

Using geophysical survey results in the inference of aquifer vulnerability measures

Niels B. Christensen* and Anders V. Christiansen

Department of Geoscience, University of Aarhus, Nordre Ringgade 1, 8000 Aarhus C, Denmark (E-mail: nbc@geo.au.dk)

Received February 2021, revision accepted May 2021

ABSTRACT

The physical parameter derived from the inversion of electromagnetic surveys, the distribution of subsurface conductivity, is interesting in itself only in very few instances. In most cases, the conductivity distribution will have to be interpreted in terms of the target properties of the survey, for example: a geological interpretation of lithology; a hydrogeological interpretation of hydraulic conductivity; a biohazard/geotechnical interpretation of polluted/not-polluted ground; and/or an archaeological interpretation of manmade/natural finds. The parameters of interest in these categories are often called derived products, indicating that the parameter of interest is not the same as the parameter whose distribution is found in the inversion process of the geophysical data. The interpretation process can be done in a wide variety of ways; from a predominantly cognitive approach based on professional experience, to an application of rigorous quantitative relations found from scientific endeavours.

In most practical situations, the number of locations with independently measured information on the derived product is considerably smaller than the number of geophysics locations. It is precisely this sparsity of primary information on the derived product that encourages the use of geophysical inversion results as a sort of qualified interpolator through a formulation of a correlation between a geophysical parameter and the parameter characterising the derived product. In this paper, a general, quantitative approach to deriving the parameter of interest is presented using statistical analytic measures and an advanced use of an interpolation method that takes uncertainties into account. The approach is demonstrated in a field example from Ølgod, Denmark, where the cumulated clay thickness in the upper 30 m is estimated using a combination of borehole drilling records and an airborne transient survey.

Key words: Aquifer vulnerability, Derived products, Geophysics.

INTRODUCTION

Apart from the situation of grounding electrical infrastructure elements, such as power lines and transformer stations, the subsurface conductivity is not interesting in itself; it needs to be interpreted/translated into values or categories of the parameter of interest (PI). Examples are found in essentially every published paper on geophysical surveys with areal cover-

age. The geophysical method referred to in this paper is an airborne transient electromagnetic (EM) survey, but the methodologies of this paper are relevant for other types of combination of a geophysical survey with areal coverage and sparse, direct measurements of the PI.

The most common way of deriving estimates of the PI is to formulate a relationship between the parameter of interest and a geophysical parameter derived from the EM inversion and then ‘calibrate’ the relationship. That is, determine the parameters defining a functional relation, based on the measured

*E-mail: nbc@geo.au.dk

values of the PI and the corresponding values of the geophysical parameter at the PI locations.

EM and resistivity data are commonly used in hydrological investigations to derive information of importance to hydraulic modelling and related environmental characterizations. Hinnell *et al.* (2010) presented an overview of approaches used to extract hydrological information from geophysical data. They focus on hydrogeophysical applications in which a hydrological model parameter (e.g., water content and solute concentration) is indirectly estimated using a geophysical method. Several other efforts of applying geophysical methods in hydrogeological context could be mentioned here (Gonzales *et al.*, 2019; Commer *et al.*, 2020; De Carlo *et al.*, 2020). In the same context, Hubbard and Rubin (2000) present an overview of available correlations used to link geophysical data with hydrological model parameters. Linking electrical conductivity to hydrological parameters is, in many cases, a two-step procedure where conductivity is first interpreted in terms of lithology, and then subsequently the lithology is linked to hydraulic properties. Gunnink and Siemon (2009), Jørgensen *et al.* (2012) and Jørgensen *et al.* (2013) build a full three-dimensional lithological clay-sand model from an airborne EM investigation targeted for further hydrological use. Kirsch *et al.* (2003) use geophysical data to produce an aquifer vulnerability index based on the electrical conductance in depth intervals (the product of electrical conductivity with layer thickness) to estimate clay content. In a study by Christiansen *et al.* (2014), some of the elements of the above approach are implemented to predict clay thickness over a survey area from measurements in boreholes with the aim of deriving a measure of aquifer vulnerability to pollutants infiltrating from the surface. An intuitively obvious application is one where the electrical conductivity is used as a proxy for pore water salinity, for example, Rhoades *et al.* (1989), Ley-Cooper *et al.* (2008) and Shevnev *et al.* (2007) amongst many others. There are numerous studies where porosity is linked with seismic velocities and their variation with lithology (e.g.; Han *et al.*, 1986; Klimentos, 1991; Graeves *et al.*, 1996), and Paasche *et al.* (2006) use tomographically inverted seismic velocity, radar attenuation and velocity data to arrive at porosity estimates.

A CONCEPTUAL ANALYSIS OF THE AIMS OF THIS STUDY

The issue addressed in this paper is the prediction of the cumulated clay thickness in the upper 30 m of the ground in a survey area as a proxy for the vulnerability of an underly-

ing aquifer towards infiltration of unwanted substances from the surface: for example, pesticides, herbicides, nitrates, phosphates, and other types of pollutants that might degrade the quality of the underlying aquifer. In the survey area, a limited number of boreholes are found where the clay thickness in the upper 30 m is estimated from drilling reports. However, the borehole coverage is quite sparse, and to assist in the interpolation/extrapolation of the clay thickness over the entire survey area, the conductivity models from the inversion of an airborne transient electromagnetic survey which cover the area much more densely are brought into play. This is done by assuming that a correlation exists between the cumulated clay thickness and the mean conductivity of the inversion models in the same depth interval.

Prediction of a parameter value at locations where it has not been measured is a problem of interpolation/extrapolation: the predicted value is found as a weighted sum of the measured data. There are a plethora of methods of interpolation, but there are three fundamental parameters that determine the weights: the uncertainty of the primary data (if available at all); a function of distance between the primary data and the location of the predicted data; and a selection criterion on which the primary data should come into play. As an example, different types of kriging come with and without including the data uncertainty and with various selection criteria and distance functions. Natural Neighbour interpolation is built on a Delaunay triangulation and involves the voronoi cell in which the location of the predicted data is situated and the surrounding cells, but does not take the data uncertainty into account. If the interpolation problem is solved as an inverse problem – as is the case with the Lateral Parameter Correlation method to be introduced below – the uncertainty of the primary data is included as a data error covariance matrix, and the distance issue is taken care of by including a model covariance matrix. In the case of including a second data set consisting of data correlating with the primary data, an approach will have to be formulated to the three basic issues mentioned above, and the correlation must be quantified in some way.

Background

In Denmark, 99% of the drinking water comes from untreated groundwater. The fact that we have access to clean groundwater is a very precious asset, and a comprehensive mapping effort has taken place over the past decades to gain insight into the presence and state of this resource and to develop dynamic hydrogeological models that would provide a scientific basis for sustainable abstraction in the future.

The National Groundwater Mapping Programme was started in 1999 and the first phase of the program ended in 2015. Note that 40% of the Danish land area was involved in the investigations, using hydrogeophysical, geological and hydraulic sciences, and a large number of methods and procedures (Thomsen *et al.*, 2004; Møller *et al.*, 2009a, 2009b). The aims of the program were to locate and estimate the potential of all the most important aquifers in Denmark, to assess their vulnerability to pollution, and, based on these findings, to establish political and administrative measures to protect the groundwater. At the moment, new areas are being included in an extension of the project.

The issue of vulnerability is a very complex one; it inherits the complexity of the geological setting in an area created and disturbed by several consecutive glaciations, and, as a consequence, the complexity and spatial variability of infiltration from the surface. However, assessing the vulnerability of aquifers is crucial for a sustainable administration of the natural resources, no matter the complexity of the task. It should be regarded as a ‘must’ in a program of this scope.

Concepts and assumptions

The first postulate regarding the vulnerability issues is: Clay thickness is a good proxy for the level of protection. However, this statement is a contested one, and for good reasons. It is known – and was studied as part of the National Program – that clay formations, especially near-surface formations of moraine clays, are fractured, meaning that such clay formations do not offer the degree of protection that can be evaluated based on their thickness alone (Reynolds and Kueper, 2001; Aguilar-López *et al.*, 2020).

The next postulate is that clay thickness can be estimated from borehole drilling reports. However, drilling reports span a large range of reliability that depends on the drilling method, the geological expertise of the drilling crew, whether the borehole was cored for later close inspection, etc. This means that clay thicknesses estimated from boreholes come with a large uncertainty that necessarily also must be assessed.

The third postulate that needs to be addressed is that attributes that correlate well with the clay thickness can be extracted from inversion models of electromagnetic data with areal coverage and serve as a proxy for the clay thickness and thereby assist in a more reliable interpolation/extrapolation of the primary borehole clay thickness data. In this paper, we have suggested the electrical conductance (or equivalently: the mean conductivity) of the depth interval in question as a proxy. However, clays are not just clays; they can have quite

varied electrical conductivities, so clay thickness is not always a simple, increasing function of mean conductivity.

With the uncertainties outlined above, both conceptually and with regard to the quality of measured/estimated data, it could be argued that a task of estimating clay thickness as a proxy for the degree of aquifer protection is unreasonably optimistic and that any results should be taken with a large pinch of salt. We agree! However, as mentioned, estimating the vulnerability of aquifers is crucial for a sustainable abstraction of groundwater, so it is a given task! It simply has to be done! Taking all of the above into account we have responded to this inevitability by presenting methods that we find appropriately simple, considering all the uncertainties, but also adequately robust and productive. An important aspect has also been to present a methodology with some generality and applicability in other situations with a challenge of estimating derived products.

METHODS AND TOOLS

The methods/approaches used in this study will not be presented in a dedicated theoretical section of the paper. They will be introduced along the way with a discussion about their suitability in connection with the field example and the actual task of predicting clay thickness over an entire survey area from a combination of clay thickness data, as estimated from borehole drilling reports, and the mean conductivity found in inversion results of a airborne transient EM (TEM) survey. The methods are presented in more detail in two Appendices.

