The author's response and corrections to referee #1

Generel comments:

(1) comments from referees/public	(2) author's response,	(3) author's changes in manuscript
The authors have defined a framework that they call HYTEB and it is presented as a great tool to better understand the role of geophysics in hydrology and it is supposed to be very flexible. Based on the way that this work is presented I was expecting to see a statement by the authors that the HYTEB platform is available for download for any academic users that are interested. I don't find this, which means that I can't judge anything about what HYTEB can do (only a most generic flowchart is shown in Figure 1), I can only judge one specific synthetic study and the quality of the assumptions made and the validity of the findings. The impact of this manuscript would be much more important if this software would be available	Our plan has always been to make HyTEB available to the public. However, when writing and submitting the manuscript we had not yet uploaded HyTEB because of its incomplete documentation. It is still not documented as well as we intend to, but to demonstrate our intention to make HyTEB available, the current version can now be downloaded from https://github.com/Nikolaj-KC/HYTEB. That is, we have only uploaded software that we developed. Software developed by others, for example AarhusInv, PEST, MODFLOW, TPROGS, BLOCKSIS, etc. must be purchased or downloaded from websites of the respective developers. With HyTEB can also be downloaded Python scripts demonstrating how to set up HI, SHI and JHI as it is done in the manuscript	No changes made in manuscript. However, python scripts for setting up the modeling framework can now be found on github.
I find it frustrating to see a lot of general statements, such as, "Much of the lack of value of the geophysical data arises from a mistaken faith in the power of the petrophysical model : : :.". This suggests that this is faith is somehow common to the "hydrogeophysics" community, while in fact I can only assume that this faith is the past faith of one of the authors. Indeed, this over reliance was prelavent until some ten years ago, but this has been solved. Most people that work in hydrogeophysics understand that a smoothness-constrained (Occam-style) geophysical inversion will lead (by construction) to a field that is smoother than reality. This implies that petrophysical relationships between geophysical and hydrogeological properties cannot be used to map a geophysical model into a hydrogeological model. Here are some of the papers from 10 years ago that clearly show this…	The reviewer is correct - some statements are too general and misrepresent that a faith in petrophysical relationships is broadly held in the hydrogeophysics community. We will revise such statements to be less general and refer to the mentioned references where relevant. However, it should also be pointed out that the references cited deal with interpretation of tomographic data that provide a high degree of resolution, thereby allowing for interpretation of spatial variability in petrophysical relationships. In large scale applications, this type of data is generally not available. In the example we imagine that a constant relationship exists, so for the entire catchment true resistivity gives true hydraulic conductivity when using the relationship. This is indeed naive compared to many real investigations, but it makes a case where EM measurements have the best possible chance to resolve change of lithology and change of hydraulic conductivity. If the applied data and inversion approaches do not produce groundwater models that make good predictions in this case, other (for example more dense) data or modeling/inversion approaches should therefore be used for this type of case. It is possible that it is obvious to the reviewer that the used approaches would not work very well and that we should not have used the true petrophysical relationship, but we did not hear this warning or criticism from any of the highly qualified international	We have been going through the manuscript and removed statements like: "Much of the lack of value of the geophysical data arises from a mistaken faith in the power of the petrophysical model." (Has been removed from both the "abstract" and "summary and conclusion".)

