 3.3 Model setup

497 The 3D translator function grid has a horizontal discretization of 1 km, with 16 nodes in the x-direction
498 and 18 nodes in the y-direction. Vertically the model spans from 100 m above sea level (asl) (highest
499 surface elevation) to 120 m below sea level (bsl). The vertical discretization is 4 m for layers asl and 8
500 m for layers bsl, which results in 40 calculation intervals. Hence, in total, the model grid holds
501 16x18x40=11,520 translator functions each holding two parameters. Translator functions in the 3D grid
502 situated above terrain, below DOI of the resistivity models, and outside geophysical coverage does not
503 contribute at all, and are only included to make the translator function grid regular for easier
504 computation/bookkeeping. The effective number of translator functions, is therefore close to 5,200.

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

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

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

Deleted: Data were inverted single site using a

Deleted: depending on the number of layers needed to fit the data

Deleted: masl

Deleted:

Deleted:

Deleted: mbs

Deleted: above sea level

Deleted: elow

Deleted: ea

Deleted: evel

Deleted: Node points in the t

Deleted: s

Deleted:

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

Deleted: , while the horizontal

Deleted: regularization

Deleted: is set to a

Deleted: s allowe

Deleted: d

Deleted: translator

Deleted: a

Deleted: ies

Deleted: variations of

Deleted:

Deleted: s

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

Deleted:

575 and 8 m between 0 and 120 m_bsl. CF-model values range between 0 and 1 and have therefore not been
576 standardized. The resistivity values have been log transformed and standardized by first subtracting the
577 mean and then dividing by four times the standard deviation. The standardization of the resistivity was
578 performed in this way to balance the weight between the two variables in the clustering. A five cluster
579 delineation is presented for the Norsminde case in the result section.

580 3.4 Results

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

587 For the deeper part of the model (deeper than 10 m bsl) the translator functions are less varying. This
588 corresponds well to the general geological setting of the area with relatively homogenous clay sequences
589 in the deeper part, but it is also a result of very limited borehole information for the deeper model parts.

590 The general geological setting of the area is also clearly reflected in the translator function in the
591 horizontal slices in Figure 8a and b. The eastern part of the area with lowest m_{low} values (dark blue in
592 Figure 8a) and lowest m_{up} values (light blue/green in Figure 8b) corresponds to the area where the
593 Palaeogene highly conductive clays are present. In the western part of the area the cross section intersects
594 the glacial complex, where the clays are mostly tills, and higher m_{low} and m_{up} values are needed to get the
595 optimum translation.

596 The resistivity cross section in Figure 7c and the slice section in Figure 8c reveal a detailed picture of the
597 effect of the geological structures seen in the resistivity data. Generally, a good correlation with the
598 boreholes is observed. Translating the resistivities we obtain the CF-model presented in Figure 7d and
599 Figure 8d. The majority of the voxels in the CF-model have values close to 0 or 1. This is expected since
600 the lithological logs are described binary clay/non clay, and Ψ_{log} values not equal to 0 or 1 can only occur
601 if both clay and non-clay lithologies are present in the same calculation interval in a particular borehole.

602 Evaluating the result in Figure 7d and Figure 8d, it is obvious that the very resistive zones are translated
603 to a CF-value close to 0 and the very conductive zones are translated to CF-value close to 1. Focusing on
604 the intermediate resistivities (20-60 Ω m) it is clear that the translation of resistivity to CF is not one-to-
605 one. For example, the buried valley structure (profile coordinate 6500-8500m, Figure 7d) has mostly
606 high-resistive fill with some intermediate resistivity zones. In the CF-section these intermediate resistivity

Deleted: and discussion

Deleted: (3

Deleted: Figures 7

Deleted:

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

Deleted:

Deleted: below

Deleted: elevation -

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 Ω m and firm and dry clay tills can have relatively high resistivities in the range of 80 Ω m. Furthermore, changes in resistivity occur within the same geological unit due to changes in the pore water resistivity as described by Archie's law. The second issue concerns the resolution of the true formation resistivity in the resistivity models. Lithological logs contain point information with a good and uniform vertical resolution. Contrary, AEM data provide a good spatial coverage, but the vertical resolution for the EM resistivity models is relatively poor and not necessarily returning the true resistivity of the formation. Especially thin high-resistivity layers (sand layers) at great depth are poorly resolved by the EM-methods making geological interpretation difficult. By allowing spatial variation in the translator function we can, to some degree, resolve weak layer indications in the resistivity models lithologically correct while also accounting for variations in the pore water resistivity and other resistivity changes within the same lithological description.

