Referee #1: The author's response and corrections

As the old reviewer 1, I have now read the new version of the manuscript and the author responses to the two reviews. The new manuscript is a vast improvement and most sweeping statements of the previous version have been removed. It is overall a nice contribution that will be of interest to hydrogeophysicists and hydrogeologists around the world. I don't agree with all the responses made by the reviewers in their response letter, but I find mostly that the new version of the manuscript is mature and convincing.

I would like to make a few more comments.

(1) comments from referees	(2) author's response,	(3) author's changes in manuscript
The response letter states that the HYTEB code is now accessible from a given website. I strongly suggest that this information is also made in the paper, for example, in the acknowledgements.	We will add a link and put a comment in the acknowledgements saying that the software is available on github.	Link and comment added to acknowledgement
Note that data are plural, so it should not be "this" data, but "these" data. For example, page 2, line 3. There are a few cases like this (e.g., page 8, line 14).	Will be corrected	done
Page 6, line 13: Either keep the sentence as it is, but remove "objectively", or keep "objectively" and state that you can investigate the worth of geophysical data for a specific data integration/calibration/joint inversion procedure. The approach provide objective results for a given workflow, it does not provide objective insights about the use of geophysics in hydrogeology. For example, I would most likely address the problem very differently, but if I would use HYTEB, then I would get the same results as the authors.	We have removed "objectively" from the sentence	done
Page 6, line 28: Consider to remove "electric data" as they are not discussed in the paper.	We have removed "electric" from the sentence	Done
Page 9, first few lines. Clarify that if a positive or negative log-log relationship exist ad a site it would be expected to have a significant scatter.	We have added two sentences on this.	We now write: "Both positive and negative relationships have been reported, and there can be significant uncertainty in the relationship (e.g. Mazáč et al. 1985)".

Section 2.6. I assume that you refer here to data that are not used in the calibration. It would be good to make this very clear. This was at least not clear in my reading. If it would be the data used in calibration, then there would be no need for a separate section as this comparison is made during calibration/inversion.	We will add a few clarifying sentences.	Few changes in section 2.2 and section 2.6. In the latter we now explain: "A prediction is a state-variable different from the calibration data, for example a forecast."
Page 12, line 13. You don't make measurements of "electrical resistivity". You use either inferred electrical resistivities or use data sensitive to electrical resistivity.	Text will be modified.	We changed from "measurements of electrical resistivity" to "conductivity inferred from of electrical resistivity".
Page 14, lines 11 and 18. Please decide if you have 6 or 7 predictions of interest.	Right. We have 7 different types of predictions.	Six changed to seven
Page 15, line 9. I suggest replacing "system" with "property".	Word will be replaced	Done
Page 17, line 5 and 8. Your TEM data are primarily sensitive to porosity, so the sentence reads strangely to a geophysicist. Here, the authors have chosen to not account for comments made by reviewer 2. This is a bit unfortunate. It would be better to say that electrical resistivity is sensitive to porosity, but that this relationship is not incorporated within the present work.	The text will be modified.	Line 5-8 on page 17 have been changed to: "Electrical resistivity is sensitive to porosity, but that is not incorporated in the relationship within the present work. Therefore porosity cannot be estimated from the hydrological and geophysical data available here, we always use the reference porosity field for making model predictions." The sentence in the parenthesis has been removed. (suggested by referee #3)
I have some problems to understand the joint inversion. The third term of equation 4 is fitted by varying the hydraulic conductivity field that in turn leads to varying electrical resistivity using equation 1. This is a joint inversion with a perfectly known petrophysical relationship. My confusion arises with equation 5. It seems that Kmf is entirely linked to the model that you are estimating, so how can it stabilize the	In joint inversion, hydrological and geophysical models are set up separately, but hydrological and geophysical parameters are estimated simultaneously through a joint objective function (see eq. [4]). Then we are employing a parameter coupling between the geophysical and hydrologic model (as suggested by Herckenrath et al 2013a). In the regularization objective function (eq[5]) Kmf is the estimated hydraulic conductivity from the hydrological model and Kjoint is K inferred from the	We have added some text in section 3.5.1 which hopefully clarifies our JHI model setup.

solution? I understand how this is done for the sequential inversion, in which the previously inverted geophysics is used to regularize the hydrological inversion (see equation 6). It seems to me that something is missing in the description here and I wouldn't know how to implement this joint inversion using equation 4 and 5. The regularization term needs to come from other sources of information or assumptions than those used in the data misfit function. This is indeed what is done in the so-called sequential inversion.	estimated electrical resistivities from the geophysical models. Eq [5] is implemented as preferred difference regularization which allows information to migrate between the hydrological and geophysical model. This approach was also demonstrated by Vilhelmsen et al 2014 and Herckenrath et al 2013a Our approaches couples the geophysical model and hydrological model through parameters in oppose to approach e.g Hinnel et al. (2010) who focused on simulated hydrologic state variables. We will clarify the explanation of the joint inversion.	
page 23, line 15: It should be "structural".	Word will be replaced	Done

Referee #3: The author's response and corrections

The paper entitled ”A framework for testing the use of electromagnetic data to reduce the prediction error of groundwater models” studies the potential improvement brought by geophysical data to the calibration of hydrogeological models through sequential and joint inversion. Globally, I think this study interesting because there are not a lot of published papers showing coupled inversion of geophysical and hydrogeological data at this scale and for very heterogeneous systems. I was not reviewer for the first round of review, but I have the feeling that the authors did not take in consideration all the remarks of the reviewers to improve the manuscript, performing only a minor to moderate revision.

Therefore, I still have some major concerns about the paper:

General comments

(1) comments from referees	(2) author's response,	(3) author's changes in manuscript
1.It is still confusing if the manuscript is about HYTEB or a comparison study of sequential and joint inversion. As pointed out by the Editor in the early stage of the review process, the methodology (behind HYTEB) described here is common to a lot of studies considering geophysical data or hydrogeological modeling (i.e. considering synthetic benchmarks to assess a new technique). A technical paper (probably in a different journal) about HYTEB would be interesting if it was presenting an end-product and demonstrated with various examples (different geophysical data and modeling tools, different inversion procedures, etc.). For such a paper, a much more thorough description of the way HYTEB works would be needed. Is HYTEB very flexible if one wants to use another forward simulator or inversion packages? How can specific petrophysical relationships be introduced, relating other parameters (for example porosity and resistivity)? How easily can the user define his own objective function? What about transport (e.g. time-lapse geophysical data)? What are the limitations (stochastic inversion, use of McMC, multiple-point geostatistics, etc.)? On a scientific basis, we are more interested in the outcomes of the comparison study. I suggest to limit the reference to HYTEB to the methodology	We must obey that the editor and reviewers seem to agree on this.	We have limited of our reference to HYTEB outside the methodology section. We have also reduced its mentioning in the Abstract, and we have changed the title of the manuscript to “Testing alternative uses of electromagnetic data to reduce the prediction error of groundwater models”. Finally, we have changed the title and somewhat the content of subsection 1.3.

section and avoid further references in the text (stating that the methodology has been implemented in a software called HYTEB, available at …). Many sentences referring to author choices are followed by comments in parenthesis saying that it is not limited by HYTEB. This should be avoided, since the paper is actually not about HYTEB.		
2. Considering the first comment, I think that the authors should focus on their comparison of sequential and joint inversion. They should propose a more thorough literature review and discussion of both sequential and joint inversion approach to highlight the novelty of their approach (is it the comparison between the techniques, the level of heterogeneity considered, the scale, the use of EM data? Given the cited references, I would suggest to insist on the heterogeneity and the number of parameters to optimize). As an example, the paper of Lochbuhler et al. (2013) describes a structure-coupled inversion in terms of penalizing cross-gradient between geophysical and hydraulic parameters as traditionally done in joint geophysical inversion, avoiding the need for a petrophysical relationship, which is quite different of the proposed approach. On another level, the authors do not discuss stochastic inversion (e.g. Irving and Singha, 2010, which is also a joint inversion with transport data) or the possibility to use a Bayesian technique in the sequential inversion which inherently introduce some uncertainty in the petrophysical relationship (Dubreuil-Boisclair et al. 2011, Ruggeri et al., 2013, 2014, for an application to sequential hydrogeological inversion, see Hermans et al., 2015).	Thank you for the advice on "focus" which have tried to follow. We have added a section in the introduction briefly mentioning examples of stochastic methods (including the recommended Irving and Singha, 2010, Ruggeri et al., 2013, 2014). However, we argue for why we find deterministic methods methods most relevant in our context.	See new addition to sections 1.1 and 1.2, respectively.
3.The importance of the choice of the petrophysical relationship has already been raised previously. I agree with the authors that this is a	We understand the reviewer's argument. We acknowledge that it can be problematic to use a petrophysical relationship to combine	We modified eq. (6) to allow variable weighting of the terms in the summation. We redid SHI using weights determined from the

simplification that can be accepted for a synthetic case (direct link between hydraulic conductivity and resistivity) to simulate a realistic geophysical data set. However, even if such a relationship is assumed to build the data and perform the joint inversion approach (where the updated hydraulic conductivity are transformed in resistivity to simulate the geophysical data), it is clearly not valid the other way around in the sequential inversion. The limited resolution of geophysical method and inversion will degrade the relationship. This will not necessarily degrade the results. Indeed, if we assume that the inversion of geophysical data smoothed the resistivity distribution, then the range of variation of resistivity in the inverted model is small and high/low hydraulic conductivity will not be represented in the hydrogeological model. This should be taken into account in the sequential approach (which will also influence the JHI-G results)	inversion of geophysical and hydrological models. We already touch upon these problems in the Introduction (p4 L9-14, L18-21).. We also realize that in for example SHI we could have weighted the individual terms of the sum on the r.h.s. of eq. (6) in accordance with the uncertainty of the corresponding resistivity estimate giving $\log(K_{seq,i})$. We expect this would have eliminated some of the concerns of the reviewer of geophysics loosing resolution. We have therefore made a supplementary SHI run where we changed weighting of terms in the summation of eq. 6. We briefly summarize these results in section 3.5.1.	variance of corresponding resistivity estimates (from the geophysical inversion). Text added after eq. (6) briefly summarizes the results. In section 3.5.2 we added a few lines of text explaining the result of SHI depends on the choice of initial parameter values. We did not mention this in the previous manuscript version.
4.The authors seem to use some estimates of resistivity to deduce the hydraulic conductivity to constrain the joint inversion approach. Clarification is needed, since the basic idea of joint inversion is to avoid geophysical inversions so that there is no error occurring from the "bad" recovery of geophysical parameter after inversion. Adding such a constraint would therefore reintroduce some error due to the limited resolution of geophysics and limit the improvement brought by the method.	As also said in the answer to referee #1: In joint inversion, hydrological and geophysical models are set up separately, but hydrological and geophysical parameters are estimated simultaneously through a joint objective function (see eq. [4]). Then we are employing a parameter coupling between the geophysical and hydrologic model (as suggested by Herckenrath et al 2013a). In the regularization objective function (eq[5]) Kmf is the estimated hydraulic conductivity from the hydrological model and Kjoint is K inferred from the estimated electrical resistivities from the geophysical models. Eq [5] is implemented as preferred difference regularization which allows information to migrate between the hydrological and geophysical model.	We have added some text in section 3.5.1 which hopefully clarifies our JHI methodology.

