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)

Abstract

Large-scale 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 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 performance statistics from comparable hydrological 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.

1 Introduction

Large-scale distributed 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).

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 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.

Lithological borehole logs are the fundamental data source for constructing hydrostratigraphic models. The modelling process is often cognitive, but also two-point geostatistical (He et al., 2013; Strebelle, 2002), 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, 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. 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 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 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 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. 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) 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). 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.

2 Materials and Methods

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 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 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. 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 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. 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.

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 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.

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

m_{low} and m_{up} are the model parameters of the translator function, $W(\rho)$, that translates resistivity, ρ, 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 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).

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

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

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 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.

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) 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 (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. Parameterization of spatial variable drain time constant relies on direct drainage flow measurements, and 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 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 $5 \cdot 10^{-5}$ m^{-1} for the entire domain.

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 is unknown; hence all K values have been calibrated. Vertical K values are tied to horizontal K with an anisotropy

266 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
267 fraction value of a zone.
268 Storage parameters were set to a priori values and not calibrated.

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

274 The aggregated objective function, Φ, shown in Eq. 2 is the sum of the scaled objective function for head
275 and discharge. The subjective weight, w_s, was determined through trial and error by starting numerous
276 calibration runs; w_s was chosen to be 0.8.

277 $$\Phi = w_s \sum_{i=1}^{N_h} \left(\frac{h_{sim,i} - h_{obs,i}}{\sigma_i} \right)^2 + (1 - w_s) \sum_{i=1}^{N_q} \left(\frac{q_{sim,i} - q_{obs,i}}{\sigma_i} \right)^2 \qquad \text{(Eq. 2)}$$

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

285 Uncertainty of stream discharges is mainly due to translation from water stages to discharge (daily mean
286 discharges). Uncertainties originate from infrequent calibration of rating curve, ice forming on streams and
287 especially stream bank vegetation (Raaschou, 1991). Errors can be as large as 50%. Blicher (1991) estimates
288 errors of 5% and 10% on the water stage measurement and rating curve respectively. In cases of very low

289 stream flows (1 L/s) Christensen et al. (1998) assigned a standard deviation of 200% while flow of 50 L/s
290 and 5-10 L/s are assigned standard deviations of 5% and 25% respectively. We have assigned an error of
291 20% to all stream discharge observations.

292 3 Results and discussion

293 First we show results for the hydrological performance of 11 hydrostratigraphic cluster models consisting of
294 1-11 zones. Secondly details of the cluster analysis for the case of a 5-cluster hydrostratigraphy are shown.
295 Finally the cluster model hydrological performance is benchmarked with comparable hydrological models.

296 3.1 Calibration and validation of hydrological model

297 Figure 6 shows the weighted RMSE of model performances for hydrostratigraphic cluster model consisting
298 of 1 to 11 zones, head and discharge respectively is shown in Figure 6a and Figure 6b. The 1-cluster model is
299 a homogeneous representation of the subsurface resulting in a uniform K field. The 1-cluster model
300 represents a situation where we have no information about the subsurface. Increasing the number of clusters
301 to represent the subsurface successively adds more information from geophysical and lithological data to the
302 calibration problem. The weights used to calculate weighted RMSE are the same weights as used in Eq. 2.

303 Head and discharge contribute by approximately 2/3 and 1/3 of the total objective function. From the 1-
304 cluster to the 2-cluster model, weighted RMSE for discharge is reduced by more than a factor 2. No
305 significant improvement of the fit to discharge data is observed for more than 2 clusters. Fit to head data
306 improves almost by a factor of 2 from the 1-cluster to the 2-cluster model. Improvement of the fit to head
307 data continues up to the 5-cluster representation of the subsurface. Improvements are a factor of 3 from the
308 1-cluster to the 5-cluster model. Beyond the 5-cluster model, the fit to head observations stagnates. The 7-
309 cluster and 9-cluster hydrostratigraphic models perform worse than the 3-cluster model. The 8-, 10-, and 11-
310 cluster models obtain an equally good or better fits to head data compared to the 5-cluster model.

