

# **Hawke's Bay 3D Aquifer Mapping Project:** Heretaunga Plains, 3D hydrostratigraphic modelling

June 2022

Hawkes Bay Regional Council Publication No. 5583

Environmental Science

## **Hawke's Bay 3D Aquifer Mapping Project: Heretaunga Plains, 3D hydrostratigraphic modelling**

June 2022  
Hawkes Bay Regional Council Publication No. 5583

Prepared By:  
Aarhus University HydroGeophysics Group  
N Foged

For: Hawke's Bay Regional Council

Reviewed by:  
A Christiansen, Aarhus University HydroGeophysics Group  
S Harper, Hawke's Bay Regional Council  
SG Cameron, GNS Science  
ZJ Rawlinson, GNS Science



# Heretaunga Plains, 3D hydrostratigraphic modelling

Hawke's Bay 3D Aquifer Mapping Project, June 2022





## TABLE OF CONTENTS

<b>1. Introduction .....</b>	<b>2</b>
<b>2. Area and Data .....</b>	<b>4</b>
2.1 HSM area .....	4
2.2 SkyTEM data.....	5
2.3 Borehole data .....	6
<b>3. Methodology .....</b>	<b>8</b>
3.1 ACT-modeling.....	9
3.2 Clustering of CF-data and resistivity .....	12
3.3 Direct sampling .....	12
<b>4. Technical Specifications .....</b>	<b>15</b>
4.1 ACT setup .....	15
4.2 Clustering setup .....	17
4.3 Direct sampling setup.....	18
<b>5. Results and Deliverables .....</b>	<b>19</b>
5.1 Modeling result examples.....	19
5.2 Deliverables.....	22
<b>6. Discussion .....</b>	<b>24</b>
<b>7. Acknowledgments .....</b>	<b>25</b>
<b>8. References .....</b>	<b>26</b>



## 1. INTRODUCTION

Hydrostratigraphic modeling, Heretaunga Plains	
Project	<b>Hawke's Bay 3D Aquifer Mapping Project, Heretaunga Plains, New Zealand</b>
Key persons	<p><b>HGG, Geosciences, Aarhus University, Denmark</b> Project leader: Associated prof., PhD, <i>Anders V. Christiansen</i></p> <p><b>Modeling and reporting</b> Senior geophysicist, PhD, <i>Nikolaj Foged</i></p> <p><b>GNS Science, New Zealand</b> Project lead/management: <i>Zara Rawlinson</i></p>
Area	<b>Heretaunga Plains, ~ 660 km<sup>2</sup></b>

The Hawke's Bay 3D Aquifer Mapping Project (3DAMP) is a three-year initiative (2019–2022) jointly funded by the Provincial Growth Fund (PGF), Hawke's Bay Regional Council (HBRC) and GNS Science (GNS). The project applies SkyTEM technology to improve mapping and modeling of groundwater resources within the Heretaunga Plains, Ruataniwha Plains and Poukawa and Otane Basins. 3DAMP involves collaboration between HBRC, GNS and the Aarhus University HydroGeophysics Group (HGG).

SkyTEM data were collected in the Hawke's Bay region during January/February 2020 (SkyTEM Australia Pty Ltd 2020). Using these data, resistivity models were developed for the Heretaunga Plains by Rawlinson et al., (2021).

This report primarily explains and documents the process of utilising a hydrostratigraphic modeling concept (HSM) for unconsolidated sediments to compile realizations of 3D hydrostratigraphic models (HS-models) for the Heretaunga Plains model area (see Figure 1), based on the SkyTEM survey resistivity mapping results (Rawlinson et al., 2021) and available lithological logs (Tschirter et al., 2022). The modeling was primarily carried out by HGG, but with essential inputs, and evaluation from GNS.

Detailed structural information of the subsurface is fundamental for the development of hydrological models. Structural information applied in the HSM is primarily derived from the geophysical mapping results (SkyTEM-derived resistivity models) with support from lithological logs. However, some parts of the model area were either not covered by the SkyTEM method due to infrastructure, or the SkyTEM data were discarded due to extensive infrastructural noise, which results in resistivity



information gaps in the modeling area. Geostatistical methods were used to fill these data gaps and to incorporate relevant uncertainty measures in the construction of the HS-models. The output of the HSM concept is an ensemble of plausible 3D HS-models.

Section 2 of this report provides an overview of the data utilised for the HSM. Section 3 provides a description of the procedures and methods of the HSM-concept, while technical specifications for the HSM setup are stated in Section 4. Finally, Section 5 provides a few examples from the modeling and documents the formats of the deliverables.



## 2. AREA AND DATA

### 2.1 HSM area

The Heretaunga Plains SkyTEM survey area, and the corresponding Heretaunga model area (defined by Sahoo et al., 2022), covers approximately 660 km<sup>2</sup> and is located on the east coast of the North Island of New Zealand (see Figure 1). HSM was applied within the Heretaunga model area.