	This approach was also demonstrated by Vilhelmsen et al 2014 and Herckenrath et al 2013a Our approaches couples the geophysical model and hydrological model through parameters in oppose to approach e.g Hinnel et al. (2010) who focused on simulated hydrologic state variables.	

Specific comments

(1) comments from referees	(2) author's response,	(3) author's changes in manuscript
P2.L25-29. The authors describe here one way to tackle the hydrogeological problem, which is the deterministic approach. They ignore the stochastic approach to solve the problem and propose predictions with uncertainty estimate (see Zhou et al., 2014).	See response to General comment 2.	See new addition in section 1.2.
P3.L20. See also Xu and Valocchi for Bayesian inference to manage structural error.	We have not included this reference since it does not deal with application of geophysics in groundwater modeling.	Nothing changed.
P3. The prediction can also be uncertain because (iv) the non-unicity of the solution. Many models can explain the data, but could produce different predictions (v) Error in the prior distribution of parameter (for example in this case the pilot-point method will generate a smooth model, whereas the truth is much more heterogeneous and can only be represented by a prior corresponding to the one used to generate the true model). To calibrate the model while conserving the prior, methods were developed such as the gradual deformation (Roggero and Hu, 1998) or probability perturbation method (Caers, 2003).	We agree. To some extent we find that the mentioned sources of uncertainty, iv and v, have to do with structural uncertainty and structural error of the model, and that they therefore can fall within our categories (ii) and (iii) in the manuscript. Not to lengthen the manuscript we decided to just reformulate one sentence, thereby admitting that other sources of uncertainty also exist.	Sentence changed to: "However, the prediction will be uncertain for various reasons of which we will emphasize three."
P3.L21-L26. You don't include isotopic or chemical data, why mentioning them? HYTEB does not seem to be able to integrate such analysis.	HYTEB can include such data although we did not demonstrate it here. We therefore disagree and find it relevant to mention the data here	Nothing changed

	since they can be considered important in many studies.	
P5.L3. See major comment 2. With probabilistic relationship, uncertainty related to inversion and petrophysics is included and the proposed hydrological model is in accordance with the data.	We slightly reformulate this sentence.	We have reformulated this sentence to read: “Furthermore, with the SHI approach used in the following the geophysical models cannot be easily updated to conform to the hydrologic observations.”
P5.L4, “extract more information” is not clear, especially given the outcomes of this study. More information on the hydrogeological parameter? More accurate information because the regularization step is avoided?	Not to give a lengthened explanation, we will remove this.	We have removed “that extract more information from the data sets” from the sentence
P5.L13. Note that the assumed relationship can be considered uncertain (see Irving and Singha, 2010).	We agree. In a study that we are currently finishing the relationship is considered uncertain.	Nothing changed on P5 L13 Few changes on P9 L1-2
P5.L20-25. You don’t answer all of those questions in the paper.	We agree.	We have removed the questions not dealt with in this paper.
P5.L12-23. I would not call HYTEB a framework, since the described framework has been applied in many studies in hydrogeophyiscs. It is a software or a platform.	OK.	We have changed framework to platform. We have also changed the title of the manuscript to “Testing alternative uses of electromagnetic data to reduce the prediction error of groundwater models”.
P6.L28. Delete “mainly electric and”, only electromagnetic is shown.	OK.	“Electrical” is removed
P7.L7. HYTEB instead of HYTEM.		Done
P7L28. Are T-Progs and Blocksis integrated to HYTEB? What about other metods such as multiple-point geostatistics (SNESIM, Direct sampling)? Can they also be used in a stochastic inversion framework? Why is this step necessary, can we directly simulate a continuous field with sgsim (step 2)?	Well, software developed by others is not directly integrated into HYTEB, but it can be called from there. HYTEB is a tool for automatizing all the logistics and book-keeping of all of the different independent softwares used for doing the model setup, parametrization, calibration, analyzing etc.	Nothing changed
P8.L28. If it is not the case, why would one use geophysical data to calibrate a hydrogeological model?	There could be cases where geophysics (e.g. resistivity) informs about lithology but only poorly about values of hydraulic conductivity. In such a case the geophysical reference system should not be generated to correlate with hydraulic conductivity.	Nothing changed.

P9.L5. Random error is not able to represent all sources of error measurement such as systematic error linked to the measuring device. Remove “to represent all sources of error”.	We agree. We have changed the text to also include method specific error. Time-domain electromagnetic measurements as used in the present study is influenced by both Gaussian noise and a background noise contribution (see Auken et al. 2008)	“…can be corrupted by adding method specific and random error.”
P10.L13-14. Zones and pilot-point are two types of parameterizations, but they are not any type. Does HYTEB works only with deterministic optimization approaches (where the number of parameter must be relatively small?)	If the review mean overdetermined inversion (more observations than parameters) then no. With HYTEB we can also deal with highly parametrized models. Currently PEST is used as optimization software and this sets the limitation of number of parameters that can be estimated. In our demonstration example we are estimating more than 500 parameters!	Nothing changed.
P14.L11. Your framework does not assess uncertainty in the prediction, it calculates one prediction for the calibrated model. So the actual influence of the geophysical data on the uncertainty of the estimate cannot be estimated. Uncertainty estimates would require a stochastic approach.	The reviewer is correct. Text is modified.	Text changed to: “the addition of geophysical data for groundwater model calibration is likely affect the error of model predictions differently”.
P15.L19. See my major comment 3. The relationship between resistivity after inversion and hydraulic conductivity should be checked as well to verify that.	What we mean by this sentence is that when there is a perfect relationship between resistivity and hydraulic conductive (a relationship not contaminated by noise) then there will be maximum possible information in the geophysical data about hydraulic conductivity. This does not mean that the inverted geophysical model (resistivities) will give perfect estimates of hydraulic conductivity.	We have changed the wording of this paragraph to better explain what we mean.
P15.L24-26. Delete this kind of statement (see major comment 1), here and later.	OK.	Done.
19. P16.L15. Better give the noise level than refer to another study.	The text will be modified.	Done
20. P16.L25. What is the true correlation length? Given the categorical nature of the true model, the variogram is probably not able to capture the	See table 1 for true correlation lengths. We partly agree. From figure 5 it is seen that the	Nothing changed.

heterogeneity and the pilot-point method not well suited for inversion.	pilot points are able to resolve the large scale variations. However, the small-scale variations are not resolved by the pilot points.	
21 .P16.L30. Your results show the contrary.	We disagree. Figure 5 shows that the pilot point parametrization for SHI and JHI fairly well resolves structures of the shallow layers. However, we admit that the transitions are smoothed which may or may not influence the prediction to some degree.	Nothing changed.
22. P17.L7-8. Delete what is in parenthesis.		Done
23. P17L19. Delete, Traditionally, it suggests that scientists rely only on what was done in the past and do not improve their methods.	We agree	Done
P17L12. It is not clear how mu is optimized, line search method?	Optimization of mu is based on a linearized model approximation.	We modified the text to read “based on a linearized model approximation” and “(for details, see Doherty, 2010)”.
P17L14. Do you mean equation 2 (not 4)?	Yes,	Corrected.
P19L5L8. This is not clear to me. Why in a joint inversion would you have two different estimates of K? I expect that a K value is obtained trough the inversion process, then, based on this K field, you can calculate the outputs of the model (heads, discharge, and simulated geophysical data through the petrophysical relationship). There is no need to estimate the hydraulic conductivity from the resistivity, it is the other way around. Actually, the whole point of joint inversion is to avoid the regularization of geophysical data (and therefore smoothing) to estimate the resistivity. See major concern 4. Please clarify.	See our reply to general comment #4.	Nothing changed here.
P20L16. Does it verify equation 1?	The reviewer misunderstands what we are saying. See our answer to general comments #4 on the way we have implemented the joint inversion.	Nothing changed.
P21L15-17. You don’t study the effect of conceptual	The reviewer misunderstands what we are	Nothing changed

models, boundary conditions, etc. Why would you use such characteristic to group them?	saying. We are grouping on model predictions. But one could instead prefer to group on conceptual models, boundary conditions etc.	
P22L16. Of course, the minimization problem is dependent on the starting model, this is a well known fact. Also, the procedure does not ensure that you find a global optimum.	We did not know beforehand that the results would be so sensitive to the "starting model", but we checked it by using alternative initial values, and we present the results. (For overdetermined problems in groundwater modeling it is often found that the minimization finds a unique minimum no matter the initial parameter values,)	Nothing changed.
P23L13-L16. This shows that the hydraulic conductivity field has little influence on the steady state results. The influence of the boundary conditions and pumping rate are probably preponderant.	It shows that many different hydraulic conductivity fields can fit the data. This also appears from the fitting results in Fig. 4.	Nothing changed.
P23L18-L20. This is expected given the smooth model recovered by the pilot-point method. Basically, Figures 5 and 6 show the same results since the spatial distribution is unequivocally determined by the value at the pilot-points and the variogram. What about the comparison of hydraulic conductivity at other locations? This would highlight the limitations of the pilot-point method to recover the true conductivity field.	Figure 5 is important because it illustrates how much structure comes out of the model calibration. We could replace figure 6 with layer-dependent density plots made from hydraulic conductivity values of each grid element in the layer. However, this makes a much larger figure than the present Figure 6, and the new shows similar patterns as the old. However, the reviewer is correct that the new figure also illustrates that interpolation from pilot points smears out the conductivity field. The same is shown by Figure 5. We therefore prefer to keep the present Figures 5 and 6.	Nothing changed.
P24L12-L14. A petrophysical relationship accounting for the limitations of EM inversion (for example by comparing co-located measurements after inversion) could reduce this and make a more realistic synthetic benchmark. See my major comment 3.	We agree.	No changes made here, but the changes mentioned below regarding the reviewers comment on P29L15 at least partly accounts for this comment.
P24L22-L23. Not sure about that. According to	We thank the reviewer for catching a mistake.	Figure 6 has been corrected.