Deleted: to

Deleted: if more

Deleted: cal

Deleted: layers

Deleted:

660 zones are translated to zones of high clay content, consistent with the lithological log at profile coordinate
661 7,000 m that contains a 25 m thick clay layer. The CF-section sharpens the layer boundaries compared to
662 the smooth layer transitions in the resistivity section. The integration of the resistivity data and
663 lithological logs in the CF-procedure results in a high degree of consistency between the CF-results and
664 the lithological logs, as seen in the CF-section in Figure 7d.

Deleted: concepts

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

Deleted: Figure 9

Deleted: Figure 9

Deleted: Figure 9

Deleted: Figure 9

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

Deleted: Figure 10

Deleted: ,

Deleted: Figure 10

Deleted: Figure 10

Deleted: ning

Deleted: Figure 10

Deleted: Figure 10

Deleted: ping

Deleted: Figure 10

678 The CF-model does not differentiate between clay types, contrary the EM-resistivity data that have a
679 good resolution in the low resistivity range and therefore, to some degree, are able to distinguish between
680 clay types. This results in the two-part clustering of the low resistivity (>20 Ω m) values as seen in Figure
681 10a.

Deleted: (

Deleted: Figure 10

Deleted:)

682 **4 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 Ω m) values as seen in (Figure 9Figure 9a).¶

683 **4.1 Translator function, grid and discretization**

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

688 The first issue includes spatial changes in the pore water resistivity, the degree of water saturation, and/or
689 contents of clay minerals for the sediments described lithologically as clay. The spatial variation in the

Deleted: ;

717 pore water resistivity on this modeling scale is probably relatively smooth and small and will therefore
718 only have a minor impact on the resistivity to lithology/clay fraction translation. Even in the case with
719 larger fluctuations in the pore water resistivity (e.g. present of saline pore water) the translator function
720 will automatic adapt to this as long as we have borehole information available that resembles the changes
721 and the basic assumption that the clay rich formations are more conductive than coarse-grained sediments
722 is fulfilled.

Deleted: amp
Deleted: ,
Deleted: ,

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

Deleted: This in the (rare) case where thatre
Deleted: d
Deleted: in
Deleted: is
Deleted:
Deleted:
Deleted: even
Deleted: if the
Deleted: water table wer
Deleted: e deeper the effect might not be important as
Deleted: we have

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

733 The second issue concerns the resolution of the true formation resistivity in the resistivity models.
734 Lithological logs contain point information with a good and uniform vertical resolution. Contrary, AEM
735 data provide a good spatial coverage, but the vertical resolution is relatively poor and decreasing with
736 depth. Detailed geological layer sequences might only be represented by an average conductivity or only
737 have a weak signature in the resistivity models. By allowing spatial variation in the translation we can, to
738 some degree, resolve weak layer indications in the resistivity models by utilizing the vertically detailed
739 structural information from the lithological logs via the translator function.

740 The resolution in the final CF-model is strongly correlated to the resolution in the resistivity model, since
741 the resistivity dataset contribute with the majority of the information. In general EM-methods are
742 sensitive to absolute changes in the electric conductivity, which makes the resolution in the low resistivity
743 end superior to the resolution of high-resistivity contrasts. The diffusive behavior of EM-methods results
744 in a decreasing horizontal and vertical resolution capability with depth, and the vertical resolution
745 capability furthermore strongly depends on the layer sequence. A sequence of thin lithological layering
746 may therefore be represented as a single resistivity layer with an average conductivity, which is obviously
747 challenging for the geological interpretation. The horizontal resolution strongly depends on the
748 sample/line density of the geophysical measurements, but the footprint of a single measurement sets the
749 lower limit for the horizontal resolution. The Norsminde airborne SkyTEM survey is conducted with a

Deleted: over
Deleted:

777 [very dense line spacing giving a very high lateral resolution, which could actually support a finer](#)
778 [horizontal discretization \(25-50 m\) in the CF-model. The 100 m horizontal discretization of the CF-model](#)
779 [and cluster-model was selected to match the computational grid setup of a subsequent groundwater](#)
780 [model. A detailed overview of resolution capabilities of the Norsminde SkyTEM survey is given by](#)
781 [Schamper et al. \(2014b\) including an extensive comparison to borehole data](#)

Deleted:

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

Deleted: ,

Deleted: ,

Deleted:

Deleted: .

