



A Regional Scale Hydrostratigraphy Generated from Geophysical Data of Varying Age, Type, and Quality

Troels Vilhelmsen¹ · Pernille Marker² · Nikolaj Foged¹ · Thomas Wernberg³ · Esben Auken¹ · Anders Vest Christiansen¹ · Peter Bauer-Gottwein² · Steen Christensen¹ · Anne-Sophie Høyer⁴

Received: 4 October 2016 / Accepted: 9 January 2018 /

Published online: 1 November 2018

© Springer Nature B.V. 2018

Abstract

In the present study, we show how persistent management and collection of hydrological and geophysical data at a national scale can be combined with innovative analysis methods to generate decision support tools for groundwater and surface water managers. This is exemplified by setting up a regional scale groundwater model in an area with geophysical data of varying age, type, and quality. The structure for the regional model is derived from a newly developed resistivity clay-fraction cluster analysis. This modelling strategy can be used in combination with local detailed geological modelling thus utilizing the detailed expertise locally, while securing a cost-effective (price vs. performance) solution to the numerical simulations of the regional scale water balance. In this way we avoid unwanted boundary effects on the local model simulations due to the presence of artificial numerical boundaries located proximate to the areas of interest. In this application, it is particularly important that boundary conditions are remote, due to the presence of a dense network of buried valley structures. Simulated impacts of groundwater abstraction from two existing well-fields spread through the valley system far beyond the local focus areas of the study.

Keywords Geophysics · Groundwater modelling · Groundwater management · Data integration · MODFLOW-USG · Buried valleys

✉ Troels Vilhelmsen
troels.norvin@geo.au.dk

¹ Department of Geoscience, Aarhus University, Høegh-Guldbergs Gade 2, 8000 Aarhus C, Denmark

² DTU Department of Environmental Engineering, Bygningstorvet Bygning 115, 2800 Kgs. Lyngby, Denmark

³ NIRAS, Ceres Allé 3, 8000 Aarhus C, Denmark

⁴ GEUS, C.F. Møllers Allé 8, 8000 Aarhus C, Denmark

1 Introduction

Water resource managers are often faced with the challenge of evaluating how their decisions manifest themselves throughout the hydrological system. This could relate to groundwater abstraction and the impact it may have on local streams and wetlands or to evaluating the risk of groundwater contamination. For that purpose, groundwater models are routinely being used for decision support. However, especially for the scenarios where groundwater abstraction is evaluated, the areal extent of such numerical models must be large enough to avoid unwanted effects from boundary conditions (BC). Depending on the type of BC applied, simulating the effects of groundwater abstraction using groundwater models with too small an areal extent is prone to either over- or underestimation of the drawdown. This problem is even more profound when dealing with groundwater abstraction from buried valley structures. For example, van der Kamp and Maathuis (2012) present a case where an aquifer test was conducted in a buried valley for a period of 8 days. This resulted in a drawdown of 20 cm at a distance of 13.4 km. Had such data been evaluated in a hydrological model, the need for distant BC would be highly important, while having a relatively fine numerical discretization in the vicinity of the pumping well.

Two problems can easily be identified when setting up groundwater models covering large areas. The first is dealing with the computational demands related to the numerical setup. This limitation has often led to the use of relatively coarse numerical resolutions in regional groundwater flow models. This problem has been partly solved with the emergence of unstructured groundwater model codes, such as e.g. MODFLOW-USG (Panday et al. 2015). MODFLOW-USG can have a fine numerical discretization locally around specific areas of interest, while maintaining a coarser discretization regionally. The second problem is setting up the hydrostratigraphy. Traditionally, this is a manual and time-consuming task. These models often contain a large degree of detail and they incorporate the knowledge of the geologists, who make the interpretations (e.g. Jørgensen et al. 2013; Kessler et al. 2009; Ross et al. 2005). When managers are on a tight financial budget, it may be hard to argue that a larger portion of the budget should be spent on defining the regional structure despite its importance. This is especially true, when governing structural features, likely to have a high impact on the modelled responses, exist near the study's focus area.

Studies of regional groundwater flow in glacially dominated areas are quite common in the northern hemisphere (Artimo et al. 2008; Atkinson et al. 2014; Herzog et al. 2003). The aquifer systems in these areas or subparts of these areas are often dominated by the sand/clay distribution. Recently, a cost-effective method called the clay-fraction (CF)/Resistivity clustering method was developed to define structural input for groundwater models (Foged et al. 2014; Marker et al. 2015). In these studies, the hydrostratigraphy was based solely on data collected using the SkyTEM system (Sørensen and Auken 2004). However, such dense datasets are rarely available for regional scale areas. In Denmark, geophysical data are stored in a national database called GERDA (Møller et al. 2009). The contents of this database have been expanded over a long period, and for the present study area, it contains data collected over a period of nearly 30 years using five different geophysical methods. These data are now available to groundwater managers, and collectively they form a rich dataset. Regional structural models can therefore potentially be developed based on this dataset. Using the CR/resistivity clustering method, we document the value of this national geophysical dataset for groundwater managers. This could act as inspiration to other countries to develop similar systems or act as an inspiration to an alternative method to develop structures for regional scale groundwater models.

2 Study Area

The app. $\sim 470 \text{ km}^2$ focus area of the study is located northwest of the city of Aarhus, Denmark (Fig. 1). It has been subject to several geological and geophysical surveys (Danielsen et al. 2003; Høyer et al. 2015; Kronborg et al. 1990; Sandersen and Jørgensen 2003) due to its interesting geological structures, and rich groundwater resources. A subpart of the areas has been the focus area of the HyGEM research project, with the main goal of developing methods for direct and automated incorporation of geophysical data into groundwater models.

2.1 Regional Hydrogeology

The hydrogeology in the area is dominated by the presence of buried valley structures. Here, these valleys can be mapped effectively using the Transient Electromagnetic Method (TEM) (Christiansen et al. 2006) because they are incised into low resistivity Paleogene clay deposits (Jørgensen et al. 2003). Several mapping campaigns have been conducted in the area using conventional ground based TEM methods, PATEM (Sørensen et al. 2000), and the SkyTEM systems. Boreholes show that the buried valleys primarily have infill containing Quaternary sand, gravel, and till deposits. Locally, Miocene sandy units are overlying the Paleogene clay layers.