the cross-section, JHI-H performs better.	Figure 6 is wrong: the bottom row shows results for JHI-H (not JHI-G as the figure says), and the row above shows results for JHI-G (and not JHI-H).	
P25L1-L2. Figure 7. Isn't it possible to show the results for all ten realizations at all locations with some color code, as you did for figure 6?	We did try that previously but the plots became very crabbed and difficult to read. We therefore prefer to keep the figure as it is and let Figure 8 summarize result statistics across the ten realizations.	Nothing changed.
P25L21-L26. Can you see a particular reason for that? If this can be observed for the 10 models, then it is not related to the geological heterogeneity. Points 8, 9 and 10 are very close to the pumping well what suggests an influence of the pumping, maybe related to grid discretization?	The ten system realizations have similarities because they were all generated from the same set of borehole information. This may be what causes the tendency to either overpredict or underpredict hydraulic head at a given point. Anyway, the mean prediction error for points 2, 8, 9, and 10 is not much different from the mean error for the other six head predictions. We have therefore deleted these lines.	Lines deleted.
P26L7. I think head recovery at location 1 are underpredicted, as they fall below the identity line and the model prediction is on the Y-axis. Please verify.	The reviewer is correct. The point tends to fall below the line.	We changed overpredict to underpredict.
P26L21. Replace above by below?		Done
P26L31. "Explanation for this predictive degradation is given above". Where? It is not clear.	The text will be modified for clarification.	(Explanation for this predictive degradation is given above) Changed to (Explanation for this predictive degradation is given above, in the end of the first paragraph of this subsection.)
P27L9-L13. I would remove, this is highly speculative. The results seem very similar for all models and JHI-G and SHI-G are as good as the other, or even slightly better. This also shows that the mean error is a poor choice here, because the overprediction of 8 points	After careful reconsideration we agree with the reviewer.	L9-11 deleted,

counterbalances the huge underprediction of the two other points for HI-T, leading to the smallest error. I would rephrase the whole interpretation of the discharge prediction to be coherent with figure 7.		
P27L30-P28L4. This behavior is expected. Using the pilot-point method as calibrating method, smoothing is introduced in the distribution of K. If it may be sufficient to explain steady-state head and discharge, it affects much more flow paths and therefore transport (and in this case recharge, as it is linked to hydraulic conductivity). To get reliable transport estimate, a more geologically realistic distribution of K should be obtained, i.e. one which is consistent with the prior.	We partially agree.	P28L1-4 changed to: "these poor predictive performances must be because the calibrated hydraulic conductivity field is too smooth (and the hydraulic connectivity therefore exaggerated) in the calibrated models."
P28L8. The framework is not new, the software might be.		We changed "framework" to "platform".
P29L15. "…erroneous when the sensitivity of the TEM data with respect to resistivity is low". In this case, geophysical data should be disregarded. See Beaujean et al. (2014) for an example of model filtering for sequential hydrogeological inversion using geophysical data.	We partly agree and have added and changed sentences.	We added: "Such uncertain, potentially very erroneous, resistivity estimates should be discarded (or filtered out) from the SHI or JHI." Furthermore we slightly changed some of the following sentences.
P29L18-23. This is actually common sense, basically what you suggest is to always perform synthetic modeling before acquiring data, which is a common practice in research. Also, the fact that the use of the chosen petrophysical relationship will lead to error in using inverted resistivities is straightforward and does not require a full synthetic case.	How should we reply to this? We realize that in principle it would have been better to estimate a spatially dependent relationship between resistivity and hydraulic conductivity during the inversion. We did try to estimate a depth dependent relationship (i.e. depth dependent a and b in eq. 1) as part of JHI but this did not improve prediction results – on the contrary. We decided not to mention this in the manuscript since it is already long.	As mentioned in our reply to the comment just above, we changed some of the text.

~~A framework for t~~Testing alternative ~~the~~ uses of 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, USA.}

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

Abstract

Despite that geophysics is being used increasingly, it is often unclear how and when the integration of geophysical data and models can best improve the construction and predictive capability of groundwater models. This paper ~~presents~~ uses a newly developed **HY**drogeophysical **TE**st-**B**ench (HYTEB) – which is a collection of geological, groundwater and geophysical modeling and inversion software – to ~~wrapped to make a platform for generation and consideration of multi-modal data for objective hydrologic analysis. It allows for flexible 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 relationships and approaches to correlating geophysical and hydrologic parameters. With HYTEB we We~~ ~~study~~ demonstrates 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 (clay). The synthetic three dimensional reference system is designed so there is a perfect relationship between hydraulic conductivity and electrical resistivity. For this system 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 (in most cases) geophysics-based regularization.

For the studied system and inversion approaches it is found that that resistivities estimated by sequential hydrogeophysical inversion (SHI) or joint hydrogeophysical inversion (JHI) should be used with caution as estimators of hydraulic conductivity or as regularization means for subsequent hydrological inversion. The limited groundwater model improvement obtained by using the geophysical data probably mainly arises from the way these~~this~~ data are used here: the alternative inversion approaches propagate geophysical estimation errors into the hydrologic model parameters. It was expected that JHI would compensate for this, but the hydrologic data were apparently insufficient to secure such compensation. With respect to reducing model prediction error, it depends on the type of prediction whether it has value to include geophysics in a joint or sequential hydrogeophysical 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~~our 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 impact 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. In this paper we will focus on 3D models typically used for decision support on large spatial scale (from tens to thousands of square-kilometers) with a heterogeneous and possibly complex geology. Deterministic groundwater modelling is generally used in such cases because the model simulation time is too long to make it feasible to use stochastic modeling. We will therefore mainly focus on deterministic groundwater modeling in the following.

A deterministic 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 whereby its reasonableness can be tested by comparing forward 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 prediction will be uncertain for various ~~a number of~~ reasons of which we will emphasize three. (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 1998; 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 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 Airborne Electromagnetic Methods (AEM) because they can be collected quickly, densely, and at a relatively low cost for the very large spatial coverage (Steuer et al. 2008; Viezzoli et al. 2010b;

Abraham et al. 2012; Sànchez et al. 2012; Refsgaard et al. 2014; Munday et al. 2015). 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; Abraham et al. 2012; Oldenborger et al. 2013), to assess aquifer vulnerability (Refsgaard et al. 2014; Foged et al. 2014a), to map saltwater intrusion (Fitterman and Deszcz-Pan 1998; Viezzoli et al. 2010b; Lawrie et al. 2012; Herckenrath et al. 2013b), and to map freshwater resources (Steuer et al. 2008; Sànchez et al. 2012; Munday et al. 2015). 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 geophysical data are used to define hydrostratigraphic units or subregions or to parameterize the model, the data are often inverted. 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; Moysey et al. 2005; Linde et al. 2006; Singha and Gorelick 2006; Singha and Moysey 2006; Hinnell et al. 2010).

The simplest inversion approach is sequential hydrogeophysical inversion (SHI). In the first step of this approach, the geophysical data are inverted independent of the hydrologic data or model. In the second step, the inverted geophysical properties are used to zonate or directly parameterize the hydrologic model (Hubbard et al. 1999; Dam and Christensen 2003; Seifert et al. 2007; Koch et al. 2009; Di Maio et al. 2013; Marker et al. 2015). This is based on the assumption that the geophysical responses are sensitive to some of the same structures and property distributions that the hydrologic data are sensitive to. Using the SHI approach has built-in challenges. In the first step, the geophysical inversions are typically stabilized by using regularization and smoothing constraints that do not reflect real physical conditions (Day-Lewis 2005; Linde et al. 2006; Singha and Gorelick 2006; Singha and Moysey 2006). Therefore one must be cautious when using such geophysical property estimates to infer hydraulic zones or property estimates to be used in the second step of the SHI (Day-Lewis 2005; Slater 2007; Hinnell et al. 2010). Furthermore, with the SHI approach used in the following the geophysical models cannot be easily updated to conform to the hydrologic observations. Such updating is allowed by the SHI approach of Dam and Christensen (2003).

Two inversion 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).

Applications of sequential and joint version of hydrologic and geophysical data using a petrophysical relationships in groundwater modeling have been demonstrated by Herckenrath et al. (2013a) and Vilhelmsen et al. (2014). Herckenrath et al. (2013a) were comparing a SHI with JHI for a large-scale groundwater model using ground-based EM data. Vilhelmsen et al. (2014) were demonstrating a method for joint inversion of aquifer test data, magnetic resonance sounding data, and ground-based

Field Code Changed

Formatted: English (U.S.)

Formatted: English (U.S.)

Field Code Changed

electromagnetic data. For synthetic benchmarking, both of these studies were using a simple model with few layers with constant parameter values, and they were evaluating the model performance by the means of improved model parameter values and reduction of parameter uncertainty. However, as concluded by e.g. Zhou et al. (2014) the goal of groundwater model calibration is not just parameter identification, but also to increase the model's prediction capability.

Petrophysical relationships between hydraulic conductivity and electrical resistivity are difficult to establish, because such translation tend to be site, scale and facies specific (Hyndman and Tronicke 2005; Slater 2007). Therefore methods have been developed that do not rely on a petrophysical relationship between a property of the geophysical model and the property of the hydrological model. For example, Marker et al. (2015) developed a deterministic SHI approach where spatially dense AEM data is first inverted; then combined with scarce lithological data from boreholes to form a clay fraction model for the sediments of the subsurface (Foged et al. 2014b); subsequently cluster analysis of electrical resistivity and clay-fraction values is used to generate one or more structural model realizations for the subsurface (Foged et al. 2014a); finally the structural model(s) are used in the groundwater model that is calibrated against hydrological data. He et al. (2014,2015) developed a similar methodology where the transition probability method is used to generate the structural model realizations from borehole data and AEM estimated resistivities. The methodologies of Marker et al. (2015) and He et al. (2015) were both used by them on large spatial scale. However, the number of calibrated parameters was kept small by assuming hydraulic conductivity to be constant within an entire deposit or structure. The methodologies are thus stochastic in term of generating structure but deterministic in term of representing and estimating property fields.