311 The blue lines in Figure 6 illustrate mean standard deviation on log(K) values of the cluster models based on
312 the post-calibration standard deviation of log(K) for each K zone. Beyond the 4- and 5-cluster models the
313 precision of the estimated K values decrease. The mean standard deviations on log(K) for the 4- and 5-
314 cluster models are 0.12 and 0.15. The corresponding widths of the 95% confidence intervals are between
315 15% and 90% of the estimated K value for 3 out of 4 zones and 3 out of 5 zones, respectively. Beyond the 5-
316 cluster model mean standard deviations on log(K) are between 0.17 and 0.27, and corresponding width of the
317 95% confidence intervals are largely above 100% for all but two zones.

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

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

326 The hydrostratigraphic models are constructed under the assumption that subsurface structures governing
327 groundwater flow can be captured by structural information contained in clay fraction values (derived from
328 lithological borehole data) and geophysical resistivity values. If this is true, an asymptotic improvement of
329 the data fit would be expected for increasing cluster numbers. However, as shown in Figure 6, this is not
330 strictly the case: Weighted RMSE of the 7-cluster and 9-cluster models is higher than weighted RMSE of the
331 3-cluster, 6-cluster and 8-cluster models, respectively. The likely explanation is that increasing number of
332 clusters does not correspond to pure cluster sub-division, but also to relocation of cluster interfaces in the 3D
333 model space. We expect the difference in hydrological performance to be due to changes in interface
334 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. 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 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 respectively. The hydrological models analysed in this study generally under-simulate the average discharge.

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 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 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 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.

3.3 Benchmarking hydrological performance

Table 1 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 270003 the 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. With respect to head, the model under-simulates in the elevated parts of the domain (head above 50 m), see Figure 11. The head values below 20 m represent the Boulstrup tunnel valley, where head is fitted the best. With weighted RMSE for head of 1.63 and 1.85 the model is almost 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.

Figure 12a-b show distributed head results. Generally hydraulic head in the tunnel valley is disconnected from the elevated terrain (Figure 12a), and groundwater overall flows towards the sea. Figure 12b shows errors (obs-sim) between observed and simulated heads for 1995-1999. The largest errors 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

411 comparability with respect to data density and quality, and hydrostratigraphy. The model performances are
412 compared based on RMSE and ME of simulated heads, see 2, as these statistics are reported in the studies.
413 The horizontal discretization of the models is 100 m, 200 m, and 500 m, and the models cover between 202
414 km^2 and 3500 km^2. We can see that the 5-cluster model is comparable with the other models.

415 3.4 Advantages and limitations

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

420 The risk of misinterpretation of AEM data due to effects of saturation, water quality, depth and material
421 dependent resolution, and vertical shielding, are higher with an automatic approach compared to a cognitive
422 approach, as these effects may be identified by a geologist during the modelling process. AEM data can be
423 integrated into geological models using cognitive methods, for example as presented by Jørgensen et al.
424 (2013), who provide an insightful discussion of the pros and cons of automatic versus cognitive geological
425 modelling from AEM data.

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

432 The hydrostratigraphic cluster model presented in this paper does not represent a lithological model, but has
433 the advantage of incorporating close to all the structural information contained in the large AEM data sets in
434 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. 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.

4 Conclusion

We have presented an automated workflow to parameterize and calibrate a large-scale hydrological model based on AEM and borehole data. The result is a competitive hydrological model that performs 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 3- and 5-cluster models give the optimal performance.

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

462 # 5 Appendix A: Observation errors

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

464 $$\sigma^2_{total} = \sigma^2_{meas} + \sigma^2_{elev} + \sigma^2_{\text{int}} + \sigma^2_{hetereo} + \sigma^2_{unknown}$$

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

467 Head measurements are typically carried out with dip-meter, and occasionally pressure transducers are used.
468 Information about which measurement technique has been used for the individual observations is not clear
469 from the Jupiter database. It is assumed that dip-meters have been used and σ_{meas} has been determined to be
470 0.05 m for all observations.