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

Deleted: s

796 [The CF-procedure supports both uncertainty estimates on the input data, on the output translator](#)
797 [functions, and on the final CF-model. Generally, the uncertainties in the CF-model are closely related to](#)
798 [the borehole density and quality, as well as resolution and density of the resistivity models. The](#)
799 [calculation and estimation of input and output uncertainties is described en detail in Christiansen et al.](#)
800 [\(2014\).](#)

801 **4.2 Clustering and validation**

802 [For the clustered 3D-model each cluster represents some unit with fairly uniform characteristics. It could](#)
803 [be hydrostratigraphic units where the hydraulic conductivity of the cluster units is determined through a](#)
804 [subsequent groundwater model calibration, typically constrained by hydrological head and discharge data.](#)
805 [Groundwater model calibration of the Norsminde 3D-cluster model has been performed with a](#)
806 [preliminary positive outcome, but more experiments are needed before drawing final conclusions. In this](#)
807 [process one needs to evaluate the cluster validity, i.e. how many clusters the data can support. Cluster](#)
808 [validity can be assessed with various statistical measures \(e.g. Halkidi et al., 2002\). The number of](#)

Deleted: are

816 [clusters resulting in the best hydrological performance might also be used as a measure of cluster validity.](#)
817 [The validity of the clusters and the resulting groundwater model is still to be explored in more detail.](#)

818 **5 CONCLUSION**

819 We have presented a [procedure](#) to produce 3D clay-fraction models, integrating the key sources of
820 information in a well-documented and objective way.

821 The [CF-procedure](#) combines lithological borehole information with geophysical resistivity models in
822 producing large scale 3D clay fraction models. The integration of the lithological borehole data and the
823 resistivity models is accomplished through inversion, where the optimum resistivity to clay fraction
824 function minimizes the difference between the observed clay fraction from boreholes and the clay fraction
825 found through the geophysical resistivity models. The [CF-procedure](#) allows for horizontal and lateral
826 variation in the resistivity to clay fraction translation, with smoothness constraints as regularization. The
827 spatially varying translator function is the key to achieve consistency between the borehole information
828 and the resistivity models. The [CF-procedure](#) furthermore handles uncertainties on both input and output
829 data.

830 The [CF-procedure](#) was applied to a 156 km² survey with more than 700 boreholes and 100,000 resistivity
831 models from an airborne survey. The output was a detailed 3D clay fraction model combining resistivity
832 models and lithological borehole information into one parameter.

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

837 With the [CF-procedure](#) and clustering we aim at building 3D models suitable as structural input for
838 groundwater models. Each cluster will then represent a hydrostratigraphic unit and the hydraulic
839 conductivity of the units will be determined through the groundwater model calibration constrained by
840 hydrological head and discharge data.

841 The 3D clay fraction model can also be seen as a binomial geological sand-clay model by interpreting the
842 high and low CF-values as clay and sand respectively, as the color scale for the CF-model example in
843 Figure 7 and Figure 8 indicated. Integration and further development of the CF-model into more complex
844 geological models have been carried out with success ([Jørgensen et al., 2013b](#)).

Deleted: AND OUTLOOK

Deleted: concept

Deleted: t

Deleted: concept

Deleted:

Deleted: inversion concept

Deleted: ,

Deleted: concept

Deleted: concept

Deleted: -concept

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: en

Deleted: (Jørgensen et al., 2013c)

868 **6 ACKNOWLEDGEMENTS**

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

877 **REFERENCES**

- 878 [Auken, E. and A. V. Christiansen, Layered and laterally constrained 2D inversion of resistivity data: *Geophysics*, 69, 3, 752-761, 2004.](#)
879
- 880 [Auken, E., A. V. Christiansen, B. H. Jacobsen, N. Foged, and K. I. Sørensen, Piecewise 1D Laterally
881 Constrained Inversion of resistivity data: *Geophysical Prospecting*, 53, 497-506, 2005.](#)
- 882 [Auken, E., A. V. Christiansen, C. Kirkegaard, G. Fiandaca, C. Schamper, A. A. Behroozmand, A. Binley,
883 E. Nielsen, F. Effersø, N. B. Christensen, K. I. Sørensen, N. Foged, and G. Vignoli, An overview of a
884 highly versatile forward and stable inverse algorithm for airborne, ground-based and borehole
885 electromagnetic and electric data: *Exploration Geophysics*, 1-13, DOI: 10.1071/EG13097, 2014.](#)
- 886 [Auken, E., A. V. Christiansen, J. A. Westergaard, C. Kirkegaard, N. Foged, and A. Viezzoli, An
887 integrated processing scheme for high-resolution airborne electromagnetic surveys, the SkyTEM system:
888 *Exploration Geophysics*, 40, 184-192, 2009.](#)
- 889 [Bussian, A. E., Electrical conductance in a porous medium: *Geophysics*, 48, 9, 1258-1268, 1983.](#)
- 890 [Carle, S. F. and G. E. Fogg, Transition Probability-Based Indicator Geostatistics: *Mathematical Geology*,
891 28, 4, 453-476, 1996.](#)
- 892 [Christiansen, A. V. and E. Auken, A global measure for depth of investigation: *Geophysics*, 77, 4,
893 WB171-WB177, 2012.](#)
- 894 [Christiansen, A. V., N. Foged, and E. Auken, A concept for calculating accumulated clay thickness from
895 borehole lithological logs and resistivity models for nitrate vulnerability assessment: *Journal of Applied
896 Geophysics*, 108, 69-77, 2014.](#)
- 897 [Clavier, C., G. Coates, and J. Dumanoir, Theoretical and experimental bases for the dual-water model for
898 interpretation of shaly sands: *Society of Petroleum Engineers Journal*, 24, 2, 153-168, 1984.](#)
- 899 [Daly, C. and J. K. Caers, Multi-point geostatistics - an introductory overview: *First Break*, 28, 9, 39-47,
900 2010.](#)