	geophysicists we talked to at the early stage of this example investigation. Now in hindsight we can all be much cleverer.	
The authors present one case-study that relies on many strong assumptions and they are then surprised that the results of the inversions are biased. Of course they are. The reference field is an indicator field and the authors invert for a continuous field that is maximally smooth (while the true field has maximum entropy!). They reduce the true parameter field of 1.2 million pixels to some 500 parameters that are solved for using strong smoothness-constrained inversion. So, there is no chance for an of the approaches to lead to unbiased results if the small-scale matters and if sharp interfaces matters. This is specific to HYTEB and it has nothing to do with the value of geophysics. Also, the authors rely on a linearized gradient-based optimization method that is prone to be stuck in local minima. This is demonstrated by the different results obtained when using a homogeneous starting model or the true field. This is to be expected and it is due to limitations in the different aspects of the inversion workflow. My criticism is not that these choices are made (simplifications are needed), but I am against making sweeping generalized statements about the value (lack of value) of geophysics and the bias caused by geophysics. Data that are not properly handled will always lead to bias, but this is not the fault of the data. The findings presented here are valid in this specific synthetic case-study, for the chosen inversion framework, experimental design, and methods used. There are no findings here that permit the authors to make statements about the value of geophysics in general. Indeed, the authors use the geophysical models/parameters as proxies of hydraulic conductivity, which is clearly not the case in a real setting.	The reviewer makes some good points. Our intention in using a complex synthetic model was to balance complexity with the advantage of knowing the 'true' condition so that we could assess model/data performance. In our opinion, the cardinal sin of synthetic model analyses is to use them to show that data/models/analyses ARE likely to be successful beyond the tested conditions. In too many cases, a simplified analysis is used to overextend the likely value of data or models. In this case, we have tried to faithfully represent the standard practice of hydrologists in constructing models. The approach shown is a simple, but still common, approach of representing complex systems with simple (smooth) models. We will be sure to revise the text to incorporate the reviewers very good point – that data are only useful if handled correctly. But, we disagree that it is not meaningful to demonstrate that geophysical data, incorporated into a model following common practice, does not add value. Furthermore, let us clarify the following details. Indicator fields were generated by TPROGS to have maximum entropy; but variation in hydraulic conductivity (and resistivity) within an indicator field was generated to be smooth. The resistivity contrasts between indicator fields (lithologies) are so large that EM measurements could be hoped to spatially resolve/map the lithologies at least at shallow depth even though we parameterize by using (interpolate from) 550 pilot points (but of course there will be some interpolation error and also some smoothing error from the regularization). That at least shallow lithological structure can be resolved this way in this case is to some extent confirmed by Figure 4. However, the data and the inversion approaches are found not sufficient to estimate unique hydraulic conductivity fields that make good groundwater model predictions of for example head recovery. This finding is apparently obvious to the reviewer, but it was not obvious to us or the geophysicists with whom we cooperate within HyGEM before the experiment was actually made. Now we are conducting new experiments with HyTEB where EM data are only used for mapping spatial structure of the indicator fields, whereas hydraulic conductivity fields within the structures are estimated subsequently from other data. Is this a better way of applying EM data in connection with groundwater modeling? It is likely, but *we* don't know for sure yet. And it will be case specific, as we have already emphasized, see for example: Introduction page 9605 lines 19-26; Summary and conclusions, page 9632, lines 10-13.	In the revision we emphasize several places that the results and findings are case specific for the demonstration case. For example, the following sentences are taken from "Discussion and conclusions": "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 the studied system, this shows that resistivities estimated by SHI or JHI must be used with caution as estimators of hydraulic conductivity or as regularization means for subsequent hydrological inversion. In this case, the use of the absolute relationship between hydraulic conductivity and true electrical resistivity led to an over-reliance on the use of inferred resistivities to populate the model's hydraulic conductivity field" "So another important insight from the HYTEB analysis is that recharge should be parameterized and estimated in a different way than it was done in the demonstration example. Alternatively HYTEB could be used to consider adding other types of data to better constrain recharge rates." Another example, in the Abstract it is said: "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 this data are used here: the alternative inversion approaches propagate geophysical estimation errors into the hydrologic model parameters."

The authors state that the case-study is highly realistic and typical of Northern Europe (meaning Denmark). I would then like the authors to use all the data that they have access to in Denmark to show that a log-log resistivity/permeability relationship is valid in this type of settings and on these scales. Also, it is essential to clarify the correlation coefficient of this relationship and to use it in the inversion. The authors criticize the use of petrophysical relationships and then they use a relationship that is certainly not likely to be valid.	The hydrogeological system that we are studying (glacial deposits and a buried valley) is common not only to Denmark but also to parts of northern America. See for example the following reference. Clayton L, Attig JW, Mickelson DM (1999) Tunnel channels formed in Wisconsin during the last glaciation. Geol Soc Am Spec Pap 337:69–82 Wright HE (1973) Tunnel valleys, glacial surges and subglacial hydrology of the Superior lobe, Minnesota. In Black RF, Goldthwaite RP, Willman HB (ed) The Wisconsinan Stage. Memoir 136, Boulder, CO, Geol Soc Am 136:251–276 Estimating the relationship between resistivity and hydraulic conductivity of Danish sediments is beyond the scope of this manuscript. (This is actually studied by some of our partners in the HYGEM project.) However, we would point out again that our approach is consistent with our philosophy of using synthetic models. Namely, we contend that it is most useful to adopt the most favorable (yet reasonable) relationships and conditions in the analysis. Then, any shortcomings of the data/model analysis can be stated more forcefully. For example, in this case, uncertainties or nonuniqueness in the petrophysical relationship would only LESSEN the value of geophysical data. The real limitation in using synthetic models is when favorable assumptions are made and then used to support/advance methods.	The new citations added to line: p9613 L10
The title should be changed by removing electric data (only electromagnetic data are shown	We will remove “electric” from the title.	Done
The authors should use data for data and models for models. It is very misleading to call the EM inversion results for data. I have highlighted some of this confusion below	The reviewer is right. We will go through the manuscript and remove confusions between geophysical “model” and “data”.	Done See for example line two of the abstract.
The abstract is somewhat convoluted and cryptic. I don’t think it is easily understandable outside an expert group. I would suggest that the authors focus on the results of their study and avoid making more “philosophical statements” that are poorly motivated by the presented case-study. Overall, I don’t see the need to include HYTEB (except if access is granted to the reader), why not only present the work as the synthetic case study it is? There is nothing wrong with that and the results are interesting. It seems in the introduction that HYTEB is the answer to one of the most important questions in hydrogeophysics, but the reader is only presented by a simple workflow and a synthetic case-study (+ statements that HYTEB is	Since we now give the reader access to HyTEB we understand from the reviewer that it will be OK to still mention HyTEB in the Abstract? Having received the remaining reviews we will consider revising the Abstract to be more concise and clear.	We have reduced and reformulated the abstract.