There are mainly two methods that will be brought into play: The Lateral Parameter Correlation method (LPC); and the Continuous Wavelet Transform (CWT) methodology. The LPC was developed as a technique for lateral correlation of one-dimensional earth models after individual inversions of the data from a survey area (Christensen and Tølbøll, 2009; Christensen, 2016b). In the present context, the LPC is used mainly as an advanced interpolation tool, and it is introduced in Appendix A. It is characterized by being able to take both the data uncertainties and the spatial correlation between the input data into account and it permits all data to influence all other data. The spatial correlation is defined by the model covariance matrix of the inversion setup defining the LPC method and introduces a smoothing of user-defined strength.

The CWT methodology is most often used to find the ‘natural’ layer boundaries in a complex log (e.g., Hill and Uvarova, 2018; Hill *et al.*, 2020) or as a tool to find ‘natural’ clusters/bins in a set of parameters (Christensen, 2018), and it is introduced in Appendix B. In essence, the CWT is a tool to

reduce the complexity of a given data set to a user-defined level, approximating the original data set with a piecewise constant function defined by the interval (bin) boundaries and the resulting mean values within the bins. A unique property of the CWT analysis is that it offers a hierarchy of importance between the bin boundaries.

Nomenclature

A word about nomenclature: in the following text, the following symbols are used:

T_c^{BH} The cumulated clay thickness of the upper 30 m of the ground estimated from borehole drilling reports.

ΔT_c^{BH} The estimated absolute uncertainty of T_c^{BH} estimated from borehole drilling reports.

T_c^{GFX} The predicted/interpolated value of clay thickness at the geophysics locations.

ΔT_c^{GFX} The estimated absolute uncertainty of T_c^{GFX} .

C_m^{GFX} The mean conductivity of the upper 30 m of the ground calculated from the inversion models of the airborne TEM data as the cumulated conductance (sum of conductivity times layer thickness) divided by 30 m.

C_m^{BH} The mean conductivity values interpolated to the borehole positions.

Interpolation methods of the plots

Efforts have been made in all of the figure captions to make it clear how the plots have been produced, and the reader should be aware of the following:

1 All of the rendering of the plot frames in the figures, whether they display raw data or processed/predicted data, have been produced in MATLAB using the Natural Neighbour (NN) interpolation option, (e.g., Sibson, 1981; Belikov *et al.*, 1997; Ledoux *et al.*, 2005). All interpolation methods are based on some assumption about the lateral correlation between points. For every point where a value is to be predicted, the primary data are weighted in a way that depends on the distance between the selected point and a selection of the primary data – or all of them. The NN interpolation can be described as a generalization of a linear interpolation scheme. It is quite robust with no or very few artefacts. Interpolation is possible only to points within the convex hull of the set of primary data points, and it does not take the uncertainty of the primary data into account.

2 In some of the figures, raw data have first been subjected to an interpolation – either to the original data positions or interpolated/extrapolated to other positions – using the LPC

method that takes both the data value and the data uncertainty into account and introduces a lateral correlation, that is, a lateral smoothing. The strength of the spatial correlation is user-definable, and in this paper it is decided using an ‘Occam’s razor’ principle: the error-normalized residual between raw and LPC-interpolated data is of the order of unity, meaning that the result is an even-handed compromise between acknowledging the information contents of the data and the wish to produce a smooth result that does not depend too much on data outliers.

FIELD EXAMPLE: USING AIRBORNE EM DATA TO ESTIMATE CLAY THICKNESS

Borehole data and geophysics data

We present a field case from the area of Ølgod, Denmark, which has been the object of dense geophysical mapping with electrical, electromagnetic and seismic methods in connection with the KOMPLEKS project 2008–2012 aimed at mapping large-scale geological structures to improve hydraulic modelling (Høyer *et al.*, 2011, 2013a, 2013b, 2014; He *et al.*, 2013). In this paper, we focus on the airborne transient electromagnetic survey carried out with the SkyTEM system in 2006 and 2009 (Sørensen and Auken, 2004; Auken *et al.*, 2006, 2009, 2017) and inverted with the WorkBench program (Auken *et al.*, 2014). Throughout the field example, the parameter that we are interested in and wish to predict over the entire survey area in the best possible way is the cumulated thickness of clay within the upper 30 m of the ground, T_c , based on the borehole data, T_c^{BH} , and the mean conductivity of that depth range calculated from results of inversion of the TEM data, C_m^{GFX} . This is one of the standard measures in the national mapping effort.

The location of the survey area, the TEM flight lines and the borehole positions are shown in Fig. 1. All coordinates in this paper are given as UTM coordinates for zone ED50/32N. The primary data volume consists of 15,948 TEM soundings and 168 boreholes with estimates of T_c^{BH} . Note that the final edited coverage of TEM soundings displays fairly large ‘holes’ where data sets have been discarded due to cultural electromagnetic (EM) coupling.

The data quality of T_c^{BH} varies considerably from borehole to borehole, depending on the age of the bore, the drilling method and the bore crew’s level of expertise in describing the borehole material. Every borehole drilling report has been carefully examined by an experienced geologist to estimate T_c^{BH} and ΔT_c^{BH} . This resulted in defining four

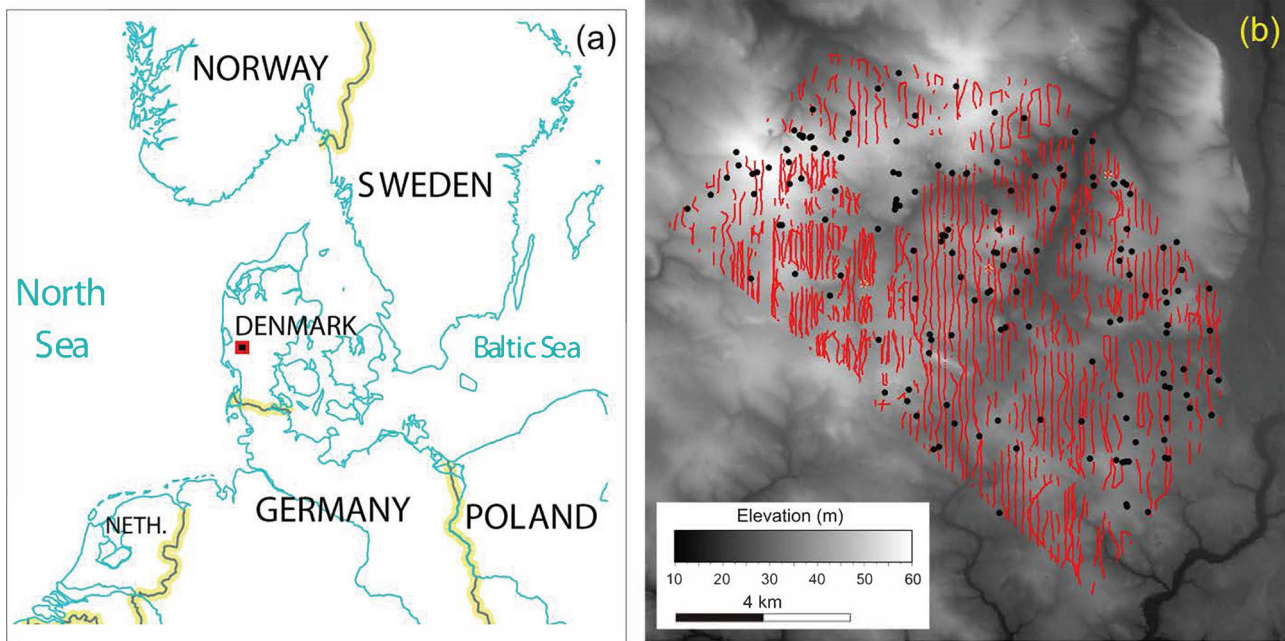


Figure 1 (a) The location of the Ølgod survey area. (b) Flight lines of the airborne TEM mapping in red together with the borehole positions where clay thickness is estimated marked by black dots. The survey area lies in the UTM coordinate interval of: E 470000 to 484000 and N 6174000 to 6188000. The geographical coordinates of the centre of the survey area are: Lat 55.7741° and Long 8.6334°.

categories of ΔT_c^{BH} of 3, 6, 9 and 15 m, reflecting the reliability of the drilling reports. The values of T_c^{BH} and the uncertainty ΔT_c^{BH} are plotted in Fig. 2a and b, respectively.

All the data used in this study are publicly available from the GERDA data repository administered by GEUS: The Geological Survey of Denmark and Greenland, and can be retrieved at: <https://eng.geus.dk/products-services-facilities/data-and-maps/national-geophysical-database-gerda>.

In the context of this paper, we seek to develop a methodology that permits an estimate of the clay thickness over the survey area, which is an important parameters in the subsequent geological-hydrogeological interpretation of the area. The details of the actual interpretation can be found in Høyer *et al.* (2013b), which also shows plots correlating seismic profiles and resistivity profiles from the EM survey with overlapping borehole lithology logs.

Predicting clay thickness through interpolation

The most simple approach to predicting T_c^{GFX} over the entire survey area is to just interpolate T_c^{BH} without involving geophysical information. Figure 2a shows the result of using a Natural Neighbour (NN) interpolation method on the raw T_c^{BH} data, and Fig. 2b shows the uncertainty, ΔT_c^{BH} . Figure 2c shows a map of the interpolated borehole clay thickness using

a more advanced interpolation tool: The Lateral Parameter Correlation (LPC) method. The LPC method produces interpolated values taking both ΔT_c^{BH} and the lateral position of the borehole data into account, see Appendix A. The LPC formulates the prediction as an inversion problem involving the identity mapping where the uncertainty is included as a data covariance matrix. The LPC method also includes a model covariance matrix that imposes a user-defined smoothness on the result. The standard deviation of the model covariance matrix determines the degree of smoothness and has been chosen so that the error-weighted residual between T_c^{BH} and the predicted value of the clay thickness at the borehole positions is approximately unity. Throughout this paper, the criterion of an error-weighted residual of approximately unity, an ‘Occam’s razor’ principle, has been used to define the strength of the lateral correlation of the LPC interpolations. A more comprehensive presentation of the LPC method is given in Appendix A. The clay thickness values of Fig. 2c are predicted to all borehole and geophysics positions over the area before plotting, but do not involve any geophysical information.

