Hydrol. Earth Syst. Sci. Discuss., 12, 9599–9653, 2015
www.hydrol-earth-syst-sci-discuss.net/12/9599/2015/
doi:10.5194/hessd-12-9599-2015

Hydrology and Earth System Sciences
Open Access
Discussions

A framework for testing the use of electric and electromagnetic data to reduce the prediction error of groundwater models

N. K. Christensen[1], **S. Christensen**[1], **and T. P. A. Ferre**[2]

[1]Department of Geoscience, Aarhus University, Aarhus, Denmark
[2]Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona, USA

Received: 18 August 2015 – Accepted: 4 September 2015 – Published: 24 September 2015

Correspondence to: N. K. Christensen (phda.nikolaj.kruse@geo.au.dk)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Despite geophysics is being used increasingly, it is still unclear how and when the integration of geophysical data improves the construction and predictive capability of groundwater models. Therefore, this paper presents a newly developed **HY**drogeophysical **TE**st-**B**ench (HYTEB) which is a collection of geological, groundwater and geophysical modeling and inversion software wrapped to make a platform for generation and consideration of multi-modal data for objective hydrologic analysis. It is intentionally flexible to allow for simple or sophisticated treatments of geophysical responses, hydrologic processes, parameterization, and inversion approaches. It can also be used to discover potential errors that can be introduced through petrophysical models and approaches to correlating geophysical and hydrologic parameters. With HYTEB we study alternative uses of electromagnetic (EM) data for groundwater modeling in a hydrogeological environment consisting of various types of glacial deposits with typical hydraulic conductivities and electrical resistivities covering impermeable bedrock with low resistivity. It is investigated to what extent groundwater model calibration and, often more importantly, model predictions can be improved by including in the calibration process electrical resistivity estimates obtained from TEM data. In all calibration cases, the hydraulic conductivity field is highly parameterized and the estimation is stabilized by regularization. For purely hydrologic inversion (HI, only using hydrologic data) we used Tikhonov regularization combined with singular value decomposition. For joint hydrogeophysical inversion (JHI) and sequential hydrogeophysical inversion (SHI) the resistivity estimates from TEM are used together with a petrophysical relationship to formulate the regularization term. In all cases, the regularization stabilizes the inversion, but neither the HI nor the JHI objective function could be minimized uniquely. SHI or JHI with regularization based on the use of TEM data produced estimated hydraulic conductivity fields that bear more resemblance to the reference fields than when using HI with Tikhonov regularization. However, for the studied system the resistivities estimated by SHI or JHI must be used with caution as estima-

HESSD
12, 9599–9653, 2015

tors of hydraulic conductivity or as regularization means for subsequent hydrological inversion. Much of the lack of value of the geophysical data arises from a mistaken faith in the power of the petrophysical model in combination with geophysical data of low sensitivity, thereby propagating geophysical estimation errors into the hydrologic model parameters. With respect to reducing model prediction error, it depends on the type of prediction whether it has value to include geophysical data in the model calibration. It is found that all calibrated models are good predictors of hydraulic head. When the stress situation is changed from that of the hydrologic calibration data, then all models make biased predictions of head change. All calibrated models turn out to be a very poor predictor of the pumping well's recharge area and groundwater age. The reason for this is that distributed recharge is parameterized as depending on estimated hydraulic conductivity of the upper model layer which tends to be underestimated. Another important insight from the HYTEB analysis is thus that either recharge should be parameterized and estimated in a different way, or other types of data should be added to better constrain the recharge estimates.

1 Introduction

1.1 Using hydrologic models for decision support

Groundwater models are commonly constructed to support decision-makers in managing groundwater resources. The model can, for example, be used to predict the ramifications of changes in groundwater pumping on hydraulic head and wellhead protection areas or to predict the fate and transport of groundwater pollution. In general terms, process models are used to base predictions of interest on all of the knowledge that we have about the physical/chemical system and the driving key processes.

A groundwater model is based on a conceptual model that encapsulates prior knowledge of important physical and chemical conditions and processes of the complex real world system. The conceptual model is translated into a numerical groundwater model

HESSD
12, 9599–9653, 2015

whereby its reasonableness can be tested by comparing model simulations with field observations. If the conceptual model appears reasonable, the groundwater model is calibrated by adjusting model parameters until simulated values fit corresponding field observations sufficiently well. The calibrated model is subsequently used to make predictions (Reilly, 2001; Reilly and Harbaugh, 2004). However, the model will be wrong and the predictions uncertain for a number of reasons. (i) Model calibration is done by fitting uncertain data. The calibrated parameters will therefore also be uncertain and this uncertainty is propagated to the model predictions (Hill, 1999; Moore and Doherty, 2006; Tonkin et al., 2007). A model's predictive uncertainty will only be reduced by calibration if the information content of the calibration dataset constrains the parameter values that significantly influence the prediction (Harvey and Gorelick, 1995; Feyen et al., 2003; Franssen et al., 2003). Thus this source of uncertainty can only be reduced by collecting more or more accurate data of type(s) and location(s) that constrain parameter values important to the prediction. The data will typically be hydrologic or hydraulic, but it can also be geophysical. (ii) Because of scarcity and lack of sensitivity of data, there will always be small scale heterogeneity that cannot be resolved. A groundwater model will therefore always contain small scale structural errors, which may not cause bias in predictions but may still cause large prediction uncertainty (Cooley, 2004; Cooley and Christensen, 2006; Refsgaard et al., 2012). (iii) A model is also prone to possess large-scale structural errors that can cause significant bias and uncertainty of estimated parameters and simulated predictions (Doherty and Welter, 2010; Doherty and Christensen, 2011; Refsgaard et al., 2012). This bias and uncertainty can be reduced by collecting data that resolve the large-scale structures of the studied hydrogeological system, which can then be accurately represented in the model. This can, for example, be spatially dense geophysical data sets.

Model errors will lead to errors and uncertainties in predictions of interest. One of the key questions to address in creating models for decision support is: which additional data are most likely to improve key predictions? The types of data available for use in hydrologic analysis are increasingly diverse, including physical, chemical, isotopic, and

HESSD
12, 9599–9653, 2015

geophysical data. In light of this complexity, it can be very difficult to compare the likely contributions of diverse data to model-based decision support.

1.2 Informing hydrologic models with geophysics

Over the last three decades, noninvasive geophysical methods have been used increasingly to construct groundwater models (Hubbard and Rubin, 2000; Vereecken et al., 2004). This is particularly true for data collected by the Airborne Electromagnetic Method (AEM) because they can be collected quickly, densely, and at a relatively low cost for the very large spatial coverage (Viezzoli et al., 2010b; Faneca Sànchez et al., 2012; Refsgaard et al., 2014). Large-scale AEM (or ground-based EM) investigations have been used to delineate aquifers, aquitards, and buried valleys or other structures containing aquifers (Auken et al., 2003; Sandersen and Jørgensen, 2003; Jørgensen et al., 2003; Seifert et al., 2007; Andersen et al., 2012), to assess aquifer vulnerability (Refsgaard et al., 2014; Foged et al., 2014), to map saltwater intrusion (Fitterman and Deszcz-Pan, 1998; Viezzoli et al., 2010b; Herckenrath et al., 2013b), and to map freshwater resources (Steuer et al., 2008; Faneca Sànchez et al., 2012). The main drawbacks of electromagnetic (EM) data are: (1) ambiguity in relating electrical properties to hydraulic properties; and (2) reduced lateral and vertical resolution with depth. The former effect can limit the quantitative use of geophysical data for parameterizing groundwater models. The latter effect makes identification of deep structures difficult (Danielsen et al., 2003; Auken et al., 2008), which will have different influences on predictions that are dominated by shallower or deeper flow paths.

Geophysical data must be related to properties or states of hydrologic relevance to use them in constructing hydrologic models. Whether the data are used to define hydrostratigraphic units or subregions or to parameterize the model, geophysical inversion is required. The way in which hydrologic and geophysical data are inverted and integrated can impact the extraction of information from geophysical data (Dam and Christensen, 2003; Day-Lewis, 2005; Hinnell et al., 2010).

HESSD
12, 9599–9653, 2015

The simplest approach to incorporating geophysical data is through sequential hydrogeophysical inversion (SHI). In this approach, the geophysical data are inverted independent of the hydrologic data or model. The inverted geophysical properties are then used to zonate or directly parameterize the hydrologic model (Hubbard et al., 1999; Seifert et al., 2007; Koch et al., 2009; Di Maio et al., 2013; Marker et al., 2015). The geophysical responses are sensitive to some of the same structures and property distributions that the hydrologic data are sensitive to. This means that potentially the hydrologic data can contain information about the geophysical system. SHI has the drawback that it does not allow the hydrologic data to play back and influence on the estimation of the geophysical model. Furthermore, with SHI it is difficult to quantify the uncertainty of groundwater model predictions because different assumptions, smoothing, and inversion approaches may have been used to invert the hydrological and geophysical data (Doherty et al., 2010; Menke, 2012).

Two alternatives to SHI that extract more information from the data sets are coupled hydrogeophysical inversion (CHI) and joint hydrogeophysical inversion (JHI) (Hinnell et al., 2010). For both alternatives, the hydrologic and geophysical data sets are inverted simultaneously. In CHI, the simulated response of one model (e.g. the hydrologic model) is used as input to constrain the other model (e.g. the geophysical model). (For example, during the inversion a water table simulated by the hydrologic model is used to constrain the depth of a layer boundary of the estimated geophysical model.) CHI has been applied successfully for reducing parameter uncertainty by using ground penetrating radar and electrical resistivity tomography data in hydraulic models (Kowalsky et al., 2005; Hinnell et al., 2010). In JHI, the hydrologic and geophysical models are coupled directly through some of their parameters using assumed relationships among the geophysical and/or hydrologic parameters (Hyndman et al., 1994). For EM data, JHI is typically done using a relationship between hydraulic conductivity and electrical resistivity inspired by Archie's law (Archie, 1942; Revil and Cathles, 1999; Purvance and Andricevic, 2000; Slater, 2007). Application of JHI for simultaneous inversion of

HESSD
12, 9599–9653, 2015

hydrologic and geophysical data has been demonstrated by Linde et al. (2006), Herckenrath et al. (2013a) and Vilhelmsen et al. (2014).

It is intuitively clear that geophysics can offer valuable information for improved groundwater modeling for decision making. However, many important questions are yet unanswered. For example: for a complex hydrogeological system what type(s) of data will be most valuable to collect, and how should they be collected; how does the value of geophysical data depend on data quality; how much can be gained by using CHI or JHI instead of SHI; can some or all inversion approaches lead to biased parameter estimates or model predictions, and under what circumstances; and how well should a petrophysical relationship be known to do JHI? Many if not all of these questions will depend on the actual hydrogeological setting as well as on what types of prediction are going to be made by the groundwater model. Furthermore, all sources of uncertainty (inversion artifacts, measurement density, measurement uncertainty, uncertainty in petrophysical relationships, etc.) may interact in different ways for different hydrogeologic settings and for different predictions of interest.

1.3 Hydrogeophysical test-bench

As discussed above, the types of data available for use in hydrologic analysis are increasingly diverse in type, accuracy, and resolution. This is not least caused by the development of new geophysical instruments and methods. The worth of various types of geophysical data to hydrologic analysis will be case specific; it will not only depend on the hydrogeologic system under study and the type, location and accuracy of the geophysical data, but also on the types of predictions to be made by the groundwater model. Before the geophysical data are actually collected in a specific investigation it is therefore important to objectively examine how much they can be expected to reduce groundwater model prediction error and uncertainty and how they can best be used for this purpose. This examination is not straight forward because it requires both hydrogeologic and geophysical understanding and competences.

HESSD
12, 9599–9653, 2015

To allow a thorough examination we have developed a cross-disciplinary, flexible framework for making experiments to objectively examine the worth of geophysical data for improvement of groundwater model predictions in potentially complex environments. The idea is to build synthetic experiments that have similarity with the actual hydrogeological and geophysical systems to be investigated, the types of data to potentially be collected, and the types of models to potentially be used. The flexibility of the framework allows easy investigation of the data worth when using alternative data sampling and alternative modeling or inversion strategies. Because of the similarity between the synthetic and the actual systems, the conclusions from the synthetic study can be transferred to actual investigation. The framework is called HYTEB, which is an abbreviation of **HY**drogeophysical **TE**st-**B**ench. The novelty of HYTEB is that it builds on a merge of software from different disciplines such as stochastic hydrogeological modeling, groundwater modeling, geophysical modeling, and advanced highly parameterized inversion using SHI, CHI or JHI.

