Contributions to uncertainty related to hydrostratigraphic modeling using Multiple-Point Statistics

Adrian A.S. Barfod[1,2], Troels N. Vilhelmsen[2], Flemming Jørgensen[1], Anders V. Christiansen[2], Anne-Sophie Høyer[1], Julien Straubhaar[3], Ingelise Møller[1]

[1]Derpartment of Groundwater and Quaternary Geological Mapping, Geological Survey of Denmark & Greenland (GEUS), C.F. Møllers Allé 8, 8000 Aarhus C, Denmark
[2]Hydrogeophysics Group, Department of Geoscience, Aarhus University, C.F. Møllers Allé 4, 8000 Aarhus C, Denmark
[3]Centre d'Hydrogéologie et de Géothermie (CHYN), Université de Neuchâtel, Switzerland

Correspondence to: Adrian A.S. Barfod (adrian.s.barfod@gmail.com)

1

Contents

15

20

25

30

35

40

Abstract.

Forecasting the flow of groundwater requires a hydrostratigraphic model, which describes the architecture of the subsurface. State-of-the-art Multiple-Point Statistical (MPS) tools are readily available for creating models depicting subsurface geology. We present a study of the impact of key parameters related to stochastic MPS simulation of a real-world hydrogeophysical dataset from Kasted Denmark using the *snesim* algorithm. The goal is to study how changes to the underlying datasets propagate into the hydrostratigraphic realizations when using MPS for stochastic modeling. This study focuses on the sensitivity of the MPS realizations to the geophysical soft data, borehole lithology logs, and the Training Image (TI). The modeling approach used in this paper utilizes a cognitive geological model as a TI to simulate ensemble hydrostratigraphic models. The target model contains three overall hydrostratigraphic categories, and the MPS realizations are compared visually, as well as quantitatively using mathematical measures of similarity. The quantitative similarity analysis is carried out exhaustively, and realizations are compared with each other as well as with the cognitive geological model.

The results underline the importance of geophysical data for constraining MPS simulations. Relying only on borehole data and the conceptual geology, or TI, results in a significant increase in realization uncertainty. The airborne transient electromagnetic SkyTEM data used in this study cover a large portion of the Kasted model area, and are essential to the hydrostratigraphic architecture. On the other hand, the borehole lithology logs are sparser, and only 410 boreholes were present in this study. The borehole lithology logs infer local changes in the immediate vicinity of the boreholes, thus providing limited large-scale structural information. Lithological information is, however, important for the interpretation of the geophysical responses. Finally, the importance of the TI was studied. An example was presented where an alternative geological model from a neighboring area was used to simulate hydrostratigraphic models. It was shown that as long as the geological settings are similar in nature, the realizations, although different, still reflect the hydrostratigraphic architecture. If a TI containing a biased geological conceptualization is used, the resulting realizations will resemble the TI and contain less structure in particular areas, where the soft data show almost even probability to two or all three of the hydrostratigraphic units.

1 Introduction

Geological models are important from both a societal and economic perspective, since they are used to locate essential natural resources, such as freshwater, oil, metals, rare earth minerals *etc*. Additionally, they are used in risk assessment related to natural hazards, such as earthquakes, sinkholes, volcanic eruptions, and landslides. Building 3D models depicting real-world subsurface geology is no trivial task. Information from multiple sources is required, *i.e.* conceptual geological understanding, geological information, lithology logs and geophysical data. Such data are sparse, uncertain and redundant. Dataset gaps force geoscientists to make uncertain predictions or estimates, which carries over into the resulting geological model. During the modeling procedure, such problems are dealt with as best as possible. Gaps in knowledge will render the resulting model uncertain, and quantifying such uncertainty is essential to making better use of the models, and to making better predictions.

A common approach for building geological models is cognitive modeling (e.g. Jørgensen et al., 2013; Royse, 2010). Here, the dataset containing borehole lithology logs and geophysical models are co-interpreted by a professional with experience in the fields of geoscience, geophysics, and geological modeling, with a relevant regional conceptual model in mind. This modeling approach is deterministic, and results in a single model realization. These specialists are trained in assessing the uncertainty of the underlying structures, and qualitative uncertainty estimates are often made on the structural model. *E.g.* indicating different levels of uncertainty in different subparts of the model domain. However, qualitative uncertainty estimates are difficult to carry over into the subsequent analysis, and the effect of the uncertainty of the geological model can therefore not be quantified in the resulting forecasts. If the forecasts are based on a single geologic model, the prediction does not encase the full complexity of the problem. Alternatively, if the model uncertainty can be quantified, it enables the option to include it

in the forecast. However, quantifying the uncertainty in a cognitive modeling approach is difficult and tedious (Seifert et al., 2012). Another approach is stochastic modeling using Multiple-Point Statistics (MPS) methodologies, *e.g.* Comunian *et al.* (2012), Ferré (2017), He *et al.* (2016), Okabe and Blunt (2005), Pirot (2017). MPS provides a framework which can integrate geophysical and borehole information, as well as conceptual geological information via a so-called Training Image (TI). Multiple model realizations are created from the dataset. The resulting model ensemble reflects the uncertainty related to the underlying datasets and overall modeling procedure.

We present a study of the uncertainty related to stochastic hydrostratigraphic MPS modeling of a hydrogeophysical dataset from Kasted, Denmark (Figure 1). The goal is to understand the consequences of modifying the underlying MPS setup to reflect some of the biases related to a real-world hydrogeophysical dataset and study the propagation of the uncertainty into the hydrostratigraphic models. We show how uncertainty related to resistivity data, measured with the airborne transient electromagnetic SkyTEM system (Sørensen and Auken, 2004), and borehole lithology logs influences the hydrostratigraphic modeling realizations. Two readily available MPS tools are showcased. The first tool is the Direct Sampling (DS) method for reconstruction of incomplete datasets (Mariethoz and Renard, 2010). The other MPS tool is the "Single Normal Equation Simulation" (*snesim)*, which is used for stochastic hydrostratigraphic modeling (Strebelle, 2002). The stochastic models will be divided into 6 overall cases, or 8 sub-cases. The first case is the basic modeling setup, which uses SkyTEM resistivity models as soft data, boreholes as hard data and a cognitive 3D hydrostratigraphic model as a TI. The remaining cases are then modified versions of the basic modeling setup, which are designed to reflect different types of modeling uncertainty. In other words, one of the overall goals of this study is to improve the Kasted model by using stochastic ensemble modeling to quantify the uncertainty of the model, such as suggested by Ferré (2017) and Pirot (2017).

In this study we will use two overall types of data, *i.e.* geophysical data and borehole data. Associated with these data types are the definitions of hard and soft data. Typically, hard data is considered certain information without an associated uncertainty, while soft data is uncertain information, which can be associated with an uncertainty. Geophysical data are typically considered soft data (Strebelle, 2002). Geophysical data are spatially dense and provides a smeared image of the overall subsurface geology. Resolution decreases with depth, and diminishes at a specific depth, which is dependent on the geophysical method. Geophysical instruments portray bulk physical properties of the subsurface. Although geophysical data provides spatially dense information, it is not possible to exhaustively sample the subsurface. The density of the geophysical data will affect the final uncertainty. The raw geophysical data goes through a processing and modeling step, where the raw data are translated into geophysical models. During this step incorrect measurements, due to instrument error or interference, are identified and removed, further decreasing the geophysical information density. Such incomplete data can either be reconstructed or used as is during modeling. We present both cases and show the advantages/disadvantages of both in relation to MPS modeling. Another consideration in regards to modeling of geophysical data is the choice of inversion scheme and thereby the choice of a priori information (*e.g.*, Ellis and Oldenburg, 1994; Tarantola and Valette, 1982). Here, several approaches can be taken which yield different geophysical models. A common inversion schemes for Airborne Electromagnetic (AEM) data, such as SkyTEM data, is the "Spatially Constrained Inversion" (SCI) (Viezzoli et al., 2008). However, this inversion approach does not represent the subsurface properly, *e.g.* layer boundaries are smeared and extreme values are not represented properly. A so-called sharp inversion scheme, suggested by Vignoli *et al.* (2015), tackles such issues. Therefore, the choice of inversion scheme influences the hydrostratigraphic model and should be considered as an integral step in the hydrostratigraphic modeling process.

The other source of data in this study is borehole lithology logs, which are commonly considered to be "ground truth" or hard data (e.g. Gunnink and Siemon, 2015; Tahmasebi et al., 2012). However, lithology logs are also uncertain. In this study the boreholes are divided into 5 quality groups, of which only boreholes above a chosen threshold are used. Generally the uncertainty of borehole lithology logs relates to a number of parameters, such as: drilling methods, the frequency with which

sediment samples are collected, precision with which the location is measured, the purpose of the borehole, the choice of drilling contractor *etc.* – see Barfod *et al.* (2016) and He *et al.* (2014) for more detail. The resolution of borehole lithology logs is especially dependent on the sampling method. If a core is extracted for the entirety of the borehole, the resolution is, in principal, unlimited. However, this is expensive. It is more common to use either an auger drill, rotary drill or a cable tool, which yields a relatively limited resolution, compared to core drilling, depending on how samples are collected and handled.

In this study data is modeled using *snesim* MPS framework. The *snesim* method is used to create geologic models and study the uncertainty related to the geophysical data, lithology logs, and conceptual geological model (TI). It is carried out on a hydrogeophysical dataset from Kasted, Denmark. Since subsurface hydraulic flow is largely controlled by geological heterogeneity (e.g. Feyen and Caers, 2006; Fleckenstein et al., 2006; Fogg et al., 1998; Gelhar, 1984; LaBolle and Fogg, 2001; Zhao and Illman, 2017), accurate geological models are crucial to accurate predictions of hydraulic flow. Geological units, however, contain additional complexities not related to hydrologic units; therefore, from here on, the concept of hydrostratigraphic units will be used. A detailed definition of hydrostratigraphic classification is given by Maxey (1964).

The *snesim* modeling scheme resulted in a total of 400 hydrostratigraphic models. Visual comparison of such a large number of realization ensembles is tedious and subjective, but offers an overall understanding of the geological realism of the models (Barfod et al., 2018). Barfod *et al.* (2018) present a comparison of 3D hydrostratigraphic models using the Modified Hausdorff Distance (d_{MH}). However, the d_{MH} was proved to be computationally expensive. Therefore, an alternative computationally feasible distance measure is presented in this paper. The distance measure used in this paper is based on Euclidean Distance Transforms (EDT) (Maurer et al., 2003). Generally, numerous mathematical method for comparing images exist in the computer vision literature, *e.g.* Image Eclidean Distance (IMED) (Liwei Wang et al., 2005; Xiaofeng and Wei, 2008) and Scale Invariant Feature Transform (SIFT) (Lowe, 2004). However, these alternative distance measures are, to our knowledge, an unexplored research avenue within comparison and uncertainty analysis of ensembles of 3D hydrostratigraphic models.

2 The Kasted study area

The Kasted area is located northwest of Aarhus, Denmark (Figure 1A), and has also been presented by Barfod *et al.* (2018), Marker *et al.* (2017), and Høyer *et al.* (2015). The regional geology of the Kasted area is dominated by a Quaternary buried valley complex with complex abutting relationships between the individual valleys. The buried valleys are infilled with a combination of till and glacial meltwater deposits. The valleys are incised into the substratum, which consists of hemipelagic clay. The regional geology has been described in detail by Høyer *et al.* (2015), who created a detailed cognitive geological model of the area.

An important geological feature of the Danish subsurface are buried tunnel valleys (e.g. Jørgensen and Sandersen, 2006; Sandersen et al., 2009). The geological heterogeneity varies considerably across Denmark, and can in some places be quite complicated, such as in the Egebjerg area (Jørgensen et al., 2010). In the Kasted area, the main tunnel valleys are clearly outlined in the geophysical dataset thanks to the significant resistivity contrasts between the infill of the buried valleys and the underlying Paleogene clay (Høyer et al., 2015). Therefore, the Kasted survey is ideal for studying the uncertainty related to stochastic hydrostratigraphic modeling using MPS methods.

The survey covers an area of 45 km^2, and is composed of a spatially dense SkyTEM survey with a total of 333 line km (Sørensen and Auken, 2004), with a line spacing of 100 m. The resulting SkyTEM soundings have been processed according to the description by Auken *et al.* (2009). Finally two sets of geophysical models were produced using either the smooth Spatially Constrained Inversion (SCI) models (Constable et al., 1987; Viezzoli et al., 2008), or the sharp SCI (sSCI) models (Vignoli et al., 2015). Furthermore, there are 948 boreholes scattered throughout the Kasted survey area, each with a corresponding lithology log of a varying quality. The quality assessment presented by He *et al.* (2014) and Barfod *et al.* (2016)

is used to divide the boreholes into quality groups, ranking between 1-5. Only 410 of the boreholes are above the selected
165 quality threshold, *i.e.* within quality groups 1-3, and contain lithological information relevant to this study. An overview of the
dataset is found in Figure 1C and described in further details in Barfod *et al.* (2018) and Høyer *et al.* (2015).
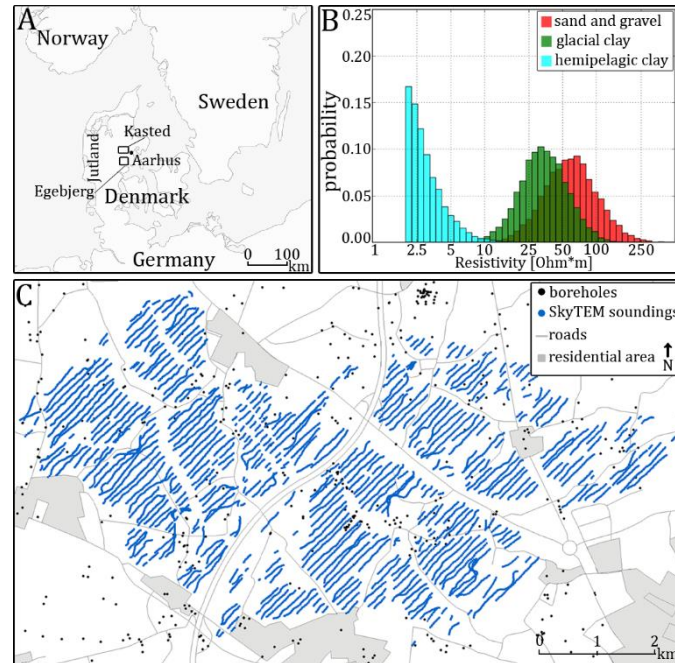

Figure 1: *The Kasted survey area and resistivity-hydrostratigraphic relationship histograms. **A** shows the geographical location of the
Kasted survey area and the Egebjerg model used as a secondary TI. **B** shows the reconstructed resistivity-hydrostratigraphic relationship*
170 *histograms for the three main hydrostratigraphic unit categories based on SkyTEM resistivity models and borehole lithology logs. **C** shows
the Kasted survey with the SkyTEM sounding and borehole locations.*

3 Methods

3.1 Multiple-Point Statistics (MPS) and single normal equation simulation (*snesim*)

The Multiple-Point Statistics (MPS) framework stems from the general geostatistics framework. Here, Multiple-Point (MP)
175 information from a Training Image (TI) is used to condition simulations to probable geological patterns (Journel and Zhang,
2007). The TI thus provides a conceptual geological understanding of a given area, and can be viewed as a database containing
probable geological patterns, which are used to condition the MPS simulation. The choice of TI is an important step in any
MPS setup, and influences the realization results, as will be illustrated. The TI does not need to carry locally accurate
information, *i.e.* the TI does not need to spatially or geographically overlap with real-world geological units, and can be purely
180 conceptual in nature. Together with the TI, it is also possible to use geophysical datasets for constraining MPS simulations,
resulting in realizations that reflect real-world regional geology. Today, MPS is a widely used tool, which is used in a variety
of geoscience fields, including, but not limited to: reservoir modeling (*e.g.* Okabe and Blunt, 2004; Strebelle and Journel,
2001), hydrology (e.g. Le Coz et al., 2011; Hermans et al., 2015; Høyer et al., 2017), geological modeling (e.g. de Iaco and
Maggio, 2011).