Full stochastic approaches have been developed. For example, Ruggeri et al. (2013) developed a Bayesian simulation approach to estimate the hydraulic conductivity field from measurements of geopysical parameters. So far the method has only been used for estimation along short profile lines (Ruggeri et al. 2013, 2014), and the Bayesian scheme has not yet been extended to also involve hydrological data and groundwater modeling. (Such extension will probably turn out to be yet computationally infeasible.) Furthermore, the methodology of Ruggeri et al. (2013) requires existence of a petrophysical relationship between electrical resistivity and hydraulic conductivity; the relationship does not need to be known in which case a data set of collocated measurements of electrical resistivity and hydraulic conductivity are required. Such a data set is rare for investigations of the systems and scale considered here. Stochastic inversion approaches have been developed that invert electricial resistivity (ER) and hydrological data. Irving and Singha (2010), for example, describe a Markov-chain-Monte-Carlo (McMC) method to estimate a binary hydraulic conductivity field from ER and tracer test data. So far this McMC method seems to be applicable only for small domains and very few unknown parameters (Irving and Singha, 2010).

In the following we therefore keep focus on use of geophysics in connection with determinstic groundwater modeling. We particularly focus on application and comparison of SHI and JHI when used in connection with groundwater investigation of large domains with three-dimensional heterogenous hydrogeological and geophysical systems.

~~It is intuitively clear that~~ spatially dense ~~geophysics can offer valuable information for improved groundwater modeling for decision making. However, many important questions are yet unanswered. For example: for a complex~~ 3D ~~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 Testing the worth of using geophysics~~Hydrogeophysical test-bench~~

It is intuitively clear that spatially dense geophysics can offer valuable information for improved groundwater modeling for decision making. However, many important questions can be raised. For example: how beneficial is it to collect and use EM data in groundwater modeling for a complex 3D hydrogeological system; how much can be gained by using for example JHI instead of SHI for model calibration; are both inversion approaches prone to lead to biased parameter estimates or model predictions; and what model prediction types will benefit from using EM data in connection with the model development and calibration. The answers to these questions will to some (large?) extend depend on the actual hydrogeological setting as well as on what types of prediction are going to be made by the groundwater model.

~~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.~~

To ~~allow a thorough examination~~provide such answers we have developed a cross-disciplinary, flexible ~~framework~~ platform to ~~objectively~~ examine the worth of geophysical data for improvement of groundwater model predictions in potentially complex environments. The ~~idea~~ platform ~~is~~ can be used 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~~ platform allows easy investigation of the data worth when using alternative data sampling and alternative modeling or inversion strategies. Because of the supposed similarity between the synthetic and the actual systems, the conclusions from the synthetic study can be transferred to actual investigation. The ~~framework~~ platform 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. HYTEB can also support examination of use of geophysics in a stochastic groundwater modeling context (which will be demonstrated in a manuscript in preparation).

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, Section 3 describes the case study, Section 4 presents the results, while Section 5 makes a summary and draw conclusions.

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

Our primary objective in developing ~~HYTEM~~ 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 systems are known and the influences of different sources of error can be investigated. We use physical and geophysical forward simulators 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 Figure 1~~Figure 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~~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 bedrock (not shown). Such categorical geological settings can, for example, be 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 reference groundwater system, 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 Figure 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 (for example forecasts) simulated for this system are called "reference predictions".

2.3 Step 3 – Generation of reference geophysical system and 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). Both positive and negative relationships have been reported, and there can be significant uncertainty in the relationship (e.g. Mazáč et al. 1985). 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 method specific and random error ~~to represent all sources of measurement error~~. Ideally a 3D code should be used. Codes for 3D computation of TEM responses have 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 1D code, where the 1D geophysical model is created from the reference system by pseudo-3D sampling, that is by taking the logarithmic average of the cells within the radius of the EM foot print at a given depth. Modeling TEM in 1D can be problematic in connection with mineral exploration, but for sedimentary environments the 1D approach should work well (Auken et al. 2008; Viezzoli et al. 2010a). In HYTEB we use AarhusInv (Auken et al. 2014) to simulate electromagnetic instrument responses.

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

2.4 Step 4 – Model construction and parameterization

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, it may be argued that 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 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, for example zones, pilot points, or combinations hereof.

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 reference 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.

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 of step 2. (A prediction is a state-variable different from the calibration data, for example a forecast.) For each prediction, this produces 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, 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 reference 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)

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 bedrock (fat clay) with very low electrical resistivity (Wright 1973; Piotrowski 1994; Clayton et al. 1999; Jørgensen and Sandersen 2006). The case is designed to have a perfect relationship between hydraulic conductivity and electrical resistivity. This is chosen to make a best possible case for resolving change of lithology and change of hydraulic conductivity ~~from measurements~~inferred from of 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 Figure 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 degrees and runs in the N-S direction from the coast and 5 km inland (Doherty and Christensen, 2011, used a steeper 21 degree 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 (Figure 2~~Figure 2~~). This borehole stratigraphy was used to condition all generated geological system realizations.

Realizations of the 3D geological model were generated on a uniform rectangular grid. The cells of the grid have horizontal dimensions of 25 m x 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 3D 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~~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~~Figure 3~~ shows one of these realizations.

3.2 Reference 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=2487.5m and y=1912.5 and the pumping rate is 0.015 m^3s^{-1}. The well screens the deepest 10 meters 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 simulation method (Deutsch and Journel 1998) with the geostatistical parameters given in Table 1~~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~~Table 2~~; Figure 2~~Figure 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 MODFLOW-2000 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 Reference 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 for groundwater model calibration is likely affect the error of ~~to have different effects on~~ model predictions differently~~different model predicions of interest.the uncertainties of different predictions of interest~~. To illustrate this, we present ~~six~~ seven types of predictions of interest in Table 3~~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 (Figure 2~~Figure 2~~ and Table 4~~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 Figure 2~~Figure 2~~ and Table 4~~Table 4~~. Type 5 is the travel time of a particle flowing with the groundwater from the location where 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 x 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) from the 25 x 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 Reference geophysical system and data – step 3

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

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

where *K* is the hydraulic conductivity (m/s) derived from resistivity, ρ is the electrical resistivity (ohm-m), and $\beta_1 = log_{10}(1e^{-12})$ and $\beta_2 = log_{10}(4)$ are empirical shape 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. Eq. (1)~~(1)~~ was used to compute the resistivity within each cell of the geological system from the corresponding cell hydraulic conductivity.

Using a perfect relationship to generate resistivity from ~~between~~ hydraulic conductivity ~~and resistivity~~ must be characterized as the ideal case because in this case electrical resistivity data can be expected to provide maximal information about hydraulic conductivity. In practive, ~~W~~when possible, estimation of hydraulic conductivity from electrical resistivity is usually based on a site specific noisy 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.)~~

3.3.1 Geophysical data set

It is assumed that measurements of the geophysical system are conducted at 77 uniformly distributed locations within the domain (Figure 2~~Figure 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 × 40 m^2 square transmitter loop. Measurements are gathered from about 10 microseconds to 10 milliseconds using a steady current of 20 Amperes, which gives a magnetic moment of 32000 Am^2 which, for the studied environment, would provide a penetration depth of around 250 meters (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 degrees (West GF and Macnae JC 1991). In other words, the sounding loses resolution with depth because of its increasing footprint. In the following, we use the 1D simulation code AarhusInv (previously called em1Dinv; Auken et 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 consisting of: (i) Gaussian 3 % noise contribution and (ii) "background" contribution with a value of 3nV/m^2 according to the noise model suggested by (Auken et al. 2008). The noise-perturbed data were subsequently processed as field data (Auken et al. 2009).

3.4 Model construction and parameterization (Step 4)

The groundwater model uses the true boundary conditions except that recharge is to be estimated together with hydraulic conductivity. Because the reference groundwater and geophysical systems were 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 (Figure 2~~Figure 2~~). Kriging is used for spatial interpolation (here using the correct correlation lengths) from the pilot points to the model grid. This kind of parametrization creates smooth transition in hydraulic conductivity which may seem problematic to use in the current case where there are "categorical" (lithological) shifts in the reference fields. However, because the property contrasts between categories are so large and the geophysical data and the pilot points so many, it is expected that the categorical shifts in property value can be fairly well resolved by the used interpolation.

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~~Electrical resistivity is sensitive to porosity, but that is not incorporated in the relationship within the present work. Therefore porosity cannot be estimated from the hydrological and geophysical data available here, we always use the reference 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 1D, following the state of practice (Viezzoli et al. 2010a; Auken et al. 2014)

3.5 Model calibration by inversion (step 5)

Formatted: English (U.S.)

~~Traditionally, c~~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

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

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; n_h and n_r are the number of head, discharge observations, respectively. However, equation (2)~~(2)~~ 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

$$\Phi_t = \Phi_m + \mu \cdot \Phi_r \qquad (3)$$

where ϕ_t is the total objective function, ϕ_m is the measurement objective function given by (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, based on a linearized model approximation, during each optimization iteration making ϕ_m equal to a user specified target value (for details, see Doherty 2010). In this case, for ϕ_m defined by (~~4~~2), 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).~~

Method 2 is joint hydrogeophysical inversion (JHI). For JHI, the hydrological model and the geophysical models are set up separately, but hydrological and geophysical parameters are estimated simultaneously ~~through~~ by minimization of a joint objective function where ~~. The hydrologic and geophysical models are joint directly through using~~ the regularization term uses an assumed relationship between electrical resistivity and hydrological ~~resistivity~~conductivity. The minimized objective function is of the same form as (3)~~(3)~~, but the measurement and regularization terms are different. For Method 2 the measurement objective function is defined as

$$\Phi_{m,joint} = {n_h}^{-1}\sum_{i=1}^{n_h}\left(\frac{h_{obs,i}-h_{sim,i}}{\sigma_{h,i}}\right)^2 + {n_r}^{-1}\sum_{i=1}^{n_r}\left(\frac{r_{obs,i}-r_{sim,i}}{\sigma_{r,i}}\right)^2 + {n_{tem}}^{-1}\sum_{i=1}^{n_{tem}}\left(\frac{V_{obs,i}-V_{sim,i}}{\sigma_{tem,i}}\right)^2 \qquad (4)$$

where n_{TEM} are of TEM observations, respectively. The first two terms on the right hand side of equation (4)~~(4)~~ are identical to the terms in (2)~~(2)~~. 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 equation (4)~~(4)~~ 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

$$\Phi_{r,joint} = \mu \cdot \sum_{i=1}^{n_{kpar}}\left(\log_{10}(K_{joint,i}) - \log_{10}(K_{mf,i})\right)^2 \qquad (5)$$