Correlation between borehole clay thickness and mean conductivity

The scarcity of primary T_c^{BH} data and the dense areal coverage with TEM soundings invites the use of the latter to

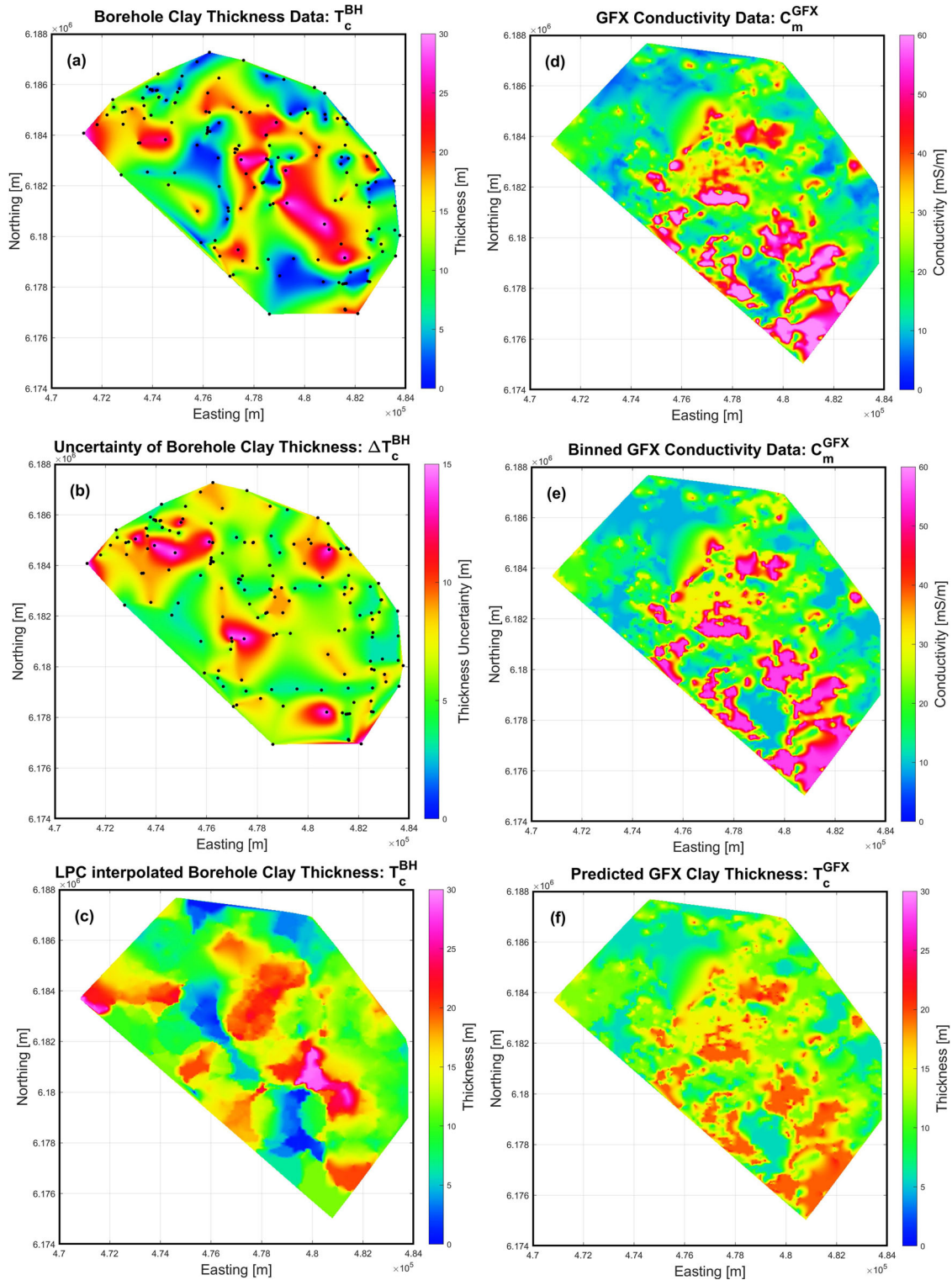


Figure 2 (a) Simple NN plot of the borehole clay thickness data, T_c^{BH} , over the survey area. (b) Simple NN plot of the uncertainty of the borehole clay thickness data, ΔT_c^{BH} , over the survey area. (c) Plot of the predicted borehole clay thickness, T_c^{BH} , over the entire area using the LPC method on the primary borehole data. (d) Simple NN plot of the mean conductivity, C_m^{GFX} , over the survey area. (e) Simple NN plot of the binned mean conductivity over the survey area, C_m^{GFX} . (f) Simple NN plot of the clay thickness over the survey area, T_c^{GFX} , derived from the binned mean value of the conductivity. In all of the subplots, the unit of clay thickness is metres, and the unit of mean conductivity is mS/m.

Table 1 The table shows the correlation coefficients between T_c^{BH} and C_m^{BH} for a series of selection criteria based on $\Delta T_c^{BH} < T_c^{BH} >$ and $< C_m^{BH} >$ indicate the mean values (bias) of T_c^{BH} and C_m^{BH} . For every selection criterion, the table also shows the number of boreholes involved

BH Selection	# of BH	$< T_c^{BH} > [m]$	$< C_m^{BH} > [mS/m]$	Correlation
All BH	168	11.56	22.53	0.42
$\Delta T_c^{BH} \leq 9 \text{ m}$	156	11.33	22.61	0.46
$\Delta T_c^{BH} \leq 6 \text{ m}$	96	12.11	23.57	0.53
$\Delta T_c^{BH} \leq 3 \text{ m}$	52	13.67	24.97	0.64

assist in improving the prediction of T_c^{BH} over the entire survey area. This requires defining a parameter derived from the geophysics models that correlates well with T_c^{BH} . In our case, the geophysical parameter that is deemed to correlate the best with clay thickness – and thereby be the best one to guide an extrapolation of T_c^{BH} – is the mean conductivity in the depth interval 0–30 m, C_m^{GFX} . Among the lithologies present in the Ølgod area, clay formations have the highest conductivity, and it is expected that T_c^{BH} would be an increasing function of C_m . Figure 2d shows a plot of C_m^{GFX} , the mean conductivity in the top 30 m, derived from the TEM inversion models.

In the Danish open geophysics database: GERDA (Møller *et al.*, 2009a, 2009b), data, system parameters and inversion results of all recent geophysical surveys in Denmark can be found, but unfortunately the posterior uncertainty of C_m^{GFX} is not part of the stored information. Neither is the posterior covariance matrix of the inversions from which we could have derived the uncertainty. However, experience from inversions of similar TEM data sets with another inversion program (Christensen, 2016a) indicates that the relative uncertainty of the mean conductivity in the top 30 m is around 0.2 with only small variations. In the depth range of 0–30 m, there are no problems at all with depth penetration or resolution (Christiansen *et al.*, 2014; Christensen, 2021), and, furthermore, the mean conductivity in a certain depth interval is often the best determined parameter in a TEM sounding.

TEM sounding locations do not necessarily coincide with the borehole positions, so to estimate the correlation between the two parameters, C_m^{GFX} values have been interpolated to all borehole positions. The interpolation is carried out using the Lateral Parameter Correlation method, see Appendix A, on $\log(C_m^{GFX})$ with a relative uncertainty of 0.2. Figure 3, plot frames a–d, shows plots of T_c^{BH} as a function of C_m^{BH} . The figure reveals a quite erratic picture from which it is not easy to infer a correlation. A strategy sometimes suggested when using geophysical measurements to improve the interpolation of T_c^{BH} is to fit a simple smooth function that maps C_m^{BH} into T_c^{BH} (Christiansen *et al.*, 2014). However, it does not seem justifiable to fit the erratic data in Fig. 3 with a smooth func-

tion, and instead we shall use statistical tools to express the relationship in an adequately simple way, thereby also avoiding having to justify the many assumptions lying behind the choice of a smooth function.

To investigate the correlation between T_c^{BH} and C_m^{BH} , a simple statistical analysis was carried out for different selection criteria based on the absolute error, ΔT_c^{BH} . The results are seen in Table 1. It is seen that the mean values of T_c^{BH} and C_m^{BH} do not change much for the different selections, and that the correlation coefficient, as expected, increases when the selection becomes more restrictive. The values of the correlation coefficients indicate a weak, but not vanishing correlation, highlighting the challenge to choose a proper data processing and a proper quantitative definition of the mapping from C_m^{BH} to T_c^{BH} .

Instead of postulating a simple smooth functional relationship, the method that we shall advocate is to approximate the mapping with a simple piecewise constant function. To do so, an appropriate number of conductivity intervals need to be defined within which a weighted mean value of T_c^{BH} will be found, so the question naturally arises: What is the best way of selecting the number of intervals and their boundaries? To choose the interval boundaries for the conductivity, the second of the methods that play a major role in the methodology suggested in this paper: the Continuous Wavelet Transform (CWT), is used on all values of C_m^{GFX} . The method is outlined in Appendix B where its use in finding ‘natural’ intervals for a parameter is explained, see also Christensen (2018).

In Fig. 4, the CWT spectrum (scalogram) pertaining to the binning of C_m^{GFX} and the mean values of C_m^{GFX} for each of the bins are shown (see Appendix B for a more detailed description). It is seen from the CWT spectrum that the cases of two, four and six bins are quite distinctly defined while the number of bins increases rapidly from eight bins and upwards when reducing the averaging only slightly. An additional constraint on the binning, demanding that the relative difference in bin averages must be larger than 0.2, has the result that both the 8- and 10-bin cases will reduce to six bins. For these reasons, we have chosen six bins.