1.4 Objectives

The paper has the following objectives. First, it will present the important elements and steps in use of HyTEB. Since HyTEB and its use is interdisciplinary, the presentation and the following case study introduce geophysicists to the methods, challenges, and purposes of groundwater modeling, and groundwater modelers to some of the challenges of using mainly electric and electromagnetic data for groundwater model calibration purposes. Second, HYTEB is used to examine the worth of adding a ground based time-domain electromagnetic data set to a hydrological data set when making a groundwater model for a glacial landscape of a kind that is typical to parts of Northern Europe and North America. It is investigated if the worth of adding the geophysical data depends on the type of groundwater model prediction as well as on whether the geophysical and hydrological data are inverted sequentially or jointly. Section 2 of this paper describes the elements of HYTEB and how they are used, Sect. 3 describes the

HESSD
12, 9599–9653, 2015

case study, Sect. 4 presents the results, while Sect. 5 makes a summary and draw conclusions.

2 The elements and concept of HYTEB (HYdrogeophysical TEst-Bench)

Our primary objective in developing HYTEB is to provide a synthetic environment that allows users to determine the value of geophysical data and, further, to investigate how best to use those data to develop groundwater models and to reduce their prediction errors. We suggest that this can best be investigated by using a synthetic case study for which the "generated synthetic", in the following termed "reference", hydrologic and geophysical systemsare known and the influences of different sources of error can be investigated. We use physical and geophysical response models to generate measurements that would be collected from the reference systems in the absence of noise. We then examine the influence of measurement error and other sources of error on model predictions of interest. By repeating this for different synthetic system realizations (i.e. for different reference systems) and for different data sets it becomes possible to statistically quantify the worth of the various data for improving the predictions of interest. The work flow of HYTEB is shown in Fig. 1. The procedure is divided into 6 steps, which will be described separately and briefly in the following subsections.

2.1 Step 1 – generation of geological realization

The first step is to generate a synthetic realization of the type of geological system under study. The generation can be made conditional on lithological data from boreholes. The borehole data can be imaginary, a real data set, or a combination of data, hydrogeologic structure, and geostatistics. Figure 1, step 1, displays an example of a generated system consisting of categorical geological deposits on a plain as well as in a valley buried under a part of the plain. The deposits are underlain by impermeable bed rock (not shown). Such categorical geological settings can, for example, be

HESSD
12, 9599–9653, 2015

generated using T-PROGS (Carle, 1999) or BlockSIS (Deutsch, 2006). The spatial discretization used for the geological realization also defines the spatial discretization of the numerical model used to simulate groundwater flow or any other process model that a user decides to integrate into HYTEB.

2.2 Step 2 – generation of groundwater system, hydrological data set, and predictions

Using the same spatial discretization as in step 1, the second step is to define the boundary conditions and the hydraulic and solute transport property values for the generated geological system. The hydraulic and solute transport properties can include, for example, hydraulic conductivity, specific storage, and effective porosity. For categorical deposits (as in Fig. 1) the value of each type of property will typically vary among categories as well as within each category. Such variation can be simulated as categorical random fields by using e.g. SGSIM (Deutsch and Journel, 1998) or FIELDGEN (Doherty, 2010). The generated realization of boundary and property values is used in a numerical simulator of groundwater flow and solute transport to simulate a set of state variables to be used in step 5 as hydrologic observations used for model calibration; random error is typically added to this observation data to represent all sources of noise that corrupt real observations. The numerical simulator is also used to simulate a set of predictions that are considered of particular interest to the study. We have implemented MODFLOW-2000 (Harbaugh et al., 2000) as the numerical simulator of groundwater flow and MODPATH (Pollock, 1994) to simulate solute transport by particle tracking.

In the following, the numerical simulators using the boundary conditions and property values that represent the system realization are called “the reference groundwater system” and the predictions simulated for this system are called “reference prediction”.

HESSD
12, 9599–9653, 2015

2.3 Step 3 – generation of geophysical system and geophysical data set

The third step is to define the property values of the geophysical system corresponding to the geological realization generated in step 1. Like the hydraulic properties, the geophysical properties can be considered and simulated as categorical random fields. A geophysical property of relevance can, for example, be the electrical resistivity of the spatially variable geological deposits. For some geological systems, it is found or assumed that there is correlation between electrical resistivity and hydraulic conductivity. In this case, the hydraulic and geophysical property fields must be generated to be dependent. Various empirical petrophysical relationships between hydraulic conductivity and electrical resistivity have been proposed (Slater, 2007). It is common to use a linear log-log relationship which is given some theoretical support by Purvance and Andricevic (2000). Having defined the property values of the geophysical reference system, the geophysical instrument responses are simulated to produce a noise-free geophysical data set that can be corrupted by adding random error to represent all sources of measurement error. Ideally a 3-D code should be used. 3-D computation of electrical responses can be efficiently modelled (Günther et al., 2006; Rücker et al., 2006). Codes for 3-D computation of TEM responses have also been developed (e.g. Árnason, 1999), but the computation is impractical and burdensome. As a practical alternative we suggest to simulate TEM responses by a 1-D code, where the 1-D geophysical model is created from the reference model by pseudo-3-D sampling, that is by taking the logarithmic average of the cells within the radius of the EM foot print. Modeling TEM in 1-D can be problematic in connection with mineral exploration, but for sedimentary environments a 1-D approach works well (Auken et al., 2008; Viezzoli et al., 2010a). In HYTEB we use AarhusInv (Auken et al., 2014) to simulate electrical and electromagnetic instrument responses.

In the following, the geophysical simulator using the actual realization of geophysical parameter values is called the “reference geophysical system”.

HESSD
12, 9599–9653, 2015

2.4 Step 4 – make and parameterize models

In this step, the synthetic data are used to constrain parameter estimation for a groundwater model of the reference groundwater system. Each property of the real groundwater and geophysical systems needs to be parameterized in the groundwater model. This step thus corresponds to the construction of a groundwater model of a real field system on the basis of the available real data. In the synthetic case, the groundwater model can be discretized exactly as the "reference groundwater system" or it can use a coarser discretization. Here we adopt the former alternative to reduce numerical discretization error. However, this effect could be examined if it were of interest to a particular study.

In studies of real systems, the groundwater model is often constructed to consist of zones of uniform hydraulic properties. The subdivision into zones is typically done subjectively by an expert on the basis of geological, hydrological, and geophysical data (Seifert et al., 2007; Di Maio et al., 2013). This principle can also be used to define zones of a model of the synthetic groundwater system by using the synthetic lithological data from boreholes used in step 1, the hydrological data set generated in step 2, and geophysical models estimated by inverting the geophysical data sets generated in step 3. In this case, the geophysical data must be inverted between step 3 and step 4. The inverted data are used either in step 4 to support parameterization of the groundwater model or in step 5 for groundwater model calibration. To avoid over-reliance on the geophysical data, they should not be used in both steps 4 and 5. If the geophysical data are used in step 4, they must be inverted before inverting the hydrological data (carried out in step 5); this is an example of sequential hydro-geophysical inversion (SHI).

An alternative parameterization approach uses the concept of pilot points (Certes and de Marsily, 1991) to parameterize the property fields and to let the data determine the variation of the model property fields (e.g. Doherty, 2003). Pilot point approaches result in a smooth property variation within the model domain (Doherty, 2003) rather

HESSD
12, 9599–9653, 2015

than sharp zonal parameter fields. Pilot points can be used in combination with zones e.g. to represent property variation within categorical deposits.

HYTEB allows any type of parameterization, zones, pilot points, or combinations hereof. In the following demonstration we chose pilot points.

It is emphasized that in the following we use the term "groundwater model" for a simulator that is set up, parameterized, and calibrated to make "model predictions" of states occurring in the real groundwater system. States occurring in (i.e. simulated for) the reference groundwater system are here termed "reference predictions". The objective of model calibration is to make the model predictions as similar as possible to the reference predictions.

2.5 Step 5 – calibrate the model(s)

The fifth step is to calibrate the groundwater model by using the data set produced in step 2 to estimate the model parameters. The step may also include estimation of geophysical model parameters on the basis of the data sets produced in step 3. The simultaneous estimation of the hydrologic and geophysical parameters can be done by using either the coupled (CHI) or joint (JHI) hydro-geophysical inversion approaches (Hinnell et al., 2010; Vilhelmsen et al., 2014). When the number of parameters is large compared to the number of data, the minimization can be aided by using a regularization technique (for example singular value decomposition or Tikhonov regularization); see Oliver et al. (2008) for an overview. For this purpose, and for JHI, we use PEST or BeoPEST (Doherty, 2010). An advantage of CHI and JHI is that by inverting the hydrologic and geophysical models simultaneously, they are subject to the same regularization effects and all of the data are fitted simultaneously by both model types.

2.6 Step 6 – simulate model predictions, then repeat steps 1–6

After successful calibration, the groundwater model is used to make model predictions equivalent to the reference predictions as in step 2. For each prediction, this produces

HESSD
12, 9599–9653, 2015

one value computed by a calibrated model that can be compared with the equivalent reference value. It is not possible to make meaningful inference about a model's ability to make a specific prediction from just one experiment. To test the reproducibility the experiment made through, steps 1 to 6 needs to be repeated a number of times. Each repetition involves generation of a new realization of the geological system and the corresponding groundwater and geophysical systems, new data sets (i.e. new reference systems), model calibration, and predictions. The number of repetitions should be sufficient to provide a basis for making consistent statistical inference on the model prediction results.

2.7 Step 7 – evaluate model prediction results

When steps 1 to 6 have been completed, an ensemble of pairs of model prediction and equivalent reference prediction are plotted to evaluate the model performance. As discussed by Doherty and Christensen (2011), if the plotted data do not scatter around the identity line, it indicates bias in the model prediction. If the intercept of a regression line through the scatter of points deviates from zero it indicates consistent bias in the prediction due to consistent errors in null space parameter components omitted from the parameterized groundwater model; if the slope of the regression line deviates from unity it indicates parameter surrogacy incurred through model calibration (see Doherty and Christensen (2011) for further explanation).

Ultimately, calibrated models are used to make predictions of interest. These predictions are generally in the future and may describe the response of the system to alternative management actions. The calibrated model, or model ensemble, can be used to predict future hydrologic responses to near-term actions, thereby providing information critical to informed decision making. Increasingly, these decisions consider both the accuracy (bias) and the uncertainty of model predictions in a probabilistic framework (Freeze et al., 1990; Feyen and Gorelick, 2005; Nowak et al., 2012).

HESSD
12, 9599–9653, 2015

3 Demonstration model

We demonstrate the use of HYTEB through a synthetic case focusing on making three types of model predictions that are commonly useful for groundwater management: (i) hydraulic head; (ii) head recovery and change of groundwater discharge related to abandoning pumping from a well; and (iii) the recharge area and the average age of groundwater pumped from that well. The synthetic demonstration model used here is, to a large degree, inspired by the model of Doherty and Christensen (2011). The hydrogeological setting of the model domain is typical for large areas of northern Europe and North America: a glacially formed landscape with a buried tunnel valley eroded into impermeable bed rock with very low electrical resistivity. The deposits above the bedrock are glacial of different types. For the sake of clarity, the synthetic model will be described in the section below, and the exceptions and changes from the setup of Doherty and Christensen (2011) will be highlighted. Each HYTEB step will be presented in order following Fig. 1.

3.1 Generation of geological system realizations (step 1)

The domain is rectangular, 7 km north–south (N–S) and 5 km east-west (E-W). It is capped by 50 m of glacial sediments deposited as gently N–S elongated layered structures composed of sand, silt or clayey till. The bedrock consists of impermeable clay with a horizontal top surface in most of the catchment, but a 150 m deep and 1500 m wide valley has been eroded into it in the central part of the domain (Doherty and Christensen (2011) used a 1000 m wide valley). The valley has sloping sides with an angle of approximately 17° and runs in the N–S direction from the coast and 3.5 km inland (Doherty and Christensen (2011) used a steeper 21° slope). The valley is filled with glacial sediments deposited in highly N–S elongated layered structures consisting of gravel, sand, silt or clayey till. The exact stratigraphy is only known at the locations of 35 synthetic boreholes of varying depth (Fig. 2). This borehole stratigraphy was used to condition all generated geological system realizations.