In (5)~~(5)~~, $K_{mf,i}$ is the estimate of the hydraulic conductivity at the i[th] pilot point of the groundwater model; $K_{joint,i}$ is also an estimate of hydraulic conductivity but inferred ~~, but this estimate is calculated~~

from the estimated electrical resistivity at the same depth and location by using equation (1)~~(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~~modified from Dam and Christensen (2003). First, the geophysical model parameters (electrical resistivities) are estimated on the basis of the geophysical data. Secondly, the groundwater model 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 1D models are inverted independently using AarhusInv (Auken et al. 2014), but vertical constraints were used to stabilize the inversion of each 1D model (Constable et al. 1987). In the second step, the estimated electrical resistivities are used to constrain the subsequent hydrologic inversion, which is carried out as minimization of equation (3) where the measurement objective function ϕ_m is defined by equation (2) while the preferred difference regularization term is defined by

$$\Phi_{r,seq} = \mu \cdot \sum_{i=1}^{n_{kpar}} \omega_i \left(\log_{10}(K_{seq,i}) - \log(K_{mf,i})\right)^2 \qquad (6)$$

As in (5)~~(5)~~, $K_{mf,i}$ 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 equation (1)~~(1)~~. In (6), ω_i is a weight that can be varied between the terms of the summation in (6). The results presented in the following were obtained by using a weight of 1.0 for all preferred differences. (We also tried using weights determined as $\omega_i = \left(\beta_2 V(\rho_i)\right)^{-1}$ where $V(\rho_i)$ is the variance of the log-resistivity estimate used to compute $\log_{10}(K_{seq,i})$. Changing the weights did not change the estimation results much; some prediction errors did reduce, others increased, but we found no general improvement by using $\omega_i = \left(\beta_2 V(\rho_i)\right)^{-1}$ instead of the simple choice $\omega_i = 1$.) ~~In this case~~For method 3, 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 reference (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 supposedly best possible outcome of HI. In the second run, termed HI-H, we assumed a homogeneous initial hydraulic conductivity field with K equal to $1 \times 10^{-6}\ m/s$ 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 reference (true) parameter values for hydraulic conductivity and electrical resistivity at the pilot points. As above this is done to show the supposedly best possible outcome of JHI. In the second run, termed JHI-H, we used a constant hydraulic conductivity of $1 \times 10^{-6}\ m/s$ and a constant electrical resistivity of 40 ohm-m 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 ohm meter 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 (1) to produce the JHI-G initial values of hydraulic conductivity at the hydraulic conductivity pilot points.

For method 3 (SHI), we present results from only ~~ran~~ one inversion sequence, termed SHI-G. First we ran the independent geophysical inversions using a homogeneous half space of 40 ohm meter as the initial model. Subsequently we used the estimated resistivities together with relation (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. (As for JHI, we also tried using an initially homogeneous hydraulic conductivity field for SHI; this gave a more blurred estimated hydraulic conductivity field than what is presented later. So, like it is shown later for JHI, the result of SHI was found to depend on the choice of initial parameter values.)

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 that different parts of the sensitivity matrix 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~~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.

4 Results

4.1 Model Calibration

Figure 4~~Figure 4~~ shows the measurement objective function value, Φ_m, obtained for the various groundwater model calibration cases and for the 10 different system realizations. It also shows the separate terms of the objective function. We aimed at using weights that would make each term contribute by a value of approximately 1.0. For HI and SHI there are two terms, quantifying fit to head data and fit to the flux measurement, respectively; the results in Figure 4 show that the head data are fitted as intended while the flux measurement is fitted more closely than intended. This fitting picture is also seen for JHI. JHI tends to produce better fit to the hydrologic data than HI and SHI.

For JHI the objective function (4)~~(4)~~ has a third term quantifying fit to decay data of the TEM measurements. Figure 4 indicates that the actually used weighting for JHI ended by producing slightly better fit to the hydrologic data than to the TEM data. It also shows that for JHI the fit to the hydrologic data is not strongly dependent on the choice of initial parameter values; JHI-T for example did not always produce better fits than JHI-G or JHI-H. That JHI-T, JHI-G, and JHI-H lead to different fits (and different parameter estimates) shows that the JHI minimization problem may not be unique. However, we did not investigate if PEST parameters could have been set differently to thereby make the JHI minimization unique.

For two realizations HI-T produced much worse fit to the hydrologic data than HI-H (Figure 4~~Figure 4~~): the HI-T minimization got stuck at a local minimum where a parameter adjustment improving the fit to head deteriorated the fit to the flux measurement. We did not investigate if PEST parameters could have been set differently to overcome this problem.

4.2 Estimated hydraulic conductivity fields

Figure 5~~Figure 5~~ shows the reference hydraulic conductivity fields of the uppermost six layers and a representative cross section for one of the 10 chosen system realizations. 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 system. Figure 6~~Figure 6~~ shows corresponding pilot-point-by-pilot-point scatter plots of reference versus estimated hydraulic conductivity. Except when noted specifically, the results in Figure 5~~Figure 5~~ and Figure 6~~Figure 6~~ for this realization are typical for all 10 chosen system realizations.

The second and third rows of Figure 5~~Figure 5~~ show results for the two hydrologic inversion (HI) runs. Inversion HI-T, which used true (reference) 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 reference field but the estimated values vary much less than the reference 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 Figure 5~~Figure 5~~ illustrates that for inversion HI-H, which used homogeneous initial hydraulic conductivity fields, there is almost no structural~~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 6~~Figure 6~~ shows that both HI-T and HI-H underestimate hydraulic conductivities 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 Figure 5~~Figure 5~~ shows hydraulic conductivity fields estimated by the sequential geophysical approach (SHI-G). For the upper layers, the true (reference) 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 6~~Figure 6~~ shows that for low-permeability deposits, the range of estimated log-hydraulic conductivities is twice as large as the reference 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 reference values, and the estimated values tend to be orders of magnitude too small (Figure 6~~Figure 6~~). 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 (Figure 6~~Figure 6~~) 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 reference electrical 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 Figure 5~~Figure 5~~ 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 (reference) 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 6~~Figure 6~~ shows that estimated versus reference hydraulic conductivity values plot nicely along the identity line for JHI-T. The resolution of structures (Figure 5~~Figure 5~~) and the quality of the K estimates (Figure 6~~Figure 6~~) deteriorate for JHI-H and JHI-G, both of which use less informative initial parameter values. Figure 5~~Figure 5~~ visually indicates that JHI-G resolves structures better than JHI-H. For sand and gravel deposits Figure 6~~Figure 6~~ 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 reference values vary within five orders of magnitude.

4.3 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 section 3.2.2. Figure 7~~Figure 7~~ 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 Figure 7~~Figure 7~~ 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 pumping well (Figure 2~~Figure 2~~), groundwater discharge to the river after pumping has stopped, and recharge area of the pumping well.

Figure 8~~Figure 8~~ 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 model predicted value for each prediction of interest averaged over the ten geological realizations considered. The prediction results are discussed individually below.

4.3.1 Head prediction

All calibrated groundwater models appear to be fairly good predictors of hydraulic head. Nearly ~~U~~unbiased head prediction is exemplified by the plots in the first column of Figure 7~~Figure 7~~ 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 appears to be 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 (Figure 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 8~~Figure 8~~ 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 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.

4.3.2 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 Figure 7~~Figure 7~~. The two plots in the top of the second column indicate that head recovery at location 1 tends to be ~~overpredicted~~ underpredicted by the models calibrated by purely hydrologic inversion (HI-T and HI-H). The third plot in this column indicates that SHI-G slightly reduces the bias seen in the HI-based model. Finally, the last three plots in the second column of Figure 7~~Figure 7~~ 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 section 2.7). The appearances of scatter plots for head recovery at locations 2 to 7 are similar to that for recovery at location 1 (Figure 7~~Figure 7~~).

The second plot in the third column of Figure 7~~Figure 7~~ 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~~ below the identity line while the tenth point falls far above the identity line. Generally, the plots indicate a consistent ~~overprediction~~ underprediction of head using HI-based inversion. The remaining plots in the third column show that recovery prediction at location 8 turns out to be too large for the models calibrated with geophysical data (JHI) or by using geophysics based regularization (SHI).

Figure 8~~Figure 8~~ 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 Figure 7~~Figure 7~~. On the contrary, for recovery prediction 8, located within the buried valley, both Figure 7~~Figure 7~~ and Figure 8~~Figure 8~~ show that including the geophysics in the groundwater modelling 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, in the end of the first paragraph of this subsection.) 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.

4.3.3 Discharge prediction

The scatter plots in the fourth column of Figure 7~~Figure 7~~ 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 by~~in~~ the SHI or JHI inversions used here (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 reference (true) values (JHI-T). That use of geophysical data is not important to improve this prediction~~This is confirmed by the relative error magnitudes for discharge shown in Figure 8~~Figure 8~~.

4.3.4 Recharge area and other particle tracking predictions

The plots in the fifth column of Figure 7~~Figure 7~~ 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 realization 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 Figure 7~~Figure 7~~ 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 must be because the calibrated hydraulic conductivity field is too smooth (and the hydraulic connectivity therefore exaggerated) in the calibrated models. ~~is similar to that just given for the prediction of the well's recharge area. Figure 7 and~~ Figure 8~~Figure 8~~ show that use of TEM data does not improve the model performance with respect to prediction of groundwater age and particle travel time.

5 Discussion 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 paper presents a newly developed ~~framework~~ platform that allows for such an application- and method-specific examination of the potential value of using geophysical data and models to develop a groundwater model and improve its predictive power. We call the ~~framework~~ platform ~~a~~ **HY**drogeophysical **TE**st-**B**ench (HYTEB). HYTEB allows for 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 used to correlate geophysical and hydrologic parameters. We use HYTEB to work with rather complex, fairly realistic but synthetic systems. In this work we strive at (and recommend) balancing complexity with the advantage of knowing the 'true' system or condition to assess model/data performance, and at avoiding to overextend the likely value of data or models beyond the tested conditions.

Our recommended way 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. HYTEB is used to investigate to what extent groundwater model calibration and, often more importantly, model predictions can be improved for this kind of setting by including in the calibration process electrical resistivity estimates obtained from TEM data in two different ways: by using either sequential hydrogeophysical inversion (SHI) or joint hydrogeophysical inversion (JHI). 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.)~~ The results are compared to those obtained by a groundwater model calibrated by purely hydrologic inversion (HI).

The calibrated groundwater models are parameterized by many pilot points that should allow a reasonable resolution of the hydraulic and geophysical property fields at depths where the properties are resolved by the data. Using PEST (Doherty, 2010), Tikhonov or geophysical regularization is used to stabilize the HI, SHI, and JHI inversion problems. In this case, JHI tends to produce the best fit to the data while SHI and HI produce comparable fits.