flexible).		
There is no information at all about if the data used in the calibration is adequately fitted. For a meaningful comparison, all data should have a weighted RMS of 1. If the misfit varies due to the optimizer is getting stuck in local minima, then this will affect results, but these results will then be due to an inappropriate calibration. I request that WRMSE, chi-square or similar metrics are presented throughout. The study has little value (in my eyes) without this information.	This is a valid point and we will present the calibration results	We have added a new section ("4.1 Model calibration ") and a figure summarizing the model calibration results.
The paper is rather lengthy and it could be shortened. Many sentences are repeated with small variations in various places in the text and I don't see the value of talking about HYTEB. The paper is not about HYTEB, it is about one synthetic case-study. A shorter paper would make the study more attractive to read	We appreciate the reviewer's opinion on this. We will consider this criticism in detail when we have received the remaining reviews	We would like to keep section 2 about "HYTEB" since the software now can be found on github. Furthermore, we have deleted some repeated text and shortened some sections (for example "Abstract" and "Summary and conclusions").

Answers and corrections to the specific comments:

(1) comments from referees/public	**(2) author's response,**	**(3) author's changes in manuscript**
Smaller comments: 9600, line 2: Add "that" before geophysics.	The text will be modified.	Done
9600, line 3: I would write "data and models". In fact, the sequential approach integrates a geophysical model (not data).	The text will be modified to "data and models" and emphasize "models" when talking about SHI.	Done
9600, line 4: Adding "Therefore" can seem a bit too strong here (even more so in light of the actual content of the paper). I would suggest removing it	The text will be modified	Done
9600, line 11: I would replace "and approaches to correlating" with "used to correlate".	The text will be modified	Done
9600, line 15: Perhaps state that the bedrock is clay. Personally I expected resistive igneous rocks.	The text will be modified.	We have added (clay) to emphasis that the bedrock is impermeable and have low resistivity
9600, line 17: These "resistivity estimates" forms a model that here are assumed to be "data" with well-known consequences of this choice.		We did not find reason to change this sentence.
9600, line 25: "be minimized uniquely", what does this mean? I don't follow	Meaning of the "objective function could be minimized uniquely"? We mean that the minimization (which the reviewer calls optimization) may not end in a unique, global minimum but in a local minimum. We thought that meaning was obvious. Does the reviewer have a better suggestion?	Nothing changed

9601, line 2: Rephrase (here and elsewhere) statements about mistaken faith. This is a sweeping statement that makes little sense. Sure, garbage in equals garbage out. This is misleading because you the authors have studied the value of the geophysical model, not the actual value of the data. The data are not responsible for being misused! I don't think there is any value in discrediting geophysics in this way. Also, the sentence does not make sense as it mixes data and models throughout the sentence.	As said before, we will rephrase.	The abstract and "summary and conclusion" section have been reduced and reformulated.
9601, line 20: Replace "ramifications" with "impact"	The text will be modified	Done
9602, line 5: This is obvious. All models are wrong and anyone involved in modeling and inversion should realize this. Remove "the model will be wrong and".	The text will be modified	Done
9602, line 16: Sensitivity to what?	To "small scale heterogeneity". Isn't this obvious when reading the entire sentence?	Nothing changed
9603, lines 6-7: Remove "the" and write "Methods". There are probably 100s of different AEM methods. The referencing in this section is very local and the authors could consider what has been done outside of Denmark and Northern Germany.	The text will be modified as recommended	The following references have been added: Abraham et al. 2012; Lawrie et al. 2012; Oldenborger et al. 2013; Munday et al. 2015. Two of them are from the U.S., and two are from Australia.
9603, line 25: This is not true, geophysical inversion is not required. Not the case in coupled inversion (say work by Mike Kowalsky) and indeed not the case for some interesting approaches, such as, this one: Data-domain correlation approach for joint hydrogeologic inversion of time-lapse hydrogeologic and geophysical data By: Johnson, Timothy C.; Versteeg, Roelof J.; Huang, Hai; et al. GEOPHYSICS Volume: 74 Issue: 6 Pages: F127-F140 Published: NOV-DEC 2009	We will rephrase from "geophysical inversion is required" to "geophysical data are often inverted".	Done
9603, line 27: Also cite:	We will include the suggested references if we find them relevant	All citatations have been added.
9604, line 1: Not correct, the SHI approach incorporates results from an inversion model, not from the data. The distinction is important and explains why the joint inversion approach works better.	We agree and will change this sentence to: "The simplest approach is sequential hydrogeophysical inversion (SHI)."	Done
9604, line 9: This is not the main reason while the sequential approach fails. Read papers suggested above.	We will carefully read the papers and consider rewriting this and following lines.	Again, Based on the suggested references we have added some text explaining the drawback of SHI. P9604 L9
9604, lines 11-12: Yes, but approaches exists to deal with this, say the paper by Moysey et al. cited above or the work by Lochbühler et al. (2014) Conditioning of Multiple-Point Statistics Facies Simulations to Tomographic Images By: Lochbuehler, Tobias; Pirot, Guillaume; Straubhaar, Julien; et al. MATHEMATICAL GEOSCIENCES Volume: 46 Issue: 5 Special Issue: SI Pages: 625-645	We will read the paper and consider reformulation.	We have decided to delete this sentence.