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

902

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

904

905 [Geonics Limited, http://www.geonics.com/index.html: 2012.](http://www.geonics.com/index.html)

906 [Halkidi, M., Y. Batistakis, and M. Vazirgiannis, Clustering validity checking methods: Part II: Sigmod Record, 31, 19-27, 2002.](#)

907

908 [Härdle, K. W. and L. Simar, Applied Multivariate Statistical Analysis: Springer, 2012.](#)

909 [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.](#)

910

911

912

913 [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.](#)

914

915

916 [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.](#)

917

918

919 [Hotelling, H., Analysis of a complex of statistical variables into principal components: Journal of Educational Psychology, 24, 417-441, 1933.](#)

920

921 [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.](#)

922

923

924 [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.](#)

925

926

927 [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.](#)

928

929 [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.](#)

930

931

932

933 [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 and salinity distribution critical for the assessment of future sea water intrusion in response to sea level rise: Hydrology and Earth System Sciences, 16, 1845-1962, 2012.](#)

934

935

936

937 [Møller, L. 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.](#)

938

939

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

943 [Pebesma, E. J. and C. G. Wesseling, Gstat: A Program for geostatistical Modelling, Prediction and](#)
944 [Simulation: Computers & Geosciences, 24, 1, 17-31, 1998.](#)

945 [Raiber, M., P. A. White, C. J. Daughney, C. Tschirter, P. Davidson, and S. E. Bainbridge, Three-](#)
946 [dimensional geological modelling and multivariate statistical analysis of water chemistry data to analyse](#)
947 [and visualise aquifer structure and groundwater composition in the Wairau Plain, Marlborough District,](#)
948 [New Zealand: Journal of Hydrology, 436-437, 13-34, 2012.](#)

949 [Rasmussen, E. S., K. Dybkjær, and S. Piasecki, Lithostratigraphy of the upper Oligocene - Miocene](#)
950 [succession of Denmark: Geological Survey of Denmark and Greenland Bulletin, 22, 1-92, 2010.](#)

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

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

958 [Sandersen, P., F. Jørgensen, N. K. Larsen, J. H. Westergaard, and E. Auken, Rapid tunnel-valley](#)
959 [formation beneath the receding Late Weichselian ice sheet in Vendsyssel, Denmark: BOREAS, 38, 4,](#)
960 [834-851, DOI: 10.1111/j.1502-3885.2009.00105.x, 2009.](#)

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

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

967 [Seifert, D., T. O. Sonnenborg, J. C. Refsgaard, A. L. Højberg, and L. Trolborg, Assessment of](#)
968 [hydrological model predictive ability given multiple conceptual geological models: Water Resources](#)
969 [Research, 48, 6, DOI: 10.1029/2011WR011149, 2012.](#)

970 [Sen, P. N., Electrochemical origin of conduction in shaly formations: Society of Petroleum Engineers:](#)
971 [Presented at 62nd Annual Technical Conference and Exhibition, 1987.](#)

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

974 [Stafleu, J., D. Maljers, J. L. Gunnink, A. Menkovic, and F. S. Busschers, 3D modelling of the shallow](#)
975 [subsurface of Zeeland, the Netherlands: Geologie en Mijnbouw/Netherlands Journal of Geosciences, 90,](#)
976 [4, 293-310, 2011.](#)

977 [Strebel, S., Conditional simulation of complex geological structures using multiple-point statistics:](#)
978 [Mathematical Geology, 34, 1, 1-21, 2002.](#)

979 [Triantafylis, J. and S. M. Buchanan, Identifying common near-surface and subsurface stratigraphic units](#)
980 [using EM34 signal data and fuzzy k-means analysis in the Darling River valley: Australian Journal of](#)
981 [Earth Sciences, 56, 535-558, 2009.](#)

982 [Turner, A., Challenges and trends for geological modelling and visualisation: Bulletin of Engineering](#)
983 [Geology and the Environment, 65, 2, 109-127, 1-5-2006.](#)

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

986 [Waxman, M. H. and L. J. M. Smits, Electrical Conductivities in Oil-Bearing Shaly Sands: Society of](#)
987 [Petroleum Engineers Journal, 8, 107-122, 1968.](#)

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

990 [Wu, j., Advances in K-means Clustering: A Data Mining Thinking: Springer,2012.](#)
991
992