HESSD
12, 9599–9653, 2015

Realizations of the 3-D geological model were generated on a uniform rectangular grid. The cells of the grid have horizontal dimensions of 25 m × 25 m and 10 m thickness, so the overall dimensions of the grid are $(n_x, n_y, n_z) = (200, 280, 20)$, giving a total of 1 200 000 cells. The categorical depositional geology of the 3-D model grid was simulated using T-PROGS (Carle, 1999). The proportions and mean lengths for the different categories of sediments are provided in Table 1. The bedding is represented as a maximally disordered system using "maximum entropy" transition frequencies (Carle, 1999).

A total of 1000 geologic system realizations were generated. These categorical realizations were all conditioned on the same stratigraphy for the 35 boreholes, but are otherwise independent. Figure 3 shows one of these realizations.

3.2 Groundwater system, data, and predictions (step 2)

The groundwater system is bounded to the south by a large freshwater lake (specified head), while the other lateral boundaries are closed (no flux). The flow is steady state and driven by recharge caused by the difference between precipitation and evapotranspiration. The local recharge depends on the type of sediment at the surface (because this is assumed to influence evapotranspiration). Most of the groundwater discharges into the lake directly from the subsurface, but approximately 35 % discharges into a straight stream running 3.5 km inland S-N in the middle of the domain from the southern boundary (coast). (The setup used by Doherty and Christensen (2011) did not include a stream.) Furthermore, groundwater is pumped from a deep well located in the south-central part of the buried valley. The well is located at $x = 2487.5$ m and $y = 1912.5$ and the pumping rate is $0.015\,m^3\,s^{-1}$. The well screens the deepest 10 m of the valley in a laterally extensive body of sand and gravel.

Within each category of sediment, the hydraulic conductivity varies as a horizontally correlated random field. The same is the case for porosity and recharge. The random fields were generated by Fieldgen (Doherty, 2010) using the sequential Gaussian sim-

HESSD
12, 9599–9653, 2015

ulation method (Deutsch and Journel, 1998) with the geostatistical parameters given in Table 1.

3.2.1 Hydrological data set

All 35 boreholes have been constructed as monitoring wells; each well screens the deepest 10 m (deepest cell) of sand registered in the borehole (Table 2; Fig. 2). For each realization, groundwater flow was simulated as confined using MODFLOW-2000 (Harbaugh et al., 2000). The corresponding set of values for the hydrological observations, consisting of hydraulic head in the 35 wells and the river discharge, were extracted from the model output. Independent Gaussian error with zero mean and 0.1 m standard deviation was added to the true head values to produce the head observations. Gaussian error with zero mean and a standard deviation corresponding to 10 % of the true river discharge was added to the discharge to produce the stream flow observation used for model calibration.

3.2.2 Predictions

Collecting and using new geophysical data is likely to constrain some groundwater model parameters more than others. Different predictions of interest will have different sensitivities to different model parameters. As a result, the addition of geophysical data is likely to have different effects on the uncertainties of different predictions of interest. To illustrate this, we present six types of predictions of interest (Table 3).

Prediction types 1 to 3 relate to steady-state flow conditions with groundwater being pumped from the deep well in the buried valley. This is the same situation for which the hydrological dataset was generated. Type 1 concerns head prediction at ten locations (Fig. 2 and Table 4). Type 2 is the size of the recharge area of the pumping well. Type 3 is the average age of the groundwater pumped from the well.

Prediction types 4 to 7 relate to a new steady-state long after pumping from the well has been stopped. Type 4 is head recovery at the ten locations given in Fig. 2 and

HESSD
12, 9599–9653, 2015

Table 4. Type 5 is the travel time of a particle flowing with the groundwater from the location that it enters the system at the northern domain boundary ($x = 2500$, $y = 6975.5$, $z = 0$) until it exits the system either into the lake (at the southern boundary) or into the stream. Type 6 is the relative location of the exit point of that particle defined as the Euclidean distance between the reference and the model predicted endpoint in a three dimensional space. Type 7 is groundwater discharge into the stream.

The prediction types 1, 4 and 7 were simulated by MODFLOW-2000 (Harbaugh et al., 2000). The other prediction types were simulated by forward particle tracking using MODPATH version 5 (Pollock, 1994) and MODFLOW-2000 results. Types 5 and 6 were simulated by tracking a single particle with MODPATH. Types 2 and 3 were simulated by placing particles in a horizontally uniform 25 m grid at the surface (i.e. releasing one particle at the surface at the center of each model cell) and tracking them forward in time until they reached either the river, the southern boundary, or the pumping well. Each particle represents an area of $25 \times 25\,m^2$. The number of particles ending in the pumping well thus defines the well's recharge area. The average groundwater age is computed as the weighted average of the travel time for all of the particles captured by the well. The weight for a particle is calculated as the recharge rate (in $m^3\,s^{-1}$) from the 25×25 m^2 surface area represented by the particle divided by the pumping rate. This sum of all weights adds to one because water only enters the model through the uppermost layer.

3.3 Geophysical system and data – step 3

In the demonstration example, the geophysical system of interest is electrical resistivity of the subsurface. For simplicity it is assumed that there is a direct relationship between hydraulic conductivity and electrical resistivity. The relationship is of the form

$$\log_{10}(K) = \beta_1 + \beta_2 \cdot \log_{10}(\rho) + e \quad (1)$$

where K is the hydraulic conductivity ($m\,s^{-1}$), ρ is the electrical resistivity (Ωm), e is random Gaussian noise, and $\beta_1 = \log_{10}(1e^{-12})$ and $\beta_2 = \log_{10}(4)$ are empirical shape

HESSD
12, 9599–9653, 2015

factors that are constant within the model domain. The shape factor values reflect conditions where, for example, clay has low electrical resistivity and also low hydraulic conductivity, and sand has high electrical resistivity and high hydraulic conductivity. Equation (3) was used to compute the resistivity within each cell of the geological system from the corresponding cell hydraulic conductivity.

Using a direct relationship between hydraulic conductivity and resistivity must be characterized as the ideal case because electrical resistivity data can provide maximal information about hydraulic conductivity. When possible, estimation of hydraulic conductivity from electrical resistivity is usually based on a site specific linear log-log relationship (see e.g. Mazáč et al., 1985; Revil and Cathles, 1999; Purvance and Andricevic, 2000; Slater, 2007), which has been found to be a positive relationship in some cases (Urish, 1981; Frohlich and Kelly, 1985), and a negative relationship in other cases (Worthington, 1975; Heigold et al., 1979; Biella et al., 1983). (A more complicated, or less certain, relationship between electrical resistivity and hydraulic conductivity could also have been chosen for the demonstration; HYTEB is designed to have no such limitation.)

Geophysical data set

It is assumed that measurements of the geophysical system are conducted at 77 uniformly distributed locations within the domain (Fig. 2) using a ground based time domain electromagnetic system (TEM). It is assumed that the TEM system uses a receiver loop centered inside a $40 \times 40\,m^2$ square transmitter loop. Measurements are gathered from about 10 µs to 10 ms using a steady current of 20 A, which gives a magnetic moment of $32\,000\,Am^2$ which, for the studied environment, would provide a penetration depth of around 250 m (Danielsen et al. 2003). For this system the electromagnetic field is propagating down- and outwards like smoke rings increasing with depth at an angle of approximately 30° (West and Macnae, 1991). In other words, the sounding loses resolution with depth because of its increasing footprint. In the following, we use the 1-D simulation code AarhusInv (previously called em1-Dinv; Auken et

HESSD
12, 9599–9653, 2015

al., 2014) to simulate the geophysical responses. To mimic the loss of resolution with layer depth we use the logarithmic average resistivity of all model cells inside the radius of the foot print at a given depth. To obtain the geophysical data set, the simulated data were contaminated with noise according to the noise model suggested by (Auken et al., 2008):

$$V_{\text{resp}} = V \cdot \left(1 + G(0,1) \cdot \left[\text{STD}_{\text{uni}}^2 + \left(\frac{V_{\text{noise}}}{V}\right)^2\right]^{1/2}\right) \quad (2)$$

where V_{resp} is the perturbed synthetic data, V is the synthetic noiseless data, $G(0, 1)$ is standard Gaussian noise (with zero mean and unit standard deviation), and $\text{STD}_{\text{uni}}^2$ is uniform noise variance. V_{noise} is the background noise contribution given by

$$V_{\text{noise}} = b \cdot \left(\frac{t}{10^{-3}}\right)^{-1/2} \quad (3)$$

where t is the gate center time in seconds, and $b = 1\,\text{nV}\,\text{m}^{-2}$ is the noise level at 1 ms. Experience has shown that in many parts of the world this number ranges between 1 and $5\,\text{nV}\,\text{m}^{-2}$ when using a stack size of 1000 transients (Auken et al., 2008). The uniform standard deviation, which accounts for instrument and other non-specified noise contributions, is set to 3 % for dB/dt responses. After the data were perturbed with noise, it was processed as field data. This was done using an auto processing function that assumes that time domain electromagnetic fields are always decaying, sign shifts only happen in off-center configurations, and data with large uncertainty is removed because the perturbation caused them to noisy to be applied in the further analysis (Auken et al., 2009).

In general, this example demonstrates that HYTEB is designed to accommodate a sophisticated level of insight regarding geophysical data. This places a high demand on users, but it is a key element of the framework because it results in more realistic investigations of the potential benefits and limitations of geophysical data.

HESSD
12, 9599–9653, 2015

3.4 Model parameterization (step 4)

The groundwater model uses the true boundary conditions except that recharge is to be estimated together with hydraulic conductivity. Because the synthetic groundwater and geophysical systems are generated with correlation between hydraulic conductivity and electrical resistivity, the hydraulic conductivity is parameterized by placing pilot points in each of the 20 layers at the locations where a geophysical sounding has been made. However, pilot points are excluded at depths of the impermeable bedrock. The number of pilot points used for hydraulic conductivity therefore totals 550 (Fig. 2). Kriging is used for spatial interpolation (here using the correct correlation lengths) from the pilot points to the model grid.

Recharge is parameterized by assuming a linear log-log relationship between recharge and hydraulic conductivity of the uppermost layer. The two shape factors of the log-log relationship are chosen as parameters to be estimated; they are assumed to be constant within the model domain. The total number of parameters for estimating recharge from hydraulic conductivity is thus two.

Because porosity cannot be estimated from the hydrological and geophysical data available here, we always use the true porosity field for making model predictions. (The effects of porosity uncertainty, and determining the likely value of adding a geophysical method that could infer porosity, could have been included but is beyond the scope of this example application of HYTEB.) A geophysical model is set up for every location of the 77 TEM soundings. Each geophysical model is parameterized to have a fixed number of layers equal to one plus the number of groundwater model layers above bedrock; the layers above bedrock all have fixed 10 m thickness while the bedrock is assumed to be of infinite thickness. The estimated parameters of the model are the resistivity within each model layer. The total number of parameters for the 77 geophysical models is thus 627. The model responses were simulated using AarhusInv neglecting lateral heterogeneity. In other words, the inverse model is 1-D, following the state of practice (Viezzoli et al., 2010a; Auken et al., 2014)

HESSD
12, 9599–9653, 2015

3.5 Model calibration by inversion (step 5)

Traditionally, calibration of geophysical and groundwater models are conducted independently. However, for our demonstration problem, we want to explore the amount of "hydraulic" information contained within the geophysical dataset. We will do this by applying three different calibration methods.

