Dear Mr. Carrera,

We have decided to compile a revised version of our paper according to the comments from the three anonymous reviewers.

As previously discussed we have added co-author Anne-Sophie Høyer to the author list, since she was somehow forgotten.

In particular, we have rearranged the paper so that a dedicated methods section is now present, which focuses on introducing the different methods, such as the Tau model, MPS, and distance methods. This resulted in the methods being introduced before they are used. The mathematical formulations have been cleaned up and shortened. Generally, the referee comments have been addressed and following you should find our response and a complete list of all corrections.

Best,
Adrian S. Barfod

Response to referees, hess-2017-734

Firstly, the authors would like to thank the three anonymous referees for taking their time to read the manuscript and providing detailed and constructive comments. The comments, questions and suggestions are addressed in the following response.

***RC: Referee Comments**
***AC:** Author Comments

NOTE: All line numbers in this document refer to the revised paper and not the "manuscript changes document" or any previously submitted versions of this paper.

Response to anonymous referee #1

General comments:
'*The main subject of the article is studying the impact of datasets used to do hydrostratigraphic modeling with the MPS framework. The authors build a "Base Case" using snesim approach, with a cognitive model as Training Image (TI), borehole data as hard data and geophysical resistivity data (SkyTEM) as soft data. Then, they present different modeling cases: using a different TI, using an incomplete resisitivity grid instead of a full resistivity grid, using borehole data as soft data instead of hard data, inverting resistivity data with a sharp inversion model instead of a smooth inversion model. The authors assess qualitatively and quantitatively the impact of changing each of these parameters. The main contribution of the article is the method to compare quantitatively the great number of geostatistical realizations (400). The method used is based on Analysis of Distance, with the Euclidean Distance Transform (EDT) algorithm applied to measure the "distances" between realizations. The distances serve as measures of similarity between the different cases and also between cases and the TI. According to the reviewed article, the EDT is straight forward method to assess the dissimilarity between realizations that can help in the quantification of the uncertainty of the 2D and 3D models. The smaller the distance, the more similar the realizations and thus, the smaller the impact of the changed parameter on the modeling results. It is a contribution because not many hydrogeology articles are found on the "metrics" for comparing geostatistical realizations. Plus, distance measures are discussed on MPS literature but for their use in pattern modeling from training images (Gregoire Mariethoz and Caers, 2015; Honarkhan, 2011), not for their use in uncertainty estimation. In the recent book from Mariethoz and Caers (2015), called "Multiple-point geostatistics: stochastic modeling with training images" the use of distance transforms for uncertainty purposes is not mentioned. Furthermore, in the review papers on MPS methodology, the study of the sensitivity of the model prediction to TIs and underlying datasets is suggested as an important research avenue (Hu and Chugunova, 2008).*
The paper is well written, with good story-telling. Even though several cases are presented, the structure is logical and the discussion about the results of each case is clear thanks to the images presented. I would agree with the publication of this article because the method seems to be a contribution with the uncertainty appraisal of the MPS results.'

RC1: '*Although the proposition to use distances to assess uncertainties is interesting, it seems to me that the simple EDT is not the most adequate to capture the differences between realizations. To give one example, in*

figure 9 we can see that the results from "Case 1a" and "Case 1b" are very different (basically, sand to the west and clay to the east for Case 1b while Case 1a is heterogeneous in the whole model). Nevertheless, in figure 11 both cases present the same distance to the cognitive geological model. The qualitative assessment by visual means remains necessary. Did you consider using more robust methods for comparing patterns in images which take into account the positioning of the events (spatial relations) and that are less affected by scaling, rotation and translation (e.g. SIFT, IMED)?'

AC1: The usage of the simple EDT for computation of the 'distance' between different realizations is a computationally feasible method for comparing 400 realizations each containing 229*133*39 cells (1,187,823 cells). The EDT method was therefore used because it thoroughly compares the average mismatch between the different realizations. However, we agree that the method is a bit simplistic for comparing heterogeneous geological objects. Other methods were in fact considered, but the EDT-based approach was the most suitable for getting started with research into comparing stochastic MPS modelling. It would be interesting to study, but was left out of the paper to keep focus on the large number of cases and assessment of the uncertainty related to them

Changes to the manuscript:
Lines 141-144: the IMED and SIFT distances are mentioned as alternative methods for comparing 3D hydrostratigraphic MPS realizations.

RC2: '*On the Kasted TI and the conceptual TI we observe channels filled with one facies, without internal variation. How come there are these intercalations of sand and clays in the simulation results?*'

AC2: This is due to the fact that the simulations are probabilistic in nature, and are based on random processes. At the beginning of the simulation a random path is drawn so that the simulation grid is filled by visiting each grid cell only once. The fine-scale patterns are partly due to the hard data which are inserted into the simulation grid before the simulation commences, and is excluded from the simulation path. As the grid is filled out, the hard borehole data might suggest a certain category but the soft data suggests another category. As the grid is filled out the overall category from the soft data dominates and if the random path visits the grid cells near the hard data point towards the end of the random path, then we are left with a small intercolation (Hansen et al., 2018). The Intercolations are also inherent in the simulations without hard data, and is mainly due to process randomness related to how the snesim algorithm draws from a cumulative density function (CDF) (REF: Strebelle 2002, snesim paper)

Changes to the manuscript:
The above comment did not result in any changes to the manuscript.

RC3: '*As mentioned in the discussion, the global target proportions of the units could have been replaced by the vertical proportions. It would have been interesting to see the results of these realizations with the vertical proportions, but I understand that the authors don't have them (time constraints?). What are the statistics of the results? How are those global proportions respected in each case? How the change of the parameters impacts the global target statistics?*'

AC3: The simulations have not been run with the vertical proportions since this would constitute an entire set of cases on its own, and due to the length of the paper and time constraints this was not considered. Regarding the compilation of the global proportions statistics, this has not been reported since the focus of this paper was to compare the results based on distance based similarities. However, provided it is not too time consuming it could still be a possibility to report the average global proportion statistics for the different cases.

Changes to the manuscript:
Due to the general length of the paper, we would like to avoid adding an extra case and expanding the paper to contain the global proportions statistics for each realization, therefore we have made no changes to the manuscript.

RC4: '*The article relies on other papers for most of the methodology, but still gives some small descriptions. Nevertheless, nothing is presented on the Direct Sampling Method used for filling the gaps on the resistivity grid. This seems to be missing in the methodology. Also, the choice of the Tau value for resistivity and boreholes (2 and 1) is not argued or referenced. Why 2 for resistivity and 1 for boreholes?*'

AC4: We see your points. The choice of Tau model was purely based on a series of tests, which are not mentioned and should be mentioned. Different combinations of Tau values were exhaustively tested, and the chosen values resulted were chosen since they resulted in simulations in which a smooth transition between borehole conditioned areas and non-borehole conditioned areas was seen.

Changes to the manuscript:
Lines 215-231: A dedicated section, section 3.2, has been added describing the usage of Direct Sampling Method for reconstructing incomplete geophysical datasets
Lines 193-214: A dedicated section, section 3.1.1, introducing the Tau model has been added
Lines 508-511: The exhaustive tests for choosing the tau model parameters are described

RC5: '*What could the authors infer about the impact of the datasets in areas where data is less dense? The study case in Denmark has good data coverage, both for geophysical surveys and for borehole data.*'

AC5: When the dataset is less dense than the Kasted dataset, the simulations have less soft data for conditioning. The less conditioning data which is available, the more the simulation relies on the TI for conditioning. An extreme of this scenario is when no conditioning data is used, and the realization is entirely unconditioned (e.g. Strebelle and Journel, 2000). Therefore, the choice of TI would impact the simulation result in a larger degree if less conditioning data is present.

Changes to the manuscript:
Lines 697-701: A sections discussing the good data coverage and non-stationarity has been added

Technical comments:

- It is not indicated what "SkyTEM" stands for.

Author response: It is a system name for an airborne transient electromagnetic system, i.e. the name of a company.

Resulting changes:
Line 54: the sentence is changed to “The airborne transient electromagnetic SkyTEM data …”

- Page 4: “Two approaches are taken”? … we are expecting a second approach…
  Author response: Yes, the other common approach should be included in this paragraph.
  Resulting changes: Lines 221-223: the other common approach is described

- Page 7: Realizations THAT reflect the real world
  Author response: P. 7 line 15: this will be corrected in the revised paper.
  Resulting changes: Line 180: the sentence now reads: “…realizations that…”

- Page 15: if the grid is too sparse, then limited or no information is present which can help reconstruct missing patterns is present (repetition of “is present”)
  Author response: P. 15 line 9: this will be corrected in the revised paper.
  Resulting changes: Lines 452-453: The repetition of the word present has been removed

- Page 29: “increased”
  Author response: P. 29 line 18: this will be corrected in the revised paper.
  Resulting changes: Lines 652: “increase” has been changed to “increased”

- Page 21: comparing a “realization” (no “s”)
  Author response: P. 21 line 27: this will be corrected in the revised paper.
  Resulting changes: Line 292: the “s” has been removed from “realizations”

- Page 35: Journel, A. G.: "Combining Knowledge From Diverse Sources: An Alternative to Traditional Data, , 34(5), 2002". (The name of the Journal is missing, “Mathematical Geology”)
  Author response: The mentioned reference will correctly contain the journal name in the revised paper.
  Resulting changes: The journal name has been added

Response to anonymous referee #2

General comments:

'*The manuscript "Contributions to uncertainty related to hydrostratigraphic modeling using Multiple-Point Statistics" presents an interesting study where the uncertainty related to the input data required by a multiple-point statistics (MPS) simulation framework is investigated. The research described in the manuscript, although focused on a specific case study in Denmark, could have a broader applicability and would probably be of interest for the HESS readers.*

Nevertheless, I believe that the manuscript contains some major issues that should be addressed by the authors before its publication. In particular, my concerns are related to three aspects: 1) The structure of the manuscript, 2) some missing details/discussion about important aspects of the parameterization of the methodology, 3) the way mathematical relationships are expressed.'

RC1: '*Manuscript structure: A number of techniques are used within the manuscript to complete the quite complex simulation framework. Some of them are used multiple times and in different contexts (for example, the tau model). Therefore, putting their description in a separate section "Methods" would be much more helpful and would help the reader in orienting himself inside a quite complex work-flow. At the moment, the description of the methods is spread all around the manuscript, sometimes together with the results, quite often with some repetition, which makes reading the manuscript not a smooth task. A clear example of this "breaking the rhythm" of the manuscript is for example at page 18. Also, here the description of the technique is made at the wrong place, because the method was already applied some step before in the work-flow. Another example is at page 21, where a 2D example is introduced to explain the EDT.*

In addition, the comparison methods (EMR-maps,...) and the distances (EDT...) definitions would deserve a separate section, maybe just after or within the "methods" section. There are also many locations, in particular in the "Results" sections, where too many details which would be more appropriate for the "Discussion" section are anticipated (see for example pages 27-28, lines 6, 10-11).'