471 Well elevations are referenced using different techniques. The elevation can be determined from a 1:25000
472 topographic map, by levelling or by differential GPS. The inaccuracies for using topographic maps and
473 DGPS measurements are in the order of respectively 1-2 m and centimetres. The Jupiter database can have
474 information about the referencing techniques, but this information is rarely supplied. An implicit information
475 source is the number of decimal places the elevations have in the database. Elevation information is supplied
476 with 0, 1 or 2 decimal places. For the wells where the reference technique is available (checked for cases
477 with topographic map and DGPS only) the decimal places reflect accuracy of the referencing technique used.
478 From this information decimal places of 0, 1 and 2 have been associated with σ_{elev} of 2 m, 1 m and 0.1 m
479 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 Δx 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 the Palaeogene tunnel valley. Hydraulic gradients of the Miocene/Glacial west and the Palaeogene tunnel valley are between 0.001-0.002. The Miocene/Glacial area and the Palaeogene tunnel valley areas were thus considered as one with σ_{int} of 0.07 m. The glacial tectonic area has an estimated hydraulic gradient of 0.01 and thus associated with σ_{int} of 0.6 m.

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

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

6 Acknowledgements

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

7 References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

622 Kemna, A., Kulessa, B. and Vereecken, H.: Imaging and characterisation of subsurface solute
623 transport using electrical resistivity tomography (ERT) and equivalent transport models, J. Hydrol.,
624 267(3-4), 125–146, doi:10.1016/S0022-1694(02)00145-2, 2002.

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

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

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

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

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

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

642 Madsen, H.: Parameter estimation in distributed hydrological catchment modelling using automatic
643 calibration with multiple objectives, Adv. Water Resour., 26(2), 205–216, doi:10.1016/S0309-
644 1708(02)00092-1, 2003.

645 Makkink, G. F.: Testing the Penman formula by means of lysimeters, J. Inst. Water Eng., 11, 277–
646 288, 1957.

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

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

653 Park, H., Scheidt, C., Fenwick, D., Boucher, A. and Caers, J.: History matching and uncertainty
654 quantification of facies models with multiple geological interpretations, Comput. Geosci., 17(4),
655 609–621, doi:10.1007/s10596-013-9343-5, 2013.

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

658 Paasche, H. and Tronicke, J.: Cooperative inversion of 2D geophysical data sets: A zonal approach
659 based on fuzzy c-means cluster analysis, Geophysics, 72(3), A35–A39, doi:10.1190/1.2670341,
660 2007.

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

664 Refsgaard, J. C., Højberg, A. L., Møller, I., Hansen, M. and Søndergaard, V.: Groundwater
665 modeling in integrated water resources management--visions for 2020., Ground Water, 48(5), 633–
666 48, doi:10.1111/j.1745-6584.2009.00634.x, 2010.

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

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

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

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

678 Sandersen, P. B. E., Jørgensen, F., Larsen, N. K., Westergaard, J. H. and Auken, E.: Rapid tunnel-
679 valley formation beneath the receding Late Weichselian ice sheet in Vendsyssel, Denmark, Boreas,
680 38(4), 834–851, doi:10.1111/j.1502-3885.2009.00105.x, 2009.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

725

726

727 Table 1 Calibration and validation statistics for the temporally split sample consisting of observations from 35 wells,
728 which have observations both in the calibration and validation period, and discharge stations 270003 and 270002.