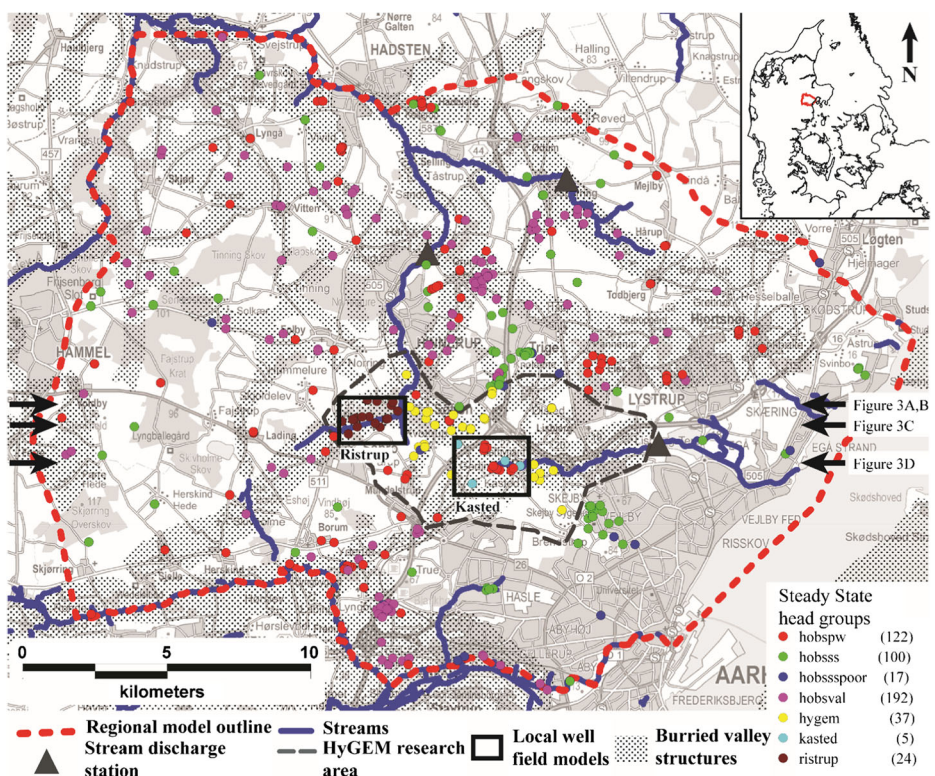


Fig. 1 Location of the regional scale study area located north-west of the town of Aarhus, Denmark. The figure also marks the location of observation data used in the hydrological model, major streams, local areas, and major buried valley structures

The importance of the valley structures for the drinking water supply is evident from the fact that approximately 70% of the screened boreholes in the area are located within known buried valley structures (Fig. 1). Minor aquifers also exist locally on the plateaus between the buried valleys. However, these aquifers are mainly utilized for groundwater abstraction to single households or smaller communities.

The two local focus areas of this study are shown in Fig. 1 (Ristrup and Kasted). Both of these areas are of major interest due to the presence of existing well fields. The geological structures have been modeled in detail (Høyer et al. 2015). According to descriptions from boreholes in the east-west running valley near Ristrup, the aquifer consists of 100 m of coarse sand and gravel. Despite the thick sequence of coarse grained deposits, the yield of the aquifer is limited by the valley structure. It is therefore expected that the influence of the pumping will spread throughout the valley system, and to a lesser extent into the aquifers located at the plateaus to the north and south. The hydraulic connection from the well-field towards the northeast through the valley structure is limited by the presence of a hydraulic barrier, which will further increase the drawdown.

The hydrostratigraphy near the Kasted well-field is more complex. Based on the findings of Høyer et al. (2015), at least five generations of valley structures are found in the area. Some valleys act as aquifers, and others are filled up with clay acting as aquitards or aquicludes. The system of valleys connects the well-fields near Kasted and Ristrup, which leaves the important question of how these two well-fields interact despite their distance of 5 km, and also how the expected pumping induced drawdown influences local streams and wetlands.

2.2 Hydrological Data Availability

For the purpose of the present study, we have utilized two types of hydraulic data. Namely, steady state head elevations and stream base flows (Fig. 1). The hydraulic heads were extracted from the Danish national well database, JUPITER (Møller et al. 2009). The quality of the head data varies significantly between boreholes, spanning from poor quality data from water supply wells to high quality data from wells that are part of the nationwide groundwater-monitoring program. The uncertainty of the hydraulic heads is also affected by the accuracy of the borehole datum and coordinates. Boreholes mapped using differential GPS, are associated with less uncertainty than those mapped using topographical maps. For the purpose of model calibration, the uncertainty for each hydraulic head measurement was estimated, such that error-based weighting could be applied in the analysis. To assist post-calibration analysis, observations were subdivided into 7 groups. First, we excluded the boreholes that were considered to be outliers (hobssspoor) from the analysis. Second, the boreholes were divided into two groups. Those located inside a valley structure (hobssval) and the ones located on plateaus (hobssss). Approximately 25% of the wells are used for groundwater abstraction (hobsspw). Using these in the groundwater model calibration can be problematic due to limitations in the numerical resolution of the model. In the present case, this will affect the calculated head elevations in the vicinity of the pumping well, thus making the simulated values a poor representation of the field measurements. The subset was not excluded from the analysis, but it was down-weighted compared to the remaining observation groups. The three remaining groups were based on either their location within the HyGEM research area (hygem), or the two local scale well-field models (Ristrup or Kasted).

Stream discharge measurements were available from three gauging stations. Two stations had time series spanning 20 years, while the last station had a span of 3 years. To estimate the

base flow for each of the gauging stations the methodology described by Arnold and Allen (1999) was used resulting in base flow variation with time. These base flow records were used to estimate the temporal average base flow, and the associated uncertainty.