AC1: Under the preparation of the manuscript we have worked iteratively on the structure of the manuscript and ended up with this structure as the most reader friendly consisting of a large Material and methods section (section 2), divided into the study area (2.1 The kasted study area), a general method description section (2.2 Multiple-Point Statistics (MPS) and single normal equation simulation (snesim)), a detailed description of the MPS modeling set-up with the methods related to each case (2.3 MPS modeling setup) and lastly a section of the methods used for comparing the simulation results. Generally, the style chosen for this manuscript was to explain some of the methods utilized in the, as mentioned, quite complex simulation framework was to use practical examples. This often aids the reader with figures and a purpose for applying the given method.
It does seem warranted to create a separate section for the general methods.

Changes to the manuscript:

The old section 2. Materials and Methods are split up in three sections: 2 The Kasted study area, 3 Methods, 4 MPS modeling setup.

Lines 145-171: The section describing the Kasted study area has been assigned to its own section, "2. The Kasted study area"

Lines 171-322: A dedicated methods section has now been added, "3. Methods", which together with the new section "4. MPS modeling setup" replaces the methods part of the old materials and methods section, "2. materials and methods section".

Lines 193-214: A sub-section describing the Tau model has been added, “3.1.1 The Tau model: Combining conditional probabilities”
Lines 215-231: A description of the direct sampling method for reconstructing incomplete datasets has been added, “3.2 Reconstructing incomplete datasets using Direct Sampling”
Lines 232-322: The old section “2.4 Comparing simulation results” has been moved to the new Methods section
Lines 325-524: The section describing the MPS setup and the simulation cases has been assigned to its own section 4. MPS modeling setup cleaned for method descriptions.

RC2: *‘Section "Basic modeling set-up”: This section is somehow quite confusing, because the authors mix the description of the "Basic modeling set-up" with the Egebjerg TI description. I suggest to better separate the description of the various cases.’*

AC2: This was a mistake. The reference to Table 2 for some reason included text from the section above or below Table 2.

Changes to the manuscript:
Line 406: The reference to Table 2 has now been fixed and the Egebjerg TI is not mentioned under section “4.1 Basic modeling setup”

RC3: ‘*Tau model usage: The tau model represents one of the crucial steps of the methodology, because it is used to take into account the soft constraints provided by geophysics, but also to combine the "borehole probability" with the SkyTEM one (Fig.7). Although some information about the tau weights are provided (i.e., in appendix), I would suggest to discuss at least briefly their choice. For example, many of the considerations made by the authors would be strongly influenced by the choice of the tau weights (see for example line 32, page 30). Some insights about the choice of these weights are provided by Allard et al (2012, DOI: 10.1007/s11004-012-9396-3). Also, what happens when the weights are_1? (see pp18, equation 2).’*

AC3: Here we refer to the Author Comment 4 (AC4) in the response to anonymous referee #1. Briefly, in relation to the choice of Tau parameters for the Tau model it should be mentioned that a series of tests have actually been carried out to select the final Tau values.

Changes to the manuscript:
Lines 508-511: The exhaustive tests for choosing the tau model parameters are described

RC4: ‘*Case studies labelling: The provided table that summarizes all the case studies is of course useful, but overall into the manuscript (for example, in figure captions), there is very often a redundancy and some of the details of the different methods, which are repeated multiple times.*
Maybe you should reference much more often to Table 1 and to the "codes" like "Case 1a", "Case 1b" only, and avoid repeating the detailed differences. One example of these repetition can be observed in Figure captions (see for example Fig.9, page 25).’

AC4: This would certainly make the paper more concise in certain parts. However, it would also require the reader to constantly avert his/her attention to Table 1.

Changes to the manuscript:
Lines 578-581: The figure text related to Figure 9 has now been simplified and a reference to Table 1 is added

RC5: '*Introduction pp4, lines 20-XX: Here I would also mention the problems related to the solution of the inverse problem (IP) in itself. By the way, this also somehow motivates your efforts in trying two different inversion techniques, like SCI and sSCI.*'

AC5: This is a good point.

Changes to the manuscript:
Lines 111-119: A brief description of the problems related to the inversion process have been added to the introduction

RC6: '*pp5, lines 27-: Here I believe you are already providing too much details for an introduction.*'

AC6: Okay, we will provide less detail

Changes to the manuscript:
Line 136-144: Less detail is now provided in the mentioned paragraph

RC7: '*Mathematical formulation: The mathematical formulation is often cumbersome, because very often long text lines are used to define quantities and as subscripts. I strongly suggest to lighten the notation avoiding long text lines, and using the many letters provided by the alphabets. For example, N and M could be used instead of Nrealizations and Ncells; another example is the definition of Di;j (page 22). In addition, some relationship could be condensed and generalized. In this way, they could be written only once and contribute to shorten the manuscript. See for example (7) and (8) at page 21.*'

AC7: We agree that the mathematical formulations could be written in a more concise manner. However, regarding eq. (7) and (8), again, we believe it aids the reader to first see the formula with a specific case in mind and then generalize afterwards.

Changes to the manuscript:
Line 253: Equation has been simplified: "N_{reals}" changed to "N",
Line 285: Equation has been simplified: "cog.model" has been changed to "m_{cog}", "$realization_i$" has been changed to "$m_{r,i}$", the expression "$d_{EDT}^{cog.model}$" has been changed to "$d_{EDT}^{m_{cog}}$", and the expression "$d_{EDT}^{realization_i}$" has been changed to "$d_{EDT}^{m_{r,i}}$"
Line 294: The equation has been simplified: the expressions "$model_A$" and "$model_B$" have been simplified to "m_A" and "m_B"
Line 312: The equation has been simplified: the expressions "$realization_i$" and "$realization_j$" have been changed to "$m_{r,i}$" and "$m_{r,j}$"

RC8: *'TI non-stationarity: In section 3.1 but also in other parts of the manuscript the imprecision of MPS in reproducing some features is clearly depicted. However, I believe that many of the encountered problems are due to the non-stationarity of the used TI. Therefore, although of course taking into account for the geophysics helps, I would suggest to at least mention and briefly describe the role of the non-stationarity of the TIs.'*

AC8: The geophysical data is so spatially dense that TI non-stationarity is not as big of an issue as when less conditioning data is present. We therefore decided to not include this complication, to not confuse the reader. However, it might be wise to mention that this is the case and that even though the Tis are non-stationary, the spatially dense geophysical data actually helps in the matter.

Changes to the manuscript:
Lines 697-701: The non-stationarity of the Kasted model and Tis are discussed in relation to the density and overall coverage of the geophysical dataset

RC9: '*Variograms: Could you briefly mention which variogram model you used for example to create the borehole footprint (pp18, lines 3-6)?*'

AC9: This is mentioned both in the text where you indicated, but also in the Appendix under section "A1".

Changes to the manuscript:
Lines 890-891: the variogram model type (exponential) is now mentioned under Appendix "A1"

RC10: '*Fig.4: Please add a comment related to the spatial scale of the Egebjerg TI, which is quite different from the other two. Also, it would be quite nice to add a sub-figure containing the same vertical proportions for the borehole logs.'*

AC10: That is a good point! A comment related to the spatial scale of the Egebjerg TI will be added. A subfigure containing the vertical proportions of the borehole logs was not added since the available boreholes were largely drilled for the purpose of groundwater exploration and management. This means they are biased towards more permeable hydrostratigraphic layers such as coarse sands and gravels, and impermeable clay layers are generally underrepresented.

Changes to the manuscript:
Line 420: A comment has been added commenting on the fact that the Egebjerg TI is larger than the other TIs.

RC11: *'Fig.11: The label "Realization number" in the vertical axis of part A is too close to part B and is therefore misleading. Also, I believe that the results of part B could be condensed using box-plots, one box-plot for each case. In this way, the fictitious and misleading order of the "realization number" would be by-passed.'*

AC11: Yes, the Fig.11B of the figure will be given a bit more space so that it is clear that the "Realization number" is part of Fig.11A. We would prefer to not use boxplots since they mainly provide summary statistics and we would prefer that the reader can see the actual distance values instead.

Changes to the manuscript:
Lines 632-634: Sub-figure B has now been moved so that sub-figure A and B are above and below each other

Technical comments:

- pp6, line 9: Please check the order in "33 line km spatially…"
  Author response: This will be re-ordered correctly.
  Resulting changes: Lines 160-161: The sentence has been revised and re-ordered.

- pp8, line 30: It looks like the reference to Fig.4 is missing between Fig.3 and Fig.5.
  Author response: We will add a sentence with reference to Fig. 4, so it will be mentioned before Fig. 5
  Resulting changes: The figure numbers have changed, but the figures are now mentioned in the correct order in the text.

- pp9, line 5: "data is" => "data are"
  Author response: This will be corrected in the revised paper.
  Resulting changes: Line 399: The sentence now read "...data are..."

- pp18, line 18: "[2,1]" is somehow confusing with the index that you introduce some equations before... I would specify that they are float values, writing explicitly 2.0 or 1.0.
  Author response: This will be added to the revised manuscript.
  Resulting changes: Line 535: the values have now been changed to floating point values ´, *i.e.* [2.0,1.0]

- pp20, line 25: Maybe "(3)" => "(4)"?
  Author response: This will be corrected.
  Resulting changes: Line 272: The equation references have now been fixed

- pp21, equation (6): Please check for the missing i subscript to v
  Author response: There is no missing subscript. The "**u**" symbol represents a single location vector, while the "**v**" is a set of vectors which are a subset of "V". Therefore, the formula only shows the computation of the d_{EDT} at a single location. The process must be repeated for all points for the grid. Perhaps this should be mentioned in the revised paper.
  Resulting changes: None

- pp21, line 20: Here Delta appears in the formula, but not in the following text... please check.
  Author response: This will be corrected in the revised paper.
  Resulting changes: Delta has been added to the following lines: 302, 304, 305, 306, 307, 320, 321 and 326

- pp22, line17: "realizations using" => "realizations computed using" (?)
  Author response: This will be corrected in the revised paper.

Resulting changes: Line 328: Changed to: "...realizations computed using..."

- pp27, lines 6-11, 30-31: This is somehow repetitive. Please try to avoid repetitions also in other locations in the text, but in particular in this section.

  Author response: We couldn't find the particular repetition mentioned here, but will try to remove as many of such repetitions in the revised paper.

  Resulting changes: None
- pp30, line 26: "to alter...?"

  Author response: Will be corrected, however, it seems like it is a logical statement.

  Resulting changes: line 708: Text has been modified accordingly

Response to anonymous referee #3

General comments:
'*This paper is the product of a nice piece of work and is very interesting. My main concerns are related to the introduction, where some transitions and justifications are missing and to the methodological section, that goes too fast into details. I therefore suggest the following minor revisions.*'

RC1: *'In the introduction, transitions are often missing between paragraphs, for instance, page 2, between lines 14 and 15. It is a bit jumps from one topic to another. Related to the paragraph comprised between lines 3 and 14, you might cite the following paper that discuss uncertainty and bias in training images : Ferré, Ty. "Revisiting the Relationship Between Data, Models, and Decision-Making." Groundwater 55, no. 5 (2017): 604-614. Pirot, Guillaume. "Using training images to build model ensembles with structural variability." Groundwater 55, no. 5 (2017): 656-659.'*

AC1: Good points.