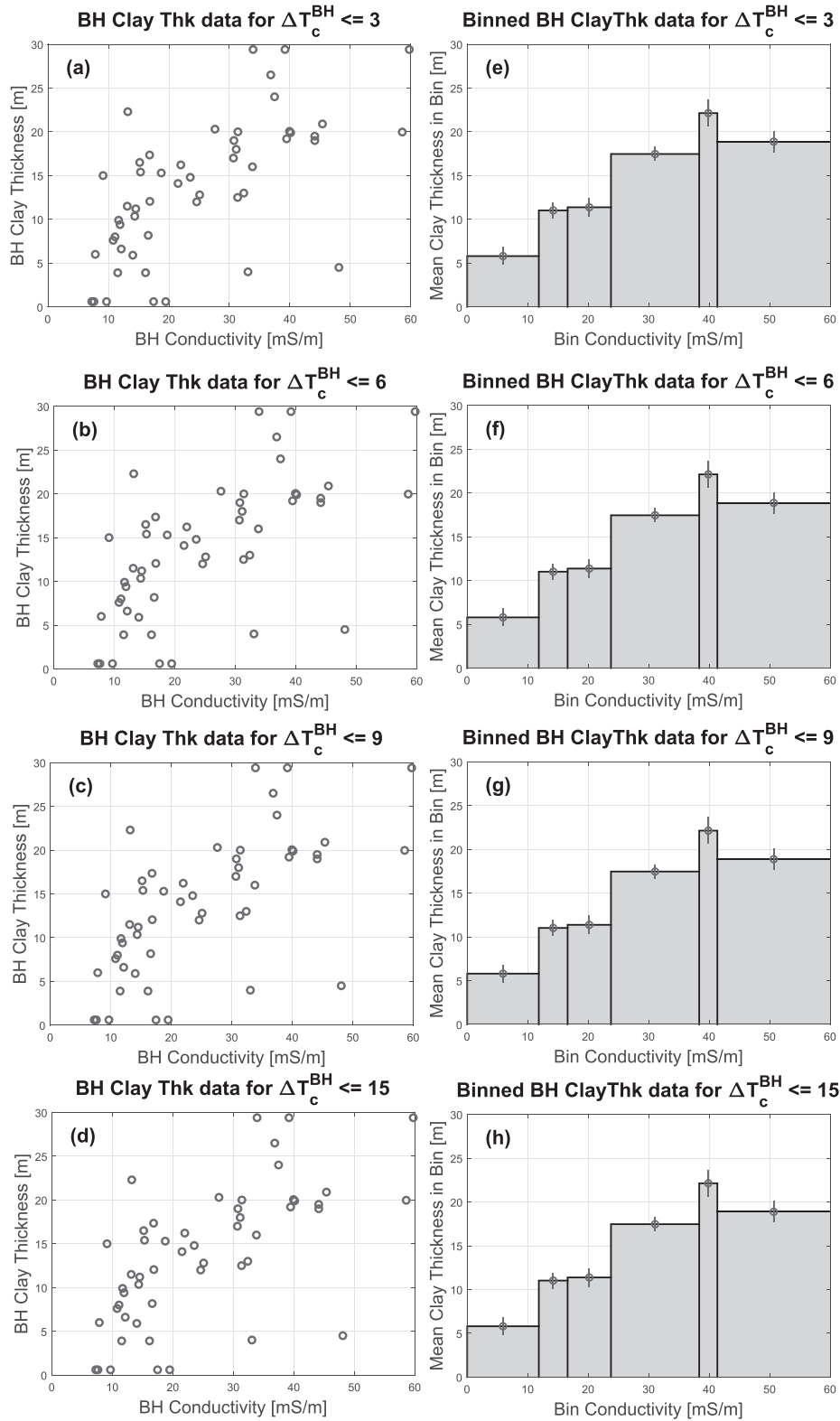


Figure 3 (a) Plot of T_c^{BH} as a function of C_m^{BH} at the borehole locations for the selection $\Delta T_c^{BH} < 15$ m (all boreholes). (e) The binned value of T_c^{BH} for each of the six bins of C_m^{FX} found through the CWT analysis. The next three rows, (b)–(e) and (f)–(h), show similar plots for the selections: $\Delta T_c^{BH} < 9$ m, $\Delta T_c^{BH} < 6$ m and $\Delta T_c^{BH} < 3$ m. The binned models of clay thickness in subplots (e)–(h) are shown as histograms with the mean value marked with circles with a vertical line indicating the standard deviation.

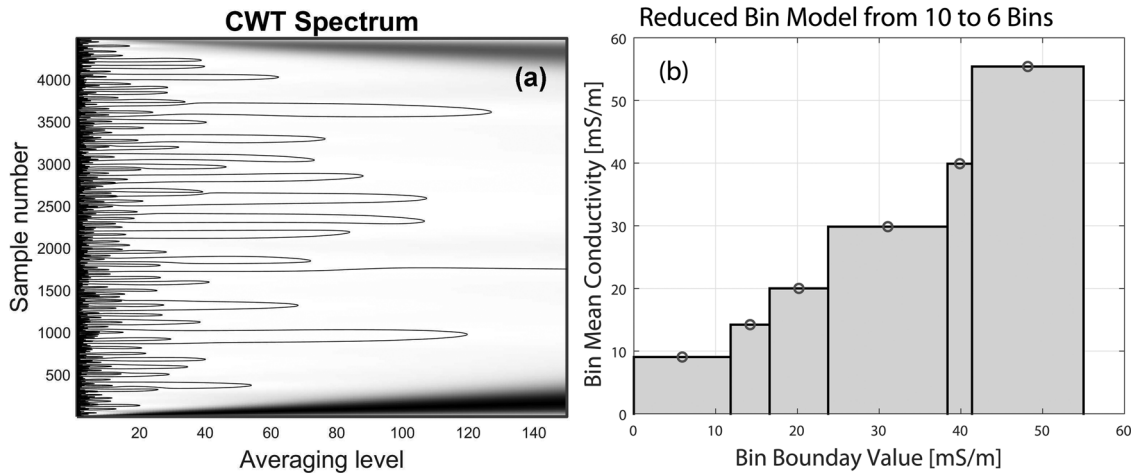


Figure 4 (a) The CWT spectrum (scalogram) of C_m^{GFX} values over the survey area. (b) The piecewise constant function approximating C_m^{GFX} when using six bins (reduced from 10 bins).

Table 2 The table shows the conductivity intervals of the six bins defined by the CWT analysis and the resulting mean values of the mean conductivity: $\langle C_m^{\text{bin}} \rangle$; plus the mean value of the clay thickness: $\langle T_c^{\text{bin}} \rangle$; and its uncertainty: $\Delta \langle T_c^{\text{bin}} \rangle$, within each of the bins shown in Fig. 4

Parameter	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Cond_Interval [mS/m]	0–11.84	11.84–16.60	16.60–23.77	23.77–38.36	38.36–41.36	41.36– ∞
$\langle C_m^{\text{bin}} \rangle$ [mS/m]	9.05	14.21	20.00	29.85	39.90	55.41
$\langle T_c^{\text{bin}} \rangle$ [m]	5.71	11.15	11.66	15.11	21.71	19.34
$\Delta \langle T_c^{\text{bin}} \rangle$ [m]	0.83	0.81	0.84	0.70	1.41	1.10

Within each of the bins, an uncertainty-weighted mean value of T_c^{BH} and its uncertainty ΔT_c^{BH} are found. The results are shown in Table 2. It is seen that $\langle T_c^{BH} \rangle$ is quite well defined with $\Delta \langle T_c^{BH} \rangle$ close to unity. The uncertainty-weighted residual between the original T_c^{BH} data and the collection of binned values is 1.42, indicating that the simplification of the piecewise constant function does not violate the original data much; it is quite reasonable. The bin boundaries and the resulting $\langle T_c^{BH} \rangle$ are also illustrated in Fig. 3e–h, where the binning and averaging results are shown for four possible selections of T_c^{BH} based on ΔT_c^{BH} . It is seen that the binned values do not change much for the different selection criteria. We take this as a confirmation of the robustness of the binning process. In the following, all T_c^{BH} data will be used.

Table 2 shows that the binning and averaging contradicts the assumption that T_c^{BH} will be a monotonically increasing function of C_m^{BH} . The sixth bin actually has a lower value of $\langle T_c^{BH} \rangle$ than the fifth bin. There can be several explanations for that: maybe the assumption that T_c^{BH} is an increasing function of C_m^{BH} is not generally valid; maybe the estimate of T_c^{BH} from the drilling reports was skewed towards lower values when T_c^{BH} was close to 30 m; and maybe there is

more than one sort of clay within the area. If some of the clays are very conductive, it is possible that less clay will be able to produce a higher C_m^{BH} . We tend to believe the latter. There are in fact several clay types in the area, both moraine clay and heavy Tertiary clays, the latter with very high conductivity, often above 100 mS/m. It is worth noting that this behaviour would have been missed is the case of using a smooth, monotonically increasing function to map the correlation between T_c^{BH} and C_m^{BH} .

A simple clay thickness prediction including both borehole and geophysical data

The simplification of the relation between T_c^{BH} and C_m^{BH} obtained in the Continuous Wavelet Transform (CWT) analysis can be used to predict T_c^{GFX} at all the geophysics locations. For each of the geophysics positions, the value of C_m^{GFX} is used to find the bin to which it belongs, and the value of $\langle T_c^{\text{bin}} \rangle$ for that bin is then used as the predicted value of T_c^{GFX} at that point. Figure 2e shows the map of binned conductivity values, and Fig. 2f shows the clay thickness predicted from the conductivity bins.

