HydroGeophysics Group
AARHUS UNIVERSITY

Hydrostructural Modelling
Bolbro Bæk - LOOP 6

MapField project, April 2021

TABLE OF CONTENTS

1. INTRODUCTION

Hydrostructural modelling, Bolbro Bæk - LOOP 6	
Project	Mapfield, Project number. 29771
Key persons	**HGG, Geosciences, Aarhus University, Denmark** Project leader: Associated prof., PhD, *Anders V. Christiansen* Modelling: Postdoc, PhD, *Niels Claes* Postdoc, PhD, *Rasmus Rumph Frederiksen* Senior geophysicist, PhD, *Nikolaj Foged*
Area	Bolbro Bæk, LOOP 6, ~752 hectares

This report primarily explains and documents the work for compiling realizations of 3D hydrostratigraphic models (HS-models) for the Bolbro Bæk area (LOOP 6).

In the MapField project thorough geophysical mapping with the tTEM-method has been carried out in six LOOP areas, which are catchments that are intensely monitored for nutrient transport and land use by the Danish Environmental Protection Agency (Miljøstyrelsen). A part of the MapField project is the development of detailed hydrological models that can be used to simulate nitrate loads to recipients on catchment scale.

Detailed 3D structural information of the subsurface is fundamental for development of both the hydrological and the geochemical models. The structural information is primarily derived from the geophysical mapping results. However, some parts of the LOOP areas are inaccessible to geophysical mapping or might suffer from low data quality, which results in information gaps. Geostatistical methods are used to remediate these data gaps and to incorporate uncertainty in the construction of the hydrostratigraphic model. The output of the hydrostratigraphic modelling concept (HSM) is an ensemble of plausible 3D hydrostratigraphic models.

In addition to geophysical data, geochemical data has been collected in these LOOP areas. The geochemical data is used together with the geophysical data to generate an ensemble of 3D redox-models that are complementary to the plausible 3D HS-models.

Chapter 2 of this report provides an overview of the data types used in the HS-modelling. Chapter 3 provides a description of the procedures and methods of the HS-modelling concept, while a technical specification for the HS-modelling setup is stated in chapter 4. Finally, chapter 5 provides a few modeling result examples and documents the formats of the deliverables.

2. AREA AND DATA TYPES

The Bolbro Bæk catchment has an area of approximately 752 ha and is located South Jutland (Denmark) between the towns of Bedsted and Hellevad (see Figure 1).
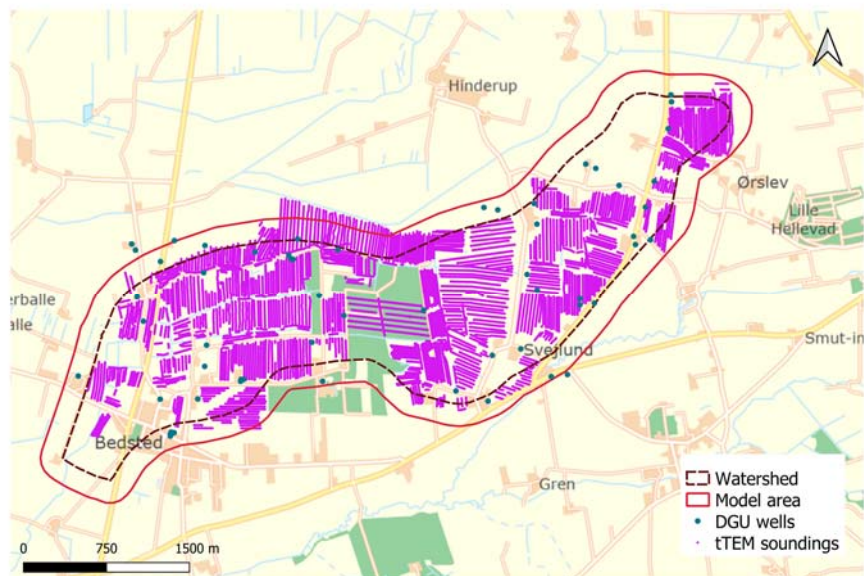

Figure 1. Catchment area. For hydrological computational reasons the HS- models are constructed with an additional 200 m buffer of the catchment area.

tTEM-data
Approximately 640 ha of the catchment area were mapped with the geophysical method tTEM in March-April 2019. The results of this mapping campaign produced the input resistivity dataset used in this project. The geophysical part of tTEM survey is reported in the *MapField - tTEM Mapping LOOP 6 Bolbro* report.

Borehole data
The borehole data used in the HS-modelling are extracted from the Danish national Jupiter borehole database, June 2020. As reported, the elevation of the borehole data is used in the Jupiter borehole database.

Digital elevation model (DEM)
The DEM applied to the resistive data and used for the HS-models is a 25m grid obtained by averaging a 0.4 m DEM grid from the Danish Agency for Data Supply and Efficiency.

3. METHODOLOGY

This chapter holds a short overall description of the HS-modelling concept with a more detailed explanation of the different components of the HS-concept in the sub-sections. The HS-modelling setup parameters for the specific area are stated in chapter 4 - *Technical Specifications*.

The overall purpose of the modelling concept is to generate an ensemble of hydrostratigraphic model realizations that honor the input data in an objective, fast/automatic, and well documented manner. This makes it possible to quantify the structural uncertainty and the uncertainty of the different model predictions in the later hydrological modelling.