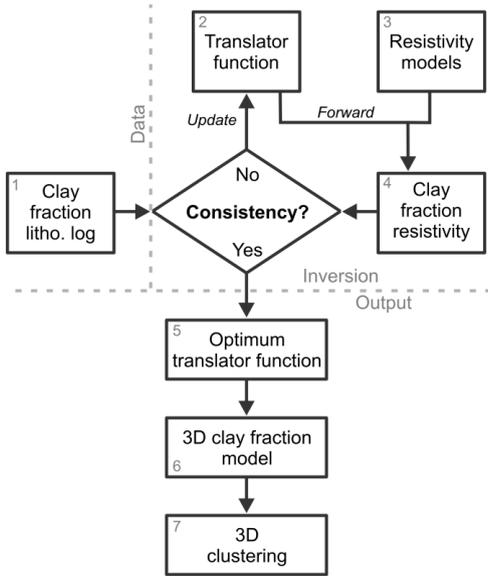
Deleted: ¶

¶
¶

996 **FIGURES AND FIGURE CAPTIONS**

997 **Figure 1**

998



999

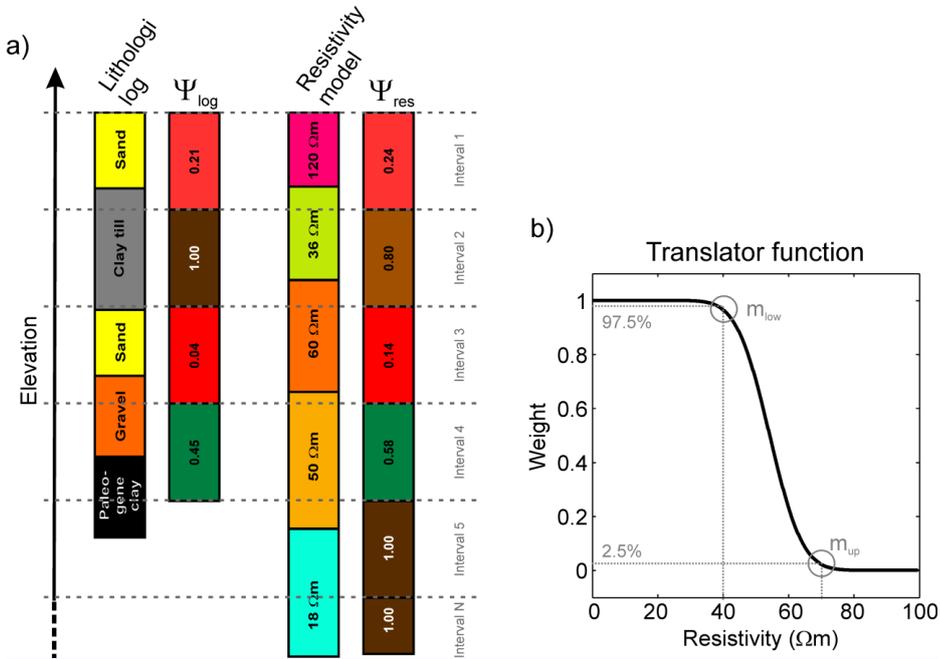
1000 Figure 1. Conceptual flowchart for the CF-procedure and clustering.

1001

Deleted: clay fraction concept
Deleted:
Deleted: .

1005

Figure 2



1006

1007 Figure 2. a) Example of how a lithological log translates into a Ψ_{log} and how a resistivity model translates
 1008 into Ψ_{res} , for a numbers of calculation intervals. The resistivity values and the resulting clay fraction
 1009 values are stated on the bars, but also indicated with colors with reference to the color scales of Figure 7.
 1010 b) The translator function returns a weight, W, between 0 and 1 for a given resistivity value. The
 1011 translator function is defined by the two parameters m_{low} , and m_{up} . In this example the m_{low} , and m_{up}
 1012 parameters are 40 Ωm and 70 Ωm respectively.

1013

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

Deleted: on

Deleted: is

Deleted: s

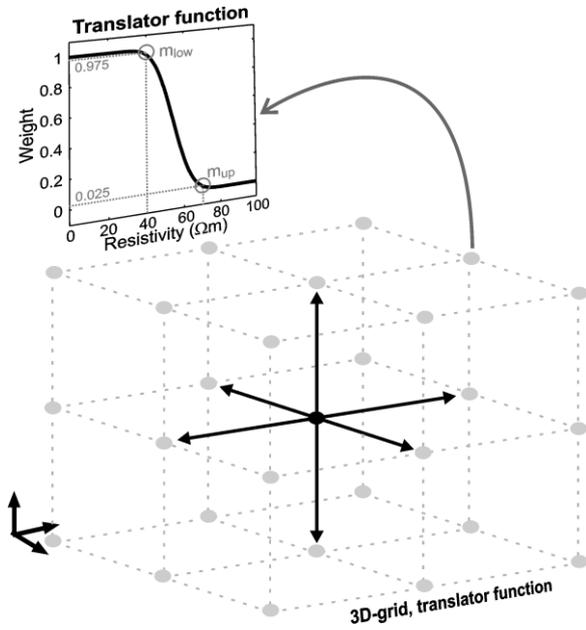
Deleted:

1025

Figure 3.