This is the simplest method of predicting T_c^{GFX} in a way that includes the geophysical information and the correlation between T_c^{BH} and C_m^{BH} as determined by the CWT binning. However, in this approach, every geophysical location is handled separately, so the spatial correlation is not taken into account; there is no lateral correlation involved. One might expect that this approach, using a discontinuous function and no spatial correlation, might result in a T_c^{GFX} map equally discontinuous and with sharp edges. However, evidently that is not the case. The C_m^{GFX} values come from a laterally correlated, that is, smooth, inversion of TEM data and, as seen in Fig. 2, the C_m^{bin} values produce a smooth map very similar to the one for C_m^{GFX} and less similar to the plot of T_c^{BH} seen in Fig. 2. The similarity between Fig. 2d and Fig. 2e testifies to the sufficiency and robustness of the six bins of the CWT analysis to describe the variability of C_m^{GFX} and thereby the interpolated values of T_c^{GFX} .

Spatially correlated clay thickness prediction using all data

In this section, a method that takes both T_c^{BH} and C_m^{GFX} into account in a way that involves the spatial correlation of all data will be presented. It is based on sorting the geophysics positions into six groups according to the bin value of C_m^{GFX} and performing the prediction of T_c^{GFX} for each group, one at a time. Subsequently, for visualization, the results are collected into one set of predicted values of T_c^{GFX} over the entire survey area. The binned values of C_m^{GFX} constitute a set where there is no overlap between any of the six bins and together they make up all of the C_m^{GFX} values of the entire survey area, that is, they are disjoint and their union is exhaustive. However, we cannot expect that the bins will be spatially compact, but that is of no importance for the prediction procedure. The distribution of bins can be gleaned from the map of binned C_m^{GFX} values in Fig. 2. The approach is built on the Lateral Parameter Correlation (LPC) method and is realized by ascribing a dummy value for T_c^{GFX} and a very large value of ΔT_c^{GFX} for all geophysics locations and including all borehole data T_c^{BH} and their uncertainties ΔT_c^{BH} . The LPC procedure will then predict a value of T_c^{GFX} at the geophysics positions (and a slightly smoothed value of T_c^{BH} at the borehole locations). In this way, the spatial correlation and the primary T_c^{BH} data are taken into account. For each of the bins, the structural information from the distribution of C_m^{GFX} is included in the procedure by penalizing the influence from the T_c^{BH} values that do not belong to the same conductivity bin as the C_m^{GFX} bin being processed. This is done by ascribing an additional data error covariance matrix to T_c^{BH} in such a way that the additional standard deviation is

proportional to the difference of the bin numbers so that the total variance ascribed to T_c^{BH} is

$$\text{var}_{\text{new}} = \text{var}_{\text{old}} + \text{var}_{\text{bindif}} = (\Delta T_c^{BH})^2 + [D_{\text{add}} \cdot (i - j)]^2, \quad (1)$$

where i denotes the number of the C_m^{GFX} bin of the geophysics data, j the bin number of the borehole data, and D_{add} is the additional uncertainty in metres (see also Appendix A). The value of D_{add} determines how much the T_c^{BH} data from neighbouring bins are down-weighted. A large value of D_{add} means a heavier down-weighting and vice versa. This means that as D_{add} increases, the result will reflect the situation where only the T_c^{BH} data from the same bin will have an influence (Fig. 5a); and for $D_{\text{add}} = 0$, the result will be the same as obtained when no binning is included, that is, C_m^{GFX} is not taken into account (Fig. 5d). Figure 5 shows the results for values of $D_{\text{add}} = \infty, 24, 6$, and 0 m. The plot for $D_{\text{add}} = 0$ is the same as the one seen in Fig. 2c.

The method outlined here fulfills the ideal criterion that both T_c^{BH} and C_m^{GFX} should have an influence on the predicted values of T_c^{GFX} , and it offers a option of weighting the relative influence of T_c^{BH} and C_m^{GFX} . However, it also comes with a requirement that the user must choose the parameter that determines the relative weight: D_{add} . Figure 5 illustrates the effect of some of the choices of D_{add} , but in essence, each user must decide for her/himself.

Comparing all of the plots in Figs 2 and 5, the assets and drawbacks of all the methods presented here can be compared and discussed in terms of how/if they fulfil the criteria for an optimal prediction of clay thickness which can be formulated as: The prediction method must take both the primary T_c^{BH} data with their uncertainty ΔT_c^{BH} and the C_m^{GFX} attribute into account in a balanced and consistent way.

The method of simple lookup in the list of $\langle T_c^{BH} \rangle$ for each of the conductivity bins gives a result that looks very much like the map of C_m^{GFX} seen in Fig. 1, however, with a smaller dynamic range – as should be expected when using average values. The method does not take spatial correlation into account and the results do not reflect the T_c^{BH} data as much as we would wish. The plots in Figure 5 illustrate that a large value of D_{add} makes the resulting map look more like the map of C_m^{GFX} with more short-wavelength details (see, Fig. 2d and e), and a small value makes the resulting map look more smooth and more similar to the map of T_c^{BH} (Fig. 2c).

Work flow overview

Now that all of the methods have been presented, it might be useful with an overview of the steps taken in the work flow. In

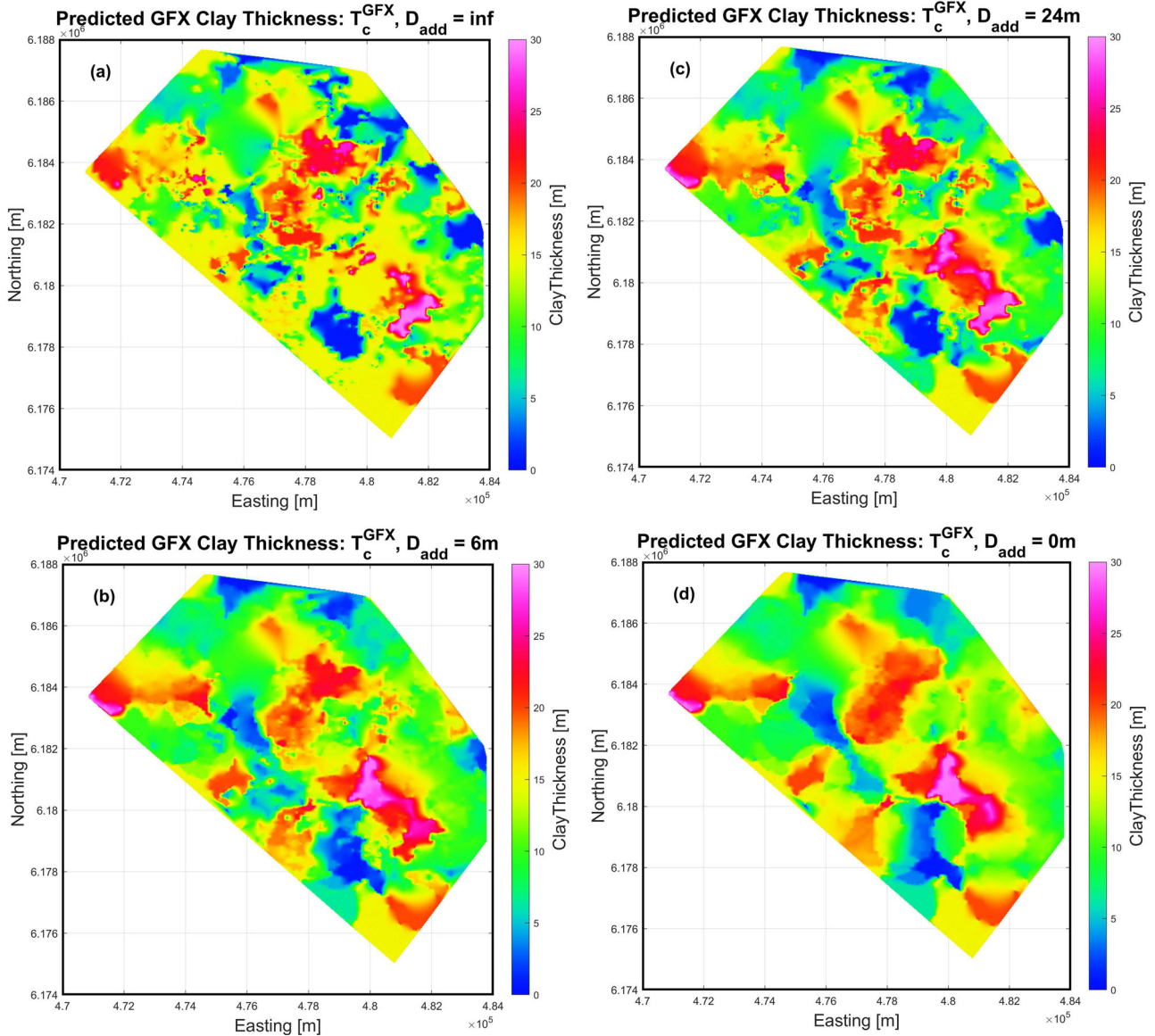


Figure 5 Maps of predicted clay thickness at the geophysics positions, T_c^{GFX} , obtained by using the binned approach and all values of the clay thickness at the borehole positions, T_c^{BH} , but for different values of D_{add} , that is, for different down-weighting of the influence of T_c^{BH} for the borehole bins not equal to the C_m^{GFX} bin. (a) A map of T_c^{GFX} for $D_{add} = \infty$, that is, when only the values of T_c^{BH} for the same bin are included. (b) A map of T_c^{GFX} for $D_{add} = 6$ m. (c) A map of T_c^{GFX} for $D_{add} = 24$ m. (d) A map of T_c^{GFX} for $D_{add} = 0$ m, that is, when all boreholes are included without penalty for differences in bins. The latter plot is the same as the one seen in Fig. 2c.

the following points, the term ‘primary data’ denotes the values of the parameter of interest measured in a limited number of locations.

- (1) Decide which of the geophysical parameters/parameter combinations would best correlate with the parameter of interest.
- (2) Calculate the geophysical parameter and interpolate it to all locations where the primary measurements of the param-

eter of interest are measured to make the subsequent correlation possible.

