## Response to the reviews

The authors thank the editor and the four referees for the comments to the manuscript. Please find all comments with answers below.

General comments from all referees are presented first, then comments given as notes in the PDF version of original manuscript are presented, and finally the marked-up manuscript are presented.

Revised figures are Figure 4, Figure 6, Figure 7, Figure 10, Figure 11, and Figure 12.

(Ref#1)

*"The only concern I have is that the paper is partly very close to some recently published papers (Foged et al., 2014, HESS, Hoyer et al., 2015, JAG). Therefore, the authors should check their paper with respect to avoiding repetition and they should focus on the differences."*

More specifically Referee#1 questions the relevance of section 2.2, the first paragraph of section 3.2 and Figure 2, Figure 4, and Figure 8, as they can all be replaced with references to Foged et al., 2014, HESS.

Section 2.2 has been deleted and a small paragraph has been added to section 2.1 in replacement.

Figure 2: Three profiles showing resistivity, clay fraction, and clusters respectively. We find this figure useful for intuitive and visual explanation of our work.

Figure 4: The basemap of geophysical and lithological information. This figure has been deleted and a reference to Foged et al., 2014, HESS is given instead.

Figure 8: Figure 8 is kept in the revised manuscript.

(Ref#2)

*"However, I agree with referee 1 that there might be some confusion about the novelty of the paper. Especially, the title of the paper is misleading since the methodology to derive the hydrostratigraphy is the subject of the paper by Foged et al. (2014) and that the field case seems to be the same. The originality of the paper here is rather to compare several models with various complexity based on their hydrogeological output. Some work should thus be done to focus the paper more on its original aspects and not on already published material (already in the introduction)."*

We have removed section '2.2 Geophysical data' and the lithological borehole descriptions in section '2.2 Hydrostratigraphic model' and added 7 lines to section '2.1 Study area' with a brief summary instead. Also Figure 5 (base map of geophysical data and borehole locations) has been deleted. We have deleted the last 6 lines of the paragraph of the CF-inversion description (The current section '2.2 Hydrostratigraphic model').

*"Another concern, from my point of view, is that the authors claim that they reduce the structural uncertainty of their model. For me, the methodology is not sufficient to do that. Indeed, once the number of*

*cluster is chosen, a unique model is drawn. It is true that this model is now based on geophysical data, but a unique model cannot be used to assess the uncertainty. Structural uncertainty has two causes: (1) the conceptualization of the geology, using alternative geological scenarios for example and (2) spatial uncertainty inside the geological scenarios itself linked to the location of the different facies and their respective relationships. Assessing structural uncertainty would require to allow for more models. It is well-known that geophysical inversion is not unique and that several models may explain the same data. This should be included in a structural uncertainty analysis which is not the case here since geophysics and clay fraction distribution are taken as certain. Moreover, the proposed methodology has two inversion steps (AEM data and CF data). Consequently, the limitations of arising from the inversion are present two times, such as the varying resolution, the effect of regularization and the risk of artifcats of inversion. Those limitations should be clearly stated in the text since it makes the model quite deterministic."*

The comments regarding structural uncertainty are appreciated, and we agree that our approach as presented in the paper is purely deterministic; one single hydrogeological cluster model is presented. We do not address model structural uncertainty in this paper. We believe that this comment on structural uncertainty arises from a misunderstanding in section '3.4 Advantages and limitation'. We have therefore deleted the sentence "Also reproducibility and especially possibility of uncertainty quantification of the hydrostratigraphic cluster model are important features.", and replaced it with "We believe that the cluster model approach presented in this paper can be extended to address structural uncertainty and its impact on hydrological predictions" further down in the same paragraph.

*"More details should be given on how the CF model is obtained. The petrophysical relationship is not shown. What are the two parameters linking clay fraction and resistivity? What are their respective ranges of variation? Are the obtained values physically plausible? With a locally dependent relationship, it is always possible to find a straightforward relationship which could subsequently affect the results. Few is said about the different scales for each data and model, borehole logs are fine-scale, geophysical data are representative of large volume and the hydrogeological model is a large scale model."*

We have added the petrophysical relationship to section '2.2 Hydrostratigraphic model' and included Figure 4 which is essential for understanding the petrophysical relationship. For a discussion about the scale difference between lithological information and AEM data we have added a sentence at the end of the same paragraph with a reference to Foged et al., 2014, HESS.

*"I do not really see the usefulness of the reference model. It is not clear how many facies are defined for this reference. Maybe a comparison of the approach with a method based on the direct interpretation of resistivity in hydrofacies would be more suited, to see the advantage of adding the CF step."*

The reference model has been removed from the manuscript. Because we believe it is relevant to benchmark the performance of the hydrostratigraphic, we have added a table (Table 2) which shows hydrological performance statistics of four comparable Danish models. The section '3.3 Benchmarking hydrological performance' therefore still exists and the last paragraph in the section refers to the models presented in Table 2.

(Jan Gunnink)

*"One issue that needs clarification is the fact that the airborne EM data is used in the inversion into resistivity, next the resistivity in an inversion together with borehole data in the Clay Fraction model and*

*finally the clustering method uses both resistivity and Clay Fraction (which is again a derivative of the resistivity). This repetitive use of the airborne EM data is of course very clever, because the resistivity datasets provides a dense 3D coverage, but it also needs some more explanation of the potential pitfalls. There is the danger of using the same data more than once that errors that are present in the data (as there always are) propagate along several ways into the end product."*

We basically start with two different and independent datasets: Resistivity from AEM and borehole lithology. Resistivity and borehole lithology are subsequently combined into the clay fraction model. However, resistivity variations are only partly due to clay content, so there is extra information in the resistivity which we want to retain for the clustering. Using the principal component analysis, we split the resistivity and clay fraction dataset into a correlated and an uncorrelated component and these two then go into the clustering algorithm. The clustering is not a data processing step but a spatial aggregation step. A formal evaluation of the error propagation into the final hydrostratigraphic cluster model is not performed in this paper, however we have added a paragraph to section '3.4 Advantages and limitations' where we address the smoothing effects that originate from the inversion steps. Also we write how we wish to address these issues with future work.

*"Also, by using the same dataset in different - but related - methods, one might end up in a "chicken-egg" situation. Is the result achieved by processing the same dataset sequentially using different methods really contributing to an improved end-product."*

This question is not easy to answer as we do not know the true hydrostratigraphy of the subsurface. The rationale in this paper is to let the hydrologic calibration decide whether or not the workflow results in a useful subsurface model in terms of being able to match hydrologic observations. We find that our approach is equally good or slightly better than the available benchmarks in that respect.

*"Especially the fact that the combination of resistivity and borehole data is both used in the Clay Fraction model and in the clustering needs some better explanation."*

The clustering is performed in a 2-dimensional attribute space. Each subsurface voxel has two attributes, clay fraction and resistivity. Please remember that we start with two independent datasets (resistivity and borehole lithology), so the use of two attribute dimensions is entirely justified. Correlations between resistivity and clay fraction are handled with the principal component analysis.

*"Are we looking at real improvements or is it merely presenting the correlated datasets in different ways? For example, why not using a simple cut-off in the Clay Fraction model to derive at clusters in the data? Because the resistivity is already used in making the Clay Fraction model, I would think that the result might not be very different. I understand that the Principal Component Analysis is crucial in the clustering method, but I am not sure that it is really contributing to a better end product. I am not saying that the cluster method is not correct to use here, but I would like to have a better understanding of the implications of the repetitive use of the same datasets."*

Yes, the CF-model alone could be used to do the clustering. However, the resistivity dataset contains signals that are not related to clay content (e.g. groundwater quality) and those should be retained in the clustering. The underlying assumption is that spatial patterns of resistivity/clay fraction are related to spatial patterns of hydraulic conductivity. We agree that empirical evidence for this hypothesis is required, which is the subject of future research.

*"Another issue with the clustering method is the fact that k-means clustering aims at producing clusters of approx. the same size. This might have an unwanted influence on the results of the clustering, since there seems no reason to assume that the clusters in the hydrostratigraphic model needs to be of equal size."*

Yes. However the clusters found for our data are not of equal size. For example, for the 5-cluster model, clusters represent 9%, 26%, 9%, 12%, and 44% respectively of the total number of voxels. The behaviour of the k-means clustering algorithm is controlled by the chosen distance measure. Here, we used the standard Euclidean distance measure. There is scope to experiment with different distance measures in the clustering in the future.

*"In the paper, another model is used as a "benchmark" or a so called reference model. This model I would not call a benchmark model, because it is also just a model! Although there is some geological knowledge inserted into this model, this knowledge is later collapsed into only four hydrological units. Keep in mind the phrase: "all models are wrong, some are useful". The results of the research in this paper indicates that for the purpose of hydrological modeling, the presented method shows better results in terms of deviation from measured groundwater heads, compared to another model."*

The reference model has been removed from the manuscript. Because we believe it is relevant to benchmark the performance of the hydrostratigraphic, we have added a table (Table 2) which shows hydrological performance statistics of four comparable Danish models. The section '3.3 Benchmarking hydrological performance' therefore still exists and the last paragraph in the section refers to the models presented in Table 2.

(Ref#4)

*"I agree with the remarks by Jan Gunnink and reviewers 1 and 2. The clay fraction concept should be explained in more details and examples because it is the link between geophysics and the geostatistic approach of this model."*

We have added the petrophysical relationship to section '2.2 Hydrostratigraphic model' and included Figure 4 which is essential for understanding the petrophysical relationship. As recommended by Referee#2 we have also added details about the regularization and how the final kriging variogram has been fitted. For a discussion about the scale difference between lithological information and AEM data we have added a sentence at the end of the same paragraph with a reference to Foged et al., 2014, HESS

*"It is good that a region at the Danish Baltic coast with complicated geological conditions due to 3 glaciations is chosen for the groundwater model. The reference model is also based on AEM data, for both models the hydraulic conductivity was found through calibration to water heads and discharge. So both models are quite similar and it is not surprising that the results are similar."*

The reference model has been removed from the manuscript. Because we believe it is relevant to benchmark the performance of the hydrostratigraphic, we have added a table (Table 2) which shows hydrological performance statistics of four comparable Danish models. The section '3.3 Benchmarking hydrological performance' therefore still exists and the last paragraph in the section refers to the models presented in Table 2.

(JG) p 1557 lines 24-25: "Could you explain what you mean by this term?" We realize that the term 'hydrostratigraphic model' has several definitions, and we have therefore added a short description of how we use the term in this paper.

(Ref#2) p 1557 lines 27-28: "Developments in geostatistics allow to take into account non-stationarity which can also be brought by indpendent data such as geophysics (for example, to define trends or various target proportions). For large-scale models, the limitation is rather the computational cost for running stochastic or MonteCarlo silmulations." We have elaborated on multiple point statistical methods for geophysical inversion in lines 62 to 65. The advantage of uncertainty estimation has been noted, and that applications commonly are at scales smaller than large-scale hydrological models. We have deleted the sentence about non-stationarity.

(Ref#2) p 1558 lines 3-4: "I agree, but there are not the only sources of information for large-scale models, geological maps are very helpful to understand the large-scale structure and geological analogs are useful for within layer variations" We have included this perspective in the end of the paragraph in which this comment was made.

(Ref#1) p 1558 lines 10-12: References have been updated

(Ref#2) p 1558 lines 16-17: "There exist some methodologies to try to quantify geological uncertainty. Besides traditional methods such as GLUE or BMA which may be time consuming since they require McMC simulations, a novel approach has been recently proposed to assess the uncertainty of geological scenarios using various kind of data prior to inversion (see Park et al. 2013, Scheidt et al. in press, Hermans et al. 2014). It is designed to work within stochastic geostatistical framework. Within scenario uncertainty can be studied using McMC or stochastic methods." Yes, we agree with this. We have therefore included this, including a reference as suggested, earlier in the introduction where geostatistical methods are discussed. Also, see the added lines about multiple point statistical methods in lines 62 to 65. However, the exact paragraph to which this comment was made, was directed not at geostatistical methods, but rather at cognitive large scale geological methods. To clarify the paragraph, we have added 'cognitive' in the beginning of the paragraph.

(JG) p 1558 line 19: "Briefly explain these concepts (Sequential, joint and coupled inversion)" We have supplied brief definitions of these concepts along with a reference for a more detailed definition. This has been rewritten with the added references suggested by Referee#2.

(Ref#1) p 1558 line 22: References have been updated

(Ref#2) p 1558 line 27-29: "The variation in space can be inherent to the deposits which are complex and data are not sufficient to draw this relationship, or to geophysical models resolution issues (E.g. Day-Lewis et al. 2005), for example the loss of resolution with depth. There inherent limitations of geophysical data which remains an indirect way to observe the subsurface." We agree that the original sentence on petrophysical relationships was brief, and did not successfully explain the problems related to petrophysical relationships. We have therefore elaborated on this to address issues related to the unknown relationship and within and between site variability for which there is rarely data available to determine.

(Ref#2) p 1559 line 22: "Probabilistic approaches can also be used, for example, in MPS, geophysical data can be use as a constraint by transforming the geophysical parameters into probability maps. This transformation is done through Bayes rule. It is largely used in geophysical data integration through classical Bayesian updating technique. Such methodology does not require a petrophysical relationship and include directly the uncertainty related to the geophysical inversion" We are not sure that we understand this comment correctly, but we are not aware of methods that entirely skip a petrophysical relationship.

(Ref#2) p 1559 line 26: "The resistivity data is already the results of an inversion process which may influence hugely the results. As I understand, you have a first inversion of resistivity, then a second inversion for clay fraction. It means that the limitations of inversions (for example the non-uniqueness of the solution) act at two level which may hugely smooth the results." We have added a section about these smoothing issues in the section '3.4 Advantages and limitations' section.

(Ref#1) p1560 lines 7-16: "Delete/rewrite this introduction as it is too short to really understand the message. On the other hand this introduction in its present form is not necessary here." This paragraph has been deleted, and the necessary content has been merged with the previous paragraph (end of the introduction)

(Ref#2) p 1561 lines 13-14: "Isn't it (annual pumping rates) problematic for calibrating the hydrogeological model?" This also depends on the location of the observation wells relative to the pumping well. However, the annual pumping rates may cause a bias in the estimated hydraulic conductivity trying to compensate for the mismatch in head due to incorrect pumping rates. The pumping rates are small compared to the overall water balance.

(JG) p 1561 lines 19-20: "What is the influence of wide-spread artificial drainage on the hydrographs that re going to be used later in the calibration / validation?" As also written in the 'Hydrological model calibration' section stream discharge is largely determined by the drainage constants.

(Ref#2) p 1563 line 17: "The petrophysical relationship should be shown here to se how the parameters are related. Are they physically based so that it is possible to control the ranges of variations of the parameter. Indeed, since the relationship can vary with space, it is always possible to find locally parameters that will fit the CF data." In section '2.2 Hydrostratigraphic model', we have included the translator function (Eq. 1), and added Figure 4 because it is essential for understanding the translator function. As recommended by Referee#2 the CF-inversion paragraph has been rewritten to include information about regularization in the CF-inversion and how the kriging model has been fitted.

(Ref#2) p 1563 line 21: "The volume represented by those measurements is quite different! What are the grid parameters for AEM and petrophysical relationship?" Foged et al. (2014) has a discussion on the difference in resolution between AEM data and borehole lithological data, and we have referred to this discussion in the revised manuscript.