Published: JUL 2014		
9604, lines 26-28: I doubt that the authors will find this type of relationships in a similar case-study in Denmark that is applied at this scale. There are better ways to do so. This is fine for a synthetic example, but it is most likely a futile approach in a real case-study.	The reviewer may be right. The validity of this type of relationship is currently being investigated by other partners in HyGEM. We do not find reason to change the text.	Nothing changed
9605, line 1: Paper of Linde et al. (2006) is on joint inversion of geophysical data, better cite: Lochbühler et al. (2013) Structure-coupled joint inversion of geophysical and hydrological data By: Lochbuehler, Tobias; Doetsch, Joseph; Brauchler, Ralf; et al. GEOPHYSICS Volume: 78 Issue: 3 Pages: ID1-ID14 Published: MAY-JUN 2013	We will change Linde et al. (2006) to the paper of Lochbühler et al. (2013).	Done
9605, line 19: Most of the geophysical methods are rather old, so I am unsure if this is the driving reason.	We are not finding good reason to change the current text.	Nothing changed
9606, line 2: I would remove "for making experiments".	The text will be modified	Done
9606, line 8: Supposed similarity". I don't think this case-study is overly realistic. Both in terms of the geostatistical model or in terms of the petrophysical relationship used. I don't suggest that it is not a good test model (it is rather good), but I wouldn't go so far that it is similar to reality. How do we know? The only thing we know is that we are always wrong.	We will change to saying "supposed similarity". The reviewer has a more negative opinion than we and others have about the realism of the synthetic system. We do not intend to discuss this in the manuscript.	Done
9607, line 10: Replace "response models" with forward simulators or just simulators. Avoid having to many different meanings of "models" in the data. Petrophysical model, forward response model, inverse model, reference model, etc.	We will go through the manuscript and change "response model" to "forward simulators" and "Petrophysical model" to "Petrophysical relationship" etc. In general our intention was to use the phrase "model" for a simplified simulator of the "true complex system" and "reference system" for the "true complex system".	Done
9607, line 11: not true, as there are significant modeling errors involved. Obviously for the 1-D EM modeling, but also for the hydrological model. However, the same simulator is used throughout (the so-called inverse crime).	The reviewer misunderstands what we are saying. Our reference system, the synthetic true system that exists in our virtual world, responds to a hydrologic input without any noise – like a system in the real world responds without noise to its input. The virtual hydrologic response comes out of a numerical simulator, but it is noise free because the simulator with all its hydraulic and hydrologic input and its discretization is the real groundwater system of the virtual world. There are no model errors whatsoever before we parameterize a model to use a limited number of parameters to simulate a response. We think our current wording in the manuscript is sufficient, but we would consider a better wording suggested by the reviewer or the	Nothing changed