		5-cluster model		
		Weighted RMSE (-)	RMSE	ME
Calibration 2000-2003	Head (m)	1.63	1.99	-0.79
	Discharge (m^3/s)	0.338	0.278	-0.0107
Validation 1995-1999	Head (m)	1.85	2.24	-0.981
	Discharge (m^3/s)	0.524	0.203	-0.0354

729

730 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

731

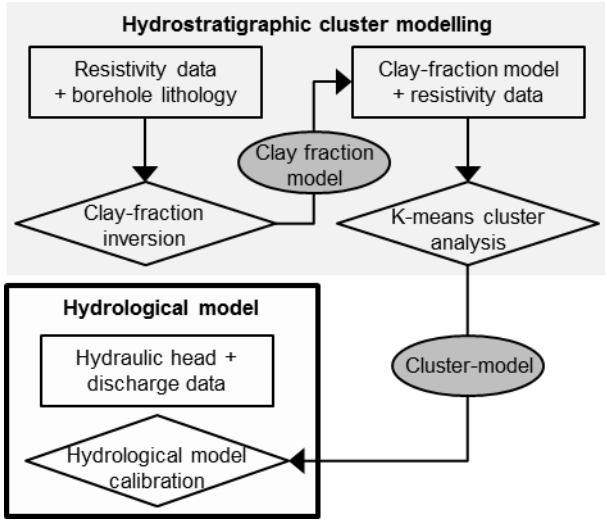
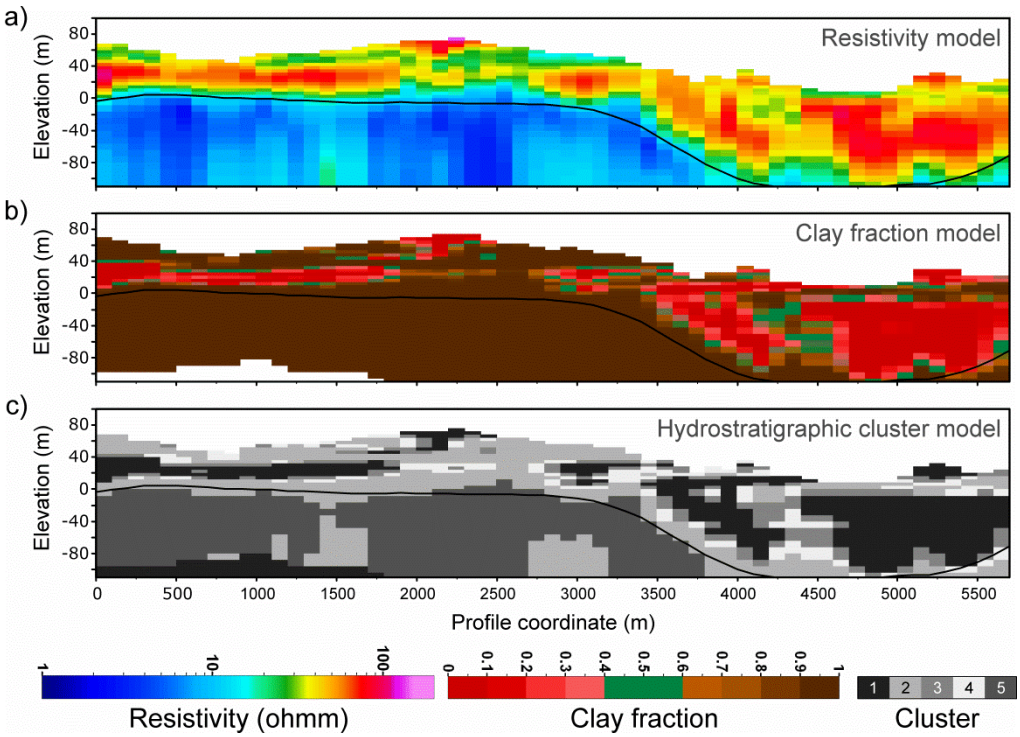


Figure 1 Workflow of the two main parts in the method. Top grey box; hydrostratigraphic cluster modelling using the structural information carried in the geophysical data and lithological information. Lower box in bold; hydrological calibration where hydraulic properties of the hydrostratigraphic zones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