Deleted: 2

1026



1027

1028 | Figure 3. The translator function and 3D translator function grid. Each node in the 3D translator function
 1029 | grid holds a set of m_{up} and m_{low} . The m_{up} and m_{low} parameters are constrained to all neighboring
 1030 | parameters, as indicated with the black arrows for the black center node.

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

Deleted: .

Deleted:

1041

Figure 4

Deleted: 3

1042

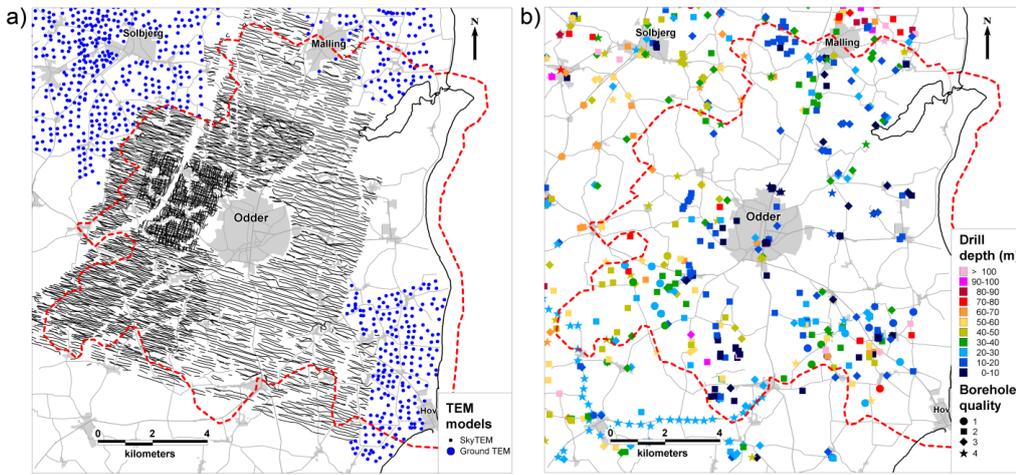


1043

1044 Figure 4. The black square marks the Norsminde survey area.

1046

Figure 5



Deleted: 4

1047

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 km²).

Deleted: Location of the r

Deleted: s

Deleted: s

Deleted: s

Deleted: . Small dots are

Deleted: models, larger scattered dots

Deleted: are

Deleted: models

Deleted: and borehole

Deleted:

Deleted: where

Deleted: modeling area

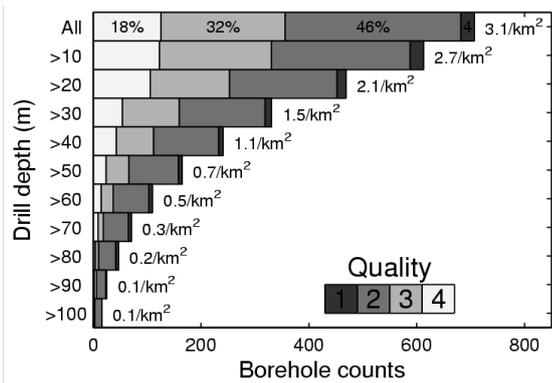
Deleted: 5

1051

1052

Figure 6

1053



1054

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 km² at the different depths. The color coding of the bars marks the borehole quality grouping.

