Multiple-point statistical simulation for hydrogeological models: 3D training image development and conditioning strategies

Anne-Sophie Høyer[1], Giulio Vignoli[1,2], Thomas Mejer Hansen[3], Le Thanh Vu[4], Donald A. Keefer[5], Flemming Jørgensen[1]

[1]Groundwater and Quaternary Geology Mapping Department, GEUS, Aarhus, 8000, Denmark
[2]Dipartimento di Ingegneria Civile, Ambientale e Architettura (DICAAR), Università di Cagliari, Cagliari, 09123, Italy
[3]Niels-Bohr Institute, University of Copenhagen, Copenhagen, 2100, Denmark
[4]I-GIS, Risskov, 8240, Denmark
[5]Illinois State Geological Survey, Champaign, IL 61820, USA

Correspondence to: Anne-Sophie Høyer (ahc@geus.dk) and Giulio Vignoli (gvignoli@unica.it)

Abstract. Most studies about the application of geostatistical simulations based on multiple-point statistics (MPS) to hydrogeological modelling focus on relatively fine-scale models and concentrate on the estimation of facies-level structural uncertainty. Much less attention is paid to the use of input data and optimal construction of training images. For instance, even though the training image should capture a set of spatial geological characteristics to guide the simulations, the majority of the research still relies on 2D or quasi-3D training images. In the present study, we demonstrate a novel strategy for 3D MPS modelling characterized by: (i) realistic 3D training images, and (ii) an effective workflow for incorporating a diverse group of geological and geophysical data sets. The study covers an area of $2\,810$ km^2 in the southern part of Denmark. MPS simulations are performed on a subset of the geological succession (the lower to middle Miocene sediments) which is characterized by relatively uniform structures and dominated by sand and clay. The simulated domain is large and each of the geostatistical realizations contains approximately 45 million voxels with size 100 m x 100 m x 5 m. Data used for the modelling include water well logs, high-resolution seismic data, and a previously published 3D geological model. We apply a series of different strategies for the simulations based on data quality, and develop a novel method to effectively create observed sand/clay spatial trends. The training image is constructed as a small 3D voxel model covering an area of 90 km^2. We use an iterative training image development strategy and find that even slight modifications in the training image create significant changes in simulations. Thus, the study underlines that it is important to consider both the geological environment, and the type and quality of input information in order to achieve optimal results from MPS modelling. In this study we present a possible workflow to build the training image and effectively handle different types of input information to perform large-scale geostatistical modelling.

1 Introduction

Simulation of groundwater flow and solute transport requires representative hydrogeological models of the subsurface. While many studies focus on estimating the spatial distribution of hydraulic properties (i.e. hydraulic conductivity, porosity), the reliable delineation of the underlying geological structural or hydrostatrigraphic model is far more important to ensure the reliability of groundwater flow predictions (e.g. Carrera, 1993; Liu et al., 2004a; Renard and Allard, 2013).

Early geostatistical methods (e.g.: Chiles and Delfiner, 1999; Deutsch, 2002) provided tools to model uncertainty in smoothly varying Gaussian properties, like porosity or permeability within targeted rock formations (e.g., oil reservoirs). These produced petrophysical models that generally improved mineral resource development, but the models were poor when abrupt changes in rock type, and hence petrophysical properties were important in defining reservoir structure. Later generations of geostatistical methods supplied tools for modelling rock types (i.e., lithology or facies), that provided incremental improvement (Stafleu et al., 2011), but still did not effectively model non-stationary patterns in rock type, and poorly captured patterns in stratigraphical units that are typically complex combinations of rock type.

Without suitable tools for developing realistic uncertainty-informed models of either non-stationary rock type or hydrostratigraphy, groundwater models are typically based on single, deterministic geological framework models that are manually constructed. Research has continued to explore methods for uncertainty-based simulation of structural heterogeneities (Huysmans and Dassargues, 2009; Kessler et al., 2013) to better estimate uncertainties in groundwater flow (Feyen and Caers, 2006; He et al., 2013; Poeter and Anderson, 2005; Refsgaard et al., 2012).

The multiple-point statistics (MPS, Guardiano and Srivastava, 1993; Strebelle, 2002) was developed with the objective to better reproduce complex geological patterns than other available geostatistical modelling techniques. The method combines the ability to condition the realizations to hard data (information without uncertainty) and soft data (information with uncertainty) with the ability to reproduce geological features characterized by statistical properties described through a so-called Training Image (TI).

Hence, the TI plays a crucial role and should be constructed to represent the structural patterns of interest (Hu and Chugunova, 2008; Maharaja, 2008; Pickel et al., 2015). However, a fundamental challenge for the use of MPS for 3D hydrogeological modelling lies in the difficulty to produce realistic 3D TIs (e.g., Pérez et al., 2014, Ronayne et al, 2008), and most of the studies in the literature are therefore based on 2D or quasi-3D TIs (Jha et al., 2014; Comunian et al., 2012; Feyen and Caers, 2006). Several researches have demonstrated the potential of the MPS to simulate 2D patterns (Hu and Chugunova, 2008; Liu et al., 2004b; Strebelle, 2002), but less attention has been given to the construction of real three-dimensional TIs.

In the present study, we describe a strategy to develop effective and realistic 3D TIs based on the a posteriori analysis of the associated unconstrained realizations.

Concerning the use of conditioning data, it is extremely important to address the fact that the data are characterized by different uncertainty levels and scales (He et al., 2014; McCarty and Curtis, 1997). In this study, available sources of

information are seismic lines, boreholes, and a pre-existing, manually constructed geo-model. Depending on their scale and uncertainty, we developed a practical way to incorporate all of them into the Single Normal Equation Simulation (SNESIM) workflow as it is implemented in SGEMS (Remy et al., 2009). The approach is general and can be readily extended to other data types.

5   In summary, the main objectives of the present study concern: (i) the iterative development of effective 3D TIs, and (ii) the optimal strategy for the simulation conditioning.

The paper is organized as follows: (i) section "2 Study area" deals with the overall geological framework of the area under investigation, while section "3 Data" describes the amount and characteristics of the different kinds of available data; (ii) section "4 Establishing framework-model constraints" goes into details about the specific geological unit targeted during the

10   stochastic simulations; (iii) "5 Defining MPS input information" is the most methodological section and consists of the presentation of the different approaches whose outputs are then compared in the subsequent section "6 Results"; (iv) the second last section, "7 Discussion", is about the assumptions/choices made across the paper and their possible limitations and future developments; (v) "8 Conclusion" is a very concise section where we simply summarize our results.

Moreover, in "Appendix A", for sake of completeness, we provide further details regarding the variability within each

15   probabilistic model and analyse additional realizations.

2 Study area

The study area covers $2\,810$ km^2, from coast to coast in the southern part of Jutland, Denmark, and northern portion of Germany (Fig. 1). A part of this area (625 km^2) has formerly undergone 3D geological modelling (the Tønder model, Fig. 1) which is presented in Jørgensen et al. (2015). In this paper, we will concentrate on the Miocene sediments within the model

20   domain. The Miocene sediments were selected because: (i) they are primarily composed of two main hydrogeological facies: clay and sand, and (ii) the unit is composed of rather uniform structures throughout the area and therefore possible to describe in a 3D TI.

A conceptual sketch of the geology in the area is shown in Fig. 2. The base of the regional groundwater flow system corresponds with the top of the very fine-grained Paleogene clays that have a gentle dip towards the southwest. The

25   overlying Miocene sediments consist of marine clay with sandy deltaic lobate layers. The Miocene deposits in Jutland has been thoroughly studied by Rasmussen et al. (2010) who based the studies on sedimentological and palynological investigations of outcrops and cores, combined with interpretation of high-resolution seismic data. During the Miocene, the coastline fluctuated generally northeast-southwest across Jutland, resulting in a deltaic depositional environment. Rasmussen et al. (2010) divided the Miocene deposits into three sand-rich deltaic units that are interfingered with prodeltaic clayey

30   units. In periods with low water levels, the coastline was situated far to the southwest and coarse-grained sediments were consequently deposited in the study area. When the water level was high, the coastline was situated further to the northeast, resulting in deposition of marine clays. The delta lobes dominate in the northeast and show smooth dips towards southwest.

In middle and upper Miocene, very fine-grained marine clays belonging to the Måde Group were deposited in the western part of the model domain.

The unconformable boundary between the Quaternary and the Miocene is often scoured by buried valleys (Fig. 2). The Quaternary deposits mainly consist of till and various meltwater deposits, but, in some places, also of interglacial and postglacial deposits. In some areas, the Quaternary deposits are heavily deformed due to glaciotectonism. In the western portion of the study area, the geology generally consists of outwash sandurs surrounding old pre-Weichselian moraine landscapes. The outwash sandurs show a low relief with a gentle dip towards west, whereas the old moraine landscape shows a more irregular relief up to 60 meters above sea-level (m.a.s.l). In the east, the younger morainic landscape from the Weichselian shows topographic changes between 40 - 90 m.a.s.l. The model area is influenced by a large fault-bounded structure (the Tønder graben structure, Fig. 1b) that offsets the deeper Top Chalk surface of about hundred meters (Ter-Borch, 1991). The faults clearly offset both the top of the Paleogene and the bottom of the Quaternary.

3 Data

3.1 Borehole data

The borehole data in the Danish portion of the study area are extracted from the Danish national borehole database, the 'Jupiter' database (http://jupiter.geus.dk), and comprise about 9 000 boreholes with lithological information from driller's logs and sample set descriptions. The Jupiter database provides information on a number of parameters about each borehole, which enables an evaluation of borehole data quality. Boreholes, like any other kinds of data, are affected by uncertainty. The level of uncertainty determines the quality/reliability of the measurements (i.e., the quality/reliability of the boreholes). Many factors impact the quality of boreholes. Just to mention a few of them: (i) the drilling methods (e.g.: rotary drilling and air lift drillings. In some cases, for example, the finer sediments can be flushed out and the driller could potentially misinterpret a clay layer as more sandy); (ii) the drilling purpose (sometimes, if the goal is to reach a specific target, the lithological description can be poor since it is not a priority); (iii) the age of the boreholes (for example, nowadays, in Denmark, samples are collected systematically every meter, and this was not the case few years ago); (iv) the presence of simultaneous wireline logging data (these kinds of ancillary information make the geological interpretation definitely more certain). During the geological modelling phase, a skilled geologist should go through all the borehole records and verify all these different pieces of information and check for inconsistencies. This is (or should be) the standard procedure to assess the quality of the boreholes and prepare the data for the subsequent geo-modelling phases. In the study area, many borehole records are of low quality.

Together with the Danish boreholes, geological information from about 500 boreholes located within the German portion of the study area and deeper than 10 m are included in the analysis. The majority of the boreholes in both the Danish and the German area are quite shallow and, in total, only 2 % of all the boreholes are deeper than 100 m (Fig. 1).

3.2 High-resolution seismic data

The seismic data are extracted from the Danish geophysical database 'Gerda' (Møller et al., 2009). Most of the seismic lines were conducted with the purpose to investigate the groundwater resources and examine the Miocene sediments (Rasmussen et al., 2007). In the study area, seismic lines for a total length of around 170 km (Fig. 1) were collected by several contractors with two slightly different acquisition systems that, for sake of clarity, we conventionally indicate here with SYS_COWI and SYS_RAMBØLL. In both cases, the lines were acquired as landstreamer high-resolution seismic data (Vangkilde-Pedersen et al., 2006) by using seismic vibrators as energy source. The frequency ranges spanned: from 50 Hz to 350 Hz, for SYS_COWI, and from 50 Hz to 400 Hz, for SYS_RAMBØLL; for both systems, the sweep was 5 s long. The receiver arrays consisted of: 95 not-equally spaced geophones (1.25 m between the first 32 geophones, and 2.5 m for the others), for SYS_COWI, and 102 geophones (1.25 m between the first 50 geophones, and 2.5 m for the others), for SYS_RAMBØLL. In both acquisition settings, the sources were fired every 10 m.

Under normal circumstances, the data are high quality between approximately 30 m to 800 m in depth. Prior to the import of the seismic data to the geological interpretation software, the elevation values are adjusted, based on an assumed constant seismic velocity of 1 800 m/s (Kristensen et al., 2015), which is a common velocity for Miocene and Quaternary deposits in Denmark (Høyer et al., 2011; Jørgensen et al., 2003). Elevation corrections are important because of the considerable effect of the topography that, even if mildly varying, along extended profile can significantly affect the quality of the final results.

4 Establishing framework-model constraints

The MPS model domain corresponds to the lower and middle Miocene sediments which are positioned below the Måde Group and above the Top Paleogene surface (Fig. 3). Inside the Tønder model area (Fig. 1), however, the results of the already existing model (Jørgensen et al., 2015) are used and therefore are not included in the MPS simulations. The Tønder model results are used because the model was constructed based on very thorough geological analyses and so considered as the best geological model obtainable in that area. The surfaces of the relevant geological boundaries: Top Paleogene, Bottom Måde and Top pre-Quaternary (Fig. 3) are constructed in the geo-modelling software package Geoscene3D (I-GIS, 2014) by using interpretation points and interpolating these data to create the corresponding stratigraphical surfaces. The top of the MPS model domain is obtained by merging the surfaces of the Bottom Måde and the Top pre-Quaternary surface (Figs. 3 and 4a). The lower boundary of the model domain corresponds to the Top Paleogene surface (Figs. 3 and 4b).

The top surface of the Miocene sediments shows a regional dip towards west with significant depressions along the Tønder graben structure and at the buried valleys. The surface elevation ranges from 20 m.a.s.l. to -460 m.a.s.l. (Fig. 4a), corresponding to depths varying between 30 m and 80 m, in the east, and of about 160 m, in the west. In the Tønder graben, the depth to the top of the Miocene ranges approximately from 260 m to 320 m and locally (associated with buried valleys) up to 460 m. The bottom surface of the Miocene sediments (Fig. 4b) varies more smoothly, and also has a small regional dip from east (near elevation -100 m.a.s.l.) to west (approximately elevation -320 m.a.s.l.), punctuated primarily by the Tønder

graben structure (elevation down to -560 m.a.s.l.). The Miocene sediments typically have a thickness varying between 50 m and 250 m, but are thinner and locally absent below the buried valleys (Fig. 2).

5 Defining MPS input information

5.1 Seismic data

5   Rasmussen et al. (2010) proposed a lithostratigraphy for the Miocene succession, a "Miocene model" (Kristensen et al., 2015), based on interpretations of seismic data that was correlated with borehole interpretations and outcrops. The observational data points used in the Miocene model were utilized in our study to define the top of individual Miocene stratigraphical formations along the seismic lines (Fig. 5a). In order to use this information in the MPS modelling, the stratigraphical interpretations from the Miocene model are translated into a binary sand/clay voxel model of one cell width
10 along each seismic line (Fig. 5b and 6).

The sand/clay distribution along the seismic lines is used as hard data in the MPS simulations. The information is used as hard conditioning since it is considered highly reliable, and since the scale of structures delineated by the seismics is comparable to the scale of the simulated output.