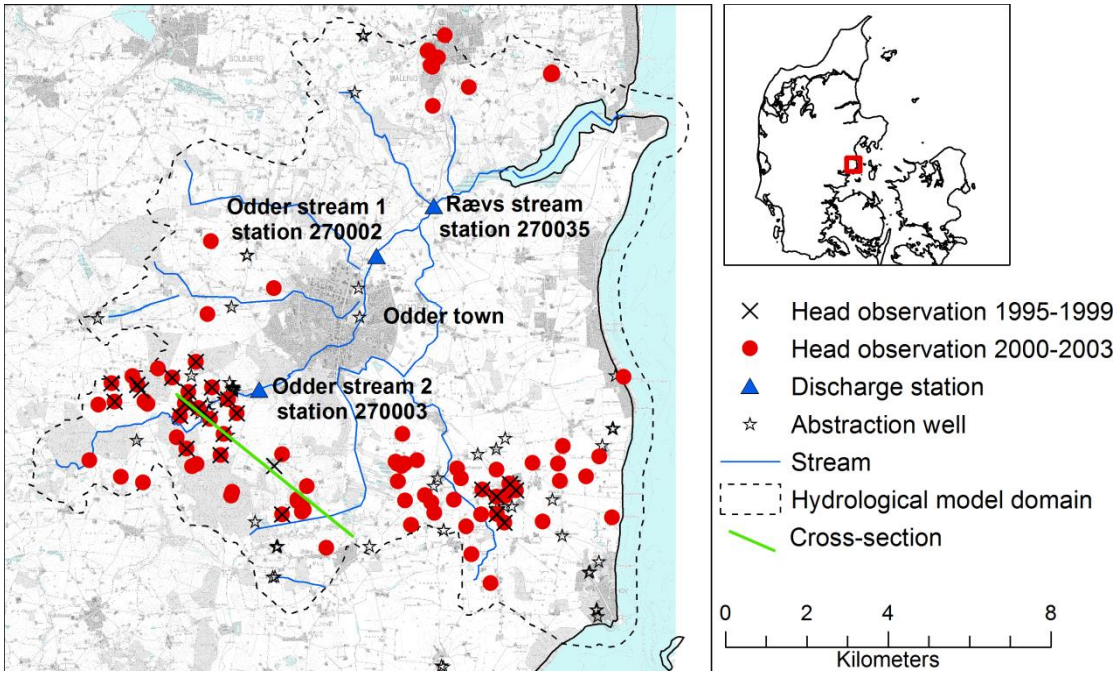


737

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

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

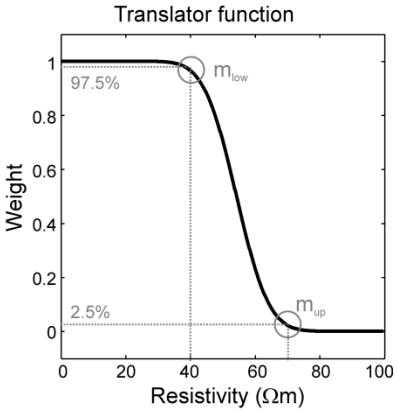
740



741

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

746

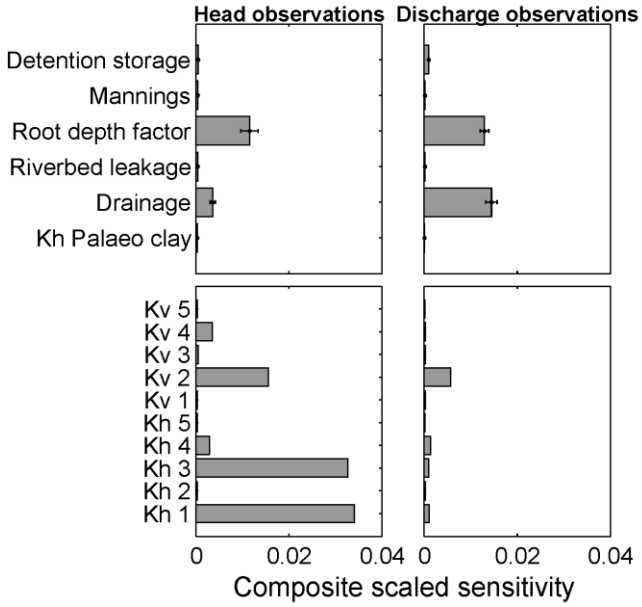


747

748 Figure 4 The translator function is the petrophysical relationship used in the CF-inversion. The parameters m_{low} and m_{up}

749 are varied to move the translator function along the resistivity axis.