Figure 2 shows a flowchart of the HS-modelling concept (HSM). The main source of structural information comes from the resistivity model output of the tTEM survey. Combining the resistivity dataset with lithological borehole data in the ACT procedure, the resistivity values are translated into clay content (CF-model), so for each resistivity layer at each resistivity model location we now have a data pair of resistivity and CF-values.
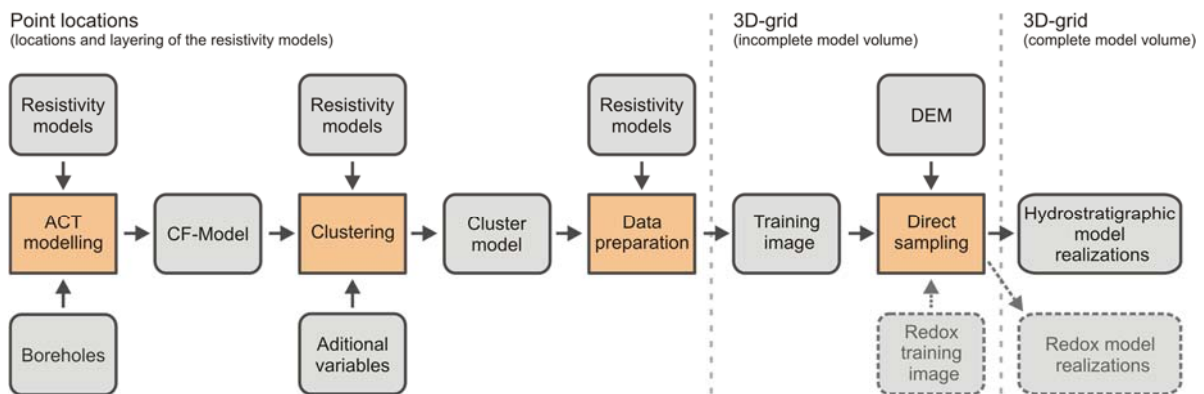

Figure 2. Flowchart of the full HS-modelling concept.

The CF–resistivity data pairs are then converted to discrete hydrostratigraphic units (HS-units) through a clustering routine. The clustering is performed because the groundwater model that is incorporating the HSM realizations of the subsurface uses HS-units to define zones of similar hydrogeological behavior.

4

To go from the HS-unit dataset, which is only present at the resistivity model positions, to the full model volume and generate multiple plausible model realizations, the direct sampling method is applied.

Direct sampling uses patterns that are recognized in the areas with high data coverage to fill out the areas with sparse or no data. The training image for the direct sampling is obtained by 3D-krigning of the resistivity data and HS-units. This results in an incomplete 3D-grid. Direct sampling uses this data as a training image to generate complete 3D HS-model grids. At the same time, direct samplings allow for incorporation of uncertainty through separation of hard and soft data.

Since the redox conditions in the subsurface are related to the hydrogeological conditions, we are using the direct sampling method to co-simulate both HS-units and redox conditions by merging the HS-units training dataset with a redox condition dataset provided by the project partners at GEUS. The resulting set of 3D hydrogeologic structural models and redox condition models retains the complex geostatistical spatial relationships that can exist between the different types of datasets within the training image, making them suitable for nitrate retention modeling at catchment scale.

3.1 ACT-modelling

The geophysical output from a tTEM survey is structural information in the resistive domain. Lithological borehole logs are often used to establish a connection between resistivity and lithology and to derive geological/hydrostratigraphic units from the resistivity model. The ACT (Accumulated Clay Thickness) concept automates this process by determining the optimum cutoff resistive values of a resistivity-to-sand/clay (hydrological permeable/impermeable) translator function, based on information from lithological logs.

The input to the concept is the accumulated meters of the clayey lithologies in depth intervals, as shown in Figure 3a, where the clay content of an interval is stated as fraction of the interval length (CF_{Log}). With the translator function in Figure 3 b, a resistivity model can also be converted into clay content in the same intervals (CF_{Res} in Figure 3a).