5.2 Existing 3D model

15 In order to ensure edge matching between the stochastic realizations and the deterministically constructed Tønder model, a buffer zone along the pre-existing model's edges is created within GeoScene3D and used as hard conditioning data for the MPS simulations. The Miocene formations interpreted in the existing Tønder model (Jørgensen et al., 2015) are translated into sand and clay. In Fig. 6, the hard conditioning data from the pre-existing model (the buffer zone) are shown together with the hard information from the seismic lines.

20 **5.3 Borehole data**

All lithological categories contained in Jupiter are simplified and divided into three groups: sand, clay and 'other'. In this study, the conditioning based on the boreholes has been conducted through the use of three different strategies. The different ways of considering the borehole information are illustrated and summarized in Fig. 7.

The simplest and most common approach for exploiting the information from the boreholes is to include them as hard
25 conditioning data. This is the first test performed in the present study (Fig. 7a). However, this simple approach does not take into consideration the uncertainty in borehole data associated with: (i) inaccuracies in the recorded information, and (ii) the resolution, or scale, differences between the information in the borehole records and the geological model cell size.

In order to address these issues, an alternative conditioning strategy is tested. In this case, the boreholes are considered as soft data (Fig. 7b): a 20-meter-long moving window is applied to each original borehole in order to average the densely
30 sampled lithological data into probabilities defined on the coarser simulation grid. Since the typical vertical dimensions of

the structures are ~20 m, the size of the averaging window is chosen accordingly. This procedure is consistent with the reasonable assumption that, in the middle of the geological formation, we are relatively certain of the lithology, while this certainty decreases as we get closer to the boundaries. Moreover, in order to take the general uncertainty in the borehole information into account, we map the resulting averaged values into a sand probability interval ranging from 80 to 20 %. In

5   this way, voxels with the highest chances of having sand are characterized by a maximum sand probability value of 80 % (in pink, Fig. 7b), whereas voxels with the lowest sand probability are associated to a value as low as 20 % (in green, Fig. 7b). In Fig. 7b, it is clear that the transition zones between sand and clay correspond to a band with sand probabilities of 40 %. This value is the marginal distribution value for sand occurrence within the investigated model volume, as it is calculated from the borehole data and consistently formalized in the TIs.

10   When borehole information is too distant, the sand/clay ratios in the stochastic realization are solely derived from the TI and the sand marginal distribution. Unfortunately, borehole information at the relevant depths is relatively sparse (due to shallow boreholes), and, at the same time, the SNESIM algorithm has a tendency to ignore such "localized" soft data (Hansen et al., submitted). Moreover, it is known that the model domain shows a slight spatial lithological variation in which the overall proportion of sand is higher in the northeast compared to the southwest (that is also clear from the borehole data). To enforce

15   this real, general trend in the realisations, the "localized" borehole probabilities are kriged into a 3D grid to be used as "diffuse" soft conditioning (Fig. 7c and Fig. 8). This would be unnecessary if SNESIM could handle soft data as local conditional probability and, here, we are suggesting a practical strategy to effectively overcome this limitation. Clearly, the kriged sand probability distribution is seen to correspond to the boreholes close to their locations (Fig. 7c), whereas areas far from boreholes (e.g., at km 6 along the section in Fig. 7c) appear white (so characterized by a sand probability equal to 40

20   %). However, in a few cases (e.g. the clay layer in the deep part of the borehole at profile distance 10 km, Fig. 7c), the borehole information does not seem to migrate into the surrounding grid. This can be explained by the influence of other boreholes in the 3D domain. The desired spatial trend, characterized by a higher sand probability in the north-east compared to the south-west, is evident in the sand probability grid (Fig. 8).

5.4 Training image

25   The training images (TIs) are constructed as 3D voxel models with the same discretisation as the entire model (100 m by 100 m laterally, and 5 m vertically) and consist in approximately 500 000 voxels, covering an area of 90 km^2. The sizes of the TIs are significantly smaller than the simulation domain, but they are large enough to cover the size of the typical structures in the simulated Miocene unit. The TIs are used by the MPS simulation algorithm to represent the basic spatial and proportion relationships of the sand and clay facies within the Miocene unit. Hence, the TIs are modelled to show deltaic sand layers

30   building out towards southwest within a larger clayey unit. The sand content in the TIs is 40 %, in accordance with the proportion of sand observed in the borehole data.

The TIs are built in Geoscene3D using the voxel modelling tools described in Jørgensen et al. (2013). In practice, each sand and clay layer is defined by so called "interpretation" points. These points are subsequently interpolated into surface grids,

defining the volumes, which are then populated with sand and clay voxels. Thus, by manually changing the locations of these interpretation points and/or creating/deleting some of them, we can have a full control over the adjustments of the TIs. In the present study, several different TIs were tested, and, for sake of simplicity, only two of them (the first attempt and the final TI) are explicitly showed in this paper (Fig. 9). The first TI (TI1 - Fig. 9a) is based on the existing 3D geological model covering an adjacent area (the Tønder model; Jørgensen et al., 2015). The first TI has been manually adjusted, during several iterations, based on the unconditional outputs. This iterative process stopped when the corresponding unconstrained realization was found satisfactory in terms of its ability to mimic the geological features we expect in the Miocene across the study area. Those expectations about the geology are based on our prior geological understanding of the area, the available seismic lines, and the few existing deep boreholes. For example, the unconditioned realizations, associated with the TIs in Fig. 9, are shown in Fig. 10. The main difference is that TI1 has more layers than the second one (TI2 - Fig. 9b). This is clearly reflected in the unconditioned simulations, in which, the realization based on TI1 (Fig. 10a) shows significantly more layers than the corresponding realization based on TI2 (Fig. 10b). The results are evaluated and compared against the structures expected from the Miocene model (Kristensen et al., 2015). Hence, to adhere to what we know about the Miocene geology, the unconstrained realization associated with the selected TI was supposed to show fewer, larger, and more compact sand structures. These characteristics are evident if we directly compare the two realizations in Fig. 10, and clearly confirmed by the study of the associated distributions in Fig.s 11, 12, 13 (see, for instance, Haralick and Shapiro, 1992). In particular, Fig. 11 quantitatively demonstrates that larger sand bodies are more frequent in the realization corresponding to TI2 (black histogram), while the realization generated by TI1 (green histogram) is characterized by a significantly higher presence of relatively small sand layers. Regarding the shape of the features of the realizations in Fig. 10, Fig. 12 highlights that elongated sand structures are more probable for the realization in Fig. 10a (in fact, eccentricity equal 1 corresponds to the degenerate case of a straight line). Jaggedness, defined as the ratio between the surface and the size of bodies, can provide a useful estimation of the compactness of the sand bodies (Fig. 13). Not surprisingly, the sand lenses in the realization associated to TI1 are more jagged than those in the other realization obtained by using TI2. Because of the higher accordance between our geological expectation of the Miocene in the area and the unconstrained realization in Fig. 10b, TI2 is selected for all MPS simulations discussed in the rest of the paper.

6 Results

In the following, we present and compare the structures of the single realizations generated by each of the different conditional strategies analysed in this study (Table 1). All the realizations are produced by using the same random seed to better appreciate the differences. For comparison, an unconditioned realization (a) is presented in the first panel of each of the figures (Figs. 14 - 16). The second panel (b) shows a realization generated by using exclusively hard conditioning. In the third panel (c) a realization is presented, in which the borehole information is directly treated as soft conditioning data. Finally, in the fourth panel (d), a realization with the sand probability borehole grid used for soft conditioning is shown.

Figure 14 shows a horizontal slice through a realization representing each of the four conditional strategies listed in Table 1. As expected, the overall size and form of the structures are comparable between the realizations in (a) – (d) since the spatial characteristics of the structures are primarily determined by the TI. In (b), (c) and (d), the realizations within the Tønder buffer zone (delimited by the dash lines, Fig. 14) are completely defined by the hard data. It can be observed that the buffer zone cannot be identified in the realizations, which further supports our final choice for the TI2. In fact, this indicates that the same kind of spatial variability and geological patterns are seamlessly present within the pre-existing Tønder model and the simulation results. In this horizontal slice, it is possible to appreciate the effect of the soft probability grid (Fig. 14d) compared to the case of "localized" borehole information (Fig. 14b and 14c): when using the probability grid, there is a significantly higher sand content and more interconnected sand body occurrence in the middle part of the study area (compare with Fig. 8).

Figure 15 shows vertical profiles through each of the tested realizations (Table 1) along a transect with several deep boreholes (for location, see Fig. 14). Like in Fig. 14, the overall structural patterns of the individual realizations appear similar. The outcome of the unconditioned realization does not agree with the borehole information, since this has not been used. For instance, this is confirmed by the borehole at profile distance 10 km in Fig. 15a. On the contrary, in the realization where the borehole information is considered as hard data (Fig. 15b), almost perfect consistency is observed between the realization and the boreholes. However, mismatches naturally occur when the individual layers in the boreholes are thinner than the simulated voxel thicknesses (e.g., by the borehole at profile distance of about 15.5 km). Another seeming mismatch in Fig. 15b occurs every time the influence of the borehole information is extremely local, sometimes limited to a single voxel corresponding to the one actually holding the borehole information. This is, for instance, seen for the borehole at profile distance of about 9 km, where the fit between the realization and the deeper part of the borehole only appears when observed at a very detailed scale.

Also in the realization where the borehole information is considered as soft data (Fig. 15c), there is a generally good fit between the realization and the borehole data, but this match is clearly less pronounced in areas where a high uncertainty is associated with the borehole data (Fig. 7b). For instance, a poorer fit compared to Fig. 15b is observed at profile distance of about 10 km. Generally, the fit to the boreholes is better when the soft conditioning information does not conflict with other statistical properties as defined by the TI or neighbouring conditioning information. An example of this conflict can be observed by the borehole at about 9 km. Here, the test with the hard data (Fig. 15b) resulted in a highly local match concentrated in a single voxel column, implying that the borehole information did not comply with neighbouring data or statistical parameters. For the same reason, the realization does not strictly fit the borehole when the borehole information is treated as soft conditioning data (Fig. 15c). The difference between (c), where the boreholes are treated as soft probability data, and (d), where the soft probability 3D grid is used (Fig. 7c), is most pronounced when considering the realization results on a larger scale than the shown in the profile in Fig. 15.

A further example highlighting the effects of the probability grid is shown in Fig. 16, where a long SW-NE profile through each of the tested realizations (Table 1) is inspected (for location, see Fig. 14). Borehole data are located further away than

the voxel size (100 m) and is therefore not shown on the profile. Again, the overall structural pattern is comparable between the different realizations. As also observed in the horizontal view (Fig. 14), and as it must be, the results in (b), (c) and (d) are fixed within the buffer zone around the Tønder model, where the hard conditioning information is used. For exactly the same reason, the realizations in (b), (c) and (d) perfectly match also where the profile crosses the seismic lines. Since the differences in the constraining strategy are related to the way the borehole data are handled, only little structural dissimilarities are seen when considering this specific profile with no borehole data. Those differences are therefore mainly controlled by the trend imposed by the soft probability grid in (d). Hence, the most pronounced difference is the higher clay content in the south-western part of (d) compared to the northeast. This spatial variation in the clay content from the west to the east is even clearer in the 3D view of the realization shown in Fig. 17. The kriged sand probability is effective in enforcing the proper spatial trend on the realization. This is evident, not only from the comparison between Fig. 17b and Fig. 8, which allows verifying, voxel-by-voxel, the accordance between the soft conditioning distribution and the final corresponding realization, but also from the results in Fig. 18. In fact, Fig. 18 shows the cross-correlations (see, for instance, Stoica and Moses, 2005) between the soft probability distribution (Fig. 8) and each of the realizations visible, respectively, e.g., in Fig. 14c and Fig. 14d (see, also, Table 1, cases (c) and (d)). As expected, the correlation with the realization (d) has a much higher and more pronounced maximum.

It is probably important to stress that our conclusions are general as all the differences highlighted for each conditioning setup are not realization-dependent. This means that they would appear consistently for every realization obtained with the same conditioning setting. For clarity, this aspect, together with the variability within each probabilistic model, are further discussed and analysed in Appendix A (and the associated Fig.s 19-22).

7 Discussion

If MPS modelling is applied to large study areas, it is typically necessary to divide the area into different regions or domains and use different TIs to properly describe the geology in each individual region. In the present study, the entire investigated area was divided into four main, vertically subdivided, units: the Quaternary, the Måde Group, the Miocene and the Paleogene. The Miocene sequence was chosen for our study since it is relatively stationary from a statistic point of view and can be easily divided into two main facies: sand and clay.

One of the challenges was the presence of the significant graben structure, which offsets the Miocene layers. In the current study, the graben structure is only visible through the morphological shape of the top and bottom of the model domain, but in the MPS realizations this has been ignored. In future studies, faults should be handled, such that the MPS simulation results are affected across the faults. A possible solution could be to use the geochron formalism (Mallet, 2004), in which the simulation could be performed in a regular grid, with geo-time as y-axis. Realizations should then be converted into depth using a geo-time to depth conversion.

The generation of 3D TIs that produce desired patterns is time consuming and difficult, which may be why many of the previous studies are based on 2D or quasi-3D TIs (Comunian et al., 2012; Cordua et al., 2016; Feyen and Caers, 2006; Strebelle, 2002). In this study, the 3D TI was created in the modelling software GeoScene3D. Since GeoScene3D is specifically designed for 3D geological modelling and hosts tools for manual voxel modelling (Jørgensen et al., 2013), the

5 creation of the TI was relatively easy and straightforward. When creating the TI in this manner, the main focus was to represent the expected geological structures and not to make it stationary. In theory, MPS implementations assume stationary TIs (Liu et al., 2004b) and MPS application to strongly non-stationary systems is currently an area of research (e.g. Honarkhah and Caers, 2012; Straubhaar et al., 2011; de Vries et al., 2009; Cuhgunova et al. 2008). The TI used for this study was generated by following the unique criterion that the unconditioned simulation (obtained by using SNESIM

10 through its implementation in SGEMS) could satisfactorily reproduce the expected geological structures. Ideally, the TI should be constructed independently from the choice of MPS algorithm, and the realizations obtained by using a specific TI should have the same spatial variability as formalized by the training image. In practice though, this is rarely the case. While a specific part of the spatial statistics may be accurately reproduced, the realizations may lack geological features that can be crucial for subsequent modelling and interpretation. This is why the choice of MPS algorithm, and the parameters used to

15 run the MPS algorithm have significant impact on the spatial structures seen on generated realizations. Hence, in practice, structural modelling should consist of choosing a TI together with a specific MPS algorithm (and the associated modelling parameters) to generate realizations capable to reflect the spatial variability that appears to be realistic from a geological perspective (Liu, 2006). Thus, the development of an effective TI involves an iterative procedure, where the realizations should be tested and evaluated. It is worth noting that even if the developed TI is not stationary, it is interpreted as stationary

20 by the algorithm we have used (SNESIM; Strebelle, 2000). The realizations showed a significant sensitivity to the actual choice of the 3D TI (Fig. 10). Thus, we highly recommend a careful evaluation of the unconditioned simulation results and subsequent, consistent TI optimisation.