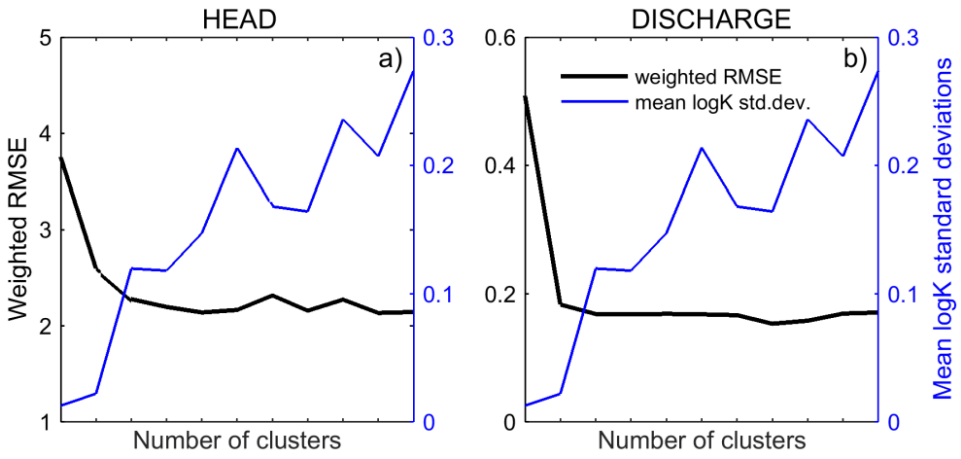
750



751

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

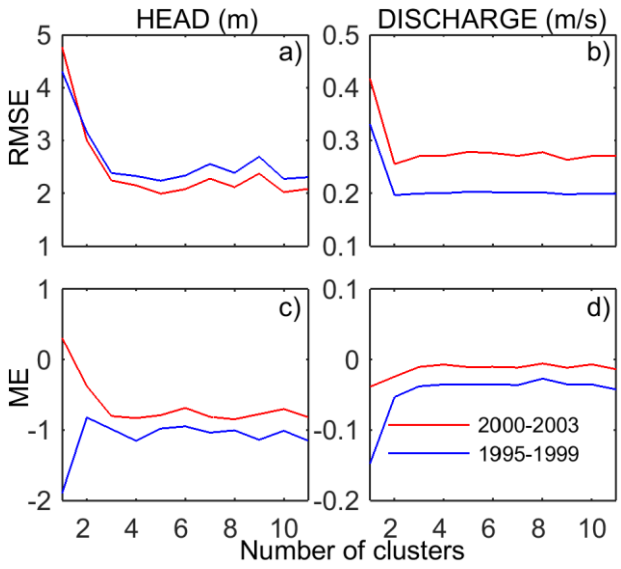
756



757

758 Figure 6 Weighted RMSE of hydrological performance of hydrostratigraphic models consisting of 1 to 11 clusters. Data is
759 shown for all calibration observations. Blue lines are mean standard deviation on log(K) values.