Changes to the manuscript:
The introduction has been cleaned up so that transitions between paragraphs are less abrupt in the revised paper.
Lines 89-101: The paragraph describing the overall goal of the paper has been moved up
Line 84: The papers mentioned above by Ty Ferré and Guillaume Pirot are now cited.

RC2: *'General justifications are given in the first paragraph of the introduction, but the authors should also justify why they chose this specific sites, and what they bring or want to improve, with regards to previous studies conducted at the Kasted site.'*

AC2: That is a good point. The main justification for using the Kasted site was that the geology of the area was fairly simple, thus the dataset is also, to a degree, simple. Furthermore, a geological model of the area had already been compiled, meaning we already had information about the hydrogeology.

Changes to the manuscript:
Lines 98-100: A reflection of what we hope to improve in regards to the Kasted model has been added
Lines 154-159: The description of the Kasted survey area has been expanded and a general justification for the usage of the Kasted data has been added

RC3: '*You should also define your notions of hard and soft conditioning clearly in the introduction*'

AC3: Okay.

Changes to the manuscript:
Lines 102-105: Our notions of hard and soft conditioning has been added

RC4: '*Then, in section 2, when presenting the study area, the historic of previous research could be explained/clarified. Page 6, line 5: one 'complex' too much? Page 6, line 14: how is defined the 'selected quality threshold?*" . '

AC4: If anonymous referee #3 is referring to the fact that we should explain previous research in the Kasted area, then the main research which has been conducted in the area is compiling the cognitive 3D geological model of the area. Therefore, we do not desire going into too much detail.

Changes to the manuscript:
Lines 154- 159: Additional information regarding the Kasted area has been added, introducing the essential feature known as tunnel valleys.
Line 150: The poor usage of the word 'complex' is corrected
Lines 163-165: References to the borehole quality assessment scheme we utilized has been added and the 'selected quality threshold' is elaborated upon

RC5: '*The main aspect, regarding the method subsection is that the reader is lost in details from the beginning. A big picture of the approach is missing. In the way of presenting, I would recommend to give first an overview of the method, and progressively go into details.*'

AC5: Perhaps it would be wise to include an overview of what we hope to achieve by stochastic modelling.

Changes to the manuscript:
The structure of the paper has been changed significantly. A new Methods section has been added which contains a generalized description of the methods before they are invoked, which clearly separates the theory behind the methods from how it is used. Furthermore, the description of the Kasted area has been given its own section. The MPS modeling setup section has got an introductory section.

RC6: '*Page 7, lines 7 to 15 are not very clear. Can you reformulate?*'

AC6: Will be reformulated in the revised paper.

Changes to the manuscript:
Lines 175-182: The paragraph has been reformulated

RC7: '*Page 8, line 26, figure 5 is called, while figure 4 has not been called.*'

AC7: We will fix this by adding a reference to Figure 4 before we call Figure 5.

Changes to the manuscript:
The figures are now invoked in the correct order

Miscellanous Corrections:

- The author list has been updated so that co-author Anne-Sophie Høyer is on the list
- Lines 183-185: References have been updated so that (Høyer et al. 2017) is now correctly categorized as a hydrology paper
- Line 268: A reference to the new tau model section has been added
- Line 210: The denominator was missing a "P" in the last formula (x_2=…)
- All figures have been updated with a larger sub-figure label
- Lines 422-424: Figure 5 text has been updated so that the usage of "slice" and "cross-section" are correct
- Lines 483-487: Figure 7 text has been updated so that the usage of "slice" and "cross-section" are correct
- Lines 506-507: A reference to Appendix A1 is has been added
- Lines 658-660: Figure 8 text has been updated so that the usage of "slice" and "cross-section" are correct
- Lines 582-586: The text has been shortened and simplified and a reference to table 1 has been added
- Line 602: The Barfod *et al.* (2018) reference has been updated since it is now a published paper
- Lines 677-683: The paragraph has been edited to make it more readable
- Line 774: "I" has been replaced with "We"

References

Hansen, T.M., Vu, L.T., Mosegaard, K., Cordua, K.S., 2018. Multiple point statistical simulation using uncertain (soft) conditional data. Comput. Geosci. 114, 1–10. https://doi.org/10.1016/j.cageo.2018.01.017

Strebelle, S., Journel, A.G., 2000. Sequential simulation drawing structures from training images. Stanford University.

Marked-up manuscript

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)

Contents

Contents 2

1 Introduction 3

2 The Kasted study area ..... 6~~5~~

3 Methods ..... 7~~6~~

3.1 Multiple-Point Statistics (MPS) and single normal equation simulation (*snesim*) ..... 7~~6~~

3.1.1 The Tau model: combining conditional probabilities ..... 8~~7~~

3.2 Reconstructing incomplete datasets using Direct Sampling ..... 9~~7~~

3.3 Comparing Simulation results ..... 9~~8~~

3.3.1 Ensemble mode ratio maps (EMR-maps) ..... 9~~8~~

3.3.2 Euclidean Distance Transforms (EDT) – measuring similarity between 3D hydrostratigraphic realizations ..... 10~~9~~

3.3.3 Evaluating the distance matrix ..... 11~~10~~

4 *MPS* modeling setup ..... 12~~11~~

4.1 Basic modeling setup ..... 16~~14~~

4.2 Case 1 – Conceptual geological understanding ..... 16~~14~~

4.3 Case 2 – Incomplete soft data ..... 20~~16~~

4.4 Case 3 – Choice of resistivity model ..... 21~~17~~

4.5 Case 4 – Borehole lithology logs ..... 23~~18~~

4.6 Case 5 – Excluding the soft resistivity data ..... 25~~19~~

5 Results ..... 25~~19~~

5.1 Visual comparison of hydrostratigraphic realizations and “Ensemble Mode Ratio”-maps (EMR-maps) ..... 28~~19~~

5.2. Quantitative comparison using differences in object based Euclidean Distances as a measure for similarity ..... 33~~22~~

6 Discussion ..... 36~~25~~

7 Conclusion ..... 39~~28~~

8 Acknowledgements ..... 39~~28~~

References ..... 40~~29~~

Appendix ..... 44~~33~~

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~~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 ~~sparse~~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 ~~setting~~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 ~~also~~ 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~~, thus making it difficult to quantify the related structural uncertainty. Furthermore, using the geological model to make predictions, the uncertainty related to the geological model carries over into the predictions.~~. 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 ~~such~~ 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~~Figure 1 Geophysical data is~~). 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 ~~our model.~~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)~~Presently, two approaches are taken when it comes to incomplete geophysical data. A common approach is to reconstruct the incomplete dataset using geostatistical tools . However, in this case it is important to emphasize that the reconstructed information is not~~

as valuable as the actual measured geophysical information, 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 viewed asconsidered 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, the "Single Normal Equation Simulation" (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).

We present a study of the uncertainty related to stochastic hydrostratigraphic MPS modeling of a hydrogeophysical dataset from Kasted, Denmark. 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 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 isThe *snesim*, which is used for stochastic hydrostratigraphic modeling. The stochastic models will be divided into 6 overall 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: Case 1 – Conceptual geological understanding, Case 2 – Incomplete soft data, Case 3 – Choice of resistivity model, Case 4 – Borehole lithology logs, and Case 5 – No soft data.

modeling scheme resulted in a total of 400 MPS realizations are created. Comparing hydrostratigraphic models. Visual comparison of such a large number of 3D hydrostratigraphic realizations visually is difficult to do in a quantitative manner. Therefore, a mathematical comparison method based on Euclidean Distance Transforms (EDT) is used. The EDT converts a 3D binary image to a continuous 3D Euclidean distance grid. 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 based distance results are computed for all of the 400 realizations, and grouped by case, to create a full distance matrix, comparable to a co-variance matrix. An example is Case s4, which studies the effect of boreholes on the hydrostratigraphic realizations. Two MPS realization ensembles are created, 50 realizations with boreholes included, and 50 realizations without borehole information. The two model ensembles can be compared in two ways: visually and

~~mathematically. Visual comparison of the~~ 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). ~~The~~. 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)~~comparison provides a quantitative overview of similarity, or dissimilarity, of the two realization ensembles~~. 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 Materials and methods**~~

~~**2.1 The Kasted study area**~~

2 The Kasted ~~survey~~study area

The Kasted area is located ~~north-west~~northwest of Aarhus, Denmark (Figure 1~~Figure 1~~A), 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 ~~333 line km~~ 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 ~~greater~~ 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 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 1~~Figure 1~~C and described in further details in Barfod *et al.* (2018) and Høyer *et al.* (2015).

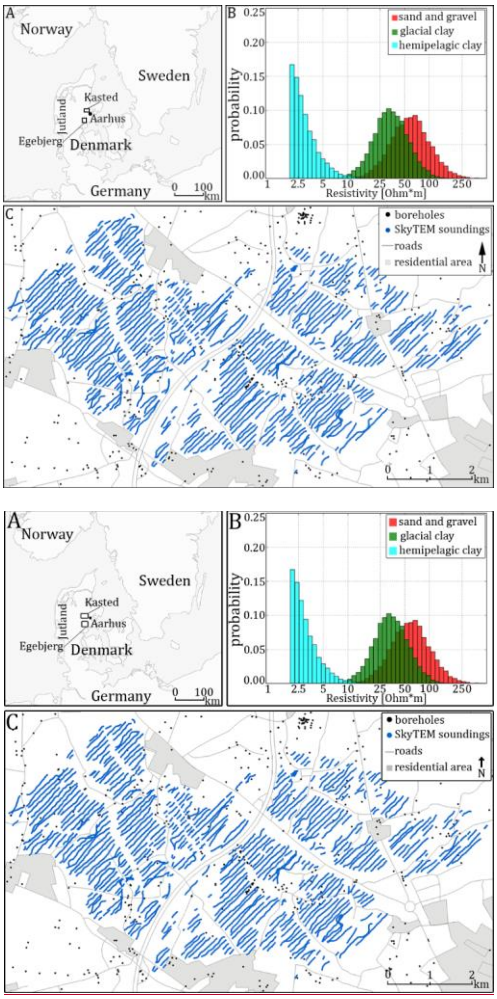


Figure 1: *The Kasted survey area and resistivity-hydrostratigraphic relationship histograms.* ***A*** *shows the geographical location of the Kasted survey area and the Egebjerg model used as a secondary TI.* ***B*** *shows the reconstructed resistivity-hydrostratigraphic relationship histograms for 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.*

~~2.2~~3 Methods

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

The Multiple-Point Statistics (MPS) framework stems from the~~, more~~ general~~,~~ geostatistics framework. Here, Multiple-Point (MP) information from a Training Image (TI) is used to condition simulations to probable geological patterns ~~from the 2D or 3D digital TI~~ (Journel and Zhang, 2007). The TI thus provides a conceptual geological understanding of a given area~~. The TI~~, and can be viewed as a database containing ~~realistic~~probable geological patterns~~. This MP information is~~, which are used to condition the MPS simulation~~, to increase the overall geological realism of the realizations~~. 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 ~~any~~ locally

accurate information, *i.e.* the TI does not need to spatially or geographically overlap with real-world geological units, and can be purely 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) ~~*etc.*~~.