(Ref#2) p 1563 line 21: "clay fraction are determined from geophysics, why not just using categories in CF but also using resistivities?" We cluster on CF-values and resistivity data, because resistivity data includes signals unrelated to clay that we want to retain in the clustering.

(Ref#2) p 1565 line 9: "How was the Paleogene clay surface determined? With AEM data? Then, it should not be de-activated, their low hydraulic conductivity should be the results of the zonation." The deeper parts of the domain (where the Paleogene clay is located) are included in the zonation, and the Paleogene clay stands out in the k-means clustering. However, because a very large part of the domain is Palaeogene clay, where we know water movement is limited, we set the lower boundary of the groundwater flow model according to the Palaeogene clay. The boundary was defined because including the entire low permeable Palaeogene clay domain is computationally very expensive.

(JG) p 1565 lines 25-26: "Could you describe / asses the effect of the fact that these variables are spatially constant on the results of the modelling?" For Norsminde we have no information about the spatial variability of the drain constant. However the reference to Hansen et al. (2013) has been elaborated on to answer this question, as spatially variable drain time constants were calibrated in that study.

(Ref#2) p 1570 lines 2-3: "The true K is unknown, increasing the number of cluster increases the SD, this is quite logical since we allow more parameters to be different" Yes. The plot is included because we are interested in seeing which number of cluster is reasonable to use, for example when does the SD increase to an unacceptable value.

(JG) p 1570 line 21: "I think you need to explain this analysis in some more detail. Explain why you think you are doing a better job than the reference model, while your Mean log(K) st.dev is higher for the 4 and 5 cluster model compared to the reference model." The reference model has been removed from the paper and replaced with a comparison with similar Danish hydrological models based on the RMSE on head.

(JG) p 1572 line 22: "In Fig. 7, RMSE is shown for the calibration period (red line), and in Fig. 6 also. The only difference is that the statistics in Fig. 7 are calculated on 35 wells. Why are the RMSE of the reference model different from Fig.6. And why using this split data set?" Figure 6 shows weighted RMSE while Figure 7 shows RMSE. The split dataset is used in order to validate the performance of the model outside of its calibration period. The validation is performed on a dataset (the 35 wells) from a different time period.

(JG) p 1573 line 26: "Explain or provide reference." We have explained what we mean in a few lines using an example.

(Ref#2) p 1576 line 4: "Your methodology does not consider structural uncertainty, since for a given number of cluster, it will yields similar results. It reduces the effect of the choice of the structure by integrating geophysical data." We agree, and this is not what we intend to say. We have clarified the text.

# ~~An automated method to build~~Performance evaluation of groundwater model hydrostratigraphy from airborne electromagnetic data and lithological borehole logs

P. A. Marker[1], N. Foged[2], X. He[3], A. V. Christiansen[2], J. C. Refsgaard[3], E. Auken[2], P. Bauer-Gottwein[1].

[1]Department of Environmental Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark.

[2]HydroGeophysics Group, Department of Geoscience, Aarhus University, Aarhus, Denmark.

[3]Geological Survey of Denmark and Greenland, Copenhagen, Denmark.

Correspondence to P. A. Marker (paam@env.dtu.dk)

Formatted: No bullets or numbering

## ~~1~~ Abstract

Large-scale ~~integrated~~ hydrological models are important decision support tools in water resources management. The largest source of uncertainty in such models is the hydrostratigraphic model. Geometry and configuration of hydrogeological units are often poorly determined from hydrogeological data alone. Due to sparse sampling in space, lithological borehole logs may overlook structures that are important for groundwater flow at larger scales. Good spatial coverage along with high spatial resolution makes airborne ~~time-domain~~ electromagnetic (AEM) data valuable for the structural input to large-scale groundwater models. We present a novel method to automatically integrate large AEM data-sets and lithological information into large-scale hydrological models. Clay-fraction maps are produced by translating geophysical resistivity into clay-fraction values using lithological borehole information. Voxel models of electrical resistivity and clay fraction are classified into hydrostratigraphic zones using k-means clustering. Hydraulic conductivity values of the zones are estimated by hydrological calibration using hydraulic head and stream discharge observations. The method is applied to a Danish case study. Benchmarking hydrological performance by comparison of ~~simulated~~performance statistics from comparable hydrological ~~state variables~~models, the cluster model performed competitively. Calibrations of 11 hydrostratigraphic cluster models with 1-11 hydraulic conductivity zones showed improved hydrological performance with

increasing number of clusters. Beyond the 5-cluster model hydrological performance did not improve. Due to reproducibility and possibility of method standardization and automation, we believe that hydrostratigraphic model generation with the proposed method has important prospects for groundwater models used in water resources management.

## ~~2~~1 Introduction

Large-scale distributed ~~integrated~~ hydrological and groundwater models are used extensively for water resources management and research. We use large-scale to refer to models in the scale of 100 km $^2$ to 1,000 km$^2$ or larger. Examples are: water resources management in water scares regions (Gräbe et al., 2012; Laronne Ben-Itzhak and Gvirtzman, 2005); groundwater depletion (Scanlon et al., 2012); contamination (Li and Merchant, 2013; Mukherjee et al., 2007); agricultural impacts on hydrogeological systems (Rossman and Zlotnik, 2013); and well capture zone delineation (Moutsopoulos et al., 2007; Selle et al., 2013).

Field Code Changed

Such models are typically distributed, highly parameterized, and depend on data availability to sufficiently represent the modelled systems. Model parameterization includes, for example, the saturated and unsaturated zone hydraulic properties, land use distribution and properties, and stream bed configuration and properties. Hydrological forcing data such as precipitation and temperature are also required. Parameters are estimated through calibration, which requires hydrological observation data commonly in the form of groundwater hydraulic heads and stream discharges. Calibration data should be temporally and spatially representative for the modelled system, and so should validation data sets.

One of the main challenges in modelling large-scale hydrogeological systems is data scarcity (Refsgaard et al., 2010; Zhou et al., 2014). Uncertainty inherent in distributed hydrological models is well known (Beven, 1989). ~~Incorrect system representation due to lack of data contributes to this uncertainty, but most important source of uncertainty in distributed groundwater models is incorrect representation of geologic structures (Refsgaard et al., 2012; Seifert et al., 2012; Zhou et al., 2014).~~Incorrect system representation due to lack of data contributes to this uncertainty, but the most important source of uncertainty in distributed groundwater

models is incorrect representation of geological structures (Refsgaard et al., 2012; Seifert et al., 2012; Zhou et al., 2014). In this paper, we refer to a 3D subsurface model that delineates the structure of the hydraulic conductivity (K) field as a hydrostratigraphic model.

Formatted: English (U.S.)

Lithological borehole logs are the fundamental data source for constructing hydrostratigraphic models. The modelling process is often cognitive, but also two-point geostatistical ~~methods are used~~ (He et al., 2013; Strebelle, 2002)~~. Geostatisical approaches are less suitable for large scale models because they assume stationarity within the modelled geological domain and hence the model area often needs to be subdivided into several stationary geological domains. Spatial inconsistent sampling pattern and scarcity make lithological borehole logs alone insufficient to capture local scale geological structures relevant for simulation of groundwater flow and contaminant transport.~~

~~Airborne time-domain electromagnetic (AEM) data is unique with respect to good spatial coverage and high resolution. AEM is the only technique that can provide high resolution subsurface information at regional scales. Geological structures and heterogeneity, which spatially scarce borehole lithology data may overlook, are well resolved in AEM data. Geophysical data and especially AEM data are commonly used to support lithological borehole information in geological mapping and modelling (Bosch et al., 2009; Høyer et al., 2011; Jorgensen et al., 2010; Jørgensen et al., 2013; Sandersen and Jorgensen, 2003).~~

, and multiple-point statistical (e.g. Park et al., 2013) methods are used. Geostatistical methods have the advantage of uncertainty estimation. Spatially inconsistent sampling pattern and scarcity make lithological borehole logs alone insufficient to capture local-scale geological structures relevant for simulation of groundwater flow and contaminant transport. Cognitive methods have the advantage of using information from geological maps to assist interpretation of larger scale geological features.

Airborne electromagnetic (AEM) data is unique with respect to good spatial coverage and high resolution. AEM is the only technique that can provide subsurface information with a resolution down to ~25 m in the horizontal and ~5 m in the vertical at regional scales (Schamper et al., 2014). Geological structures and heterogeneity, which spatially scarce borehole lithology data may overlook, are well resolved in AEM data.

Geophysical data and especially AEM data are commonly used to support lithological borehole information in geological mapping and modelling (Bosch et al., 2009; Burschil et al., 2012; Høyer et al., 2011; Jorgensen et al., 2010; Jørgensen et al., 2013; Steinmetz et al., 2014). Also multiple point statistical methods are applied to invert geophysical data, where a priori geological information is incoporated through training images (e.g. Caers and Hoffman, 2006; Lange et al., 2012; Lochbuhler et al., 2015). Although uncertainty of the estimated structures is available from the inversion, multiple point statistical methods are applied at scales smaller than large-scale hydrological models. He et al. (2014) used transitions probabilities (two point statistics) to integrate AEM data with borehole lithologial data.

Current practice for cognitive hydrostratigraphic and geological model generation faces a number of challenges: structures that control groundwater flow may be overlooked in the manual 3D modelling process; geological models are subjective, and different geological models may result in very different hydrological predictions; structural uncertainty inherent in the model building process cannot be quantified. Currently there is no standardized way of integrating high resolution AEM into hydrogeological models.

Sequential, joint and coupled hydrogeophysical inversion methods ~~have been developed and used extensively in hydrological and groundwater research~~, as defined by Ferré et al. (2009), have been developed and used extensively in hydrological and groundwater research. In sequential inversion hydrological and geophysical models and inversions are set up and performed separately (e.g. Binley et al., 2001; Kemna et al., 2002). In joint inversion hydrological and geophysical models are set up separately but hydrological and geophysical parameters are estimated simultaneously through a joint objective function (e.g. Hyndman and Gorelick, 1996; Hyndman et al., 1994; Linde et al., 2006; Vilhelmsen et al., 2014). In coupled inversion only one model is set up, the hydrological, and the geophysical data is evaluated by comparison to translated simulated hydrological states (e.g. Hinnell et al., 2010; Kowalsky et al., 2005). The methods have been applied to capture hydrological processes or estimate aquifer properties and structures from geophysical data ~~(Hinnell et al., 2010).~~. Hydrogeophysical inversion addresses hydrogeological property estimation or delineation of hydrogeological structures. In the context of large-scale groundwater models studies, Dam & Christensen (2003) and Herckenrath et al. (2013) translate between hydraulic conductivity and electrical

resistivity to estimate hydraulic conductivity parameters of the subsurface in a joint hydrogeophysical inversion framework. ~~Petrophysical relationships however are not well known and vary in space, which makes a fixed translation between geophysical and hydrological parameter space problematic.~~Petrophysical relationships, however, are uncertain, partly because of unknown physical relationship between geophysical and hydrological parameter space. The relationship may vary within and/or between field sites depending on given conditions and cannot be determined a priori. For electrical resistivity versus hydraulic conductivity, relationships suggesting both positive and negative correlation have been found (Purvance and Andricevic, 2000). Herckenrath et al. (2013) concluded that sequential hydrogeophysical inversion was preferred over joint hydrogeophysical inversion due to the uncertainty associated with the ~~translator function.~~petrophysical relationship. Structural inversions are often performed as purely geophysical inversions, where subsurface structures (that mimic geological or hydrogeological features) are favoured during inversion by choosing appropriate regularization terms. An example is the layered and laterally constrained inversion developed by Auken & Christiansen (2004), which respects vertically sharp and laterally smooth boundaries found in sedimentary geology. Joint geophysical inversions have been used extensively to delineate subsurface hydrogeological structures under the assumption that multiple geophysical data sets carry information about the same structural features of the subsurface (Christiansen et al., 2007; Gallardo, 2003; Haber and Oldenburg, 1997) but examples of successful joint hydrogeophysical inversion at larger scales are rare.

~~As a response to lack of global petro-physical relationships, clustering algorithms as an extension to structural inversion methods have been applied in geophysics (Bedrosian et al., 2007). Fuzzy c-means and k-means clustering algorithms have been used with sequential inversion schemes (Paasche et al., 2006; Triantafilis and Buchanan, 2009) and joint inversion schemes (Di Giuseppe et al., 2014; Paasche and Tronicke, 2007). These studies have focused on the structural information contained in geophysical information, and hydrogeological or geological parameters of the subsurface are assumed uniform within the delineated zones. This approach corresponds well with the common practice in groundwater modelling where degrees of freedom of the subsurface are reduced by zoning the subsurface.~~

~~We present an objective and semi-automatic method to model large-scale hydrostratigraphy from geophysical resistivity and lithological data. The method is a novel sequential hydrogeophysical inversion for integration of AEM data into the hydrological modelling process. First, resistivity data is translated into clay fraction values by inverting for the parameters of a spatially variable petrophysical relationship (Foged et al., 2014). Second, a cluster analysis is performed on the principal components of resistivity data and clay fraction values. The zones identified in the cluster analysis are assumed to have uniform hydrogeological properties, and thus form the hydrostratigraphic model. Third, the hydraulic conductivity (K) of each zone in the hydrostratigraphic cluster model is estimated in a hydrological model calibration. The hydrological performance is benchmarked against a geological reference model. Results are shown for a Danish case study.~~

As a response to lack of global petrophysical relationships, clustering algorithms as an extension to structural inversion methods have been applied in geophysics (Bedrosian et al., 2007; Doetsch et al., 2010). Fuzzy c-means and k-means clustering algorithms have been used with sequential inversion schemes (Paasche et al., 2006; Triantafilis and Buchanan, 2009) and joint inversion schemes (Di Giuseppe et al., 2014; Paasche and Tronicke, 2007). These studies have focused on the structural information contained in geophysical information, and hydrogeological or geological parameters of the subsurface are assumed uniform within the delineated zones. This approach corresponds well with the common practice in groundwater modelling where degrees of freedom of the subsurface are reduced by zoning the subsurface.

We present an objective and semi-automatic method to model large-scale hydrostratigraphy from geophysical resistivity and lithological data. The method is a novel sequential hydrogeophysical inversion for integration of AEM data into the hydrological modelling process. ~~Materials and Methods~~

~~We propose a data-driven 3D zonation method to build groundwater model hydrostratigraphy.~~ Hydrostratigraphic structures and parameters are determined sequentially by geophysical/lithological and hydrological data respectively.

As shown in Figure 1 the 3D subsurface zonation is completed in two steps; 1) ~~delineation of~~a Hydrostratigraphic Cluster Modelling part, and 2) a Hydrological Modelling part. In part 1 the hydrostratigraphic structures are delineated (see Figure 2c) through k-means cluster analysis on resistivity data (see Figure 2a) and clay fraction values (see Figure 2b~~), and 2) estimation of hydraulic parameters of the hydrostratigraphic structures in a hydrological calibration using observations of hydraulic head and stream discharge~~). To obtain clay fraction values, resistivity data is translated into clay fraction values by inverting for the parameters of a spatially variable translator function (this is the petrophysical relationship) (Foged et al., 2014). The cluster analysis is performed on the principal components of normalized resistivity data and clay fraction values. In part 2 the hydraulic conductivity (K) of each zone in the hydrostratigraphic cluster model is estimated in a hydrological model calibration using observations of hydraulic head and stream discharge. The zones identified in the cluster analysis are assumed to have uniform hydrogeological properties, and thus form the hydrostratigraphic model.