- (3) Make plots of the primary data and the geophysical parameter for visual inspection and to evaluate if step (1) and (2) seem reasonable, that is, if it seems likely that a correlation exists.
- (4) Bin the values of the geophysical parameter, thereby finding the bin boundaries. Decide the number of bins as the

smallest number that produces a map of the geophysical parameter that is sufficiently similar to a map of the unbinned values. In this paper, we suggest using the Continuous Wavelet Transform method because of its clear hierarchical structure.

(5) Distribute the primary data into the bins according to the interpolated value of the geophysics parameter at the measurement positions. Find the mean value and its standard deviation of the primary data within each bin. Evaluate again if the choices made so far seem reasonable.

(6) Produce maps of the parameter of interest based on both the geophysical and the primary data using the correlation that has been found in 4 and 5. We suggest that a flexible approach should be used in which the interpolated maps will take all bins into account or doing it bin by bin – or by applying a method where the bins are included with weights that decrease with the distance between bins.

(7) Based on additional geological/hydrogeological sources of information, decide which of the weighted interpolation schemes produces the most likely distribution of the parameter of interest.

DISCUSSION

Throughout this paper it has been argued that the uncertainties in finding valid proxies for the vulnerability of aquifers and the overall uncertainties of the computations and estimates involved justify a simple/robust practical approach. A basic tenet in this paper has been that clay thickness is a reasonable proxy for vulnerability, and that the mean conductivity in a certain depth range is a reasonable proxy for the clay thickness in that range. Based on these, we have developed a series of methodologies that permit an estimate of clay thickness in a certain depth range over a survey area based on a limited number of primary borehole clay thickness estimates and the mean conductivity in that same depth range estimated from inversion of airborne TEM data.

A key tool for the realization of these methodologies is the CWT analysis in which the airborne EM mean conductivity is clustered in six bins. In our case, the resulting piecewise constant function permits a better/more robust representation of the relation between mean conductivity and clay thickness than a monotonically increasing smooth function. In the field example of this paper, we found that clay thickness actually decreases for increasing conductivity at large conductivities and we attribute that to the presence of more than one type of clays, some with a higher and some with a lower resistivity.

Coming back to the uncertainties involved in the whole chain from vulnerability, to clay, to conductivity, it should be

considered if clay thickness is really such a good proxy for vulnerability – or if it would be better to go directly from conductivity to vulnerability. If the hydraulic and electrical conductivity of various clays can be established and geological background knowledge can supply knowledge of the clay lithology within an area, a direct link between vulnerability and conductivity might in fact be better. Heavy clays have a higher electrical conductivity and lower hydraulic conductivity – and are probably less fractured – than moraine clay, meaning that the direct link might be a more robust approach than using an unqualified clay thickness as an in-between link. However, this is not the place to expand on this subject. It would require a whole new line of investigations dedicated to the subject.

CONCLUSION

We have presented a procedure to interpolate/extrapolate clay thickness data estimated from borehole drilling reports over an entire survey area, supporting this process by using an attribute from a geophysical survey with dense areal coverage to produce more reliable results. The procedure involves mainly two methods: The Lateral Parameter Correlation (LPC) method and the Continuous Wavelet Transform (CWT). In this context, the LPC method is used as an advanced interpolating/extrapolating tool that takes all available data, their uncertainties and spatial positions into account. The procedure ensures that all data influence all other data and it offers an option of a user-defined degree of smoothing on the results. The CWT has been used to provide a natural binning of the geophysical parameter, the mean conductivity, into a fairly small number of representative bins, and to solve the clay thickness prediction problem using these bins one at a time or all of them in a weighted scheme.

This study was carried out to assist in estimating the vulnerability of aquifers, a crucial parameter needed for a sustainable abstraction of groundwater. The methods suggested here solve the clay thickness prediction problem with due regard to the data quality and other uncertainty factors in a way that is both simple and robust. However, the methodology is general and can be used in other situations of estimating derived products from a sparse set of direct measurements and a dense coverage with geophysical inversion products.

ACKNOWLEDGMENTS

We would like to thank Geological Survey of Denmark and Greenland for permitting us to use the data from the KOM-PLEKS project in this study. We are much obliged to Flemming

Jørgensen, previously the Geological Survey of Denmark and Greenland, presently with Central Denmark Region, for evaluating the borehole drilling reports, compiling the borehole data and estimating their uncertainty. Anne-Sophie Høyer Christensen and Nikolaj Foged conducted the data processing and the inversion of the airborne TEM data. We are indebted to Bo Holm Jacobsen for several in-depth discussion about the conceptual prerequisites of this study and the adequacy of the methodologies.

DATA AVAILABILITY STATEMENT

All the data used in this study are publicly available from the GERDA data repository administered by GEUS: The Geological Survey of Denmark and Greenland, and can be retrieved at: <https://eng.geus.dk/products-services-facilities/data-and-maps/national-geophysical-database-gerda>.

APPENDIX A

BRIEF INTRODUCTION TO THE LATERAL PARAMETER CORRELATION METHOD

The lateral parameter correlation method (LPC) was developed as a technique for lateral correlation of one-dimensional earth models after individual inversions of the data from a survey area. Several methods have been presented in the literature for this purpose: for example, Auken and Christiansen (2004). In this Appendix, we give a brief description of the Lateral Parameter Correlation (LPC) procedure of Christensen and Tølbøll (2009) with special emphasis on the use of the method as an advanced interpolation-extrapolation method.

Given a certain parameter \mathbf{p} as a function of spatial coordinates \mathbf{r} and with an uncertainty estimate, $\Delta\mathbf{p}$, the LPC is designed to provide a spatially correlated/smoothed version of \mathbf{p} , \mathbf{p}_{cor} , in such a way that both the uncertainties, $\Delta\mathbf{p}$, and the spatial correlation between the parameters are taken into account. The procedure is based on an inversion formulation, where $\Delta\mathbf{p}$ is included in the form of a data error covariance matrix and the spatial correlation is included in the form of a model covariance matrix that introduces a smoothing effect on the resulting \mathbf{p}_{cor} . The constrained inversion formulation ensures that all data will influence all other data. For a description of the use of the LPC method in laterally correlating inversion models, see Christensen and Tølbøll (2009) and Christensen (2016b). The forward mapping between \mathbf{p} and \mathbf{p}_{cor} is given by

$$\mathbf{p} = \mathbf{I}\mathbf{p}_{\text{cor}} + \mathbf{e}, \quad (\text{A1})$$

where \mathbf{I} is the identity matrix and \mathbf{e} is the observation error.

The model covariance matrix, \mathbf{C}_m , is defined using a broadband covariance function belonging to the von Karman family of covariance functions

$$\Phi_{v,L}(z) = \sigma_v^2 \frac{2^{1-v}}{\Gamma(v)} \left(\frac{|z|}{L} \right)^v K_v \left(\frac{|z|}{L} \right), \quad (\text{A2})$$

where K_v is the modified Bessel function of the second kind and order v , Γ is the gamma function, L is the maximum correlation length accounted for and is best chosen as a few times the maximum value of $|z|$ over the survey area, and σ_v is the standard deviation of the model covariance. Smoothing increases for smaller values of σ_v and vice versa. For $v \rightarrow 0$, the von Karman function effectively contains all correlation lengths due to the logarithmic singularity of K_0 . This broadband behaviour ensures superior robustness in the inversion, that is, model structure on all scales will be permitted if required by the data, and it makes the regularization imposed by the model covariance matrix insensitive to the discretization (Serban and Jacobsen, 2001; Christensen and Tølbøll, 2009). Maurer *et al.* (1998) demonstrate that the von Karman covariance functions possess both a smoothing and a damping aspect.

The least squares inversion solution to the mapping in (A1) is

$$\mathbf{p}_{\text{cor}} = (\mathbf{I}^T \mathbf{C}_e^{-1} \mathbf{I} + \mathbf{C}_m^{-1})^{-1} \mathbf{I}^T \mathbf{C}_e^{-1} \mathbf{p} = (\mathbf{C}_e^{-1} + \mathbf{C}_m^{-1})^{-1} \mathbf{C}_e^{-1} \mathbf{p}, \quad (\text{A3})$$

where \mathbf{C}_e is the data error covariance matrix of the uncorrelated parameters. It is assumed that all data errors, $\Delta\mathbf{p}$, are independent so that \mathbf{C}_e becomes a diagonal matrix with elements $1/(\Delta\mathbf{p})^2$. The posterior standard deviations of the correlated parameters are finally estimated as the square root of the diagonal elements of the posterior covariance matrix \mathbf{C}_{est} given as

$$\mathbf{C}_{\text{est}} = (\mathbf{C}_e^{-1} + \mathbf{C}_m^{-1})^{-1}. \quad (\text{A4})$$

As a consequence of the smoothing involved in the correlation process, the predicted values \mathbf{p}_{cor} are not equal to \mathbf{p} , the difference increasing for decreasing σ_v .

The LPC method as a tool for interpolation, extrapolation and prediction

In this paper, we shall use the LPC method as a tool for interpolation/extrapolation/prediction. There are basically two different situations in which we will use the LPC method: (1) As an advanced interpolator producing maps of measured data that reflect their uncertainty and their spatial distribution; and (2) as an interpolation tool in the situation where primary data are given only in a fraction of all the spatial

positions and we wish to predict the values at positions where there are no primary data.

The first situation can be used to produce, for example, a smoothed map of T_c^{BH} only involving the primary values of T_c^{BH} . Interpolating with the LPC method provides a means to take not only the primary values of T_c^{BH} into account, but also ΔT_c^{BH} and their spatial positions. The same approach can be used to plot a map of C_m^{GFX} . In both cases, all elements of \mathbf{p} and $\Delta \mathbf{p}$ contain only primary data. The degree of smoothing determined by the parameter σ_v can be chosen pragmatically or by requiring that the error normalized residual, R , between the original data and the predicted data:

$$R = \sum_{i=1}^N \frac{(p_i - p_i^{\text{cor}})^2}{(\Delta p_i)^2} \quad (\text{A5})$$

should be close to unity, thereby respecting the original data error $\Delta \mathbf{p}$. In this paper, the latter has been used as a guideline in choosing σ_v .