Field Code Changed

Formatted: English (United States)

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 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).

~~2.~~3 ~~*MPS*~~ .1.1 The Tau model: combining conditional probabilities

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) ~~setup~~. 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 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, $D_i, i = 1, \dots, n$, and the goal is to estimate the probability that a hydrostratigraphic unit (A) is present provided all data events:

$$P(A|D_1, \dots, D_n) \quad (1)$$

The first step is then to define the prior probability distribution, $P(A)$. Generally the Tau model can be applied to as many different probability grids as desired, but for the purpose of simplification two probability distributions are defined: $P(A|D_1)$ and $P(A|D_2)$. In this study we will consider D_1 and D_2 as 2D or 3D probability grids from different sources. As an example D_1 could be a probability grid from geophysical data and D_2 a probability grid from borehole lithology logs. The probability grids are translated into distance grids by applying the "*probability-into-distance*" transform:

$$x_o = \frac{1-P(A)}{P(A)}, x_1 = \frac{1-P(A|D_1)}{P(A|D_1)}, \text{ and } x_2 = \frac{1-P(A|D_2)}{P(A|D_2)} \quad (2)$$

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

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

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

$$P(A|D_1, D_2) = \frac{1}{1+x} \quad (4)$$

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

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

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)~~Kasted~~, 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.

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 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 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~~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 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.

Formatted: English (United Kingdom)

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 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, $P_k(\mathbf{u})$, is computed as follows:

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

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 **u** and for the *i*'th realization. The EMR values for a given cell, **u**, can then be computed as follows:

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

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 location **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, **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 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, **u**, the $r_{EMR}=1/3$ and the simulation is uncertain.

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**:

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

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 2~~Figure 2~~A-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 2~~Figure 2~~E-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 2~~Figure 2~~I-L, are used to compute an average Euclidean Distance between each realization, $m_{r,i}$, and the cognitive geological model, m_{cog}:

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

where $i \in \{1, \dots, N\}$, 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~~Figure 2~~, where the realization which is closest to the cognitive geological model (Figure 2~~Figure 2~~B, F and J) has a Δd_{EDT}-value of 240 m, while the realization, which was ranked 25th closest (Figure 2~~Figure 2~~C, G and K) has an Δd_{EDT}-value of 280 m, and lastly the realization, which was farthest (Figure 2~~Figure 2~~D, 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:

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

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.* $M_{m_A} = M_{m_B} = M$.

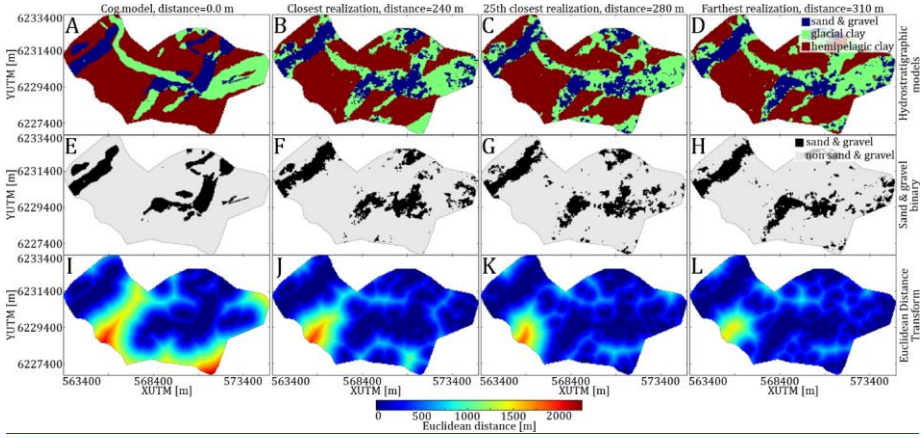


Figure 2: *A 2D example of the Euclidean Distance Transforms (EDT) as a measure for the similarity between categorical MPS realizations. In this example a set of 50 realizations, from the basic modeling setup, are compared based on the differences in EDT for sand & gravel units. **A-D** shows the hydrostratigraphic 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. (8) and is defined as follows:

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

where $i, j = \{1, \dots, N\}$, 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~~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.

4 *MPS* modeling setup

The Kasted dataset, used in this study, is comprised of a dense geophysical dataset acquired using the SkyTEM ~~dataset~~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. 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 the SkyTEM data are not used. Further details on each case follow in the coming sections and Table 1~~Table 1. However, the goal of this study is different. Since a 3D deterministic model of the area already exists, the study area, and overall geological setting, is well known. Additionally, the cognitive geological model can be utilized as a 3D TI.~~ 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. ~~The Kasted survey lithology logs reveal a combination of 59 geological categories, which are grouped together~~It is parameterized into three ~~key~~ hydrostratigraphic units:

1) *sand and gravel*: a combination of coarse lithological units, including sand till, meltwater sand, gravel and pebbles of 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. ~~Using these hydrostratigraphic categories, the~~ The Kasted survey lithology logs reveal a combination of 59 geological categories, which are translated into a set of hydrostratigraphic logs ("step 1", Figure 3~~Figure 3Step1",), and~~) using these hydrostratigraphic categories. Similarly, the 42 geological units in the cognitive geological model ~~is translated~~are divided into the three abovementioned categories (Figure 4~~Figure 4a~~A). The vertical proportions of the three category hydrostratigraphic Kasted model can be viewed in Figure 5~~Figure 5(A).~~A.

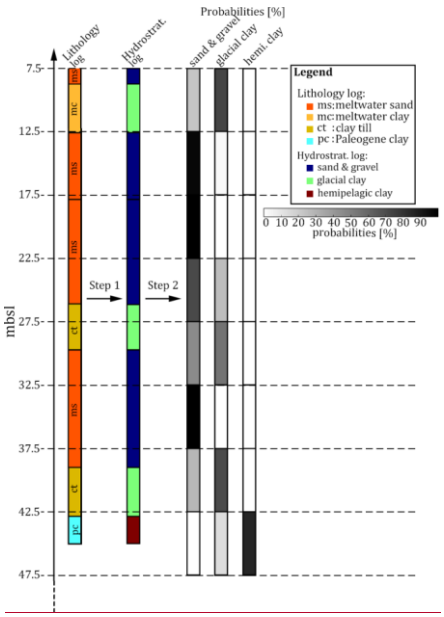


Figure 3: *A schematic diagram presenting the conversion of the lithological logs into probability logs for the three hydrostratigraphic units: sand & gravel, glacial clay and hemipelagic clay. Step 1: the lithology log is translated into a hydrostratigraphic log. Step 2: The hydrostratigraphic logs are resampled according to the vertical modeling grid intervals and an interval probability is calculated for each of the hydrostratigraphic units.*

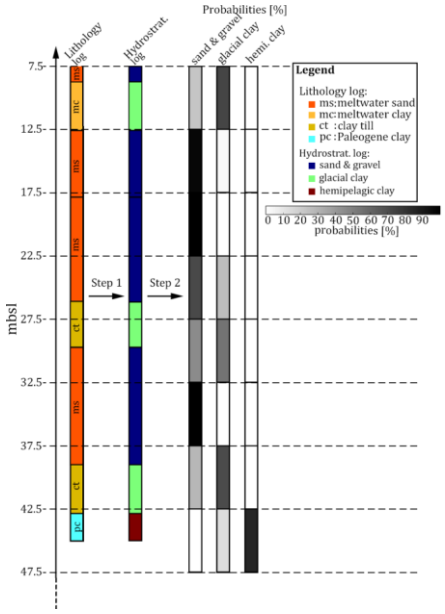
The borehole lithology logs need to be assigned to a 3D grid, which is carried out in three overall steps. The first step is to translate the borehole lithology logs into hydrostratigraphic logs using the above mentioned three categories; "Step 1" Figure 3~~Figure 3~~. The second step is then to divide 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~~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~~, by~~ . 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 6~~Figure 6~~A). The second and final step is to stochastically reconstruct the incomplete resistivity grid using DS stochastic reconstruction (Mariethoz and Renard, 2010) (Figure 7~~Figure 7~~~~, which is different from the approach in~~). The reconstruction procedure was originally presented by Barfod et al. (2018)~~.~~, however, we have made some improvements for this study. Originally,~~ in Barfod *et al.*~~ 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~~. This meant that~~, *i.e.* the resistivity grid had already been ~~slightly~~partly reconstructed~~.~~ in the proximity of the geophysical soundings. To avoid this, block Kriging estimation ~~is~~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)~~The block average resistivity value for all grid cells containing a resistivity model is estimated. The end result is an incomplete resistivity grid (A). The incomplete resistivity grid is then reconstructed using DS stochastic reconstruction (). See Barfod *et al.* for more detail on reconstruction of the incomplete resistivity grids of the Kasted survey.~~.

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

Field Code Changed



~~Figure :~~ ~~*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. Step1: 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.*~~

~~Like the resistivity models, the borehole lithology logs also 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 prior knowledge regarding the regional hydrostratigraphy; "Step 1" . 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 . 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" (e.g. Gunnink and~~

~~Siemon, 2015; Tahmasebi et al., 2012). 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 general MPS ~~setup~~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.
2) The incomplete resistivity grids (Figure 6~~Figure 6~~A) are stochastically reconstructed using Direct Sampling (DS~~), as presented by Mariethoz and Renard (A and)~~ (Figure 7~~Figure 7~~A). 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 1~~Figure 1~~B).
4) The borehole lithology logs are translated into hydrostratigraphic logs; "Step 1 " Figure 3~~Figure 3~~.
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~~Figure 23~~.
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.

Field Code Changed

Field Code Changed

~~This study 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*.~~ A total of 400 realizations are created, with 50 realizations per sub-case ~~– see for more details and a brief description of each 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.

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	---

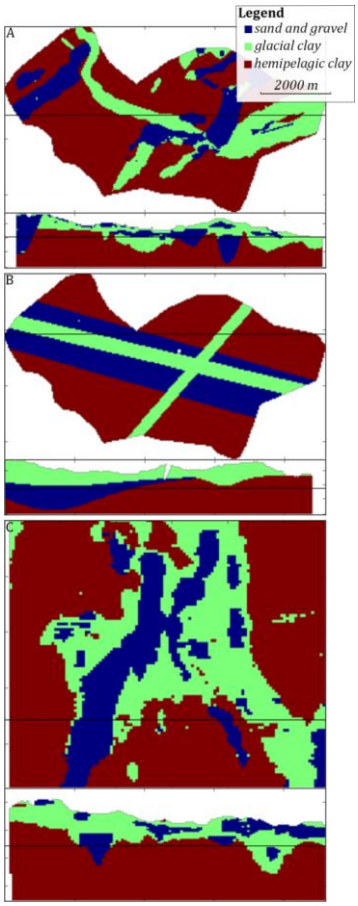
2.34.1 Basic modeling setup

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 4Figure 4A), which global proportion is listed in Table 2Table 2 and vertical proportion displayed in Figure 5Figure 5) () (A). The hard borehole logs are created from hydrostratigraphic probability logs by assigning the most probable hydrostratigraphic unit to a hard data grid.