For HI, the estimated hydraulic conductivity field turns out to be very smooth in the top layers and almost homogeneous in the deeper layers, which is expected for this type of (Tikhonov) regularization. For SHI and JHI, the estimated hydraulic conductivity field resolves much of the true 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 for the geophysical models either in the first step of the SHI scheme or during the JHI scheme can turn out to be very erroneous when the sensitivity of the TEM data with respect to resistivity is low. Such uncertain, potentially very erroneous, resistivity estimates should be discarded (or filtered out) from the SHI or JHI. ~~For the studied system, this shows~~By not doing this, we showed 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 electrical resistivity led to an over-reliance on the use of inferred resistivities to populate the model's hydraulic conductivity field. ~~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 geophysics in the model calibration. It was found that all models are good predictors of hydraulic head. However, head prediction errors tend to be reduced for models calibrated by SHI or JHI as compared to models calibrated by 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 or models (with JHI or SHI) 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 change the most). Analyzing the prediction results by the method described by Doherty and Christensen (2011) indicates that geophysics 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 method 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 geophysics. 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 here 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 these~~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~~ this ~~HYTEB analysis~~ study is that recharge should be parameterized and estimated in a different way than it was done in the demonstration example. Alternatively 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. HYTEB is available for download on https://github.com/Nikolaj-KC/HYTEB. With HYTEB can also be downloaded Python scripts demonstrating how to set up HI, SHI and JHI as it is done in the manuscript.

Formatted: Hyperlink, Font color: Auto

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

6 References

Abraham JD, Cannia JC, Bedrosian PA, Johnson MR, Ball LB, Sibray SS (2012) : Airborne Electromagnetic Mapping of the Base of Aquifer in Areas of Western Nebraska. In: U.S. Geol. Surv. Sci. Investig. Rep. 2011–5219. http://pubs.usgs.gov/sir/2011/5219/. Accessed 4 Jan 2016

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

Auken E, Christiansen AV, H.Westergaard J, Kirkegaard C, Foged N, Viezzoli A (2009) : An integrated processing scheme for high-resolution airborne electromagnetic surveys, the SkyTEM system. Explor Geophys 40(2):184–192. doi: http://dx.doi.org/10.1071/EG08128

Auken E, Christiansen AV, Kirkegaard C, Fiandaca G, Schamper C, Behroozmand AA, Binley A, Nielsen E, Effersø F, Christensen NB, Sørensen K, Foged N, Vignoli G (2014) : An overview of a highly versatile forward and stable inverse algorithm for airborne, ground-based and borehole electromagnetic and electric data. Explor Geophys 46(3):223–235. doi: 10.1071/EG13097

Auken E, Christiansen A V., Jacobsen LH, Sørensen KI (2008) : A resolution study of buried valleys using laterally constrained inversion of TEM data. J Appl Geophys 65:10–20.

Auken E, Jørgensen F, Sørensen KI (2003) : Large-scale TEM investigation for groundwater. Explor Geophys 34(3):188–194. doi: 10.1071/EG03188

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

Carle SF (1999) T-PROGS: Transition Probability Geostatistical Software. Users Manual.

Certes C, De Marsily G (1991) : 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

Clayton L, Attig JW, Mickelson DM (1999) Tunnel channels formed in Wisconsin during the last glaciation. Geological Society of America

Constable SC, Parker RL, Constable CG (1987) : Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data. GEOPHYSICS 52:289–300. doi: 10.1190/1.1442303

Cooley RL (2004) : A theory for modeling ground-water flow in heterogeneous media. In: US Geological Survey Professional Paper 1679, 220 pp. US Geological Survey Report,

Cooley RL, Christensen S (2006) : 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

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

Danielsen JE, Auken E, Jørgensen F, Søndergaard V, Sørensen KI (2003) : The application of the transient electromagnetic method in hydrogeophysical surveys. J Appl Geophys 53:181–198. doi: 10.1016/j.jappgeo.2003.08.004

Day-Lewis FD (2005) : Applying petrophysical models to radar travel time and electrical resistivity tomograms: Resolution-dependent limitations. J Geophys Res 110:B08206. doi: 10.1029/2004JB003569

Deutsch C V. (2006) : 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

Deutsch C V., Journel AG (1998) GSLIB: Geostatistical Software Library and User's Guide: Clayton V. - Oxford University Press, Second Edi. Oxford University Press

Di Maio R, Fabbrocino S, Forte G, Piegari E (2013) : 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

Doherty J (2010) PEST, Model-Independent Parameter Estimation, User Manual, 5th ed, 336 pp.,. Watermark Numerical Computing

Doherty J (2003) : Ground Water Model Calibration Using Pilot Points and Regularization. Ground Water 41:170–177. doi: 10.1111/j.1745-6584.2003.tb02580.x

Doherty J, Christensen S (2011) : Use of paired simple and complex models to reduce predictive bias and quantify uncertainty. Water Resour Res 47(12):W12534. doi: 10.1029/2011WR010763

Doherty J, Welter D (2010) : A short exploration of structural noise. Water Resour Res 46 (5):W05525. doi: 10.1029/2009WR008377

Feyen L, Gómez-Hernández JJ, Ribeiro PJ, Beven KJ, De Smedt F (2003) : A Bayesian approach to stochastic capture zone delineation incorporating tracer arrival times, conductivity measurements, and hydraulic head observations. Water Resour Res 39 (5):1126. doi: 10.1029/2002WR001544

Feyen L, Gorelick SM (2005) : Framework to evaluate the worth of hydraulic conductivity data for optimal groundwater resources management in ecologically sensitive areas. Water Resour Res 41 (3):W3019. doi: 10.1029/2003WR002901

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

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

Foged N, Marker PA, Christansen A V., Bauer-Gottwein P, Jørgensen F, Høyer A-S, Auken E (2014b) : 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

Franssen H-JH, Gómez-Hernández J, Sahuquillo A (2003) : 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

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

Frohlich RK, Kelly WE (1985) : 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

Harbaugh AW, Banta ER, Hill MC, McDonald MG (2000) MODFLOW-2000, The U.S. Geological Survey modular ground-water model: User guide to modularization concepts and the ground-water flow process. U.S. Geological Survey Open-File Report 00-92, 121 p.

Harvey CF, Gorelick SM (1995) : 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

He X, Højberg AL, Jørgensen F, Refsgaard JC (2015) : Assessing hydrological model predictive uncertainty using stochastically generated geological models. Hydrol Process 29:n/a–n/a. doi: 10.1002/hyp.10488

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

Heigold PC, Gilkeson RH, Cartwright K, Reed PC (1979) : Aquifer Transmissivity from Surficial Electrical Methods. Ground Water 17:338–345. doi: 10.1111/j.1745-6584.1979.tb03326.x

Herckenrath D, Fiandaca G, Auken E, Bauer-Gottwein P (2013a) : 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

Herckenrath D, Odlum N, Nenna V, Knight R, Auken E, Bauer-Gottwein P (2013b) : Calibrating a salt water intrusion model with time-domain electromagnetic data. Ground Water 51 (3):385–397. doi: 10.1111/j.1745-6584.2012.00974.x

Hill M (1998) : 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 Investig Rep 98–4005. doi: 10.1061/40517(2000)18

Hinnell a. C, Ferré TP a., Vrugt J a., Huisman J a., Moysey S, Rings J, Kowalsky MB (2010) : Improved extraction of hydrologic information from geophysical data through coupled hydrogeophysical inversion. Water Resour Res 46:W00D40. doi: 10.1029/2008WR007060

Hubbard SS, Rubin Y (2000) : 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

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

Hyndman D., Tronicke J (2005) : Hydrogeophysical case studies at the local scale: the saturated zone. In: Rubin Y, Hubbard SS (eds) Hydrogeophysics. Springer Netherlands, Dordrecht, pp 391–412

Hyndman DW, Harris JM, Gorelick SM (1994) : Coupled seismic and tracer test inversion for aquifer property characterization. Water Resour Res 30:1965–1977. doi: 10.1029/94WR00950

Jørgensen F, Sandersen PBE (2006) : Buried and open tunnel valleys in Denmark—erosion beneath multiple ice sheets. Quat Sci Rev 25:1339–1363.

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

Koch K, Wenninger J, Uhlenbrook S, Bonell M (2009) : 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

Kowalsky MB, Finsterle S, Peterson J, Hubbard S, Rubin Y, Majer E, Ward A, Gee G (2005) : Estimation of field-scale soil hydraulic and dielectric parameters through joint inversion of GPR and hydrological data. Water Resour Res 41 (11):W11425. doi: 10.1029/2005WR004237

Lawrie KC, Carey H, Christensen NB, Clarke J, Lewis S, Ivkovic KM, Marshall SK (2012) : Evaluating the Role of Airborne Electromagnetics in Mapping Seawater Intrusion and Carbonate-Karstic Groundwater Systems in Australia. Geoscience Australia, Canberra, http://dx.doi.org/10.11636/Record.2012.042

Linde N, Finsterle S, Hubbard S (2006) : Inversion of tracer test data using tomographic constraints. Water Resour Res 42:n/a–n/a. doi: 10.1029/2004WR003806

Marker PA, Foged N, He X, Christiansen A V., Refsgaard JC, Auken E, Bauer-Gottwein P (2015) : 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

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

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

Moysey S, Singha K, Knight R (2005) : A framework for inferring field-scale rock physics relationships through numerical simulation. Geophys Res Lett 32:L08304. doi: 10.1029/2004GL022152

Munday T, Gilfedder M, Taylor andrew r, Ibrahimi T, Ley-cooper Y, Cahill K, Smith S, Costar A (2015) : The role of airborne geophysics in facilitating long-term outback water solutions to support mining in South Australia. Water - J Aust Water Assoc 42:138–141.

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

Oldenborger GA, Pugin AJ-M, Pullan SE (2013) : Airborne time-domain electromagnetics, electrical resistivity and seismic reflection for regional three-dimensional mapping and characterization of the Spiritwood Valley Aquifer, Manitoba, Canada. Near Surf Geophys 11:63–74. doi: 10.3997/1873-0604.2012023

Oliver DS, Reynolds AC, Liu N (2008) Inverse Theory for Petroleum Reservoir Characterization and History Matching. Cambridge University Press; University Printing House, Cambridge CB2 8BS, United Kingdom

Piotrowski JA (1994) : Tunnel-valley formation in northwest Germany—geology, mechanisms of formation and subglacial bed conditions for the Bornhöved tunnel valley. Sediment Geol 89:107–141.

Pollock DW (1994) User ' s Guide for MODPATH / MODPATH-PLOT , Version 3 : A particle tracking post-processing package for MODFLOW , the U . S . Geological Survey finite-difference ground-water flow model.