The method is applied to a Danish case study, for which details and results are presented in the following sections.

## ~~3~~2 Materials and Methods

~~11 hydrostratigraphic cluster models consisting of 1-11 zones are set up and calibrated.~~

### ~~3.1~~2.1 Study area

Norsminde study area is located on the eastern coast of Jutland, Denmark, and covers a land surface area of 154 km$^2$. Figure 3 shows a map of the area delineating study area boundary, streams, and hydrological data. An overview of the geophysical and lithological data ~~is shown in Figure 4.~~can be found in Foged et al. (2014). Within 5-7 km from the sea, the land is flat and rises only to 5-10 meters above sea level. Further to the west the land ascends into an up-folded end-moraine at elevations between 50-100 meters above sea

Field Code Changed

level. The town of Odder with approximately 20,000 inhabitants is located at the edge of the flat terrain in the middle of the model domain.

Palaeogene, Neogene and Quaternary deposits characterize the area. The Palaeogene deposits are thick clays, and define the lower geological boundary. Neogene marine clays interbedded with alluvial sands overlay the Palaeogene deposits in the elevated northern and western parts of the model domain. Quaternary deposits are glacial meltwater sediments and tills found throughout the domain. ~~A large~~The WE striking Boulstrup tunnel valley (2 km by 14 km) incises the Palaeogene clay in the south (Jørgensen and Sandersen, 2006). The unconsolidated fill materials are meltwater sand and gravel, clay tills, and waterlaid silt/clay.

Groundwater is abstracted for drinking water supply, mainly from tunnel valley deposits and the elevated south-western part of the domain. The groundwater resource is abstracted from 66 abstraction wells, with a total production of 18,000-26,000 $m^3$/year, excluding smaller private wells. Maximum annual abstraction from one well is 12,400 $m^3$/year. Actual pumping variation among the 66 wells and inter-annual variation of pumping rates are unknown. Abstraction is planned locally by water works and only information about permissible annual rates has been obtained for this study.

Groundwater hydraulic heads are available from 132 wells at various depths, see ~~Figure 3~~Figure 3 for the spatial distribution. Hydraulic head data are collected from the Danish national geological and hydrological database Jupiter (GEUS, n.d.).

Average annual precipitation is 840 mm/year for the years 1990-2011. Most of the area is tile-drained. The catchment is drained by a network of 24 streams; the main stream is gauged at the three stations 270035, 270002 and 270003, see ~~Figure 3.~~Figure 3. Streams vary from ditch-like channels to meters wide streams. Low and high flows respectively are in the order of 0.05-0.5 $m^3$/s and 0.5-5 $m^3$/s. Daily stream discharge data is available from three gauging stations. Discharges are calculated from mean daily water table measurements and translated with QH curves, which are available from approximately monthly discharge measurements.

**Formatted:** Font: 11 pt, English (U.K.)
**Formatted:** Normal (Web), Justified
**Formatted:** Font: 11 pt, English (U.K.)
**Formatted:** Font: 11 pt, English (U.K.)
**Formatted:** Font: 11 pt, English (U.S.)
**Formatted:** Font: 11 pt, English (U.S.)
**Formatted:** Font: 11 pt, English (U.S.), Superscript
**Formatted:** Font: 11 pt, English (U.S.)
**Formatted:** Font: 11 pt, English (U.S.), Superscript
**Formatted:** Font: 11 pt, English (U.S.)
**Formatted:** English (U.S.)

## ~~3.2 Geophysical data~~

~~Time-domain electro-magnetic (AEM) data collected through ground and airborne surveys is available for most of the study area; brown dashed areas in Figure 4a show the extent of the ground-based surveys and dots in Figure 4a show locations of AEM soundings. AEM data was collected using the SkyTEM[101] system in 2011 (Schamper et al., 2014a). SkyTEM[101] is developed for near-surface exploration by measuring also very early time gates, which requires careful system calibration and data processing (Auken et al., 2009; Schamper et al., 2014a). Depth of investigation (DOI) (Auken et al., 2014; Christiansen and Auken, 2012) varied between 50m and 150m. The survey was completed with a flight line spacing of 100m/50m and sounding spacing of 15m (total of 1856 line km, equivalent to 106,770 1D resistivity models). Data was inverted using spatially constrained inversion (SCI), which mimics 3D distribution of subsurface resistivity by passing information through lateral and vertical constraints along and between flight lines (Viezzoli et al., 2008). In the SCI, single soundings were modelled as smooth 1D resistivity models consisting of 29 layers with fixed layer boundaries. DOI was used to terminate the resistivity models with depth. For the airborne survey an acceptable match of almost 90% was found when checking resistivity results against borehole data (Schamper et al., 2014b). Ground based TEM soundings were collected during the 1990's with the Geonics TEM47/PROTEM system, and were inverted individually as 1D layered resistivity models. Depending on optimal fit these models have 3-5 layers.~~

## ~~3.3 Hydrostratigraphic model~~

~~Geophysical and lithological data are used to zone the subsurface.~~Time-domain electro-magnetic (EM) data collected through ground and airborne surveys is available for most of the study area. The AEM survey covers 2000 line kilometers, equivalent to 106,770 1D models and was carried out with the SkyTEM[101] system (Schamper et al., 2014). Lithological information is available at approximately 700 boreholes. The borehole descriptions~~Geophysical data consists of resistivity values determined from inversion of airborne and ground-based electromagnetic data. Lithological information is represented in clay fraction values determined through inversion within the clay fraction concept (CF concept). Zonation is performed in 3D.~~

~~The CF concept is formulated as a least squares inversion problem to determine the parameters of a petro-physical relationship that translates geophysical resistivities into clay fraction values. The concept is described in detail in Foged et al. (2014) and (Christiansen et al., 2014), and only a brief introduction is given here. The inversion minimizes the difference between observed clay fraction as determined from borehole lithological logs (in the inversion this is the data) and translated clay fraction as determined from geophysical resistivity values (in the inversion this is the forward data). Clay fraction expresses relative accumulated thickness of clay material over an interval. In this context clay refers to material described as clay in lithological logs, and not clay minerals. Clay definitions include, among others, clay till, marl clay, mica clay, and silty clay. Lithological borehole information is available at approximately 700 locations, see Figure 4b. Descriptions are from the Danish Jupiter database (GEUS, n.d.) and level of detail and quality varies from detailed lithological description at 1m intervals to more simple sand, clay, till descriptions at layer interface depths. Boreholes are categorized according to quality of lithological information where 1 is the highest quality and 5 is the lowest, see Figure 4b. The quality rating is presented in (Schamper et al., 2014b).~~

~~In the CF inversion, the petrophysical relationship (in the inversion this is the forward model) is a two-parameter function defined in regular horizontal 2D grids which are vertically constrained to form a 3D model. The translator model parameters are interpolated from grid nodes to the locations of the geophysical resistivity models. Using the location specific model, geophysical resistivities are translated into clay fraction values. Using kriging the simulated clay fractions are interpolated to the location of the lithological borehole logs. The misfit between the observed and simulated clay fractions are then calculated at the borehole locations. The objective function, which contains data misfit and vertical and horizontal smoothness constraints, is optimized iteratively.~~

are from the Danish Jupiter database (GEUS, n.d.) and the level of detail and quality varies from detailed lithological description at 1m intervals to more simple sand, clay, till descriptions at layer interfaces. A thorough description of EM data collection and processing and lithological borehole information can be found in (Foged et al., 2014).

## 2.2 Hydrostratigraphic model

Geophysical and lithological data is used to zone the subsurface. Geophysical data consists of resistivity values determined from inversion of airborne and ground-based electromagnetic data. Lithological information is represented in clay fraction values determined through inversion within the clay fraction concept (CF-concept). Zonation is performed in 3D.

The CF-concept is formulated as a least squares inversion problem to determine the parameters of a petro-physical relationship (in the inversion this is the forward model) that translates geophysical resistivities into clay fraction values. The concept is described in detail in Foged et al. (2014) and Christiansen et al. (2014), and only a brief introduction is given here. The inversion minimizes the difference between observed clay fraction as determined from borehole lithological logs (in the inversion this is the data) and translated clay fraction as determined from geophysical resistivity values (in the inversion this is the forward data). Clay fraction expresses relative accumulated thickness of clay material over an interval. In this context clay refers to material described as clay in lithological logs, and not clay minerals. Clay definitions include, among others, clay till, marl clay, mica clay, and silty clay. In the CF-inversion, the translator function is a heuristic two-parameter function defined on a regular 3D grid which is constrained vertically and horizontally. Discretization is 1000 m in the horizontal and 4 m in the vertical. The translator function is a scaled inverse error function, see Eq. 1 and Figure 4.

$$W(\rho) = 0.5 \cdot erfc\left(\frac{K \cdot (2\rho - m_{up} - m_{low})}{m_{up} - m_{low}}\right), \quad K = erfc^{-1}(0.05) \qquad \text{(Eq. 1)}$$

$m_{low}$ and $m_{up}$ are the model parameters of the translator function, $W(\rho)$, that translates resistivity, $\rho$, into clay fraction. $K$ scales the error function so that $W(\rho)$ equal 0.025 and 0.975 for resistivity values equal to $m_{low}$ and $m_{up}$ respectively, see Figure 4. The parameters of the translator function vary throughout the 3D grid. The objective function, with a data misfit term and vertical and horizontal regularization term, is minimized iteratively. The regularization constraint is a measure of weighted squared difference between

Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed

$m_{low}$ and $m_{up}$ at neighbouring grid nodes, where the weighting is the regularization constraint. The final parameters of the translator function translate geophysical resistivity values into CF-values. An experimental semi-variogram is estimated from the simulated CF-values, and 2D block kriging is used to obtain a 3D CF-model. The resolution difference between lithological borehole data and AEM data is discussed in Foged et al. (2014).

Field Code Changed

Field Code Changed

Delineation of subsurface structures is performed as a k-means cluster analysis on geophysical resistivities and clay fraction values. Information contained in clay fraction values is to some extent duplicated in the geophysical resistivity values. Heterogeneity captured in the resistivity data however is simplified in the translation to clay fraction; for example till and Palaeogene clay have respectively medium and low resistivity values while the clay fraction for both materials is 1.

K-means clustering is a well-known cluster analysis which finds groups in multivariate data based on a measure of similarity between cluster members (Wu, 2012). Similarity is defined as minimum squared Euclidean distance between each cluster member and cluster centroid, summed over all cluster members. The number of clusters that the data is divided into is defined by the user. We use the k-means analysis implementation in MATLAB R2013a, which ~~use~~uses a two-phase search, batch and sequential, to minimize the risk of reaching a local minimum.

Because clay fraction values are correlated with geophysical resistivities k-means clustering is performed on principal components (PC) of the original variables. Principal components analysis (PCA) is an orthogonal transformation based on data variances (Hotelling, 1933). PCA thus finds uncorrelated linear combinations of original data while obtaining maximum variance of the linear combinations (Härdle and Simar, 2012). The uncorrelated PCs are a useful representation of the original variables as input to a k-means cluster analysis. Original variables must be weighted and scaled prior to PCA, as PCA is scale sensitive, and the lack of explicit physical meaning of the PCs makes weighting difficult. Clay fraction values are unchanged as they

range between 0 and 1. The normalized resistivity values are calculated as $\rho_{norm} = \frac{\log \rho - \log \rho_{\min}}{\log \rho_{\max} - \log \rho_{\min}}$.

Where $\rho_{\min}$ ~~is~~ and $\rho_{\max}$ is minimum and maximum resistivity values respectively.

~~Integrated hydrological~~11 hydrostratigraphic cluster models consisting of 1-11 zones are set up and calibrated.

## ~~3.4~~2.3 Hydrological model

Hydrological data are used to parameterise the structures of the hydrostratigraphic model. Stream discharges and groundwater hydraulic heads are used as observation data in the hydrological calibration.

The hydrological model is set up using MIKE SHE (Abbott et al., 1986; Graham and Butts, 2005), which is a physically based ~~integrated~~ hydrological model code simulating evapotranspiration, the unsaturated zone, overland flow and saturated flow, while stream discharge is simulated by coupling with the MIKE 11 routing model code.

### ~~3.4.1~~2.3.1 Hydrological model parameterization

The model has a horizontal discretization of 100 m x 100 m, and a vertical discretization of 5 m following topography. The uppermost layer is 10 m thick for numerical stability~~.~~, which is not expected to negatively impact river discharge as this is largely controlled by drainage. Because the model represents a catchment, all land boundaries are defined as no-flow boundary conditions following topographical highs. Constant head boundary conditions are defined for sea boundaries, and the model domain extends 500 meters into the sea. Model grid cells 10 meters below the Palaeogene clay surface have been de-activated, due to the computational burden.

The unsaturated zone and evapotranspiration (ET) ~~is~~are modelled using the 2-Layer water balance method developed to represent recharge and ET to/from the groundwater in shallow aquifer systems (Yan and Smith, 1994). The reference evapotranspiration is calculated using Makkink's formula (Makkink, 1957). Soil water characteristics of the five soil types and the associated 250 m grid product are developed and described by

Borgesen & Schaap (2005) and Greve et al. (2007), respectively. Land use data is obtained from the DK-model2009, for which root depth dependent vegetation types were developed (Højberg et al., 2010).

Stream discharge is routed using the kinematic wave equation. The stream network is modified from the DK-model2009 (Højberg et al., 2010) by adding additional calculation points and cross sections. Groundwater interaction with streams is simulated using a conductance parameter between aquifer and stream. ~~Overland flow is simulated using the Saint-Venant equations.~~Overland flow is simulated using the Saint-Venant equations (DHI, 2012, pp. 267-281). Manning number and overland storage depth is 5 $m^{1/3}$ $s^{-1}$ and 10 mm respectively. Drainage parameters, drain time constant [$s^{-1}$] and drain depth [m] are uniform in space and time~~, as parameterization~~. Parameterization of spatial variable ~~drainage parameters~~drain time constant relies on direct drainage flow measurements, and ~~(Hansen et al., 2013).~~Hansen et al. (2013) found little variability in the estimated time constants and no justification for a spatial variability judging from eight hydrological performance criteria. Drain depth is 1 m below terrain.

Saturated flow is modelled as anisotropic Darcy flow, xy-z anisotropy being restricted to the orientation of the computational model grid (DHI, 2012). A vertical anisotropy of 1/10 is assumed. The saturated zone is parameterized with the cluster models. The lower boundary of the saturated zone is defined by the surface of the Palaeogene clay, available in ~~100m~~100 m grid, and has a fixed horizontal K of $10^{-10}$ m/s. Specific yield and specific storage are fixed at 0.15 and 5~~·~~·$10^{-5}$ $m^{-1}$ for the entire domain.

### ~~3.4.2~~2.3.2 Hydrological model calibration

Forward models are run from 1990 to 2003; the years 1990-1994 serve as warm up period (this was found sufficient to obtain stable conditions); the calibration period is from 2000 to 2003 and the validation period is from 1995 to 1999.