2.3A.

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 1Figure 1A) is used as a TI (Figure 4Figure 4C). The geologic setting in Egebjerg is relevant since it is partly dominated by a buried valley complex (Jørgensen et al., 2010).


Legend
sand and gravel
glacial clay
hemipelagic clay
2000 m
A
B
C

485

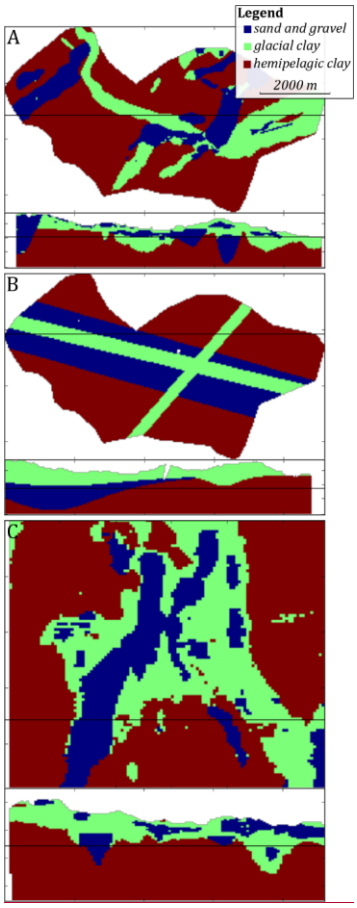


Figure 4: *An overview of the training images (TIs) which are used during MPS simulation. A horizontal ~~cross-section~~slice and vertical ~~slice~~cross-section is presented for each TI, portraying the hydrostratigraphic architecture;* ***A*** *shows the Kasted TI,* ***B*** *shows the conceptual TI,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

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 4~~Figure 4C),~~); this is referred to as Case 1a. It is clearly seen, by comparing Figure 4~~Figure 4~~A 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 4~~Figure 4~~C). The global proportions of the Egebjerg TI (Table 2~~Table 2~~) are similar to the ones found in the Kasted TI. However, the vertical proportions of the Egebjerg TI (Figure 5~~Figure 4~~5C) are different, especially in the upper part of the TI where *glacial clay* units dominate.

Table 2: *The global proportions related to each of the three TIs presented in Figure 4~~Figure 3~~4.*

	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 4~~Figure 4~~B 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 5~~Figure 5~~B). 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 5~~Figure 4~~5B).

Field Code Changed

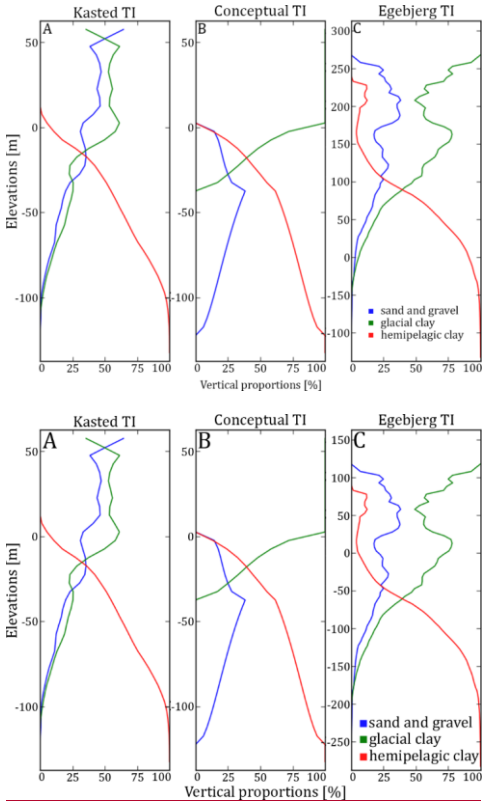


Figure 5. The vertical proportions of the three training images for each of the hydrostratigraphic categories, where **A** portrays the Kasted TI, **B** the conceptual TI, and **C** the Egebjerg TI.

~~2.3~~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 ~~is present~~. Signs of mediocre data density are seen in the incomplete grids (Figure 6~~Figure 6~~A). 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 7~~Figure 7~~A 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 1~~Figure 1~~B) (Figure 6~~Figure 6~~B-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 6~~Figure 5~~6B-D).

Field Code Changed

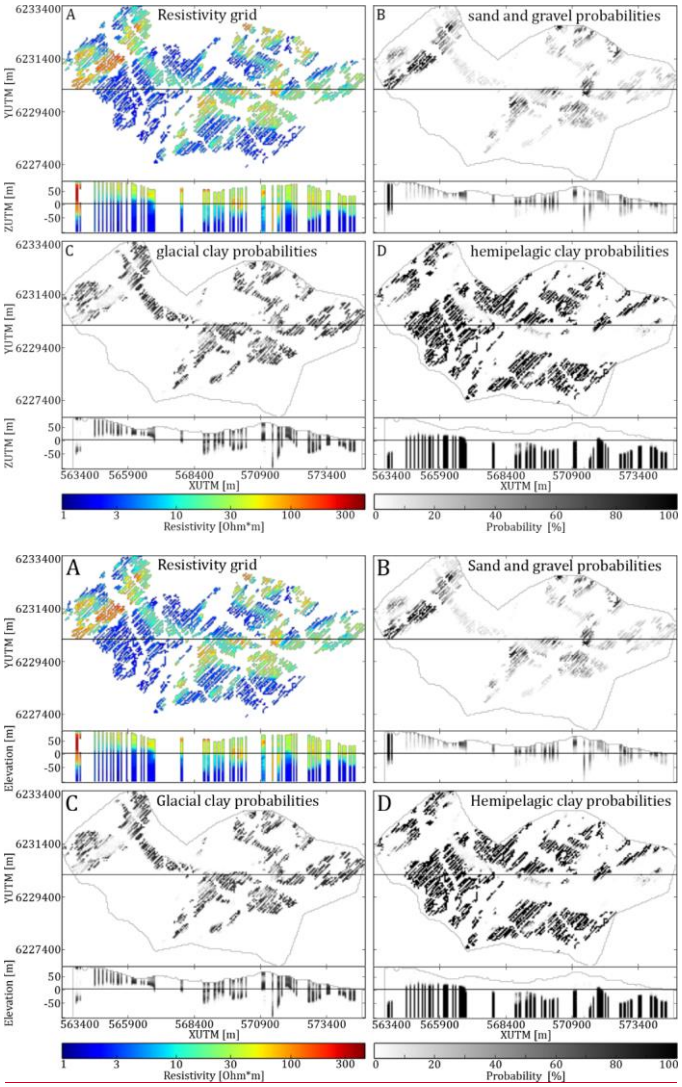


Figure 6: *A presentation of the incomplete resistivity grid. Each grid is portrayed as a horizontal ~~cross-section~~slice at 20 mbsl, and a vertical ~~slice~~cross-section intersecting at UTMY 6230100 m.* ***A*** *shows the resistivity grid which is translated into three probability grids using the resistivity-hydrostratigraphic relationship histograms (Figure 1~~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.*

~~2.3~~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 7~~Figure 6~~7B).

Field Code Changed

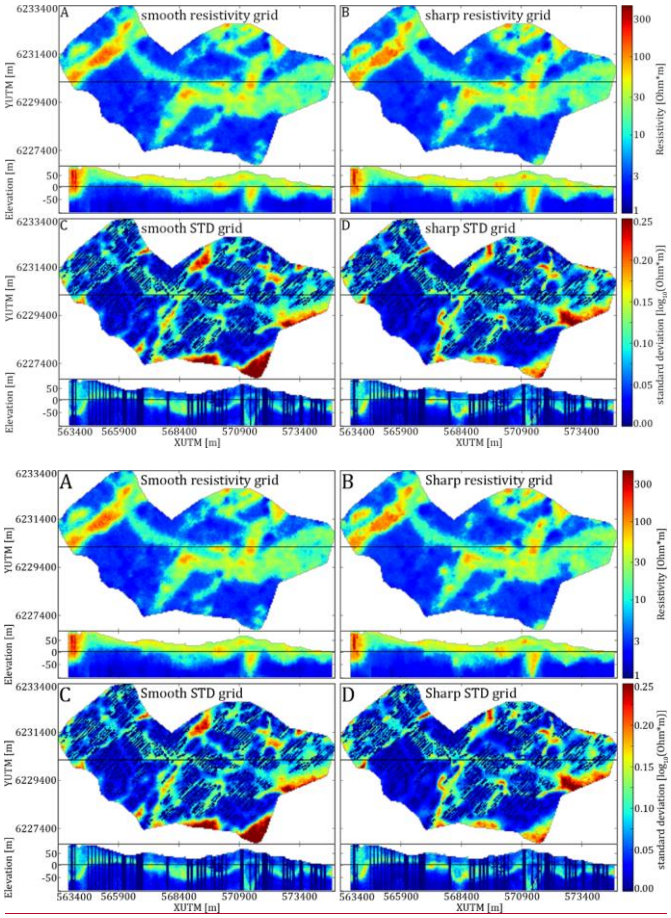


Figure 7: *An overview of the key differences between reconstructing the resistivity grid using smooth and sharp inversion resistivity models. Each grid is portrayed as a horizontal ~~cross-section~~ slice at 20 mbsl, and a vertical ~~slice~~ cross-section intersecting at UTMY 6230100 m. **A** shows 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 7~~Figure 7~~B). 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 7~~Figure 7~~C 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.

~~2.3~~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~~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 8~~Figure 8~~D-F) (For additional information see section A1 in the Appendix~~).~~). The tau model is then used to combine the SkyTEM (Figure 8~~Figure 8~~A-C) and borehole (Figure 8~~Figure 8~~D-F) probability grids (e.g. Journel, 2002; Krishnan, 2004; Remy et al., 2014). ~~The usage of the Tau model for combining soft data grids will be briefly described here. Suppose we have a set of data events,~~ $\boldsymbol{D_i}, i = 1, \dots, n$~~, and the goal is to estimate the probability that a hydrostratigraphic unit (A) is present provided all data events:~~

$$\boldsymbol{P(A|D_1, \dots, D_n)} \quad (\text{~~1~~})$$

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 5~~Figure 5~~A). Then the probability distributions~~, which are to be combined,~~ are defined: $P(A|\boldsymbol{D_1})(A|D_r)$ and $P(A|\boldsymbol{D_2})(A|D_b)$, where $\boldsymbol{D_1}D_r$ is the resistivity probability grid and $\boldsymbol{D_2}D_b$ the borehole probability grid. The 3D probability grids are translated into distance grids by applying the "*probability-into-distance*" and 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~~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 8~~Figure 8transform:~~G-I. The combined probability grids replace the smooth probability grids used in the basic setup.