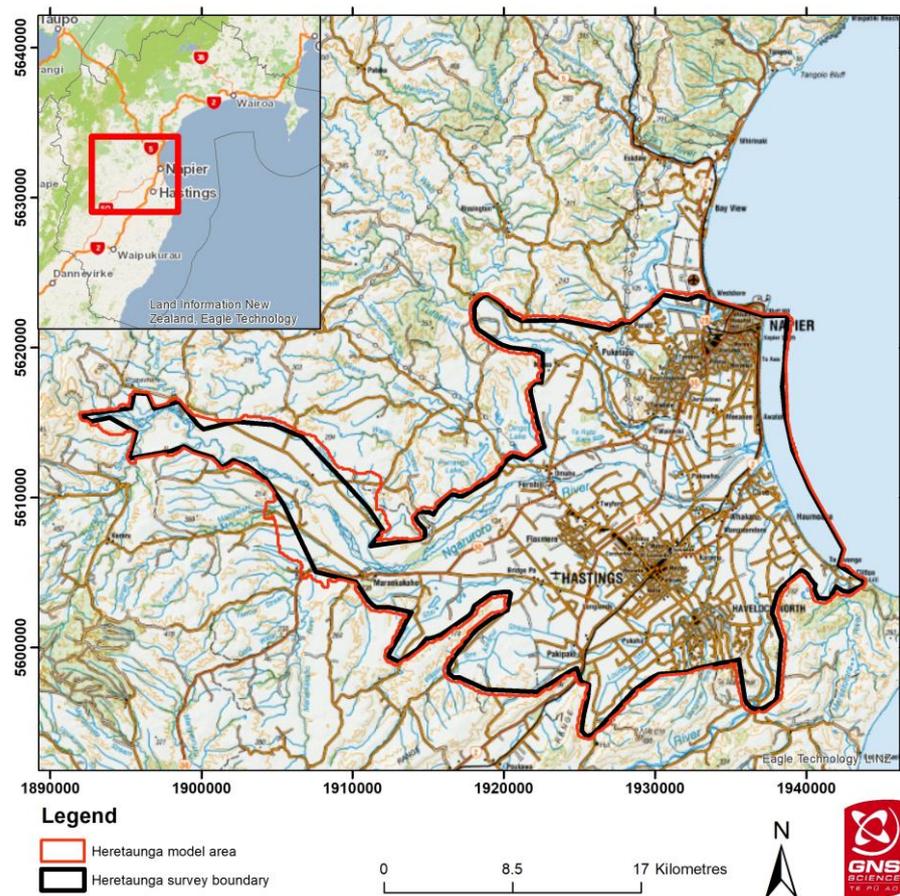


Figure 1. The Heretaunga Plains SkyTEM survey area (black polygon) and the Heretaunga model area utilised for HSM within this report (red polygon).



## 2.2 SkyTEM data

SkyTEM data were collected in the Hawke's Bay region during January/February 2020 (SkyTEM Australia Pty Ltd, 2020). Using these data, both sharp and smooth resistivity models were developed for the Here-taunga Plains by Rawlinson et al., (2021).

The green dots in Figure 2 mark the locations of the SkyTEM resistivity models utilised within the HSM. The offshore SkyTEM models and the models in magenta in Figure 2 were excluded due to the significant influence of saltwater in the resistivity models. The magenta models were selected based on areas containing resistivity values  $<8 \Omega\text{m}$  (freshwater should result in resistivity values  $>8 \Omega\text{m}$ ). These magenta areas are consistent with the locations of bores with saline water in the upper ~30 m and saline-influenced water within estuaries and river mouths (e.g., Tschirter et al., 2022). Although deeper bores, in the magenta areas, draw from a freshwater source, the entire resistivity model in these areas were excluded primary due to depth blinding limitation in the software. The exclusion of saline-influenced resistivity models was carried out since they might violate the basic assumption of the HSM: That a lower electrical resistivity corresponds to a lower hydrological conductivity and vice versa (see Section 3.1).

Additionally, as the HSM-concept was developed for modelling unconsolidated sediments, a preliminary top of basement surface (corresponding to the consolidated geology in the area such as limestone, mudstone and sandstone) was utilised to remove SkyTEM data considered to correspond to basement geology (blue dots in Figure 2). Details of the basement geology and manual surface delineation are described within Sahoo et al., (2022). The elevation of the preliminary basement surface was decreased by 60 m to account for uncertainty and used to clip the resistivity models (basement grid; Rawlinson, 2021). Although this adjustment does result in the inclusion of some basement-influenced SkyTEM data within the HSM dataset, it does exclude the bulk portion of the basement-influenced SkyTEM data. This is considered important to reduce the influence of these data on the HSM results, particularly as limestone and gravel exhibit overlapping resistivity ranges.

Vertically, the modeling area was limited upwards by the DEM (DEM grid; Farrier, 2020) and in depth by the resistivity model Standard Depth of Investigation (DOI) or else to a maximum limit of -360 m elevation (New Zealand Vertical Datum 2016). As the Standard DOI varies through the model area, the limit of -360 m elevation was chosen based on still retaining resistivity model coverage over a significant portion of the model area (see Rawlinson et al., 2021).

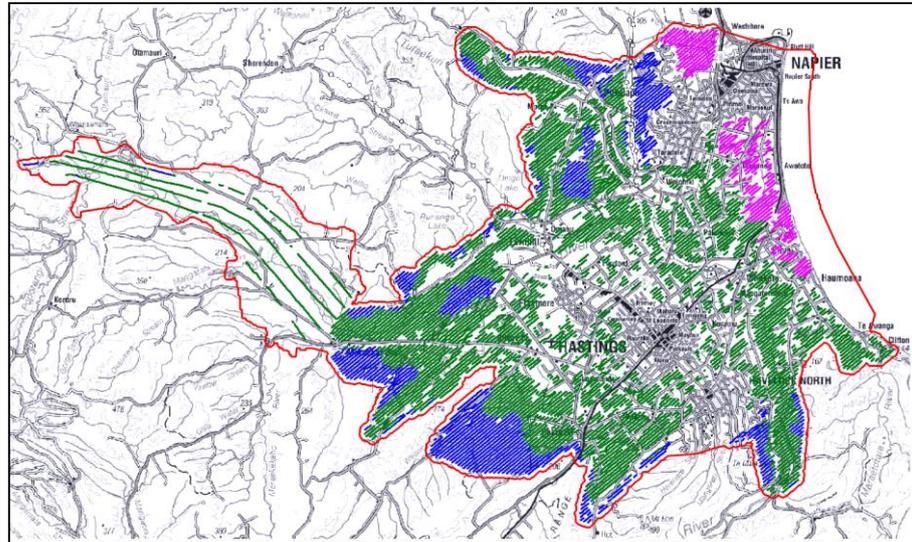


Figure 2. HSM model area (red line). Green dots mark the SkyTEM resistivity model locations utilised within the HSM; magenta dots mark resistivity model locations excluded from the HSM due to significant saltwater influence; blue dots mark resistivity model locations excluded from the HSM due to basement-influence.

### 2.3 Borehole data

A quality-coded lithological log dataset (Rawlinson, 2021; Tschritter et al., 2022) containing 4883 boreholes was assessed for use within the HSM. The locations of the boreholes are marked in Figure 3 and are color coded according to the drill depth. Most of the boreholes have drill depths <50 m and borehole information below a depth of 100 m is relatively sparse.