185 The MPS method used in this paper is known as the single normal equation simulation (*snesim*) framework (Strebelle, 2002),
and is implemented in the Stanford Geostatistical Modeling Software, or SGeMS. The *snesim* framework allows for simulating
real-world categorical geological model using a TI, constrained using soft geophysical data, and hard borehole data. The *snesim*
algorithm scans the entire *TI*, ahead of simulation, and stores the MP information contained in the TI in a search-tree database.
The MP information can then be retrieved from the database during simulation. The integration of soft geophysical data for
190 constraining the simulations is achieved by utilizing the tau model, which will be described in detail in section 3.1.1 (Journel,
2002; Krishnan, 2004). Here, the continuous soft data variable needs to be translated into a probability grid, describing the

probability of finding given geological unit based on the geophysical data; see Barfod *et al.* (2018). In order to guarantee the reproduction of geological patterns at all scales, *snesim* uses the multiple grid formulation, presented by Tran (1994).

3.1.1 The Tau model: combining conditional probabilities

195 Combining information from different sources is a frequent challenge in subsurface modeling. A fundamental challenge of the research conducted in this paper was to combine conditional probabilities from different sources. In this paper we used the common Tau model approach (Journel, 2002). The Tau model generally combines the probability values from different sources using Bayes' theorem and a set of τ-values, or τ-weights, for determining how to weight the probabilities. The choice of τ-weights is subjective, and assigning these is not a trivial task. It is recommended to run a series of exhaustive tests when

200 assigning the τ-weights.

We will now briefly introduce the Tau model; for more detail see Journel (2002). Suppose we have a set of data events, $D_i, i = 1, ..., n$, and the goal is to estimate the probability that a hydrostratigraphic unit (A) is present provided all data events:

$$P(A|\boldsymbol{D}_1, ..., \boldsymbol{D}_n) \tag{1}$$

The first step is then to define the prior probability distribution, $P(A)$. Generally the Tau model can be applied to as many

205 different probability grids as desired, but for the purpose of simplification two probability distributions are defined: $P(A|\boldsymbol{D}_1)$ and $P(A|\boldsymbol{D}_2)$. In this study we will consider $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ as 2D or 3D probability grids from different sources. As an example $\boldsymbol{D}_1$ could be a probability grid from geophysical data and $\boldsymbol{D}_2$ a probability grid from borehole lithology logs. The probability grids are translated into distance grids by applying the "*probability-into-distance*" transform:

$$x_o = \frac{1-P(A)}{P(A)}, x_1 = \frac{1-P(A|\boldsymbol{D}_1)}{P(A|\boldsymbol{D}_1)}, \text{ and } x_2 = \frac{1-P(A|\boldsymbol{D}_2)}{P(A|\boldsymbol{D}_2)} \tag{2}$$

210 Then the following distance ratio is computed using the tau model expression:

$$\frac{x}{x_0} = \prod_{i=1}^{n=2} \left(\frac{x_i}{x_0}\right)^{\tau_i}, \tau \epsilon[-\infty; +\infty] \tag{3}$$

where the tau values are assigned as follows: $[\tau_1, \tau_2]$. The final conditional probability is computed as follows:

$$P(A|\boldsymbol{D}_1, \boldsymbol{D}_2) = \frac{1}{1+x} \tag{4}$$

where the value of x is computed from eq. (2), as follows:

215 $$x = x_0 \cdot \left(\frac{x_1}{x_0}\right)^{\tau_1} \cdot \left(\frac{x_2}{x_0}\right)^{\tau_2} \tag{5}$$

3.2 Reconstructing incomplete datasets using Direct Sampling

In the field of geoscience, we are always dealing with incomplete datasets, since we cannot sample the subsurface exhaustively. Several approaches exist for dealing with incomplete datasets, of which two general approaches can be defined. A common approach is to reconstruct incomplete datasets using geostatistical tools (*e.g.* Goovaerts, 1997; Mariethoz and Renard, 2010),

220 which means that during the hydrostratigraphic modeling process no information is present in the dataset gaps. However, it is important to emphasize that the reconstructed information is not as valuable as the actual measured geophysical information. The other common approach is to just use the incomplete dataset as is. This means that no information is present in the dataset gaps during the hydrostratigraphic modeling process, which, depending on the modeling method, might result in large uncertainties.

225 In this study the MPS method called Direct Sampling (DS) is used for stochastic reconstruction of incomplete datasets (Mariethoz and Renard, 2010). The DS method uses the dataset we wish to reconstruct both as a simulation grid and a TI. This means that the patterns that are present in the incomplete dataset are inserted into the simulation grid before reconstruction. It is, according to Mariethoz and Renard (2010), important that the patterns we wish to reconstruct are actually present in the

incomplete dataset, since we are borrowing the patterns from the TI, or incomplete dataset, to stochastically reconstruct the
230 dataset. If the patterns are not present in the incomplete dataset they will, simply put, not be inferred in the reconstructed
dataset. Provided enough information on the overall patterns is available in the incomplete dataset, the DS method is a
straightforward approach for reconstructing incomplete datasets.

3.3 Comparing Simulation results

Comparing a large set of extensive 3D models is a common problem encountered in stochastic MPS modeling. A common
235 approach is visual comparison, which is not an objective or quantitative comparison method. Each equiprobable
hydrostratigraphic model in this study contains 1,187,823 cells. Furthermore, a total of 400 MPS realizations were computed,
Table 1, which makes it difficult to visually compare modeling results. This, along with advances in stochastic modeling tools
such as MPS, motivated Tan *et al.* (2014) to develop a framework in which multiple 2D or 3D realizations can be compared
quantitatively. The idea is to use a distance measure, which measures the distance between two realizations. Realizations which
240 are geometrically similar have small distance values, while dissimilar realizations have a large distance value. The comparison
techniques in this study are based on the principles presented by Tan *et al.* (2014). In this study the distances between individual
realizations are based on the Euclidean Distance Transforms (EDT) (Maurer et al., 2003). The usage of EDT as a measure for
similarity will be described in more detail below. A full distance matrix is computed containing distances between each
individual realization for all the different cases. The resulting 400 by 400 distance matrix is then interpreted by itself.

245 ### 3.3.1 Ensemble mode ratio maps (EMR-maps)

The visual comparison can be helped by creating so-called Ensemble mode ratio maps, or EMR-maps. The idea is to create a
summary map portraying the mode ratio of a given ensemble of models, ranging between 1/K and 1, where K is the number
of hydrostratigraphic categories. The EMR-maps describe the certainty of the simulation based on the resulting realization
ensemble. If the EMR-map shows a value of one, then every single realization in the present ensemble has simulated the same
250 category or, in this case, hydrostratigraphic unit. On the other hand if the EMR-map shows a ratio of 1/K the ensemble of
realizations shows equal probability for each of the K categories. Each realization is equiprobable, and the EMR values of the
categorical variables are computed from the probability distribution of a given cell with location, u. The probability that the
attribute S is equal to s_k, $P_k(u)$, is computed as follows:

$$P_k(\boldsymbol{u}) = \frac{1}{N}\sum_{i=1}^{N}\left(s_{k,i}(\boldsymbol{u}) = s_k\right) \tag{6}$$

255 Where N is the number of realizations, s_k is the state of attribute S for which we are currently computing the probability and
$s_{k,i}(\mathbf{u})$ is the state of the attribute at location $\mathbf{u}$ and for the *i*'th realization. The EMR values for a given cell, $\mathbf{u}$, can then be
computed as follows:

$$r_{EMR}(\boldsymbol{u}) = \max_{k=\{1,2,\dots,K\}}\left(P_k(\boldsymbol{u})\right) \tag{7}$$

where K is the number of categories for which the EMR value is computed, and $P_k(\mathbf{u})$ denotes the probability for category *k* at
260 location $\mathbf{u}$ computed using eq. (6).

The EMR values are then computed for each grid cell using eq. (6) and (7), which, simply put, is the occurrence ratio of the
mode category of a given ensemble containing a given number of realizations, N_{reals}. In other words, at a given location, $\mathbf{u}$, if
45 out of 50 realizations yield the same category, then the EMR-value is 0.9, and the ensemble certainty for the given cell is
high. On the other hand, with three possible lithological categories *i.e.* K=3, the lowest possible certainty is 1/K=1/3, which
265 means there is an equal probability of occurrence for each lithological category. This means that $P(s_1)=P(s_2)=P(s_3)=1/3$, and
therefore at the given location, $\mathbf{u}$, the $r_{EMR}=1/3$ and the simulation is uncertain.

8

3.3.2 Euclidean Distance Transforms (EDT) – measuring similarity between 3D hydrostratigraphic realizations

The hydrostratigraphic realizations are categorical and contain three hydrostratigraphic units. Comparing two realization grids, they first need to be transformed from a categorical grid into continuous Euclidean distance grids by using EDT (Maurer et al., 2003). The two 3D EDT grids are then compared by calculating the average difference in the respective grids. Similar images have a small average EDT distance, and dissimilar images have a large average EDT distance. The EDT computes the Euclidean Distances for all locations of a binary grid, i.e. a grid containing only two states (codes 0 and 1). The EDT map is simply the Euclidean distance of the medium depicted by the state code 1, i.e. for grid cell at location **u**:

$$d_{EDT}(\boldsymbol{u}) = \min_{v \in V}(\|\boldsymbol{u} - \boldsymbol{v}\|_2) \tag{8}$$

Where V is a set of grid cells with a state code equal to 1.

The d_{EDT} implementation presented by Maurer *et al.* (2003), uses a computationally favorable method for computing the exhaustive EDT at all locations in a binary grid.

To illustrate the d_{EDT} approach for comparing realizations a 2D example case is presented. The basic modeling setup contains 50 realizations, i.e. $N_{realizations} = 50$, which are going to be compared to the cognitive model, which in this case also happens to be the TI. The 2D example is created by selecting the horizontal cross-section at 20 mbsl, for each of the 50 basic modeling setup realizations and the single cognitive geological model (Figure 2A-D). Each of the 2D layers are transformed into 2D binary layers, portraying *sand and gravel* as the main variable, and *glacial clay* and *hemipelagic clay* as a background variable (Figure 2E-H). The 2D binary layers are then translated into 2D d_{EDT}-layers by using eq. (8) to exhaustively compute the d_{EDT} at each grid cell for all of the 50 realizations. The resulting d_{EDT} layers, of which three are seen in Figure 2I-L, are used to compute an average Euclidean Distance between each realization, $m_{r,i}$, and the cognitive geological model, m_{cog}:

$$\Delta d_{EDT}(m_{cog}, m_{r,i}) = \frac{1}{M}\sum_{j=1}^{M}\left[d_{EDT}^{m_{cog}}(u_j) - d_{EDT}^{m_{r,i}}(u_j)\right] \tag{9}$$

where $i \in \{1, ..., N\}$, N being the number of realizations, which in this case is N=50, and M bessing the number of cells in the simulation grid, or in this case, the 2D layer. The Δd_{EDT}, eq. (9), then describes the average difference of the distance to the nearest active cell in the binary grid. The 50 realizations are then ranked by the average Euclidean Distance differences, Δd_{EDT}, as seen in Figure 2, where the realization which is closest to the cognitive geological model (Figure 2B, F and J) has a Δd_{EDT}-value of 240 m, while the realization, which was ranked 25th closest (Figure 2C, G and K) has an Δd_{EDT}-value of 280 m, and lastly the realization, which was farthest (Figure 2D, H and L) has a Δd_{EDT}-value of 310 m. It should be noted that the Δd_{EDT} computation, described by eq. (9), is not limited to comparing a realization to a cognitive model, and can in fact be used to compare any pair of 3D categorical model. In fact, a generalized version of eq. (9) can be defined as follows:

$$\Delta d_{EDT}(m_A, m_B) = \frac{1}{M}\sum_{j=1}^{M}\left[d_{EDT}^{m_A}(u_j) - d_{EDT}^{m_B}(u_j)\right] \tag{10}$$