The second situation, where the LPC method is used to predict parameter values at locations where it has not been measured, arises primarily in two situations in this paper: (1) In the prediction of C_m^{BH} from C_m^{GFX} ; and (2) in the prediction of T_c^{GFX} from T_c^{BH} . In these cases, the LPC method is implemented by including all data positions, both borehole and geophysics positions, in the inversion problem and by defining data and data errors at all positions. At the positions where there are no primary data, data are given dummy values with infinite data uncertainties. The LPC method will take care of the rest, that is, use the primary data with their errors and the spatial correlations between all points to predict values at the positions where there are no primary data, imposing a degree of smoothness on the final results determined by σ_v .

A modified data error covariance matrix

In the paper, the last two approaches to predicting clay thickness at the geophysics positions involve a binning of the geophysics positions according to C_m^{GFX} and separate applications of the Lateral Parameter Correlation (LPC) method for each bin. In the first of the two, only T_c^{BH} data values at locations where C_m^{BH} is in the same bin are involved in the LPC procedure, and the previous section in this appendix explains how that is realized. The second of the two approaches also involves a separate application of the LPC method for each of the bins, but now all T_c^{BH} data are involved. Instead of including only T_c^{BH} values belonging to the same bin, now all borehole data are included, but with a redefined weighting that penalizes the difference between the geophysics and the bore-

hole bin value. This is done by adding an additional data error covariance matrix to the one defined by ΔT_c^{BH} proportional to the difference in bin numbers so that the total variance ascribed to T_c^{BH} is

$$\text{var}_{\text{new}} = \text{var}_{\text{old}} + \text{var}_{\text{bindif}} = (\Delta T_c^{BH})^2 + [D_{\text{add}} \cdot (i - j)]^2, \quad (\text{A6})$$

where i denotes the C_m^{GFX} bin number and j the bin number of C_m^{BH} . In the paper, values of D_{add} between 0 and ∞ have been investigated and the results plotted in Fig. 5.

APPENDIX B

BINNING WITH THE CONTINUOUS WAVELET TRANSFORM

Typically the continuous wavelet transform is used to find layer boundaries in log data and other time series (Davis and Christensen, 2013; Hill et al., 2020). It can also be used to simplify a data set reducing its complexity by finding its ‘natural’ categories (bins) (Christensen, 2018). Defining the mean value of the data lying within a bin to become the function value of the bin produces a piecewise constant approximation to the complexity of the original data set.

The CWT method

To introduce the methodology of the continuous wavelet transform (CWT), the basic principles will be outlined in this section illustrated with an example of using the CWT to find layer boundaries in an electrical log. More detailed presentations of the method can be found in the references cited above and in Mallat (1998), Cowan and Cooper (2003), and Cooper and Cowan (2009).

The CWT is defined as a spectrum of convolutions with a wavelet function (Mallat, 1998), the scale of which is varied:

$$W[f(u, s)] = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{s}} \Psi^* \left(\frac{t - u}{s} \right) dt, \quad (\text{B1})$$

where s is the scale, u the position, Ψ is the wavelet used and $*$ indicates complex conjugate.

In all of the following, the positions of boundaries are defined as the inflection points of the function, f , that is, where the second derivative of the function is zero: $f'' = 0$. Using the so-called ‘Mexican Hat’ wavelet (also called the ‘Gauss2’ wavelet) given as the second derivative of a Gaussian function:

$$\Psi^* = \frac{d^2}{dt^2} \left\{ \frac{1}{\sigma \sqrt{2\pi}} \cdot \exp \left(-\frac{t^2}{2\sigma^2} \right) \right\} = \frac{t^2 - \sigma^2}{\sigma^5 \sqrt{2\pi}} \cdot \exp \left(-\frac{t^2}{2\sigma^2} \right), \quad (\text{B2})$$

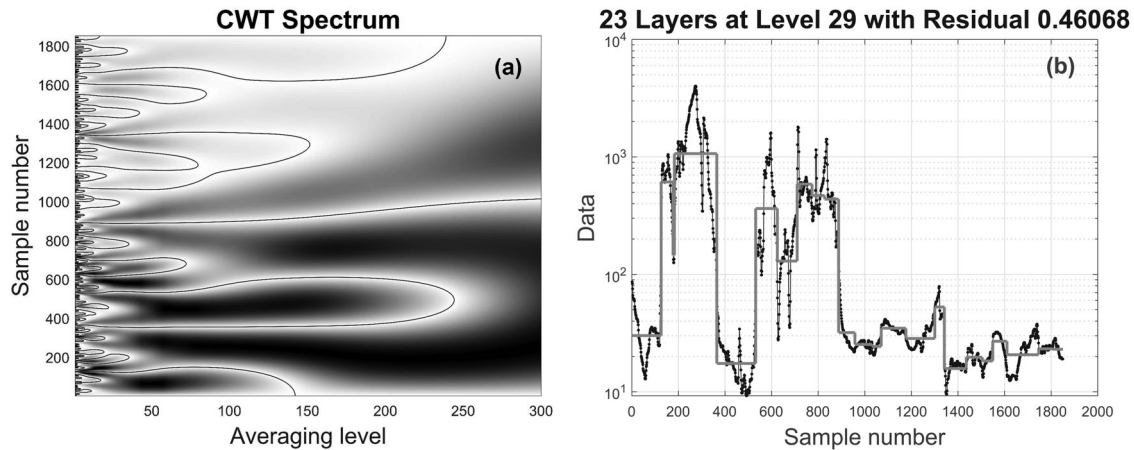


Figure B1 (a) The CWT spectrum of the electrical log. (b) A plot of the original electrical log data (black) and the piecewise constant approximation with 22 layers (grey).

where σ is the standard deviation of the distribution, now playing the role of scale length, the CWT will deliver the second derivative of f at different scales. In the presence of noise, the CWT of f at small scales will produce several zeros that are just an expression of the noise. For increasing scale length, the CWT will provide an increasingly smooth version of the second derivative, the number of zeros will decrease and only the main boundaries will survive. Eventually the user must decide the scale length or, equivalently, the number of boundaries relevant for the interpretation.

The discrete numerical implementation of CWT is done with discrete binomial filters. For increasing averaging scale, the binomial filters will become an excellent approximation to the Gauss2 wavelet. When the averaging process is completed, the zeros of the CWT spectrum are contoured, see Fig. B1. Now it is up to the interpreter to choose the appropriate averaging level, that is, choose the number of zeros and thereby the number of bins and their boundaries.

When the user has chosen the number of relevant boundaries by choosing the proper averaging level (a vertical line through the CWT spectrum), the position of the boundaries are chosen not as the actual position of the zeros at that averaging level, but by following the zero contours to the zero position at the lowest averaging level. This ensures that the position is found with the best possible precision and that the boundary thus defined does not change when/if more boundaries are added. This establishes a unique hierarchy between the boundaries which is a prominent and very desirable property of the CWT transform. Furthermore, due to the fact that boundaries on a higher averaging level do not change when more boundaries are added, it has the very valuable property

that bin interval end points do not change when more bins are considered. This means that when more complex structures are permitted by introducing more bins, the complexity is always expressed as a subdivision of already existing bins. In this respect, it differs from traditional techniques of clustering (Jain, 2010; Wu, 2012).

An example – finding layer boundaries in an electrical log

Figure B1 shows the (CWT) spectrum of the data from an electric log acquired by the Ellog logging-while-drilling method (Sørensen and Larsen, 1999), and a plot of the piecewise constant approximation to the logging data obtained by choosing 22 layers. From the CWT spectrum, it is seen that the most important layer boundaries appear well separated from each other. The choice of 22 layers is arrived at by inspecting the original log data and striking a compromise between demanding that the piecewise constant approximation should contain the important variability of the log, while at the same time avoiding fitting the noise that appears as the ‘undergrowth’ at the left side of the spectrum with very closely lying zero contours.

Binning the mean conductivity


In this paper, the CWT is used to find the natural bins of the mean conductivities of the top 30 m of the TEM inversion models and the optimal number of bins, thereby simplifying the relation between mean conductivity and clay thickness. The procedure is the following:

- 1 Collect all values of the mean conductivities from all inversion models within the survey area.

- 2 Sort the array.
- 3 Use the sorted array as input array to the CWT analysis.

The boundaries found in the CWT analysis define the end points of the natural intervals of the sorted array values, that is, the values that naturally belong together. To establish the desired correlation between mean conductivity and clay thickness, a mean value of all the primary borehole data for which the mean conductivity at the borehole position falls within the bin, is found as a noise-weighted average of the borehole clay thickness data.