Boreholes with a poor-quality rating (rating of 5; Tschritter et al., 2022) were filtered out, resulting in 4214 boreholes used for the HSM.

The HSM-concept initially utilizes a binary approach for the borehole data and requires borehole lithologies to be classified into “aquifer” or “aquitard”. Lithologies may also be classified as “unknown”, in which case they do not contribute information to the HSM (see Section 3.1). The main separation of the borehole lithologies into aquitard units and aquifer units is listed in Table 1. Classifications were undertaken based on an understanding of local conditions (e.g., Tschritter et al., 2022) and some iterations of the definitions were tested within preliminary HSM simulations.

Python scripts were used to assess the main lithology, full lithological descriptions, and comments/notes within both the lithology and borehole construction datasets to derive the final classified dataset (Rawlinson,



2021). The lithological log quality index (Tschritter et al., 2022) was used to assign different uncertainties for the lithological data within the HSM-concept (see Section 3.1).

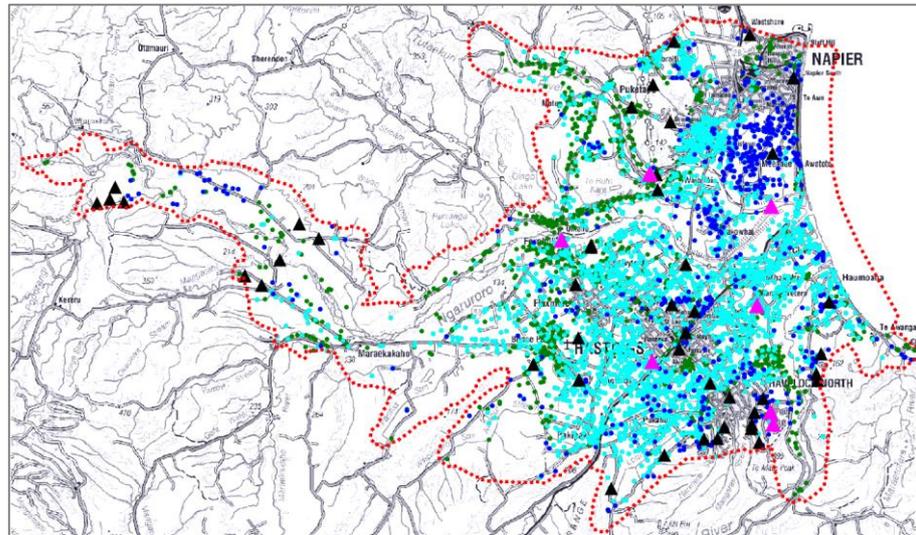


Figure 3. Borehole locations, colored by drill depth. Drill depth: Green dots 0-20 m, Cyan dots 20-50 m, Blue dots 50-100 m, Black triangles 100-200 m, Magenta triangles >200 m.

Lithology	Acronym	Aquitard/Aquifer
Aquitard	aqt	Aquitard
Ash/Pumice	as	Aquitard
Clay, pure	cl1	Aquitard
Clay, with fine grained material	cl2	Aquitard
Clay, with coarse grained material	cl3	Aquitard
Siltstone	is	Aquitard
Mudstone	ms	Aquitard
Peat/veg/wood	or	Aquitard
Shale	sh	Aquitard
Silt, with no coarse material	si1	Aquitard
Aquifer	aq	Aquifer
Sand	sa	Aquifer
Gravel	gr	Aquifer
Conglomerate	cm	Unknown
Shell	ll	Unknown
Limestone	ls	Unknown
Silt	si	Unknown
Sandstone	ss	Unknown
Volcanic rock	vo	Unknown
Unknown	x	Unknown
fill/topsoil/unknown	xx	Unknown

Table 1. Lithology classification into aquitard, aquifer, and unknown. Lithologies listed as “unknown” enter the HSM-concept as “no data”.



### 3. METHODOLOGY

This section holds a short overall description of the HSM-concept, with more detailed explanations of the different components of the HSM-concept in the sub-sections. The setup parameters for the specific area are stated in Section 4 - *Technical Specifications*.

The overall purpose of the modeling 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 any later hydrological modeling (e.g., numerical groundwater flow modeling).

Figure 4 shows a flowchart of the HSM-concept. The main source of structural information comes from the resistivity models. Combining the resistivity dataset with lithological borehole data in the Accumulated Clay Thickness (ACT) procedure, the resistivity values are translated into clay content (CF-model). This results in a data pair of resistivity and CF-values for each resistivity layer at each resistivity model location.

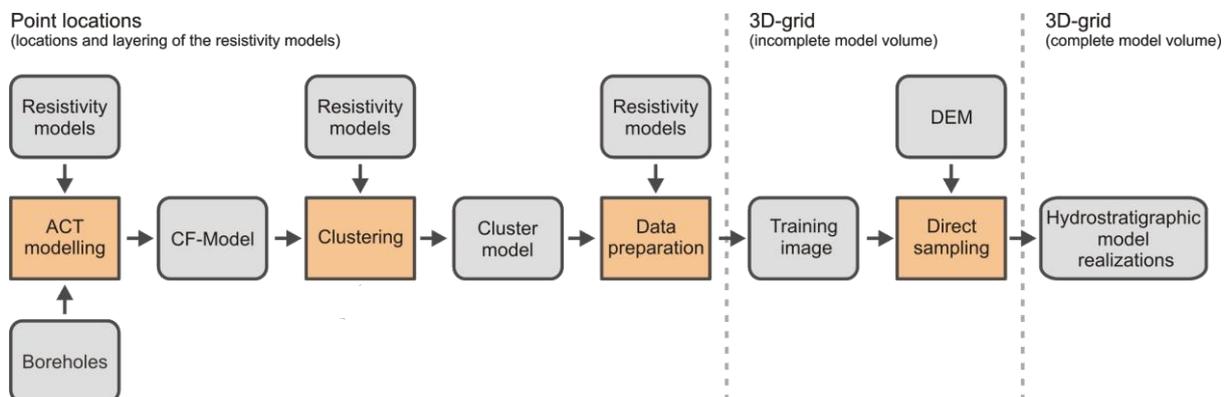


Figure 4. Flowchart of the HSM-concept.

The CF-resistivity data pairs are then converted to a number of discrete hydrostratigraphic units (HS-units) through a clustering routine.