2.3 Geophysical and Geological Data

2.3.1 Boreholes

Borehole information was extracted from the JUPITER database. The study area holds approximately 4000 boreholes (Fig. 2c), where app. 3000 boreholes have lithological logs. The majority of boreholes are shallow (<30 m), and only ~5% have a depth > 100 m. The JUPITER database also holds different borehole meta-data (drill type, age, etc.). Based on the borehole meta-data the quality of each borehole was rated into four groups, using the scheme presented by He et al. (2014). Borehole ratings were used for error based weighting in the structural modeling.

2.3.2 Geophysical Data

The study area has been intensely mapped with geophysical methods, primarily using electrical (DC) and TEM methods. The DC methods resolve shallow geological variations, while TEM methods resolve resistivity contrasts at greater depths.

The TEM data set (Fig. 2a) consists of ground-based single site soundings (TEM40) and PATEM, and two airborne SkyTEM surveys from 2003 and 2014. The TEM data were collected to delineate the aquifers and the important buried valley structures. The PATEM-system is a towed system providing continuous data along the drive path. Development and improvements of the airborne SkyTEM system have resulted in a higher density and much improved near surface resolution in the 2014 SkyTEM survey (central SkyTEM survey in Fig. 2a), than the 2003 SkyTEM survey towards the west and north.

1D layered resistivity models with 3 to 5 layers were used in the interpretation of the TEM40 data. The PATEM data and the older SkyTEM survey were also interpreted with layered 1D resistivity models, but in a lateral constrained inversion setup (Auken et al. 2005). The 2014 SkyTEM survey uses a 1D sharp model formulation (Vignoli et al. 2015) in a spatial constrained inversion setup (Viezzoli et al. 2009). In total ~3000 TEM40, 7400 PATEM, and 24,600 SkyTEM resistivity models were used.

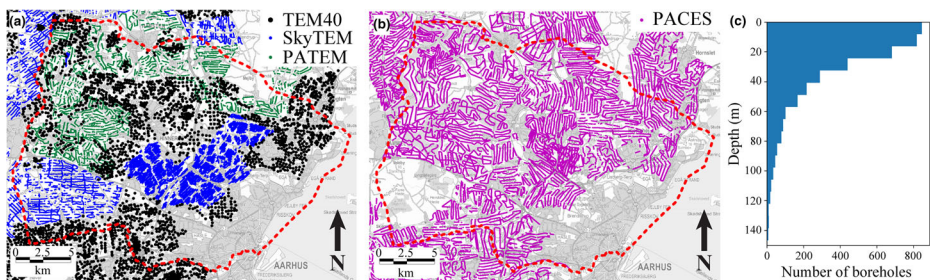


Fig. 2 **a** EM data available in the study area. TEM40 indicates single site soundings. The newest SkyTEM dataset collected in 2014 is the dense blue “cloud” of dots in the central Eastern part of the area. **b** DC data available in the study area. **c** Number of boreholes in specific depth intervals within the field area

The DC data of the study area (Fig. 2b) consist of two generations of the Pulled Array Continuous Electrical Sounding data (PACES) (Sørensen, 1996), collected in the 1990's. The PACES system provides detailed resistivity information of the upper ~20 m, and a good spatial resolution is achieved with a line density of approximately 300 m and a model spacing of 10 m along the lines. The PACES data were interpreted with 1D 3-layer models in a laterally constrained inversion setup. In total ~225,000 PACES models were used.

The combined geophysical data set was collected using at least 9 different instrument types. The data collection spans 30 years subdivided into more than 50 individual surveys. Combining these data into one seamless geophysical model was possible due to a strict and comprehensive quality control in the data collection and processing phase, as well as a common database system handling all geophysical data collected. For the TEM data this includes calibration at the Danish National TEM test-site (Foged et al. 2013).

3 Methods

The applied methods comprise three different ways of delineating the hydrostratigraphy: “cluster”, “alternative” and “local”. The first is the main method of this paper, the second is a fallback method in areas where there are no geophysical data available, and the third is a manual interpretation in two main areas of interest (Ristrup and Kasted).

3.1 Cluster Model Hydrostratigraphy

The regional scale structural model (see Fig. 1 for outline) is generated using the CF – resistivity clustering methodology. The outcome of the CF – resistivity clustering is a subdivision of the subsurface into a predefined number of categories. It is therefore an inherent assumption of the concept that throughout a given category formed by the CF and resistivity values the hydrological properties are constant. This conforms to the traditional zone based groundwater modeling approach.

In short terms the CF-modelling is generated by translating the resistivity data into a clay-fraction value for the model space divided into voxels. The optimum resistivity to clay-fraction translation function is determined by lithological borehole logs. The translation function is varying spatially in 3D thereby adapting to the borehole information, the different geological environments, etc. The basic assumption of the CF-concept is that the clay materials are more electric-conductive than the non-clay materials. This assumption is fulfilled in the study area because the non-clay materials are highly dominated by sand and gravel deposits, and because saline pore water has not been observed (Søndergaard et al. 2004).

To form the structures of the hydrological model the CF and resistivity data are delineated into zones using k-means clustering (e.g. Jain 2010; Wu 2012) by iteratively minimizing the squared Euclidean distance between the data points and their cluster centers. Details on the resistivity-CF clustering method can be found in Foged et al. (2014) and Marker et al. (2015)

Based on the overall geological understanding of the area it was decided to subdivide the cluster model by depth into two parts: The upper 30 m below terrain, and the deeper parts below a depth of 30 m. This separation corresponds approximately to the delineation between the shallow aquifers located at the plateaus and the deeper aquifers located in the buried valleys. The structural model was generated as a voxel model with a horizontal discretization of 100 by 100 m and a vertical resolution of 10 m. The optimal number of clusters in the

analysis was determined from their success in hydrological model calibration. We thereby generated model structures with between 2 and 7 clusters in each of the shallow and the deep subdomains. Each of these structural model realizations were calibrated to the hydrological dataset using the groundwater model. Here we observed a gradual improvement in data fit up until the 6 cluster model (6 cluster each subdomain). This selection is similar to the one applied by Marker et al. (2015). A cross section of this optimal cluster model is shown on Fig. 3b.