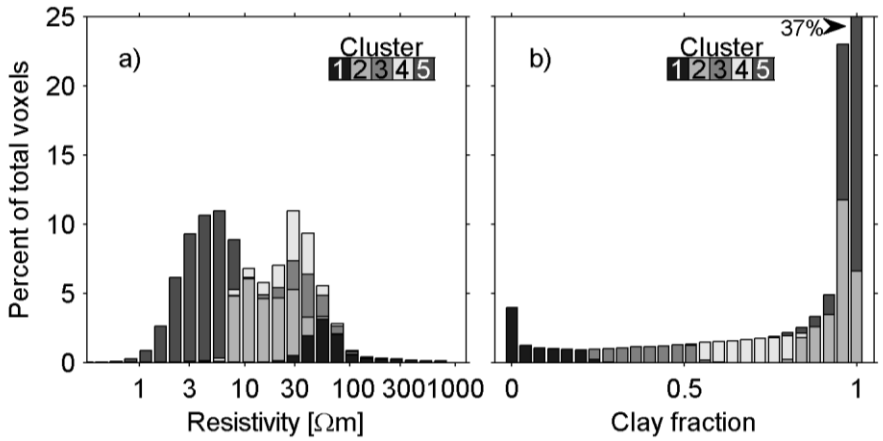
760



761

762 Figure 7 2000-2003 Calibration and 1995-1999 validation period performance statistics for the 11 hydrostratigraphic
763 cluster models consisting of 1 to 11 clusters. The top row shows RMSE and the bottom row shows ME.