At this stage, the HS-units are only present at the resistivity model positions. To populate the full model volume with HS-units and to generate multiple plausible hydrostratigraphic model realizations, the direct sampling method is applied.



The direct sampling method 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-kriging HS-units, resulting in an incomplete 3D-grid. The direct sampling then uses the patterns within the training image to generate multiple realizations of complete 3D HS-models. The incorporation of uncertainty is handled through the separation of hard and soft input data.

### 3.1 ACT-modeling

The geophysical output from a SkyTEM 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 models. The ACT (Accumulated Clay Thickness) concept automates this process by determining the optimum cutoff resistive values of a resistivity-to-aquitard/aquifer (for example, "clay"/"sand") translator function, based on information from lithological logs.

The input to the concept is the accumulated meters of the aquitard units (clay) in depth intervals, as shown in Figure 5a, where the clay content of an interval is stated as a fraction of the interval length ( $CF_{Log}$ ). With the translator function in Figure 5b a resistivity model can also be converted into clay content in the same intervals ( $CF_{Res}$  in Figure 5a).

The translator function is defined by two parameters – an Upper and a Lower resistivity value ( $m_{up}$ ,  $m_{low}$ ). With the translator function in Figure 5b, resistivity layers below 40  $\Omega m$  get a weight of  $\sim 1$ , which means that the full length of the resistivity layer is presumed to be clay/aquitard; while resistivity layers above 70  $\Omega m$  will get a weight of  $\sim 0$ , corresponding to no-clay content (e.g., sand). Hence, resistivity values between 40-70 m, in this example, will result in  $CF_{Res}$  values between 0 and 1.

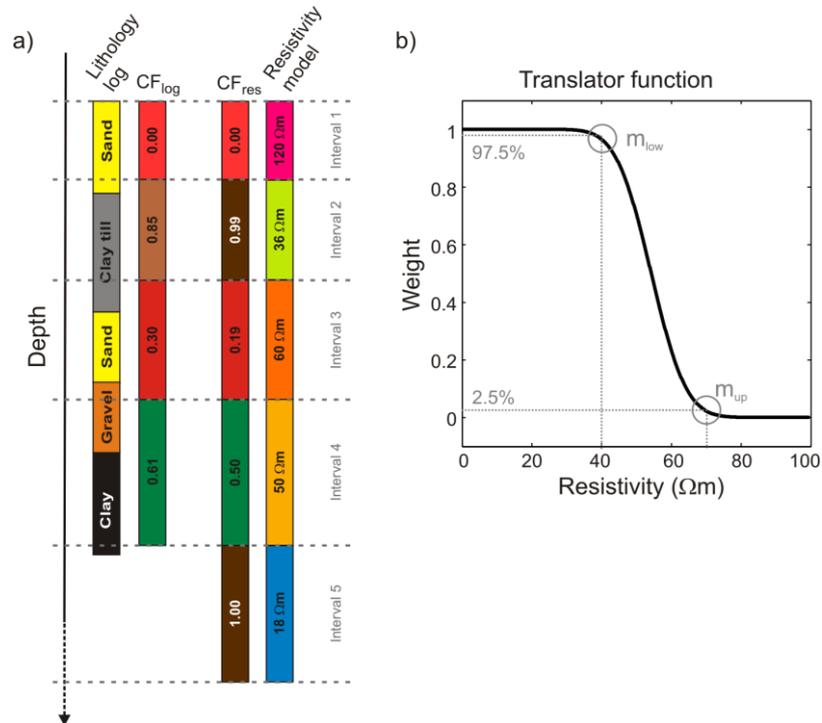


Figure 5. a) Lithological log with the corresponding CF-values, and the resistivity model and the CF values obtained by using the translator function in b). In this case, the depth intervals follow the layer discretization of the resistivity model.

Through inversion, we determine the optimum parameters of the translator functions ( $m_{up}$ ,  $m_{low}$ ) 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 ACT is that the translator function can change horizontally and vertically, adapting to the local conditions and borehole data. Therefore, a number of translator functions covering the entire survey, both vertically and horizontally, are applied. The translator functions are spatially distributed in a regular 3D grid as shown in Figure 6. The vertical discretization of the translator function grid follows the calculation intervals from Figure 5a, while the horizontal discretization is set to allow for the needed flexibility in the translation, depending on borehole density and geological complexity.

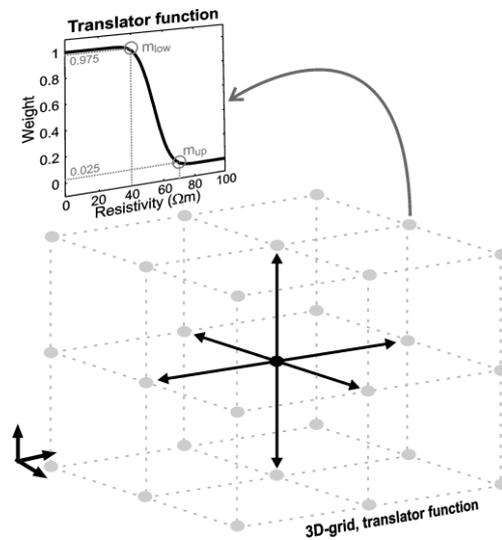


Figure 6. Translator model grid. Each node (gray dots) in the grid holds a translator function. The  $m_{up}$ ,  $m_{low}$  parameters of the translator function are constrained to the neighboring nodes laterally and vertically, as indicated with the arrows in the plot for a single node.

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 6.

The optimum translator function grid from the ACT inversion is used for a final translation of all the resistivity models to CF-values. With the ACT-concept we incorporate the key information from both the resistivity and lithological data into the CF-model.

This ACT 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. An estimated quality/uncertainty is assigned to the borehole data based on the borehole quality index. The distances between boreholes and the resistivity models are also taken into account in the weighting of the borehole data. In exceptional cases, specific boreholes are excluded or the uncertainty increased due to a clear mismatch to close neighboring borehole(s) and/or the resistivity models.

The lithological borehole information is normally extracted from a borehole database and the extraction of the clay/ aquitard units is done regardless of the type of clay, i.e. both "Tertiary lacustrine clay" and "Quaternary meltwater clay" are labeled as clay/aquitard. Lithologies marked as "unknown" correspond to no borehole information.