Composite scaled sensitivities (Hill, 2007) were calculated based on local sensitivity analyses. Figure 5 shows calculated sensitivity for selected model parameters. Sensitivities of the parameters, which are shared by the 11 cluster models, are calculated for each cluster model. The top panel in Figure 5 shows sensitivities of the shared parameters. The bars indicate the mean value of these sensitivities, and the error bars mark the

minimum and maximum value of these sensitivities. The lower panel in Figure 5 shows subsurface parameters for the 5-cluster model.

The following parameters are made a part of the model calibration;

- The root depth scaling factor, which was found sensitive, see Figure 5 top panel. Because root depth values vary inter-annually and between crop types, root depth sensitivity was determined by a root depth scaling factor, which scales all root depth values.
- The drain time constant. Especially considering discharge observations, the model shows sensitivity towards this parameter. Stream hydrograph peaks are controlled by the drainage time constant (Stisen et al., 2011; Vazquez et al., 2008).
- The river leakage coefficient.
- The horizontal hydraulic conductivities of all zones of the 11 hydrostratigraphic cluster models. Figure 5 shows sensitivity to K of the zones of the 5-cluster model. K of the zones ~~are~~is unknown; hence all K values have been calibrated. Vertical K values are tied to horizontal K with an anisotropy factor of 10. Initial horizontal K values are $10^{-4}$ m/s, $10^{-6}$ m/s or $10^{-8}$ m/s depending on mean clay fraction value of a zone.

Storage parameters were set to a priori values and not calibrated.

Calibration is performed using the Marquardt-Levenberg local search optimization implemented in PEST (Doherty, 2005). Observations are 632 hydraulic heads from 132 well filters and daily stream discharge time series from three gauging stations, see Figure 3. Observation variances are estimated, and, in the absence of information, observation errors were assumed to be uncorrelated. Objective functions for head and discharge have been scaled to balance contributions to the total objective function.

The aggregated objective function, $\Phi$, shown in Eq. ~~1~~2 is the sum of the scaled objective function for head and discharge. The subjective weight~~–~~, $w_s$, was determined through trial and error by starting numerous calibration runs; $w_s$ was chosen to be 0.8.

$$\Phi = w_s \sum_{i=1}^{N_h} \left( \frac{h_{sim,i} - h_{obs,i}}{\sigma_i} \right)^2 + (1 - w_s) \sum_{i=1}^{N_q} \left( \frac{q_{sim,i} - q_{obs,i}}{\sigma_i} \right)^2 \quad \text{(Eq. 12)}$$

Hydraulic head observation errors are determined following the guidelines following Henriksen et al. (2003). They suggest an error budget approach which accounts for contributions from 1) the measurement (e.g. with dip meter); 2) inaccuracy in vertical referencing of wells; 3) interpolation between computational nodes to observation well location; and 4) heterogeneity that is not represented in the lumped computational grid. The total error expresses the expected uncertainty between observation and corresponding simulation. The approach for estimating these uncertainties can be found in Appendix A. Total errors amount to 0.95 m, 1.4 m and 2.2 m.

Uncertainty of stream discharges is mainly due to translation from water stages to discharge (daily mean discharges). Uncertainties originate from infrequent calibration of rating curve, ice forming on streams and especially stream bank vegetation (Raaschou, 1991). Errors can be as large as 50%. Blicher (1991) estimates errors of 5% and 10% on the water stage measurement and rating curve respectively. In cases of very low stream flows (1 L/s) Christensen et al. (1998) assigned a standard deviation of 200% while flow of 50 L/s and 5-10 L/s are assigned standard deviations of 5% and 25% respectively. We have assigned an error of 20% to all stream discharge observations.

## ~~3.5 Benchmarking hydrostratigraphic cluster models~~

~~The performance of the hydrological model based on the cluster model hydrostratigraphy has been benchmarked against the hydrological performance using a reference geological model (He et al. 2015, in review). The reference geological model is based on geological interpretation and AEM data. The study area was subdivided into seven major geological elements, based on geological interpretation. The seven elements are the Palaeogene clay, the Miocene, Boulstrup tunnel valley, three glacial sequences, and a glaciotectonic complex. By collapsing the three glacial sequences into one glacial element, four hydrogeological elements were defined; the Miocene, Boultrup tunnel valley, the glacial, and the glaciotectonic complex. (He et al.,~~

~~2014) performed a geostatistical analysis with TProGS (Carle and Fogg, 1996) on the lithological information and AEM data in the Miocene element to determine a sand-clay cut-off value for geophysical resistivities. This cut-off value was used to subdivide each of the four elements into units of sand and clay. Surface geology is characterized by a clay, sand and peat unit. The horizontal distribution of clay, sand and peat is taken from the Danish National Water Resources Model (Højberg et al., 2010). Because of problems with drying filters in groundwater abstraction wells, a material with horizontal and vertical hydraulic conductivity of $3 \cdot 10^{-5}$ m/s / $3 \cdot 10^{-6}$ m/s was added around selected filter screens.~~

~~Both the reference geological model and the cluster models are used to construct hydrological models that are calibrated in the hydrological calibration framework described previously. The only difference between the hydrological models and the calibrations are the parameterization (structures and values) of the saturated zone. As for the cluster models the Palaeogene clay surface elevation defines the lower boundary of the hydrological model. A vertical anisotropy of 1/10 is assumed for all geological units. Calibration parameters of the reference model are shown in Table 1. Specific yield values are fixed 0.05 and 0.2 for clay and sand deposits respectively, and specific storage is fixed at $5 \cdot 10^{-5}$.~~

# ~~4~~3 Results and discussion

First we show results for the hydrological performance of 11 hydrostratigraphic cluster models consisting of 1-11 zones. Secondly details of the cluster analysis for the case of a 5-cluster hydrostratigraphy are shown. ~~Finally the results of benchmarking with the reference model are presented through comparison of observed and simulated state variables~~Finally the cluster model hydrological performance is benchmarked with comparable hydrological models.

## ~~4.1~~3.1 Calibration and validation of hydrological model

Figure 6 shows the weighted RMSE of ~~models consisting of~~ model performances for hydrostratigraphic cluster model consisting of 1~~-~~ to 11 zones~~. Values are shown for~~ , head and discharge respectively is shown

in ~~separate figures~~Figure 6a and Figure 6b. The 1-cluster model is a homogeneous representation of the subsurface resulting in a uniform K field. The 1-cluster model represents a situation where we have no information about the subsurface. Increasing the number of clusters to represent the subsurface successively adds more information from geophysical and lithological data to the calibration problem. ~~Horizontal dashed lines indicate weighted RMSE of the reference model.~~ The weights used to calculate weighted RMSE are the same weights as used in Eq. ~~1~~2.

Head and discharge contribute by approximately 2/3 and 1/3 of the total objective function. From the 1-cluster to the 2-cluster model, weighted RMSE for discharge is reduced by more than a factor 2. No significant improvement of the fit to discharge data is observed for more than 2 clusters. Fit to head data improves almost by a factor of 2 from the 1-cluster to the 2-cluster model. Improvement of the fit to head data continues up to the 5-cluster representation of the subsurface. Improvements are a factor of 3 from the 1-cluster to the 5-cluster model. Beyond the 5-cluster model, the fit to head observations stagnates. The 7-cluster and 9-cluster hydrostratigraphic models perform worse than the 3-cluster model. The 8-, 10-, and 11-cluster models obtain an equally good or better fits to head data compared to the 5-cluster model.

The blue lines in Figure 6 illustrate mean standard deviation on log(K) values of the cluster models ~~and reference model respectively,~~ based on the post-calibration standard deviation of log(K) for each K zone. ~~The mean standard deviation of the reference model is 0.046, and corresponds to that of the 2- and 3-cluster models.~~ Beyond the 4- and 5-cluster models the precision of the estimated K values decrease. The mean standard deviations on log(K) for the 4- and 5-cluster models are 0.12 and 0.15. The corresponding widths of the 95% confidence intervals are between 15% and 90% of the estimated K value for 3 out of 4 zones and 3 out of 5 zones, respectively. Beyond the 5-cluster model mean standard deviations on log(K) are between 0.17 and 0.27, and corresponding width of the 95% confidence intervals are largely above 100% for all but two zones.

With the combined information from weighted RMSE values and standard deviation on log(K) we are able to address over-parameterisation. The results indicate that we obtain good fit to observations without over-parameterisation with a ~~4~~3- to 5-cluster hydrostratigraphic model.

In this paper, we have discussed the performance of the cluster models as a measure of fit to hydraulic head and stream discharge observations. ~~Integrated hydrological~~Hydrological models are typically used to predict transport, groundwater age, and capture zones, which are sensitive to geological features. It is likely that the optimal number of clusters is different for these applications. An analysis, as is presented here for head and discharge, for predictive application is more difficult because observations are often unavailable.

The hydrostratigraphic models are constructed under the assumption that subsurface structures governing groundwater flow can be captured by structural information contained in clay fraction values (derived from lithological borehole data) and geophysical resistivity values. If this is true, an asymptotic improvement of the data fit would be expected for increasing cluster numbers. However, as shown in Figure 6, this is not strictly the case: Weighted RMSE of the 7-cluster and 9-cluster models is higher than weighted RMSE of the 3-~~cluster6~~cluster, 6-cluster and 8-cluster models, respectively. The likely explanation is that increasing number of clusters does not correspond to pure cluster sub-division, but also to relocation of cluster interfaces in the 3D model space. We expect the difference in hydrological performance to be due to changes in interface configuration.

It is well-known that an unsupervised k-means clustering algorithm does not result in a unique solution, due to choice of initial (and unknown) cluster centroids. We have sampled the solution spaces (200 samples) of the eleven cluster models. Clustering the principal components of geophysical resistivity data and clay fraction values into 1 to 5 clusters gives unique solutions. Clustering the principal components of geophysical resistivity data and clay fraction values into 6 to 11 clusters results in three or more solutions. The non-unique solutions however have different objective functions (squared Euclidean distance between points and centroids). In all cases the cluster model with the lowest objective function was chosen as the best solution.

Figure 7 shows RMSE and mean errors for calibration and validation periods for all 11 cluster models ~~and the reference model. Horizontal dotted lines are reference model performances.~~. Data used to calculate the statistics are a temporally split sample from 35 wells, which have observations both in the calibration and validation period, and the discharge is for stations 270002 and 270003.

The cluster models ~~as well as the reference model~~ perform similarly in 2000-2003 and 1995-1999. With respect to RMSE, Figure 7a, for head the validation period is approximately 10% worse than the calibration period. RMSE for discharge, Figure 7b, is lower in the validation, approximately a third of the calibration values. Mean errors for head, Figure 7c, are lower and higher ~~for the reference model and the cluster models~~ respectively. The hydrological models analysed in this study generally under-simulate the average discharge.

~~From Figure 7a and Figure 7c it appears that the cluster models for 3-4-5 clusters perform better than the reference model with respect to RMSE head, while they have equal performance for ME head. Recalling that the reference model and the 5-cluster model have respectively 6 and 5 degrees of freedom in the hydrological model calibration, this indicates the difference in spatial patterns of the two models.~~

### ~~4.2~~3.2 The cluster model

Figure 8 presents histograms of clay fraction values and resistivity values and how the values are represented in the five clusters, which was chosen to be the optimal number. Counts are shown as percentages of total number of pixels in the domain. The histograms in Figure 8 show that the clay fraction attribute separates high resistivity/low clay fraction (sandy sediments) from other high-resistivity portions of the domain, while the resistivity attribute separates low resistivity/high clay fraction (clayey sediments) from other high clay-fraction portions. High resistivity/low clay fraction values are represented by clusters 1, 3 and 4 and low resistivity/high clay fraction are represented by clusters 2 and 5, see Figure 8a.

Figure 9 shows the data cloud that forms the basis of the clustering. The data cloud is binned into 300 bins in each dimension and the colour of the cloud shows the bin-wise data density. We see that cluster boundaries appear as straight lines in the attribute space. Values with a low resistivity and corresponding

high clay fraction, mainly clusters 2 and 5, populate more than half of the domain. Clay is expected to dominate this part of the domain.

The results of the cluster analysis are presented with respect to geophysical resistivity and clay fraction values, while the cluster analysis is performed on the principal components (PC) of geophysical resistivity and clay fraction values. The first PC explains the information where the two original variables, log resistivity and clay fraction, are inversely correlated. This corresponds to the situation where a clay fraction of 1 coincides with a low resistivity value, and vice versa for clay fraction values of 0 and high resistivities. This is the information that we expect, i.e. our understanding of how geophysical resistivities relate to lithological information as represented by ~~our~~the translator function (Eq. 1) (defined under the assumption that variation in geophysical resistivities with respect to lithological information depends on the presence of clay materials). Thus the first principal component is the 'clay' information in the geophysical resistivities. The second PC is less straight forward to interpret. Ideally, the second PC represents the data pairs where the resistivity response is *not* dominated or explained by lithological clay material. This might reflect a situation where a low resistivity value - and its associated low clay fraction value - is a result of a sandy material with a high pore-water electrical conductivity due to elevated dissolved ion concentrations. The second PC can also be a result of the CF-conceptualisation. Clay till, categorized as 'clay' in the CF-inversion, can have electrical resistivities up to 60 Ωm (Jorgensen et al., 2005; Sandersen et al., 2009), which will yield a high clay fraction coinciding with a relatively high geophysical resistivity.

Electromagnetic methods are sensitive to the electrical resistivity of the formation, which is commonly dominated by clay mineral content, dissolved ions in the pore water and saturation. Groundwater quality data is available at numerous sites in the domain. Pore-water electrical conductivity (EC) values were gathered from the coast and inland following ~~the~~Boulstrup tunnel valley. From the coast and 12 kilometres inland values are stable around 50-70 mS/m at 28 wells with varying filter depths. Four outliers with EC ranging between 120 and 250 mS/m were identified at various locations and depths. No trend due to salinity from the coast was identified. In theory, variations in formation electrical resistivity that are *not* due to lithological changes will implicitly be taken into account by spatial variation of ~~translator function parameters~~the

translator function in the CF-inversion. If there is a region in the modelled domain where the electromagnetic signal, and the resulting resistivity value, is affected by pore water salinity (low resistivity value is due to salinity and not clay content) and there is available borehole information, the parameters of the translator function will adjust to obtain lower values in order to translate a low resistivity value to a low clay fraction value.

### ~~4.3~~3.3 Benchmarking hydrological performance

Table 1~~ gives an overview of how the performance of the two models compare. The fit to discharge data of the two models is comparable. The weighted~~ shows RMSE and ME for head and discharge based on the 5-cluster model. Weighted RMSE for discharge is below 1, indicating that discharge is over-fitted. The standard deviation of discharge is 20% of the observation, which is a conservative definition. As presented in the methods section errors may vary between 5%-50%. The 1995-1999 hydrograph and scatter plot in Figure 10 for the 270002 gauging station show good fit to data. Peak and low flows are fitted, but baseflow recession is generally not matched very well. At gauging station ~~27003~~270003 the ~~models fail~~model fails to capture dynamics and relative magnitudes of the observations. Peak as well as low flows are under-simulated, which is clearly demonstrated in the scatter plot for station 270003 in Figure 10.