A strategy to include available information into the stochastic simulation is via hard conditioning. Through hard conditioning, the realizations are forced to perfectly match the provided data. In this study, hard conditioning was used to

25 ensure a perfect correspondence between the simulation results and a former geological model available for the Tønder area. The results illustrate that this goal can be successfully reached even though the influence of the hard conditioning data remains quite local.

Also the seismic data, represented by interpretations along the seismic lines, have been treated as hard conditioning data. This can be debated since seismics is an "indirect" geophysical method that is inherently affected by uncertainty. The

30 reflections observed on seismic data represent changes in seismic velocity and/or density, but they are not necessarily related to lithological variations. Furthermore, the quality of the data can be highly varying as the resolution capability depends on the depth and, due to the uncertainty of seismic velocities used for depth conversion, also depths in the seismic sections are uncertain. In the present case, it was decided to use the interpretations along the seismics as hard conditioning data as their

level of uncertainty is assumed to be much lower than the other available data (Kristensen et al., 2015), especially at the scale required for the simulation.

Because of technical and economic limitations, the application of reflection seismics for shallow hydrostatgraphic studies is relatively recent (probably the first examples can be traced back to the '80s). However, as a consequence of the increasing and general awareness concerning the use and protection of water resources, during the last 40 years, seismics - together with several other geophysical techniques (e.g., the airborne electromagnetic methodologies) - has been applied to many, diverse hydrogeological characterizations with varying results (e.g., Francese et al. 2005; Giustiniani et al, 2008). Several of the problems in this kind of surveys lie: (i) in the presence of anthropic noise; (ii) the fact that a possible shallow water table may reflect the majority of the energy and, at the same time, mask low velocity geological features, preventing the effective reconstruction of deeper structures. Actually, in the attempt to overcome these difficulties, it has become more and more common to process and invert what is still often considered noise: the ground-roll (Strobbia, 2009). In fact, ground-roll contains valuable information about the share velocity distribution in the subsurface and is generally characterized by high-amplitude. Recently high-resolution, shallow techniques based on surface waves have been developed and tested successfully for hydrogeological investigations (Vignoli et al., 2012; Vignoli et al., 2016). A possible limitation of techniques based on surface waves concerns the availability of low frequency sources: in presence of slow sediments, low frequencies are required to reach the desired depths, but generating them is very demanding and potentially detrimental to the seismic vibrator.

In the present research, the use of the borehole data as hard conditioning data was considered suboptimal due to the low quality of many of the wells and the different discretization of the borehole data (1 m) compared to the size of the realization grid (5 m). The borehole data were therefore translated into soft probabilities by using a moving window strategy that, in one shot, takes into account the borehole information uncertainty and the different scale issue. A uniform uncertainty of 20 % has been assumed for the boreholes across the entire domain; however, in future studies, it would be straightforward to extend the present approach and locally rescale the soft probabilities according to a quality rate of each individual borehole as, e.g., the one presented in He et al. (2014). In this way, poor quality borehole would influence the simulation less than more reliable data.

One of the difficulties in the development of a proper MPS realization of the Miocene sequence was that the sand content varied spatially across the model domain making the sequence non-stationary. A 3D sand probability grid (Fig. 8) was generated in order to migrate the information further from the boreholes and constrain the geostatistical simulation to follow the spatial sand/clay trend characterizing the study area. The soft probability grid can consequently be seen as a shortcut to address the non-stationarity, such that the sand/clay content derived from the TI was overruled by the probability grid. Another possible solution could consist in the creation of different TIs to represent the end-members, and then interpolate these to obtain gradual changes dependent on the positions as discussed in Mariethoz and Caers (2014). Actually, if SNESIM could correctly handle soft data in the form of localized conditional probability, kriging the borehole probabilities would be unnecessary (Hansen et al., submitted).

In addition to the data already incorporated as (soft and hard) conditioning data in this study, dense electromagnetic (EM) surveys, on approximately half of the area, are available in the Danish geophysical database Gerda (Møller et al., 2009). An option would be to use these resistivity data in the MPS modelling as soft conditioning data, such that high electric resistivities indicate sand, while clay corresponds to low resistivity values (see, for instance, He et al., 2014; He et al. 2016).

5 However, EM data have a limited resolution capability towards thin layers, especially in the deeper parts of the investigated sequences. The modelled unit is generally present at great depths (from a range between 30 m and 80 m, in the east, to a range between 150 m and 170 m, in the west) and the resolution of the EM data is consequently quite low in most of the area. This is especially the case in the west, where the Miocene unit is located below the depth of investigation (Christiansen and Auken, 2012). Another challenge regarding the use of resistivity data for detecting sand and clay within the Miocene

10 deposits is the common occurrence of silt (Rasmussen et al., 2010). Some of the clayey formations are very silty and, while silt has small grain sizes and hydraulic conductivities, it has high resistivities. Thus, clayey formations with high silt contents might show higher resistivities than expected, leading to potentially wrong interpretations. In addition to that, during the borehole description phase, silt is often recognised as clay, and this clearly causes a mismatch between the borehole and resistivity information.

15 Geostatistical simulation methods are most commonly used to create multiple realizations whose variability represents the combined (uncertain) information; for instance, with the purpose to estimate uncertainties of structural variability (Feyen and Caers, 2006; He et al., 2013; Poeter and Anderson, 2005; Refsgaard et al., 2012) or to make probability calculations to be used for various forecasts (Stafleu et al., 2011, Christensen et al., 2017). In the majority of this study (see Appendix A), only one realization of each of the conditional strategies was presented. The reason is two-fold. Firstly, the primary goal of this

20 paper is to describe a workflow for choosing a training image, and to combine all available information into one consistent probabilistic model. And each individual realization (e.g., in Fig. 17) is, by construction, compatible with all simulation inputs (thus, the statistics from the TI, the hard data, and the soft conditioning). Secondly, the realization generated in the final test (Fig. 17) was, at the end, incorporated in an overall geological model (Meyer et al., 2016): whereas, in this case, using a single realization is considered acceptable since the objective is to study large-scale groundwater flow and saltwater

25 intrusion, the use of multiple realizations for the propagation of uncertainty into, for example, hydrological models would be outside the scope. To make groundwater predictions on a large scale, the overall distribution and connectivity of the overall structures are crucial, while the precise location of the individual structures is less important. In contrast, detailed studies, like catchment analyses, would not make sense based on a single realization (He et al., 2013) and assessments based on a significant number of different realizations should be conducted to investigate the variability of the outcome.

30 The validity of the presented workflow is demonstrated for the Miocene unit characterized by relative simplicity and presence of only two categories. This does not mean that the applicability of this approach should be limited to simple situations. This research can be considered a proof of concept and its intent is to clearly show the relevance and effectiveness of our strategy in addressing the difficulties frequently encountered, even in simple cases. We do not see any particular difficulty in extending the proposed strategy to more complex settings characterized, for example, by a larger number of

categories. In fact, if this is the case, dealing with three or more categories makes the preparation of the hard conditioning data (e.g., the interpretations of seismic lines) clearly more laborious, but conceptually not more difficult. The same is true for the way we handle the borehole as soft conditioning data: definitely, the implementation of the sliding window and the following kriging procedure are not different if a larger number of categories are involved. Finally, MPS approaches
5 (together with the associated TIs) are already routinely used in situations with more than two categories (e.g., Jones et al. 2013), and, in our approach, special emphasis is placed uniquely on the strategy for the development of effective TIs via careful analyses of the unconstrained realizations.

8 Conclusions

This study investigates strategies for MPS simulations in large 3D model domains consistent with different types of input
10 data. The strategies were tested within an area of $2\,810$ km^2 in which the Miocene unit was modelled using MPS simulation. This part of the model was chosen since the Miocene can be effectively subdivided into few categories (i.e.: sand and clay) and is relatively stationary in the investigated area. An already existing and detailed geological model (the Tønder model) was present in a part of the study area, and was used in the final comprehensive geological model. A 3D TI was constructed based on the well-known geology of the unit. The stochastic simulations were conducted using the SNESIM algorithm as it
15 is implemented in SGeMS. The final TI was developed iteratively by checking the outcomes of the corresponding unconditioned simulations, and adjusting it in order to obtain the most geologically meaningful structures in the final realizations. The previously published Tønder model and reliable seismic interpretations were used as hard conditioning data in order to preserve the associated information during the simulation. On the other hand, the boreholes were incorporated into the simulation workflow through different conditioning strategies. The first approach - the most traditional one -
20 consisted of using the boreholes as hard conditioning data. This quite standard approach was not satisfactory as borehole data can have a high degree of uncertainty. Hence, via a moving window strategy, the lithological information in the borehole was translated into a probability distribution that could address uncertainties in borehole data both from inaccuracies and scale mismatch. Unfortunately, SNESIM limits the influence of soft conditioning data to local neighbourhoods around each data value and is unable to effectively regionalize any trends that might be captured. To better
25 address this problem, we kriged the sand probability derived from the boreholes into a 3D voxel model and used that kriged sand probability as soft conditioning. By using this last approach, we managed to successfully reproduce the sand/clay trend across the simulation domain evaluated based on a visual inspection. The study shows a practical workflow to properly build TI and effectively handle input information to be successfully used for large-scale geostatistical modelling.

9 Acknowledgements

The study is a part of the research project ERGO: "Effective high-resolution geological modelling", which is funded by the Innovation Fund Denmark. In the ERGO project, 3D geological modelling tools have been developed and integrated into the software package GeoScene 3D. The study is also connected to the Geocenter project SaltCoast: "Large-scale hydrogeological modelling of saltwater intrusion in a coastal groundwater system".

We are grateful to Wolfgang Scheer, who kindly provided us the borehole information from the German part of the study area and to Torben Bach, I-GIS, and Ingelise Møller, GEUS, for fruitful discussions.

We would like to thank Anders Juhl Kallesøe, who contributed to the project by modelling the Top Pre-Quaternary surface. Finally, we would like to express our gratitude to DICAAR at University of Cagliari for the computational resources used to generate Fig.s 19-22.

References

Carrera, J., 1993, An overview of uncertainties in modelling groundwater solute transport: Journal of Contaminant Hydrology, v. 13, no. 1-4, p. 23-48.

Chiles, J.-P., and Delfiner, P., 1999, Geostatistics: Modeling spatial uncertainty, Wiley.

Christensen, N. K., Minsley B. J., and Christensen S., 2017, Generation of 3-D hydrostratigraphic zones from dense airborne electromagnetic data to assess groundwater model prediction error: Water Resources Research, v.53, p. 1019-1038.

Christiansen, A. V., and Auken, E., 2012, A global measure for depth of investigation: Geophysics, v. 77, no. 4, p. WB171-WB177.

Chugunova, T. L., and Hu, L. Y., 2008, Multiple-point simulations constrained by continuous auxiliary data: Mathematical geosciences, v. 40, p. 133-146.

Comunian, A., Renard, P., and Straubhaar, J., 2012, 3D multiple-point statistics simulation using 2D training images: Computers and Geosciences, v. 40, p. 49-65.

Cordua, K. S., Hansen, T. M., Gulbrandsen, M. L., Barnes, C., and Mosegaard, K., 2016, Mixed-point geostatistical simulation: A combination of two-and multiple-point geostatistics: Geophysical Research Letters, v. 43, no. 17, p. 9030-9037.

de Vries, L.M., Carrera, J., Falivene, O., Gratacós, O., and Slooten, L. J., 2009, Application of Multiple Point Geostatistics to Non-stationary Images: Mathematical Geosciences, v. 41, p. 29-42. The modelled unit is generally present at great depths Deutsch, C. V., 2002, Geostatistical Reservoir Modeling, Oxford University Press.

Feyen, L., and Caers, J., 2006, Quantifying geological uncertainty for flow and transport modeling in multi-modal heterogeneous formations: Advances in Water Resources, v. 29, no. 6, p. 912-929.

Francese, R., Giudici, M., Schmitt, D.R., and Zaja, A., 2005, Mapping the geometry of an aquifer system with a high-resolution reflection seismic profile: Geophysical Prospecting, v 53, p. 817-828.

Giustiniani, M., Accaino, F., Picotti, S., and Tivinella, U., 2008, Characterization of the shallow aquifers by high-resolution seismic data: Geophysical Prospecting, v. 56, p. 655-666.

Guardiano, F. B., and Srivastava, R. M., 1993, Multivariate geostatistics: Beyond bivariate moments, in Soares, A., ed., Geostatistics, Volume 1: Kluwer, Dordrecht, The Netherlands, Troia, p. 133-144.

Hansen, T. M., Vu, L. T., and Cordua, K. S., submitted, Integration of Geo-Information using sequential simulation: Geophysical Journal International.

Haralick, R. M. and Shapiro L. G., 1992, Computer and Robot Vision, Volume I, Addison-Wesley, Boston, MA.

He, X., Koch, J., Sonnenborg, T. O., Jørgensen, F., Schamper, C., and Refsgaard, J. C., 2014, Transition probability based stochastic geological modeling using airborne geophysical data and borehole data: Water Resources Research, v. 50, no. 4, p. 3147-3169.

He, X., Sonnenborg, T. O., Jørgensen, F., Høyer, A.-S., Møller, R. R., and Jensen, K. H., 2013, Analyzing the effects of geological and parameter uncertainty on prediction of groundwater head and travel time: Hydrology and Earth System Sciences, HESS, v. 10, p. 2789-2833.

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

Honarkhah, M., and Caers, J., 2012, Direct Pattern-Based Simulation of Non-stationary Geostatistical Models: Mathematical Geosciences, v. 44, no. 6, p. 651-672.

Hu, L. Y., and Chugunova, T., 2008, Multiple-point geostatistics for modeling subsurface heterogeneity: A comprehensive review: Water Resources Research, v. 44, no. 11, p. 1-14.

Huysmans, M., and Dassargues, A., 2009, Application of multiple-point geostatistics on modelling groundwater flow and transport in a cross-bedded aquifer (Belgium): Hydrogeology Journal, v. 17, no. 8.

Høyer, A.-S., Lykke-Andersen, H., Jørgensen, F., and Auken, E., 2011, Combined interpretation of SkyTEM and high-resolution seismic data: Physics and Chemistry of the Earth, v. 36, no. 16, p. 1386-1397.

I-GIS, 2014, Geoscene3D: Aarhus, Denmark.

Jones, P., Douglas, I., and Jewbali, A., 2013, Modeling Combined Geological and Grade Uncertainty: Application of Multiple-Point Simulation at the Apensu Gold Deposit, Ghana: Math. Geosci. v. 45, p. 949–965.

Jørgensen, F., Høyer, A.-S., Sandersen, P. B. E., He, X., and Foged, N., 2015, Combining 3D geological modelling techniques to address variations in geology, data type and density - an example from Southern Denmark: Computers and Geosciences, v. 81, p. 53-63.

Jørgensen, F., Lykke-Andersen, H., Sandersen, P. B. E., Auken, E., and Nørmark, E., 2003, Geophysical investigations of buried Quaternary valleys in Denmark: An integrated application of transient electromagnetic soundings, reflection seismic surveys and exploratory drillings: Journal of Applied Geophysics, v. 53, no. 4, p. 215-228.