Where the number of cells in model-A, m_A, must be equal to the number of cells in model-B, m_B, i.e. $M_{m_A} = M_{m_B} = M$.
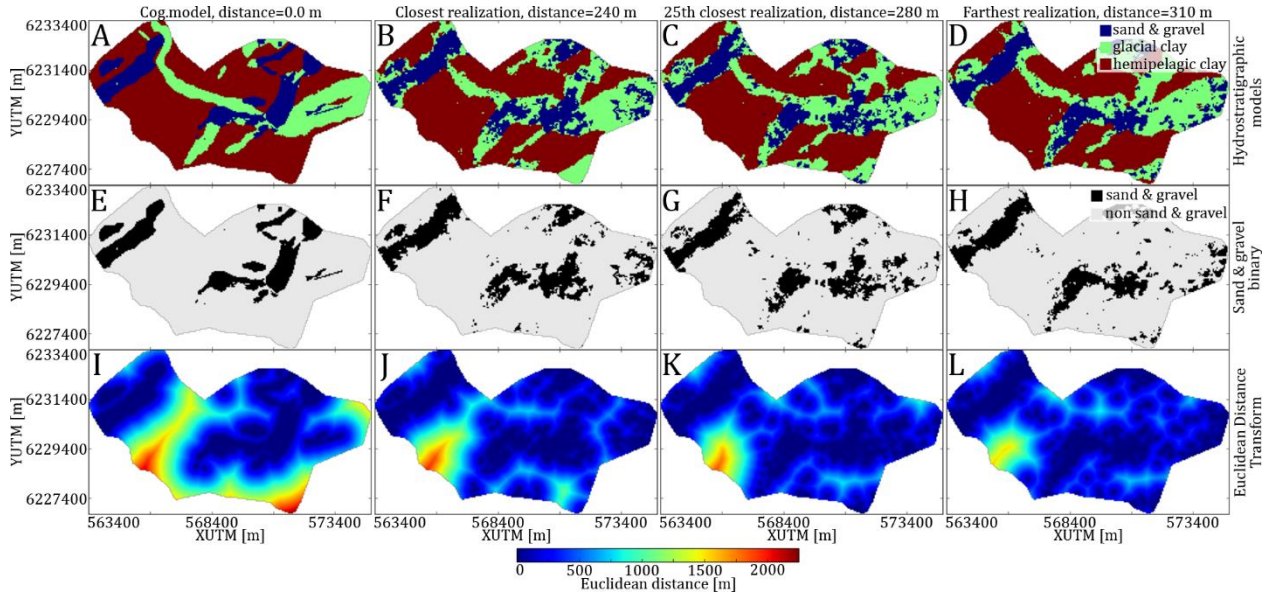
Figure 2: *A 2D example of the Euclidean Distance Transforms (EDT) as a measure for the similarity between categorical MPS realizations. In this example a set of 50 realizations, from the basic modeling setup, are compared based on the differences in EDT for sand & gravel units. **A-D** shows the hydrostratigraphic models for the TI, closest realization, 25th closest realizations, and farthest realization, respectively. **E-H** shows, in the same order as above, the binary images of the sand & gravel units of the 2D hydrostratigraphic model layers. **I-L** shows, in the same order as above, the Euclidean Distances layers computed from the 2D sand & gravel binary layers.*

From this point forward we leave the 2D example behind, and will from here on only consider Δd_{EDT} computations on 3D hydrostratigraphic grids. Furthermore, the Δd_{EDT} computations are carried out on a set of three binary grids, one for each of the three hydrostratigraphic categories. The distance value between two hydrostratigraphic grids is the summed distance for each of the three hydrostratigraphic categories, ensuring that the distance values reflect the complexities related to each of the hydrostratigraphic categories.

3.3.3 Evaluating the distance matrix

The average Euclidean Distance difference, Δd_{EDT}, from here on referred to as the "distance" between two realizations, is exhaustively computed between all realizations and compiled into an exhaustive 400 by 400 distance matrix. The distance matrix, **D**, contains all distance values between all hydrostratigraphic realizations computed using eq. (9) and is defined as follows:

$$D_{i,j} = d_{EDT}(m_{r,i}, m_{r,j}) \tag{11}$$

where $i, j = \{1, ..., N\}$, where N is the number of realizations. The distance matrix, **D**, can be evaluated directly by comparing the distances between individual realizations to each other. Another option is to summarize the distance matrix in a table representing the distances between the different cases. This is achieved by organizing the distance matrix according to which case they belong to. In this study the distance matrix is sorted according to the order of the individual cases, as in Table 1. The distance matrix can then be summarized, by computing the average distance for each group of realizations pertaining to a specific case. The concepts of distance variability and distance to cognitive model were presented by Barfod *et al.* (2018), and are also used here. The concept is that the variability pertaining to a specific case can be computed by computing the average of the distances of the 50 realizations for a given case ensemble. Another measure is the distance to the cognitive model. The distances between all realizations and the cognitive model are computed, and provides a reference point to which the realizations are compared.

10

4 *MPS* modeling setup

325   The Kasted dataset, used in this study, is comprised of a dense geophysical dataset acquired using the SkyTEM system (Sørensen and Auken, 2004), borehole lithology logs and a cognitive geological model (Høyer et al., 2015). The MPS modeling setup is similar to the one presented by Barfod *et al.* (2018). However, the goal of this study is different. It is divided into a total of 6 cases, or 8 sub-cases, which are designed to study how perturbations of the underlying MPS setup affects the hydrostratigraphic realizations using *snesim* and study the propagation of uncertainties into the hydrostratigraphic models.

330   First the basic case (case 0), from which the other cases are perturbed, uses a hydrostatigraphic simplification of the 3D cognitive model of the Kasted area as a TI, the gap-filled SkyTEM data present as smooth inversion models as soft data, and the borehole lithological logs as hard data. Then, in case 1, the TI is substituted by two other TIs. The cases 2 and 3 are related to the SkyTEM data, where the incomplete SkyTEM data and sSCI inversion models are used, respectively. The influence of the boreholes are studied in case 4 either leaving out boreholes as hard data or changing them into soft data. Finally, in case 5

335   the SkyTEM data are not used. Further details on each case follow in the coming sections and Table 1 summarizes each case. First of all, the model discretization and parameterization as well as construction of hard and soft data grids are described.

The Kasted model covers an area of 12 km by 7 km, discretized on a modeling grid with 229 by 133 by 39 cells, containing a total of 1,187,823 cells. Each cell has a size of 50 m by 50 m by 5 m. It is parameterized into three hydrostratigraphic units:

    1) *sand and gravel*: a combination of coarse lithological units, including sand till, meltwater sand, gravel and pebbles of

340   glacial origin, late glacial freshwater sand and postglacial freshwater sand

    2) *glacial clay*: this category contains silty and sandy clays, including clay till and meltwater clay of glacial origin

    3) *hemipelagic clay*: a combination of fine grained conductive clays, containing the extensive and homogeneous hemipelagic Paleogene and Oligocene clays found in Denmark.

These three categories serve the purpose of simplifying the geology of the Kasted area. The Kasted survey lithology logs

345   reveal a combination of 59 geological categories, which are translated into a set of hydrostratigraphic logs ("step 1", Figure 3) using these hydrostratigraphic categories. Similarly, the 42 geological units in the cognitive geological model are divided into the three abovementioned categories (Figure 4A). The vertical proportions of the three category hydrostratigraphic Kasted model can be viewed in Figure 5A.
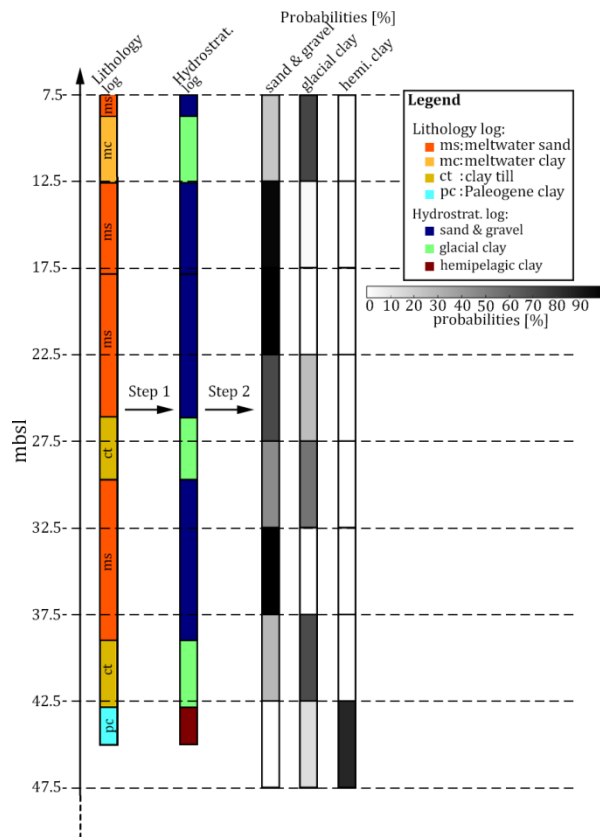
Figure 3: *A schematic diagram presenting the conversion of the lithological logs into probability logs for the three hydrostratigraphic units: sand & gravel, glacial clay and hemipelagic clay. Step 1: the lithology log is translated into a hydrostratigraphic log. Step 2: The hydrostratigraphic logs are resampled according to the vertical modeling grid intervals and an interval probability is calculated for each of the hydrostratigraphic units.*

The borehole lithology logs need to be assigned to a 3D grid, which is carried out in three overall steps. The first step is to translate the borehole lithology logs into hydrostratigraphic logs using the above mentioned three categories; "Step 1" Figure 3. The second step is then to divide the hydrostratigraphic logs into intervals identical to the vertical intervals of the model grid. At each resampled interval a probability value is directly calculated for each hydrostratigraphic unit; "Step 2" in Figure 3. The probability is simply the percentage of the given unit, which is present within the interval. Finally, the last step is to assign the hydrostratigraphic probabilities to a grid. The probability values are assigned to the grid cell in which the given hydrostratigraphic log is present. On the rare occasion that multiple logs are present within a given cell, the probabilities are combined accordingly to one representative probability value. The end result is a grid containing the borehole probability values of each hydrostratigraphic unit: *sand & gravel*, *glacial clay* and *hemipelagic clay*. It is common to view borehole lithology logs as hard information, or "ground truth". The borehole probability grid can therefore be translated into a hard data grid, by assigning the most probable hydrostratigraphic unit in each grid cell.

The 1D SkyTEM resistivity models are assigned to a 3D grid, identical to the modeling grid. The first step is to fill all grid cells containing a resistivity model. This is carried out using block Kriging and results is an incomplete resistivity grid of block average resistivities (Figure 6A). The second and final step is to stochastically reconstruct the incomplete resistivity grid using DS stochastic reconstruction (Mariethoz and Renard, 2010) (Figure 7). The reconstruction procedure was originally presented by Barfod et al. (2018), however, we have made some improvements for this study. Originally, a simple Kriging estimation approach was used to assign the resistivity models to a 3D modeling grid. This resulted in an incomplete resistivity grid, which contained resistivity information not only pertaining to grid cells containing a SkyTEM sounding, *i.e.* the resistivity grid had already been partly reconstructed in the proximity of the geophysical soundings. To avoid this, block Kriging estimation was used instead. The block Kriging method is also a variogram based estimation method, which estimates the average value of a rectangular block (Goovaerts, 1997). For more details on reconstructing incomplete resistivity grids see Barfod *et al.* (2018).

12

375 The SGeMS *snesim* framework utilizes the tau model for soft data conditioning (Journel, 2002), which requires the translation of resistivity grids into probability grids. This requires information on the regional resistivity-hydrostratigraphic relationship. Such knowledge is not always available, but if enough boreholes and electromagnetic geophysical data are available, the framework for studying the resistivity-hydrostratigraphic relationship, presented by Barfod *et al.* (2016), can be used to create a set of histograms. The resistivity-hydrostratigraphic histograms, Figure 1B, are compiled from available hydrostratigraphic

380 logs and SkyTEM resistivity models, and are presented in more detail in Barfod *et al.* (2016) and (2018). The estimated histograms (Figure 1B) are then used to directly translate each resistivity value, in a given resistivity grid, into three probabilities, one for each hydrostratigraphic unit.

The general MPS workflow can be summarized in 7 overall steps as follows:

1) Using block Kriging, the SkyTEM resistivity models are assigned to a 3D grid identical to the Kasted model grid.

385 2) The incomplete resistivity grids (Figure 6A) are stochastically reconstructed using Direct Sampling (DS) (Figure 7A). The result is an ensemble of 50 equiprobable reconstructed resistivity grids.

3) The reconstructed resistivity grids are translated into probability grids using the resistivity-hydrostratigraphic relationship histograms (Figure 1B).

4) The borehole lithology logs are translated into hydrostratigraphic logs; "Step 1*"* Figure 3.

390 5) The hydrostratigraphic logs are resampled and three probability values, one for each hydrostratigraphic unit, is directly computed at each resampled interval; "Step 2" Figure 3.

6) The borehole probabilities are assigned to a grid identical to the cognitive Kasted model grid.

7) The borehole probability grid is translated into a hard data grid, by assigning the most likely hydrostratigraphic unit to each grid cell.

395 A total of 400 realizations are created, with 50 realizations per sub-case (Table 1). In *snesim* a random number seed needs to be manually selected for each realization to initialize the random number generator and in particular define a random path through the modeling grid. The random seed convention chosen in this paper was to apply the same random seed vector to each sub-case. The vector contains 50 linearly increasing random seed numbers, ensuring consistency when comparing realizations from the individual sub-cases.