In areas with no geophysical data coverage, like below the town of Aarhus, the CF-cluster methodology could not be applied without interpolating resistivities and CF translations over large distances. Such interpolation can introduce bias into the model structure. In these parts of the model, the cluster structure was substituted by a manual interpretation adopted from the alternative model shown in Fig. 3a. Examples on how they are joined can be seen in the last part of the profiles shown in Fig. 3c and d.

3.2 Alternative Model Hydrostratigraphy

The performance of the cluster model was evaluated by comparing it to an alternative model structure adopted from the National water resource model of Denmark “the DK-model” (Henriksen et al. 2003). This structural model is relatively simple and mainly based on boreholes. For the purpose of the present study, the DK-model is slightly updated to improve

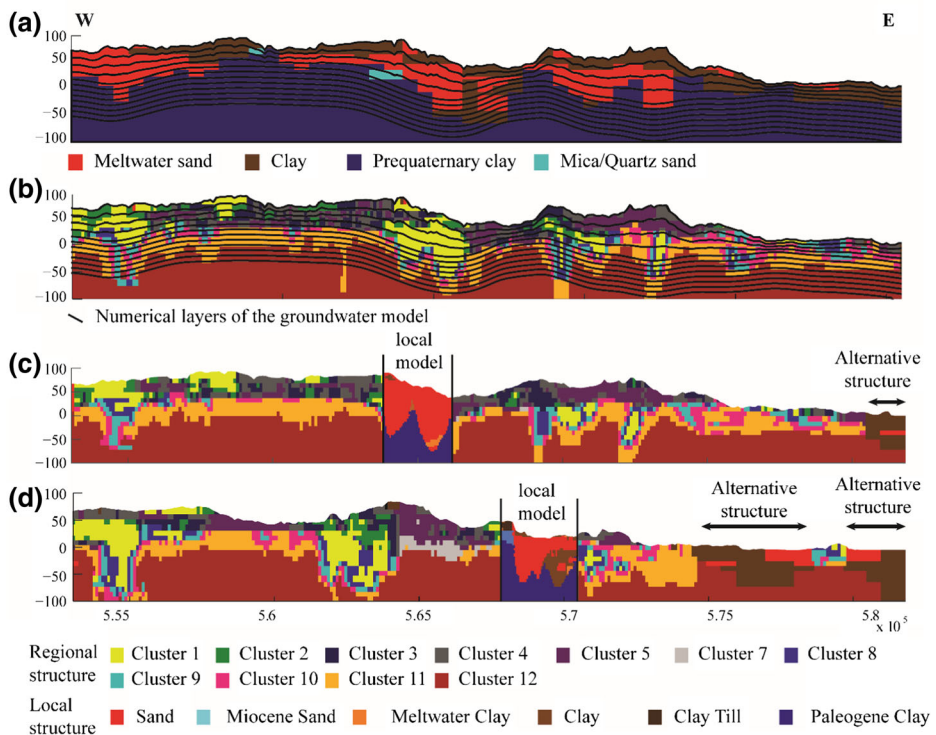


Fig. 3 Cross sections for the three model setups (location shown on Fig. 1). **a** is the model extracted from the DK-model (Henriksen et al. 2003; Sonnenborg et al. 2003), **b** is the cluster model derived from the combined analysis of borehole and geophysical data, **c** is the cross section with the local model through Ristrup, and **d** is the cross section with the local model through the Kasted site

the representation of the buried valley structures. In the regional area, only 6 different geological units are available in the model. These are: meltwater sand, clay, mica sand, quartz sand, prequaternary sand, and chalk (Fig. 3a). The model is a voxel model with a horizontal resolution of 500 m by 500 m and a vertical resolution of 10 m.

3.3 Local Model Hydrostratigraphy

Detailed and manually defined geological models have been incorporated into the two local focus areas (Fig. 1). The structure is extracted from the geological model developed by Høyer et al. (2015). Examples can be seen from Fig. 3, where an east-west cross section cuts through the areas of the Ristrup well-field (Fig. 3c) and the Kasted well-field (Fig. 3d), respectively. Both of the local scale models are voxel models with a horizontal discretization of 50 m by 50 m, and a vertical resolution of 5 m. Each voxel refers to a specific type of sediment based on an interpretation of the dominant lithology in the voxel. The lithologies comprise Paleogene clay, Miocene sand, Quaternary meltwater sand, clay till, valley infill made up of Quaternary meltwater sand, and valley infill made up of Quaternary meltwater clay or clay till. Based on a manual interpretation of the individual valleys, each structure has been given a unique type of deposit, such that sand and clay units in individual valley structures can be distinguished. Details on the model development can be obtained from Høyer et al. (2015).

3.4 Groundwater Model Setup

To utilize the higher detail of the localized well-field models, while having a coarser resolution at the regional scale to reduce model runtime, the groundwater model was set up in MODFLOW-USG.

Major stream segments were simulated using the river (RIV) package, and the DRN package were used to simulate shallow subsurface runoff and runoff from minor trenches and stream segments that may run dry. Recharge was simulated using the RCH package, and pumping wells were simulated using the WEL package. Details on these MODFLOW BC types can be found in Harbaugh et al. (2000).

In the present model we used the unstructured formulation for the horizontal discretization. In the regional scale area of the model we used a discretization of 100 m, and in the two local areas, we used a 50 m discretization. Vertical refinement was not used.