The translator function is described by two parameters – an Upper and a Lower resistivity value (m_{up}, m_{low}). With the translator function in Figure 3b, resistivity layers below 40 Ωm will get a weight of ~1 which means that the full length of the resistivity layer is presumed to be clay, while resistivity layers above 70 Ωm will get a weight of ~0 corresponding to no-clay content (sand) for this resistivity layer. Hence, resistivity values between 40-70 m, in this example, will result in CF_{Res} values between 0-1.
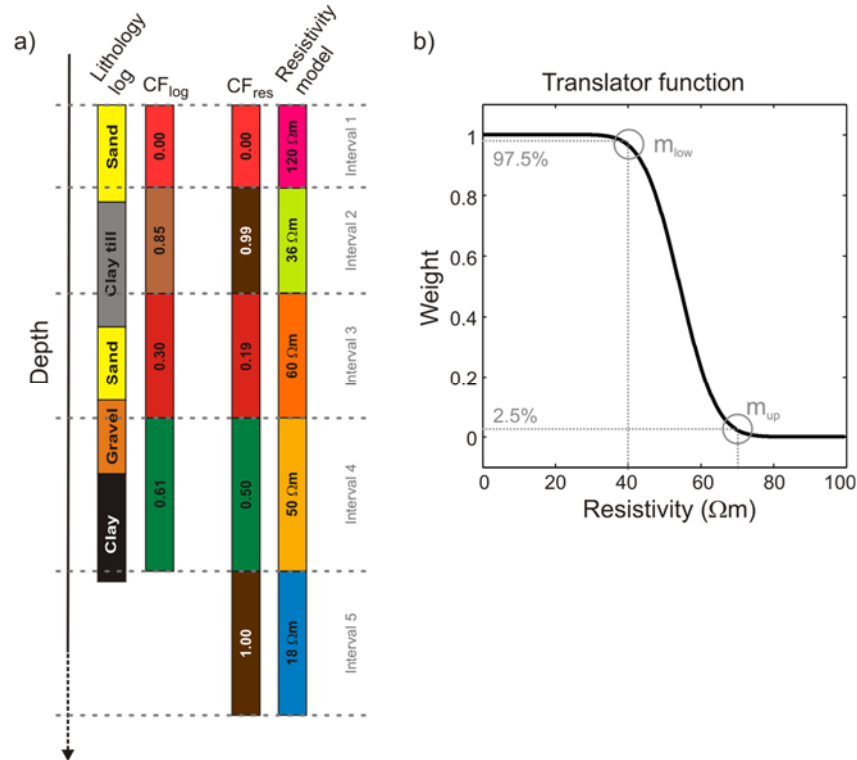
Figure 3. a) Lithological log with the corresponding CF-values, and the resistivity model and the CF values obtained by using the translator function in b). In this case, the depth intervals follow the layer discretization of the resistivity model.

Through inversion, we determine the optimum parameters of the translator functions by minimizing the difference between the clay contents calculated from the resistivity models (CF_{res}) and the observed clay contents in the lithological well logs (CF_{log}).

A key aspect in this concept is that the translator function can change horizontally and vertically, adapting to the local conditions and borehole data. Therefore, not only one translator function is used for an entire survey, but a number of translator functions spatially distributed in a regular 3D grid, as shown in Figure 4. The vertical discretization of the translator function grid follows the calculation intervals from Figure 3b, while the horizontal discretization is set to allow for the needed flexibility in the translation.
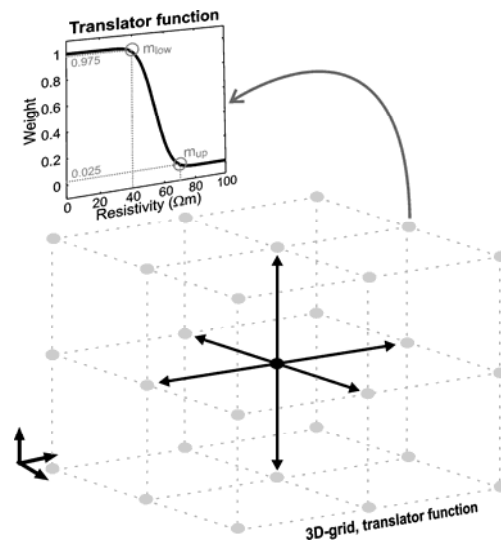
Figure 4. Translator model grid. Each node in the grid holds a translator function. The m_{up}, m_{low} parameters of the translator function are constrained to the neighboring nodes laterally and vertically, as indicated with the arrows in the plot.

To migrate information of the translator function from regions with many boreholes to regions with few or no boreholes, horizontal and vertical smoothness constraints are applied between the parameters of the translator functions as shown in Figure 4.

After we have obtained the optimum translator function grid, it is used for a final translation of all the resistivity models to CF-values, hence the CF-model incorporating the key information from both the resistivity models and the lithological logs into one parameter.

This concept also takes the uncertainties on the input resistivity models and the borehole data into account in the determination of the optimum translation. The uncertainties for the resistivity model are derived from the model parameter analysis performed during the inversion of the geophysical data. A different quality/uncertainty is assigned to the borehole data using the borehole quality index available in the Aarhus Workbench software, which is partly developed by GEUS. The borehole quality index primarily takes the following parameters into account: Drilling method and purpose, vertical sample density, the accuracy in the geographical position of borehole. The distances between boreholes and the resistivity models are also taken into account in the weighting of the borehole data. In some few cases boreholes are excluded or the uncertainty increased due to a clear mismatch to close neighboring borehole(s) and resistivity models.

The lithological borehole information is normally extracted from the PC-Jupiter borehole database and the extraction of the "clay" unit is done regardless to the type of clay; clay till, mica clay, Paleogene clay, etc.

Due to the spatially varying resistivity to clay translation, the ACT-concept can handle changes in the pore water resistivity, but only within the basis assumption of the concept that clay units are more electrically conductive than the non-clay units. This basic assumption might be challenge in case of saline pore water.

In a hydrological context, the CF-values close to 1 forms the highly impermeable units (~clay) and CF-vales close to 0 the permeable units (~sand/gravel).

A detailed description of the ACT-concept is given in Christiansen et al., (2014) and Foged, (2014).