Jørgensen, F., Møller, R. R., Nebel, L., Jensen, N.-P., Christiansen, A. V., and Sandersen, P. B. E., 2013, A method for cognitive 3D geological voxel modelling of AEM data: Bulletin of Engineering Geology and the Environment, v. 72, p. 421-432.

Kessler, T. C., Comunian, A., Oriani, F., Renard, P., Nilsson, B., Klint, K. E. S., and Bjerg, P. L., 2013, Modeling Fine-Scale Geological Heterogeneity—Examples of Sand Lenses in Tills: Groundwater, v. 51, no. 5, p. 692-705.

Kristensen, M., Vangkilde-Pedersen, T., Rasmussen, E. S., Dybkjær, K., Møller, I., and Andersen, L. T., 2015, Miocæn 3D, opdateret 2015; Den rumlige geologiske model: GEUS.

Liu, G., Zheng, C., and Gorelick, S. M., 2004a, Limits of applicability of the advection-dispersion model in aquifers containing connected high-conductivity channels: Water Resources Research, v. 40, no. 8.

Liu, Y., 2006, Using the Snesim program for multiple-point statistical simulation: Computers and Geosciences, v. 32, no. 10, p. 1544-1563.

Liu, Y., Harding, W., and Strebelle, S., 2004b, Multiple-point simulation integrating wells, three-dimensional seismic data and geology: American Association of Petroleum Geologists Bulletin, v. 88, no. 7, p. 905-921.

Maharaja, A., 2008, TiGenerator: Object-based training image generator: Computers and Geosciences, v. 34, no. 12, p. 1753-1761.

Mallet, J. L., 2004, Space–time mathematical framework for sedimentary geology: Mathematical Geology, v. 36, no. 1, p. 1-32.

Mariethoz, G., and Caers, J., 2014, Multiple-point Geostatistics: Stochastic Modeling with Training Images, Wiley.

McCarty, J. F., and Curtis, A. A., 1997, The impact of upscaling techniques in the geostatistical characterization of a hetereogeneous petroleum reservoir, The Fifth International Geostatistics Congress: Wollongong, Australia, p. 551-562.

Meyer, R., Sonnenborg, T. O., Engesgaard, P., Høyer, A.-S., Jørgensen, F., Hinsby, K., Hansen, B., Jensen, J. B., and Piotrowski, J. A., 2016, Effects of climate variability on saltwater intrusions in coastal aquifers in Southern Denmark, EGU General Assembly 2016, Volume 18: Vienna, Geophysical Research Abstracts. http://meetingorganizer.copernicus.org/EGU2016/EGU2016-458.pdf [accessed on 07 November 2016]

Møller, I., Søndergaard, V. H., Jørgensen, F., Auken, E., and Christiansen, A. V., 2009, Integrated management and utilization of hydrogeophysical data on a national scale: Near Surface Geophysics, v. 7, no. 5-6, p. 647-659.

Pérez, C., Mariethoz, G., and Ortiz, J. M., 2014, Verifying the high-order consistency of training images with data for multiple-point geostatistics: Computers & Geosciences, v. 70, p. 190-205.

Pickel, A., Frechette, J. D., Comunian, A., and Weissmann, G. S., 2015, Building a training image with Digital Outcrop Models: Journal of Hydrology, v. 531, p. 53-61.

Poeter, E., and Anderson, D., 2005, Multiple ranking and inference in ground water modelling Ground Water, v. 43, no. 4, p. 597-605.

Rasmussen, E. S., Dybkjær, K., and Piasecki, S., 2010, Lithostratigraphy of the Upper Oligocene – Miocene succession of Denmark: Geological survey of Denmark and Greenland Bulletin, v. 22, p. 92+99plates.

Rasmussen, E. S., Vangkilde-Pedersen, T., and Scharling, P., 2007, Prediction of reservoir sand in Miocene deltaic deposits in Denmark based on high-resolution seismic data: Geological Survey Of Denmark And Greenland Bulletin, v. 13, p. 17-20.

Refsgaard, J. C., Christensen, S., Sonnenborg, T. O., Seifert, D., Højbjerg, A. L., and Troldborg, L., 2012, Review of strategies for handling geological uncertainty in groundwater flow and transport modeling: Advances in Water Resources, v. 36, p. 36-50.

Remy, N., Boucher, A., and Wu, J., 2009, Applied Geostatistics with SGeMS, Cambridge, UK.

Renard, P., and Allard, D., 2013, Connectivity metrics for subsurface flow and transport: Advances in Water Resources, v. 51, p. 168-196.

Ronayne, M. J., S. M. Gorelick, and J. Caers, 2008, Identifying discrete geologic structures that produce anomalous hydraulic response: An inverse modeling approach: Water Resour. Res., v. 44, p. W08426, doi:10.1029/2007WR006635.

Stafleu, J., Maljers, D., Gunnink, J. L., Menkovic, A., and Busschers, F. S., 2011, 3D modelling of the shallow subsurface of Zeeland, the Netherlands: Netherlands Journal of Geoscience - Geologie en Mijnbouw, v. 90, no. 4, p. 293-310.

Stoica, P. and Moses R., 2005, Spectral Analysis of Signals. Prentice Hall, Upper Saddle River, NJ.

Straubhaar, J., Renard, P., Mariethoz, G., Froidevaux, R., and Besson, O., 2011, An improved parallel multiple-point algorithm using a list approach: Mathematical Geosciences, v. 43, p. 305-328.

Strebelle, S., 2000, Sequential Simulation Drawing Structures from Training Images: PhD thesis, Stanford University.

Strebelle, S., 2002, Conditional simulation of complex geological structures using multiple-point statistics: Mathematical Geology, v. 34, no. 1, p. 1-21.

Strobbia, C.L., Laake, A., Vermeer, P.L., and Glushchenko, A., Surface Waves - Use Them Then Lose Them, in Proceedings of 71st EAGE Conference & Exhibition, Amsterdam, 2009.

Ter-Borch, N., 1991, Geological map of Denmark 1:500.000. Structural map of the Top Chalk Group.

Vangkilde-Pedersen, T., Dahl, J. F., and Ringgaard, J., Five years of experience with landstreamer vibroseis and comparison with conventional seismic data acquisition, in Proceedings Symposium on the Application of Geophysics to Engineering and Environmental Problems (SAGEEP), Seattle, 2006.

Vignoli, G., Cassiani, G., Rossi, M., Deiana, R., Boaga, J.,and Fabbri, P., 2012, Geophysical characterization of a small pre-Alpine catchment: Journal of Applied Geophysics, v. 80, p. 32-42.

Vignoli, G., Gervasio, I., Brancatelli, G., Boaga, J., Della Vedova, B., and Cassiani, G., 2016, Frequency-dependent multi-offset phase analysis of surface waves: an example of high-resolution characterization of a riparian aquifer: Geophysical Prospecting, v. 64, n.1, p. 102-111.

Appendix A

Despite the discussion of the rest of the paper is realization-independent, it might be worth highlighting some of the effects of our approach on different realizations of the obtained probabilistic models. On this respect, Fig. 19 confirms what is in Fig. 11. Thus, the use of TI2, instead of TI1, consistently promotes the presence of larger sand bodies also when the

5   conditioning strategy (d) in Table 1 is adopted. In addition, the two panels of Fig. 19 demonstrate that similar size distributions are common for all the realizations generated with a specific TI. Analogous behaviours clearly appear also when we compare the other two properties analysed previously: eccentricity (Fig. 20) and jaggedness (Fig. 21). If we compare TI1's with TI2's results for each of the Fig.s 20 and 21, we can draw the same conclusion as for Fig.s 12-13. Hence, elongated and jagged sand structures are more probable in the realizations obtained by using TI1. And, again, within the

10   same probabilistic model, the distributions of eccentricity and jaggedness are very consistent between the realizations. Of course, this should not lead us to the wrong deduction that there is no variability between the elements of the probabilistic model. For example, Fig. 22 compares the conditioning approaches (c) and (d) in Table 1 in terms of e-type and variance maps: Fig.s 22b-c concern the conditioning strategy (c), while Fig.s 22e-f result from the application of approach (d). To make this comparison easier, also the corresponding (soft and hard) conditionings are shown in the same figure: in the panel

15   (a), for strategy (c) and, in the panel (d), for approach (d). Fig. 22 makes evident the role of the seismic lines and the buffer zone from the Tønder model (used as hard data in both conditioning strategies) in constraining all the realizations of both probabilistic models. This is particularly clear when we observe the vertical W-E sections (Fig. 22c-d): the only place characterized by an extremely limited variance is where the intersection with the seismic line occurs. Fig. 22 highlights also the crucial role of the 3D kriged grid of the sand probability distribution in migrating the information far from the boreholes.