Deleted: y

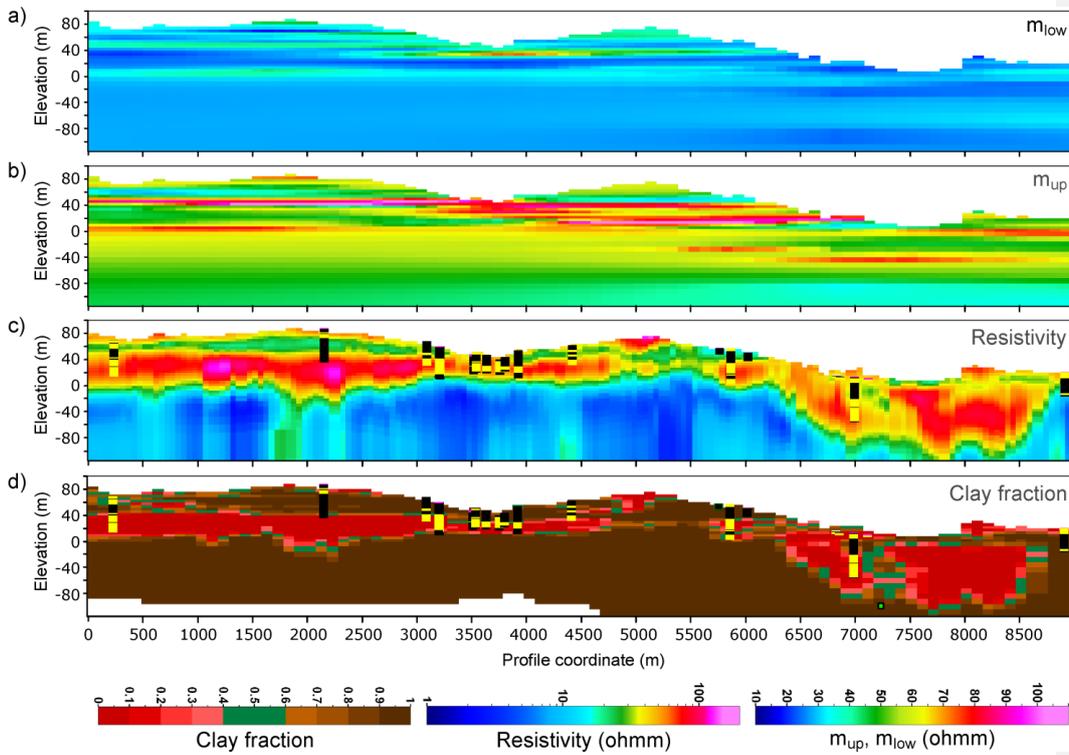
1057

1073

Figure 7

1074

Deleted: -----Page Break-----
F



1075

1076 Figure 7. Northwest-southeast cross sections (vertical exaggeration x6). Location and orientation of [the](#)
 1077 cross sections are marked in Figure 8. a) The m_{low} parameters of the translator function. b) The m_{up}
 1078 parameters of the translator function. c) The resistivity section with boreholes within 200 m of the profile
 1079 superimposed. [Black and yellow vertical bars show the position of boreholes](#): Black [blocks](#) mark the clay
 1080 layers, [and yellow blocks](#), mark sand and gravel layers. d) Clay fraction section and boreholes (same
 1081 [boreholes](#) as plotted in the resistivity section).