	editor?	
9609, line 16: The paper by Günther is on inversion remove it. Indeed, no idea why the authors write about ERT here as it is not used. I would only focus on EM.	Our initial plan was to incorporate ERT into the demonstration; however since we are not presenting any results from ERT we will remove the text about ERT and references. (ERT is allowed though by HYTEB.)	Done
9609, lines 19-20: This is clearly leading to very large errors, so please don't make statements that the modeling is highly advanced and realistic. IF such statements are to be left, then I request a comparison of the 3-D forward simulation with the true field and the one by the 1-D integrated modeling.	We agree that it would have been much better to use 3D-forward simulation but we did not have access to such code. As a substitute geophysicists recommended us to do what we did. The important thing to notice is that for making the geophysical data set we used the pseudo-3D simulation approach while for inverting it we used strictly 1D model simulation: this introduces model error in the inversion; and this is what happens in real investigations where the real system that is measured is 3D but the model/simulator applied for inversion is 1D. In this respect our study is indeed fairly realistic.	We have removed words like "sophisticated" and kept the text explaining what we have actually done.
9610, title 2.4: Please reconsider the name of the section.	The section name will be changed to "model parametrization"	The section title has been changed to "Model construction and parameterization".
9610, line 21: This statement makes little sense and it is at odds with the summary and conclusions where it is written that EM is excellent to derive the bedrock interface.	We will change wording from "they should not be used in both steps 4 and 5" to "it may be argued that they should not be used in both steps 4 and 5". Some people will have this purist viewpoint.	Done
9611, line 3: Not true, HYTEB deals as stated with zones, pilot points and combinations. This is fine, but it is a very small part of all types of parameterizations that are and can be used. Please revise.	Any parameterization that a modeler can come up with is allowed by HYTEB as long as the software necessary to generate the parameterization is made available. To clarify this we will change wording to: "HYTEB allows any type of parameterization, for example zones, pilot points, or combinations hereof".	Done
9614, line 7: Why using maximum entropy? It does not seem like a very geologically realistic choice.	Maximum entropy was only used to generate the indicator fields (with TPROGS), not the variation in hydraulic conductivity within an indicator field; this variation was generated to be smooth (by a different simulator). We cannot change this choice at this stage.	Nothing changed
9616, equation 1: State that this relationship is used here (what is the value of e, what is the resulting correlation coefficient, a value of 0.40-5 seems fair), but that this is unlikely to be valid in real settings throughout the whole model domain. It is not good to write in a hydrology context that this type of relationships this defensible, they are normally not, and there is ample literature that explains why.	The reviewer makes a good point. We here point to our answer from the general comment 4. Furthermore, notice that we in line 23 write "for simplicity" indicating that it may not be valid in real settings.	Nothing changed
9617, lines 18-27: Why not airborne as suggested in intro. Why 77 land-based, and not 5000 airborne? Probably due to computational issues with HYTEB.	This limitation is not an issue within HYTEB itself. It is rather an issue with the used inversion software, PEST. The limited number of soundings is due to the limited number	Nothing changed

	of parameters that PEST can handle and the long run time for running multiple reference systems through the "HYTEB analysis". To our knowledge there is no open-source software for handling the data density of AEM in a proper way for doing JHI.	
9618: This is highly technical and an average hydrological reader (and many geophysicists will not follow). Please simplify or explain. What is the gate center time, what is dB/dt, what is a sign shift, what is an off-center configuration?	We agree. The sentence is unnecessary "technical" and can easily be found in the reference by Auken et al. 2009. In the revised version we will delete the following "technical" sentence and keep the reference by Auken et al. 2009.	Done
9619, lines 9-10: A very important assumption is made when going from 1,2 million pixels to 550 pilot points with kriging in-between, and then using a deterministic, gradient based smoothness-constrained inversion. Many of the findings can be traced to these assumptions that are not general (they are specific to this study) and they have little to do with the data/models considered.	We think we have commented on this already. And we agree, as we have said several places in the manuscript, that the findings are case specific. But this does not make the results less interesting in our opinion.	Nothing changed
9620, equation 5: Here is an important source for some of the bias seen later.	That is true. We will add a small comment to the manuscript saying that this kind of parametrization and regularization creates smooth transition in hydraulic conductivity, which may not be fully sufficient to resolve the "categorical" shifts in reference fields. Furthermore, in a follow-up study we are using a "sequential approach" where we are using MCMC estimated resistivity probabilities in a sequential indicator simulation to construct sharp boundaries; approximately 5000 AEM soundings are used in this study.	Comment added in section 3.4
9621, line 1: The target choice should be 552. Is it reached for all inversions? The data fit really needs to be presented for all cases and comparisons are only meaningful for the same data misfit.	No, the choice should be 2 as said in the manuscript. (Because the two terms on the r.h.s. of (4) is normalized by the number of data in the group.) The calibration results will be added in the revised version.	The calibration results is added to section "results"
9621, lines 13-14: This is not consistent from inverse theory and some more motivation is needed.	It is true that it is not standard to divide each term with the number of data in the group, but, as we already say in line 14, we do it to give a balanced weight to each of the three groups of data; if we didn't, the third term would totally dominate the objective function because of this group's large number of data. As already stated, this is the subjective choice made here, and it is OK. We will keep the text as it is. This is also encouraged by Hill (1998); data weights should be used to scale observations providing meaningful results when being summed to estimate the combined objective function, or to reduce contribution from less reliant sources. Hill, M. C. (1998). Methods and Guidelines for Effective Model Calibration. U.S. Geological	Nothing changed