~~The hydraulic~~ With respect to head~~ performance statistics~~, the model under-simulates in ~~Table 2 show an improved performance~~the elevated parts of the ~~5-cluster model over the reference model: the calibration period RMSE/ME for~~domain (head above 50 m), see Figure 11. The head values below 20 m represent the ~~reference and 5-cluster model respectively~~Boulstrup tunnel valley, where head is ~~3.01 m/ 1.01 m and 1.99 m/ 0.790 m. Weighted RMSE for~~fitted the ~~reference model is 2.63 and 2.93, while~~best. With weighted RMSE for ~~the 5-cluster model is~~head of 1.63 and 1.85~~. The reference~~ the model ~~thus~~ is almost 2~~-3~~ standard deviations from fitting ~~data, while the 5-cluster model is 1-2 standard deviations from fitting~~head data. Assuming head observation error estimates are correct, this indicates model deficiencies such as structural errors and/or forcing data errors.~~ The outliers in the simulated head from the reference model, Figure 11 red diamonds, are from three wells in the elevated south-western part of the domain. Recalculating performance~~

~~statistics without the outliers gives RMSE and ME of 2.45 m and -0.647 m. Figure 11 shows that the largest differences in simulated heads between the two models are for hydraulic head below 20 m. These observations represent the tunnel valley aquifers, see also Figure 12a-b. Results indicate that the 5-cluster model performs better in the tunnel valley than the reference model.~~

Figure 12a-b show ~~simulated~~distributed head ~~(from the 5-cluster model) and the difference in simulated head between the 5-cluster model and the reference model~~results. Generally hydraulic head in the tunnel valley is disconnected from the elevated terrain (Figure 12a), and groundwater overall flows towards the sea. ~~The reference model simulates higher heads (~~Figure 12b~~) in the tunnel valley compared to the 5-cluster model. Figure 12c-d show~~ shows errors (obs-sim) between observed and simulated heads for 1995-1999. ~~Errors in the elevated terrain towards west are similar for the two models, whereas the reference model has larger~~The largest errors ~~(over-simulates head) in the tunnel valley compared to the 5-cluster model~~are found in the south eastern part of the domain, where discharge station 270003, with the worst fit, is located, see Figure 10 top row.

We have compared the hydrological performance of the Norsminde model based on the 5-cluster hydrostratigraphic model with similar Danish hydrological models. We have chosen Danish models due to comparability with respect to data density and quality, and hydrostratigraphy. The model performances are compared based on RMSE and ME of simulated heads, see 2, as these statistics are reported in the studies. The horizontal discretization of the models is 100 m, 200 m, and 500 m, and the models cover between 202 $km^2$ and 3500 $km^2$. We can see that the 5-cluster model is comparable with the other models.

## ~~4.4~~3.4 Advantages and limitations

We have presented a method for automatic generation of hydrostratigraphic models from AEM and lithological data for groundwater model applications. Other automatic methods of integrating AEM data into geological models are geostatistical methods presented by e.g. Gunnink et al. (2012), using artificial neural networks, or He et al. (2014), using transition probabilities.

~~A limitation~~The risk of ~~performing automatic integration~~misinterpretation of AEM ~~is that~~data due to effects of saturation, water quality, depth and material dependent resolution, and vertical shielding~~ effects,~~ are ~~misinterpreted, and wrongly interpreted~~higher with an automatic approach compared to a cognitive approach, as ~~geological structural information. These~~these effects may be identified by a geologist during the modelling process. AEM data can be integrated into geological models using cognitive methods, for example as presented by Jørgensen et al. (2013), who provide an insightful discussion of the pros and cons of automatic versus cognitive geological modelling from AEM data.

Geological knowledge, which can be incorporated into cognitive geological models (Royse, 2010; Scharling et al., 2009; Sharpe et al., 2007), cannot be included in automatically generated models. Geological knowledge may identify continuity/discontinuity of geological layers, or discriminate materials based on stratigraphy or depositional environment. For regional scale groundwater flow, characterisation of sedimentation patterns and sequences may not be relevant, but at smaller scales this information is valuable for transport modelling.

Field Code Changed

The hydrostratigraphic cluster model presented in this paper~~, however,~~ does not represent a lithological model, but has the advantage of incorporating close to all the structural information contained in the large AEM data sets~~.~~ in a fast and well documented way. This is not possible in practice for cognitive methods due to spatial complexity and the large amount of AEM data~~ amounts. Also reproducibility and especially possibility of uncertainty quantification of the hydrostratigraphic cluster model are important features~~. For hydrological applications hydrostratigraphic model uncertainty, and the resulting hydrological prediction uncertainty, has great value. We believe that the cluster model approach presented in this paper can be extended to address structural uncertainty and its impact on hydrological predictions. Cognitive geological model uncertainty is difficult to quantify.

The CF-model is to some degree influenced by smoothing resulting from the AEM data inversion and CF-inversion, and the finial kriging of CF-values to a regular grid. Smoothing effects causing resistivity transition zones are inconsistent with our understanding of geological interfaces. In future studies different

geophysical inversion schemes will be compared to evaluate the effect of smoothing on the final cluster model. This work will partly evaluate how the smooth transition zones impact hydrological results. We expect the geological interfaces to lie in the transition zones, but the exact location is unknown. We will address this problem by generating several cluster models that identify zonal divides at different locations in the transition zones. Hereby hydrological uncertainty as a result of the transition zones may also be assessed.

## ~~5~~4 Conclusion

We have presented an automated workflow to parameterize and calibrate a large-scale ~~integrated~~ hydrological model based on AEM and borehole data. The result is a competitive hydrological model that performs ~~equally good or even better than the reference geological model.~~satisfactory compared to similar hydrological models. From geophysical resistivity data and clay fraction values we delineate hydrostratigraphic zones, whose hydrological properties are estimated in a hydrological model calibration. The method allows for semi-automatic generation of reproducible hydrostratigraphic models. Reproducibility is naturally inherent as the method is data-driven and thus, to a large extent, also objective.

The number of zones in the hydrostratigraphic model must be determined as part of the cluster analysis. We have proposed that hydrological data, through hydrological calibration and validation, guide this choice. Based on fit to head and discharge observation and calibration parameter standard deviations, results indicate that the ~~4~~3- and 5-cluster models give the optimal performance.

Distributed groundwater models are used globally to manage groundwater resources. Today large-scale AEM data sets are acquired for mapping groundwater resources on a routine basis around the globe. There is a lack of knowledge on how to incorporate the results of these surveys into groundwater models. We believe the proposed method has potential to solve this problem.

## 65 Appendix A: Observation errors

Hydraulic head observation errors have been estimated using an error budget;

$$\sigma_{total}^2 = \sigma_{meas}^2 + \sigma_{elev}^2 + \sigma_{\text{int}}^2 + \sigma_{hetereo}^2 + \sigma_{unknown}^2$$

Quantitative estimates of the different error sources are to a large extent based on data from the Danish Jupiter database.

Head measurements are typically carried out with dip-meter, and occasionally pressure transducers are used. Information about which measurement technique has been used for the individual observations is not clear from the Jupiter database. It is assumed that dip-meters have been used and $\sigma_{meas}$ has been determined to be 0.05 m for all observations.

Well elevations are referenced using different techniques. The elevation can be determined from a 1:25000 topographic map, by levelling or by differential GPS. The inaccuracies for using topographic maps and DGPS measurements are in the order of respectively 1-2 m and centimetres. The Jupiter database can have information about the referencing techniques, but this information is rarely supplied. An implicit information source is the number of decimal places the elevations have in the database. Elevation information is supplied with 0, 1 or 2 decimal places. For the wells where the reference technique is available (checked for cases with topographic map and DGPS only) the decimal places reflect accuracy of the referencing technique used. From this information decimal places of 0, 1 and 2 have been associated with $\sigma_{elev}$ of 2 m, 1 m and 0.1 m respectively.

Errors due to interpolation depend on horizontal discretization of the hydrological model and the hydraulic gradient. Sonnenborg & Henriksen (2005, chapter 12) suggest it be estimated as $\sigma_{\text{int}} = 0.5 \cdot \Delta x \cdot J$, where $\Delta x$ is horizontal discretization and $J$ is hydraulic gradient. The model domain has been divided into three groups for which the error from interpolation has been calculated. The three areas are geologically different: north is glacial tectonically deformed; the west has similar Miocene and Glacial melt water sediments; and

626 the Palaeogene tunnel valley. Hydraulic gradients of the Miocene/Glacial west and the Palaeogene tunnel
627 valley are between 0.001-0.002. The Miocene/Glacial area and the Palaeogene tunnel valley areas were thus
628 considered as one with $\sigma_{int}$ of 0.07 m. The glacial tectonic area has an estimated hydraulic gradient of 0.01
629 and thus associated with $\sigma_{int}$ of 0.6 m.

630 Within-cell (hydrological model grid) heterogeneity affecting hydraulic head was estimated using data from
631 eight wells that are located within the same hydrological model grid. Temporally coinciding head
632 observations from the period 2001 and 2002 were used. The error is evaluated as the standard deviation of a
633 linear plane fitted through the observed heads at the eight boreholes. This has been done for three dates,
634 which gives a mean $\sigma_{hetereo}$ of 0.53 m.

635 $\sigma_{unknown}$ was set to 0.5 m.

636 ## ~~7~~6 Acknowledgements

637 This paper was supported by HyGEM, Integrating geophysics, geology, and hydrology for improved
638 groundwater and environmental management, Project no. 11-116763. The funding for HyGEM is provided
639 by The Danish Council for Strategic Research. We are thankful for the support and data provided by the
640 NiCA research project (funded by The Danish Council for Strategic Research under contract no. DSF 09-
641 067260), including SkyTEM data and the integrated hydrological model for Norsminde study area.

642 ## ~~8~~7 References

643 Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E. and Rasmussen, J.: An introduction to
644 the European Hydrological System — Systeme Hydrologique Europeen, "SHE", 2: Structure of a
645 physically-based, distributed modelling system, J. Hydrol., 87(1-2), 61–77, doi:10.1016/0022-
646 1694(86)90115-0, 1986.

Auken, E. and Christiansen, A. V: Layered and laterally constrained 2D inversion of resistivity data, Geophysics, 69(3), 752–761, doi:10.1190/1.1759461, 2004.

Bedrosian, P. A., Maercklin, N., Weckmann, U., Bartov, Y., Ryberg, T. and Ritter, O.: Lithology-derived structure classification from the joint interpretation of magnetotelluric and seismic models, Geophys. J. Int., 170(2), 737–748, doi:10.1111/j.1365-246X.2007.03440.x, 2007.

Beven, K.: Changing ideas in hydrology — The case of physically-based models, J. Hydrol., 105(1-2), 157–172, doi:10.1016/0022-1694(89)90101-7, 1989.

Binley, A., Winship, P., Middleton, R., Pokar, M. and West, J.: High-resolution characterization of vadose zone dynamics using cross-borehole radar, Water Resour. Res., 37(11), 2639–2652, doi:10.1029/2000WR000089, 2001.

Blicher, A. S.: Usikkerhed på bearbejdning af data fra vandføringsstationer. Publication nr. 1 from Fagdatacenter for Hydrometriske Data, Hedeselskabet, Viborg., 1991.

Borgesen, C. and Schaap, M.: Point and parameter pedotransfer functions for water retention predictions for Danish soils, Geoderma, 127(1-2), 154–167, doi:10.1016/j/geoderma.2004.11.025, 2005.

Bosch, J. H. A., Bakker, M. A. J., Gunnink, J. L. and Paap, B. F.: Airborne electromagnetic measurements as basis for a 3D geological model of an Elsterian incision <BR>[Hubschrauberelektromagnetische Messungen als Grundlage für das geologische 3D-Modell einer glazialen Rinne aus der Elsterzeit], Zeitschrift der Dtsch. Gesellschaft für Geowissenschaften, 160(3), 249–258, doi:10.1127/1860-1804/2009/0160-0258, 2009.

667 Burschil, T., Scheer, W., Kirsch, R. and Wiederhold, H.: Compiling geophysical and geological
668 information into a 3-D model of the glacially-affected island of Föhr, Hydrol. Earth Syst. Sci.,
669 16(10), 3485–3498, doi:10.5194/hess-16-3485-2012, 2012.

670 Caers, J. and Hoffman, T.: The Probability Perturbation Method: A New Look at Bayesian Inverse
671 Modeling, Math. Geol., 38(1), 81–100, doi:10.1007/s11004-005-9005-9, 2006.

672 Christensen, S., Rasmussen, K. R. and Moller, K.: Prediction of Regional Ground Water Flow to
673 Streams, Ground Water, 36(2), 351–360, doi:10.1111/j.1745-6584.1998.tb01100.x, 1998.

674 Christiansen, A. V., Auken, E., Foged, N. and Sorensen, K. I.: Mutually and laterally constrained
675 inversion of CVES and TEM data: a case study, NEAR Surf. Geophys., 5(2), 115–123, 2007.

676 Christiansen, A. V., Foged, N. and Auken, E.: A concept for calculating accumulated clay thickness
677 from borehole lithological logs and resistivity models for nitrate vulnerability assessment, J. Appl.
678 Geophys., 108, 69–77, doi:10.1016/j.jappgeo.2014.06.010, 2014.

679 Dam, D. and Christensen, S.: Including Geophysical Data in Ground Water Model Inverse
680 Calibration, Ground Water, 41(2), 178–189, doi:10.1111/j.1745-6584.2003.tb02581.x, 2003.

681 DHI: MIKE SHE User Manual: Reference Guide, Hørsholm, Denmark., 2012.

682 Doetsch, J., Linde, N., Coscia, I., Greenhalgh, S. A. and Green, A. G.: Zonation for 3D aquifer
683 characterization based on joint inversions of multimethod crosshole geophysical data, Geophysics,
684 75(6), G53–G64, doi:10.1190/1.3496476, 2010.

685 Doherty, J.: PEST: Model-Independent Parameter Estimation. User Manual: 5th Edition, Brisbane,
686 QLD, Australia., 2005.

687 Ferré, T., Bentley, L., Binley, A., Linde, N., Kemna, A., Singha, K., Holliger, K., Huisman, J. A.
688 and Minsley, B.: Critical Steps for the Continuing Advancement of Hydrogeophysics, Eos, Trans.
689 Am. Geophys. Union, 90(23), 200, doi:10.1029/2009EO230004, 2009.

690 Foged, N., Marker, P. A., Christansen, A. V., Bauer-Gottwein, P., Jørgensen, F., Høyer, A.-S. and
691 Auken, E.: Large scale 3-D modeling by integration of resistivity models and borehole data through
692 inversion, Hydrol. Earth Syst. Sci. Discuss., 11(2), 1461–1492, doi:10.5194/hessd-11-1461-2014,
693 2014.

694 Gallardo, L. A.: Characterization of heterogeneous near-surface materials by joint 2D inversion of
695 dc resistivity and seismic data, Geophys. Res. Lett., 30(13), 1658, doi:10.1029/2003GL017370,
696 2003.

697 GEUS: Danish national geological and hydrological database, JUPITER, n.d.

698 Di Giuseppe, M. G., Troiano, A., Troise, C. and De Natale, G.: k-Means clustering as tool for
699 multivariate geophysical data analysis. An application to shallow fault zone imaging, J. Appl.
700 Geophys., 101, 108–115, doi:10.1016/j.jappgeo.2013.12.004, 2014.