3.5.1 Three calibration methods

Method 1 estimates groundwater model parameters on the basis of hydrologic data only (HI). This estimation involves constrained minimization of the misfit between model-simulated responses and the equivalent observation data. This misfit is quantified by the measurement objective function

$$\phi_m = n_h^{-1} \sum_{i=1}^{n_h} \left(\frac{h_{obs,i} - h_{sim,i}}{\sigma_{h,i}} \right)^2 + n_r^{-1} \sum_{i=1}^{n_r} \left(\frac{r_{obs,i} - r_{sim,i}}{\sigma_{r,i}} \right)^2, \quad (4)$$

where h_{obs} and h_{sim} are observed and corresponding simulated hydraulic heads; r_{obs} and r_{sim} are observed and corresponding simulated river discharge; σ_h and σ_r are the noise levels (standard deviations) for the head and discharge data, respectively. However, Eq. (4) cannot be minimized uniquely because the number of groundwater model parameters (552) is larger than the number of measurements (36). Method 1 therefore relies on minimization of the regularized objective function

$$\phi_t = \phi_m + \mu \cdot \phi_r \quad (5)$$

where ϕ_t is the total objective function, ϕ_m is the measurement objective function given by Eq. (4), μ is a weight factor, and ϕ_r is a Tikhonov regularization term. Here, ϕ_r is defined as preferred difference regularization, where the preferred difference between neighboring parameter values is set to zero. The regularization weight factor, μ, is iteratively calculated during each optimization iteration making ϕ_m equal to a user specified

HESSD
12, 9599–9653, 2015

target value (Doherty, 2010). In this case, for ϕ_m defined by Eq. (4), the target value is set to 2 (indicating that the fitted data residuals correspond to the data noise levels).

Method 2 is joint estimation of groundwater model parameters and geophysical model parameters on the basis of both hydrologic and geophysical data (JHI). The minimized objective function is of the same form as Eq. (5), but the measurement and regularization terms are different. For Method 2 the measurement objective function is defined as

$$\phi_{m,joint} = n_h^{-1} \sum_{i=1}^{n_h} \left(\frac{h_{obs,i} - h_{sim,i}}{\sigma_{hi}} \right)^2 + n_r^{-1} \sum_{i=1}^{n_r} \left(\frac{r_{obs,i} - r_{sim,i}}{\sigma_{r,i}} \right)^2 + n_{TEM}^{-1} \sum_{i=1}^{n_h} \left(\frac{V_{obs,i} - V_{sim,i}}{\sigma_{TEM,i}} \right)^2 \quad (6)$$

where n_h, n_r and n_{TEM} are the number of head, discharge and TEM observations, respectively. The first two terms on the right hand side of Eq. (6) are identical to the terms in Eq. (4). The values of V_{obs} and V_{sim} are observed and corresponding simulated decay data from TEM. Finally, σ_{TEM} is the noise level for the TEM data. Each of the three terms on the right hand side of Eq. (6) is divided by the number of respective measurements to promote a balanced weight among the three datasets. (However, this is based on user preference and can be modified within HYTEB.) The regularized objective term for the joint approach is also preferred differences, now defined as

$$\phi_{r,joint} = \mu \cdot \sum_{i=1}^{n_{kpar}} \left(\log_{10}\left(k_{joint,i}\right) - \log_{10}\left(k_{mf,i}\right) \right)^2. \quad (7)$$

In Eq. (7), $k_{mf,i}$ is the estimate of the hydraulic conductivity at the ith pilot point of the groundwater model; $k_{joint,i}$ is also an estimate of hydraulic conductivity, but this estimate is calculated from the estimated electrical resistivity at the same depth and location by using Eq. (1). In this case, the target value of $\phi_{m,joint}$ is set equal to 3.

Method 3 is sequential parameter estimation (SHI) as proposed by Dam and Christensen (2003). First, the geophysical model parameters (electrical resistivities) are estimated on the basis of the geophysical data. Subsequently, the groundwater model

HESSD
12, 9599–9653, 2015

parameters are estimated on basis of the hydrologic data as well as the resistivity estimates that are used as regularizing prior information on the hydraulic conductivity. In the first step, the geophysical inversion is done as "smooth model" inversion (Constable et al., 1987). This means that each geophysical model has fixed 10 m layer thicknesses while the resistivity within the layers is estimated. The 77 1-D models are inverted independently using AarhusInv (Auken et al., 2014), but vertical constraints were used to stabilize the inversion of each 1-D model (Constable et al., 1987). The estimated electrical resistivities are used to constrain the subsequent hydrologic inversion, which is carried out as minimization of Eq. (5) where the measurement objective function ϕ_m is defined by Eq. (4) while the preferred difference regularization term is defined by

$$\phi_{r,seq} = \mu \cdot \sum_{i=1}^{n_{kpar}} \left(\log_{10}\left(k_{seq,i}\right) - \log\left(k_{mf,i}\right)\right)^2. \quad (8)$$

As in Eq. (7), $k_{mf,i}$ is the hydraulic conductivity at the ith pilot point of the groundwater model; $k_{seq,i}$ is the hydraulic conductivity at the pilot point calculated from the corresponding resistivity, estimated in the first step of Method 3, by using Eq. (1). In this case the target value of ϕ_m is set equal to 2.

For all three methods, the objective function is minimized iteratively by the modified Gauss–Newton method. This involves recalculation of the sensitivity matrix for each iteration, which is time consuming due to the large number of model parameters.

3.5.2 Initial parameter values

We did the following to investigate how much the choice of initial parameter values influences the parameter estimates obtained by the three inversion approaches.

For method 1 (HI), we ran two inversions. In the first run, termed HI-T, we used the true hydraulic conductivity values at each pilot point as initial values. We acknowledge that this is not a realistic occurrence but it is done as a control to show the best possible outcome of HI. In the second run, termed HI-H, we assumed a homogeneous initial

HESSD
12, 9599–9653, 2015

hydraulic conductivity field with K equal to $1 \times 10^{-6}\,\mathrm{m\,s^{-1}}$ which is equal to the true mean value of silt.

For method 2 (JHI), we ran three inversions. In the first run, termed JHI-T, we used the true parameter values for hydraulic conductivity and electrical resistivity at the pilot points. As above this is done to show the best possible outcome of JHI. In the second run, termed JHI-H, we used a constant hydraulic conductivity of $1 \times 10^{-6}\,\mathrm{m\,s^{-1}}$ and a constant electrical resistivity of 40 Ωm at the pilot points. In the third run, termed JHI-G, we first ran independent geophysical inversions (one for each sounding location) using a homogeneous half space of 40 Ωm as the starting model. The resulting estimates of electrical resistivity were subsequently used as initial parameter values for JHI-G at the resistivity pilot points, and they were used together with relation Eq. (1) to produce the JHI-G initial values of hydraulic conductivity at the hydraulic conductivity pilot points.

For method 3 (SHI), we only ran one inversion sequence, termed SHI-G. First we ran the independent geophysical inversions using a homogeneous half space of 40 Ωm as the initial model. Subsequently we used the estimated resistivities together with relation Eq. (1) to produce the initial values for hydraulic conductivity at the pilot points that were used for the hydrologic inversion carried out in step two of SHI-G.

In the demonstration example, we examine SHI and JHI approaches, without considering CHI. We made this choice not because we do not see value in examining CHI, nor because of any limitation in HYTEB for examining CHI. Rather, for clarity of presentation, we considered SHI and JHI to be more easily comparable. CHI analyses are generally more valuable when considering transient data; essentially, CHI allows the process model to replace smoothing in time when interpreting the geophysical data. Having made the choice not to examine CHI, we could use a realistic case study that did not include transient data.

HESSD
12, 9599–9653, 2015

3.5.3 Inversion software

The objective functions were minimized using BeoPEST, a version of PEST (Doherty, 2010) that allows the inversion to run in parallel using multiple cores and computers. We used a new version of BeoPEST modified by John Doherty particularly for our purpose to do gradient based minimization involving several models with each of their parameters; thus the modified BeoPEST exploits different parts of the sensitivity matrix that can be calculated by running just one of the models. However, for method 3, the geophysical data were inverted using AarhusInv (Auken et al., 2014).

3.6 Picking 10 realizations

For this demonstration, the computational burden would be overwhelming if the entire HYTEB analysis was to be carried out for each of the 1000 system realizations. We therefore sought a way to reduce the number of models to just 10 that would maintain a representative diversity of models. The strategy we used to down sample from 1000 realizations to 10 was as follows.

We first decided to group the models based on the predictions of interest. It would be reasonable to group models based on other characteristics, such as underlying conceptual model, or zonation, or imposed boundary conditions. However, we contend that for both practical and scientific applications, it is more often the predictions of models that are of primary interest than the structure or parameterization of the models. We began by creating an ensemble from the 25 predictions of interest listed in Table 3 over all 1000 realizations. We then used k-means clustering to group the prediction sets into 10 clusters within this prediction space. Because the units of the predictions varied, all predictions were whitened, or normalized, before clustering. For stability, we ran 1000 repetitions of the clustering to minimize the effects of initial cluster selection. Once the clusters were defined, we identified the prediction set that was closest to the centroids. This resulted in ten models that broadly represent the range of model behaviors, including both the range of each prediction and the correlations among predictions.

HESSD
12, 9599–9653, 2015

4 Results

4.1 Estimated hydraulic conductivity fields

Figure 4 shows the reference hydraulic conductivity fields of the uppermost six layers and a representative cross section for one of the 10 chosen system realizations (see Sect. 3.6). It also shows the corresponding estimated hydraulic conductivity fields obtained by six different inversion runs. The figure can thus be used to visually compare the estimated hydraulic conductivity fields and to judge whether they resolve the structures of the reference model. Figure 5 shows corresponding pilot-point-by-pilot-point scatter plots of reference versus estimated hydraulic conductivity. Except when noted specifically, the results in Figs. 4 and 5 for this realization are typical for all 10 chosen system realizations.

The second and third rows of Fig. 4 show results for the two hydrologic inversion (HI) runs. Inversion HI-T, which used true parameter values as initial values, produces very blurred hydraulic conductivity fields. This is caused by the used Tikhonov regularization constraint which guides the inversion to estimate a field as smooth as possible while still fitting the calibration data. The estimated field for layer one has some structural similarity with the true field but the estimated values vary much less than the true values. Similar results are seen for layers 2 to 5 while structure has disappeared from the deeper layers representing the deposits in the buried valley. Similar results were achieved for three other realizations. For the remaining six realizations HI-T produced very blurred hydraulic conductivity fields for all model layers, having essentially no resemblance to the structure of the reference fields. The third row of Fig. 4 illustrates that for inversion HI-H, which used homogeneous initial hydraulic conductivity fields, there is almost no structurel similarity between the estimated and reference hydraulic conductivity fields, and for most layers the estimated field appears to be almost homogeneous. However, the cross sections show that the structure with high hydraulic conductivity in the bottom of the burried valley is resolved to some degree by both HI-T and HI-H. Figure 5 shows that both HI-T and HI-H underestimate hydraulic conductivities

HESSD
12, 9599–9653, 2015

for high-permeability deposits (sand and gravel) but overestimate for low-permeability deposits (silt and clay). For HI-H, the range of estimated conductivities is the same for high-permeability and low-permeability deposits. For HI-T, there is a small difference between the two ranges – they are slightly shifted in the correct directions compared to HI-H.

The fourth row of Fig. 4 shows hydraulic conductivity fields estimated by the sequential geophysical approach (SHI-G). For the upper layers, the true structures can be recognized, but the resolution decreases with depth. The cross section shows that the true structures of the upper five layers can be identified to some degree from the estimated fields. Because of loss of resolution, the structures cannot be identified inside the buried valley. Figure 5 shows that for low-permeability deposits, the range of estimated log-hydraulic conductivities is twice as large as the true range of values, and the horizontal scatter around the identity line is considerable. For high-permeability deposits, the range of estimated values is much larger than the range of true values, and the estimated values tend to be orders of magnitude too small (Fig. 5). This happens because the resistivities estimated from the TEM data in the first step of the SHI scheme often turn out to be too small if the resistivity at depth is high. This is a well-known result from the fact that the sensitivity of TEM data with respect to layers of high resistivity reduces with depth, which causes problems of equivalence for the geophysical models. (This has been demonstrated and discussed by Auken et al. (2008) for a similar type of geological system.) When resistivity estimates that are too small are used to regularize the second hydrologic inversion step of the SHI scheme, the hydraulic conductivity estimates are likely to be too small as well. Similarly, hydraulic conductivity estimates are too high in some high-resistivity parts of the shallow layers (Fig. 5) because the resistivity estimated from TEM tends to be too high due to low sensitivity of the TEM data. For the studied system, this shows that resistivities estimated by independent TEM data inversion must be used with caution as estimators of hydraulic conductivity or as regularization means for subsequent hydrological inversion. In this case, the absolute relationship between hydraulic conductivity and true electri-

HESSD
12, 9599–9653, 2015

cal resistivity led to an over-reliance on the use of inferred resistivities to populate the model's hydraulic conductivity field.