Borehole data are often relatively sparse, especially for the deeper part in a tTEM survey area, and furthermore is it primarily the resistivity values close to and between the m_{up} and m_{low} parameter of the translator function that drive the determination of the translator function. Therefore, the constraint setup and start values of the translator function become relatively important, and often more ACT inversions are performed and evaluated to obtain a working translator function/CF-model.

3.2 Clustering of CF-data and resistivity

Clustering of CF-values and resistivity is performed to transform the datasets to discrete hydrostratigraphic units (HS-units) that are afterwards used in a groundwater model.

For dimensionality reduction of the multivariable cloud obtained by the CF-data combined with the resistivity-data, and to obtain a categorical dataset, we used K-means clustering (Wu, 2012). More variables, such as depth or the translator function from the ACT scheme, can also be included in the clustering. The clustering is not directly performed on the variables, but on their principal components (PCA) after normalization. The PCA is performed to obtain independent variables, which is a condition for the clustering routine. Details about the used clustering scheme can be found in Vilhelmsen et al., (2019).

3.3 Direct sampling

Direct sampling is applied to fill out the full model volume and generate multiple plausible HS-model realizations based on the HS-units present at the resistivity model locations. A 2D example of this is shown in Figure 5, where Figure 5a shows the volume with its data gaps. Plausible full 2D

profiles are generated through direct sampling, as demonstrated in Figure 5b and c.
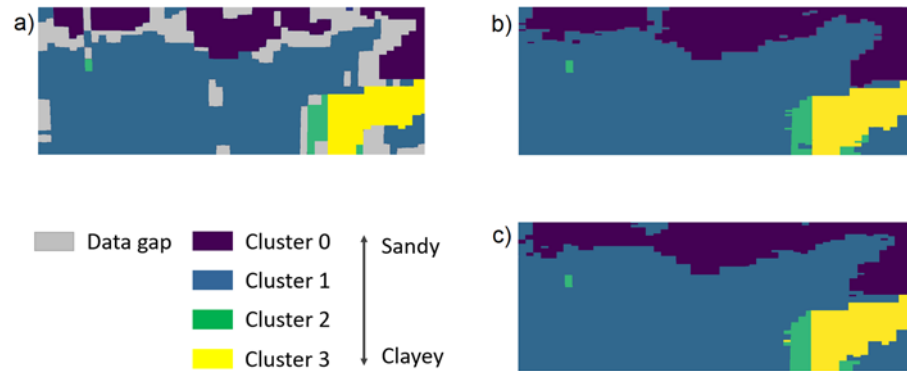

Figure 5. Example of the use of direct sampling to fill out full model volumes and generate multiple realizations. a) 2D profile with data gaps that are remediated by Direct Sampling; b-c) direct sampling generated profiles from the profile in (a).

Direct sampling is a geostatistical method that falls under multi-points statistics (MPS). The higher order of statistics incorporated in MPS allows them to produce more complicated geostatistical realizations based on Training Images (TI). The direct sampling realizations are generated based on conditional geostatistical information that is obtained by scanning the TI. (Mariethoz et al., 2010)

The direct sampling method uses both hard data and soft data to guide the simulations. The hard data points in the simulation grid are the points that are set to a specific value and are constant in between the simulations. They are the anchor points of the simulation and are based either on measurements or on other data points that have a high probability. The soft data in the simulations are included in a training image to guide the simulation or supplementary probability distributions if these are available. Generally, hundreds or more simulations are generated. For each simulation, the grid cells in between the hard data are filled out by following a different simulation path that searches the training data for a similar pattern as can be found in the simulation grid around the cell that is simulated. This results in an ensemble of model realizations that reflect the uncertainty included in the data used for the direct sampling method and has the same geostatistical characteristics as the training image.

The training image and hard data are derived from the clustering analysis and the resistivity dataset. These clustering and resistivity datasets are points distributed in space with categorical variables (HS-units) and resis-

tivity values and are geographically located at the geophysical model positions. The HS-units and resistivity data are projected to the 3D-grid using indicator kriging for the discrete clustering of groups of data and ordinary kriging for log-transformed resistivity data (Goovaerts, 1997). The goal of this kriging procedure is not to fill out too much of the modeling grid, but rather to transfer the data points to a regular grid, which is used in the direct sampling procedure. The search radius of the kriging procedure is therefore limited to a short length. Because of larger correlation lengths in the horizontal plan than in the vertical plan, the search radius in the horizontal plan is set to be larger than in the vertical plan, resulting in a search-ellipsoid. The top of the 3D grid is delineated by the surface elevation data.