400 **Table 1.** *An overview table showing information on the MPS cases along with information on number of realizations for each case / sub-case, and a brief description of each case.*

Case name	Sub-case names	Num. realizations	Total num. realizations	Case description
Basic setup	Basic modeling setup	50	50	*The basic setup uses boreholes as hard data, smooth resistivity models as soft data, and the cognitive Kasted model as a TI*
Case 1a	a) Egebjerg TI	50	100	*Two different TIs are used to study the uncertainty related to the choice of TI, which reflects the conceptual geological understanding*
Case1b	b) Conceptual TI	50		
Case 2	Incomplete soft data grid	50	50	*The uncertainty related to the reconstruction of the resistivity grid is studied by running simulations with an incomplete resistivity grid*
Case 3	Sharp resistivity models	50	50	*The sharp resistivity models are used for simulations instead of the smooth models, to study how the choice of resistivity model influences the hydrostratigraphic models*
Case 4a	a) No borehole data	50	100	*Simulations are run without hard data, to see how much the hard data influences the results*

Case 4b	b) Soft borehole data	50		*The borehole data are used as soft information instead of hard by combining the borehole probability grid with the SkyTEM probability grid using the Tau model*
Case 5	No soft resistivity data	50	50	*Simulations are run using only the hard data and the cognitive Kasted TI*
Total	---	---	400	---

4.1 Basic modeling setup

The basic modeling setup is designed to act as the base from which all other cases are built. The different sub-cases are simply modified versions of the basic modeling setup, each designed to study how modification to the base setup relates to hydrostratigraphic MPS modeling. The basic modeling setup uses the borehole data as hard information, SCI models with smooth inversion constraints as soft data, and the cognitive hydrostratigraphic Kasted model as a TI (Figure 4A), for which the global proportions are listed in Table 2 and vertical proportions are displayed in Figure 5A.

4.2 Case 1 – Conceptual geological understanding

The basic modeling setup uses the actual cognitive geological model of the Kasted survey area as a TI (Høyer et al., 2015). In Denmark, it is common practice to build 3D cognitive geological models of the near-subsurface. Many cognitive models exist and are publicly available. Such models can easily be adapted and used as 3D TIs to simulate new survey areas, provided the geological settings are similar. Case 1 is divided into two sub-cases. The first sub-case, Case1a, uses the basic setup, but in place of the cognitive Kasted model, the cognitive geological model of the Egebjerg area (Figure 1A) is used as a TI (Figure 4C). The geologic setting in Egebjerg is relevant since it is partly dominated by a buried valley complex (Jørgensen et al., 2010).
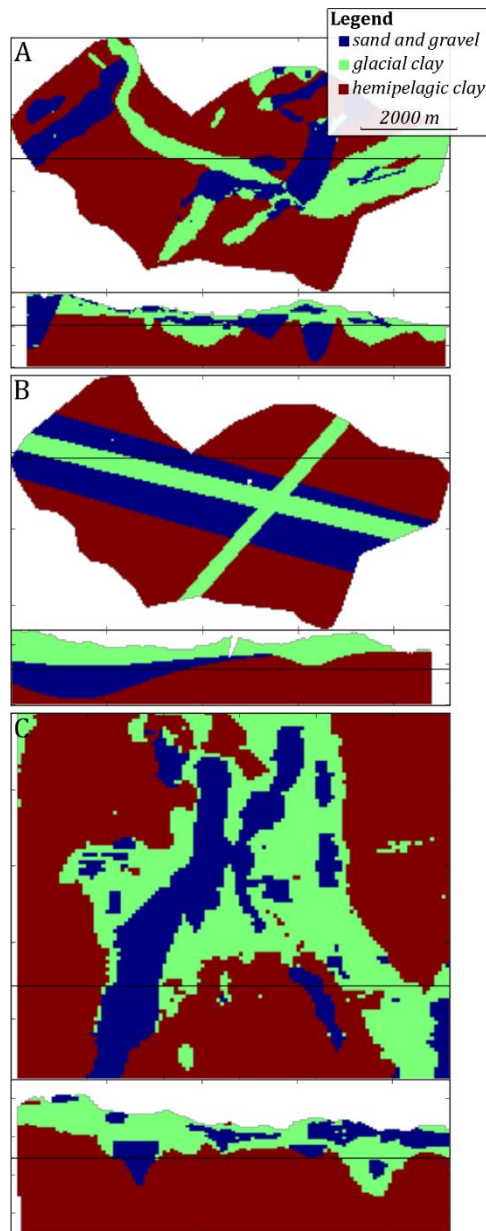
14

Figure 4: *An overview of the training images (TIs) which are used during MPS simulation. A horizontal slice and vertical cross-section is presented for each TI, portraying the hydrostratigraphic architecture; **A** shows the Kasted TI, **B** shows the conceptual TI, and **C** shows the Egebjerg TI which is generally larger than the Kasted model.*

420   The Egebjerg model consists of a total of 72 geological units which are categorized accordingly to reflect the three hydrostratigraphic units of the Kasted hydrostratigraphic model. Egebjerg additionally contains undesired features, such as local Miocene complexes. Two such local geological environments, which do not reflect the geological setting of the Kasted area, are present. One is found south of the buried valley complex, and the other to the west. By cropping the model and rotating it 90 degrees counter-clockwise, a relevant TI without undesired geological architecture is produced (Figure 4C); this

425   is referred to as Case 1a. It is clearly seen, by comparing Figure 4A and C that the Kasted and Egebjerg TIs are different. The Kasted TI is smaller, and contains smooth geological features, while the Egebjerg model is larger and contains coarse, block-like geological features. The important features, in relation to hydrostratigraphic modeling, are the buried valley complexes, which are present in the Egebjerg model (Figure 4C). The global proportions of the Egebjerg TI (Table 2) are similar to the ones found in the Kasted TI. However, the vertical proportions of the Egebjerg TI (Figure 5C) are different, especially in the

430   upper part of the TI where *glacial clay* units dominate.

	sand and gravel	glacial clay	hemipelagic clay
Kasted *TI*	0.17	0.21	0.62
Conceptual *TI*	0.17	0.22	0.61
Egebjerg *TI*	0.10	0.22	0.68

The second sub-case, Case1b, utilizes a purely conceptual TI. The conceptual TI is created by using a set of hyperbolic secant functions to populate a 3D matrix and is purely mathematical in nature. The conceptual TI can be seen in Figure 4B and is designed to have three overall buried valleys eroded into a *hemipelagic clay* substratum. There are two narrow and shallow *glacial clay* valleys, and a broad and deep *sand and gravel* valley. One of the glacial *clay valleys* is a younger valley, which is eroded into the older *sand and gravel* valley, and run roughly parallel to each other. The last *glacial clay* valley is almost orthogonal to the other valleys, and also erodes into the *sand and gravel* valley. The upper part of the TI contains a cover layer of glacial clay (Figure 5B). The simple conceptual TI is designed to contain the main geological architecture of the Kasted area, namely the buried valley complexes. The *sand and gravel* valley, trending west-northwest – east-southeast, was chosen on purpose to study what happens when over-simplified and smooth MP information is added to a TI. The global proportions of the conceptual TI are consistent with the other TIs, while the vertical proportions for *sand and gravel* and *glacial clay* units show a significantly different pattern (Figure 5B).
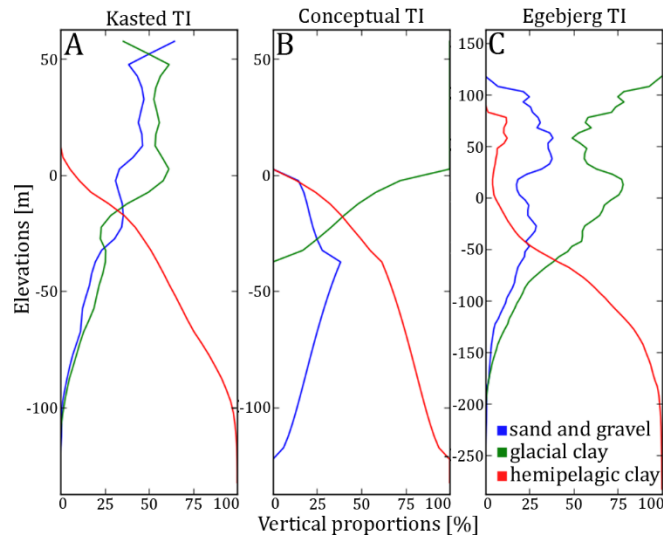

Figure 5. The vertical proportions of the three training images for each of the hydrostratigraphic categories, where **A** portrays the Kasted TI, **B** the conceptual TI, and **C** the Egebjerg TI.

4.3 Case 2 – Incomplete soft data

During reconstruction of the resistivity grid, it is assumed that the patterns in the incomplete data set contain information regarding the content of the data set gaps. This is true only when the incomplete grid contains a sufficient amount of data. Sufficient, in this case, means that the parameter space is sampled densely enough to reflect the patterns we wish to reconstruct (Mariethoz and Renard, 2010). If the grid is too sparse, then limited or no information is present which can help reconstruct missing patterns. Signs of mediocre data density are seen in the incomplete grids (Figure 6A). Artifacts from the DS reconstruction are present in the completed resistivity grids. The resistive valley to the west in the horizontal slices and vertical cross-sections in Figure 7A and B reveals a striated pattern. An alternative to reconstructing the resistivity grid beforehand is to use the incomplete resistivity grids for simulation, meaning no information is present in the resistivity dataset gaps. Grid cells containing a resistivity model are translated into three probability values using the resistivity-hydrostratigraphic relationship histograms (Figure 1B) (Figure 6B-D). Areas without soft resistivity data rely on the TI during simulation,

emphasizing the fact that no actual information is present between soundings. The overall setup is identical to the basic setup; the only difference is the reconstructed soft data grids are interchanged for the incomplete soft data grid (Figure 6B-D).
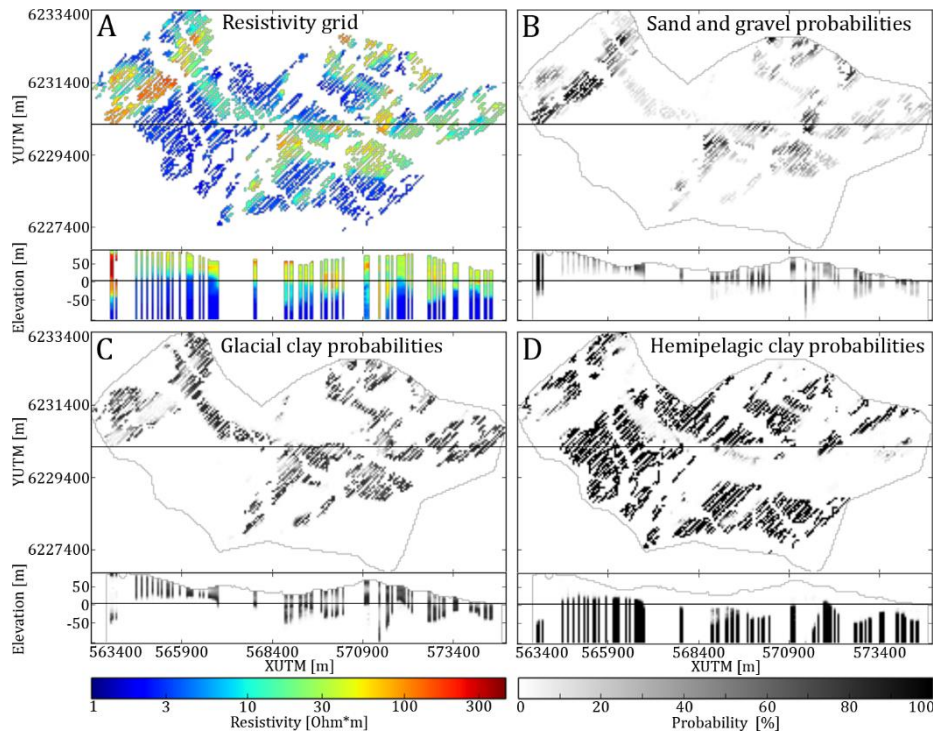
Figure 6: *A presentation of the incomplete resistivity grid. Each grid is portrayed as a horizontal slice at 20 mbsl, and a vertical cross-section intersecting at UTMY 6230100 m. **A** shows the resistivity grid which is translated into three probability grids using the resistivity-hydrostratigraphic relationship histograms (**Figure 1***B). Grid cells without SkyTEM soundings are not assigned a probability value. **B-D** show the sand and gravel, glacial clay, and hemipelagic clay probability values, respectively.*

4.4 Case 3 – Choice of resistivity model

The choice of inversion algorithm results in different SkyTEM resistivity models. The purpose of this case is to study how using sSCI (Vignoli et al., 2015) models influences the modeling results. A common inversion approach is SCI where a smooth regularization is used (Constable et al., 1987). Such resistivity models have a smooth transition from resistive to conductive features, and *vice versa*. Geological layer boundaries are rarely smooth in nature, meaning such soft transitions in resistivities seldom reflect reality. Furthermore, extreme resistivity values are not presented correctly in the smooth model inversions.

Vignoli *et al.* (2015) propose an alternative SCI approach, employing a "Minimum Gradient Support" regularization term instead. Such sSCI models produces resistivity models with sharp layer boundaries and a better representation of extreme values. The setup in Case 3 is identical to the basic setup, except that the SCI models are interchanged for sSCI models. The DS grid reconstruction is then conducted on the sSCI models, which are then translated into probability grids. Finally, these grids are used as soft data for simulation using the *snesim* method.