	Survey, Denver, Colorado, Water-Resources Investigations Report 98-4005.	
9626: Better results would have been obtained if inverting for the parameters in the petrophysical relationship (see paper by Linde et al. (2006) cited above).	The reviewer is probably right, but this is beyond the scope of this study. In a follow-up study we are currently using a spatially varying petrophysical relationship.	Nothing changed
9626, lines 26-29: This is expected and it is not a new finding.	That is probably true, but we are not aware of references finding this. Can the reviewer help us so we can include such references? Still we will mention the finding because it may not be common knowledge within the hydrological community.	Nothing changed
The content of the rest of the paper is fine, but the manuscript needs to be revised throughout to be in line with the comments made above. This is especially true for the Summary and conclusions section. It would also be good to shorten this section.	We will shorten the "Summary and conclusion" section	The summary and conclusion section has been shortened. Besides, we have removed all the summary text and changed the name of this section to "Discussion and conclusion". (suggested by referee #2)

The author's response and corrections to referee #2

Overall Comments:

(1) comments from referees/public	(2) author's response,	(3) author's changes in manuscript
Some of the largest challenges with coupled or joint inversion are linking geophysical measurements to hydrological parameters of interest. In this manuscript, the authors almost entirely neglect this with the justification of demonstrating an example (resistivity is assumed to have a direct relationship to K, porosity is assumed to be known). How is it possible to know that the absence of reliable hydrologic output parameter prediction isn't due to the poor petrophysical relationship?	In the example we imagine that a constant relationship exists, so for the entire catchment true resistivity gives true hydraulic conductivity when using the relationship. This is indeed naive compared to many real investigations, but it makes a case where EM measurements have the best possible chance to resolve change of lithology and change of hydraulic conductivity. This implies that a "poor" prediction is not due to the petrophysical relationship, but due to the limitations of the geophysical model and inversion approach, initial parameters of the inversion etc.	See following "change".
After all, if a fully synthetic system is designed and then converted between hydrologic and geophysical properties using an empirical or semi-empirical petrophysical model the petrophysical model may be incorrect. How can the authors justify enforcing a link between hydraulic conductivity and resistivity, but not porosity (as Kozeny-Carman would require)?	As we state at page 9619 line 16, porosity cannot be estimated from the hydrological and geophysical data available here. We therefore made the subjective choice not to include porosity as a part of the petrophysical relationship nor estimating this parameter during the different model calibration. As we state at page 9617 line 13-16, a more complicated, or less certain, relationship between electrical resistivity and hydraulic conductivity (and porosity) could have been chosen, but we made the simpler choice (with no influence from porosity) because it makes a case where EM measurements together with the hydrological data have the best possible chance to resolve	To make completely clear we now write in the beginning of section 3: "The case is designed so there is 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 of electrical resistivity."

	change of lithology and change of hydraulic conductivity. We will emphasize this in the revised manuscript.	
A discussion section is absent.	See our answer to the second last question of referee #2	The summary and conclusion section has been rewritten. Summary part has been deleted, and the section is now named “discussion and conclusion”
The authors helpfully identified previous simultaneous inversion examples “Linde et al. (2006), Herckenrath et al. (2013a) and Vilhelmsen et al. (2014),” and it would be helpful to relate these current results to the past examples. Alternatively, if the results of this investigation cannot be related to past experiments due to the highly synthetic nature of the study, then I question it’s relevance to a hydrology journal and suggest consideration of an engineering journal may be more appropriate to document the development of the HYTEB computational environment.	We disagree. Our demonstration focus on hydrological predictions and whether they can be improved by using geophysical data in two different ways (SHI and JHI) to support parameter estimation of a groundwater model. In this respect our demonstration differs from Herckenrath et al. (2013a) and Vilhelmsen et al. (2014) who only focus on parameter estimation and not on hydrological predictions. So we find that this study is highly relevant to a hydrological journal, and even more than the mentioned previous studies.	Nothing changed
I suggest including a table of all symbols and definitions. There are many symbols used in this manuscript, and some of them are ambiguous (for example, small sigma may be used to refer to electrical conductivity or standard deviation, although I think it is always standard deviation in this manuscript.)	We will consider including a table of all symbols and definitions if the editor finds it necessary.	We have not added a table of symbols, but we have avoided ambiguous use of symbols.

Line Comments:

(1) comments from referees/public	(2) author's response,	(3) author's changes in manuscript
Replace all “worth” with “value.”	We are not finding good reason to change the current text, because the phrase “worth” is in line with related literature on the topic. See for example Doherty et al. (2010). Approaches to Highly Parameterized Inversion: A Guide to Using PEST for Model-Parameter and Predictive-Uncertainty Analysis. Scientific Investigations Report 2010–5211, USGS	Nothing changed
P9604, L:9: “play back” idiom. Consider replacing.	The text will be modified and enhanced according to the suggestions made by referee #1.	The text has been re-formulated.
P9604, L:9-10: The wording in this sentence is awkward. Suggest rephrasing.	The text will be modified and enhanced according to the suggestions made by referee #1.	The text has been re-formulated.
P9613, L. 26: “Fig. 2” The text on this figure is hard to read and in some cases overlapping. I suggest redrawing for clarity.	We will enhance the layout off figure 2	Done
P9616, L23: I understand that assuming a relationship between res and K is handy for simplicity, but it is also highly unrealistic. What will be the impact when a realistic relationship must be used when incorporating field data? How should that relationship be developed in order to work properly within this modeling framework?	The referee is asking all the right and relevant questions. One of the advantages of HYTEB is providing the framework for doing such experiments. However, estimating the relationship between resistivity and hydraulic conductivity for field data is beyond the scope of this manuscript. (This is actually studied by some of our	We think this is now made clear in the revised “Discussion and conclusions”.