3.5 Voxel to Layer Based Model Structure

The vertical discretization of the groundwater model is different from that of the structural model in a way similar to the one applied in the numerical groundwater model code MIKE SHE (DHI 2009, pp. 44–47). Vertical and horizontal effective hydraulic conductivities were thereby calculated within each cell of the groundwater model depending on the relative elevation of the structural model layers compared to the groundwater model layers. This has some advantages as the numerical layers do not have to follow boundaries of specific geological units, and it is thus straightforward to represent geological layers that pinch out. Because the structural models are voxel based, it can be avoided to include more layers in the hydrological model than necessary from both a hydrological and a computational (forward runtime) perspective. The up scaled hydro-stratigraphy model was imported into MODFLOW using the LPF-package (Harbaugh et al. 2000).

The representation of the numerical layers in the groundwater model can be seen in Fig. 3, where the black lines represent layer boundaries of the groundwater model. The effective hydraulic property values are vertically averaged from the structural grid to the numerical grid using the weighted arithmetic mean value

$$K_{h,i} = \frac{\sum_{j=1}^n z_{StrL,n} K_{StrL,n}}{\sum_{j=1}^n z_{StrL,n}} \quad (1)$$

where $z_{StrL,n}$ is the thickness of the part of the structural model layer number n lying within the numerical cell number i , and $K_{StrL,n}$ is the hydraulic conductivity of the unit type within the corresponding structure. The vertical hydraulic conductivity is calculated as the geometric mean of the vertical hydraulic conductivity of the structural units:

$$K_{v,1} = \frac{\sum_{i=1}^n z_{StrL,n}}{\sum_{i=1}^n \frac{z_{StrL,n}}{K_{StrL,n}}} \quad (2)$$

3.6 Groundwater Model Calibration

The groundwater model was calibrated using nonlinear least squares inversion as implemented in PEST (Doherty 2016). Details on this methodology can be found in e.g. Aster et al. (2005) or Doherty (2015). For the present study, the objective function is formulated as

$$\Phi = \sum_{i=1}^n [\omega_{h,i} (h_o - h_s)_i]^2 + \sum_{j=1}^m [\omega_{q,j} (q_{o,j} - q_{s,j})_j]^2 \quad (3)$$

where $h_{o,i}$ is the i 'th hydraulic head observation, and $h_{s,i}$ is the simulated equivalents. $q_{o,j}$ is the j 'th base flow measurement, and $q_{s,j}$ is the simulated equivalent, $\omega_{h,i}$ is the weight associated with head observation i , and $\omega_{q,j}$ is the weight associated with base flow observation j . The head weights were assigned by taking into account the evaluated quality of the boreholes, as mentioned in section describing hydrological data availability.

For the regional scale cluster model, the parameterization consisted of horizontal hydraulic conductivities of all clusters except for the cluster constituting the lower limit of the aquifer systems. Vertical hydraulic conductivities were tied to their horizontal counterparts with a ratio of 1/10. The stream bed conductance was estimated with a linear scaling for each river cell in the model. This was done to maintain the relative conductance between model cells containing long and short stream segments, respectively. The spatial distribution of the recharge was taken from the output of the national water resource model of Denmark (Henriksen et al. 2003), and it was calibrated using a constant scaling factor.

For the alternative structure taken from the DK- model, the parameters constituted hydraulic conductivity of each geological unit except for the Paleogene clay, which was fixed in the inversion. Similar to the cluster structure, we used a fixed ratio between horizontal and vertical hydraulic conductivities. The remaining parameters were the same as for the cluster structure model.

The local scale models shown are voxel models set up with a predefined lithology subdivided into six types. Within each local area, parameters for hydraulic conductivity were estimated for these lithologies independently.

In cases where the applied parameterization scheme resulted in an ill posed inversion problem, the inversion was regularized using singular value decomposition (Doherty 2015).

4 Results

4.1 Calibration Results

Figure 4 presents the results obtained using the two different regional structural model approaches. Figure 4a for the DK-model structure and Fig. 4b for the cluster structure. The two major groups are hobssval and hobsss (observation count in brackets). The DK-model structure is performing relatively well for the observations located outside the valley structures with a mean error (ME) close to zero, and a relatively low root mean square error (RMSE). On the contrary, poor performance is seen for observations located inside valley structures. This is not surprising, because the structure of the DK-model was primarily built from boreholes, which due to their low density have a poor resolution of the valley structures. The model has a relatively high bias in the simulated heads within the HyGEM research area. This is likely to be caused by the geological complexity of this area.

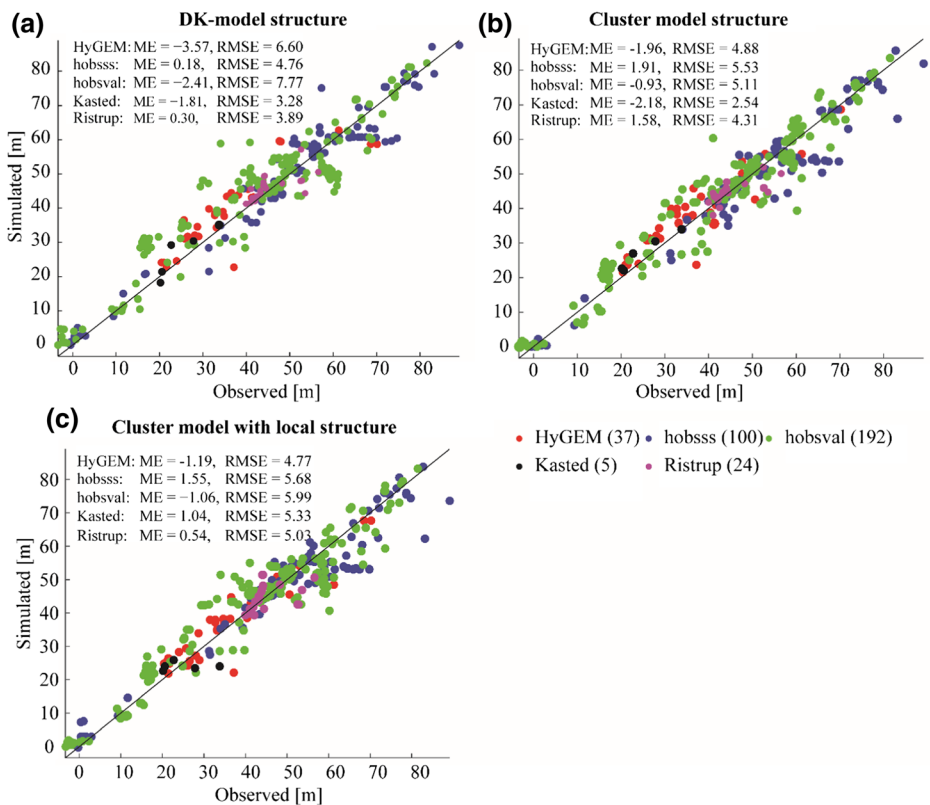


Fig. 4 Observed versus simulated hydraulic heads. **a** shows the calibration results for the DK-model structure. **b** shows the results for the regional cluster model structure (**c**) shows the results obtained by combining the regional cluster structure with the local structures