The sharp resistivity models are different from the smooth models, but no particularly sharp layer boundaries are reflected in the reconstructed resistivity grid (Figure 7B).
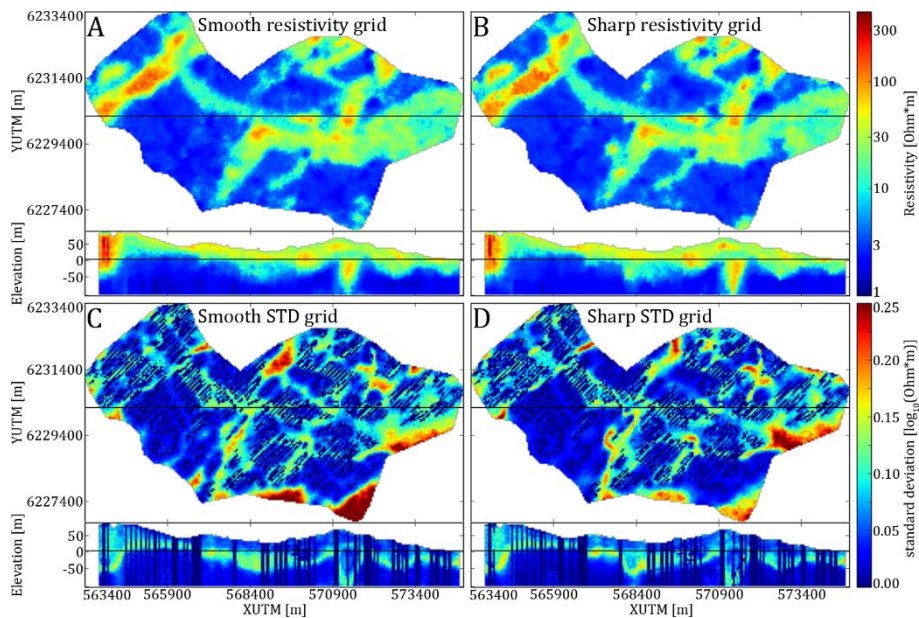
17

Figure 7: *An overview of the key differences between reconstructing the resistivity grid using smooth and sharp inversion resistivity models. Each grid is portrayed as a horizontal slice at 20 mbsl, and a vertical cross-section intersecting at UTMY 6230100 m. A shows the smooth reconstructed resistivity grid. B portrays the sharp reconstructed resistivity grid. C shows the standard deviation calculated from 50 stochastic reconstructions of the smooth resistivity grid. D shows the standard deviation calculated from 50 stochastic reconstructions of the sharp resistivity grid.*

One of the obvious differences is found in the resistivity patterns of the *sand & gravel* valley to the far west of the survey area. The valley itself is not significantly different, however, the small resistive patch, west of the large valley, is more pronounced in the sharp model and has an overall more pronounced fingerprint (Figure 7B). The sharp resistivity models better estimate the "true" bulk resistivity values of specific geological units, such as the resistive patch accentuated here. The ensemble standard deviation grid, Figure 7C and D, show a general reduction in the ambiguity of the reconstructed sharp resistivity models. This is clear from the reduction areas with large standard deviation, red colors, which are overall reduced in size.

4.5 Case 4 – Borehole lithology logs

This case is dedicated towards how the borehole data is handled, and how it influences the hydrostratigraphic modeling results. The hard borehole data is normally sparse, relative to geophysical data. Boreholes are commonly considered "ground truth" since they directly sample the subsurface sediments or petrological units. This case is divided into two sub-cases. The first sub-case, Case 4a, portrays what happens when hard data is not included in the *snesim* simulation. The model setup is therefore identical to the basic MPS setup, but without including the borehole data.

The second sub-case, Case 4b, incorporates the borehole lithology logs as soft data. The certainty of a lithological log varies depending on a range of factors, *e.g.* drilling method, the purpose of the borehole, and sampling frequency (*e.g.* Barfod et al., 2016; He et al., 2014). The hydrostratigraphic probability logs, introduced in the basic modeling setup ("Step 2" Figure 3), are utilized in place of the hard borehole grid. The boreholes are assigned a lateral footprint, so the information is not only found at the borehole locations. The borehole footprint is assigned by creating a grid where the borehole probability values have been estimated in a radius of 200 m around each borehole using simple Kriging with a search radius of 200 m and a mean of 1/K=1/3, where K is the number of unique hydrostratigraphic units (Figure 8D-F) (For additional information see section A1 in the Appendix). The tau model is then used to combine the SkyTEM (Figure 8A-C) and borehole (Figure 8D-F) probability grids (e.g. Journel, 2002; Krishnan, 2004; Remy et al., 2014). The first step is then to define the prior probability distribution, $P(A)$, which in this case are the vertical proportions taken from each layer of the cognitive Kasted TI (Figure 5A). Then the probability distributions are defined: $P(A|D_r)$ and $P(A|D_b)$, where D_r is the resistivity probability grid and D_b the borehole probability grid. The 3D probability grids are translated into distance grids by applying the "*probability-into-distance*" and

18

computing the distance ratio using equations (2) and (3), where the tau values were assigned based on a series of exhaustive tests. The final tau values were selected based on the criteria that the transitions in areas where both borehole and resistivity information is available should be as smooth as possible in the resulting combined probability grid, as seen in Figure 8. Based on the tests, the resulting tau values were: $[\tau_r, \tau_b] = [2.0, 1.0]$. The final conditional probability was computed using eq. (4) and resulted in the three hydrostratigraphic probability grids is seen in Figure 8G-I. The combined probability grids replace the smooth probability grids used in the basic setup.
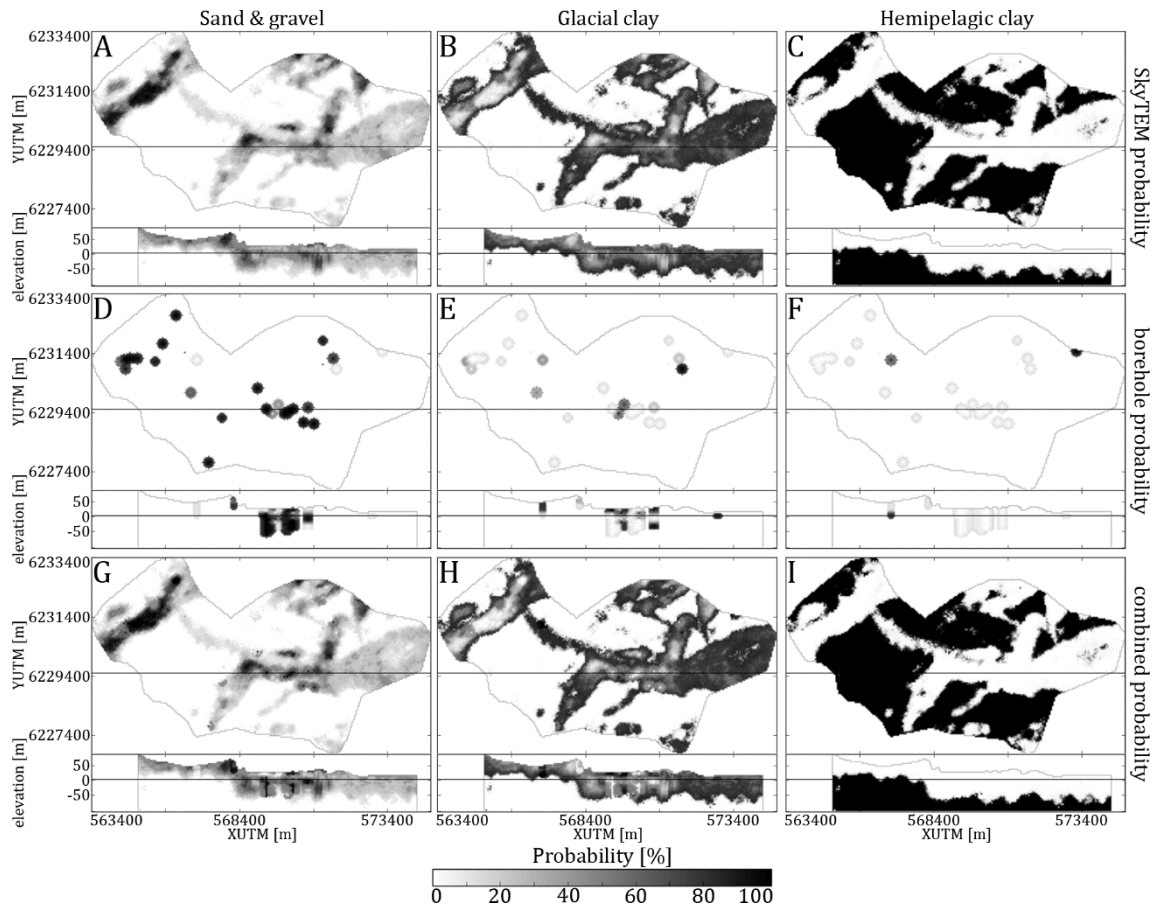

Figure 8: *A visual representation of the Tau model procedure for combining the soft resistivity and borehole grids. Each grid is portrayed as a horizontal slice at 20 mbsl, and a vertical cross-section intersecting at UTMY 6230100 m; **A**-**C** shows the sand and gravel, glacial clay, and hemipelagic clay probability maps, respectively, for one DS reconstructed resistivity grid, **D**-**F** shows the 200 m radius Kriged borehole probability, and **G**-**I** shows the combined resistivity grid, which has been combined using a tau model with the values:[τr, τb]=[2,1].*

4.6 Case 5 – Excluding the soft resistivity data

The final case, Case 5, illustrates the consequences of not including the soft SkyTEM resistivity information in the MPS simulation routine. The basic setup is simply run without the inclusion of soft data, *i.e.* the setup only uses the cognitive Kasted TI and hard borehole information.

5 Results

5.1 Visual comparison of hydrostratigraphic realizations and "Ensemble Mode Ratio"-maps (EMR-maps)

For each of the presented cases two hydrostratigraphic realizations are presented (Figure **9**), along with an EMR-map (Figure 10). The EMR-maps show the occurrence ratio of the most likely simulated category for each grid cell based on 50 realizations. The two realizations and EMR-map of the basic modeling setup, Figure **9**A and Figure 10A, reveal the same overall trends as the cognitive geological model, Figure 4A. Namely the western *sand and gravel* valley striking ~N40°E, the *glacial clay* valley striking ~E30°S, the large mixed *sand and gravel* and *glacial clay* valley striking ~N20°E to the south, and the small subsidiary

19

glacial clay valley striking ~N50°E to the south. However, even though the main hydrostratigraphic architecture of the

530 cognitive geological model is similar, there are still differences between the *snesim* realizations and the cognitive geological model. The cognitive model shows clear-cut, smooth, and ordered hydrostratigraphic units. The basic modeling setup realizations reveal sporadic and random patterns. The *sand and gravel* units are placed in small lumps throughout the *glacial clay* units, but are not present within the homogenous *hemipelagic clay*. Patches of uncertain g*lacial clay* units are, however, found in the homogeneous *hemipelagic clay*, especially in the southeast corner of the Kasted survey area (Figure **9**A) (Figure

535 10A). The same sporadic picture is seen in the vertical slices of the realizations (Figure **9**A), although, here an additional trend is revealed. The *sand and gravel* valley to the far west and at XUTM 570900 m, are not consistently filled with *sand and gravel* (Figure **9**A) as in the cognitive geological model (Figure 4A). Furthermore, the EMR-map reveals that the valley margins are subject to a larger degree of ambiguity (Figure 10A), in fact at some locations the r_{EMR}-value is close to 1/3, which means that for the model ensemble the occurrence of either hydrostratigraphic unit is possible.

540 The Case 1a realizations (Figure **9**B) (Figure 10B), which use the Egebjerg TI (Figure 4C), show the same overall trends as in the basic modeling setup. The subset of buried valleys mentioned above are present, however, an obvious difference is the coarse and block-like appearance of Case 1a realization ensemble. This appearance is similar to the block-like appearance of the Egebjerg *TI* (Figure 4C). Furthermore, the horizontal slice of the realizations and EMR-map reveals that the *glacial clay* dominated area to the east has a generally larger occurrence ratio, and is thus more certain. The realizations are clearly

545 influenced by the choice of TI, especially when Case1b is also considered (Figure **9**C). The hydrostratigraphic realizations of Case1b (Figure **9**C) (Figure 10C), clearly depict the same overall buried valley trends, but the valleys in the central part of the model are largely filled with the opposite of the valley filling hydrostratigraphic units. Furthermore, the occurrence ratio seems quite low in certain areas, such as to the south of the model, which means the ambiguity has increased. Finally, the realizations also reveal an absence of small-scale patterns, which corresponds to the conceptual TI, which only contains homogenous

550 hydrostratigraphic units.

The importance of reconstructing the incomplete resistivity grid is seen in the Case 2 (Figure **9**D) (Figure 10D). The two realizations in Figure **9**D show the main buried valley features, *e.g.* the western *sand and gravel* valley. However, the hydrostratigraphic units are sporadic, especially in areas with no data. Patches of *sand and gravel* and *glacial clay* are randomly spread throughout the presented horizontal slice and vertical cross-section (Figure **9**D). The EMR-map also reveals an increase

555 in low occurrence ratios in areas without soft data (Figure 10D).

The uncertainty related to the choice of geophysical modeling procedure is portrayed by Case 3. Here, *snesim* realizations are constrained to sharp resistivity models. Generally, the realizations (Figure **9**E) are quite similar to the basic modeling setup realizations (Figure **9**A). However, a key difference is the significant reduction or absence of patches of *glacial clay* in the homogeneous *hemipelagic clay*. In fact only one patch is found in the first realization (Figure **9**E) in the southwest corner,

560 while it is not present in the second realization, and the EMR-map further reveals a reduction of the occurrence ratios generally, especially along the southern margin of the realizations (Figure 10E).

Case 4 shows the influence that the hard data has on the hydrostratigraphic realizations in two sub-cases: Case 4a, where *snesim* simulations are run without hard data, and Case 4b, where the borehole data is treated as soft information. Figure **9**F and G shows two hydrostratigraphic realizations without hard data and with soft borehole data, respectively. These

565 realizations do not differ significantly from the basic modeling setup realizations and in fact are quite similar. One key difference is the central *glacial clay* valley striking ~E30°S, which does not contain any *sand and gravel* to the west (Figure **9**F and G). The EMR-maps reveal that without boreholes (Figure 10F) the occurrence ratios generally decrease, making the realizations more ambiguous. The usage of the borehole data as soft information also seems to reduce the occurrence ratios

compared to the basic modeling setup. Generally, leaving out the borehole data, or treating it as soft data, results in local
570 changes in areas with a high density of boreholes.