The last three rows of Fig. 4 show hydraulic conductivity fields estimated by the three joint hydrogeophysical inversion runs (JHI-T, JHI-H and JHI-G), respectively. JHI-T, which used true parameter values as initial values, resolves the true structures of the upper five layers well while the estimated field of layer six is blurred; the cross section shows that the true structures within the burried valley are also resolved to some degree. Figure 5 shows that estimated versus true hydraulic conductivity values plot nicely along the identity line for JHI-T. The resolution of structures (Fig. 4) and the quality of the K estimates (Fig. 5) deteriorate for JHI-H and JHI-G, both of which use less informative initial parameter values. Figure 4 visually indicates that JHI-G resolves structures better than JHI-H. For sand and gravel deposits Fig. 5 shows wider horizontal scatter for JHI-G than for JHI-H. It also shows that estimated hydraulic conductivity for sand and gravel tends to be much too small for both JHI-G and JHI-H (the explanation of which is similar to that given for SHI above), and that particularly JHI-H cannot resolve variations in hydraulic conductivity within the buried valley: the estimated values vary only within roughly an order of magnitude whereas the true values vary within five orders of magnitude.

4.2 Prediction results

For each of the ten chosen geological realizations, each of the six calibrated groundwater models were used to make the model predictions described in Sect. 3.2.2. Figure 6 shows five examples of scatter plots of reference predictions versus calibrated model predictions; each plot shows ten points, each of which corresponds to a particular geological realization selected by the clustering. Each plot also gives the mean error of the prediction (ME) calculated from the ten model predictions. The five predictions represented in Fig. 6 are head in the capping layer at location 1, head recovery at location 1, head recovery within the deeper part of the buried valley at location 8 near the pump-

HESSD
12, 9599–9653, 2015

ing well (Fig. 2), groundwater discharge to the river after pumping has stopped, and recharge area of the pumping well.

Figure 7 shows the mean absolute relative error (MARE) for the 25 model predictions made by models calibrated with six inversion approaches. The relative error magnitudes are calculated as the absolute value of the difference between the reference and predicted value for each prediction of interest averaged over the ten geological realizations considered. The prediction results are discussed individually below.

4.2.1 Head prediction

All calibrated groundwater models appear to be fairly good predictors of hydraulic head except in the buried valley near the pumping well. Unbiased head prediction is exemplified by the plots in the first column of Fig. 6 for which the points scatter around the identity line. This indicates that all calibrated models make unbiased prediction of hydraulic head at location 1. However, the scatter around the identity line is larger for HI calibrated models than for JHI calibrated models. This indicates that the use of geophysical data in JHI reduces the uncertainty of this head prediction as compared to the HI calibrated models. The scatter plots for the other head predictions are similar to those shown for location 1 with the following exceptions. For head prediction 2 (Fig. 2) the points tend to fall above the identity line for all calibrated models, indicating a consistent overprediction in this prediction whether or not geophysical data are used in the calibration process. For head predictions 8, 9 and 10, which are inside the buried valley, the points also tend to fall below the identity line for HI and SHI calibrated models while they plot closer to the identity line for the JHI calibrated models. Use of geophysical data and the JHI approach thus reduce bias and uncertainty of these head predictions.

Figure 7 shows that for all head predictions except at location 2, the use of geophysical data with SHI-G, JHI-H and JHI-G reduces the prediction error when compared to the HI based predictions. It also shows that the relative error magnitude is smaller for head predictions than for most other prediction types. Only change of discharge prediction has a relative error magnitude comparable to the head predictions. The small

HESSD
12, 9599–9653, 2015

relative head prediction errors are likely due to the fact that this type of prediction is similar to the head data used for model calibration. Only the location differs between prediction and calibration heads.

4.2.2 Head recovery prediction

Head recovery due to cessation of pumping is a type of prediction that turns out to be biased for all calibrated models. This is exemplified by the results shown in the second and third columns of Fig. 6. The two plots in the top of the second column indicate that head recovery at location 1 tends to be overpredicted by the models calibrated by purely hydrologic inversion (HI-T and HI-H). The third plot in this column (SHI-G) indicates that some of the bias in the HI-based model prediction may be reduced slightly by using geophysical data in a sequential approach. Finally, the last three plots in the second column of Fig. 6 show that all the models calibrated by JHI appear to be better predictors for this head recovery than the HI and SHI-G based models. The quality of this model prediction appears to be unaffected by the choice of initial parameter values used for JHI. However, for JHI the points tend to scatter around a line with an intercept less than zero and a slope larger than unity. The former indicates consistent bias in the prediction probably due to consistent errors in null space parameter components omitted from the parameterized groundwater model; the latter probably indicates parameter surrogacy incurred through model calibration (see Sect. 2.7). The appearances of scatter plots for head recovery at locations 2 to 7 are similar to that for recovery at location 1 (Fig. 6).

The second plots in the third column of Fig. 6 indicate that head recovery at location 8 within the deeper part of the buried valley is predicted fairly well for nine out of ten geological realizations when the model is calibrated by hydrologic inversion (HI-H); however, the nine points tend to fall slightly above the identity line while the tenth point falls far above the identity line. Generally, the plots indicate a consistent overprediction of head using HI-based inversion. The remaining plots in the third column show that recovery prediction at location 8 turn out to be too large for the models calibrated

HESSD
12, 9599–9653, 2015

with geophysical data, no matter whether this is done by SHI or by JHI. These plots indicate that use of the geophysical data introduces further bias in the prediction of head recovery within the buried valley: if a line is visually fitted through the points, the apparent non-zero intercept indicates bias (see Sect. 2.7). The scatter plots for head recovery at locations 9 and 10, also inside the buried valley, are similar to those for location 8.

Figure 7 shows that for recovery predictions 1 to 7, the use of geophysical data with SHI-G, JHI-H and JHI-G reduces the prediction error when compared to the HI based predictions. For recovery 1, this is confirmed by the scatterplots in column two of Fig. 6. On the contrary, for recovery prediction 8, located within the buried valley, both Figs. 6 and 7 show that using geophysical data with either SHI-G, JHI-H or JHI-G tends to increase the prediction error as compared to HI-H and HI-T. Depending on the choice of initial parameter values, a similar result is seen for recovery predictions 9 and 10. (Explanation for this predictive degradation is given above.) It is finally noted that recovery prediction 2 benefits from use of geophysical data while head prediction at the same location does not, and that the relative error magnitude is larger for recovery predictions than for head predictions. This is likely because head recovery depends on a different stress situation than that represented by the head calibration data.

4.2.3 Discharge prediction

The scatter plots in the fourth column of Fig. 6 indicate that discharge to the river without pumping is overpredicted except for the HI-T and JHI-T based models. Further, this is a type of model prediction that is not improved by including geophysical data in the inversion (compare for example the HI-H plot with the JHI-G plot). If anything, the results for the ten realizations indicate that use of geophysical data may bias discharge prediction unless joint inversion is used with initial parameter values being equal (or close) to the true values (JHI-T). That use of geophysical data is not important to improve this prediction is confirmed by the relative error magnitudes for discharge shown in Fig. 7.

HESSD
12, 9599–9653, 2015

4.2.4 Recharge area and other particle tracking predictions

The plots in the fifth column of Fig. 6 are for the recharge area prediction. Except for JHI-T and JHI-G, the points in all plots appear to fall along an almost vertical line; the scatter along the vertical axis is much longer than the scatter along the horizontal axis, indicating that all of these models are a poor, highly biased predictor of the pumping well's recharge area. Including TEM data in the model calibration only improves this model prediction for JHI-T and JHI-G. Further analysis shows that at least part of the reason for the poor prediction is that the estimated areal average recharge for the model domain in all cases is too low. Lower estimated recharge rates requires a larger predicted recharge area to balance the rate of water pumped from the pumping well. For the JHI-T models, the estimated areal recharge amounts to about two thirds of the actual average recharge. For the JHI-H models the estimated recharge tends to be less than half (for one model realizations as low as one third) of the actual area. The estimated areal recharge for the other models is between the JHI-T and JHI-H estimates. It should be mentioned that all calibrated models sufficiently fit the river discharge measurement; the underestimated recharge means that the simulated discharge to the lake turns out to be too small (typically less than half of the actual discharge to the lake; for one calibrated model there is almost no simulated lake discharge).

It is finally mentioned that the scatter plots look similar to those in column 6 of Fig. 6 for the prediction of average age of groundwater pumped from the well and for the prediction of particle travel time. The explanation for these poor predictive performances is similar to that just given for the prediction of the well's recharge area. Figures 6 and 7 show that use of TEM data does not improve the model performance with respect to prediction of groundwater age and particle travel time.

HESSD
12, 9599–9653, 2015

5 Summary and conclusions

It is intuitively clear that geophysics can offer valuable information for improved groundwater modeling, but for an actual investigation it is often unclear how, at what cost, and to what extent modeling can be improved by adding geophysical data. This reduces to the question: what type and abundance of geophysical (and other) data should be gathered, and how can they be used optimally? This can be clarified by doing controlled experiments. For large spatial scales, these questions can best (only?) be done by synthesizing both the actual hydrogeological setting and the alternative data gathering, doing modeling experiments using these data, and comparing modeling results with the known "synthetic reality" for the alternative choices. This paper presents a newly develop framework that allows for such an application- and method-specific examination of the potential value of using geophysical data to develop a groundwater model and improve its predictive power. We call the framework a **HY**drogeophysical **TE**st-**B**ench (HYTEB). HYTEB allows for sophisticated treatment of hydrologic and geophysical data and inversion approaches. It can be used to examine the combined use of hydrologic and geophysical data, including model parameterization, inversion, and the use of multiple geophysical or other data types. It can also be used to discover potential errors that can be introduced through petrophysical models and approaches to correlating geophysical and hydrologic parameters.

The advantage of using HYTEB is demonstrated by synthesizing a hydrogeological environment that is typical to parts of northern Europe and northern America, consisting of various types of glacial deposits covering low-permeability (in practice impermeable) bedrock of Tertiary clay, which has a surface with the form of a plateau with a deep valley buried by the glacial deposits. The bedrock has low electrical resistivity in significant contrast to the higher resistivity of the glacial deposits. The resistivity of the glacial deposits varies with the grain size and clay content. TEM data is often collected (now by airborne systems) to map this type of environment because it is ideal for mapping the depth to the top of the Tertiary bedrock, including the depth and location

HESSD
12, 9599–9653, 2015

of buried valleys (shown by Auken et al., 2008). The bedrock structure represented in groundwater models for this type of environment is therefore often constructed primarily on the basis of TEM data while the groundwater models are calibrated on the basis of hydrologic data only. Here we use HYTEB to investigate to what extent groundwater model calibration and, often more importantly, model predictions can be improved by including in the calibration process electrical resistivity estimates obtained from TEM data. For simplicity we assumed that the resistivity correlates with hydraulic conductivity and that the relationship is constant and known. (But notice that with HYTEB we could have assumed differently.)

In all calibration cases, the hydraulic conductivity field was parameterized by 550 pilot points. The pilot point estimates for hydraulic conductivity together with two estimated shape factors estimate the recharge field. The pilot points for hydraulic conductivity coincide with the pilot points for resistivity to allow maximum extraction of information from the TEM data about hydraulic conductivity. To estimate this many groundwater model parameters from only 36 hydraulic head and discharge data requires regularization. For purely hydrologic inversion (HI, only using hydrologic data) we used Tikhonov regularization combined with singular value decomposition. For joint hydrogeophysical inversion (JHI) and sequential hydrogeophysical inversion (SHI) the resistivity estimates from TEM were used together with the petrophysical relationship to formulate the regularization term. In all cases, the regularization stabilized the inversion.

It was found that the estimation results depend on the type of regularization used as well as on the choice of initial parameter values; neither the HI nor the JHI objective function could be minimized uniquely. SHI or JHI with regularization based on the use of TEM data produced estimated hydraulic conductivity fields that bear more resemblance to the true fields than when using HI with Tikhonov regularization. However, even for the studied case, for which there is a perfect and known relationship between hydraulic conductivity and electrical resistivity, the hydraulic conductivity field estimated by the inversion methods used here depends on the choice of initial parameter values. Not surprisingly it is found best to use true parameter values as initial values, but the true

HESSD
12, 9599–9653, 2015

values will, of course, never be known. However, HYTEB could be used to dig deeper into this finding. That is, how well does the initial field have to reflect the reference field? Are there specific components of the field that need to be captured in the initial estimate?