Due to the spatially varying resistivity to clay translation, the ACT-concept can handle changes in the pore water resistivity, but only within the basic assumption of the concept that clay units are more electrically conductive than the non-clay units. This basic assumption may not be correct in case of saline pore water.

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

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

Borehole data are often relatively sparse, especially for the deeper part of a survey area, and it is furthermore 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. This means that the constraint setup and starting 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 could afterwards be used within a groundwater model.

For a 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 performed directly on the resistivity CF variables, but on their principal components after normalization. Using the principal components ensure that the variables are uncorrelated, which is a condition for the clustering routine. Details about the clustering scheme used 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 7, where Figure 7a shows the volume with its data gaps. Plausible full 2D profiles are generated through direct sampling, as demonstrated in Figure 7b and c.

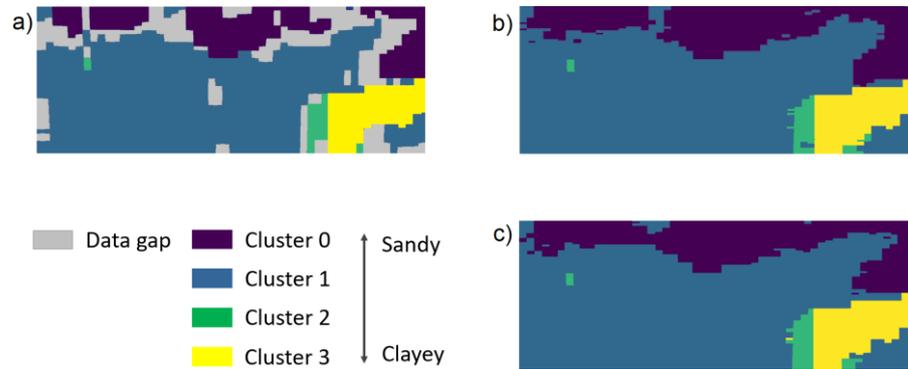


Figure 7. Examples of direct sampling procedure used to fill out full model volumes and generate multiple realizations. a) 2D profile with data gaps that are remediated by Direct Sampling; b-c) direct sampling generated profiles from the profile in (a).

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 points that are set to a specific value and do not change in the different simulations (anchor points). The hard data 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 resistivity values, and are geographically located at the geophysical model positions. The HS-units and resistivity data are interpolated to a 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 3D modeling grid, but 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 direction than in the vertical direction, the horizontal search radius is set larger than the vertical, 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 are collected. This incomplete 3D grid becomes the training image for the direct sampling simulations.

The hard data for the simulations are selected as the most certain data points in the training image and the selection is based on a silhouette index computation. The silhouette index is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). Figure 8a shows an example of how the resistivity and CF data are split up into five clusters, with the associated silhouette index in Figure 8b. Only cluster data with a high silhouette index enter as hard data, which corresponds to the cluster data points that are closest to their respective cluster center. A silhouette index threshold value for hard data selection could for example be the mean silhouette index.

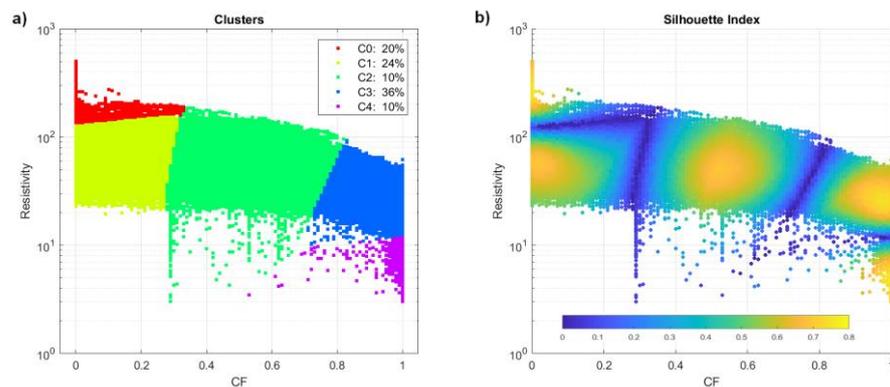


Figure 8. Five cluster example. Plot a) Input CF and resistivity data (x,y axis) and the resulting cluster group (dot color). b) Corresponding silhouette index (SilScor). Only data with a high silhouette index are used as hard data.



## 4. TECHNICAL SPECIFICATIONS

This section contains the parameter types, specific parameter settings, etc. for the HSM area.

### 4.1 ACT setup

The vertical discretization of the ACT translator function grid follows the initial layer discretization of the resistivity models, with the exception that the first seven layers (upper 10.5 m) were merged into 3 layers. Due to the size of the survey area, the ACT was conducted in two areas, each with a horizontal translator node spacing of 900 m.

Prior to the clustering, the output from the two ACT areas were merged together. Figure 9 shows the translator model grids of the two ACT areas.

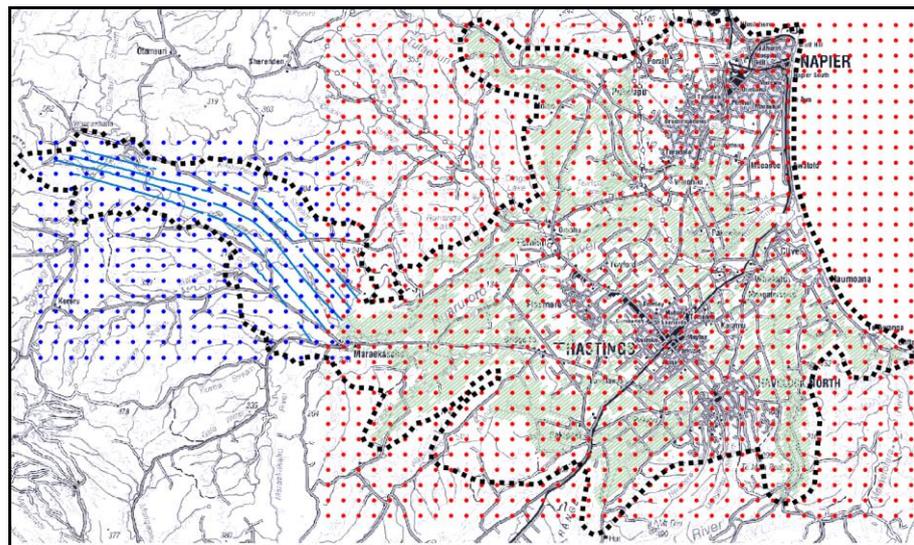


Figure 9. Red (area 1) and blue (area 2) dots mark the translator function grid in the horizontal plane of the two ACT areas. The small light blue and light green dots mark the SkyTEM resistivity model locations.