The final case, Case 5, illustrates the importance of the SkyTEM soft data. The *snesim* simulations are run using only hard
data and the cognitive geological model as a TI. The output realizations (Figure **9**H) portray smooth and large-scale
hydrostratigraphic units. The hydrostratigraphic architecture of the buried valleys is not reflected. However, the *sand and
gravel* valley, to the west, does seem to protrude slightly in the realizations (Figure **9**H) and EMR-map reveals a significant
575 decrease in the occurrence ratio, and thus an increase in the ambiguity of the model ensemble (Figure 10H).
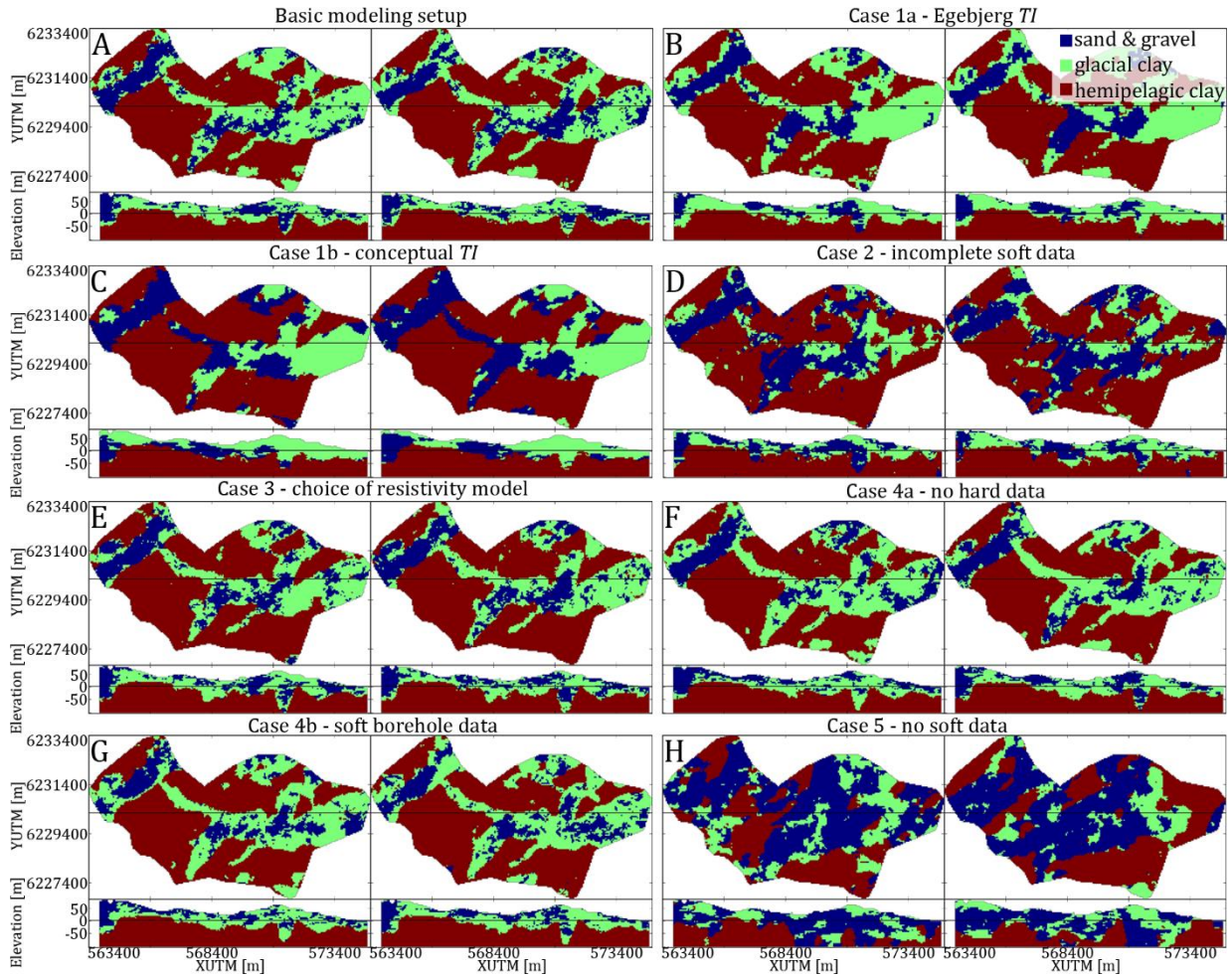
Figure 9: *Each case is displayed by two realizations, realization #1 of 50, and realization #30 of 50. Each realization is portrayed as a horizontal slice at 20 mbsl, and a vertical cross-section intersecting at UTMY 6230100 m. A shows the realization results for the Basic modeling setup, B shows the realization results for Case 1a, C shows the realization results for Case 1b, , D shows the results for Case 2, ,*
580 *E shows the results for Case 3, F shows the results for Case 4a, G shows the results for Case 4b, and H shows the results for Case 5. For more details on individual cases the reader is referred to* Table 1.
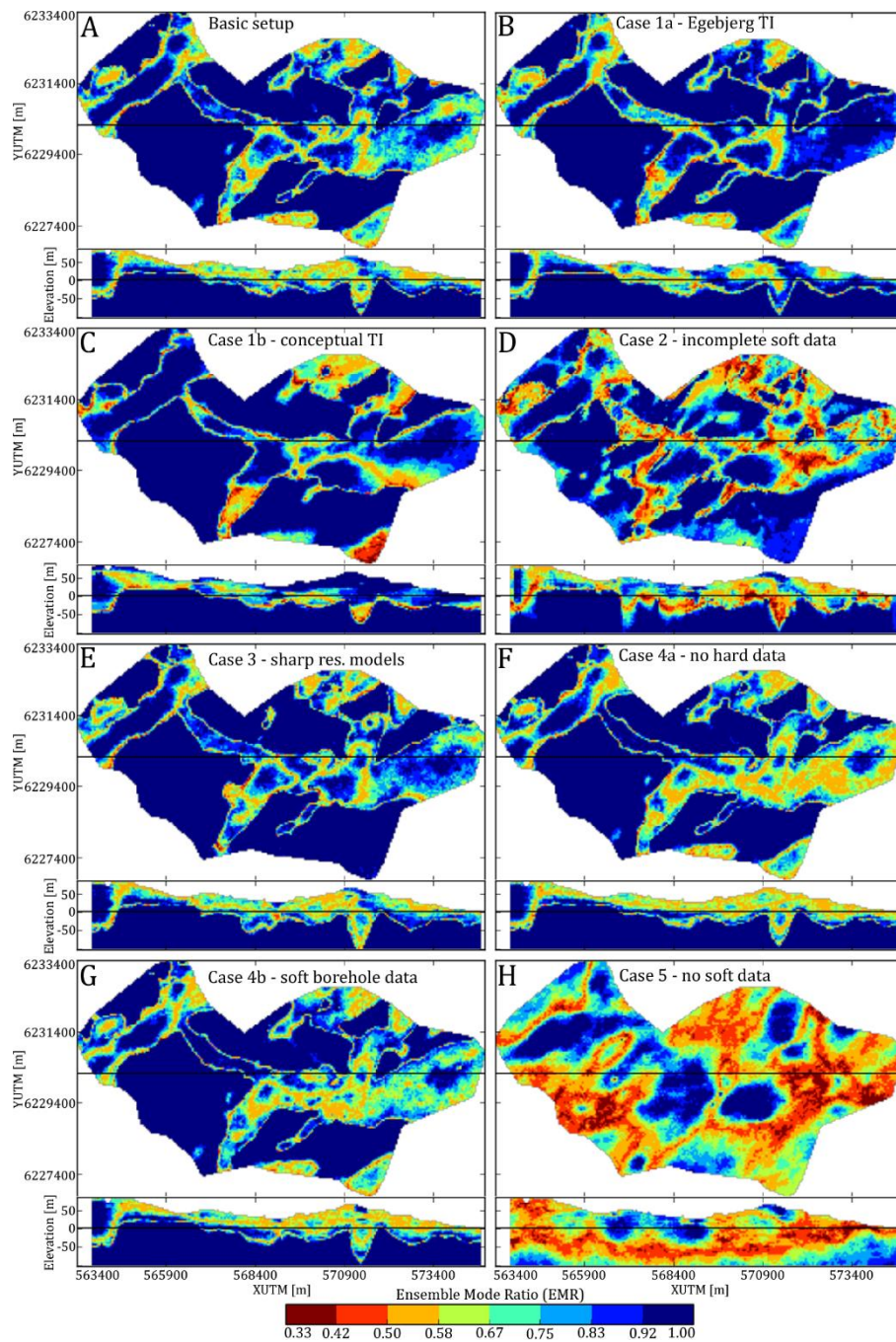
Figure 10: *A presentation of the "ensemble mode ratio" (EMR) maps, computed for the different case ensembles of hydrostratigraphic models. Each EMR map is presented as a horizontal slice centered on 20 mbsl, and a vertical cross-section intersecting at UTMY 6230150m; A-H presents the EMR-type uncertainty map for each of the different cases, which are summarized in* Table 1. *The EMR values portray how certain the ensemble of MPS realizations are, i.e. if $r_{EMR}=1/3$ then the realization is uncertain, and we have equal probability of finding either hydrostratigraphic unit since $P(s_1)=P(s_2)=P(s_3)=33\%$. On the other hand if $r_{EMR}=1$, then each realization of the given ensemble contains the same hydrostratigraphic unit at the given grid cell.*

5.2. Quantitative comparison using differences in object based Euclidean Distances as a measure for similarity

The distances between each of the 400 realizations have been computed using eq. (8) and (10). The full distance matrix is presented in Figure 11A. The distances between each realization and the cognitive geological models have also been computed and plotted in Figure 11B. To aid the interpretation of the distance matrix and distances to the cognitive model a summary table, Table 3, has been compiled.

The basic modeling setup constitutes a common *snesim* setup, with the geophysical data as soft data, boreholes as hard data, and a 3D geological conceptualization encased in a TI. The ensemble average variability is computed according to the equations presented by Barfod *et al.* (2018), and the resulting ensemble average variability is 10.1 m, with an average distance

22

to the cognitive model of 24.3 m. This means that the Euclidean distance mismatch between the individual realizations related to basic modeling setup is 10.1 m, and the average difference in Euclidean distance to the nearest active cell between the realizations and the cognitive model was 24.3 m.

The 3D geological conceptualization contained in the TI influences the final hydrostratigraphic realizations as illustrated in Case 1, which is divided into two sub-cases: Case 1a and Case 1b. In Case 1a, using a 3D cognitive geological model from the Egebjerg area as a TI for hydrostratigraphic simulation increases the average distance to the cognitive model to 24.9 m (Figure 11A) (Table 3). Furthermore, the average variability has increased to 13.6 m (Figure 11B) (Table 3). The other sub-case revolve around using an entirely conceptual geological model as a TI. The conceptual TI was designed to reflect the overall geology, yet still contains some bias. The results reflect the bias, with increased distances to the cognitive geological model, which are now centered on 25.6 m (Figure 11A) (Table 3). The ensemble variability has increased to 14.8 m (Figure 11B) (Table 3).

The importance of proper reconstruction of the incomplete resistivity grid is illustrated in Case 2, where the incomplete resistivity grid was used for simulation. The resulting realization ensemble have a large ensemble variability centered on 24.1 m (Figure 11A) (Table 3). The distance to the cognitive geological model is also large, with an average value of 33.1 m (Figure 11B) (Table 3).

In Case 3 the sharp SCI models were used for simulation in place of the smooth SCI models. The realizations related to Case 3 were the closest to the cognitive model with an average value of 21 m (Figure 11B) (Table 3). The variability of Case 3 realization ensemble, *i.e.* the distances between the realizations pertaining to Case 3, is small with an average value of 9.4 m – see Table 3. Recalling the raw hydrostratigraphic realizations (Figure **9**E) and the EMR-map (Figure 10E), the large reduction in distances could partly be related to the removal of non-*hemipelagic* clay units along the southern border of the model and an overall increase in confidence along the southern and southeastern border of the model.

The influence of the boreholes lithology logs on the hydrostratigraphic realizations is reflected in Case 4, which is divided into two sub-cases. In the first sub-case, Case 4a, the borehole information is not used as hard data, and the realizations are created only using soft geophysical data and the Kasted TI. However, the borehole data is still used for creating the resistivity-hydrostratigraphic histograms (Figure 1B), which are used for creating the probability grids. The ensemble average variability is 10.7 m (Figure 11A) (Table 3) and the average distance to the cognitive model is 24.3 m (Figure 11B) (Table 3). In the second sub-case, Case 4b, the boreholes are used as soft information to reflect the uncertainty of the borehole information. The ensemble average variability is 10.9 m, and the average distance to the cognitive model is 24.3 m. This illustrates how the *snesim* realizations are not particularly sensitive towards the sparse borehole hard data.