20   This confirms, once more, the validity of the suggested way to handle the well information.
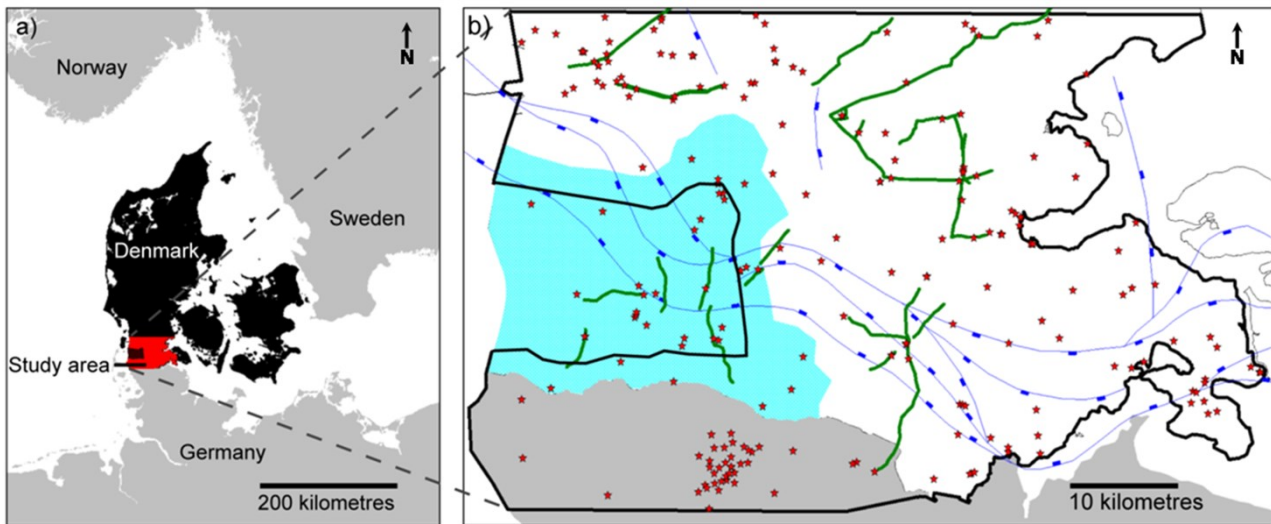
Figure 1: a) Location of the study area, b) Map of the model area (the solid black line) shown together with the data and the fault structures delimiting the Tønder graben: red stars indicate boreholes deeper than 100 m, green lines the position of the seismic lines; the Tønder graben structure is marked in blue (Ter-Borch, 1991). The turquois shading marks the model area of the Tønder model (Jørgensen et al., 2015); the grey zone in map (b) represents Germany.
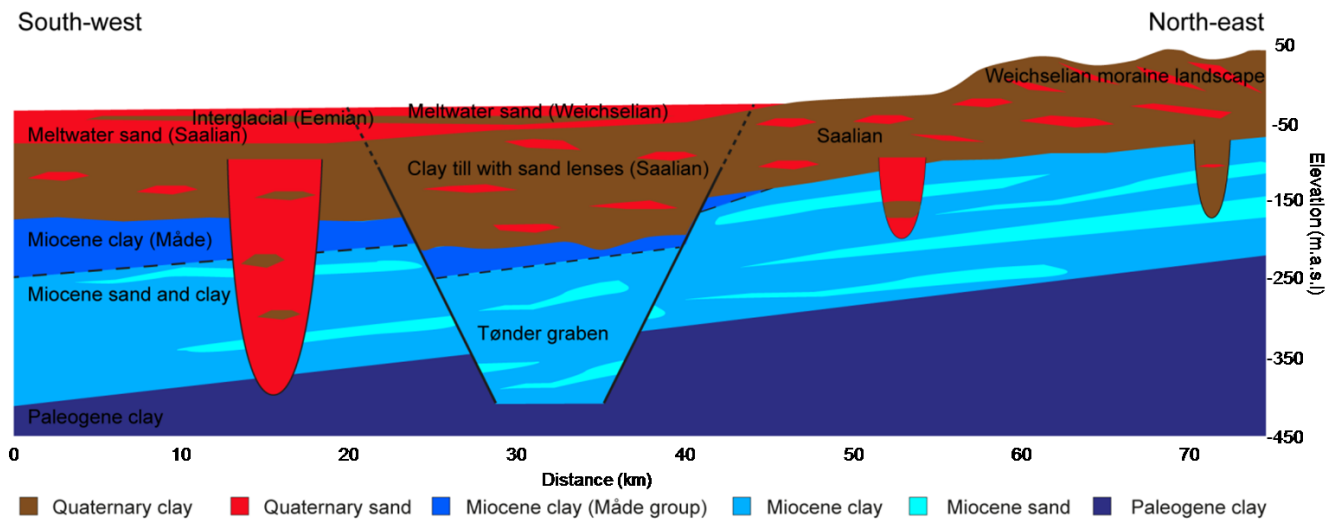
Figure 2: Conceptual sketch of the geology in the study area.
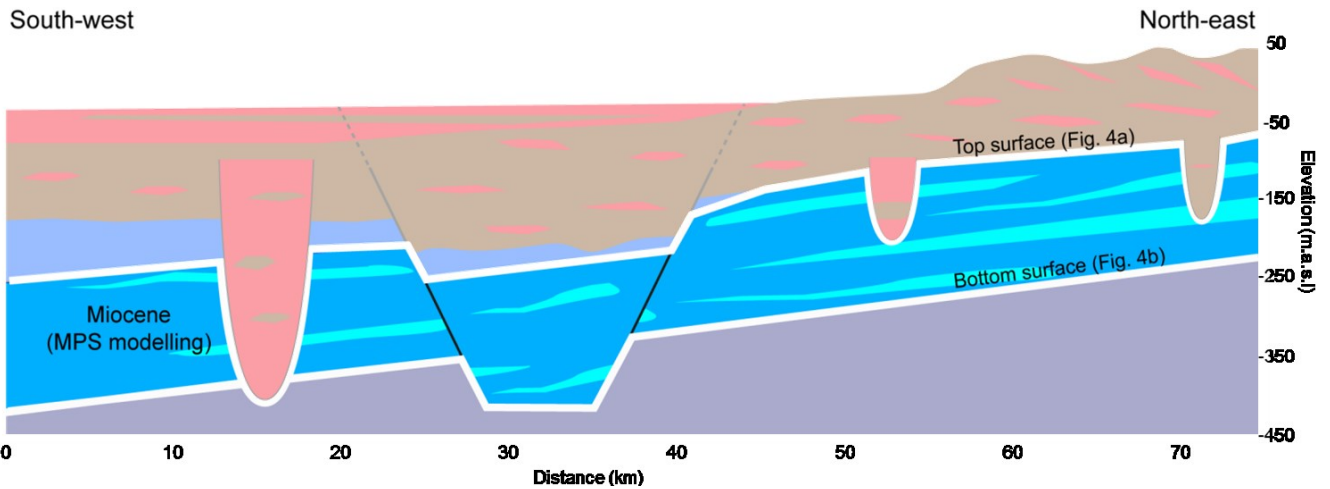
Figure 3: Sketch of the MPS model domain based on the conceptual sketch of the geology in Fig. 2: Top and Bottom of the Miocene unit is outlined by thick white lines. The surfaces are shown in Fig. 4a and 4b.
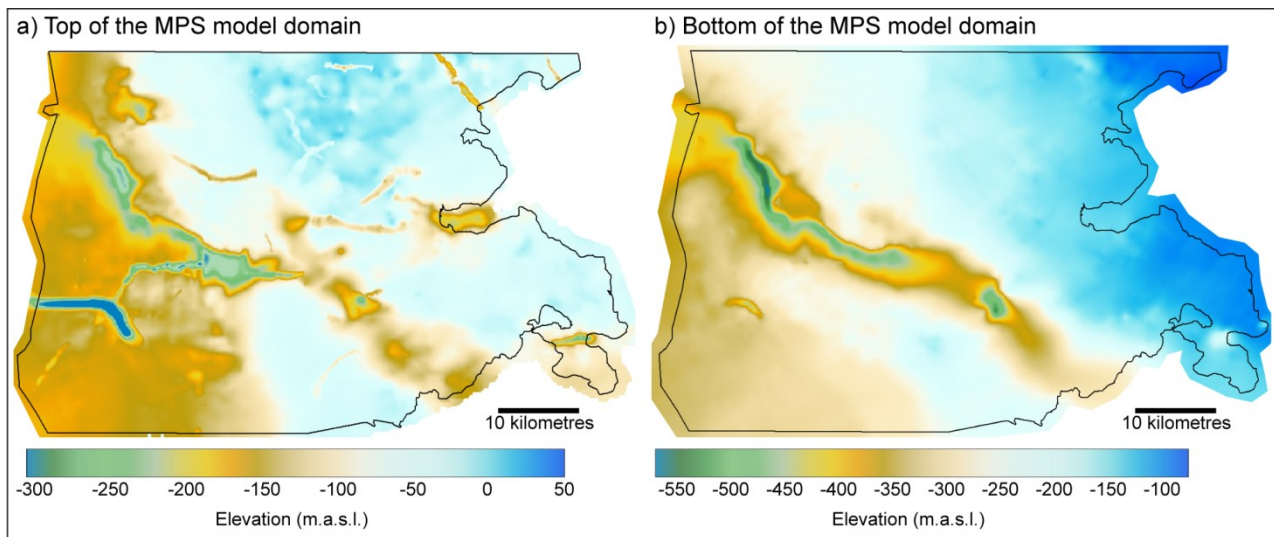
Figure 4: Elevation of the top and bottom of the MPS model domain (the Miocene unit). a) The modelled surface defining the top of the MPS model domain. The surface is obtained by merging the bottom of the Måde Group and the top pre-Quaternary (see Fig. 3). b) The modelled surface of the bottom of the MPS model domain, which consists of the Top Paleogene clay. Note: Different colour scales.
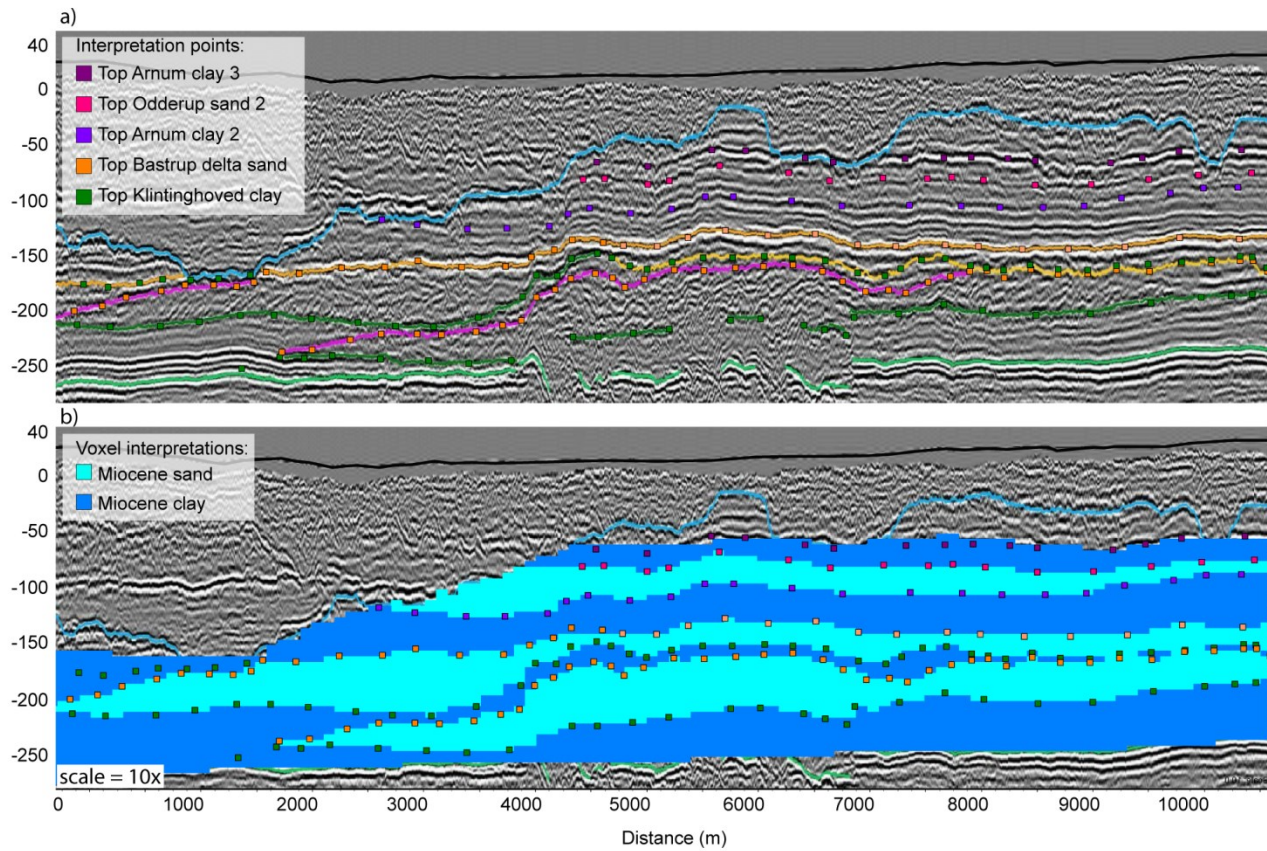
Figure 5: Preparation of seismic data input exemplified by a selected seismic section (for position, see Fig. 6). a) The seismic section shown together with coloured horizons and interpretation points derived from the Miocene model (Kristensen et al., 2015). b) The seismic section shown together with the generated "2D voxel model". Vertical exaggeration = 10x.
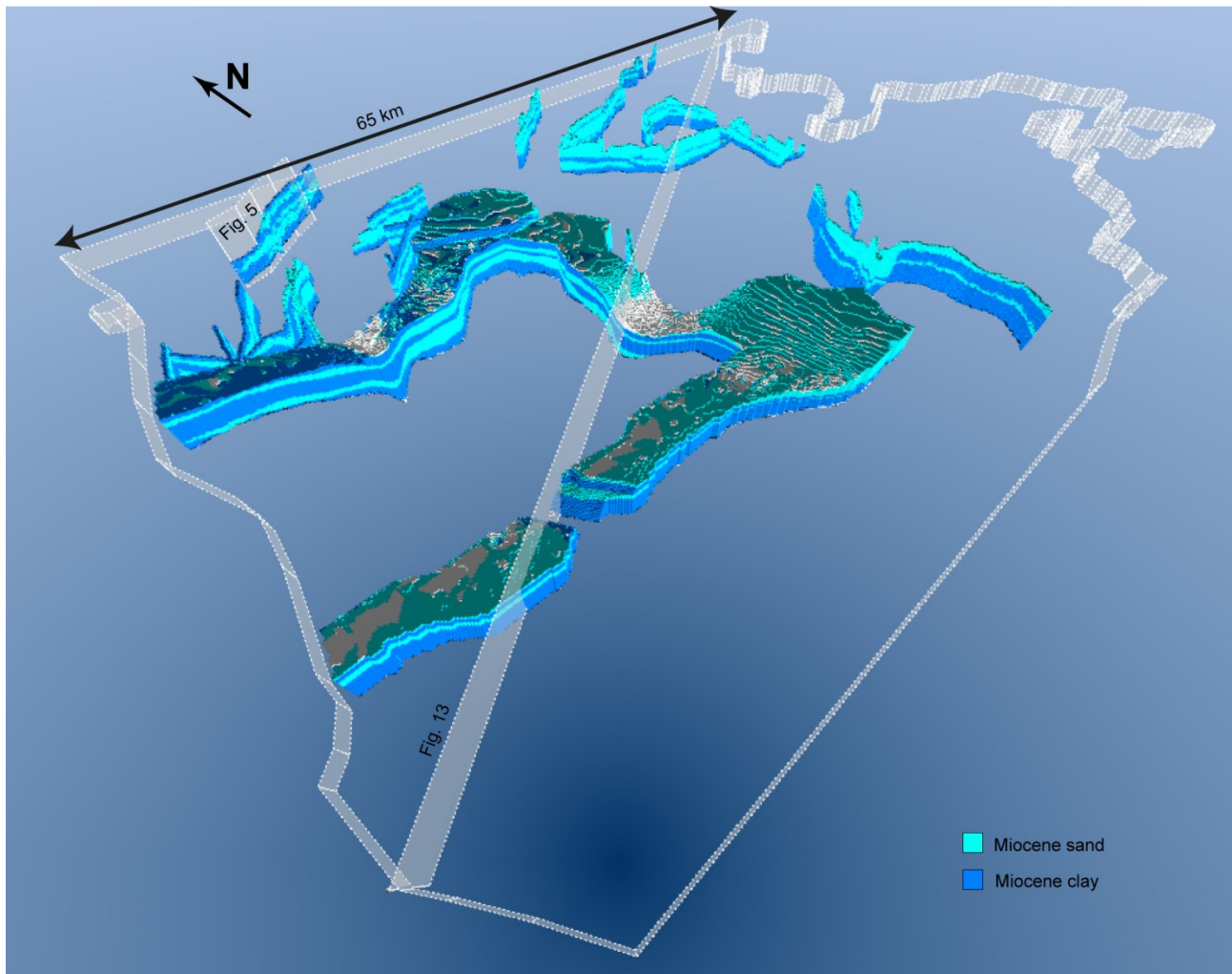