For HI, the estimated hydraulic conductivity field turns out to be very smooth in the top layers and almost homogeneous in the deeper layers. For SHI and JHI, the estimated hydraulic conductivity field resolves some of the reference structures in the shallow layers while less or, in the deeper part, no structure is resolved inside the buried valley. However, the estimated hydraulic conductivities are orders of magnitude wrong in some parts of the model. This occurs because the resistivities estimated from the TEM data either in the first step of the SHI scheme or during the JHI scheme can turn out to be either too small or too large when the sensitivity of the TEM data with respect to resistivity is low. For the studied system, this shows that resistivities estimated by SHI or JHI must be used with caution as estimators of hydraulic conductivity or as regularization means for subsequent hydrological inversion. In this case, the use of the absolute relationship between hydraulic conductivity and true electrical resistivity led to an over-reliance on the use of inferred resistivities to populate the model's hydraulic conductivity field. That is, much of the lack of value of the geophysical data arose from a mistaken faith in the power of the petrophysical model in combination with geophysical data of low sensitivity, thereby propagating geophysical estimation errors into the hydrologic model parameters. In other words, even when there is a correlation between electrical resistivity and hydraulic conductivity reliance on the relationship can lead to errors. This is exactly the kind of insight that can be gained from the use of HYTEB before collecting geophysical or other data that would be difficult or impossible to infer without this integrated platform.

With respect to reducing model prediction error, it depends on the type of prediction whether it has value to include geophysical data in the model calibration. It was found that all models are good predictors of hydraulic head. However, head prediction errors

HESSD
12, 9599–9653, 2015

tend to be reduced by models calibrated using both hydrologic and geophysical data (by SHI or JHI) as compared to models calibrated by only using hydrologic data (HI).

When the stress situation is changed from that of the hydrologic calibration data, then all models make biased predictions of head change. Use of geophysical data (with SHI or JHI) reduces error and bias of head prediction at shallow depth but not in the deep part of the buried valley near the pumping well (where the stress field was changed most). Analyzing the prediction results by the method described by Doherty and Christensen (2011) indicates that the geophysical data helps to reduce parameter null space as well as parameter surrogacy for parameters determining the shallow part of the hydraulic conductivity field. In hindsight, this is obvious since the TEM data better resolves the shallow variations in glacial deposits' resistivity than the variations inside the deep buried valley.

For model prediction of change of discharge to the stream, there is no improvement in using the TEM data. HI based prediction results are comparable to SHI and JHI based results.

All models are a very poor predictor of the pumping well's recharge area and groundwater age. The reason for this is that distributed recharge is estimated during the model calibration together with distributed hydraulic conductivity. Recharge is parameterized by assuming a linear log-log relationship between recharge and hydraulic conductivity of the upper model layer; two shape factors of the relationship are treated as parameters that are calibrated together with the pilot point parameters for hydraulic conductivity and (for JHI) resistivity. It was assumed that the shape factors could be estimated because stream discharge data were included in the calibration data set. All models fit this data, but the estimated areal recharge turned out to be two thirds or less of the actual areal recharge. The predicted recharge area of the pumping well and the predicted age of the pumped water therefore turn out to be much too large. So another important insight from the HYTEB analysis is that recharge should be parameterized and estimated in a different way than it was done in the demonstration example. Alterna-

HESSD
12, 9599–9653, 2015

tively HYTEB could be used to consider adding other types of data to better constrain recharge rates.

Acknowledgements. The presented work was supported by HyGEM, Integrating geophysics, geology, and hydrology for improved groundwater environmental management, Project no. 11-15 116763. The funding for HyGEM is provided by The Danish Council for Strategic Research. John Doherty is thanked for making modifications of BeoPEST. Finally, we thanks Troels N. Vilhelmsen, Esben Auken, and Anders V. Christiansen for their participation in discussions and sharing of experiences during the initial phase of the project.

References

Andersen, T. R., Poulsen, S. E., Christensen, S., and Jørgensen, F.: A synthetic study of geophysics-based modelling of groundwater flow in catchments with a buried valley, Hydrogeol. J., 21, 491–503, doi:10.1007/s10040-012-0924-5, 2012.

Archie, G. E.: The Electrical Resistivity Log as an Aid in Determining Some Reservoir Characteristics, Trans. AIME, 146, 54–62, doi:10.2118/942054-G, 1942.

Auken, E., Jørgensen, F., and Sørensen, K. I.: Large-scale TEM investigation for groundwater, Explor. Geophys., 34, 188–194, doi:10.1071/EG03188, 2003.

Auken, E., Christiansen, A. V., Jacobsen, L. H., and Sørensen, K. I.: A resolution study of buried valleys using laterally constrained inversion of TEM data, J. Appl. Geophys., 65, 10–20, 2008.

Auken, E., Christiansen, A. V., Westergaard, J., Kirkegaard, C., Foged, N., and Viezzoli, A.: An integrated processing scheme for high-resolution airborne electromagnetic surveys, the SkyTEM system, Explor. Geophys., 40, 184–192, doi:10.1071/EG08128, 2009.

Auken, E., Christiansen, A. V., Kirkegaard, C., Fiandaca, G., Schamper, C., Behroozmand, A. A., Binley, A., Nielsen, E., Effersø, F., Christensen, N. B., Sørensen, K., Foged, N., and Vignoli, G.: An overview of a highly versatile forward and stable inverse algorithm for airborne, ground-based and borehole electromagnetic and electric data, Explor. Geophys., 46, 223–235, doi:10.1071/EG13097, 2014.

Biella, G., Lozej, A., and Tabacco, I.: Experimental Study of Some Hydrogeophysical Properties of Unconsolidated Porous Media, Ground Water, 21, 741–751, doi:10.1111/j.1745-6584.1983.tb01945.x, 1983.

HESSD

12, 9599–9653, 2015

Carle, S. F.: T-PROGS: Transition Probability Geostatistical Software, Users Manual, Hydrologic Sciences Graduate Group, University of California, Davis, USA, 1999.
Certes, C. and de Marsily, G.: Application of the pilot point method to the identification of aquifer transmissivities, Adv. Water Resour., 14, 284–300, doi:10.1016/0309-1708(91)90040-U, 1991.
Constable, S. C., Parker, R. L., and Constable, C. G.: Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data, Geophysics, 52, 289–300, doi:10.1190/1.1442303, 1987.
Cooley, R. L.: A theory for modeling ground-water flow in heterogeneous media, US Geological Survey Professional Paper 1679, 220 pp., US Geological Survey, 2004.
Cooley, R. L. and Christensen, S.: Bias and uncertainty in regression-calibrated models of groundwater flow in heterogeneous media, Adv. Water Resour., 29, 639–656, doi:10.1016/j.advwatres.2005.07.012, 2006.
Dam, D. and Christensen, S.: Including Geophysical Data in Ground Water Model Inverse Calibration, Ground Water, 41, 178–189, doi:10.1111/j.1745-6584.2003.tb02581.x, 2003.
Danielsen, J. E., Auken, E., Jørgensen, F., Søndergaard, V., and Sørensen, K. I.: The application of the transient electromagnetic method in hydrogeophysical surveys, J. Appl. Geophys., 53, 181–198, 2003.
Day-Lewis, F. D.: Applying petrophysical models to radar travel time and electrical resistivity tomograms: Resolution-dependent limitations, J. Geophys. Res., 110, B08206, doi:10.1029/2004JB003569, 2005.
Deutsch, C. V.: A sequential indicator simulation program for categorical variables with point and block data: BlockSIS, Comput. Geosci., 32, 1669–1681, doi:10.1016/j.cageo.2006.03.005, 2006.
Deutsch, C. V. and Journel, A. G.: GSLIB: Geostatistical Software Library and User's Guide: Clayton V., 2nd Edn., Oxford University Press, 1998.
Di Maio, R., Fabbrocino, S., Forte, G., and Piegari, E.: A three-dimensional hydrogeological–geophysical model of a multi-layered aquifer in the coastal alluvial plain of Sarno River (southern Italy), Hydrogeol. J., 22, 691–703, doi:10.1007/s10040-013-1087-8, 2013.
Doherty, J.: PEST, Model-Independent Parameter Estimation, Watermark, Numerical Computing, 2010.
Doherty, J.: Ground Water Model Calibration Using Pilot Points and Regularization, Ground Water, 41, 170–177, doi:10.1111/j.1745-6584.2003.tb02580.x, 2003.

HESSD
12, 9599–9653, 2015

Doherty, J. and Christensen, S.: Use of paired simple and complex models to reduce predictive bias and quantify uncertainty, Water. Resour. Res., 47, W12534, doi:10.1029/2011WR010763, 2011.

Doherty, J. and Welter, D.: A short exploration of structural noise, Water Resour. Res., 46, W05525, doi:10.1029/2009WR008377, 2010.

Doherty, J. E., Fienen, M. N., and Hunt, R. J.: Approaches to Highly Parameterized Inversion: Pilot-Point Theory, Guidelines, and Research Directions, US Geological Survey Scientific Investigations Report 2010-5168, US Geological Survey, p. 36, 2010.

Faneca Sànchez, M., Gunnink, J. L., van Baaren, E. S., Oude Essink, G. H. P., Siemon, B., Auken, E., Elderhorst, W., and de Louw, P. G. B.: Modelling climate change effects on a Dutch coastal groundwater system using airborne electromagnetic measurements, Hydrol. Earth Syst. Sci., 16, 4499–4516, doi:10.5194/hess-16-4499-2012, 2012.

Feyen, L. and Gorelick, S. M.: Framework to evaluate the worth of hydraulic conductivity data for optimal groundwater resources management in ecologically sensitive areas, Water Resour. Res., 41, W3019, doi:10.1029/2003WR002901, 2005.

Feyen, L., Gómez-Hernández, J. J., Ribeiro, P. J., Beven, K. J., and De Smedt, F.: A Bayesian approach to stochastic capture zone delineation incorporating tracer arrival times, conductivity measurements, and hydraulic head observations, Water Resour. Res., 39, 1126, doi:10.1029/2002WR001544, 2003.

Fitterman, D. V. and Deszcz-Pan, M.: Helicopter EM mapping of saltwater intrusion in Everglades National Park, Florida, Explor. Geophys., 29, 240–243, doi:10.1071/EG998240, 1998.

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., 18, 4349–4362, doi:10.5194/hess-18-4349-2014, 2014.

Franssen, H.-J. H., Gómez-Hernández, J., and Sahuquillo, A.: Coupled inverse modelling of groundwater flow and mass transport and the worth of concentration data, J. Hydrol., 281, 281–295, doi:10.1016/S0022-1694(03)00191-4, 2003.

Freeze, R. A., Massmann, J., Smith, L., Sperling, T., and James, B.: Hydrogeological Decision Analysis: 1. A Framework, Ground Water, 28, 738–766, doi:10.1111/j.1745-6584.1990.tb01989.x, 1990.

HESSD
12, 9599–9653, 2015

Frohlich, R. K. and Kelly, W. E.: The relation between hydraulic transmissivity and transverse resistance in a complicated aquifer of glacial outwash deposits, J. Hydrol., 79, 215–229, doi:10.1016/0022-1694(85)90056-3, 1985.

Günther, T., Rücker, C., and Spitzer, K.: Three-dimensional modelling and inversion of dc resistivity data incorporating topography – II. Inversion, Geophys. J. Int., 166, 506–517, doi:10.1111/j.1365-246X.2006.03011.x, 2006.

Harbaugh, A. W., Banta, E. R., Hill, M. C., and McDonald, M. G.: MODFLOW-2000, The U.S. Geological Survey modular ground-water model: User guide to modularization concepts and the ground-water flow process, US Geol. Surv. Open File Report 00-92, US Geological Survey, 121 pp., 2000.

Harvey, C. F. and Gorelick, S. M.: Mapping Hydraulic Conductivity: Sequential Conditioning with Measurements of Solute Arrival Time, Hydraulic Head, and Local Conductivity, Water Resour. Res., 31, 1615–1626, doi:10.1029/95WR00547, 1995.

Heigold, P. C., Gilkeson, R. H., Cartwright, K., and Reed, P. C.: Aquifer Transmissivity from Surficial Electrical Methods, Ground Water, 17, 338–345, doi:10.1111/j.1745-6584.1979.tb03326.x, 1979.