~~$x_o = \frac{1-P(A)}{P(A)}$, $x_1 = \frac{1-P(A|\boldsymbol{D_1})}{P(A|\boldsymbol{D_1})}$, and $x_2 = \frac{1-(A|\boldsymbol{D_2})}{P(A|\boldsymbol{D_2})}$~~

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

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

~~where the tau values are chosen as follows: $[\tau_{resistivity}, \tau_{borehole}] = [2,1]$. The final conditional probability is computed as follows:~~

$$P(A|D_1, D_2) = \frac{1}{1+x}$$

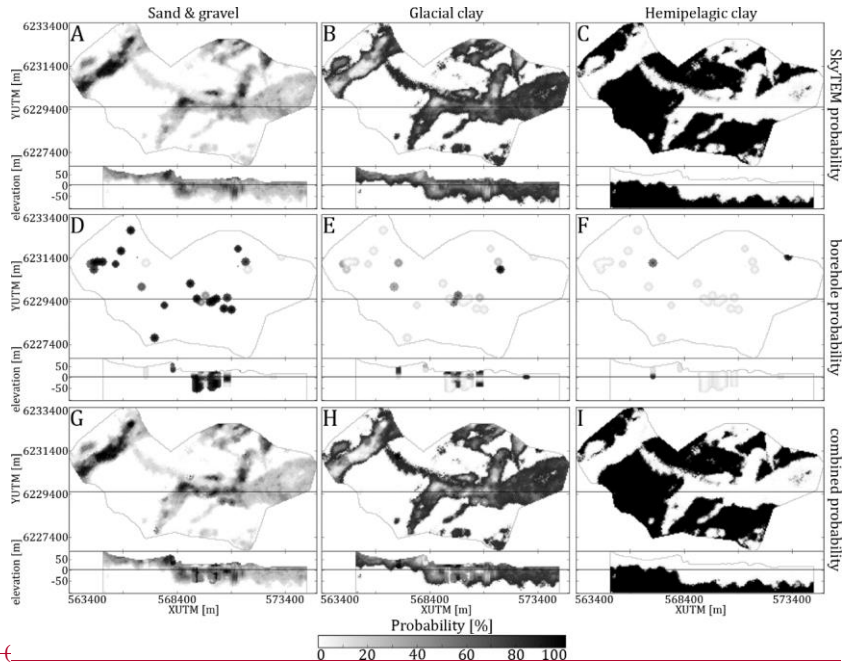
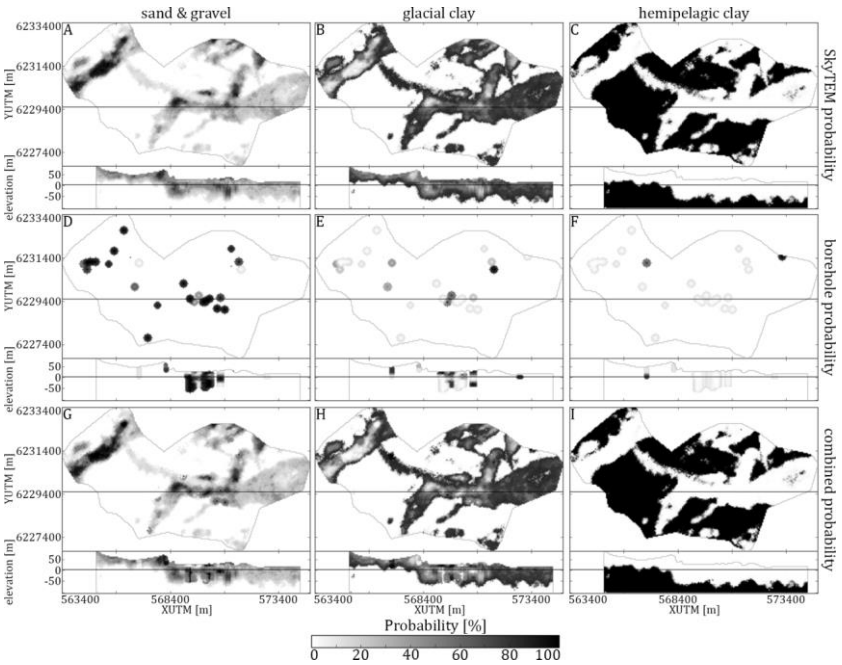


Figure 8)

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

$$x = x_0 \cdot \left(\frac{x_1}{x_0}\right)^{\text{resistivity}} \cdot \left(\frac{x_2}{x_0}\right)^{\text{borehole}}$$

The resulting combination of the three hydrostratigraphic probability grids is seen in G–I. The combined probability grids are used in the basic setup, in place of the smooth probability grids.



~~Figure~~ : *A visual representation of the Tau model procedure for combining the soft resistivity and borehole grids. Each grid is portrayed as a horizontal ~~cross-section~~ slice at 20 mbsl, and a vertical ~~slice~~ 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:[~~$\tau_{resistivity}$~~, ~~$\tau_{borehole}$~~ τ_r, τ_b]=[2,1].*

~~2.3~~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.

~~2.4. Comparing Simulation results~~

~~Comparing a large set of extensive 3D models is a common problem encountered in stochastic MPS modeling. A common 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, , 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 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.~~

2.4.1 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 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, $\boldsymbol{u}$. The probability that the attribute S is equal to s_k, $P_k(\mathbf{u})$, which is computed as follows:

$$\boldsymbol{P_k(u) = \frac{1}{N_{reals.}} \sum_{i=1}^{N_{reals}} (s_{k,i}(u) == s_k)} \quad (\,)$$

where 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:

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

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 location $\mathbf{u}$ computed using eq. (4).

The EMR-values are then computed for each grid cell using eq. (3) and (5), 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, if 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 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.

2.4.2 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 a EDT (Maurer et al., 2003). The EDT computes the Euclidean Distances for all locations of a binary grid, *i.e.* a grid containing only two states. For a given grid cell, with n-dimensional location vector $\boldsymbol{u}$, the Euclidean Distance is computed between location $\boldsymbol{u}$, and all other locations in the grid, $\boldsymbol{v_i}, i = \{1, \dots, N_{cells}\}$, with $\boldsymbol{u \neq v_i}$ and N_{cells} being the number of cells contained in the grid:

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

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 (A-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 (E-H). The 2D binary layers are then translated into 2D d_{EDT}-layers by using eq. (6) 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 I-L, are used to compute an average Euclidean Distance between each realization, $realization_i$, and the cognitive geological model, $cog.model$:

$$\overline{\Delta d_{EDT}}(cog.model, realization_i) = \frac{1}{N_{cells}} \sum_{j=1}^{N_{cells}} \left[d_{EDT}^{cog.model}(u_j) - d_{EDT}^{realization_i}(u_j) \right] \quad (6)$$

where $i \in \{1, \dots, N_{realizations}\}$, with $N_{realizations}$=50, and N_{cells} being the number of cells in the, in this case, 2D layer. The $\overline{d_{EDT}}$, eq. (7), 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, $\overline{d_{EDT}}$, as seen in the , where the realization which is closest to the cognitive geological model (B, F and J) has a $\overline{d_{EDT}}$ value of 240 m, while the realization which was ranked 25th closest (C, G and K) has an $\overline{d_{EDT}}$ value of 280 m, and lastly the realization which was farthest (D, H and L) has a $\overline{d_{EDT}}$ value of 310 m. It should be noted that the $\overline{d_{EDT}}$ computation, described by eq. (7), is not limited to comparing a realizations 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. (7) can be defined as follows:

$$\overline{\Delta d_{EDT}}(model_A, model_B) = \frac{1}{N_{cells}} \sum_{j=1}^{N_{cells}} \left[d_{EDT}^{model_A}(u_j) - d_{EDT}^{model_B}(u_j) \right] \quad (6)$$

Where the number of cells in model$_A$ must be equal to the number of cells in model$_B$, *i.e.* $N_{cells}^{model_A} = N_{cells}^{model_B} = N_{cells}$.

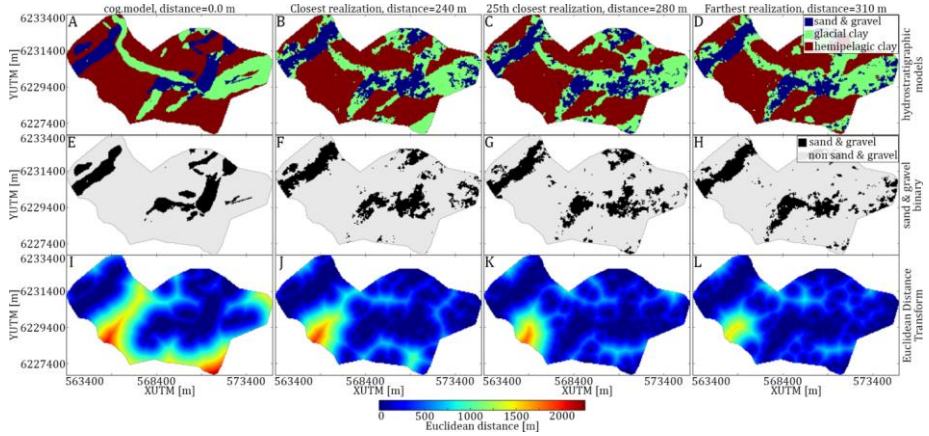


Figure 5: *A 2D example of the Euclidean Distance Transforms (EDT) as a measure for the similarity between categorical MPS realizations. In this example a set of 50 realizations, from the basic modeling setup, are compared based on the differences in EDT for sand & gravel units. **A-D** shows the hydrostratigraphic 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 $\overline{d_{EDT}}$ computations on 3D hydrostratigraphic grids. Furthermore, the $\overline{d_{EDT}}$ 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.

2.4.3 Evaluating the distance matrix

The average Euclidean Distance difference, $\overline{d_{EDT}}$, 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 using Equation (6) and is defined as follows:

$$\boldsymbol{D}_{i,j} = \overline{d_{EDT}}(realization_i, realization_j)$$

~~where $i, j = \{1, \dots, N_{realizations}\}$. 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 . 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.* , 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.~~

~~3~~5 Results

~~3~~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~~Figure 9~~), along with an EMR-map (Figure 10~~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~~Figure 9~~A and Figure 10~~Figure 10~~A, reveal the same overall trends as the cognitive geological model, Figure 4~~Figure 4~~A. 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 *glacial clay* valley striking ~N50°E to the south. However, even though the main hydrostratigraphic architecture of the 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 ~~south-east~~southeast corner of the Kasted survey area (Figure 9~~Figure 9~~A) (Figure 10~~Figure 10~~A). The same sporadic picture is seen in the vertical slices of the realizations (Figure 9~~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~~Figure 9~~A) as in the cognitive geological model (Figure 4~~Figure 4~~A). Furthermore, the EMR-map reveals that the valley margins are subject to a larger degree of ambiguity (Figure 10~~Figure 10~~A), 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.