The cluster model, performs relatively well for observations located within valley structures, but relatively poorly for observations located outside valley structures. In the HyGEM research area, the cluster model significantly out-performs the DK-model structure.

The results for the cluster model, including the two locally refined areas, can be seen in Fig. 4c. Surprisingly, the model performance deteriorates in the Ristrup area. This is caused by poor representation of a single well located at the edge of the valley structure. The hydraulic gradient between this well and wells in the valley structure is steep, indicating poor hydraulic connection. However, in the model this borehole is located within the valley structure. Moreover, this borehole is constructed with three wells screening different depths, and three observation points are thereby available at this location, resulting in high penalty in the statistics with poor representation. The three observations can be seen as the purple dots with observed values around 55 m. The remaining observations for the Ristrup area are located close to the identity line, indicating good performance. Using the localized models, we see an improved performance in the ME, while, the RMSE deteriorates.

4.2 Selected Model Forecasts

Two types of model forecasts are made using the regional cluster model containing the two locally detailed models. The first forecast is the recovery of the hydraulic head after cessation of groundwater abstraction at both well-fields. The simulated recovery is shown on Fig. 5a. The recovery is largest near the pumping wells, and spreads through the buried valley structures. The recovery stretches far beyond the two local well-field models. Based on the model results, it is also likely that the well-fields will interact. This is noticed from the simulated recovery from the Kasted well field, which stretches through the north-west striking buried valley structure towards the Ristrup well-field.

Due to strict regulations of groundwater abstraction's effect on surface water, it is often of interest to analyze how this affects streams and wetlands. The second forecast is therefore the increase in outflow to the surface due to cessation of the abstractions. Figure 5b indicates that the abstractions have limited influence on stream discharge around the Ristrup well field. The effect is most profound in the vicinity of the stream, where the drawdown is largest. Figure 5c shows the result for the Kasted well field. Based on the modelled scenario, the effect on the surface discharge is more profound here. Figure 5c also shows the trench systems that were dug mostly before the establishment of the Kasted well field. The simulated increase of outflow around the trenches supports the validity of the forecast.

5 Discussion

We have presented an application of the CF-resistivity cluster method to generate structural input to regional scale groundwater models. Compared to previous applications of the method, this study applies abundant resistivity data obtained during several field campaigns using a variety of different geophysical methods. The entire data collection spans a period of nearly 30 years. Except for the new SkyTEM survey, the model was developed solely on the basis of existing data, this including geophysical, geological, and hydrological data. It demonstrates the large payoffs for water resource management by a continued effort to develop and update geological, hydrological, and geophysical databases.

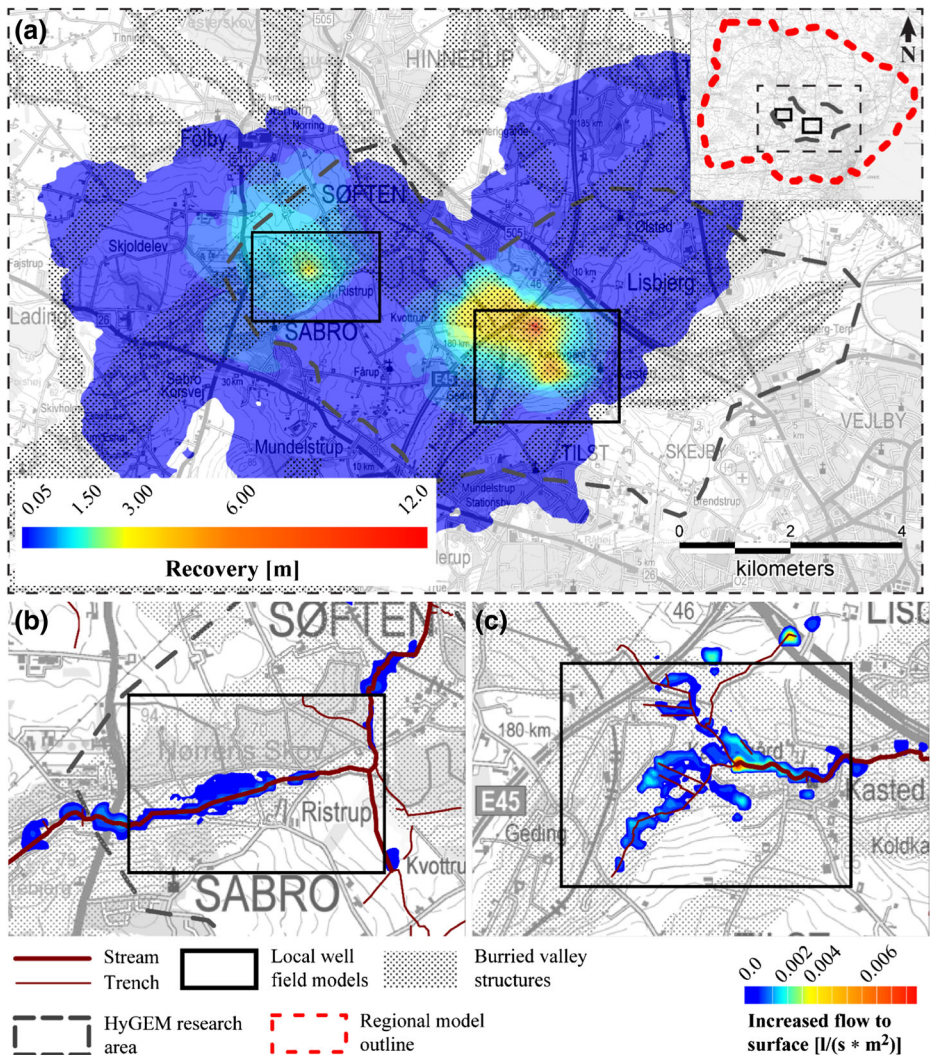


Fig. 5 a Forecast of head increase caused by cessation of abstraction from the well-fields at Kasted and Ristrup. Forecast showing an estimate of the expected increased flux to the surface after ceasing groundwater abstraction at the two well-fields for (b) Ristrup and (c) for Kasted