	partners in the HYGEM project.) However, here we intend to analyze under the most favorable system conditions (where there is a perfect relationship between hydraulic conductivity and resistivity), how well can a groundwater model make predictions when it is developed and calibrated from geophysical and hydrological data as it is done here. Any shortcomings of the data/model analysis would only be worse if there were uncertainties or nonuniqueness in the petrophysical relationship. As said to the other reviewer, the real limitation in using synthetic models is when favorable assumptions are made and then used to support/advance methods. Here we are not really successful even under favorable conditions, indicating that other (for example more dense) data or modeling/inversion approaches should be used for this type of case. We will emphasize this in the conclusion of the revised manuscript.	
P9619, L16-17: This is a bit confusing –porosity is a key and critical parameter. How is it justified to assume it is known? Also, it seems like the Archies type relationship for porosity might be more reliable than estimating K from resistivity, so why is K the one calculated and porosity?	We have answered this question previously.	Nothing changed
P9619, L21-24: Since the numbers of layers in the geophysical model is linked to the number of layers in the synthetic geological model, does this mean it is required to know the number of geologic units in a real scenario a priori?	No, but again it is done to simplify as well as to make conditions favorable.	Nothing changed
P9624, L10: How computationally intensive was it really? What kind of limitation might this pose for general users to HYTEB?	Making the calibration and predictions for 10 system realizations parallelized onto 24 CPU's took approx. 7 days for JHI-T, JHI-H, and JHI-G; 5 days for HI-T and HI-H; and approx. 2 days for SHI. This makes a total of approx. 2 weeks. We do not see this as a serious limitation.	Nothing changed
P9621, L18: It appears here that hydraulic conductivity is now represented as lowercase-k, rather than uppercase-K as in table 1. Is this significant? An error? What is the difference between these k's?	The referee is correct. This is an error. We will change to use uppercase-K everywhere. As said in lines 18-19, K_joint is K inferred from geophysics, and K_mf is K inferred from hydrology (and used in the groundwater model).	Done
P9627, L21: "Figure 6" the figures have a lot of overlapping points and numbers – hard to decipher overall. Suggest re-drawing for clarity.	We will remove the numbering on this figure. The idea behind this is to identify bias and scatter of the prediction around the unity line as suggested by Doherty and Christensen (2011).	Done
P9627, L25: "Mean Error" Can the ME value reported on each panel of Figure 6 be interpreted as "Smaller is better"? In other words, would it be possible to interpret these results as "for each parameter, the model prediction with the smallest ME is the most well resolved"? If so, perhaps placing an identifying mark on each panel of this figure matrix would help the reader see more easily which is performing best and second best for each parameter? I	On figure 7 we have summarized our findings from figure 6. Figure 7 highlights (with red) the predictions for which SHI or JHI reduce the prediction error compared to HI (hydraulic observations only).	Nothing changed

think it would enhance clarity.		
P9628, L11-12: “the scatter around the identity line is larger for HI calibrated models than for JHI calibrated models” it is really hard to tell!	For the head_1 prediction (and the other head predictions that are not shown in the figure) we find it visually fairly distinct that the points plot closer to the identity line for JHI and SHI than for HI. However, we will change wording in lines 13-14 to: “However, the scatter around the identity line appears to be larger for HI calibrated models than for JHI calibrated models.”	Done
P9632, L1: The purpose of the long summary text is unclear and conclusions are nearly absent. I suggest removing the summary text and instead focus on developing a clear, concise conclusions section.	The referee has a good point. We will remove as much text as possible from the summary part and make the conclusions part more concise. As mentioned by referee #1, this paper is already rather lengthy. We are therefore very reluctant to add a focused discussion section to the manuscript. The present “Summary and conclusion” section also has some discussion element in it. We will consider if we should keep it that way.	We have deleted the summary part, shortened and made this section more concise. Furthermore, we have renamed this section to “Discussion and conclusions”, since we find that this section already has some discussion element in it.
Table 1: The caption for the figure needs to be improved and the definition of each parameter needs to be included. I see the table referenced on p. 9614 line 6 for the first time, and no clear definitions of the symbols in the table are included there either in the immediate vicinity. K is clearly hydraulic conductivity, I presume “R” is resistivity given equation 1 on 9691, line 25, however in eq. 1, the Greek symbol rho is used. Typically R is “Resistance,” not a physical property. I presume the last symbol is phi for porosity, but how is this calculated, or how does this value link with the K-to-resistivity transform? Clearly all three must be linked somehow (P9619, L16 would suggest that this is not the case – this should be expanded upon, justified, and rectified).	The referee is right; we will explain the symbols in the caption. To clarify, in table 1 K is for hydraulic conductivity, R is for recharge to the groundwater model, and phi for porosity.	Done