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, 4043–4060, doi:10.5194/hess-17-4043-2013, 2013a.

Herckenrath, D., Odlum, N., Nenna, V., Knight, R., Auken, E., and Bauer-Gottwein, P.: Calibrating a salt water intrusion model with time-domain electromagnetic data, Ground Water, 51, 385–397, doi:10.1111/j.1745-6584.2012.00974.x, 2013b.

Hill, M.: Methods and guidelines for effective model calibration; with application to UCODE, a computer code for universal inverse modeling, and MODFLOWP, a computer code for inverse modeling with MODFLOW, Water-Resources Investigtion Report 98-4005, Water-Resources Investigtion, doi:10.1061/40517(2000)18, 1999.

Hinnell, A. C., Ferré, 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, W00D40, doi:10.1029/2008WR007060, 2010.

Hubbard, S. S. and Rubin, Y.: Hydrogeological parameter estimation using geophysical data: a review of selected techniques, J. Contam. Hydrol., 45, 3–34, doi:10.1016/S0169-7722(00)00117-0, 2000.

HESSD
12, 9599–9653, 2015

Hubbard, S. S., Rubin, Y., and Majer, E.: Spatial correlation structure estimation using geophysical and hydrogeological data, Water Resour. Res., 35, 1809–1825, doi:10.1029/1999WR900040, 1999.

Hyndman, D. W., Harris, J. M., and Gorelick, S. M.: Coupled seismic and tracer test inversion for aquifer property characterization, Water Resour. Res., 30, 1965–1977, doi:10.1029/94WR00950, 1994.

Jørgensen, F., Sandersen, P. B. E., and Auken, E.: Imaging buried Quaternary valleys using the transient electromagnetic method, J. Appl. Geophys., 53, 199–213, doi:10.1016/j.jappgeo.2003.08.016, 2003.

Koch, K., Wenninger, J., Uhlenbrook, S., and Bonell, M.: Joint interpretation of hydrological and geophysical data: electrical resistivity tomography results from a process hydrological research site in the Black Forest Mountains, Germany, Hydrol. Process., 23, 1501–1513, doi:10.1002/hyp.7275, 2009.

Kowalsky, M. B., Finsterle, S., Peterson, J., Hubbard, S., Rubin, Y., Majer, E., Ward, A., and Gee, G.: Estimation of field-scale soil hydraulic and dielectric parameters through joint inversion of GPR and hydrological data, Water Resour. Res., 41, W11425, doi:10.1029/2005WR004237, 2005.

Linde, N., Binley, A., Tryggvason, A., Pedersen, L. B., and Revil, A.: Improved hydrogeophysical characterization using joint inversion of cross-hole electrical resistance and ground-penetrating radar traveltime data, Water Resour. Res., 42, W12404, doi:10.1029/2006WR005131, 2006.

Marker, P. A., Foged, N., He, X., Christiansen, A. V., Refsgaard, J. C., Auken, E., and Bauer-Gottwein, P.: Performance evaluation of groundwater model hydrostratigraphy from airborne electromagnetic data and lithological borehole logs, Hydrol. Earth Syst. Sci., 19, 3875–3890, doi:10.5194/hess-19-3875-2015, 2015.

Mazáč, O., Kelly, W. E., and Landa, I.: A hydrogeophysical model for relations between electrical and hydraulic properties of aquifers, J. Hydrol., 79, 1–19, 1985.

Menke, W.: Geophysical Data Analysis: Discrete Inverse Theory, 3rd Edn.: MATLAB Edition, Academic Press, Boston, 2012.

Moore, C. and Doherty, J.: The cost of uniqueness in groundwater model calibration, Adv. Water Resour., 29, 605–623, doi:10.1016/j.advwatres.2005.07.003, 2006.

Nowak, W., Rubin, Y., and de Barros, F. P. J.: A hypothesis-driven approach to optimize field campaigns, Water Resour. Res., 48, W06509, doi:10.1029/2011WR011016, 2012.

HESSD
12, 9599–9653, 2015

Oliver, D. S., Reynolds, A. C., and Liu, N.: Inverse Theory for Petroleum Reservoir Characterization and History Matching, Cambridge University Press, University Printing House, Cambridge, UK, 2008.

Pollock, D. W.: User's Guide for MODPATH/MODPATH-PLOT, Version 3: A particle tracking post-processing package for MODFLOW, the U.S. Geological Survey finite-difference groundwater flow model, US Geological Survey Open-File Report 94-464, USGS Publications Warehouse, 248 pp., 1994.

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

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

Refsgaard, J. C., Auken, E., Bamberg, C. A., Christensen, B. S. B., Clausen, T., Dalgaard, E., Effersø, F., Ernstsen, V., Gertz, F., Hansen, A. L., He, X., Jacobsen, B. H., Jensen, K. H., Jørgensen, F., and Jørgensen, L. F.: Nitrate reduction in geologically heterogeneous catchments – a framework for assessing the scale of predictive capability of hydrological models, Sci. Total Environ., 468–469, 1278–1288, doi:10.1016/j.scitotenv.2013.07.042, 2014.

Reilly, T. E.: Techniques of Water-Resources Investigations of the United States Geological Survey, Book 3, Applications of Hydraulics, in: System And Boundary Conceptualization In Ground-Water Flow Simulation, US Geological Survey, Denver, CO, USA, 2001.

Reilly, T. E. and Harbaugh, A. W.: Guidelines for Evaluating Ground-Water Flow Models, US Geological Survey Scientific Investigations Report 2004-5038, Version 1.01, USGS Publications Warehouse, 2004.

Revil, A. and Cathles, L. M.: Permeability of shaly sands, Water Resour. Res., 35, 651–662, doi:10.1029/98WR02700, 1999.

Rücker, C., Günther, T., and Spitzer, K.: Three-dimensional modelling and inversion of dc resistivity data incorporating topography - I. Modelling, Geophys. J. Int., 166, 495–505, doi:10.1111/j.1365-246X.2006.03010.x, 2006.

Sandersen, P. B. E. and Jørgensen, F.: Buried Quaternary valleys in western Denmark-occurrence and inferred implications for groundwater resources and vulnerability, J. Appl. Geophys., 53, 229–248, 2003.

HESSD

12, 9599–9653, 2015

Seifert, D., Sonnenborg, T. O., Scharling, P., and Hinsby, K.: Use of alternative conceptual models to assess the impact of a buried valley on groundwater vulnerability, Hydrogeol. J., 16, 659–674, doi:10.1007/s10040-007-0252-3, 2007.

Slater, L.: Near Surface Electrical Characterization of Hydraulic Conductivity: From Petrophysical Properties to Aquifer Geometries – A Review, Surv. Geophys., 28, 169–197, doi:10.1007/s10712-007-9022-y, 2007.

Steuer, A., Siemon, B., and Eberle, D.: Airborne and Ground-based Electromagnetic Investigations of the Freshwater Potential in the Tsunami-hit Area Sigli, Northern Sumatra, J. Environ. Eng. Geophys., 13, 39–48, doi:10.2113/JEEG13.1.39, 2008.

Tonkin, M., Doherty, J., and Moore, C.: Efficient nonlinear predictive error variance for highly parameterized models, Water Resour. Res., 43, W07429, doi:10.1029/2006WR005348, 2007.

Urish, D. W.: Electrical resistivity-hydraulic conductivity relationships in glacial outwash aquifers, Water Resour. Res., 17, 1401–1408, doi:10.1029/WR017i005p01401, 1981.

Vereecken, H., Hubbard, S., Binley, A., and Ferre, T.: Hydrogeophysics: An Introduction from the Guest Editors, Vadose Zone J., 3, 1060–1062, doi:10.2113/3.4.1060, 2004.

Viezzoli, A., Munday, T., Auken, E., and Christiansen, A. V.: Accurate quasi 3D versus practical full 3D inversion of AEM data – the Bookpurnong case study, Preview, 2010, 23–31, doi:10.1071/PVv2010n149p23, 2010a.

Viezzoli, A , Tosi, L., Teatini, P., and Silvestri, S.: Surface water-groundwater exchange in transitional coastal environments by airborne electromagnetics: The Venice Lagoon example, Geophys. Res. Lett., 37, L01402, doi:10.1029/2009GL041572, 2010b.

Vilhelmsen, T. N., Behroozmand, A. A., Christensen, S., and Nielsen, T. H.: Joint inversion of aquifer test, MRS and TEM data, Water Resour. Res., doi:10.1002/2013WR014679, in press, 2014.

West, G. F. and Macnae, J. C.: Physics of the Electromagnetic Induction Exploration Method, in: Electromagnetic Methods in Applied Geophysics – Part A and Part B, edited by: Nabighian, M. N., Society of Exploration Geophysicists, Tulsa, USA, 1991.

Worthington, P. F.: Quantitative geophysical investigations of granular aquifers, Surv. Geophys., 2, 313–366, doi:10.1007/BF01447858, 1975.

HESSD

12, 9599–9653, 2015

Table 1. Geostatistical parameters for stochastic hydraulic field employed by the hydraulic model.

	$\log_{10}(K)$			$\log_{10}(R)$			$\log_{10}(\varphi)$		
Category	μ	a	σ^2	μ	a	σ^2	μ	a	σ^2
Gravel	−3.00	200.	0.0227	−8.20	200.	0.007752	−0.60	200.	0.000428
Sand	−4.00	200.	0.0227	−8.20	200.	0.007752	−0.60	200.	0.000428
Silt	−6.00	200.	0.0227	−8.60	200.	0.007752	−0.74	200.	0.000428
Clay	−7.00	50.	0.122	−8.82	50.	0.007752	−1.00	50.	0.000428

HESSD
12, 9599–9653, 2015

Table 2. Location and screen layer of boreholes with head measurements for model calibration.

Location				Location				Location			
Borehole	*X* (m)	*Y* (m)	Screened layer	Borehole	*X* (m)	*Y* (m)	Screened layer	Borehole	*X* (m)	*Y* (m)	Screened layer
well_1	3692	6100	4	well_13	2375	4127	19	well_25	1460	2064	5
well_2	2375	5824	8	well_14	1155	3905	3	well_26	2506	2024	20
well_3	850	5662	4	well_15	2616	3720	20	well_27	2611	1990	18
well_4	4308	5602	3	well_16	2394	3637	19	well_28	2468	1750	20
well_5	2717	5570	6	well_17	4073	3565	4	well_29	2893	1741	9
well_6	1201	5550	4	well_18	2828	3498	12	well_30	4255	1632	4
well_7	2144	5477	8	well_19	2140	3421	10	well_31	2542	1482	20
well_8	2384	5006	16	well_20	2412	3184	20	well_32	2357	1047	5
well_9	2634	4830	14	well_21	665	3042	4	well_33	900	705	5
well_10	1174	4583	3	well_22	2311	2823	13	well_34	2838	649	11
well_11	4243	4506	4	well_23	2884	2379	6	well_35	2384	400	12
well_12	2708	4330	15	well_24	2421	2231	20				

HESSD
12, 9599–9653, 2015

Table 3. Different types of model predictions with and without a pumping well.

With pumping (the flow situation when calibrating)	Without pumping
1. Head at 10 locations	4. Head recovery at 10 locations
2. Recharge area	5. Particle travel time
3. Average groundwater age	6. Relative particle endpoint
	7. River discharge

HESSD
12, 9599–9653, 2015

Table 4. Head and head recovery prediction points and screen layer.