Not including the geophysical soft data in the *snesim* simulations, Case 5, resulted in the largest ensemble average variability of 40.0 m (Figure 11A) (Table 3). The average distances between Case 5 realizations and the cognitive model was 59.3 m. This means that the realizations of Case 5 are the most different from the rest of the realizations. The *snesim* realizations are sensitive towards not including the geophysical data, or using the incomplete resistivity grid. This underlines the importance of the geophysical soft data in relation to hydrostratigraphic modeling using the *snesim* methodology.
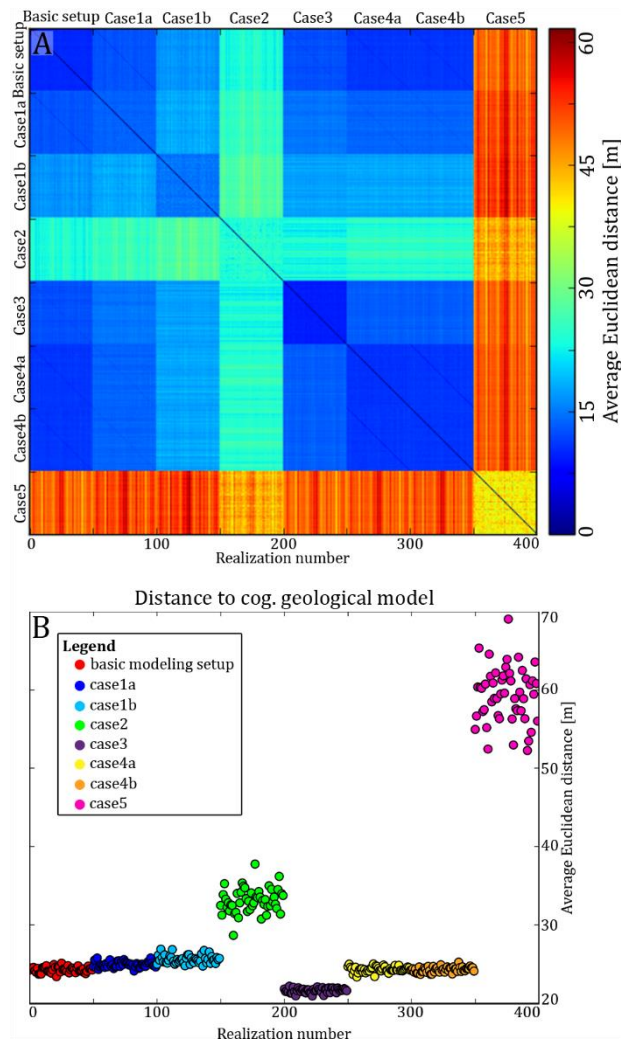
Figure 11: *A presentation of the average Euclidean distance calculations. **A** shows the full distance matrix, **B** shows the average Euclidean distances between each individual hydrostratigraphic realization and the cognitive geological model.*

Table 3: *A summary table showing the average distance value for each 50 by 50 square representing a given case in the distance matrix* (**Figure 11**A). *The final column, labelled "Distance_{cog}", summarizes the distances to the cognitive geological model, presented as the average of each colored point cloud in* **Figure 11**B. *The distances in parenthesis represent ensemble variabilities, and the remaining values represent average distances between different ensembles. The unit of the average distances is meters.*

Distance [m]	Basic setup	Case1a	Case1b	Case2	Case3	Case4a	Case4b	Case5	Distance_{cog}
Basic setup	(10.1)	12.9	16.9	24.0	12.7	11.1	11.2	49.6	24.3
Case1a	12.9	(13.6)	18.3	26.0	15.1	14.0	13.9	51.7	24.9
Case1b	16.9	18.3	(14.8)	27.9	17.8	18.1	18.1	52.5	25.6
Case2	24.0	26.0	27.9	(24.1)	23.5	25.0	24.9	45.2	33.1
Case3	12.7	15.1	17.8	23.5	(9.4)	13.6	13.6	49.6	21.6
Case4a	11.1	14.0	18.1	25.0	13.6	(10.7)	11.1	50.7	24.3
Case4b	11.2	13.9	18.1	24.9	13.6	11.1	(10.9)	50.6	24.3
Case5	49.6	51.7	52.5	45.2	49.6	50.7	50.6	(40.0)	59.3

6 Discussion

The cognitive geological model was created based on smooth SkyTEM resistivity models and lithological logs (Høyer et al., 2015) as well as the conceptual geological understanding of the area. The model was simplified from a full 3D geological model containing a total of 42 unique geological units, to a hydrostratigraphic model containing only 3 hydrostratigraphic units. The cognitive geological model, although detailed and extensive, is not the "true" geological model. The ensemble realizations should not directly reflect the cognitive model, yet the cognitive model can be thought of as a reference point in modeling space, which we would prefer our models to resemble.

The results revealed the importance of the SkyTEM dataset. Not including the resistivity models in the MPS simulations, Case 5, yielded realizations which were both the least similar to the cognitive geological model, and with the largest variability between the individual realizations. Including the incomplete resistivity grid, Case 2, improved the realization results compared to not including them at all. Yet, the ensemble variability was large and resulting realizations were ranked second least similar to the cognitive geological model. The realization ensemble which was closest to the cognitive geological model belongs to Case 3. Here, the resistivity grid was reconstructed from the sharp SCI models, which, in this case, increased the fingerprint of resistive extreme values, which in turn results in less ambiguous reconstructed resistivity grids; compare Figure 7C and D. It should be noted that the usage of block Kriging for assigning the sharp resistivity models to the modeling grid, resulted in smoothing of sharp vertical boundaries otherwise found in sSCI models. These three cases together reveal the importance of the geophysical soft data when using the *snesim* setup presented in this study.

In relation to Case 5, it can be argued that even though the SkyTEM resistivity models are not used as soft data, they are still included indirectly since the TI, or cognitive geological model, was created using smooth SkyTEM resistivity models. However, the realizations related to Case 5, revealed an ensemble of realizations, which did not replicate the overall geological architecture, implying the importance of using the SkyTEM models as soft data.

On the other hand the cases related to studying the sensitivity towards borehole information, Case 4a and Case 4b, revealed that the large-scale hydrostratigraphic architecture was not changed significantly. The distance measure used in this study observes similarities or dissimilarities of large-scale hydrostratigraphic architecture, and is not sensitive towards local changes in small-scale patterns. The amount of geophysical information is relatively large, meaning the relative influence of (few) borehole data becomes less significant. This does *not* mean that the borehole data are not important; they both contain locally

accurate information and are used to estimate the regional resistivity-hydrostratigraphic relationship (Figure 1B). In other surveys, where the contrast between geophysical and lithological information is smaller, the importance of the borehole data will likely increase. In relation to this study, such small-scale changes are insignificant. Yet if the realizations are to be used for flow simulations or predictions on a smaller scale, such smaller scales might suddenly have an important impact on prediction accuracy. Additionally, if such small-scale patterns are important, the size of the model grid-cells should be smaller to accommodate simulations of these variations. Discretizing hydrostratigraphic and groundwater models with relatively small grid-cells can be CPU demanding, depending on the total number of grid cells.

In case 4b, the borehole data was used as soft information as in the study by Høyer *et al.* (2017). This was done since boreholes are associated with uncertainty related to a number of factors, as described above. Therefore, the soft borehole probability values derived during the assigning of the boreholes to the modeling grid are combined with the SkyTEM-based probability grids using the Tau model. This approach enables the borehole probability to alter the final probability-grid, while still conditioning the SkyTEM data. Combining the information rather than letting the borehole data count as "ground truth", *i.e.* hard data, allows the borehole data to influence the realizations, especially if the soft borehole information disagrees with the soft geophysical data (Figure 8).

The conceptual geological understanding has always been considered an integral part of geological modeling. In this case the conceptual geological understanding is implemented via the TI, which makes it easy to change the underlying conceptual geological understanding of a given model. A total of 3 different TIs were used for simulation in this study, the Kasted, Egebjerg, and conceptual TIs. The results showed that models simulated using the Egebjerg TI, Case1a, portrayed the same overall hydrostratigraphic architecture. This opens for the possibility of using 3D cognitive geological models as TIs for new survey areas, as long as the geological settings are similar. One key difference between the models, however, was the more block-like and coarse nature of the realizations using the Egebjerg TI, due to the coarseness of the Egebjerg TI. An important observation is that when a spatially dense and extensive geophysical dataset, such as SkyTEM, is present, the *snesim* realizations are not as sensitive towards the choice of TI, when the TI is relatively similar to the expected scenario. However, as illustrated in Case 1b picking a TI which has significantly different vertical proportions (Figure 5), which do not match the soft data, the TI dominates the realizations in places where the soft data do not display a high probability for a specific hydrostratigraphic unit. In Figure **9**C and Figure 10C, it can be seen that the *glacial clay valley*, both present in the soft data variable (Figure 8B) and the cognitive Kasted geological model (Figure 4A), is represented as *sand and gravel*. This leads to the conclusion that one needs to pay attention to the construction of the TI, as also witnessed in the study by Høyer *et al.* (2017). Furthermore, the large-scale and homogenous nature of the hydrostratigraphic architecture in the conceptual TI, results in realizations, which reflects the homogeneity. In comparison with the realizations based on the TIs derived from cognitive models, the realizations do not contain small-scale patterns.

The Kasted model and TI is influenced by non-stationarity, which has not been dealt with in the MPS setup. Even though the models are influenced by non-stationarity the simulations result in models, which overall resemble the cognitive model, *e.g.* Cases 1-4. However, once the geophysical data is removed in Case 5, the resulting MPS models are increasingly random and are heavily influenced by non-stationarity. It is important to note that the increasing amounts of soft geophysical data generally decrease the effects of non-stationarity, due to the increased conditioning of the soft data.

The reconstruction of the resistivity grid is an important step of the MPS setup presented in this study. This was illustrated in Case 2, where the incomplete resistivity grid was used instead of the reconstructed resistivity grid, resulting in larger realization variability and distance to the cognitive geological model. These realizations could have been improved by increasing the prior knowledge provided to *snesim* before simulation. One such option is to provide so-called vertical proportions, in place of solely the target global proportions. The global proportions simply give a percentage fraction of the different hydrostratigraphic

units in the outcome realizations. The vertical proportions are defined for each simulation grid layer, and determine the proportions as a function of depth. This makes sense if the different units in the realizations are clearly linked to geological units, which in turn have clear stratigraphic layering. In our case, this would have impacted the realizations by not allowing the presence of *hemipelagic clay* at the top of the model. Furthermore, *sand and gravel* and *glacial clay* would not be allowed at the bottom of the model. However, vertical proportions were not used in any other cases, and were therefore not used in Case 2. The usage of vertical proportions for conditioning could also improve the results of Case 5.

Part of the considerations of this study was to utilize the *DeeSse* code for Direct Sampling (DS) simulation. Whereas Chugunova and Hu (2008) present a MPS method for constraining a categorical simulation by a continuous auxiliary data, DS is much more flexible and allows multivariate simulations reproducing spatial statistics within and between variables, which can be categorical or continuous. DS requires the construction of a multivariate TI, and it is important that every variable reflects the spatial relationship to be modeled. In our context, one could envision creating a bivariate TI consisting of a hydrostratigraphic model and a continuous auxiliary variable reflecting an AEM data set. It is no trivial task, and, to our knowledge, no studies have been presented where an auxiliary variable is created for a 3D AEM data set. However, Lochbühler *et al.* (2014) presented a generalized example on creating an auxiliary variables for tomographic images, *i.e.* 2D images. Generally, the requirements for the geophysical modeling procedure are twofold. Firstly, the categorical TI needs to be populated with resistivity values, *e.g.* as in Christensen *et al.* (2017) where a Bayesian McMC algorithm is used to create 1D resistivity models drawn from a posterior probability distribution. This is no straightforward task. Secondly, the populated resistivity model then ideally needs to be forward modelled using full 3D forward modeling code, which is computationally expensive. Alternatively, approximate 1D forward modeling is also an option. The correct system parameters of the AEM instrument and data processing paramenters have to be taken into account. Thirdly, the synthetic data obtained by forward modeling must be inverted using the same procedure as the field data set. To our knowledge, such usage of an auxiliary variable for constraining soft geophysical models is not widespread within the domain of AEM geophysical methods. In this study, DS was only used to reconstruct the incomplete AEM dataset (a univariate case with the data set as a TI, and hard data; see section 3.2) and the *snesim* method was used for hydrostratigraphic modeling, due to its usage of the τ-model (Journel, 2002). The τ-model proved a more straightforward approach when combined with the method for creating resistivity-hydrostratigraphic histograms presented by Barfod *et al.* (2016).

The study presented by Barfod *et al.* (2018) used the alternative modified Hausdorff distance (MHD) measure for comparing realizations. Due to the computational burden of the method, it was difficult to create exhaustive distance computations, *i.e.* where all information from individual realizations is used. The usage of differences in EDT of binary translations of the categorical realizations for comparing the individual realizations proved to be a more computationally feasible approach. In this paper an efficient algorithm for computing the EDT was used (Maurer et al., 2003). This computationally advantageous approach for computing the distance between two realizations allows for a full analysis of the realizations. Each realization is then compared based on each of the hydrostratigraphic categories and on the entire 3D objects, resulting in a detailed comparison. The resulting distance matrix (Figure 11A) was able to differentiate between the realizations pertaining to the different cases. The random number seed between cases was chosen so the first realization of each case has the same random seed; the second realization has the same seed, *etc*. This can be seen in the distance matrix (Figure 11A), where off diagonal cases have a smaller distance values along the diagonal within the given 50 by 50 sub-matrix. An example is the 50 by 50 sub-matrix between the basic setup and Case 1a, where the diagonal is clearly marked by lower distances relative to the remaining sub-matrix.

7 Conclusion

A hydrogeophysical data set from Kasted in Denmark was used for stochastic hydrostratigraphic simulation using the *snesim* algorithm. The main goal of this study was to improve our understanding of ensemble hydrostratigraphic modeling variability related to stochastic *MPS* modeling. The study was divided into 8 sub-cases designed to reflect the impact related to key components of the hydrostratigraphic modeling setup, *i.e.* the TI, borehole lithology logs, and SkyTEM resistivity models. The results revealed that the hydrostratigraphic realizations were sensitive first and foremost to the geophysical dataset due to its extensive nature. Not including the geophysical data in the realizations resulted in an average Euclidean distance variability of 40 m and a distance to the cognitive model of 59 m, which was, by far, the largest distance of all realizations. Furthermore, the geophysical modeling procedure influences the resulting realizations. It was shown that choosing so-called sharp inversion models (sSCI), in place of smooth inversion models (SCI), resulted in a realization ensemble which had similar distance based variabilities, 9.4 m and 10.1 m, respectively. However, using sSCI models decreased the distance to the cognitive geological model from 24.3 m, to 21.6 m. The choice of a TI containing a relevant geological conceptualization is important. The cognitive Egebjerg model was used as a TI to simulate the hydrostratigraphic Kasted model, which yielded similar realizations to the case where the cognitive Kasted model was used as a TI. The Egebjerg TI contained relevant geological architecture, but if a conceptual TI is introduced containing significantly different vertical proportions, the resulting realizations will reflect these differing vertical proportions. Finally, it was seen that the borehole lithology logs did not significantly influence the realizations. The lithology logs only carry information in the immediate vicinity of the borehole, and are sparse in comparison to the resistivity data. The boreholes therefore only have a minor influence on the realizations. The comparison measures used here mainly compare the overall large-scale architecture components of the realizations, and do not reflect small-scale changes. In relation to this study the usage of the lithology logs as hard data does not show a significant impact on the MPS realizations. However, if the hydrostratigraphic models are used for predicting groundwater flow the boreholes might be important. However, it should be mentioned that the resistivity-hydrostratigraphic histograms, which are used extensively in this research, are created from the borehole information.