The kriging procedure results in a 3D grid with data gaps in the areas where no geophysical data is collected. This incomplete 3D grid becomes the training image for the direct sampling simulations. The hard data for the simulations contains only the data points with a relatively high probability. This dataset is derived from the training image (which is soft data) by using a threshold in the conditional probability distribution (Figure 6). Figure 6 shows an example of the conditional probability distribution for four clusters groups. In this case, the 20 % most certain HS unit-resistivity pairs (data above the red line) for each HS-unit are used as hard data. They will be used as anchor points of the simulation, while the direct sampling method is used to fill out the holes in between these points by scanning the training image for both the HS-unit and the resistivity data.
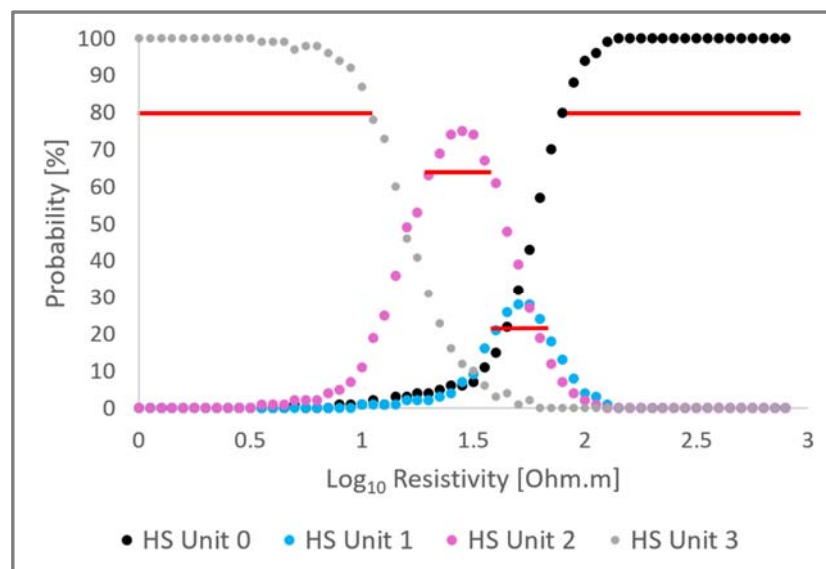

Figure 6. Example of conditional probability plot for resistivity-HS-units with an 80 percent threshold (red line).

The HS-unit resistivity training image is passed to hydrogeochemistry experts (project partners at GEUS). They develop a 3D training image for the redox conditions in the subsurface based on our training image and hydrogeochemical observations from the catchment area. More information about the construction of this training image can be found in (Kim et al., 2021). Together with this redox TI, they estimate a probability for the estimated redox condition for every grid cell. This is used to filter between hard data and soft data. Only the grid cells with a probability above a specific threshold are kept as hard data (anchor points for the simulation).

Combining the three training data sets allows for simultaneous simulation of the three variables in each grid point as a vector, instead of simulating them as separate variables. This approach allows for retaining the complex geostatistical spatial relationships that can exists between the different datasets in the resulting generated 3D models. (Mariethoz et al., 2010)

The resulting ensemble of simulated complete 3D-models contains per realization: A HS-model, a redox condition model, and a resistivity model.

4. TECHNICAL SPECIFICATIONS

This chapter contains the parameter types, specific parameter settings, etc. for the Bolbro Bæk area.

4.1 ACT setup

The vertical discretization of the ACT translator function grid follows the initial layer discretization of the resistivity models, hence additional vertical smoothing is avoided. The sharp resistivity model inversion result was used rather than the smooth resistivity model, again to optimize the sharpness in the final structures model. The resistivity models were masked below the estimated depth of investigation. Table 1 provides detailed ACT setup information.

The locations of boreholes with lithological information for the survey area are plotted in Figure 7 , with colors indicating the drill depth, together with the tTEM survey lines (in green) and the translator function grid (orange squares). As seen in the figure the borehole information is relatively sparse in some parts of the tTEM survey, especially for deeper part of the model.
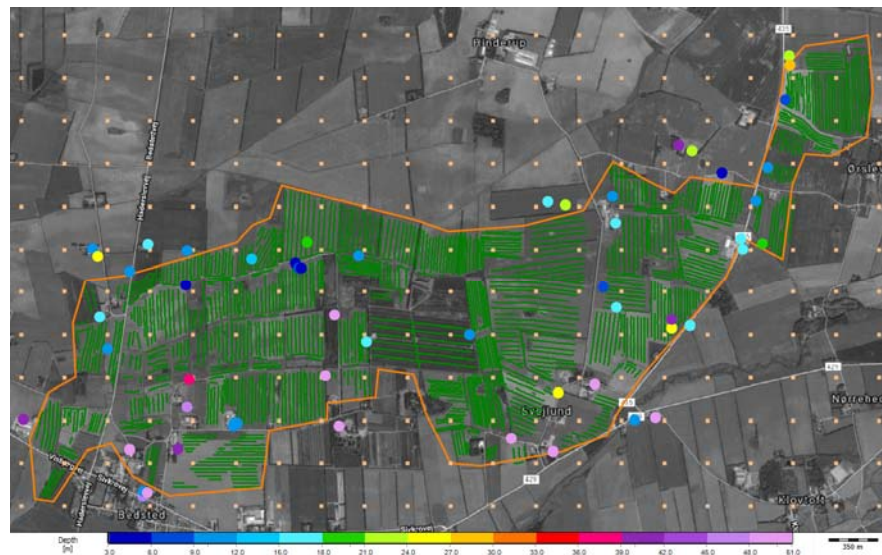

Figure 7. Bolbro Bæk survey area. Green dots mark the location of the tTEM resistivity models. Colored dots are the borehole locations with the color indicating the drill depth (see color scale), and orange squares mark the translator function grid in the horizontal plan.