The ACT setup for the two areas was the same, with the exception that the western area only holds 23 layers due to a shallower depth to basement and lack of deep borehole information. Table 2 provides detailed ACT setup information.

The sharp resistivity model result was used instead of the smooth resistivity model, since the sharp model provides a sharper definition of the



major resistivity structures compared to the smooth resistivity model, which is beneficial in the later clustering of the data.

The locations of boreholes with lithological information for the survey area are plotted in Figure 3 with colors indicating the drill depth. For the ACT, boreholes with no resistivity model within 500 m were filtered out, resulting in approximately 3900 boreholes being used for the ACT modeling. Boreholes >500 m from a resistivity model were reintroduced in the clustering (Section 4.2).

As mentioned, the drill depth is 0-50 m for the majority of the boreholes, and the borehole information is very sparse at depths of >100 m. The resulting translator model for the deeper part is, therefore, to a large extent determined by the starting values for the translator model in the ACT inversion. Starting values for the translator model were set uniformly for the entire area, and were based partly on resistivity histograms for the lithology groups and partly on preliminary ACT inversions that tested different starting models.

ACT setup		
Input	Type	
Resistivity model input	SkyTEM, Sharp SCI inversion result (Rawlinson et al., 2021)	
Resistivity model blinding	DOI standard	
Borehole input	Classified lithological logs from GNS (Rawlinson, 2021)	
Borehole quality	Rating 1-5 (1=best)	
Uncertainty, rating, % of interval	R1=10%, R2=15%, R3=30%, R4=50%, R5=excluded	
Translator model type	Complementary error function	
Search radius, borehole to res. model	500 m	
Translator function grid	Vertical	Horizontal
Discretization	Layer thicknesses of resistivity model*	400 m
Model constraints, $m_{up}$ and $m_{low}$	2.0	1.8
Model layers, max depth (area 1;2)**	28 layers, 389 m; 23 layers, 206 m	
Translator function	Upper	Lower
Start values	30 $\Omega$ m	45 $\Omega$ m
Prior (factor)	8 (very weak)	

Table 2. Detailed information for the ACT setup. \* First seven resistivity layers (upper 10.5 m) were merged into 3 layers. \*\* see Figure 9 for the area locations.

Normally, borehole information is not included directly in the clustering. Here, large gaps were present in the model domain due to cities/towns and coupling in the SkyTEM data. In these areas, numerous boreholes were present, without being a part of the ACT due to the 500 m search radius limit. We therefore reintroduced the boreholes not initiated in the ACT in the later clustering. To do this, we need both a CF and a resistivity



value. The CF-values for the reintroduced boreholes were calculated as for the CF input data for the ACT: a resistivity value was estimated by using the inverse of the local translator model, with CF-values as input, thus a resistivity value was produced as the output.

## 4.2 Clustering setup

Before the clustering was performed, the datasets were prepared in the following manner:

1. If more than one resistivity layer enters an ACT interval, the mean resistivity value of the interval is used. The mean resistivity value was calculated as the inverse of the mean conductivity.
2. The resistivity values were truncated to avoid outliers having a large impact on the later data scaling.
3. The resistivity data were log-transformed.
4. All datasets were 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 that were already within the [0-1] range.
5. Optional: Weights can be added to the different data types (not applied in this instance).
6. The Principal components (PC) of the data types were calculated.

The K-means clustering was implemented in the `sklearn.clustering` Python library with `k++` means initialization (Pedregosa et al., 2011). Preliminary clustering were undertaken to determine the optimal parameters, particularly the number of clusters utilised. The final settings for the clustering are listed in Table 3.

Data preparation		
Data types	Resistivity	CF-values
Truncation [min, max]	[3, 500]	---
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	5	

Table 3. Data preparation and clustering setup.



### 4.3 Direct sampling setup

#### Training images

The kriging process to compile training images was performed in SGEMS software with variogram parameters that were estimated based on the resulting data cloud from the clustering procedure. Parameters for the kriging process are listed in Table 4.

	Resistivity data	Cluster data
Data transform	Log	none
Interpolation	Ordinary kriging	Indicator kriging
Vertical resampling	5 points per ACT layer	
Conditioning data points, [min, max]	[0, 32]	
Search ellipsoid, [min, mid, max]	[200, 200, 20] m	

Table 4. Kriging parameters for training image generation.

#### Direct Sampling parameters

The settings for direct sampling are included in Table 5. See Section 2 for further details on the area polygon, as well as top and bottom grids utilised.

Direct Sampling grid setup	
Model grid vertical discretization, elevation interval	10 m, [-360; +430] m
Model grid horizontal discretization	100 m
Model grid top, bottom blanking (z)	DEM grid, basement grid
Model grid area blanking (x,y)	Heretaunga model area
Direct Sampling parameters	
Simulation variable(s)	Clusters
Training image	Clusters
Hard data	From training image
Hard data threshold	Mean silhouette index = 0.58
Maximum numbers of neighboring nodes	20
Distance threshold	0.05
Max scan fraction of training images	40%
Number of realizations	500

Table 5. Direct Sampling parameters



## 5. RESULTS AND DELIVERABLES

Using the SkyTEM-derived sharp resistivity models and classified and quality-coded lithological information, a CF-model was created using the ACT concept - resulting in CF-values at the SkyTEM resistivity model locations. The CF-model was further clustered into five HS-units, and with the direct sampling routine, 500 model realizations covering the full 3D volume of the Heretaunga model area were generated.

This section shows a few modeling result examples and provides both an overview and a detailed format description of the deliverables.