1082

Deleted: s

Deleted: borehole colors

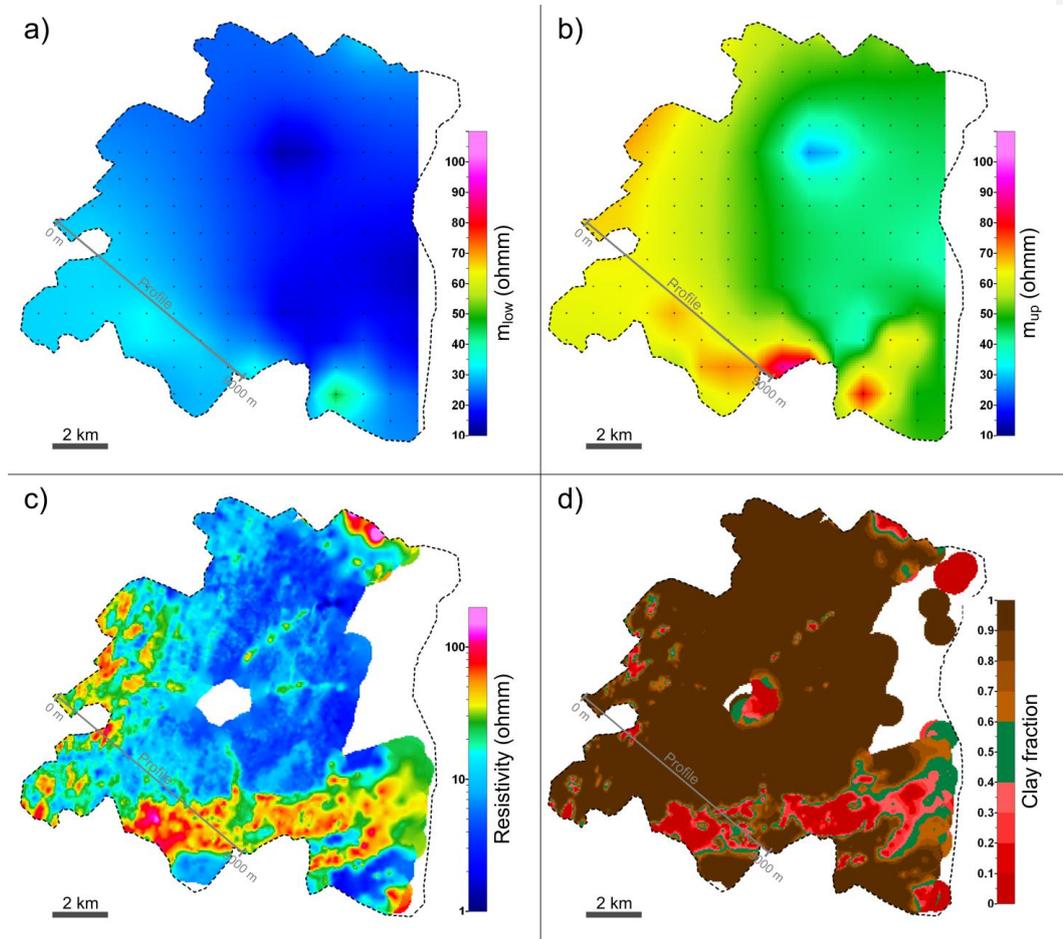
Deleted: while

Deleted: colors

1090
1091

Figure 8

Deleted: 7



1092

1093 Figure 8. Horizontal slices at 2 m_bsl cropped to the catchment area (dashed line). a) The m_{low} parameters
1094 of the translator function superimposed with the 1 km translator function grid (black dots). b) The m_{up}
1095 parameters of the translator function superimposed with the 1 km translator function grid (black dots). c)
1096 Resistivity slice (interpolated). Note that no EM-data is available around the town of Odder (see Figure
1097 5a) resulting in a “hole” in the resistivity map. d) Resulting CF-model. The hole in the resistivity map is
1098 here partly closed because CF-values from boreholes are available in this area.

Deleted: Interpolated r

Deleted: No

Deleted: town

Deleted: center

Deleted:

Deleted: center

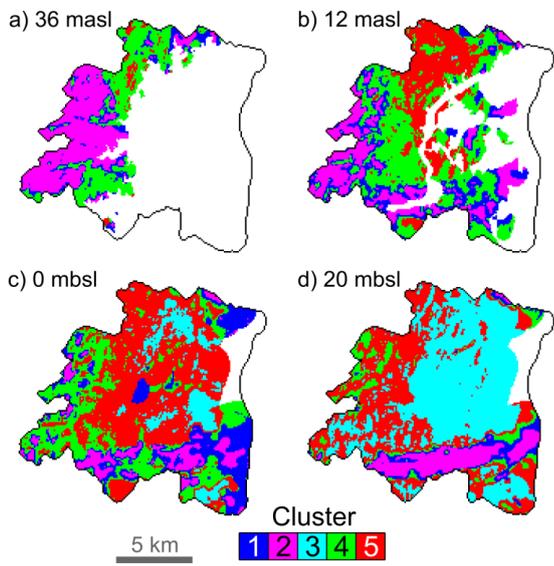
Deleted: since

Deleted: is

Deleted: ————— Page Break —————

1109 **Figure 9.**

1110



1111

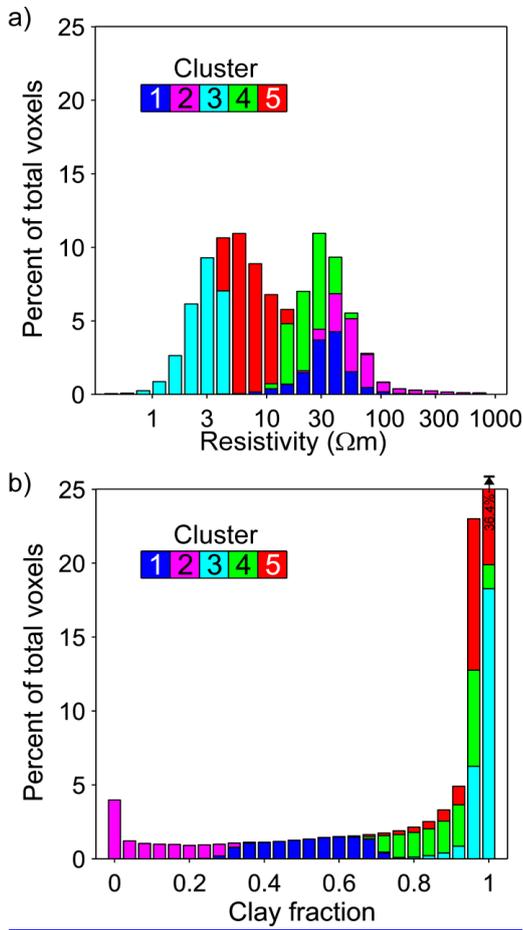
1112 Figure 9. Horizontal slices in four depths of the 3D cluster model.

1114

Figure 10

Deleted: 9

1115



1116

1117 Figure 10. Cluster statistics. The histograms show which data from the original variables make up the five
1118 clusters. a) The distribution of the resistivity data in the five clusters. b) The distribution of the CF data in
1119 the five clusters.

1120