ACT setup

Input	Type	
Resistivity model input	tTEM, Sharp SCI inversion result	
Resistivity model blinding	DOI standard	
Borehole input	DK - PC Jupiter	
Unknown borehole lithology	b, x , o, u, v, m	
Borehole rating scheme	Rating 1-4, Aarhus Workbench	
Uncertainty, rating, % of interval	R1=10%, R2=15%, R3=30%, R4=50%	
Translator model type	Compl. error function	
Search radius, borehole to res. model	300 m	
Translator function grid	**Vertical**	**Horizontal**
Discretization	Layer thicknesses, resistivity model	300 m
Model constraints	2.0	1.0
Model layers, max depth	24 layers down to 74.9 m	
Translator function	**Upper**	**Lower**
Start values	40 Ωm	60 Ωm
Prior	No prior	No prior
Inversion output		
Data residual	0.59	
Number of iterations	9	

Table 1. Detailed information for the ACT setup.

4.2 Clustering setup

Before the clustering is performed, the datasets are prepared in following order:

1. If more resistivity layers enter an ACT layer interval, the mean resistivity value of the interval is used. The mean resistivity value is calculated as the inverse of the mean conductivity.
2. The resistivity values (or mean resistivity values) are truncated to avoid that outliers have large impact on the later data scaling.
3. The resistivity data is log-transformed.
4. All datasets are min/max scaled to values between [0-1] as implemented by the *sklearn.preprocessing.MinMaxScaler* Python library, except the CF-values from the CF-model which are already in the [0-1] range.
5. Optional: Weights can be added to the different data types.
6. The Principal components (PCA) of the data types are calculated.

K-means clustering as implemented in the *sklearn.clustering* Python library with k++ means initialization. (Pedregosa et al., 2011)

Setting for the different steps are listed in Table 2.

Data preparation		
Data types	Resistivity	CF-values
Truncation [min, max]	[3, 3000]	---
Log Transform	Yes	No
Min/max scaling	[0,1]	As if [0,1]
Data weights	1	1
Clustering		
Clustering type	K-means	
Number of clusters	4	

Table 2. Data preparation and clustering setup.

4.3 Direct sampling setup

Training images

The kriging process to compile training images is performed in SGeMS software with variogram parameters that are estimated based on the resulting data cloud from the clustering procedure. Parameters for the kriging process are listed in Table 3.

	Resistivity data	Cluster data
Data transform	Log	non
Interpolation	Ordinary kriging	Indicator kriging
Vertical resampling	5 points per ACT layer	
Conditioning data points, [min, max]	[0, 3]	
Search ellipsoid, [min, mid, max]	[30, 30, 3] m	

Table 3. Kriging parameters for training image.

Direct Sampling parameters

Setting for direct sampling are included in Table 4.

Direct Sampling parameters	
Model grid vertical discretization	2 m
Model grid horizontal discretization	25 m
Training image	Log-Res, Clusters, Redox
Hard data	From Training image
Hard data Log-Res, Cluster conditional probability threshold	70 %
Hard data redox, probability threshold	85 %
Maximum numbers of neighboring nodes	20
Distance threshold for each variable	0.05
Max scan fraction	60 %
Output	
Number of realizations	500
All grids contains simulations of	Log-Res, cluster, redox zone

Table 4. Direct Sampling parameters

5. MODELLING EXAMPLES AND DELIVERABLES

This chapter shows a few modelling examples and provide an overview of the deliverables and a detailed format description of the deliverables.

5.1 Modelling examples

This section provides some visual examples of the modelling results. Figure 8 shows an example of the sliced model in the horizontal plane. While Figure 9 shows an example of the sliced model in the vertical plane.
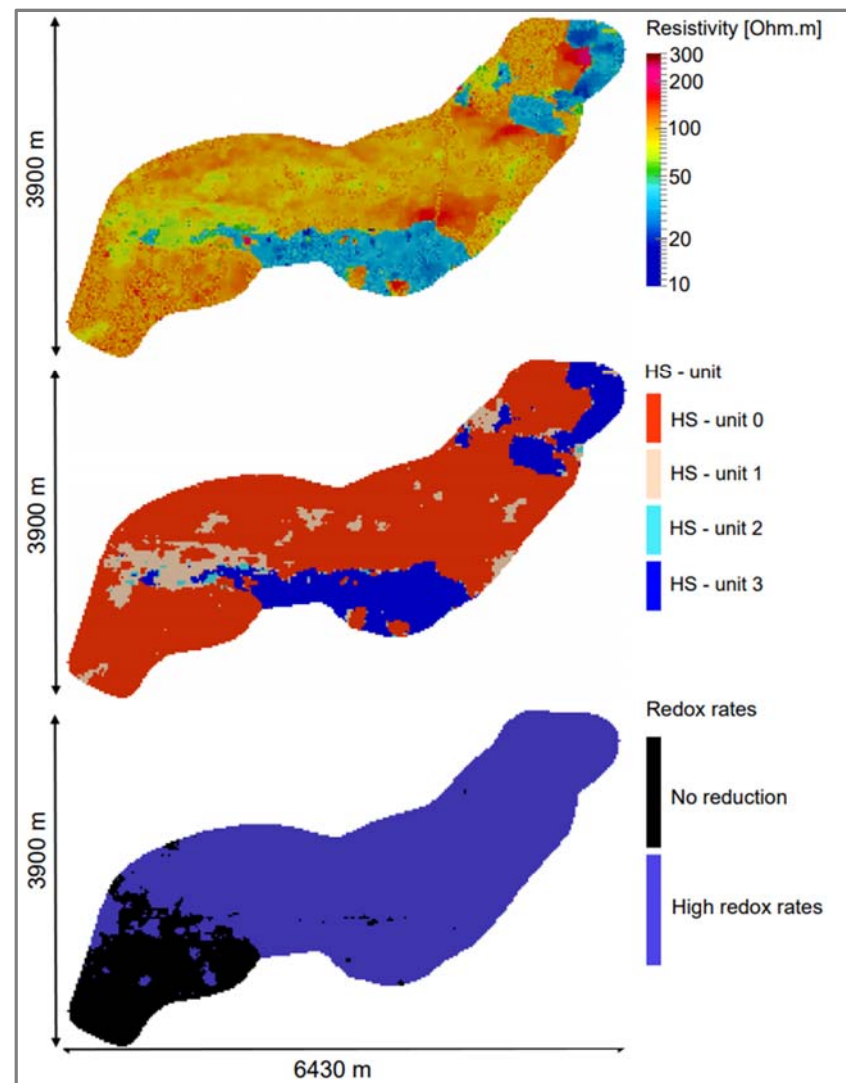