### 5.1 Modeling result examples

This section provides some visual examples of the modeling results. Figure 10 shows an example of a horizontal slice through the model at elevation -20 m, with the resistivity and CF values (plot a-b), and two cluster realizations (plot c-d). Figure 11 shows a W-E cross-section example.

The results are consistent with the expected hydrogeology based on current knowledge (e.g., Tschirner et al., 2022). It can be seen in Figure 10 that the top part of the basement geology is included within the HS-model (CF-model 1 and clusters 3 & 4).

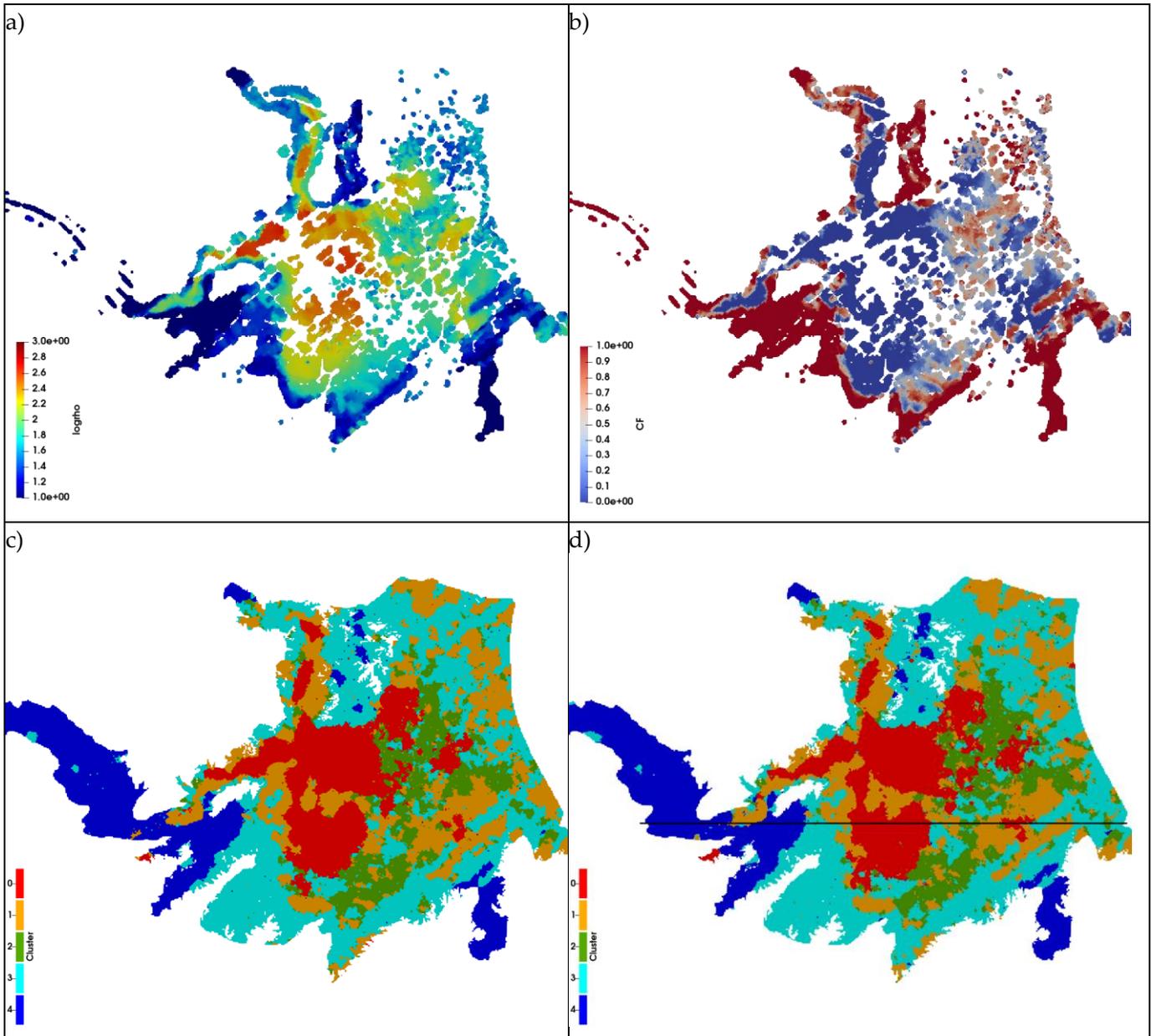


Figure 10 Horizontal slice, elevation -20 m. a) Resistivity model, b) CF-model (~1=aquitard; ~0=aquifer), c) and d) two cluster simulations (cluster 4 is the most clay rich/impermeable unit, while cluster 0 is the most permeable/lowest clay content unit). The black line in d) marks the location of the cross section in Figure 11.