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

Refsgaard JC, Auken E, Bamberg CA, Christensen BSB, Clausen T, Dalgaard E, Effersø F, Ernstsen V, Gertz F, Hansen AL, He X, Jacobsen BH, Jensen KH, Jørgensen F, Jørgensen LF, et al (2014) : 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

Refsgaard JC, Christensen S, Sonnenborg TO, Seifert D, Højberg AL, Troldborg L (2012) : 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

Reilly TE (2001) : 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. U.S. Geological Survey, Denver, CO, USA.,

Reilly TE, Harbaugh AW (2004) Guidelines for Evaluating Ground-Water Flow Models. U.S. Geological Survey Scientific Investigations Report 2004-5038—Version 1.01

Revil A, Cathles LM (1999) : Permeability of shaly sands. Water Resour Res 35:651–662. doi: 10.1029/98WR02700

Sànchez M, Gunnink JL, van Baaren ES, Oude Essink GHP, Siemon B, Auken E, Elderhorst W, de Louw PGB (2012) : 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

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

Seifert D, Sonnenborg TO, Scharling P, Hinsby K (2007) : 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

Singha K, Gorelick SM (2006) : Effects of spatially variable resolution on field-scale estimates of tracer concentration from electrical inversions using Archie's law. GEOPHYSICS 71:G83–G91. doi: 10.1190/1.2194900

Singha K, Moysey S (2006) : Accounting for spatially variable resolution in electrical resistivity tomography through field-scale rock-physics relations. GEOPHYSICS 71:A25–A28. doi: 10.1190/1.2209753

Slater L (2007) : 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

Steuer A, Siemon B, Eberle D (2008) : 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

Tonkin M, Doherty J, Moore C (2007) : Efficient nonlinear predictive error variance for highly parameterized models. Water Resour Res 43 (7):W07429. doi: 10.1029/2006WR005348

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

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

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

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

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

West GF, Macnae JC (1991) Physics of the Electromagnetic Induction Exploration Method. In: Electromagnetic Methods in Applied Geophysics, Part A, edited by Nabighian, M.N., Society of Exploration Geophysicists, Tulsa.

Worthington PF (1975) : Quantitative geophysical investigations of granular aquifers. Geophys Surv 2:313–366. doi: 10.1007/BF01447858

Wright HE (1973) The Wisconsinan Stage. Geological Society of America

Zhou H, Gómez-Hernández JJ, Li L (2014) : Inverse methods in hydrogeology: Evolution and recent trends. Adv Water Resour 63:22–37. doi: 10.1016/j.advwatres.2013.10.014

Table 1. Geostatistical parameters for stochastic hydraulic field employed by the hydraulic reference model. K is for hydraulic conductivity (ms^{-1}), R is for recharge (ms^{-1}) to the groundwater model, and phi for porosity. μ is mean value to the log10 of K, a is range for small-scale variability, and σ^2 the sill. The semivariograms are exponential.

	$log_{10}(K)$			$log_{10}(R)$			$log_{10}(\varphi)$		
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

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

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

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

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

Head pred. point	Location X(m)	Location Y(m)	Screen	Head pred. point	Location X(m)	Location 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

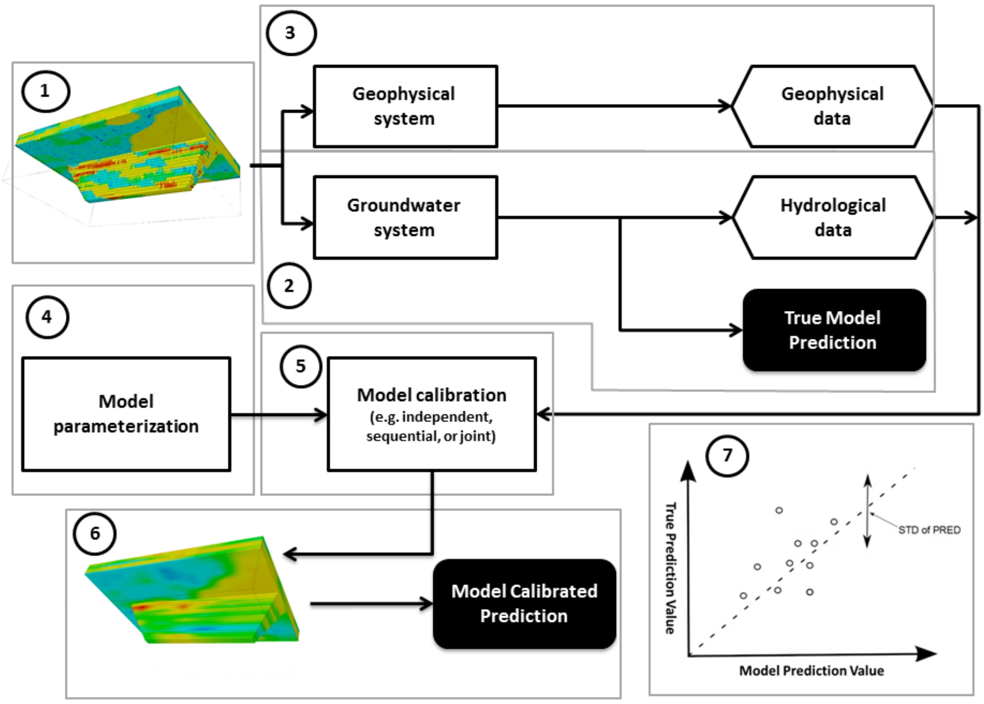

Figure 1: Workflow of the HYTEB. Each numbered dashed box marks a major step in the work flow. In parts 1 and 5 the red, yellow, blue and green colors indicate different categories (types) of geological deposits; color variation within each category (in part 6) indicates variation in hydraulic conductivity

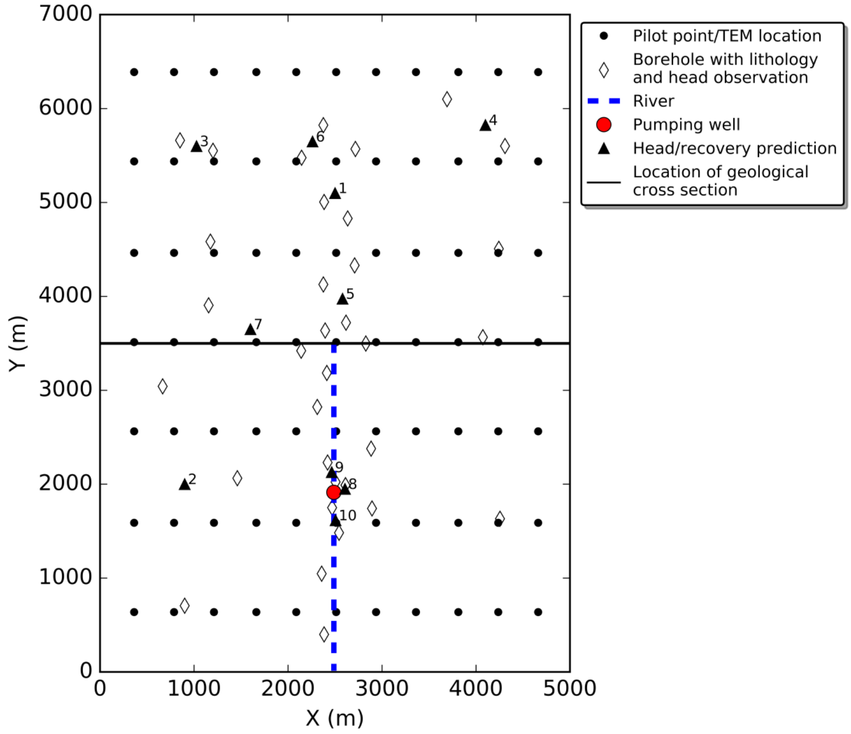

Figure 2: A map of locations of boreholes, a pumping well, geophysical data, pilot points, predictions of interest and location of a geological cross-section. (The positions of the pilot points and geophysical measurements are coincident.)

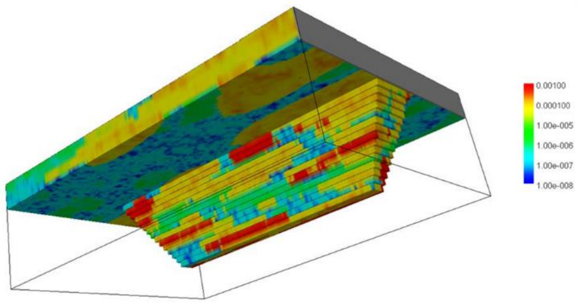

Figure 3. Hydraulic conductivity field for one of the model realizations. (Red shades are for gravel, yellow for sand, green for silt, and cyan/blue for clay.)

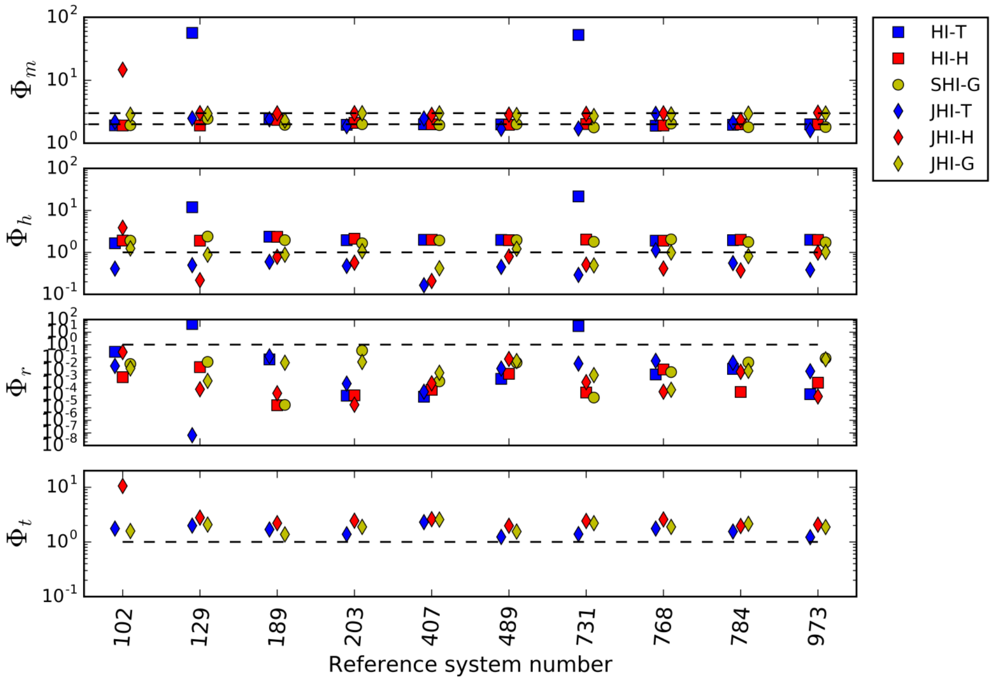

Figure 4: Measurement objective function value obtained for the various groundwater model calibration cases and for the 10 different system realizations. The two dashed lines in the top plot indicate the target value for the various model calibrations: the upper dashed line is the target value for the JHI, and the lower dashed line is the target value for HI and SHI. The dashed line in the three lower plots similarly marks the respective target value.