ORCID

Niels B. Christensen 

<https://orcid.org/0000-0003-0770-4708>

REFERENCES

- Aguilar-López, J.P., Bogaard, T. & Gerke, H.H. (2020) Dual-permeability model improvements for representation of preferential flow in fractured clays. *Water Resources Research*, 56, 1–20.
- Auken, E. & Christiansen, A.V. (2004) Layered and laterally constrained 2D inversion of resistivity data. *Geophysics*, 69, 752–761.
- Auken, E., Pellerin, L., Christensen, N. B. & Sørensen, K. I. (2006) A survey of current trends in near-surface electrical and electromagnetic methods. *Geophysics*, 71(5), G249–G260.
- Auken, E., Christiansen, A. V., Westergaard, J. A., Kirkegaard, C., Foged, N. & Viezzoli, A. (2009) An integrated processing scheme for high-resolution airborne electromagnetic surveys, the SkyTEM system. *Exploration Geophysics*, 40, 184–192.
- Auken, E., Christiansen, A.V., Kirkegaard, C., Fiandaca, G., Schamper, C., Behroozmand, A.A. *et al.* (2014). An overview of a highly versatile forward and stable inverse algorithm for airborne, ground-based and borehole electromagnetic and electric data. *Exploration Geophysics*, 46, 223–235.
- Auken, E., Boesen, T. & Christiansen, A.V. (2017) A review of airborne electromagnetic methods with focus on geotechnical and hydrological applications from 2007 to 2017. *Advances in Geophysics*, 58, 47–93.
- Belikov, V.V., Ivanov, V.D., Kontorovich, V.K., Korytnik, S.A. & Semenov, A.Y. (1997) The non-Sibsonian interpolation: a new method of interpolation of the values of a function on an arbitrary set of points. *Computational Mathematics and Mathematical Physics*, 37(1), 9–15.
- De Carlo, L., Caputo, M.C., Masciale, R., Vurro, M. & Portoghesi, I. (2020) Monitoring the drainage efficiency of infiltration trenches in fractured and karstified limestone via time-lapse hydrogeophysical approach. *Water* 2020, 12, 1–21.
- Christensen, N.B. & Tølbøll, R.J. (2009) A lateral model parameter correlation procedure for one-dimensional inverse modelling. *Geophysical Prospecting*, 57, 919–929.
- Christensen, N.B. (2016a) Fast approximate 1D modelling and inversion of transient electromagnetic data. *Geophysical Prospecting*, 64, 1620–1631.
- Christensen, N.B. (2016b) Strictly horizontal lateral parameter correlation for 1D inverse modelling of large data sets. *Near Surface Geophysics*, 14, 391–399.
- Christensen, N.B. (2018) Interpretation attributes derived from AEM inversion models using the continuous wavelet transform. *Near Surface Geophysics*, 16, 665–678.
- Christensen, N.B. (2021) Resolution attributes for AEM inversion models. *A Conceptual Analysis and Novel Measures*. Submitted.
- Christiansen, A.V., Foged, N. & Auken, E. (2014) A concept for calculating accumulated clay thickness from borehole lithological logs and resistivity models for nitrate vulnerability assessment. *Journal of Applied Geophysics*, 108, 69–77.
- Commer, M., Pride, S.R., Vasco, D.W., Finsterle, S. & Kowalsky, M.B. (2020) Imaging of a fluid injection process using geophysical data—A didactic example. *Geophysics* 2020, 85, 1–16.
- Cooper, G.R.J. & Cowan, D. (2009) Blocking geophysical borehole log data using the continuous wavelet transform. *Exploration Geophysics*, 40(2), 233–236.
- Cowan, D.R. & Cooper, G.R.J. (2003) Wavelet analysis of detailed drillhole magnetic susceptibility data, Brockman Iron Formation, Hamersley Basin, Western Australia. *Exploration Geophysics*, 34, 87–92.
- Davis, A.C. & Christensen, N.B. (2013) Derivative analysis for layer selection of geophysical borehole logs. *Computers & Geosciences*, 60, 34–40.
- Gonzales Amaya, A., Ortiz, J., Durán, A. & Villazon, M. (2019) Hydrogeophysical methods and hydrogeological models: basis for groundwater sustainable management in Valle Alto (Bolivia). *Sustainable Water Resources Management*, 5, 1179–1188.
- Graeves, R.J., Lesmes, D.P., Lee, J.M. & Toksöz, M.N. (1996) Velocity variations and water content estimated from multi-offset ground penetrating radar. *Geophysics*, 61, 683–695.
- Gunnink, J.L. & Siemon, B. (2009) Combining airborne electromagnetic and drillings to construct a stochastic 3D lithological model. *EAGE 15th European Meeting of Environmental and Engineering Geophysics*, Dublin, Ireland.
- Han, D.H., Nur, A. & Morgan, D. (1986) Effects of porosity and clay content on wave velocities in sandstones. *Geophysics*, 51, 2093–2107.
- He, X., Sonnenborg, T.O., Jørgensen, F., Høyer, A.-S., Møller, R.R. & Jensen, K.H. (2013) Analyzing the effects of geological and parameter uncertainty on prediction of groundwater head and travel time. *Hydrology and Earth System Sciences*, 17, 3245–3260.
- Hill, E.J. & Uvarova, Y. (2018) Identifying the nature of lithogeochemical boundaries in drill holes. *Journal of Geochemical Exploration*, 184, 167–178.
- Hill, E.J., Pearce, M.A. & Stromberg, J.M. (2020) Improving automated geological logging of drill holes by incorporating multiscale spatial methods. *Mathematical Geosciences*, 53, 21–53.
- Hinnell, A.C., Ferré, T.P.A., Vrugt, J.A., Huisman, J.A., Moysey, S., Rings, J. *et al.* (2010) Improved extraction of hydrologic information from geophysical data through coupled hydrogeophysical inversion. *Water Resources Research*, 46.

- Hubbard, S.S. & Rubin, Y. (2000) Hydrogeological parameter estimation using geophysical data. A review of selected techniques. *Journal of Contaminant Hydrology*, 45, 3–34.
- Høyer, A.-S., Lykke-Andersen, H., Jørgensen, F. & Auken, E. (2011) Combined interpretation of SkyTEM and high-resolution seismic data. *Physics and Chemistry of the Earth*, 36, 1386–1397.
- Høyer, A.-S., Møller, I. & Jørgensen, F. (2013a) Challenges in geophysical mapping of glaciotectionic structures. *Geophysics*, 78, B287–B303.
- Høyer, A.-S., Jørgensen, F., Piotrowski, J.A. & Jakobsen, P.R. (2013b) Deeply rooted glaciotectionism in western Denmark: geological composition, structural characteristics and the origin of Varde hill-land. *Journal of Quaternary Science*, 28, 683–696.
- Høyer, A.-S., Jørgensen, F., Lykke-Andersen, H. & Christiansen, A.V. (2014) Iterative modelling of AEM data based on geological a priori information from seismic and borehole data. *Near Surface Geophysics*, 12, 635–650.
- Jain, A.K. (2010) Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31, 651–666.
- Jørgensen, F., Scheer, W., Thomsen, S., Sonnenborg, T.O., Hinsby, K., Wiederhold, H. *et al.* (2012) 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.
- Jørgensen, F., Møller, R.R., Nebel, L., Jensen, N., Christiansen, A.V. & Sandersen, P. (2013) A method for cognitive 3D geological voxel modelling of AEM data. *Bulletin of Engineering Geology and the Environment*, 72, 421–432.
- Kirsch, R., Sengpiel, K.P. & Voss, W. (2003) The use of electrical conductivity mapping in the definition of an aquifer vulnerability index. *Near Surface Geophysics*, 1, 13–20.
- Klimentos, T. (1991) The effects of porosity-permeability-clay content on the velocity of compressional waves. *Geophysics*, 56, 1930–1939.
- Ledoux, H., Gold, C.M. & Fisher, P. (2005) An efficient natural neighbour interpolation algorithm for geoscientific modelling. In: Fisher, Peter F. (Ed.) *Developments in spatial data handling*. Berlin: Springer, pp. 97–108.
- Ley-Cooper, A.Y., Macnae, J.C. & Tweed, S. (2008) Estimating subsurface porosity and salt loads using airborne geophysical data. *Near Surface Geophysics*, 6, 31–37.
- Mallat, S. (1998) *A wavelet tour of signal processing*. Cambridge, MA: Academic Press.
- Maurer, H., Holliger, K. & Boerner, D.E. (1998) Stochastic regularization: smoothness or similarity? *Geophysical Research Letters*, 25(15), 2889–2892.
- Møller, I., Søndergaard, V.H. & Jørgensen, F. (2009a) Geophysical methods and data administration in Danish groundwater mapping. *Geological Survey of Denmark and Greenland Bulletin*, 17, 41–44.
- Møller, I., Søndergaard, V.H., Jørgensen, F., Auken, E. & Christiansen, A.V. (2009b) Integrated management and utilization of hydrogeophysical data on a national scale. *Near Surface Geophysics*, 7, 647–659.
- Paasche, H., Tronicke, J., Holliger, K., Green, A.G. & Maurer, H. (2006) Integration of diverse physical-property models: subsurface zonation and petrophysical parameter estimation based on fuzzy c-means cluster analyses. *Geophysics*, 71, H33–H44.
- Reynolds, D.A. & Kueper, B.H. (2001) Multiphase flow and transport in fractured clay/sand sequences. *Journal of Contaminant Hydrology*, 51, 41–62.
- Rhoades, J.D., Manteghi, N.A., Shouse, P.J. & Alves, W.J. (1989) Soil electrical conductivity and soil salinity: new formulations and calibrations. *Soil Sciences Society of America Journal*, 53, 433–439.
- Serban, D.Z. & Jacobsen, B.H. (2001) The use of broadband prior covariance for inverse palaeoclimate estimation. *Geophysical Journal International*, 147, 29–40.
- Shevnev, V., Mousatov, A., Ryjov, A. & Delgado-Rodriguez, O. (2007) Estimation of clay content in soil based on resistivity modelling and laboratory measurements. *Geophysical Prospecting*, 55, 265–275.
- Sibson, R. (1981) A brief description of natural neighbor interpolation. In: Barnett, V. (Ed.) *Interpreting multivariate data*. Chichester: John Wiley, pp. 21–36.
- Sørensen, K.I. & Larsen, F. (1999) Ellog Auger drilling: 3-in-one method for hydrogeological data collection. *Ground Water Monitoring & Remediation*, 19(4), 97–101.
- Sørensen, K.I. & Auken, E. (2004) SkyTEM - a new high-resolution helicopter transient electromagnetic system. *Exploration Geophysics*, 35, 191–199.
- Thomsen, R., Søndergaard, V.H. & Sørensen, K.I. (2004) Hydrogeological mapping as a basis for establishing site-specific groundwater protection zones in Denmark. *Hydrogeology Journal*, 12, 550–562.
- Wu, J. (2012) *Advances in K-means clustering - a data mining thinking*. Berlin–Heidelberg: Springer.