The Case 1a realizations (Figure 9~~Figure 9~~B) (Figure 10~~Figure 10~~B), which use the Egebjerg TI (Figure 4~~Figure 4~~C), 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 4~~Figure 4~~C). 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 influenced by the choice of TI, especially when Case1b is also considered (Figure 9~~Figure 9~~C). The hydrostratigraphic realizations of Case1b (Figure 9~~Figure 9~~C) (Figure 10~~Figure 10~~C), 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 ~~to~~ 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 hydrostratigraphic units.

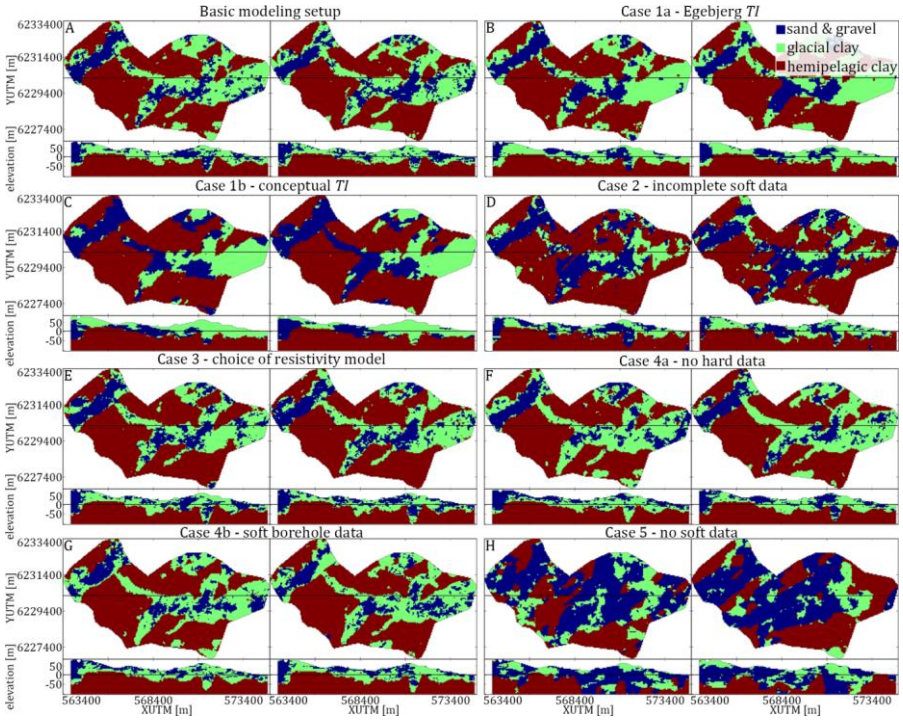
The importance of reconstructing the incomplete resistivity grid is seen in the Case 2 (Figure 9~~Figure 9~~D) (Figure 10~~Figure 10~~D). The two realizations in Figure 9~~Figure 9~~D show the main buried valley features, *e.g.* the western *sand and gravel* valley. However, the hydrostratigraphic units 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 ~~and vertical slice ((~~Figure 9~~Figure 9~~D). The EMR-map also reveals an increase in low occurrence ratios in areas without soft data (Figure 10~~Figure 10~~D).

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~~Figure 9~~E) are quite similar to the basic modeling setup realizations (Figure 9~~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~~Figure 9~~E) in the ~~south-west~~southwest corner, 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 10~~Figure 10~~E).

Field Code Changed

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~~Figure 9~~F and G shows two hydrostratigraphic realizations without hard data and with soft borehole data, respectively. These 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~~Figure 9~~F and G). The EMR-maps reveal that without boreholes (Figure 10~~Figure 10~~F) 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 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~~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~~Figure 9~~H) and EMR-map reveals a significant decrease in the occurrence ratio, and thus an increase in the ambiguity of the model ensemble (Figure 10~~Figure 10~~H).


Basic modeling setup
Case 1a - Egebjerg TI
Case 1b - conceptual TI
Case 2 - incomplete soft data
Case 3 - choice of resistivity model
Case 4a - no hard data
Case 4b - soft borehole data
Case 5 - no soft data
A
B
C
D
E
F
G
H
sand & gravel
glacial clay
hemipelagic clay
YUTM [m]
elevation [m]
XUTM [m]
6233400
6231400
6229400
6227400
50
0
-50
563400
568400
573400

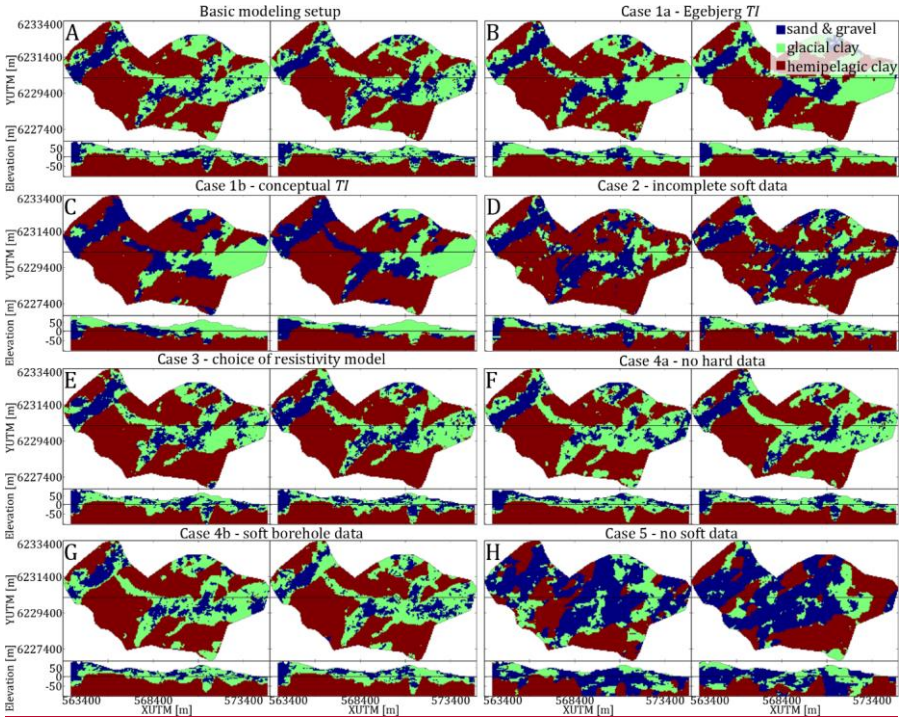
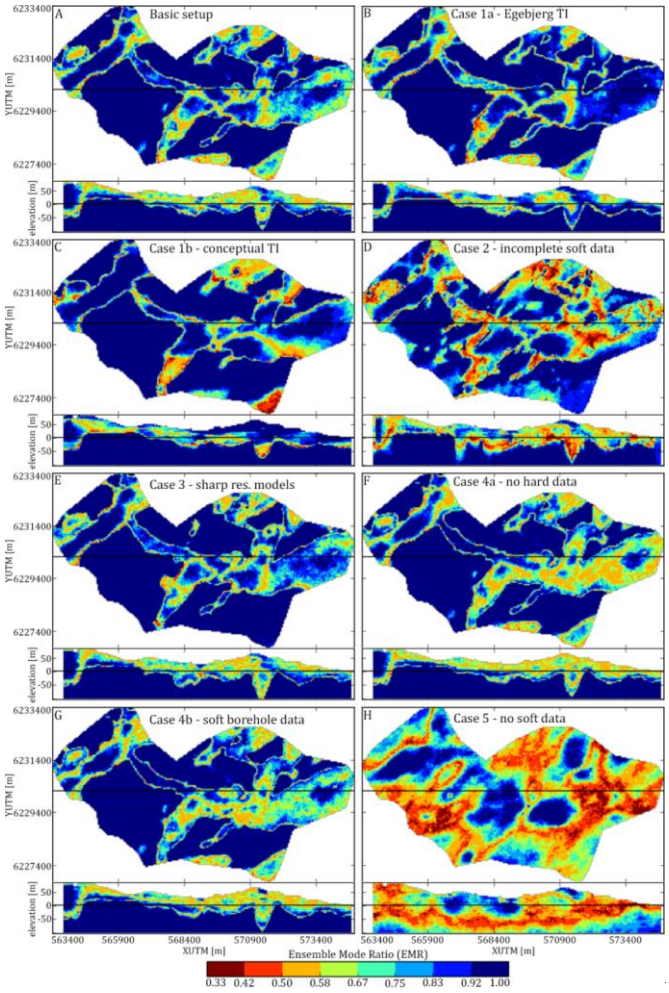


Figure 9: *Each case is displayed by two realizations, realization #1 of 50, and realization #30 of 50. Each realization is portrayed as a horizontal slice at 20 mbsl, and a vertical cross-section intersecting at UTMY 6230100 m. A shows the realization results for the Basic modeling setup, B shows the realization results for Case 1a ~~which uses the Egebjerg TI.~~, C shows the realization results for ~~the~~ Case 1b ~~in which the conceptual TI was used for simulation~~, , D shows the results for Case 2 ~~where the simulation is constrained to an incomplete soft data grid~~, , E shows the results for Case 3 ~~where the simulation is constrained to the sharp resistivity models~~, F shows the results for Case 4a ~~where the simulation are run without borehole information~~, G shows the results for Case 4b ~~where the simulations are constrained to soft borehole data~~, and H shows the results for Case 5. For more details on individual cases the reader is referred to* Table 1~~Table 1 where the simulation is run without the soft resistivity data~~.