Figure 6: 3D view of the hard conditioning data (the buffer zone around the Tønder model and the seismic lines), seen from south-west. Positions of the seismic line in Fig. 5 and the profile in Fig. 16 are marked on the figure. Vertical exaggeration = 10x.
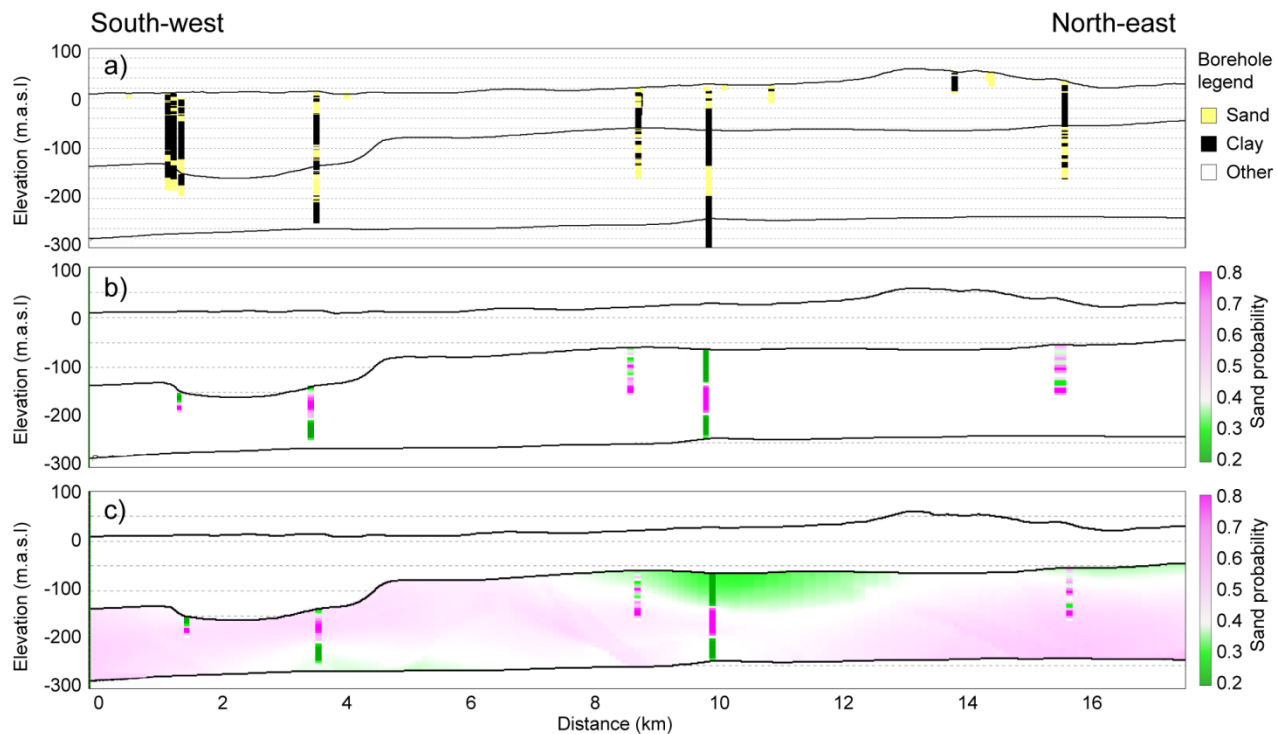
Figure 7: Three different ways to handle borehole information illustrated along a profile in the study area (for position, see Fig. 14). The resulting realizations are shown in Fig. 12. a) The sand/clay occurrence in the borehole is considered as hard data. b) The sand/clay occurrence is translated into sand/clay probability and then considered as soft data. c) The sand probabilities from the boreholes are kriged into a 3D grid to be used as soft conditioning. Vertical exaggeration 8x.
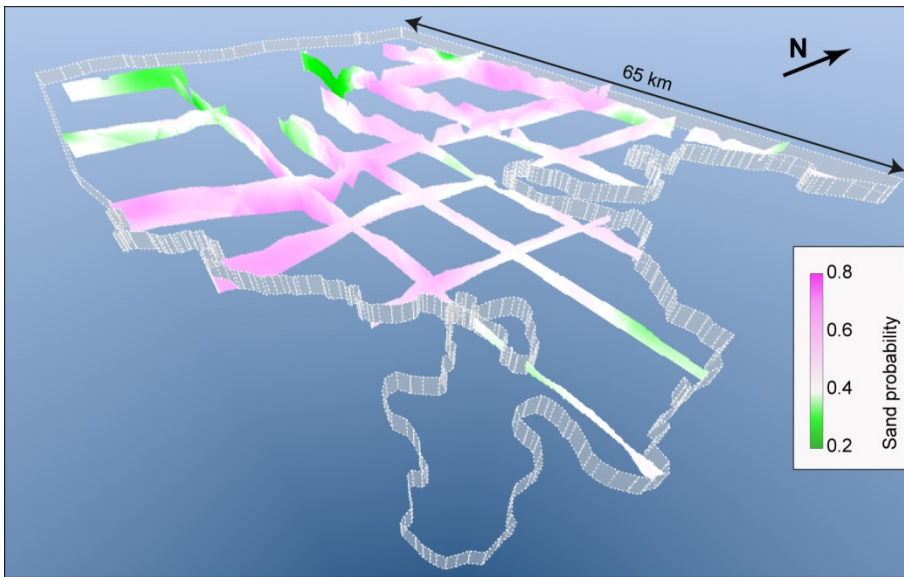
Figure 8: 3D fence view into the sand probability grid, seen from south-east. Vertical exaggeration = 10x.
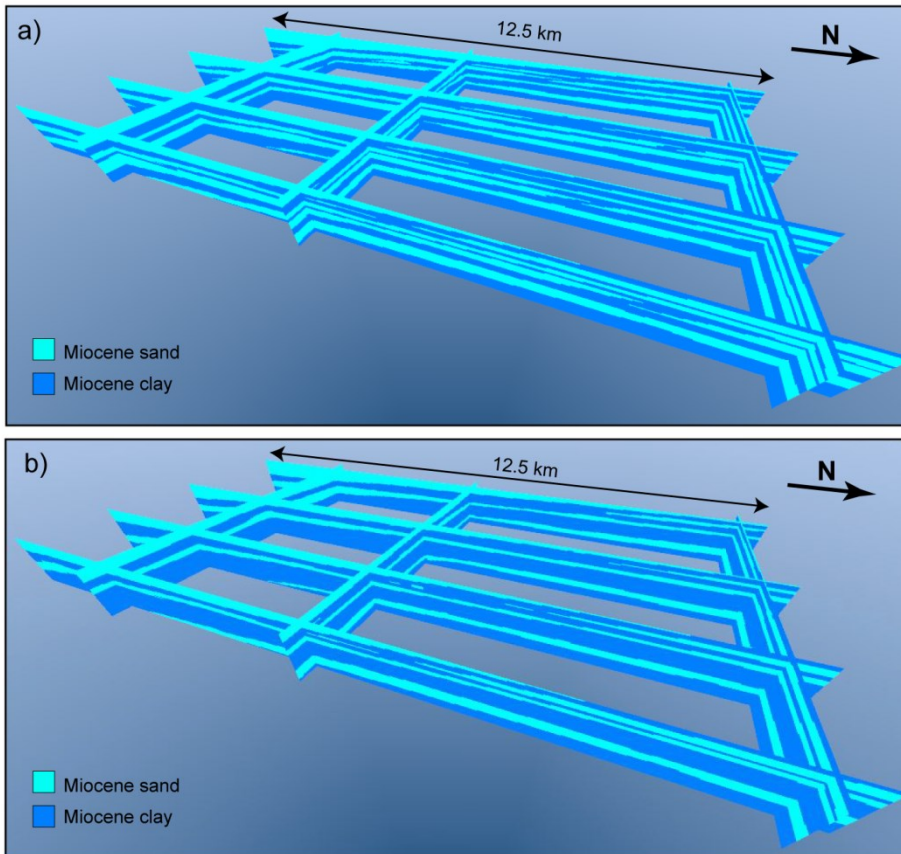
Figure 9: N-S and E-W slices through the two tested 3D training images seen from north-east. a) The first TI (TI1), used to generate the unconditioned realization in Fig. 10a. b) The second TI (TI2), used for the unconditioned realization in Fig. 10b. Vertical exaggeration = 3x.
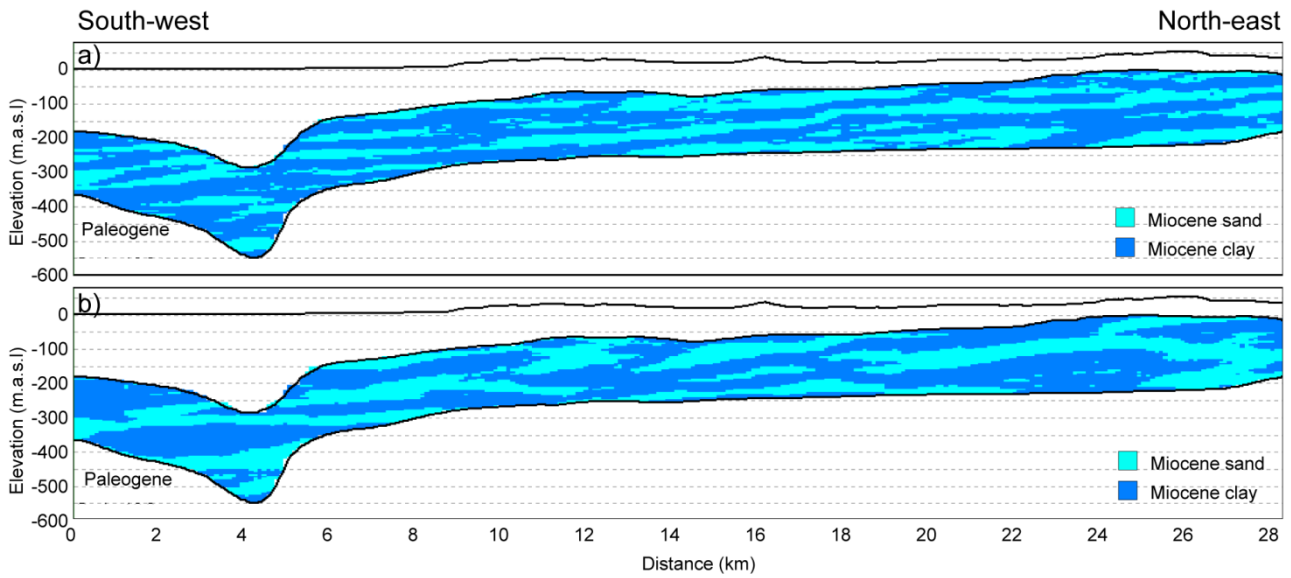
Figure 10: Profile through the unconditioned realizations obtained by using the two different TIs shown in Fig. 4. a) Realization based on TI1 (Fig. 9a). b) Realization based on TI2 (Fig. 9b). Vertical exaggeration = 8x. For location of the profile, see Fig. 14.
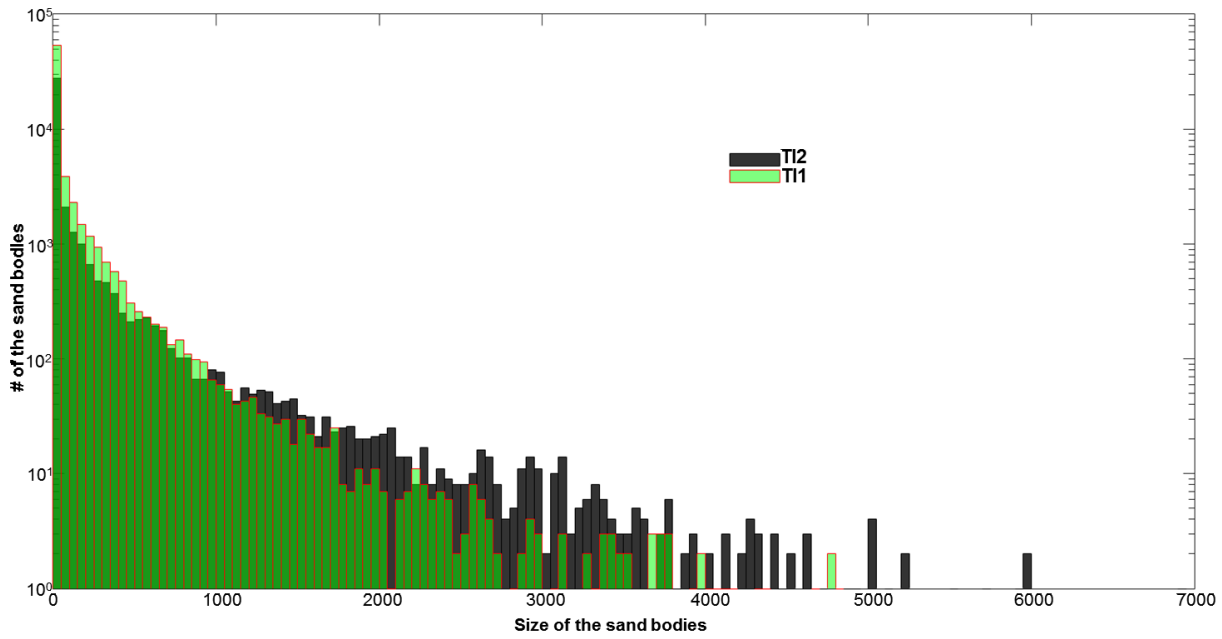
Figure 11: Sizes of the sand bodies of the unconditioned realizations obtained by using the two different TIs shown in Fig. 9.
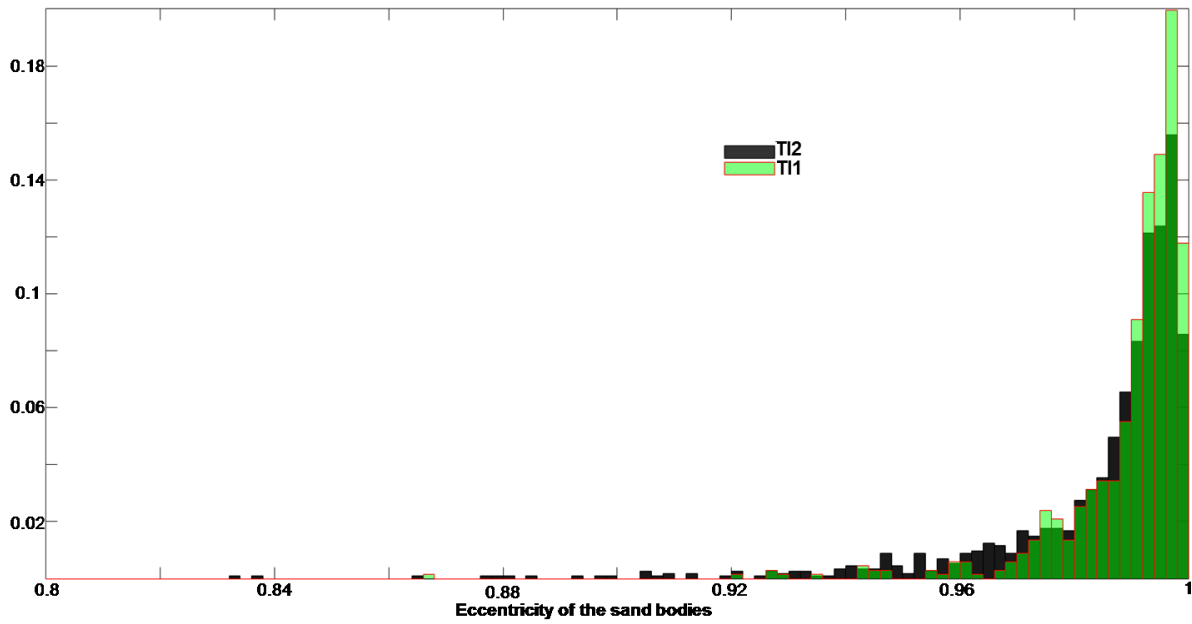
Figure 12: Probability of the eccentricity of the sand bodies (with a size larger than 1000) for the unconditioned realizations obtained by using the two TIs in Fig. 9.
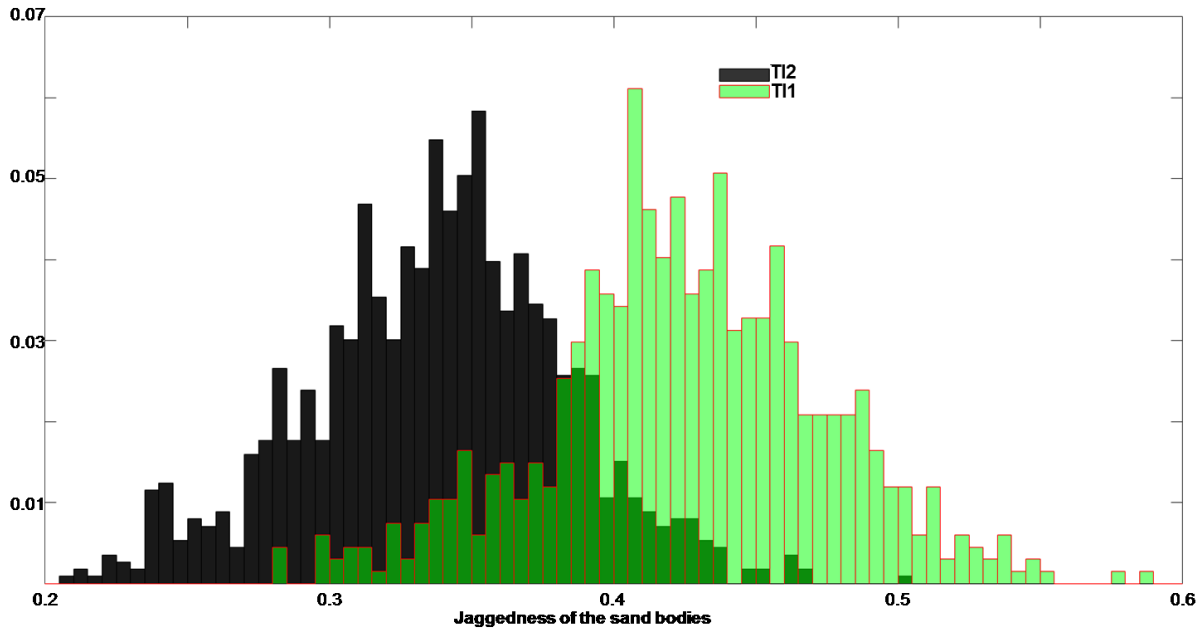