The CF-resistivity cluster method builds on the assumptions that the distribution of resistivities can be related to the water bearing and confining units. In the investigated area, the aquifers are primarily located in Quaternary sand and gravel deposits. These deposits tend to have high resistivities. On the contrary, the aquitards and aquicludes are related to Quaternary till clays and Paleogene clays that have intermediate and low resistivity values, respectively. These conditions conform with conditions reported previously by Mazac et al. (1985). The methodology is less likely to be useful in areas where resistivity intervals of aquifers and aquicludes overlap.

In the case presented, the clustering was made from resistivity grids interpolated using kriging. Such interpolation tends to create smooth transitions between points of different

values. This can be particularly problematic in areas where the distance between soundings is large. Intermediate resistivity intervals are therefore produced when interpolating between low and high values. Such features are therefore a product of the interpolation and not the lithology. This can cause problems in the CF-resistivity cluster method as it is applied here when intermediate resistivity features caused by interpolation are linked to intermediate resistivity intervals related directly to lithology. For the present study such overlapping resistivity intervals appeared between deposits located near terrain and interpolation artefacts generating shell-like features around the buried valley structures. To eliminate the effect it had on the resulting model, we performed the clustering separately on the deposits near terrain and the ones located at depth.

One of the main advantages of the CF-resistivity clustering methodology is the limited time it takes to make a structural model directly from data. Time consumption for setting up structural models for regional scale groundwater models can thereby be reduced to days compared to manual interpretations that take months of preparation. This also facilitates making fast model updates in the case where new data acquisition has been made as well as it facilitates making a direct value of data assessment by comparing model outcomes.

6 Conclusion

In this paper we have shown how the CF-resistivity clustering methodology (Foged et al. 2014; Marker et al. 2015) can be applied to generate groundwater model structure in an area covered by a dense geophysical dataset collected over 30 years using multiple instruments. This is possible and straightforward due to the Danish database systems containing geological and hydrological data, and inverted geophysical models. Hopefully, this study will act as inspiration to other countries to develop similar systems, such that data can come to good use for groundwater management years after collection.

Using the presented method, we could set up the hydrostratigraphy on a regional scale (~470 km²) within days, where a manually built model covering the same area would require months. Such time saving on defining the regional model structure frees resources that can be focused on local details around areas of primary interest. We exemplified this by combining the regional scale groundwater model with two embedded local scale models. These local models were based on a detailed geological model developed in a previous study (Høyer et al. 2015).

Regional scale groundwater model simulations obtained with the generated cluster structure were compared to simulations made by a model using a manually interpreted regional scale structure. We found the simulation results of these two models to be comparable. For two observation subcategories, the performance of the cluster model clearly exceeded that of the reference model. This was particularly true in the subarea containing the newly acquired SkyTEM dataset. This illustrates the value of the densely covering geophysical data compared to the more sparsely collected data in the remaining area. The cluster model also outperformed the reference model in simulation of hydraulic heads at wells located in the buried valley structures.

The study illustrates the need for a regional scale groundwater model, in areas that contain aquifers in buried valleys subject to large groundwater abstractions. Because aquifer responses transmit quickly through the valley systems, the outer model boundary conditions must be placed carefully and at large distances. Using MODFLOW-USG makes it straightforward to

update the regional model by embedding local detailed models. The regional model can thereby be used as basis for multiple local scale management scenarios.

Acknowledgements This study was financed by the HyGEM project, project no. 11-116763. The funding is provided by Innovation Fund Denmark.