Figure 8 Profiles for resistivity, HS-units and redox state in the horizontal plane for z = 22m above sea-level.

HS – unit 3 is the unit with the most clayey material, while lower HS – units indicate units with lower clay content. Zones where no reduction occurs are only located near the surface of the model volume.
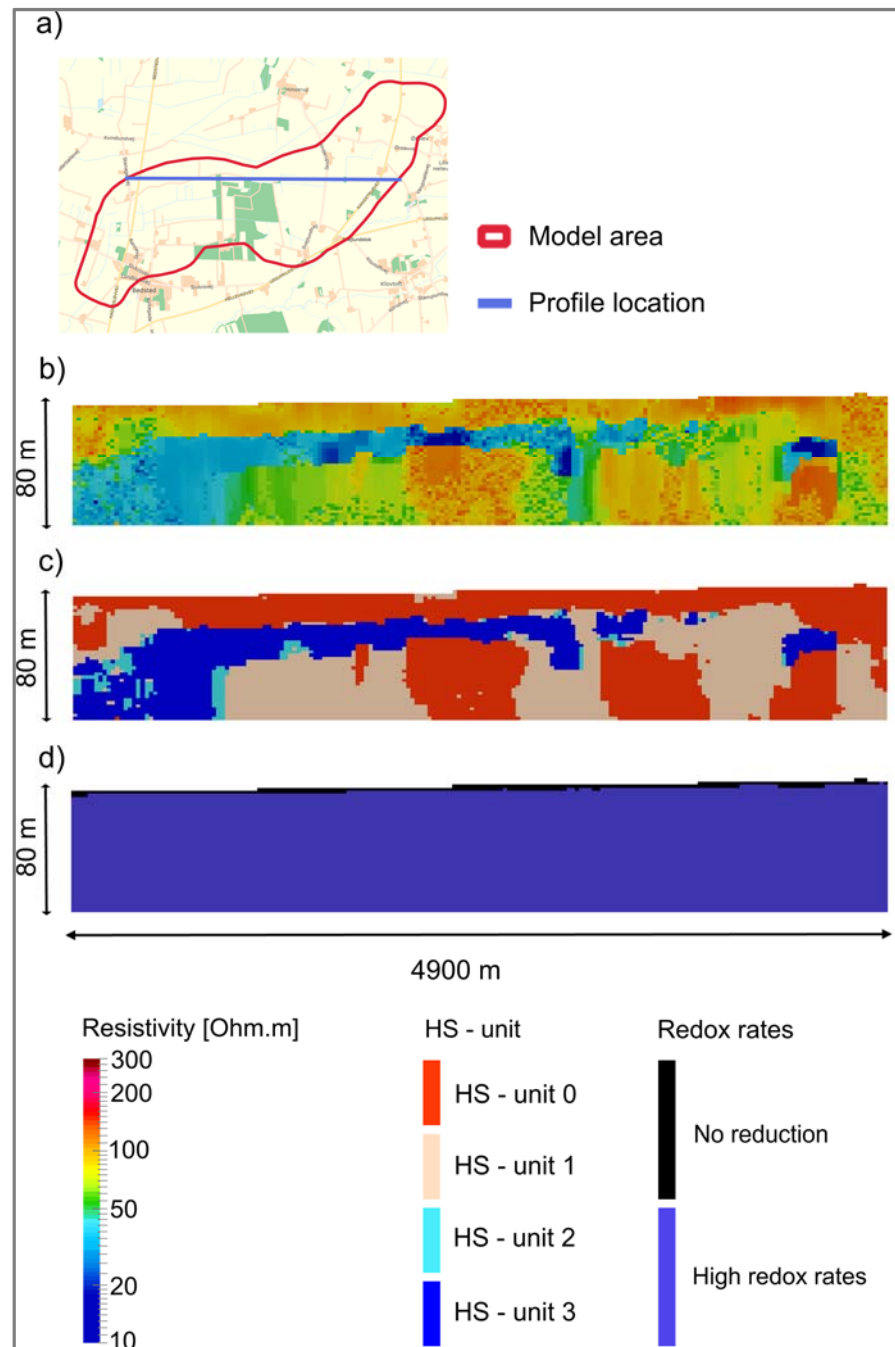

Figure 9. a) Location of the profile; Vertical slices of the profile for Y-UTM = 6103200. a) Resistivity; b) HS - units; c) Redox conditions.