Formatted: Font: Not Bold


A
Basic setup
B
Case 1a - Egebjerg TI
C
Case 1b - conceptual TI
D
Case 2 - incomplete soft data
E
Case 3 - sharp res. models
F
Case 4a - no hard data
G
Case 4b - soft borehole data
H
Case 5 - no soft data
YUTM [m]
6233400
6231400
6229400
6227400
elevation [m]
50
0
-50
563400
565900
568400
570900
573400
XUTM [m]
Ensemble Mode Ratio (EMR)
0.33 0.42 0.50 0.58 0.67 0.75 0.83 0.92 1.00

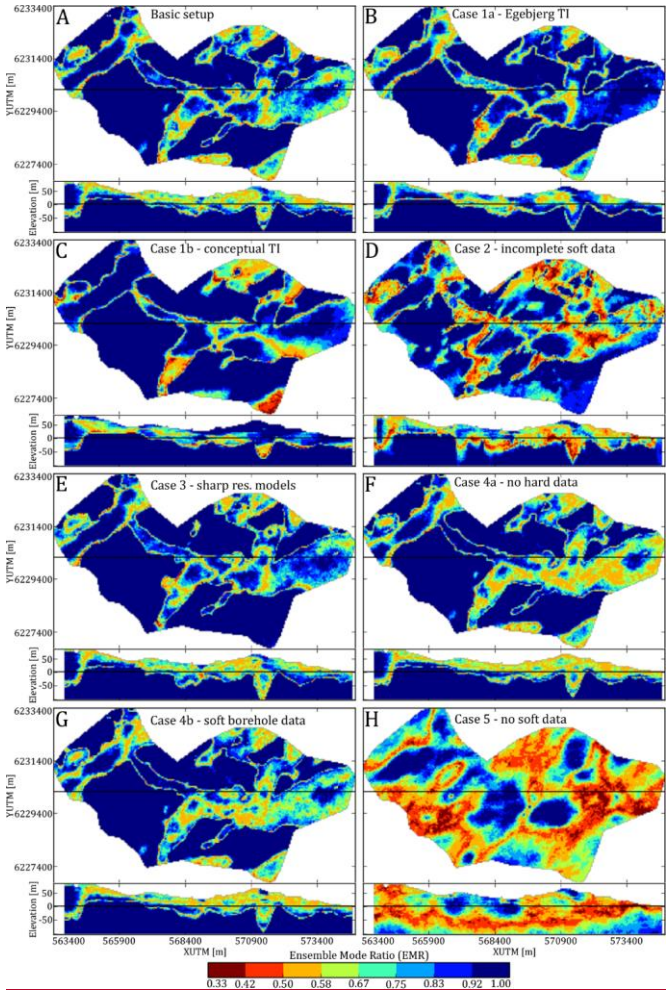


Figure 10: *A presentation of the "ensemble mode ratio" (EMR) maps, computed for the different case ensembles of hydrostratigraphic models. Each EMR map is presented as a horizontal slice centered on 20 mbsl, and a vertical cross-section intersecting at UTMY 6230150m; A-H presents the EMR-type uncertainty map for each of the different cases, which are summarized in* Table 1~~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.*

Formatted: Font: Not Bold

~~3~~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. (~~6~~8) and (~~7~~10). The full distance matrix is presented in Figure 11~~Figure 11~~A. The distances between each realization and the cognitive geological models have also been computed and plotted in Figure 11~~Figure 11~~B. To aid the interpretation of the distance matrix and distances to the cognitive model a summary table, Table 3~~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

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.

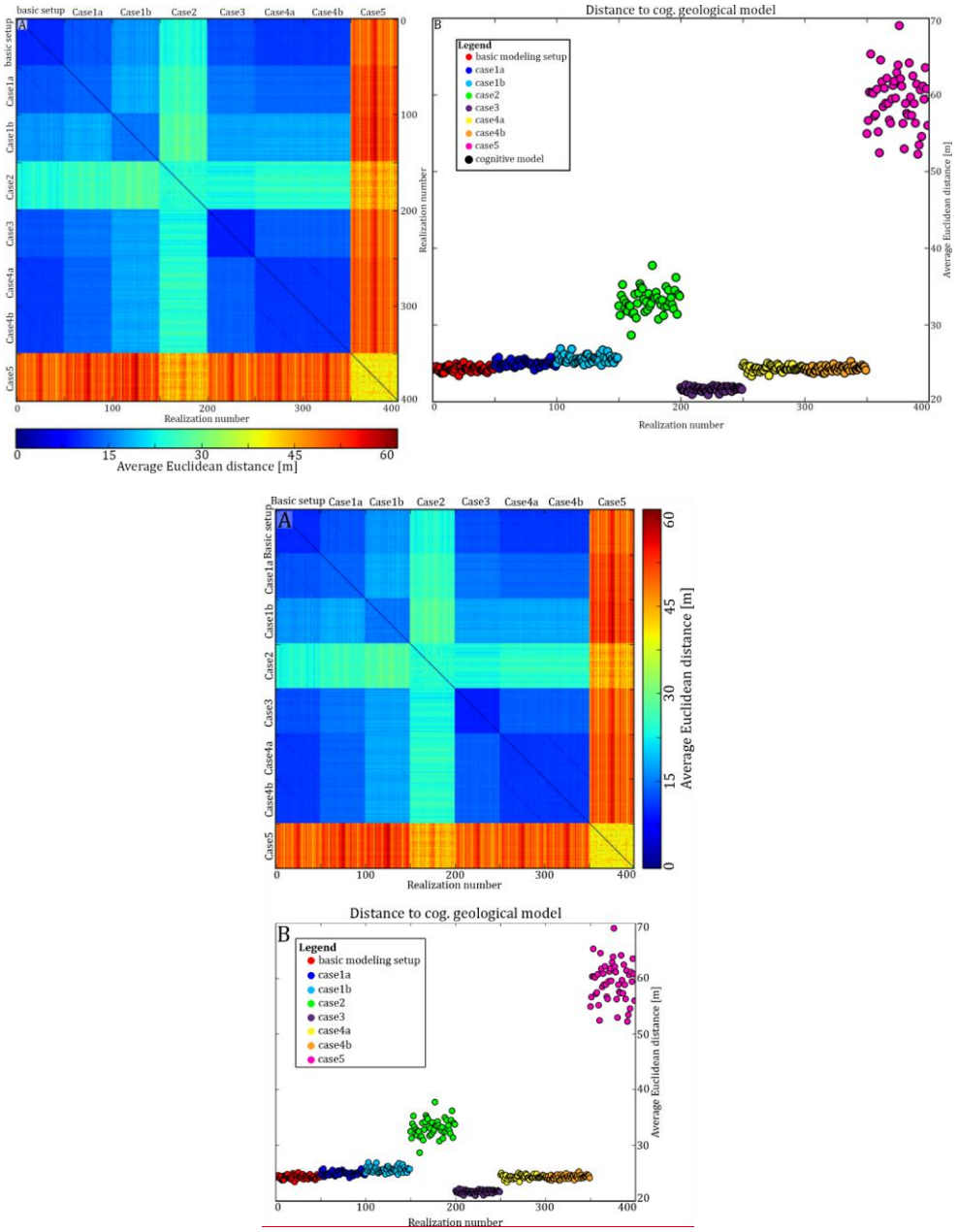
The 3D geological conceptualization contained in the TI influences the final hydrostratigraphic realizations as illustrated in Case 1, which is divided into two sub-cases: Case 1a and Case 1b. In Case 1a, using a 3D cognitive geological model from the Egebjerg area as a TI for hydrostratigraphic simulation increases the average distance to the cognitive model to 24.9 m (Figure 11~~Figure 11~~A) (Table 3~~Table 3~~). Furthermore, the average variability has increased to 13.6 m (Figure 11~~Figure 11~~B) (Table 3~~Table 3~~). The other sub-case ~~revolves~~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 11~~Figure 11~~A) (Table 3~~Table 3~~). The ensemble variability has increased to 14.8 m (Figure 11~~Figure 11~~B) (Table 3~~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 11~~Figure 11~~A) (Table 3~~Table 3~~). The distance to the cognitive geological model is also large, with an average value of 33.1 m (Figure 11~~Figure 11~~B) (Table 3~~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 11~~Figure 11~~B) (Table 3~~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~~Table 3~~. Recalling the raw hydrostratigraphic realizations (Figure **9**~~Figure 9~~E) and the EMR-map (Figure 10~~Figure 10~~E), 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 1~~Figure 1~~B), which are used for creating the probability grids. The ensemble average variability is 10.7 m (Figure 11~~Figure 11~~A) (Table 3~~Table 3~~) and the average distance to the cognitive model is 24.3 m (Figure 11~~Figure 11~~B) (Table 3~~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 11~~Figure 11~~A) (Table 3~~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.



820 **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~~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~~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

46 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, ~~increase~~increased the fingerprint of resistive extreme values, which in turn results in less ambiguous reconstructed resistivity grids; compare Figure 7~~Figure 7~~C 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 ~~Case5~~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 1~~Figure 1~~B). 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.

~~The idea of using~~ In case 4b, the borehole data was used as soft information~~,~~ as ~~seen with Case 4b, was motivated~~ in the study by Høyer *et al.* (2017)~~the fact that borehole information is often considered as hard information, *i.e.* they are not~~. This was done since boreholes are associated with ~~an uncertainty value. In this study the borehole dataset contains information from a variety of sources, and can be associated with a degree of~~ 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~~Figure 78~~).

Field Code Changed

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 ~~alter.~~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~~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**~~Figure 9~~C and Figure 10~~Figure 10~~C, it can be seen that the *glacial clay valley*, both present in the soft data variable (Figure 8~~Figure 8~~B) and the cognitive Kasted geological model (Figure 4~~Figure 4~~A), 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 ~~*snesim*~~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. ~~This DS implementation uses an alternative method for constraining continuous auxiliary data, presented by~~Whereas Chugunova and Hu (2008)~~. The~~ present a MPS method ~~requires the construction of multivariate TI, where~~for constraining a categorical ~~TI is coupled with~~simulation by a continuous auxiliary ~~variable, which is used during simulation to constrain the soft~~ data ~~variable.~~, 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 ~~the auxiliary~~every variable reflects ~~both the complications related to the given geophysical method, as well as the local petrophysical~~ the spatial relationship to be modeled. In our context, one could envision creating a bivariate TI consisting of a hydrostratigraphic ~~relations. Creating such~~model and a continuous auxiliary variable reflecting an ~~auxiliary variable, which spatially overlaps with the TI, and reflects the 3D~~ 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~~, which uses~~ 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 11~~Figure 11~~A) 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 11~~Figure 11~~A),

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.

~~5~~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.

~~6~~8 Acknowledgements

~~I~~We would like to thank Senior Research Engineer Celine Scheidt of Stanford University for pointing us in the direction of using Euclidean Distance Transforms for comparing MPS realizations. Three anonymous reviews and editor are thanked for valuable comments clarifying the manuscript. This study is supported by HyGEM, Integrating geophysics, geology, and hydrology for improved groundwater and environmental management, project no. 11- 116763. The funding for HyGEM is provided by The Danish Council for Strategic Research.

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 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 modeling using multiple-point statistics and airborne transient electromagnetic methods, Hydrol. Earth Syst. Sci., 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 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. 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.

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 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.

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 determining groundwater protection, Near Surf. Geophys., 13(1), 45–60, doi:10.3997/1873-0604.2014044, 2015.

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

stochastic geological modeling using airborne geophysical data and borehole data, Water Resour. Res., 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, , 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. Stanford Center for Reservoir Forecasting Annual Report. Stanford University, Stanford, CA,.

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.

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.

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.

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.

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), 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.

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, doi:10.1016/0098-3004(94)90069-8, 1994.

Viezzoli, A., Christansen, A. V., Auken, E. and Sørensen, K..: Quasi-3D modeling of airborne TEM data by spatially constrained inversion, Geophysics, 73(3), F105-F113, 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.

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.

Appendix

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

Variogram model type:
Exponential

$$mean_{SK} = \sout{1/3}\ 1/3$$

Search ellipsoid:

	Max	Med	Min
Ranges	200	200	10
Angles	0	0	0

Variogram:

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]

Marginal cdf:

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