Compliance with Ethical Standards

Conflict of Interest None

References

- Arnold J, Allen P (1999) Automated methods for estimating baseflow and ground water recharge from streamflow records. *J Am Water Resour Assoc* 35(2):411–424
- Artimo A, Saraperä S, Ylander I (2008) Methods for integrating an extensive geodatabase with 3D modeling and data management tools for the Virttaankangas artificial recharge project, southwestern Finland. *Water Resour Manag* 22(12):1723–1739
- Aster RC, Borchers B, Thurber CH (2005) *Parameter estimation and inverse problems*. Elsevier Academic Press, Amsterdam
- Atkinson L, Ross M, Stumpf A (2014) Three-dimensional hydrofacies assemblages in ice-contact/proximal sediments forming a heterogeneous 'hybrid' hydrostratigraphic unit in Central Illinois, USA. *Hydrogeol J* 22(7):1605–1624
- Auken E, Christiansen AV, Jacobsen BH, Foged N, Sørensen KI (2005) Piecewise 1D laterally constrained inversion of resistivity data. *Geophys Prospect* 53:497–506
- Christiansen AV, Auken E, Sørensen KI (2006) The transient electromagnetic method. In: Kirsch R (ed) *Groundwater geophysics. A tool for hydrogeology*. Springer, New York, pp 179–224
- Danielsen JE, Auken E, Jørgensen F, Søndergaard VH, Sørensen KI (2003) The application of the transient electromagnetic method in hydrogeophysical surveys. *J Appl Geophys* 53(4):181–198
- DHI (2009) MIKE SHE user manual, volume 1: user guide. DHI, Hørsholm
- Doherty J (2015) *Calibration and uncertainty analysis for complex environmental models*. Watermark Numerical Computing, Brisbane
- Doherty J (2016) *PEST, Model-Independent Parameter Estimation. User Manual Part I: PEST, SENSAN and Global Optimisers*, Brisbane
- Foged N, Auken E, Christiansen AV, Sørensen KI (2013) Test site calibration and validation of airborne and ground based TEM systems. *Geophysics* 78(2):E95–E106
- Foged N, Marker PA, Christiansen AV, Bauer-Gottwein P, Jørgensen F, Høyer A-S, Auken E (2014) Large scale 3D-modeling by integration of resistivity models and borehole data through inversion. *Hydrol Earth Syst Sci* 18:4349–4362
- Harbaugh AW, Banta ER, Hill MC, McDonald MG (2000) *MODFLOW-2000, The U.S. Geological Survey Modular Ground-Water Model - User Guide to Modularization Concepts and the Ground-Water Flow Process*, Reston, Virginia
- He X, Koch J, Sonnenborg T, Jørgensen F, Schamper C, Refsgaard J (2014) Transition probabilitybased stochastic geological modeling using airborne geophysical data and borehole data. *Water Resour Res* 50: 3147–3169
- Henriksen H, Trolborg L, Nyegaard P, Sonnenborg T, Refsgaard J, Madsen B (2003) Methodology for construction, calibration and validation of a national hydrological model for Denmark. *J Hydrol* 280(1–4): 52–71
- Herzog B, Larson D, Abert C, Wilson S, Roadcap G (2003) Hydrostratigraphic modeling of a complex, glacial-drift aquifer system for importation into MODFLOW. *Ground Water* 41(1):57–65
- Høyer A-S, Jørgensen F, Sandersen PBE, Viezzoli A, Møller I (2015) 3D geological modelling of a complex Buried-Valley network delineated from borehole and AEM data. *J Appl Geophys* 2015(122):94–102
- Jain A (2010) Data clustering: 50 years beyond K-means. *Pattern Recogn Lett* 31(8):651–666
- Jørgensen F, Sandersen PBE, Auken E (2003) Imaging buried quaternary valleys using the transient electromagnetic method. *J Appl Geophys* 53:199–213

- Jørgensen F, Møller RR, Nebel L, Jensen N, Christiansen AV, Sandersen P (2013) A method for cognitive 3D geological voxel modelling of AEM data. *Bull Eng Geol Environ* 72(3–4):421–432
- Kessler H, Mathers S, Sobisch H (2009) The capture and dissemination of integrated 3D geospatial knowledge at the British Geological Survey using GSI3D software and methodology. *Comput Geosci* 35(6):1311–1321
- Kronborg C, Bender H, Bjerre R, Friberg R, Jacobsen HO, Kristiansen L, Rasmussen P, Sørensen PR, Larsen G (1990) Glacial stratigraphy of east and Central Jutland. *Boreas* 19:273–287
- Marker PA, Foged N, He X, Christiansen AV, Refsgaard A, Auken E, Bauer-Gottwein P (2015) Performance evaluation of groundwater model hydrostratigraphy from airborne electromagnetic data and lithological borehole logs. *HESS* 19:3875–3890
- Mazac O, Kelly WE, Landa I (1985) A hydrogeophysical model for relations between electrical and hydraulic properties of aquifers. *J Hydrol* 79(1–2):1–19
- Møller I, Verner H, Søndergaard VH, Flemming J, Auken E, Christiansen AV (2009) Integrated management and utilization of hydrogeophysical data on a national scale. *Near Surf Geophys* 7(5–6):647–659
- Panday S, Langevine CD, Niswonger RG, Ibaraki M, Hughes JD (2015) MODFLOW-USG version 1.3.00: An unstructured grid version of MODFLOW for simulating groundwater flow and tightly coupled processes using a control volume finite-difference formulation: U.S. Geological Survey Software Release, 01 December 2015
- Ross M, Parent M, Lefebvre R (2005) 3D geologic framework models for regional hydrogeology and land-use management: a case study from a quaternary basin of southwestern Quebec, Canada. *Hydrogeol J* 13(5–6):690–707
- Sandersen PBE, Jørgensen F (2003) Buried quaternary valleys in western Denmark—occurrence and inferred implications for groundwater resources and vulnerability. *J Appl Geophys* 53(4):229–248
- Søndergaard V, Thomsen R, Dyrso O, Nyholm T, Fuglsang E, Thorling L, Misser PV, Hansen B (2004) Redegørelse for grundvandsressourcerne i Århus-Nord området. Vandforsyning - Delrapport 1 (In danish), Aarhus Amt
- Sonnenborg T, Christensen B, Nyegaard P, Henriksen H, Refsgaard J (2003) Transient modeling of regional groundwater flow using parameter estimates from steady-state automatic calibration. *J Hydrol* 273(1–4):188–204
- Sørensen KI (1996) Pulled Array Continuous Electrical Profiling. *First Break*, 14:85–90
- Sørensen K, Auken E (2004) SkyTEM – a new high-resolution helicopter transient electromagnetic system. *Explor Geophys* 35:191–199
- Sørensen K, Auken E, Thomsen P (2000) TDEM in groundwater mapping - a continuous approach. In: *Proceedings of the symposium on the application of geophysics to engineering and environmental problems*, 2000, pp 485–491
- van der Kamp G, Maathuis H (2012) The unusual and large drawdown response of Buried-Valley aquifers to pumping. *Ground Water* 50(2):207–215
- Viezzoli A, Auken E, Munday T (2009) Spatially constrained inversion for quasi 3D modelling of airborne electromagnetic data - an application for environmental assessment in the lower Murray region of South Australia. *Explor Geophys* 40:173–183
- Vignoli G, Fiandaca G, Christiansen AV, Kirkegaard C, Auken E (2015) Sharp spatially constrained inversion with applications to transient electromagnetic data. *Geophys Prospect* 63(1):243–255
- Wu J (2012) *Advances in K-means clustering: a data mining thinking*. Springer, New York