Location				Location			
Head pred. point	X (m)	Y (m)	Screen	Head pred. point	X (m)	Y (m)	Screen
pred_1	2500	5100	5	pred_6	2260	5650	5
pred_2	900	2000	4	pred_7	1600	3650	5
pred_3	1025	5600	5	pred_8	2606	1950	19
pred_4	4100	5825	4	pred_9	2464	2128	20
pred_5	2580	3975	15	pred_10	2505	1615	15

HESSD
12, 9599–9653, 2015

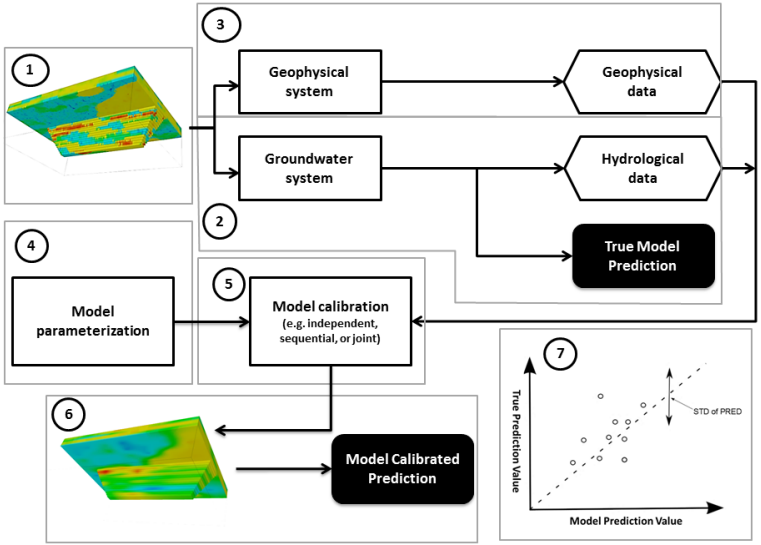

Figure 1. Workflow of the HYTEB. Each numbered dashed box marks a major step in the work flow. In parts 1 and 6 the red, yellow, blue and green colors indicate different categories (types) of geological deposits; color variation within each category indicates variation in hydraulic conductivity.

HESSD
12, 9599–9653, 2015

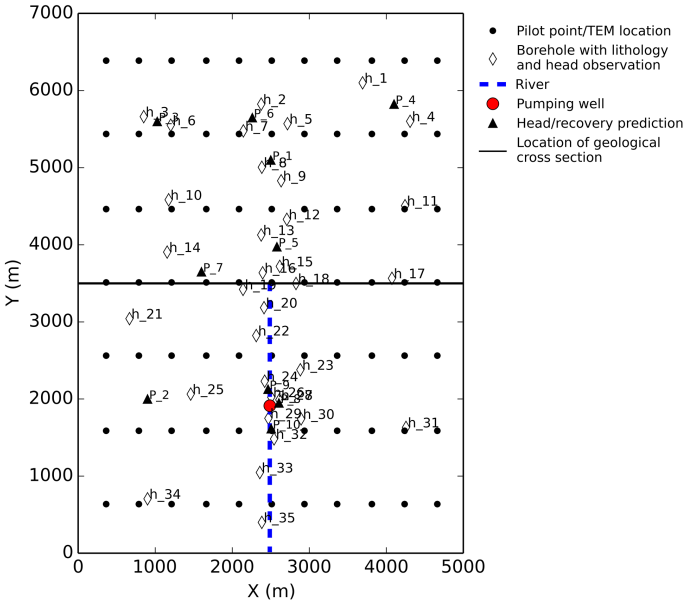

Figure 2. A map of locations of boreholes, a pumping well, geophysical data, pilot points, predictions of interest and location of a geological cross-section. (The positions of the pilot points and geophysical measurements are coincident.)

HESSD
12, 9599–9653, 2015

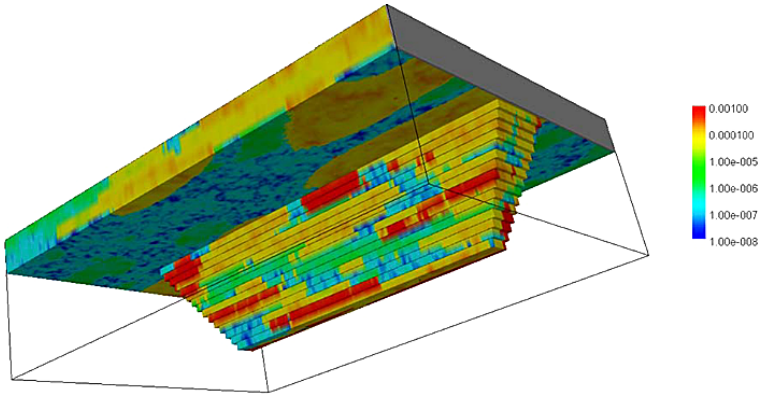

Figure 3. Hydraulic conductivity field for one of the model realizations. (Red shades are for gravel, yellow for sand, green for silt, and cyan/blue for clay.)

HESSD
12, 9599–9653, 2015

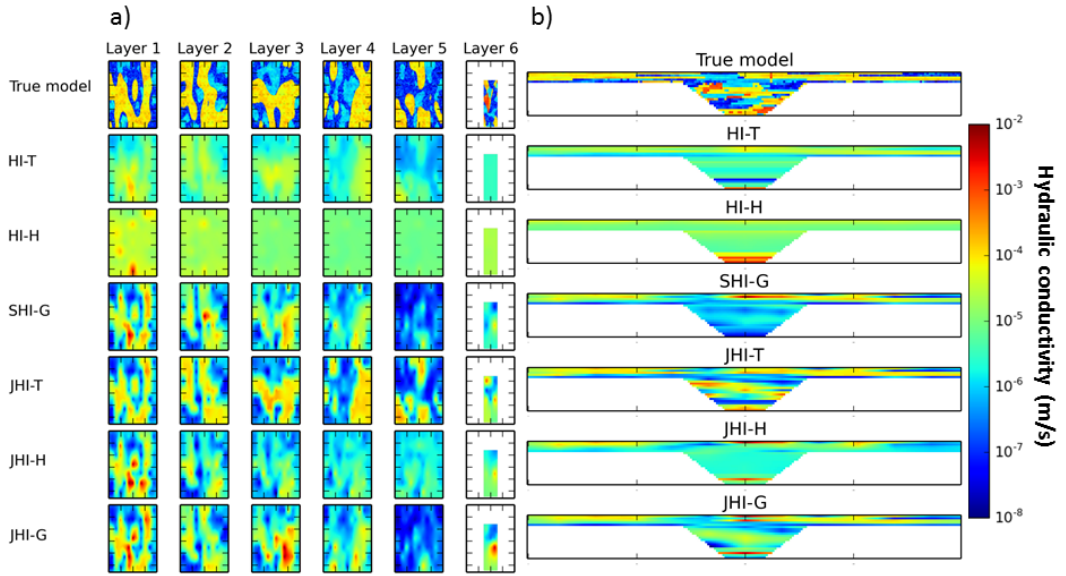

Figure 4. True and estimated hydraulic conductivity fields for model realization number 189: **(a)** shows the fields for layers 1 to 6 ; **(b)** shows the field along an east-west cross section in the middle of the domain.

HESSD
12, 9599–9653, 2015

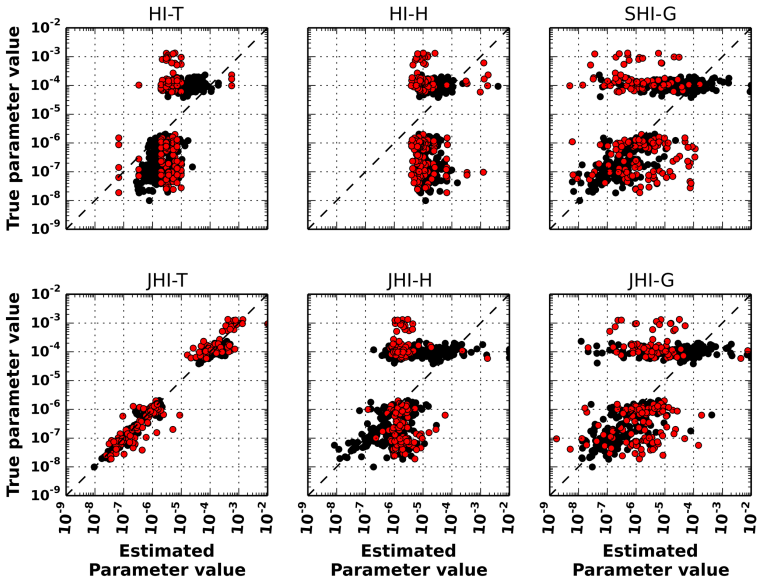

Figure 5. Pilot-point-by-pilot-point scatter plot of true versus estimated hydraulic conductivity for the six inversion runs. Black dots are estimated parameter values from the capping part of the model, while the red dots are estimated parameter values within the buried valley.

HESSD
12, 9599–9653, 2015

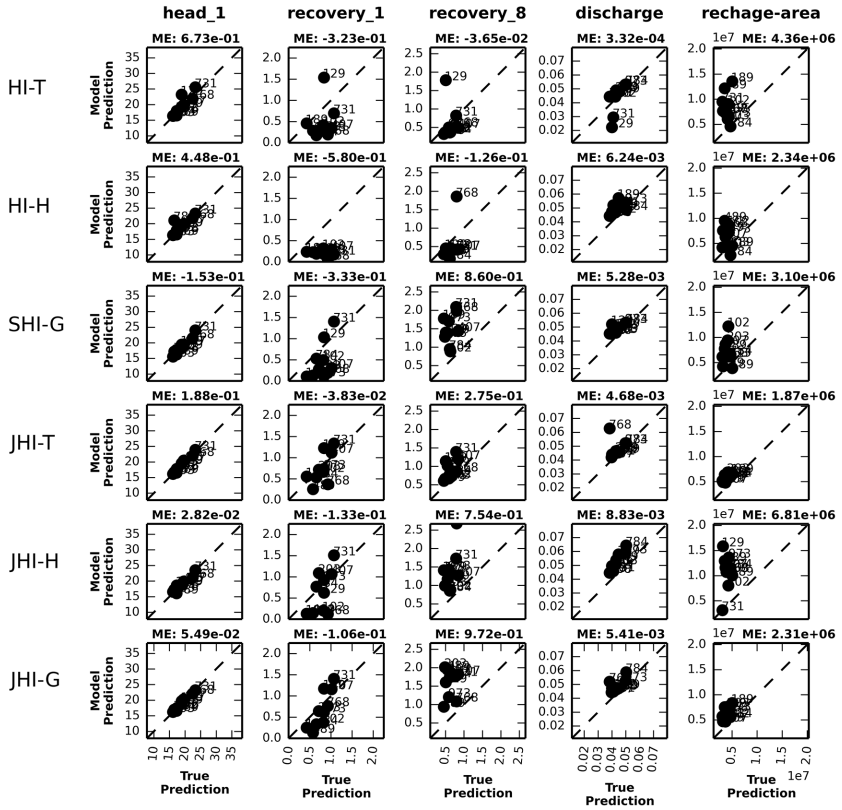

Figure 6. Scatter plots of calibrated model prediction versus the true model prediction using results from the ten geological realizations. The plots in the first column is for head in the capping layer at location 1, the second column is for head recovery in the capping layer at location 1, the third column is for head recovery within the buried at location 8 (Fig. 2), The fourth column is for groundwater discharge to the river after pumping has stopped and fifth column is for recharge-area to the pumping well. ME quantifies the mean prediction error calculated on basis of the ten realizations.

HESSD
12, 9599–9653, 2015

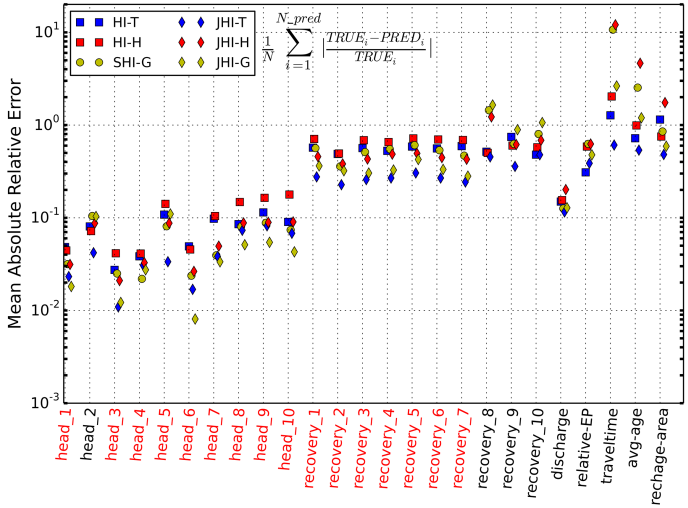

Figure 7. Mean absolute relative prediction error calculated from the ten geological realization results. The symbol type indicates the inversion approach and the symbol color indicates the initial parameter values used when calibrating the groundwater model. Red labels at *x* axis highlight prediction errors that are reduced by using TEM data for model calibration.

HESSD
12, 9599–9653, 2015