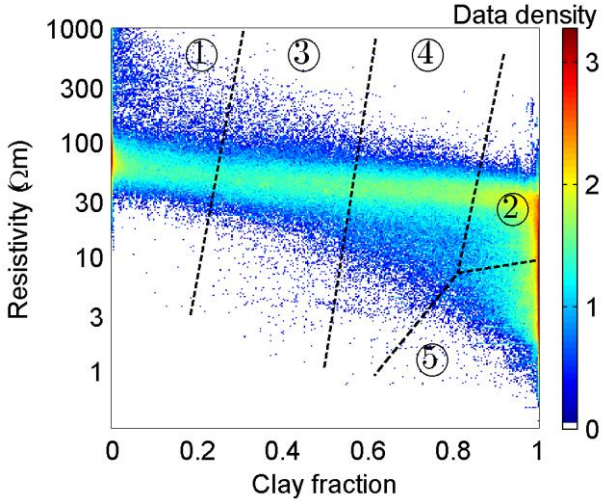
764



765

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

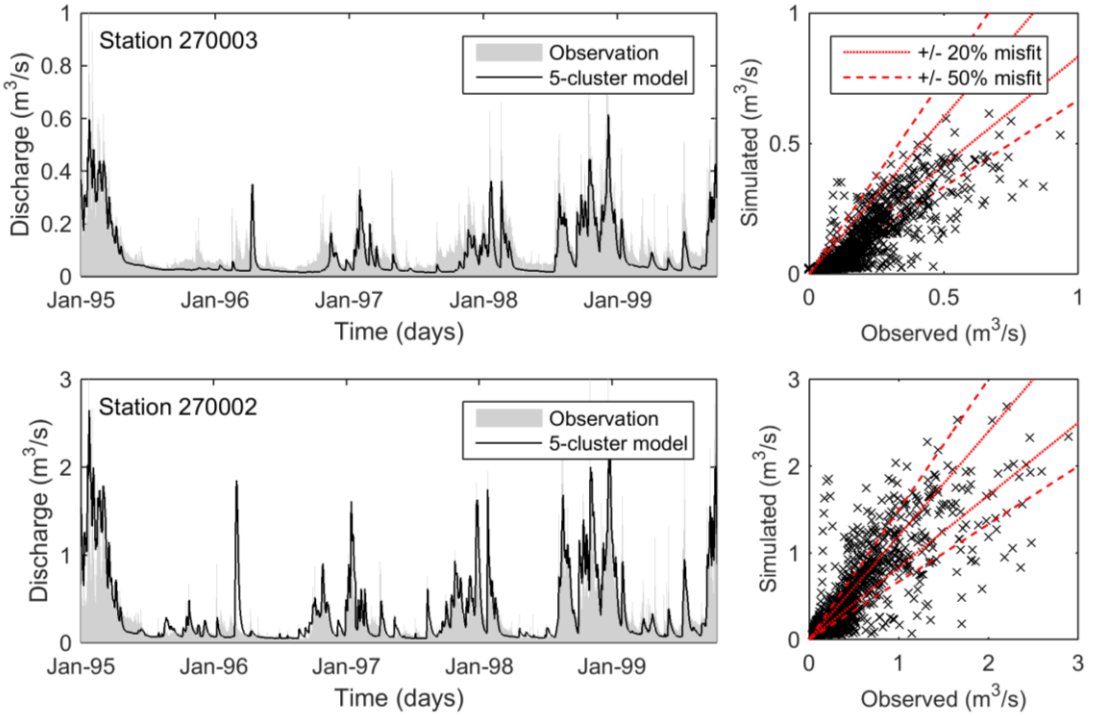
769



770

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

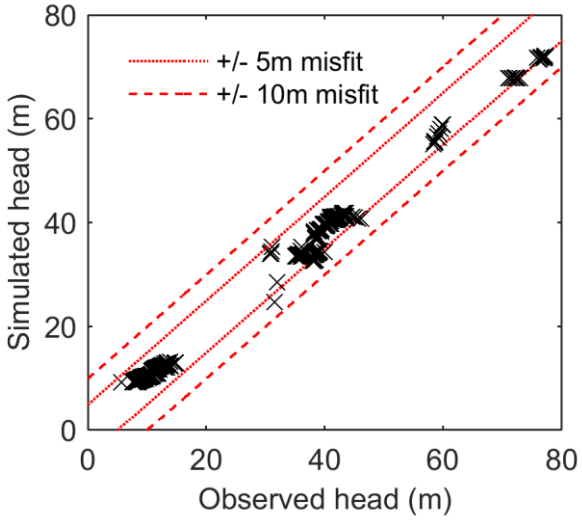
774



775

776 Figure 10 Observed and simulated stream discharge at stations 270003 (top row) and 270002 (bottom row) from the
777 1995-1999 validation period. To the left stream discharge hydrographs are shown and to the right scatter plots of
778 observed vs simulated values. In the scatter plots the dotted and dashed red lines mark misfits of 20% and 50%
779 respectively.

780

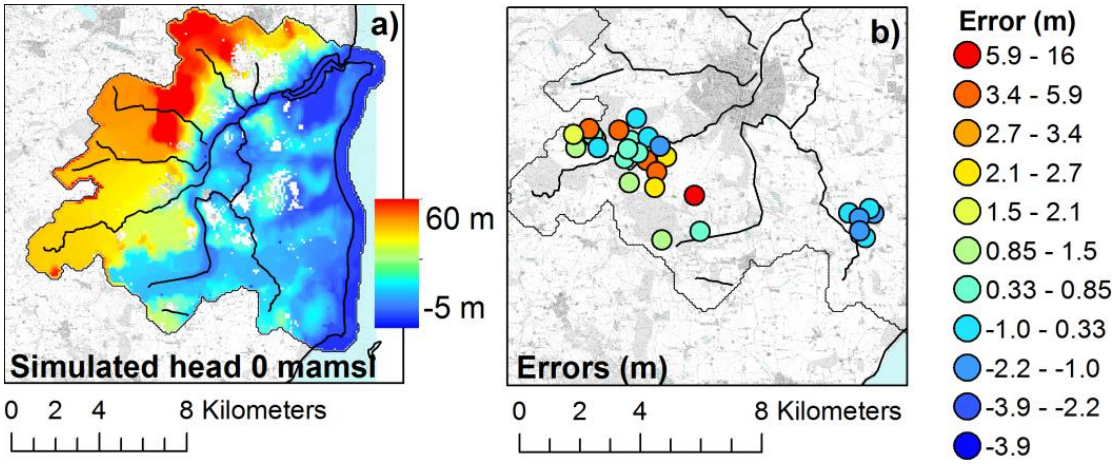


781

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

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

784



785

786 Figure 12 Distributed head results for the validation period 1995-1999. (a) 5-cluster model simulated hydraulic head at

787 July 27 1997 at 0 mamsl. (b) Errors (observed-simulated) between observed and simulated head..

788