701 Graham, D. N. and Butts, M. B.: Flexible integrated watershed modeling with MIKE SHE, in
702 Watershed Models, edited by V. P. Singh and D. K. Frever, pp. 245–272, CRC Press., 2005.

703 Greve, M. H., Greve, M. B., Bøcher, P. K., Balstrøm, T., Breuning-Madsen, H. and Krogh, L.:
704 Generating a Danish raster-based topsoil property map combining choropleth maps and point
705 information, Geogr. Tidsskr. J. Geogr., 107(2), 1–12, doi:10.1080/00167223.2007.10649565, 2007.

706 Gräbe, A., Rödiger, T., Rink, K., Fischer, T., Sun, F., Wang, W., Siebert, C. and Kolditz, O.:
707 Numerical analysis of the groundwater regime in the western Dead Sea escarpment, Israel + West
708 Bank, Environ. Earth Sci., 69(2), 571–585, doi:10.1007/s12665-012-1795-8, 2012.

709 Gunnink, J. L., Bosch, J. H. A., Siemon, B., Roth, B. and Auken, E.: Combining ground-based and
710 airborne EM through Artificial Neural Networks for modelling glacial till under saline groundwater
711 conditions, Hydrol. Earth Syst. Sci., 16(8), 3061–3074, doi:10.5194/hess-16-3061-2012, 2012.

712 Haber, E. and Oldenburg, D.: Joint inversion: a structural approach, Inverse Probl., 13(1), 63–77,
713 doi:10.1088/0266-5611/13/1/006, 1997.

714 Hansen, A. L., Refsgaard, J. C., Christensen, B. S. B. and Jensen, K. H.: Importance of including
715 small-scale tile drain discharge in the calibration of a coupled groundwater-surface water catchment
716 model, Water Resour. Res., 49(1), 585–603, doi:10.1029/2011wr011783, 2013.

717 He, X., Sonnenborg, T. O., Jørgensen, F., Høyer, A.-S., Møller, R. R. and Jensen, K. H.: Analyzing
718 the effects of geological and parameter uncertainty on prediction of groundwater head and travel
719 time, Hydrol. Earth Syst. Sci., 17(8), 3245–3260, doi:10.5194/hess-17-3245-2013, 2013.

720 He, X., Koch, J., Sonnenborg, T. O., Jørgensen, F., Schamper, C. and Christian Refsgaard, J.:
721 Transition probability-based stochastic geological modeling using airborne geophysical data and
722 borehole data, Water Resour. Res., 50(4), 3147–3169, doi:10.1002/2013WR014593, 2014.

723 He, X., Højberg, A. L., Jørgensen, F. and Refsgaard, J. C.: Assessing hydrological model predictive
724 uncertainty using stochastically generated geological models, Hydrol. Process., n/a–n/a,
725 doi:10.1002/hyp.10488, 2015.

726 Henriksen, H. J., Troldborg, L., Nyegaard, P., Sonnenborg, T. O., Refsgaard, J. C. and Madsen, B.:
727 Methodology for construction, calibration and validation of a national hydrological model for
728 Denmark, J. Hydrol., 280(1-4), 52–71, doi:10.1016/s0022-1694(03)00186-0, 2003.

729 Herckenrath, D., Fiandaca, G., Auken, E. and Bauer-Gottwein, P.: Sequential and joint
730 hydrogeophysical inversion using a field-scale groundwater model with ERT and TDEM data,
731 Hydrol. Earth Syst. Sci., 17(10), 4043–4060, doi:10.5194/hess-17-4043-2013, 2013.

732 Hill, M. C.: Effective groundwater model calibration : with analysis of data, sensitives, predictions,
733 and uncertainty, Wiley-Interscience., 2007.

734 Hinnell, A. C., Ferre, T. P. A., Vrugt, J. A., Huisman, J. A., Moysey, S., Rings, J. and Kowalsky,
735 M. B.: Improved extraction of hydrologic information from geophysical data through coupled
736 hydrogeophysical inversion, Water Resour. Res., 46, doi:10.1029/2008wr007060, 2010.

737 Hotelling, H.: Analysis of a complex of statistical variables into principal components, J. Educ.
738 Psychol., 24, 417–441, doi:10.1037/h0071325, 1933.

739 Hyndman, D. W. and Gorelick, S. M.: Estimating lithologic and transport properties in three
740 dimensions using seismic and tracer data: The Kesterson aquifer, Water Resour. Res., 32(9), 2659–
741 2670, doi:10.1029/96wr01269, 1996.

742 Hyndman, D. W., Harris, J. M. and Gorelick, S. M.: COUPLED SEISMIC AND TRACER TEST
743 INVERSION FOR AQUIFER PROPERTY CHARACTERIZATION, Water Resour. Res., 30(7),
744 1965–1977, doi:10.1029/94wr00950, 1994.

745 Härdle, W. K. and Simar, L.: Applied multivariate statistical analysis, 3rd ed., Springer., 2012.

746 Højberg, A. L., Nyegaard, P., Stisen, S., Troldborg, L., Ondracek, M. and Christensen, B. S. B.:
747 DK-model2009. Modelopstilling og kalibrering for Midtjylland, GEUS, København., 2010.

748 Høyer, A.-S., Lykke-Andersen, H., Jørgensen, F. and Auken, E.: Combined interpretation of
749 SkyTEM and high-resolution seismic data, Phys. Chem. Earth, Parts A/B/C, 36(16), 1386–1397,
750 doi:10.1016/j.pce.2011.01.001, 2011.

751 Jorgensen, F., Sandersen, P., Auken, E., Lykke-Andersen, H. and Sorensen, K.: Contributions to the
752 geological mapping of Mors, Denmark - A study based on a large-scale TEM survey, Bull. Geol.
753 Soc. DENMARK, 52, 53–75, 2005.

754 Jorgensen, F., Moller, R. R., Sandersen, P. B. E. and Nebel, L.: 3-D geological modelling of the
755 Egebjerg area, Denmark, based on hydrogeophysical data, Geol. Surv. DENMARK Greenl. Bull.,
756 (20), 27–30, 2010.

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

760 Jørgensen, F., Møller, R. R., Nebel, L., Jensen, N.-P., Christiansen, A. V. and Sandersen, P. B. E.:
761 A method for cognitive 3D geological voxel modelling of AEM data, Bull. Eng. Geol. Environ.,
762 72(3-4), 421–432, doi:10.1007/s10064-013-0487-2, 2013.

763 Kemna, A., Kulessa, B. and Vereecken, H.: Imaging and characterisation of subsurface solute
764 transport using electrical resistivity tomography (ERT) and equivalent transport models, J. Hydrol.,
765 267(3-4), 125–146, doi:10.1016/S0022-1694(02)00145-2, 2002.

766 Kowalsky, M. B., Finsterle, S., Peterson, J., Hubbard, S., Rubin, Y., Majer, E., Ward, A. and Gee,
767 G.: Estimation of field-scale soil hydraulic and dielectric parameters through joint inversion of GPR
768 and hydrological data, Water Resour. Res., 41(11), doi:10.1029/2005wr004237, 2005.

769 Lange, K., Frydendall, J., Cordua, K. S., Hansen, T. M., Melnikova, Y. and Mosegaard, K.: A
770 Frequency Matching Method: Solving Inverse Problems by Use of Geologically Realistic Prior
771 Information, Math. Geosci., 44(7), 783–803, doi:10.1007/s11004-012-9417-2, 2012.

772 Laronne Ben-Itzhak, L. and Gvirtzman, H.: Groundwater flow along and across structural folding:
773 an example from the Judean Desert, Israel, J. Hydrol., 312(1-4), 51–69,
774 doi:10.1016/j.jhydrol.2005.02.009, 2005.

775 Li, R. and Merchant, J. W.: Modeling vulnerability of groundwater to pollution under future
776 scenarios of climate change and biofuels-related land use change: a case study in North Dakota,
777 USA., Sci. Total Environ., 447, 32–45, doi:10.1016/j.scitotenv.2013.01.011, 2013.

778 Linde, N., Finsterle, S. and Hubbard, S.: Inversion of tracer test data using tomographic constraints,
779 Water Resour. Res., 42(4), doi:10.1029/2004wr003806, 2006.

780 Lochbuhler, T., Vrugt, J. A., Sadegh, M. and Linde, N.: Summary statistics from training images as
781 prior information in probabilistic inversion, Geophys. J. Int., 201(1), 157–171,
782 doi:10.1093/gji/ggv008, 2015.

783 Madsen, H.: Parameter estimation in distributed hydrological catchment modelling using automatic
784 calibration with multiple objectives, Adv. Water Resour., 26(2), 205–216, doi:10.1016/S0309-
785 1708(02)00092-1, 2003.

786 Makkink, G. F.: Testing the Penman formula by means of lysimeters, J. Inst. Water Eng., 11, 277–
787 288, 1957.

788 Moutsopoulos, K. N., Gemitzi, A. and Tsihrintzis, V. A.: Delineation of groundwater protection
789 zones by the backward particle tracking method: theoretical background and GIS-based stochastic
790 analysis, Environ. Geol., 54(5), 1081–1090, doi:10.1007/s00254-007-0879-3, 2007.

791 Mukherjee, A., Fryar, A. E. and Howell, P. D.: Regional hydrostratigraphy and groundwater flow
792 modeling in the arsenic-affected areas of the western Bengal basin, West Bengal, India, Hydrogeol.
793 J., 15(7), 1397–1418, doi:10.1007/s10040-007-0208-7, 2007.

794 Park, H., Scheidt, C., Fenwick, D., Boucher, A. and Caers, J.: History matching and uncertainty
795 quantification of facies models with multiple geological interpretations, Comput. Geosci., 17(4),
796 609–621, doi:10.1007/s10596-013-9343-5, 2013.

797 Purvance, D. T. and Andricevic, R.: On the electrical-hydraulic conductivity correlation in aquifers,
798 Water Resour. Res., 36(10), 2905–2913, doi:10.1029/2000WR900165, 2000.

799 Paasche, H. and Tronicke, J.: Cooperative inversion of 2D geophysical data sets: A zonal approach
800 based on fuzzy c-means cluster analysis, Geophysics, 72(3), A35–A39, doi:10.1190/1.2670341,
801 2007.

802 Paasche, H., Tronicke, J., Holliger, K., Green, A. G. and Maurer, H.: Integration of diverse
803 physical-property models: Subsurface zonation and petrophysical parameter estimation based on
804 fuzzy c-means cluster analyses, Geophysics, 71(3), H33–H44, doi:10.1190/1.2192927, 2006.

805 Refsgaard, J. C., Højberg, A. L., Møller, I., Hansen, M. and Søndergaard, V.: Groundwater
806 modeling in integrated water resources management--visions for 2020., Ground Water, 48(5), 633–
807 48, doi:10.1111/j.1745-6584.2009.00634.x, 2010.

808 Refsgaard, J. C., Christensen, S., Sonnenborg, T. O., Seifert, D., Hojberg, A. L. and Troldborg, L.:
809 Review of strategies for handling geological uncertainty in groundwater flow and transport
810 modeling, Adv. Water Resour., 36, 36–50, doi:10.1016/j.advwatres.2011.04.006, 2012.

811 Rossman, N. R. and Zlotnik, V. A.: Review: Regional groundwater flow modeling in heavily
812 irrigated basins of selected states in the western United States, Hydrogeol. J., 21(6), 1173–1192,
813 doi:10.1007/s10040-013-1010-3, 2013.

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

817 Raaschou, P.: Vejledning i Bearbejdning af data fra vandføringsstationer. Publication nr. 7 from
818 Fagdatacenter for Hydrometriske Data, Hedeselskabet, Viborg., 1991.

819 Sandersen, P. B. E., Jørgensen, F., Larsen, N. K., Westergaard, J. H. and Auken, E.: Rapid tunnel-
820 valley formation beneath the receding Late Weichselian ice sheet in Vendsyssel, Denmark, Boreas,
821 38(4), 834–851, doi:10.1111/j.1502-3885.2009.00105.x, 2009.

822 Scanlon, B. R., Faunt, C. C., Longuevergne, L., Reedy, R. C., Alley, W. M., McGuire, V. L. and
823 McMahon, P. B.: Groundwater depletion and sustainability of irrigation in the US High Plains and
824 Central Valley., Proc. Natl. Acad. Sci. U. S. A., 109(24), 9320–5, doi:10.1073/pnas.1200311109,
825 2012.

826 Schamper, C., Jørgensen, F., Auken, E. and Effersø, F.: Assessment of near-surface mapping
827 capabilities by airborne transient electromagnetic data — An extensive comparison to conventional
828 borehole data, GEOPHYSICS, 79(4), B187–B199, doi:10.1190/geo2013-0256.1, 2014.

829 Scharling, P. B., Rasmussen, E. S., Sonnenborg, T. O., Engesgaard, P. and Hinsby, K.: Three-
830 dimensional regional-scale hydrostratigraphic modeling based on sequence stratigraphic methods: a
831 case study of the Miocene succession in Denmark, Hydrogeol. J., 17(8), 1913–1933,
832 doi:10.1007/s10040-009-0475-6, 2009.

833 Seifert, D., Sonnenborg, T. O., Refsgaard, J. C., Hojberg, A. L. and Troldborg, L.: Assessment of
834 hydrological model predictive ability given multiple conceptual geological models, Water Resour.
835 Res., 48, doi:10.1029/2011wr011149, 2012.

836 Selle, B., Rink, K. and Kolditz, O.: Recharge and discharge controls on groundwater travel times
837 and flow paths to production wells for the Ammer catchment in southwestern Germany, Environ.
838 Earth Sci., 69(2), 443–452, doi:10.1007/s12665-013-2333-z, 2013.

839 Sharpe, D. R., Russell, H. A. J. and Logan, C.: A 3-dimensional geological model of the Oak
840 Ridges Moraine area, Ontario, Canada, J. Maps, 239–253, 2007.

841 Sonnenborg, T. O. and Henriksen, H. J.: Håndbog i grundvandsmodellering, GEUS, København.,
842 2005.

843 Steinmetz, D., Winsemann, J., Brandes, C., Siemon, B., Ullmann, A., Wiederhold, H. and Meyer,
844 U.: Towards an improved geological interpretation of airborne electromagnetic data: a case study
845 from the Cuxhaven tunnel valley and its Neogene host sediments (northwest Germany),
846 Netherlands J. Geosci., 94(02), 201–227, doi:10.1017/njg.2014.39, 2014.

847 Stisen, S., Sonnenborg, T. O., Hojberg, A. L., Troldborg, L. and Refsgaard, J. C.: Evaluation of
848 Climate Input Biases and Water Balance Issues Using a Coupled Surface-Subsurface Model,
849 Vadose Zo. J., 10(1), 37–53, doi:10.2136/vzj2010.0001, 2011.

850 Strebelle, S.: Conditional simulation of complex geological structures using multiple-point
851 statistics, Math. Geol., 34(1), 1–21, doi:10.1023/A:1014009426274, 2002.

852 Triantafilis, J. and Buchanan, S. M.: Identifying common near-surface and subsurface stratigraphic
853 units using EM34 signal data and fuzzy k-means analysis in the Darling River valley, Aust. J. Earth
854 Sci., 56(4), 535–558, doi:10.1080/08120090902806289, 2009.