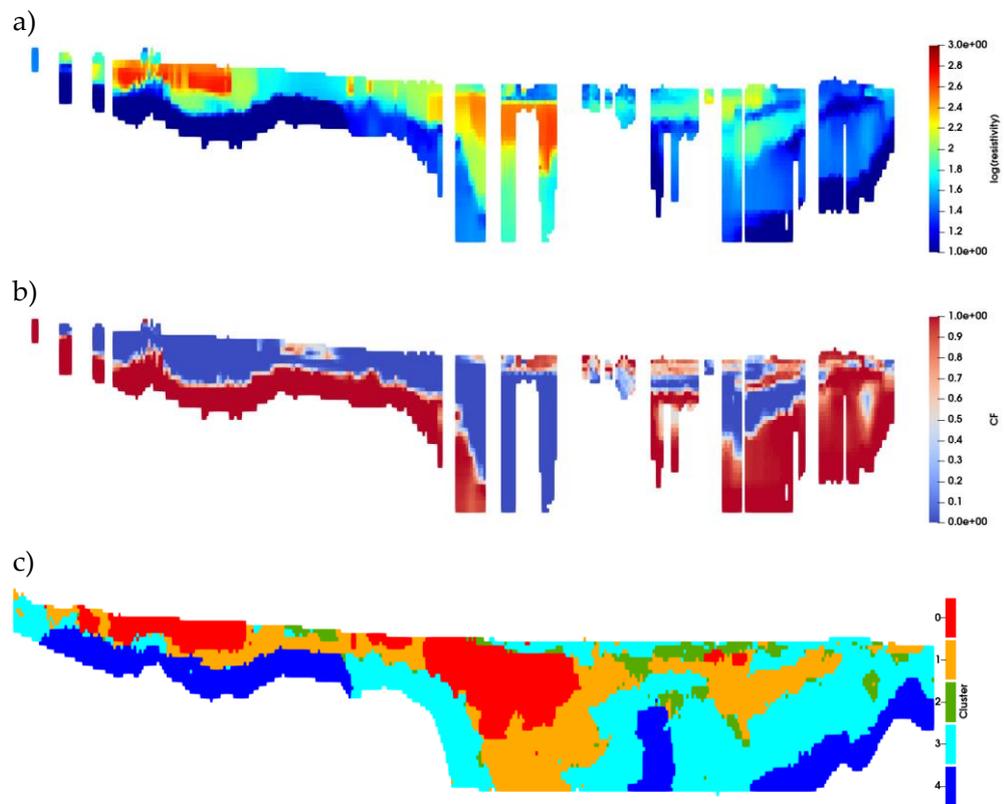


Figure 11. W-E cross-section example. a) Resistivity model, b) CF-model ( $\sim 1$ =aquitard;  $\sim 0$ =aquifer), c) cluster simulations (cluster 4 is the most clay rich/impermeable unit, while cluster 0 is the most permeable/lowest clay content unit). The location of the section is marked in Figure 10d.



## 5.2 Deliverables

This section provides a format description of the different digital deliverables. Note that all stated UTM coordinates are geo-referenced to the coordinate system: NZGD2000 New Zealand Transverse Mercator 2000 (EPSG 2193).

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\DS\ModelRealizations\

### CF-model (\*.gct-file)

The gct-files hold information about the clay contents ( $CF_{Res}$  values) for the translation of the resistivity 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 6.

Column 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-Y 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 resistivity of interval, ( $\Omega m$ )
Upper	$m_{up}$ parameter of the translator function at resistivity model position, ( $\Omega m$ )
Lower	$m_{low}$ parameter of the translator function at resistivity model position, ( $\Omega m$ )

Table 6. Format of gct-file.

### 3D-model grid, hard/soft data file (\*.gslib file)

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



The gslib-file consists of header lines with metadata, followed by the variable data in columns (one line for every node in the grid), as described in Table 7.

Line #	Variable	Description
1	1-3	Number of x,y,z nodes
	4,5,6	x,y,z node distances (m)
	7,8,9	x,y,z coordinate (utm) of the center of the bottom left front node
2	1	Number of variables
3	1	Name of Variable
4-end	1	Variable data (one column per variable)

*Table 7. Format of the gslib-file. 3D-grid format of the HS-models, and hard/soft data.*

The order of the variable data follows the grid, starting from the front bottom left node, first increasing columns, then rows, and after one layer is completed, moving up one layer. NaN values in grids are stated as -9999999.

For each model realization, a gslib-file is generated. The HS-unit values range from 0-4 with 0 the least clayey material and 4 the most clayey unit.



## 6. DISCUSSION

The HSM aims at producing HS-units for unconsolidated sediments. Due to a cautious approach being taken with the basement masking in this work (60 m depth added to ensure all unconsolidated material was included within the HSM), further masking of the resulting 3D-model grids with respect to e.g. the basement grid must be expected before incorporating the HSM result into a groundwater flow model.

Due to the limited lithological bore log data with depth, determination of the resistivity to CF translation in the ACT concept is more uncertain with depth. The uncertainty estimates for the CF-values are calculated based on the obtained local resistivity to CF translator function and the CF uncertainty estimates, therefore, they primarily reflect the position of the resistivity value on the translator function. Thus, resistivity values located on the sloping part of the translator function results in a high CF uncertainty, while resistivity values at the flat part of translator function (low and high resistivities) result in a low CF uncertainty. The limited deep borehole information also indirectly results in the uncertainty for HS-units increasing with depth.



## **7. ACKNOWLEDGMENTS**

This work has been jointly funded by the New Zealand Government's Provincial Growth Fund, Hawke's Bay Regional Council and GNS Science's Strategic Science Investment Fund (Ministry of Business, Innovation & Employment).



## 8. 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.
- Farrier T., 2020, 3D Aquifer Mapping Project DEM Version 2. Napier (NZ): Hawke's Bay Regional Council.
- Foged, N., P. A. Marker, A. V. Christiansen, P. Bauer-Gottwein, F. Jørgensen, A.-S. Høyer, and E. Auken, 2014, Large scale 3D-modeling by integration of resistivity models and borehole data through inversion: *Hydrology and Earth System Sciences*, **18**, 4349-4362.
- Goovaerts, P., 1997, *Geostatistics for natural resources evaluation*, Applied geostatistics series: Oxford University Press.
- 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.
- Rawlinson, ZJ., 2021, Personal communication: datasets provided in October 2021. Senior Hydro-Geophysicist, Hydrogeology & Geophysics; GNS Science, Wairakei, NZ.
- Rawlinson, ZJ., Foged, N., Westerhoff, RS., Kellett, RL. 2021 Hawke's Bay 3D Aquifer Mapping Project: Heretaunga Plains SkyTEM data processing and resistivity models. GNS Science consultancy report 2021/93. 90 p.
- Sahoo, TR., Rawlinson, ZJ., Kellett, RL., 2022, Hawke's Bay 3D Aquifer Mapping Project: Delineation of major hydrogeological units within the Heretaunga Plains from SkyTEM-derived resistivity models. Lower Hutt (NZ): GNS Science. in prep. Consultancy Report 2022/30.
- SkyTEM Australia Pty Ltd, 2020, Acquisition and processing report: SkyTEM helicopter EM survey, Hawkes Bay, NZ. Malaga (AU): SkyTEM Australia Pty Ltd. 33 p. Report AUS 10056. Prepared for Hawke's Bay Regional Council.



- Tschritter, C., Kellett, R.L., Rawlinson, Z.J., Griffin, A.G., 2022, Hawke's Bay 3D Aquifer Mapping Project: Heretaunga Plains data and model inventory. Wairakei (NZ): GNS Science. 96 p. Consultancy Report 2021/113.
- 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 Resources 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.