Figure 13: Probability of the jaggedness of the sand bodies (with a size larger than 1000) for the unconditioned realizations obtained by using the TI1 and TI2 in Fig. 9.
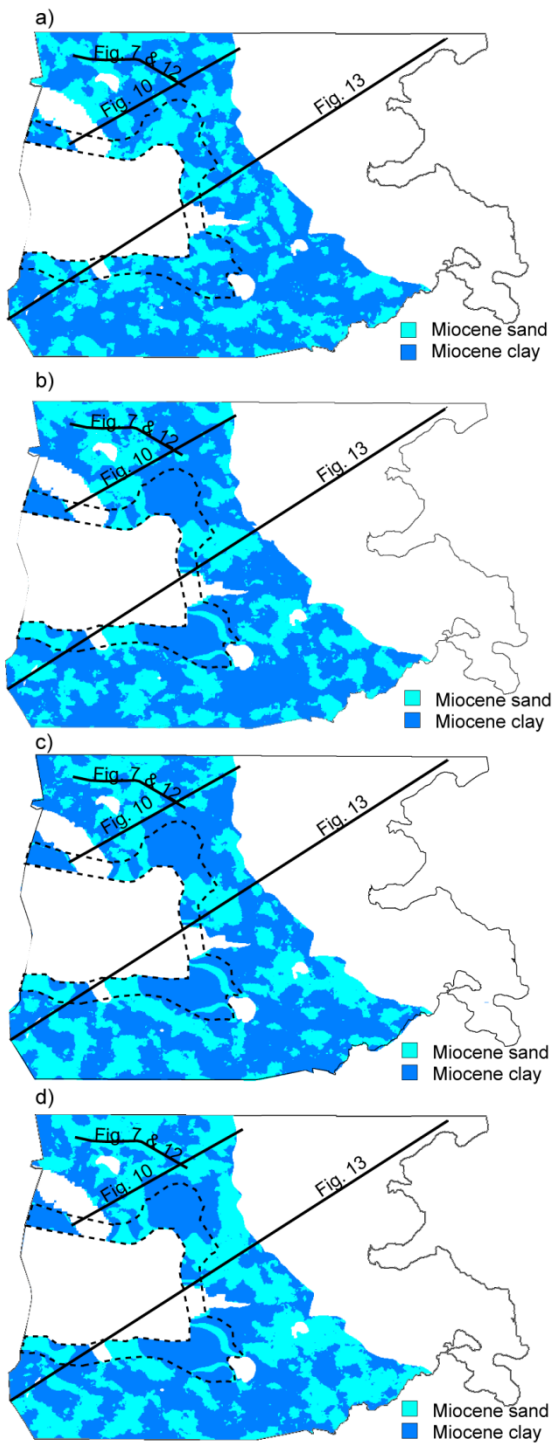
Figure 14: Horizontal slice at elevation -188 m through the realizations generated by using different conditional strategies (see Table 1). Position of the profiles in Figs. 7, 10 & 12 and 13 are showed. The dashed region encapsulates the buffer around the Tønder model, which has been used as hard conditioning information in (b), (c) and (d).
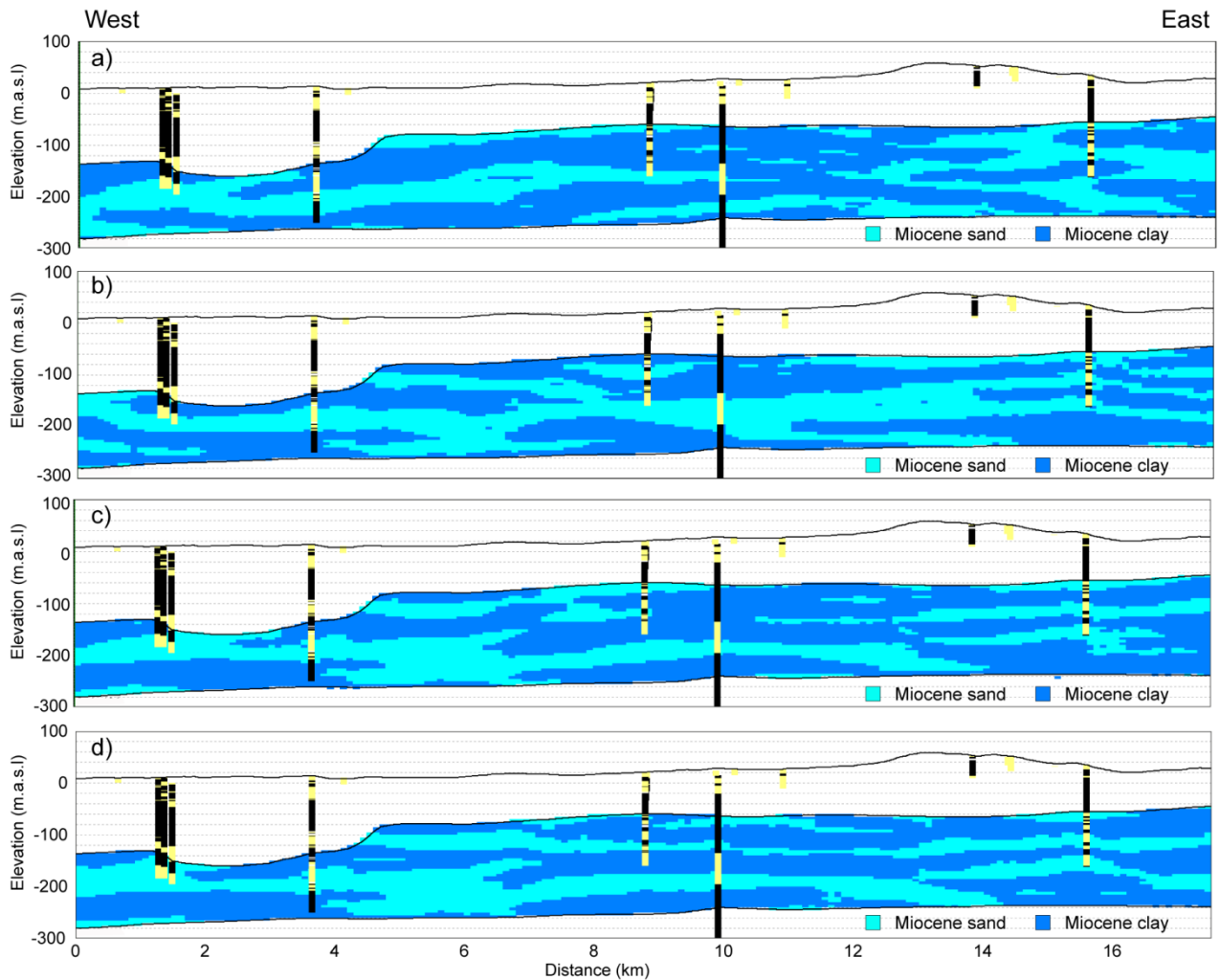
Figure 15: Profile through the realizations generated by using different conditional strategies (see Table 1). The borehole information used for conditioning is illustrated in Fig. 7: a) unconditioned; b) borehole as hard data (Fig. 7a); c) borehole as soft data (Fig. 7b); d) sand probability grid derived by boreholes as soft data (Fig. 7c). Sand/clay information from the original boreholes are showed within a buffer of 100 m (yellow = sand, black =clay, white = 'other'). For location of the profile, see Fig. 14. Vertical exaggeration = 8x.
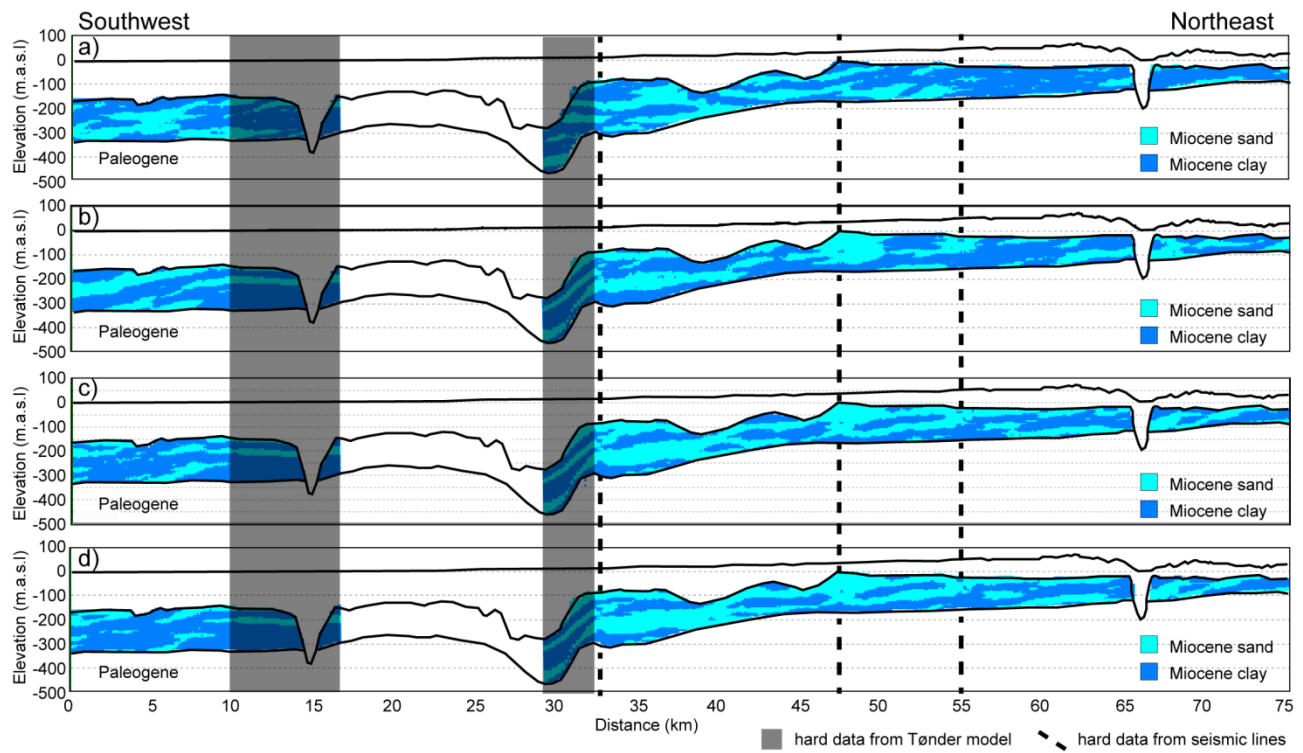
Figure 16: Long SW-NE profile through the realizations generated by using different conditional strategies (see Table 1). For location of the profile, see Fig. 14. Vertical exaggeration = 15x.
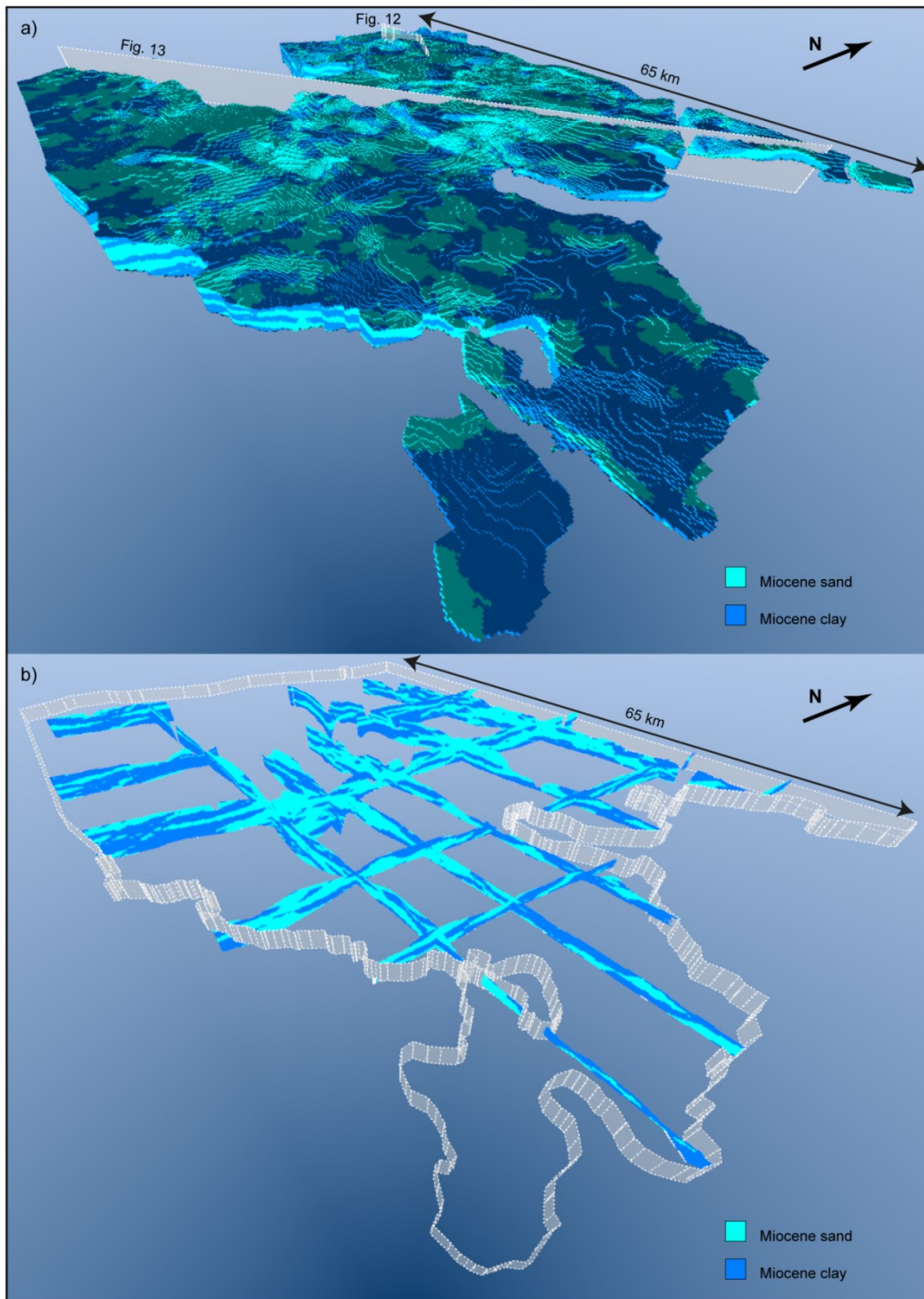
Figure 17: 3D view of the final realization (test (d), Table 1). The associated 3D probability grid is shown in Fig. 8. a) All voxels are plotted and the location of the profile in Figs. 12 and 13 are shown. b) The associated fence view. Vertical exaggeration = 10x.
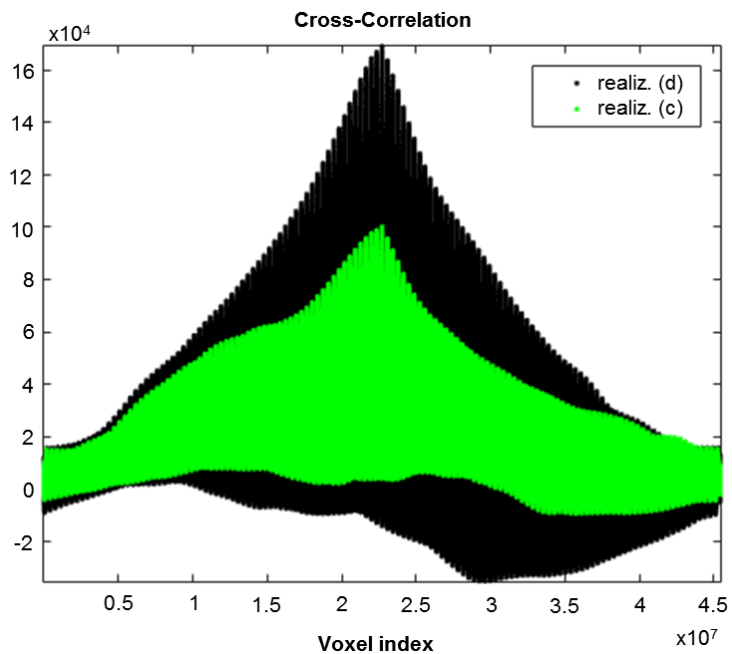
Figure 18: Cross-correlation between the soft probability distribution (Fig. 8) and each of the realizations showed, for example, in Fig. 14c (real. (c)) and Fig. 14d (real. (d)), and described in Table 1 (cases (c) and (d)).