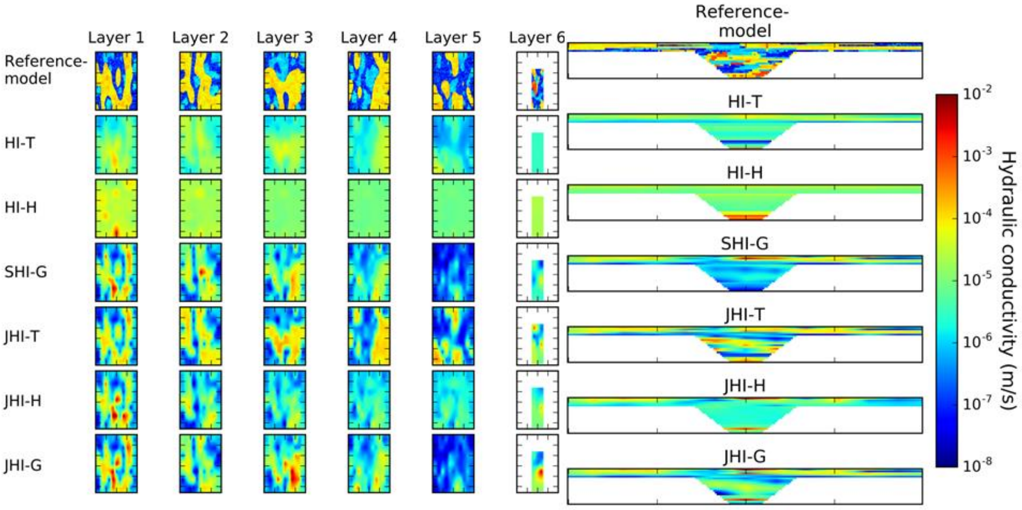

Figure 5: Reference and estimated hydraulic conductivity fields for model realization number 189: a) shows the fields for layers 1 to 6 ; b) shows the field along an east-west cross section in the middle of the domain.

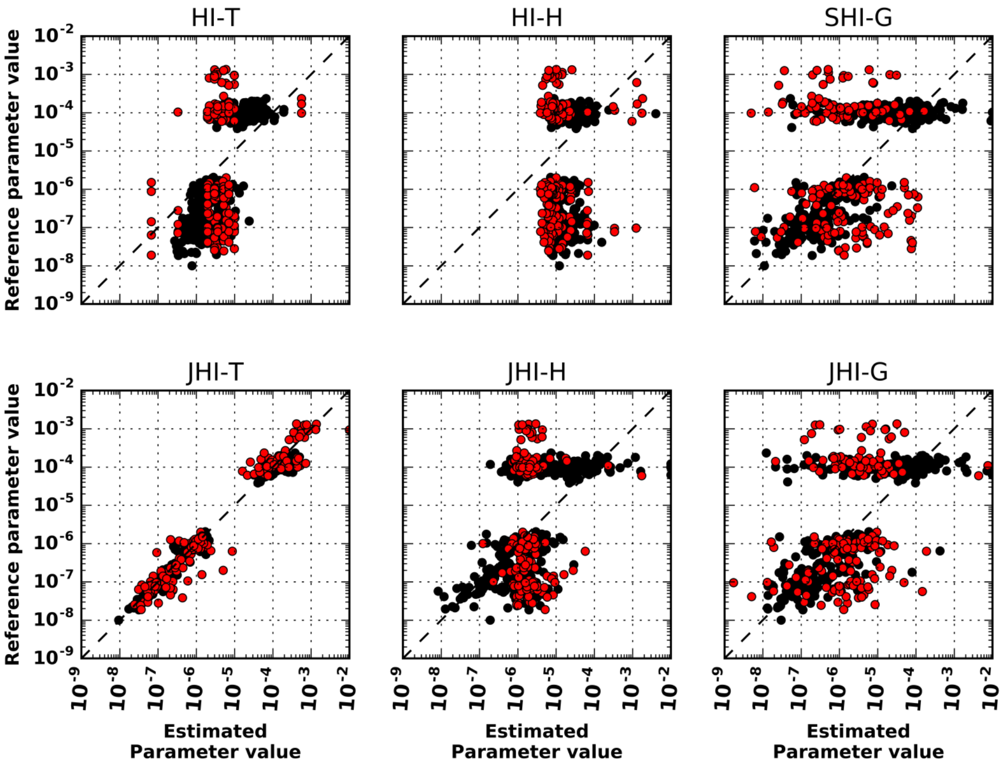

Figure 6. Pilot-point-by-pilot-point scatter plot of reference versus estimated hydraulic conductivity for the six inversion runs. Black dots are estimated parameter values from the capping part of the model, while the red dots are estimated parameter values within the buried valley.

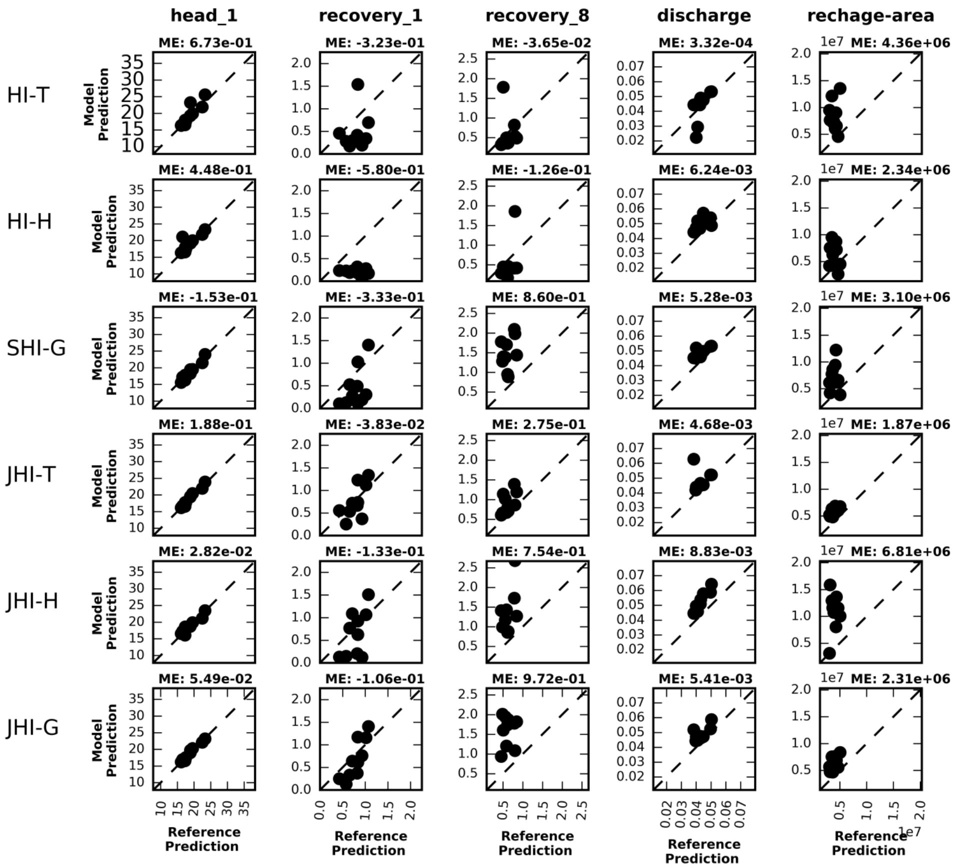

Figure 7: Scatter plots of calibrated model prediction versus reference prediction for five predictions (explained in body text). Each plot shows results from ten system realizations.

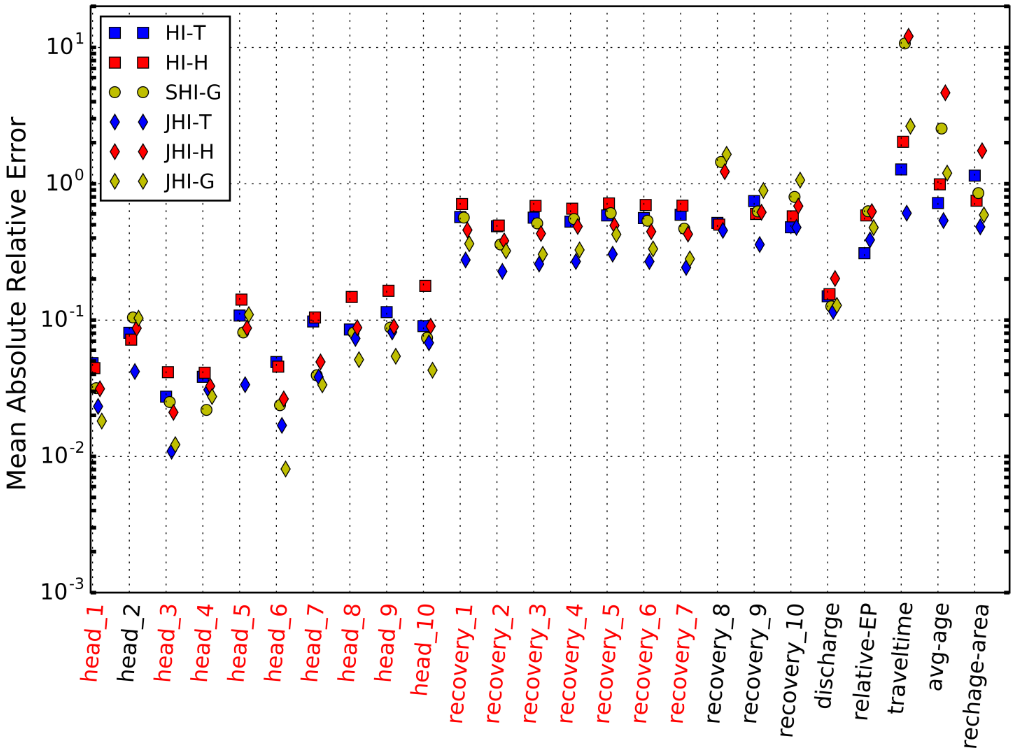

Figure 8: Mean absolute relative prediction error calculated from the ten geological realization results. The symbol type indicates the inversion approach and the symbol color indicates the initial parameter values used when calibrating the groundwater model. Red labels at x-axis highlight prediction errors that are reduced by using TEM data and TEM models for groundwater model calibration.