855 Vazquez, R. F., Willems, P. and Feyen, J.: Improving the predictions of a MIKE SHE catchment-
856 scale application by using a multi-criteria approach, Hydrol. Process., 22(13), 2159–2179,
857 doi:10.1002/hyp.6815, 2008.

858 Vilhelmsen, T. N., Behroozmand, A. A., Christensen, S. and Nielsen, T. H.: Joint inversion of
859 aquifer test, MRS, and TEM data, Water Resour. Res., 50(5), 3956–3975,
860 doi:10.1002/2013WR014679, 2014.

861 Wu, J.: Advances in K-means Clustering, Springer Berlin Heidelberg., 2012.

862 Yan, J. and Smith, K.: SIMULATION OF INTEGRATED SURFACE-WATER AND GROUND-
863 WATER SYSTEMS - MODEL FORMULATION, WATER Resour. Bull., 30(5), 879–890, 1994.

864 Zhou, H. Y., Gomez-Hernandez, J. J. and Li, L. P.: Inverse methods in hydrogeology: Evolution
865 and recent trends, Adv. Water Resour., 63, 22–37, doi:10.1016/j.advwatres.2013.10.014, 2014.

866

867

Formatted: Normal (Web), Line spacing: single

Table 1 Calibration parameters of the reference model. Six parameters are calibrated (column 1) to which the remaining six parameters have been tied (columns 3). Initial parameters values of free and tied are shown in columns 2 and 4.

| Parameter, free | Initial parameter value (m/s) | Parameter, tied | Initial parameter value (m/s) |
|---|---|---|---|
| Glacial sand | $3 \cdot 10^{-5}$ | Miocene sand | $3 \cdot 10^{-5}$ |
| Glacial clay | $3 \cdot 10^{-8}$ | | |
| Valley sand | $3 \cdot 10^{-5}$ | Glaciotectonic sand | $3 \cdot 10^{-5}$ |
| Valley clay | $3 \cdot 10^{-8}$ | Glaciotectonic clay | $3 \cdot 10^{-8}$ |
| Palaeogene clay | $1 \cdot 10^{-10}$ | Miocene clay | $3 \cdot 10^{-8}$ |
| Surface clay | $3 \cdot 10^{-8}$ | Surface sand | $3 \cdot 10^{-5}$ |
| | | surface peat | $3 \cdot 10^{-7}$ |

~~Table 2 Calibration~~ and validation statistics for the temporally split sample consisting of observations from 35 wells, which have observations both in the calibration and validation period, and discharge stations 270003 and 270002 ~~and 270003~~.

| | | ~~Reference model~~ | | | 5-cluster model | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | Weighted RMSE (-) | RMSE | ME | ~~Weighted RMSE (-)~~ | ~~RMSE~~ ~~ME~~ |
| **Calibration** ~~2000-2003~~ 2000-2003 | Head (m) | ~~2.63~~ | ~~3.01~~ | ~~-1.01~~ | 1.63 | 1.99 | | -0.~~790~~79 | |
| | Discharge ($m^3/s$) | ~~0.326~~ | ~~0.267~~ | ~~-0.0259~~ | 0.338 | 0.278 | | -0.0107 | |
| **Validation** ~~1995-1999~~ 1995-1999 | Head (m) | ~~2.93~~ | ~~3.32~~ | ~~-0.816~~ | 1.85 | 2.24 | | -0.981 | |
| | Discharge ($m^3/s$) | ~~0.446~~ | ~~0.180~~ | ~~-0.0501~~ | 0.524 | 0.203 | | -0.0354 | |

Deleted Cells
Deleted Cells
Deleted Cells
Deleted Cells
Deleted Cells
Deleted Cells
Merged Cells
Merged Cells
Merged Cells
Merged Cells
Deleted Cells
Deleted Cells
Deleted Cells
Merged Cells
Merged Cells
Merged Cells
Merged Cells
Deleted Cells
Deleted Cells
Deleted Cells

Table 2 Performance statistics of four Danish hydrological models that are comparable to the Norsminde model

| Study | RMSE (m) | Mean error (m) | Horizontal discretization | Model size | Code | Comment |
|---|---|---|---|---|---|---|
| 5-cluster model | 1.99 m | -0.79 m | 100 m | 156 $km^2$ | MIKE SHE | |
| (Stisen et al., 2011) | 3.9 m | 1.2 m | 500 m | 3500 $km^2$ | MIKE-SHE | Mean of calibration using 7 different calibration setups |
| (Seifert et al., 2012) | 3.03 m – 6.34 m | -1.17 m – 0.605 m | 200 m | 465 $km^2$ | MIKE-SHE | Min and max of calibration of 6 different geological models |
| (He et al., 2015) | 4.85 m | - | 100 m | 101 $km^2$ | MIKE-SHE | Mean using borehole based geology |
| (Madsen, 2003) | 1.08 m | 0.19 m | - | 440 $km^2$ | MIKE-SHE | Balanced Pareto optimum |

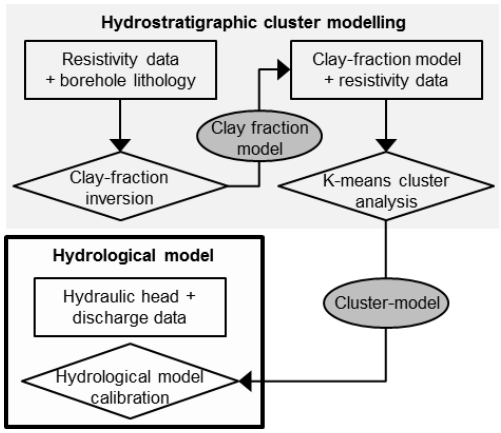
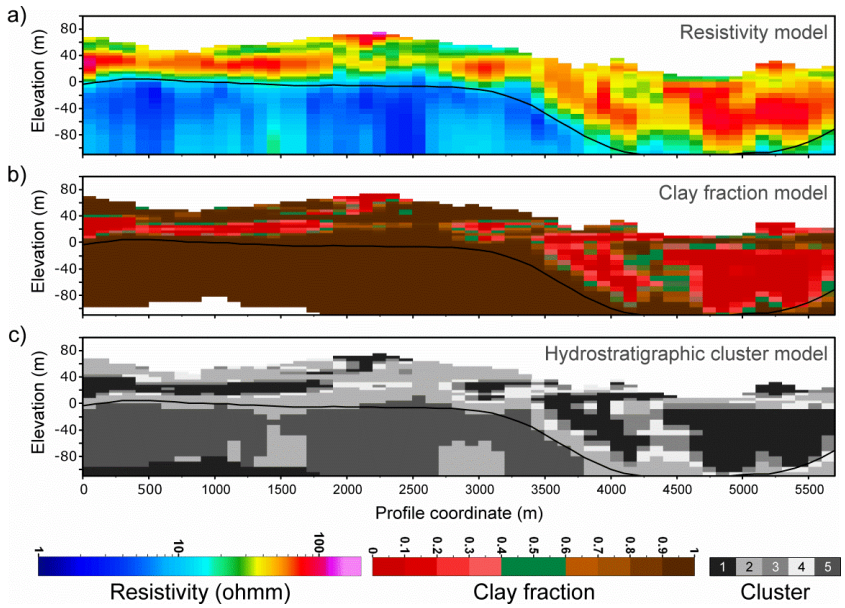



Figure 1 Workflow of the two main parts in the method. Top grey box; hydrostratigraphic cluster modelling using the structural information carried in the geophysical data and lithological information. Lower box in bold; hydrological calibration where hydraulic properties of the hydrostratigraphic zones are estimated using hydrological data.




883

884 Figure 2 Northwest-southeast profiles (vertical exaggeration x5), location is marked in Figure 3. a) Resistivity model, b)

885 clay fraction model, and c) hydrostratigraphic cluster model for the 5-cluster case.

886

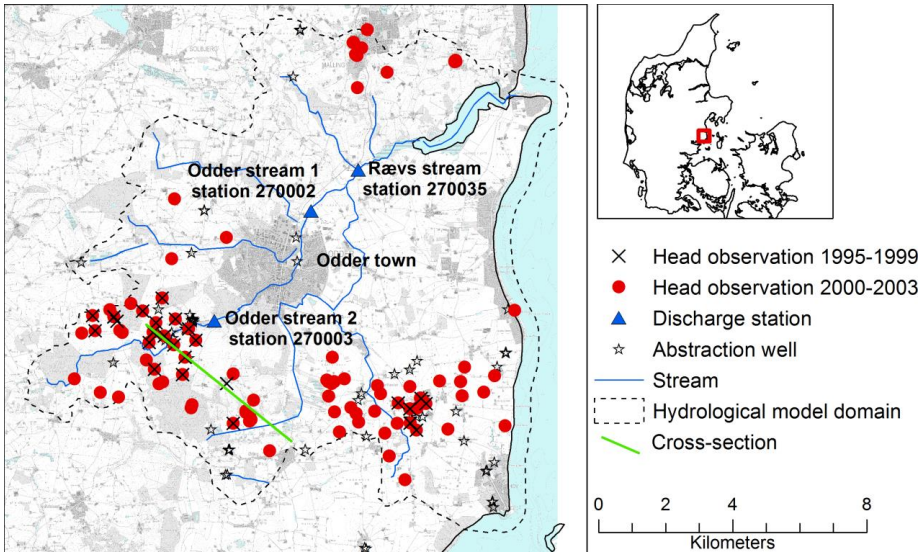



Figure 3 Map of Norsminde study area. The map shows the location of the three discharge gauging stations (blue triangles) along the main river, hydraulic head observations for the calibration period (red dots) and the validation period (black crosses), and abstraction wells (stars). The black dashed line delineates the model domain of the hydrological model.

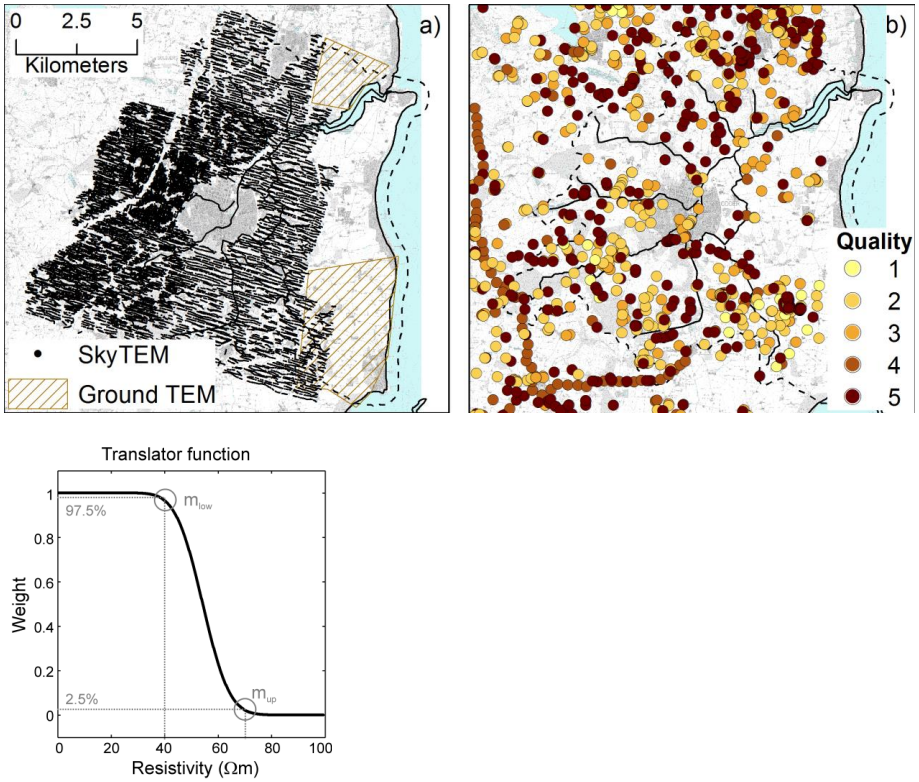



Figure 4~~. Maps showing geophysical and lithological data in the Norsminde area. a) Airborne electromagnetic soundings (small black points) and the extent of the less dense ground-TEM surveys (dashed brown polygons). b) Lithological boreholes used in the clay fraction inversion. Colours indicate quality of the lithological description. The black dashed line delineates the model domain of the hydrological model.~~

The translator function is the petrophysical relationship used in the CF-inversion. The parameters $m_{low}$ and $m_{up}$ are varied to move the translator function along the resistivity axis.

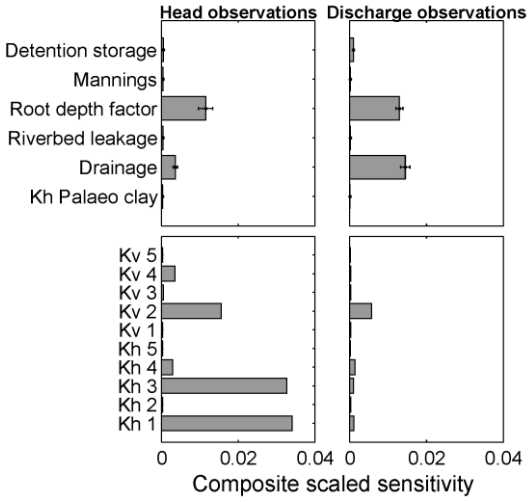



Figure 5 Composite scaled sensitivity values of selected parameters in the hydrological model. Sensitivities are shown for head and discharge observation separately. The two top plots show average, minimum and maximum sensitivity of the 11 hydrostratigraphic cluster models. The two lower plots show sensitivity of subsurface parameters given a 5-cluster model. Kh is horizontal hydraulic conductivity and Kv is vertical hydraulic conductivity.