28

775 **References**

Auken, E., Christansen, A. V., Westergaard, J. H., Kirkegaard, C., Foged, N. and Viezzoli, A.: An integrated processing scheme for high-resolution airborne electromagnetic surveys, the SkyTEM system, Explor. Geophys., 40(2), 184–192, doi:10.1071/EG08128, 2009.

Barfod, A. A. S., Møller, I. and Christansen, A. V.: Compiling a national resistivity atlas of Denmark based on airborne and
780 ground-based transient electromagnetic data, J. Appl. Geophys., 134, 199–209, doi:10.1016/j.jappgeo.2016.09.017, 2016.

Barfod, A. A. S., Møller, I., Christiansen, A. V., Høyer, A.-S., Hoffimann, J., Straubhaar, J. and Caers, J.: Hydrostratigraphic modelling using multiple-point statistics and airborne transient electromagnetic methods, Hydrol. Earth Syst. Sci. Discuss., 22, 3351–3373, doi:https://doi.org/10.5194/hess-22-3351-2018, 2018.

Christensen, N. K., Minsley, B. J. and Christensen, S.: Generation of 3-D hydrostratigraphic zones from dense airborne
785 electromagnetic data to assess groundwater model prediction error, Water Resour. Res., 53(2), 1019–1038, doi:10.1002/2016WR019141, 2017.

Chugunova, T. and Hu, L. Y.: Multiple-Point Simulations Constrained by Continuous Auxiliary Data, Math. Geosci., 40(2), 133–146, doi:10.1007/s11004-007-9142-4, 2008.

Comunian, A., Renard, P. and Straubhaar, J.: 3D multiple-point statistics simulation using 2D training images, Comput.
790 Geosci., 40, 49–65, doi:10.1016/j.cageo.2011.07.009, 2012.

Constable, S. C., Parker, R. L. and Constable, C. G.: Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data, Geophysics, 52(3), 289–300, doi:10.1190/1.1442303, 1987.

Le Coz, M., Genthon, P. and Adler, P. M.: Multiple-Point Statistics for Modeling Facies Heterogeneities in a Porous Medium: The Komadugu-Yobe Alluvium, Lake Chad Basin, Math. Geosci., 43(7), 861–878, doi:10.1007/s11004-011-9353-6, 2011.

795 Ellis, R. G. and Oldenburg, D. W.: Applied geophysical inversion, Geophys. J. Int., 116(1), 5–11, doi:10.1111/j.1365-246X.1994.tb02122.x, 1994.

Ferré, T. P. A.: Revisiting the Relationship Between Data, Models, and Decision-Making, Groundwater, 55(5), 604–614, doi:10.1111/gwat.12574, 2017.

Feyen, L. and Caers, J.: Quantifying geological uncertainty for flow and transport modeling in multi-modal heterogeneous
800 formations, Adv. Water Resour., 29(6), 912–929, doi:10.1016/j.advwatres.2005.08.002, 2006.

Fleckenstein, J. H., Niswonger, R. G. and Fogg, G. E.: River-aquifer interactions, geologic heterogeneity, and low-flow management, Ground Water, 44(6), 837–852, doi:10.1111/j.1745-6584.2006.00190.x, 2006.

Fogg, G. E., Noyes, C. D. and Carle, S. F.: Geologically based model of heterogeneous hydraulic conductivity in an alluvial setting, Hydrogeol. J., 6(1), 131–143, doi:10.1007/s100400050139, 1998.

805 Gelhar, L. W.: Stochastic analysis of flow in heterogeneous porous media, in Fundamental of Transport Phenomena in Porous Media, edited by J. Bear and Y. M. Corapcioglu, pp. 673–717, Springer Netherlands., 1984.

Goovaerts, P.: Geostatistics for Natural Resource Evaluation, 1st ed., edited by A. G. Journel, Oxford University Press, New York., 1997.

Gunnink, J. L. and Siemon, B.: Applying airborne electromagnetics in 3D stochastic geohydrological modelling for
810 determining groundwater protection, Near Surf. Geophys., 13(1), 45–60, doi:10.3997/1873-0604.2014044, 2015.

He, X., Koch, J., Sonnenborg, T. O., Jørgensen, F., Schamper, C. and Christian Refsgaard, J.: Transition probability-based

stochastic geological modeling using airborne geophysical data and borehole data, Water Resour. Res., 50(4), 3147–3169, doi:10.1002/2013WR014593, 2014.

He, X. L., Sonnenborg, T. O., Jørgensen, F. and Jensen, H. J.: Modelling a real-world buried valley system with vertical non-stationarity using multiple-point statistics, Hydrogeol. J., 25(2), 359–370, doi:10.1007/s10040-016-1486-8, 2016.

Hermans, T., Nguyen, F. and Caers, J.: Uncertainty in training image-based inversion of hydraulic head data constrained to ERT data: Workflow and case study, Water Resour. Res., 51(7), 5332–5352, doi:10.1002/2014WR016460, 2015.

Høyer, A.-S., Jørgensen, F., Sandersen, P. B. E., Viezzoli, A. and Møller, I.: 3D geological modelling of a complex buried-valley network delineated from borehole and AEM data, J. Appl. Geophys., 122, 94–102, doi:10.1016/j.jappgeo.2015.09.004, 2015.

Høyer, A.-S., Vignoli, G., Hansen, T. M., Vu, L. T., Keefer, D. A. and Jørgensen, F.: Multiple-point statistical simulation for hydrogeological models: 3-D training image development and conditioning strategies, Hydrol. Earth Syst. Sci., 21(12), 6069–6089, doi:10.5194/hess-21-6069-2017, 2017.

de Iaco, S. and Maggio, S.: Validation techniques for geological patterns simulations based on variogram and multiple-point statistics, Math. Geosci., 43(4), 483–500, doi:10.1007/s11004-011-9326-9, 2011.

Jørgensen, F. and Sandersen, P. B. E.: Buried and open tunnel valleys in Denmark—erosion beneath multiple ice sheets, Quat. Sci. Rev., 25(11–12), 1339–1363, doi:10.1016/j.quascirev.2005.11.006, 2006.

Jørgensen, F., Møller, R. R., Sandersen, P. B. E. and Nebel, L.: 3-D geological modelling of the Egebjerg area , Denmark, based on hydrogeophysical data, Geol. Surv. Denmark Greenl. Bull., 20, 27–30, 2010.

Jørgensen, F., Møller, R. R., Nebel, L., Jensen, N. P., Christansen, A. V. and Sandersen, P. B. E.: A method for cognitive 3D geological voxel modelling of AEM data, Bull. Eng. Geol. Environ., 72(3–4), 421–432, doi:10.1007/s10064-013-0487-2, 2013.

Journel, A. G.: Combining Knowledge From Diverse Sources : An Alternative to Traditional Data, Math. Geol., 34(5), 2002.

Journel, A. G. and Zhang, T.: The Necessity of a Multiple-Point Prior Model, Math. Geol., 38(5), 591–610, doi:10.1007/s11004-006-9031-2, 2007.

Krishnan, S.: Combining individual data information : A review and the tau model., 2004.

LaBolle, E. M. and Fogg, G. E.: Role of Molecular Diffusion in Contaminant Migration and Recovery in an Alluvial Aquifer System, Transp. Porous Media, 42, 155–179, doi:10.1023/A:1006772716244, 2001.

Liwei Wang, Yan Zhang and Jufu Feng: On the Euclidean distance of images, IEEE Trans. Pattern Anal. Mach. Intell., 27(8), 1334–1339, doi:10.1109/TPAMI.2005.165, 2005.

Lochbühler, T., Pirot, G., Straubhaar, J. and Linde, N.: Conditioning of Multiple-Point Statistics Facies Simulations to Tomographic Images, Math. Geosci., 46(5), 625–645, doi:10.1007/s11004-013-9484-z, 2014.

Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints, Int. J. Comput. Vis., 60(2), 91–110, doi:10.1023/B:VISI.0000029664.99615.94, 2004.

Mariethoz, G. and Renard, P.: Reconstruction of Incomplete Data Sets or Images Using Direct Sampling, Math. Geosci., 42(3), 245–268, doi:10.1007/s11004-010-9270-0, 2010.

Marker, P. A., Vilhelmsen, T. N., Foged, N., Wernberg, T., Auken, E. and Bauer-Gottwein, P.: Probabilistic predictions using a groundwater model informed with airborne EM data, Adv. Water Resour., 103, 86–98, doi:10.1016/j.advwatres.2017.03.002,

2017.

850 Maurer, C. R., Qi, R., Raghavan, V. and Member, S.: A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions, , 25(2), 265–270, doi:10.1109/TPAMI.2003.1177156, 2003.

Maxey, G. B.: Hydrostratigraphic units, J. Hydrol., 2, 124–129, 1964.

Okabe, H. and Blunt, M. J.: Prediction of permeability for porous media reconstructed using multiple-point statistics, Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys., 70(6 2), 1–2, doi:10.1103/PhysRevE.70.066135, 2004.

855 Okabe, H. and Blunt, M. J.: Pore space reconstruction using multiple-point statistics, J. Pet. Sci. Eng., 46(1–2), 121–137, doi:10.1016/j.petrol.2004.08.002, 2005.

Pirot, G.: Using Training Images to Build Model Ensembles with Structural Variability, Groundwater, 55(5), 656–659, doi:10.1111/gwat.12556, 2017.

Remy, N., Boucher, A. and Wu, J.: Applied geostatistics with SGeMS: a user's guide, Cambridge University Press., 2014.

860 Royse, K. R.: Combining numerical and cognitive 3D modelling approaches in order to determine the structure of the Chalk in the London Basin, Comput. Geosci., 36(4), 500–511, doi:10.1016/j.cageo.2009.10.001, 2010.

Sandersen, P. B. E., Jørgensen, F., Larsen, N. K., Westergaard, J. H. and Auken, E.: Rapid tunnel-valley formation beneath the receding Late Weichselian ice sheet in Vendsyssel, Denmark, Boreas, 38(4), 834–851, doi:10.1111/j.1502-3885.2009.00105.x, 2009.

865 Seifert, D., Sonnenborg, T. O., Refsgaard, J. C., Højberg, A. L. and Troldborg, L.: Assessment of hydrological model predictive ability given multiple conceptual geological models, Water Resour. Res., 48(6), 1–16, doi:10.1029/2011WR011149, 2012.

Sørensen, K. I. and Auken, E.: SkyTEM – a new high-resolution transient electromagnetic system, Explor. Geophys., 35(3), 191–199, doi:10.1071/EG04194, 2004.

Strebelle, S.: Conditional Simulation of Complex Geological Structures Using Multiple-Point Statistics, Math. Geol., 34(1), 870 1–21, doi:10.1023/A:1014009426274, 2002.

Strebelle, S. and Journel, A. G.: Reservoir modeling using multiple-point statistics, Proc. SPE Annu. Tech. Conf. Exhib., 11, doi:10.2523/71324-MS, 2001.

Tahmasebi, P., Hezarkhani, A. and Sahimi, M.: Multiple-point geostatistical modeling based on the cross-correlation functions, Comput. Geosci., 16(3), 779–797, doi:10.1007/s10596-012-9287-1, 2012.

875 Tan, X., Tahmasebi, P. and Caers, J.: Comparing training-image based algorithms using an analysis of distance, Math. Geosci., 46(2), 149–169, doi:10.1007/s11004-013-9482-1, 2014.

Tarantola, A. and Valette, B.: Generalized nonlinear inverse problems solved using the least squares criterion, Rev. Geophys., 20(2), 219, doi:10.1029/RG020i002p00219, 1982.

Tran, T.: Improving variogram reproduction on dense simulation grids, Comput. Geosci., 20(7–8), 1161–1168, 880 doi:10.1016/0098-3004(94)90069-8, 1994.

Viezzoli, A., Christansen, A. V., Auken, E. and Sørensen, K. I.: Quasi-3D modeling of airborne TEM data by spatially constrained inversion, Soundings, 73(3), 2008.

Vignoli, G., Fiandaca, G., Christansen, A. V., Kirkegaard, C. and Auken, E.: Sharp spatially constrained inversion with applications to transient electromagnetic data, Geophys. Prospect., 63(1), 243–255, doi:10.1111/1365-2478.12185, 2015.

885 Xiaofeng, F. and Wei, W.: Centralized binary patterns embedded with image euclidean distance for facial expression recognition, Proc. - 4th Int. Conf. Nat. Comput. ICNC 2008, 4, 115–119, doi:10.1109/ICNC.2008.94, 2008.

Zhao, Z. and Illman, W. A.: On the importance of geological data for three-dimensional steady-state hydraulic tomography analysis at a highly heterogeneous aquifer-aquitard system, J. Hydrol., 544, 640–657, doi:10.1016/j.jhydrol.2016.12.004, 2017.

890

Appendix

A1. Simple Kriging parameters for creating borehole probability grids (Case 4)

<u>Variogram model type:</u>
Exponential

895

$mean_{SK} = 1/3$

<u>Search ellipsoid:</u>

	Max	Med	Min
Ranges	200	200	10
Angles	0	0	0

<u>Variogram:</u>

900 Contribution = 1

	Max	Med	Min
Ranges	1000	1000	50
Angles	0	0	0

A2. General SGeMS parameters used for the *snesim* realizations:

Property name:	value/count:
algorithm name	snesim_std
use_pre_simulated_gridded_data	0
Use_ProbField	1
ProbField_properties	count=3, value="sg_0;ct1;pc2"
TauModelObject	[1 1]
use_vertical_proportion	0
Cmin	5
Constraint_Marginal_ADVANCED	0
resimulation_criterion	-1
resimulation_iteration_nb	1
Nb_Multigrids_ADVANCED	5
Debug_Level	0
Subgrid_choice	0
expand_isotropic	1
expand_anisotropic	0
aniso_factor	NA
Use_Affinity	0
Use_Rotation	0
Nb_Facies	3
Marginal_Cdf	0.19 0.24 0.57
Max_Cond	100
Search_Ellipsoid	[750 750 0 0 0 0]

<u>Marginal cdf:</u>

	sand and gravel	*glacial clay*	*hemipelagic clay*
value	0.19	0.24	0.57

905