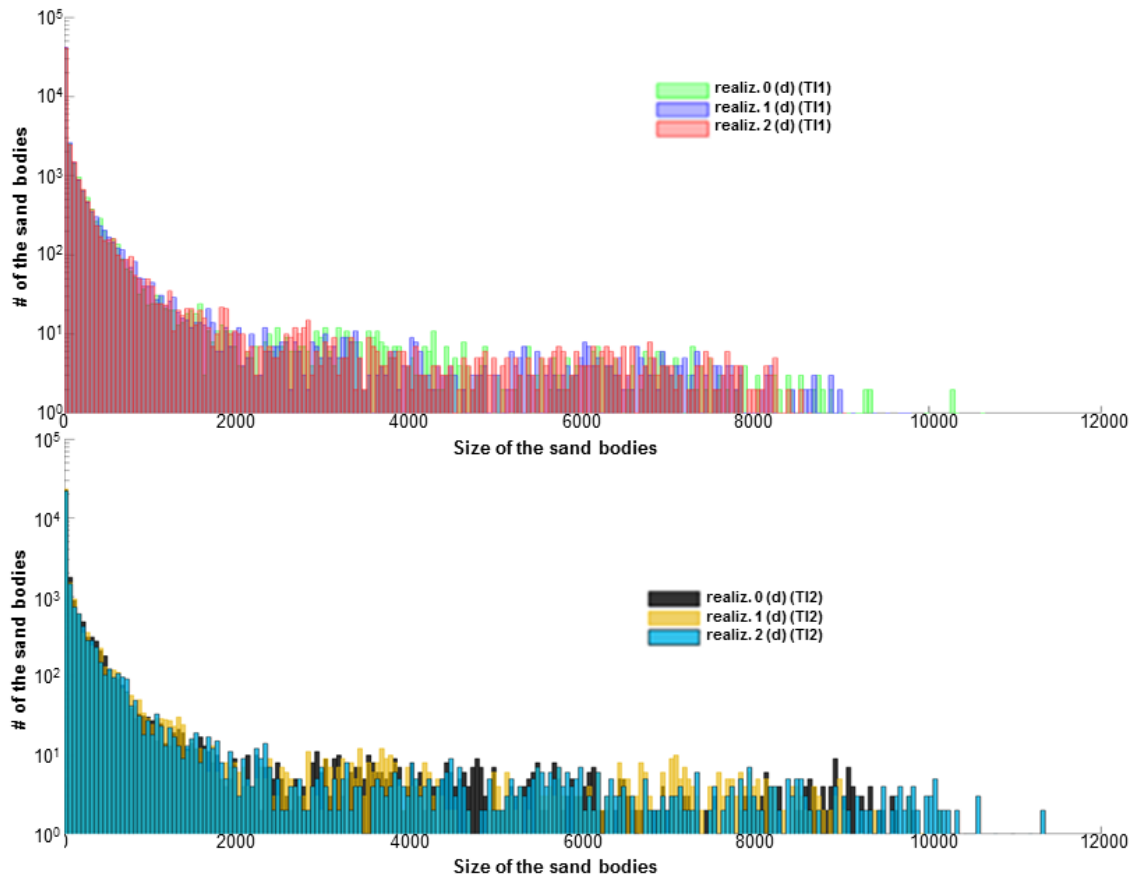
Figure 19: Sizes of the sand bodies for the realizations obtained by using the conditioning strategy (d) in Table 1 and, respectively, TI1 and TI2 (Fig. 9).
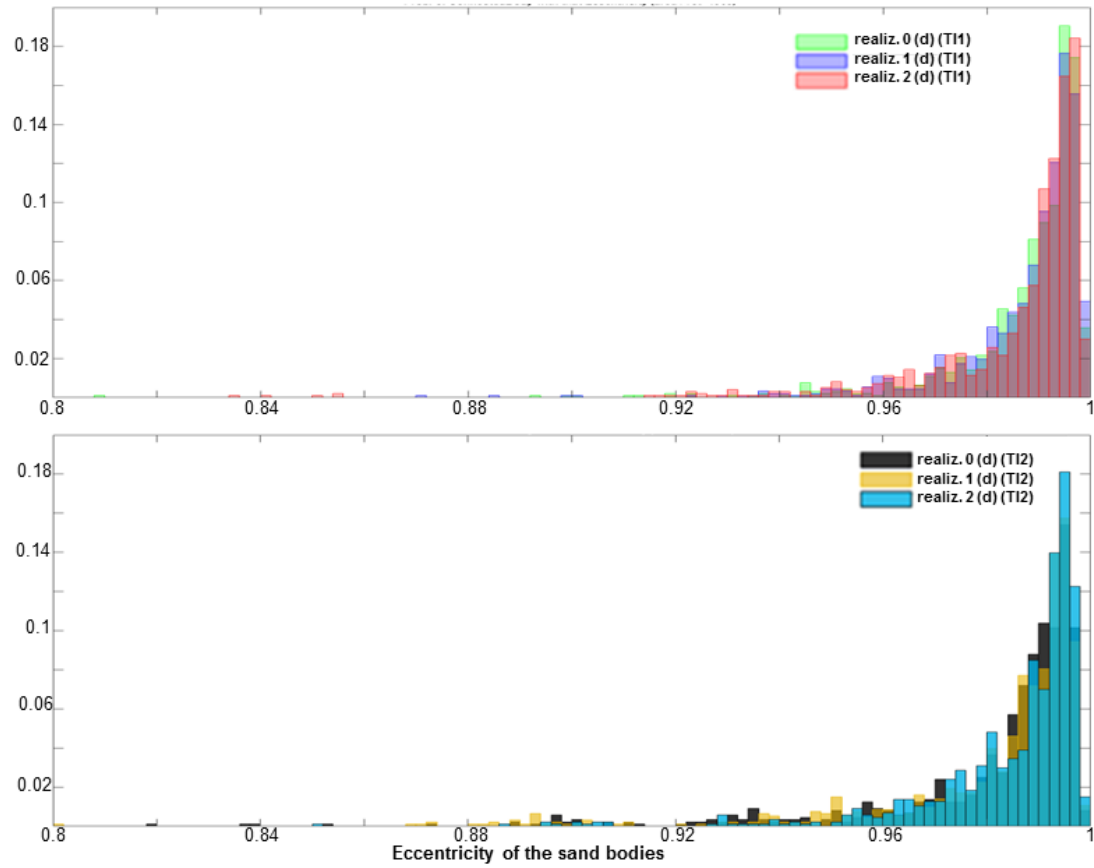
Figure 20: Probability of the eccentricity of the sand bodies (with a size larger than 1000) for the realizations obtained by using the conditioning strategy (d) in Table 1 and, respectively, TI1 and TI2 (Fig. 9).
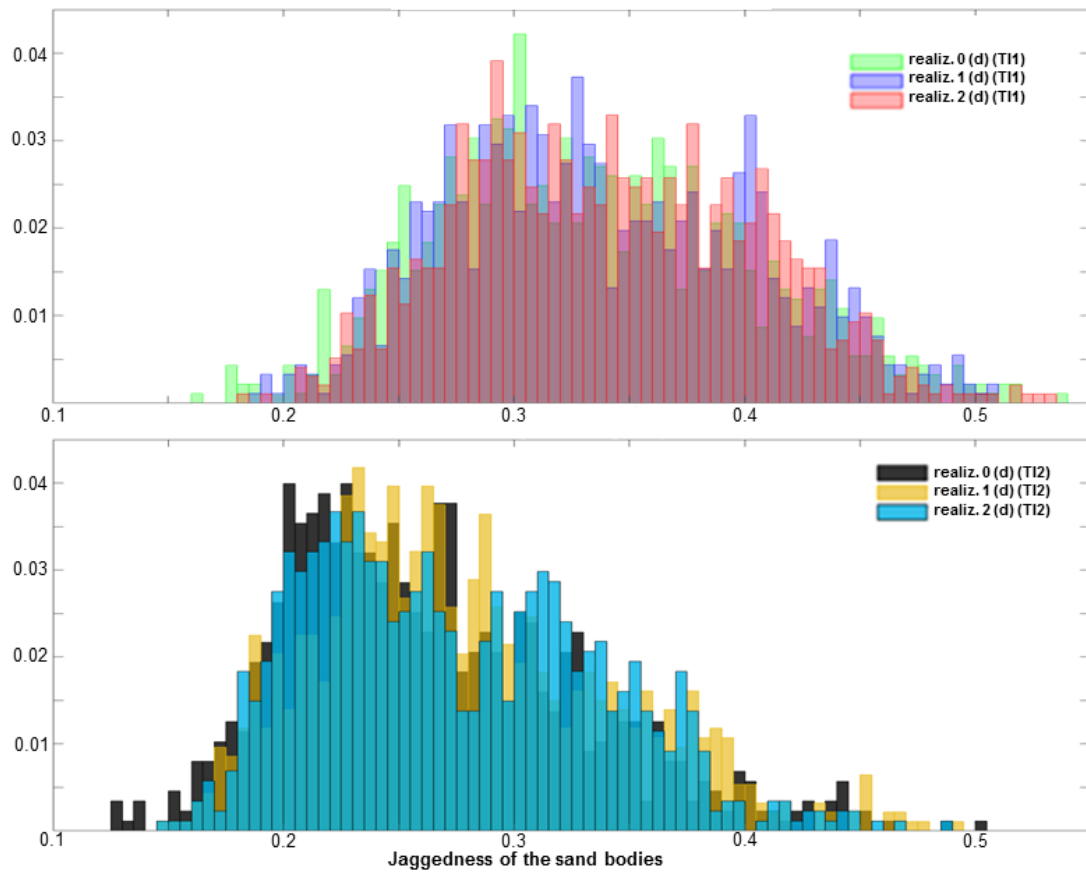
Figure 21: Probability of the jaggedness of the sand bodies (with a size larger than 1000) for the realizations obtained by using the conditioning strategy (d) in Table 1 and, respectively, TI1 and TI2 (Fig. 9).
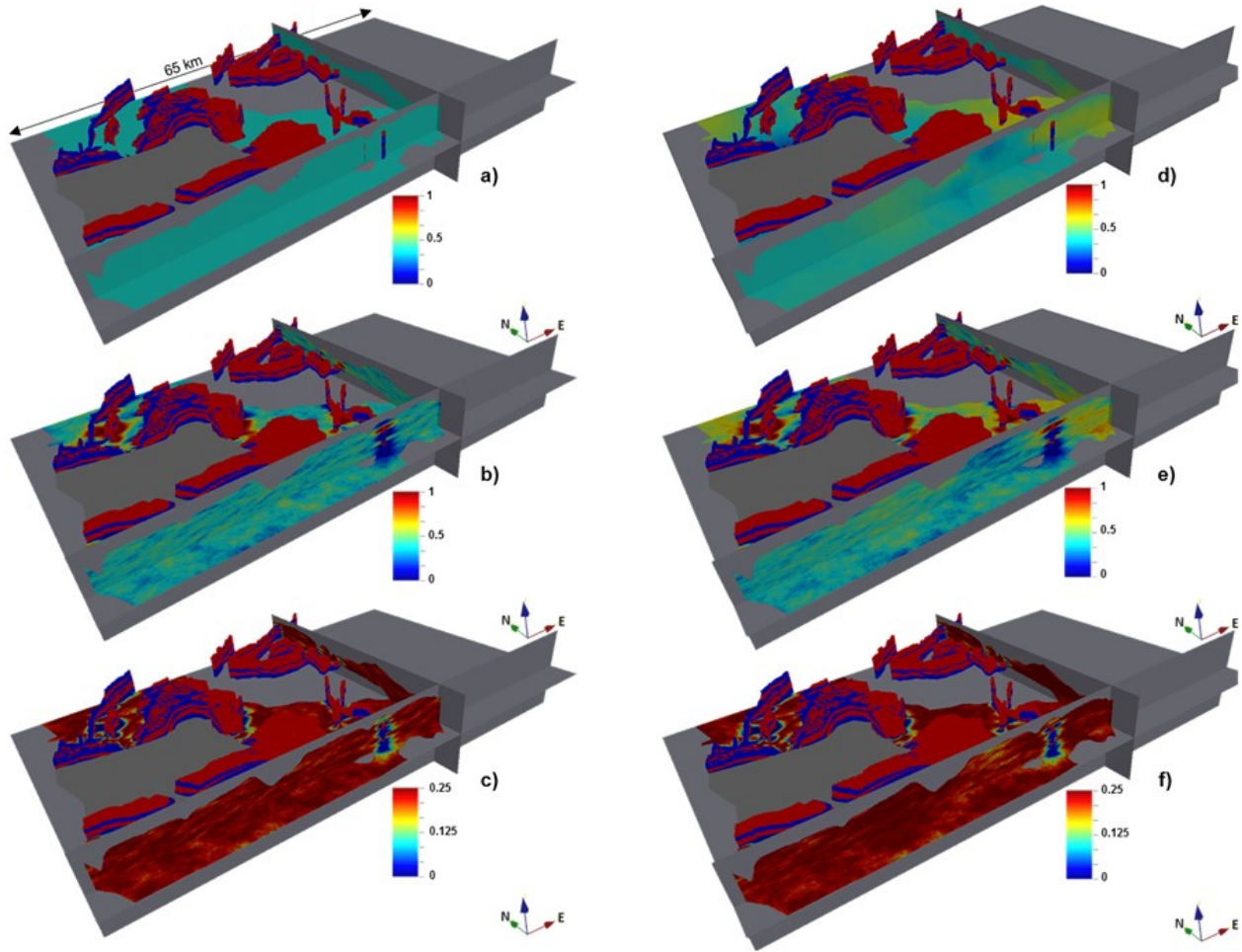
Figure 22: Comparison of the conditioning approaches (c) and (d) of Table 1 in terms of: soft and hard conditioning - panels (a) and (d); e-type map - panels (b) and (e); variance map - panels (c) and (f). To facilitate the comparison, the probabilities in panels (a) and (d) are presented in a different colour scale with respect to Fig.s 7-8. The e-type and variance maps are based on 100 realizations. In all panels, the interpretation of the seismic data, and the buffer zone around the Tønder model are explicitly shown in terms of sand and clay (the red homogeneous volumes represent the sand bodies, the blue volumes show the clay lenses). Vertical exaggeration = 20x.

Strategy	Training image	Soft data	Hard data
a	2nd (Fig. 9b)	÷	÷
b	2nd (Fig. 9b)	÷	The Tønder model Seismic interpretation Boreholes
c	2nd (Fig. 9b)	Sand probability directly from boreholes	The Tønder model Seismic interpretation
d	2nd (Fig. 9b)	3D kriged grid of the sand probability distribution from boreholes	The Tønder model Seismic interpretation

Table 1: The different conditioning strategies tested in this study. The corresponding realizations are presented in the Figs. 14 - 16.