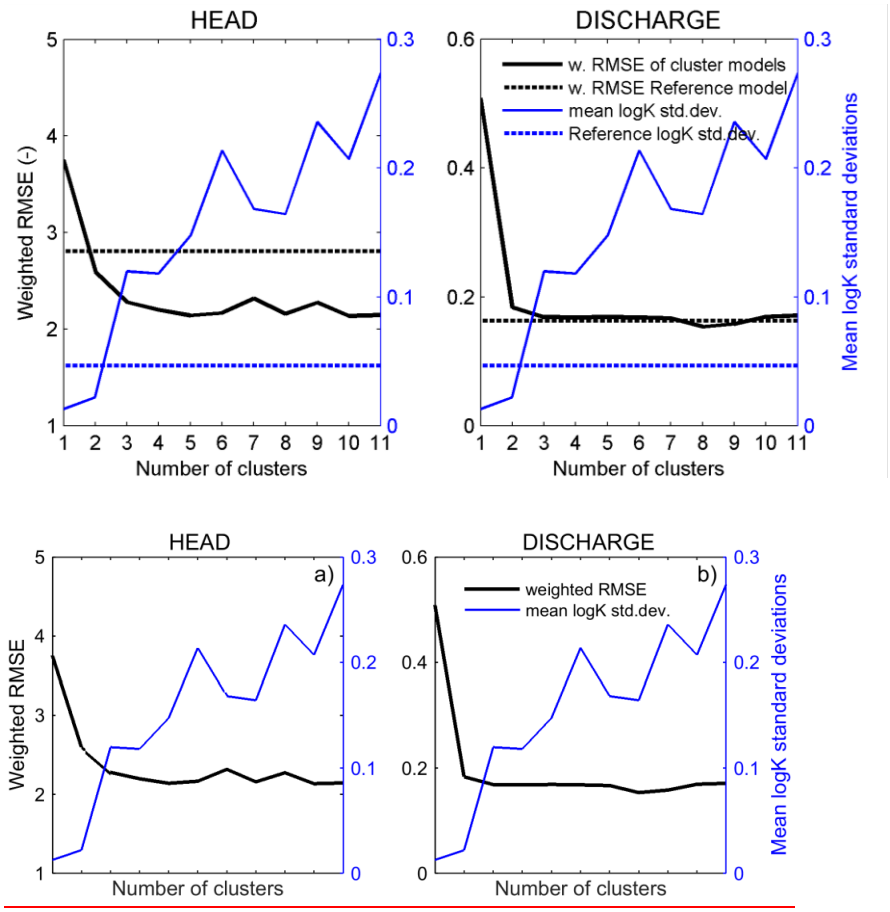
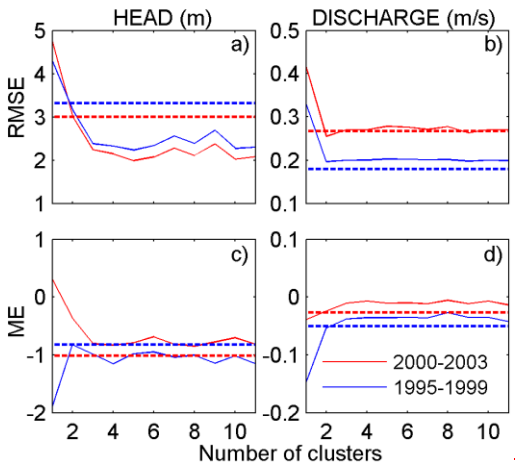



Figure 6 Weighted RMSE of hydrological performance of hydrostratigraphic models consisting of 1 to 11 clusters. Data is shown for all calibration observations. ~~Horizontal dotted lines are weighted RMSE for the reference model.~~ Blue lines are mean standard deviation on log(K) values~~; the solid line represents the cluster models and the dashed represent the reference model~~.



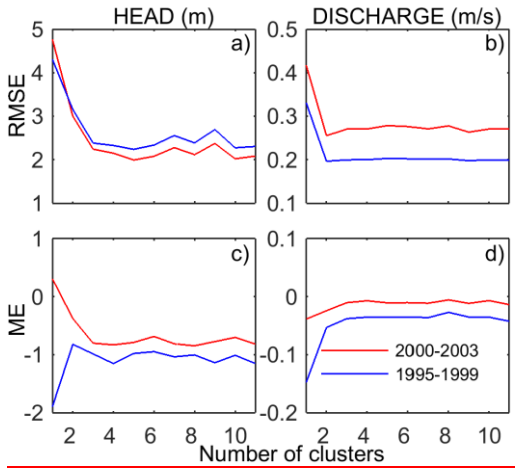




Figure 7 2000-2003 Calibration and 1995-1999 validation period performance statistics for the 11 hydrostratigraphic cluster models ~~and the reference model. Horizontal dotted lines are reference model performance statistics. Row one is RMSE and row two is~~consisting of 1 to 11 clusters. The top row shows RMSE and the bottom row shows ME.

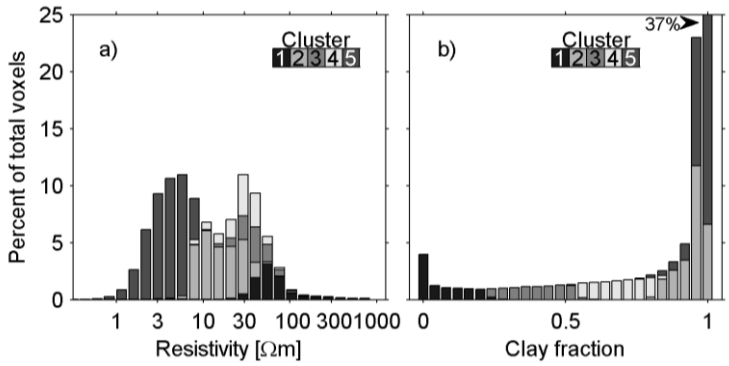
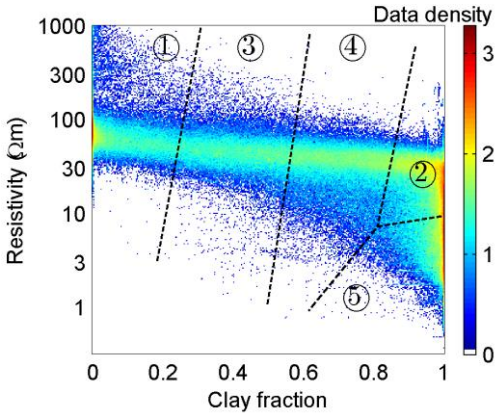


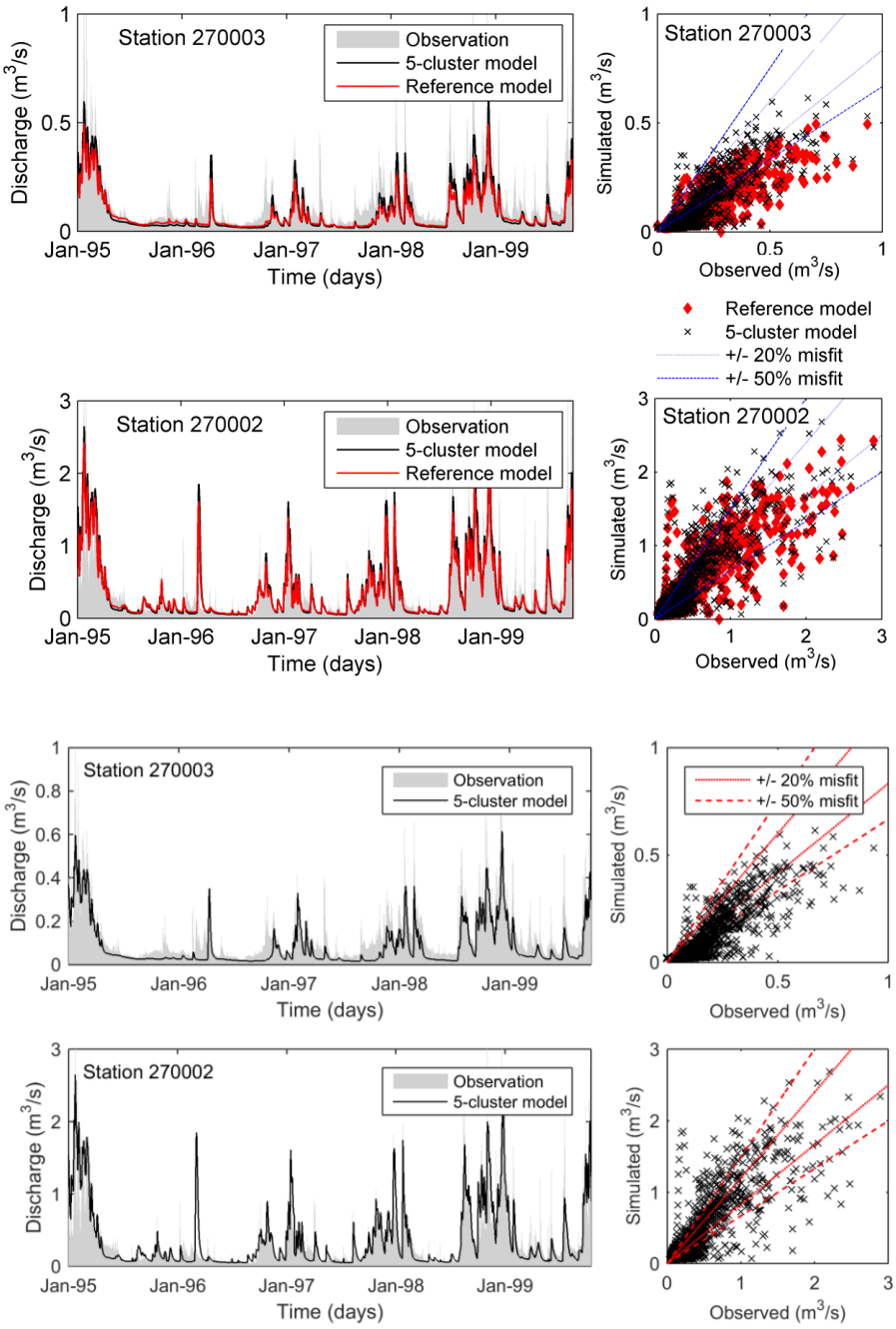

Figure 8 Histograms of a) logarithmic geophysical resistivity values and b) clay fraction values. Cluster memberships of the values are identified by ~~colours~~shades of grey and the histograms thus show how resistivity values and clay fraction values are represented in the clusters. The histograms are shown as percentage of total number of data values.




926

927 Figure 9 Data cloud of geophysical resistivity values and clay fraction values. Dotted black lines indicate cluster
928 interfaces and cluster are labelled with numbers. The cloud colour represents bin-wise data density (300 bins), which is
929 shown in logarithmic scale.

930



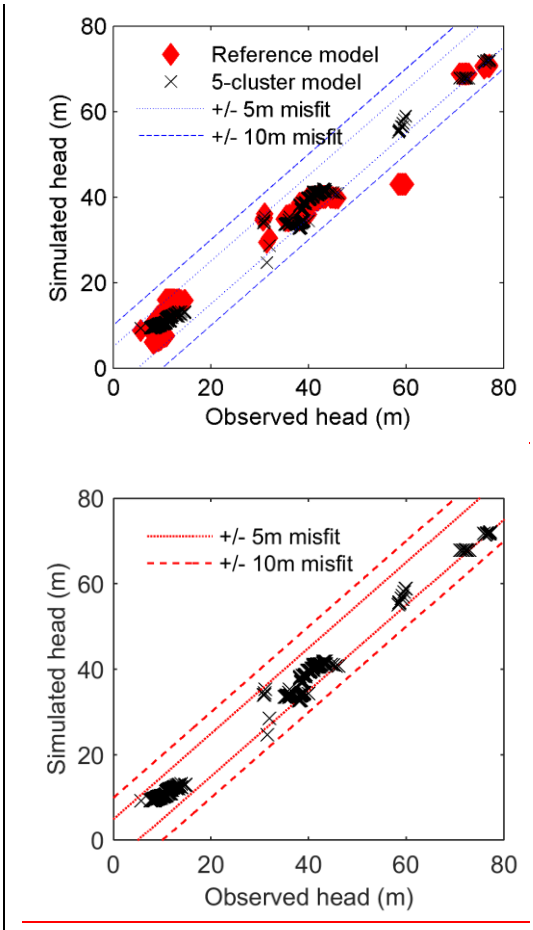

931

932

933 Figure 10 Observed and simulated stream discharge at stations 270003 (top row) and 270002 ~~and 270003~~(bottom row)

934 from the 1995-1999 validation period. To the left stream discharge hydrographs are shown and to the right scatter plots

of observed vs simulated values. In the scatter plots the dotted ~~blue~~and dashed red lines mark misfits of 20% ~~(thin line)~~ and 50% ~~(thick line).~~ respectively.



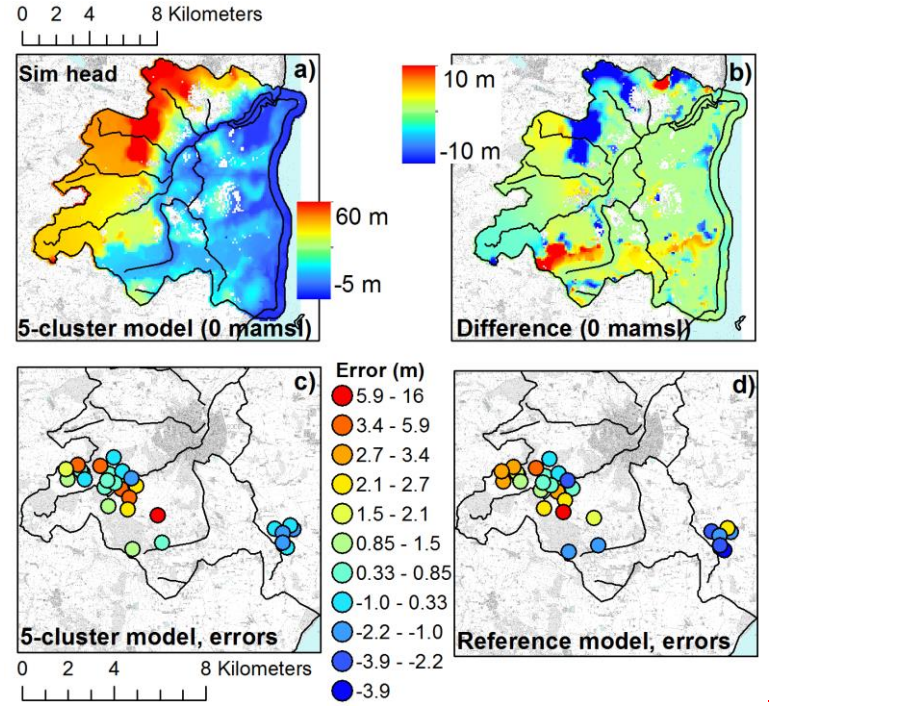

938

939

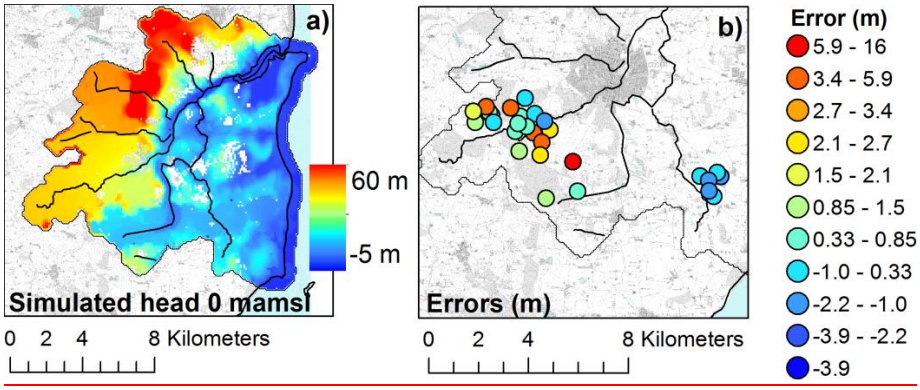
940 Figure 11 Scatter plot of observed and simulated heads values from the 1995-1999 validation period. Dashed lines mark

941 misfits larger than 10 m and dotted lines mark misfits larger than 5 m.

942




943




944

945 Figure 12 Distributed head results for the validation period 1995-1999. (a) 5-cluster model simulated hydraulic head at
946 July 27 1997 at 0 mamsl. (b) ~~Difference maps between the reference model and the 5-cluster model at 0 mamsl~~
947 ~~(difference = reference model – 5-cluster model). (c-d)~~ Errors (~~obs-sim~~observed-simulated) between observed and
948 simulated head~~; c) 5-cluster mode and d) reference model.~~..

949