16

5.2 Deliverables

This section provides a format description of the different deliverables. Note that all stated UTM coordinates are geo-referenced to the coordinate system: WGS84, UTM zone 32N (EPSG 32632).

The deliverables are:

ACT:

- CF-model   …Output\ACT\CFModel\

Direct sampling:

- The training images.   … Output \DS\TrainingImage\
- Hard data   … Output \DS\HardData\
- Model realizations   … Output \ModelRealizations\
- Redox-model realizations … Output \RedoxModel\

CF-model (*.gct-file)
The gct-files hold information about the clay contents (CF$_{Res}$ values) for the translation of the resistive models with the obtained optimum translator function, hence all the information in the gct-file is at the resistivity model positions for the different CF-intervals (layers). The format of the gct-file is documented in Table 5.

Colum header	Description
Interval	Index of calculation interval (layer number)
ModIndex	Model index. Same model index equals same model position
UTMX	Model position, UTM-X coordinate, (m)
UTMY	Model position, UTM-X coordinate, (m)
Topo	Surface elevation at model position, (m)
IntvBegin	Depth to top of layer interval, (m)
IntvEnd	Depth to bottom of layer interval, (m)
GeophCT	Clay thickness in interval for the translated resistivity value(s), (m)
GeophCTstd	Estimated uncertainty of GeophCT, absolute STD, (m)
GeophCF	Clay fraction in interval (GeophCT/interval length)
GeophCFstd	Estimated uncertainty of GeophCF, absolute STD, (fraction of interval)
MeanRes	Resistivity or mean resistive of interval, (Ωm)
Upper	m_{up} parameter of the translator function at resistivity model position, (Ωm)
Lower	m_{low} parameter of the translator function at resistivity model position, (Ωm)

Table 5. Format of gct-file.

3D-model grid, hard/soft data file (*.gslib file)
The HS-model realizations, Redox model realizations, and hard/soft data are provided in the same 3D-grid format (*.gslib).

The gslib-file consists of 5 lines with metadata, followed by a column for every point in the simulation grid.

Line #	Variable	Description
1	1	Number of columns in the simulation grid
	2	Number of rows in the simulation grid
	3	Number of layers in the simulation grid
	4	Grid size in the x direction (m)
	5	Grid size in the y direction (m)
	6	Grid size in the z direction (m)
	7	X coordinate (utm) of the center of the bottom left front cell
	8	Y coordinate (utm) of the center of the bottom left front cell
	9	Z coordinate (utm) of the center of the bottom left front cell
2	1	Number of variables
3	1	Name of Variable 1: Cluster (HS – unit)
4	1	Name of Variable 2: Log(resistivity)
5	1	Name of Variable 3: Redox state
6,7,…,end	1	HS-unit
	2	Log(resistivity)
	3	Redox state

Table 6. Format of the gslib-file. 3D-grid format of the HS-models, Redox-models, and hard/soft data.

From line 6 until the end, the lines consist of 3 columns, one for every variable. The order of the lines follow the grid, starting from the front bottom left cell, first increasing columns, then rows, and after one layer is completed, moving up one layer.

For each model realization, a gslib-file is generated. NaN values in grids are stated as -9999999.

The HS-unit values range from 0-3 with 0 the least clayey material and 3 the most clayey unit. The redox states either have a value of 1 or a value of 3. The value of 1 indicates oxidized conditions with no reduction occurring, and a value of 3 indicates a high redox rate occurring in that grid cell.

6. REFERENCES

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

Foged, N., 2014, Integration of borehole and airborne transient electromagnetic data for automatic compilation of large scale hydrogeological models, Aarhus University, Department of Geoscience, Hydrogeophysics Group.

Goovaerts, P., 1997, Geostatistics for natural resources evaluation, Applied geostatistics series: Oxford University Press.

Kim, H., P. B. E. Sandersen, R. Jakobsen, A. J. Kallesøe, N. Claes, G. Blicher - Mathiesen, N. Foged, J. Aamand, and B. Hansen, 2021, A 3D hydrogeochemistry model of nitrate transport and fate in a glacial sediment catchment: a first step toward a numerical mode: Submitted to Science of The Total Environment.

Mariethoz, G., P. Renard, and J. Straubhaar, 2010, The Direct Sampling method to perform multiple-point geostatistical simulations: Water Resources Research, **46**. http://dx.doi.org/10.1029/2008wr007621.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, 2011, Scikit-learn: Machine Learning in Python: Journal of Machine Learning Research, **12**, 2825-2830. <Go to ISI>://WOS:000298103200003.

Vilhelmsen, T. N., P. A. Marker, N. Foged, T. Wernberg, E. Auken, A. V. Christiansen, P. Bauer-Gottwein, S. Christensen, and A. Høyer, 2019, A regional scale hydrostratigraphy generated from geophysical data of varying age, type and quality: Water Ressources Management, **33**, no. 2, 15. http://dx.doi.org/https://doi.org/10.1007/s11269-018-2115-1.

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