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

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

[1]{Department of Geoscience, Aarhus University, Aarhus,Denmark}

[2]{Department of Hydrology and Water Resources, University of Arizona, Tucson, USA.}

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

Field Code Changed

Abstract

Despite that geophysics is being used increasingly, it is ~~still~~ often unclear how and when the integration of geophysical data and models~~data~~ can best improve~~s~~ the construction and predictive capability of groundwater models. ~~Therefore, t~~This paper presents a newly developed **HY**drogeophysical **TE**st-**B**ench (HYTEB) which is a collection of geological, groundwater and geophysical modeling and inversion software wrapped to make a platform for generation and consideration of multi-modal data for objective hydrologic analysis. It ~~is intentionally flexible to~~ allows for ~~simple or sophisticated~~ 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 ~~models~~ and approaches to correlating geophysical and hydrologic parameters. With HYTEB we study alternative uses of electromagnetic (EM) data for groundwater modeling in a hydrogeological environment consisting of various types of glacial deposits with typical hydraulic conductivities and electrical resistivities covering impermeable bedrock with low resistivity(clay). ~~Using a~~The synthetic three dimensional reference system is designed so there is a perfect relationship between hydraulic conductivity and electrical resistivity.~~,~~ For this system ~~It~~ 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 purely hydrologic inversion (HI, only using hydrologic~~

~~data) we used Tikhonov regularization combined with singular value decomposition. For joint hydrogeophysical inversion (JHI) and sequential hydrogeophysical inversion (SHI) the resistivity estimates from TEM are used together with a petrophysical relationship to formulate the regularization term. In all cases, the regularization stabilizes the inversion, but neither the HI nor the JHI objective function could be minimized uniquely. SHI or JHI with regularization based on the use of TEM data produced estimated hydraulic conductivity fields that bear more resemblance to the reference fields than when using HI with Tikhonov regularization. However,~~

F~~f~~or the studied system and inversion approaches ~~the~~ it is found that that resistivities estimated by sequential hydrogeophysical inversion (SHI)~~SHI~~ or joint hydrogeophysical inversion (JHI) ~~JHI must~~ should be used with caution as estimators of hydraulic conductivity or as regularization means for subsequent hydrological inversion. ~~Much of t~~The ~~lack of~~ limited ~~value~~ groundwater model improvement ~~of~~ obtained by using the the geophysical data probably mainly arises from the way this data are used here: the alternative inversion approaches propagate geophysical estimation errors ~~a mistaken faith in the power of the petrophysical relationship model in combination with Sgeophysical data of low sensitivity, thereby propagating 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~~geophysical data in the model calibration~~g. It is found that all calibrated models are good predictors of hydraulic head. When the stress situation is changed from that of the hydrologic calibration data, then all models make biased predictions of head change. All calibrated models turn out to be a very poor predictor of the pumping well's recharge area and groundwater age. The reason for this is that distributed recharge is parameterized as depending on estimated hydraulic conductivity of the upper model layer which tends to be underestimated. Another important insight from the HYTEB analysis is thus that either recharge should be parameterized and estimated in a different way, or other types of data should be added to better constrain the recharge estimates.

1 Introduction

1.1 Using hydrologic models for decision support

Groundwater models are commonly constructed to support decision-makers in managing groundwater resources. The model can, for example, be used to predict the ~~ramifications~~ 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.

A groundwater model is based on a conceptual model that encapsulates prior knowledge of important physical and chemical conditions and processes of the complex real world system. The conceptual model is translated into a numerical groundwater model whereby its reasonableness can be tested by comparing forward simulations~~ors~~ ~~model simulations~~ with field observations. If the conceptual model appears reasonable, the groundwater model is calibrated by adjusting model parameters until simulated values fit corresponding field observations sufficiently well. The calibrated model is subsequently used to make predictions (Reilly 2001; Reilly and Harbaugh 2004). However~~, the model will be wrong and~~ the prediction will be ~~s~~ uncertain for a number of reasons. (i) Model calibration is done by fitting uncertain data. The calibrated parameters will therefore also be uncertain and this uncertainty is propagated to the model predictions (Hill 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 ~~the~~ 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. 2014), 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.

Field Code Changed

Field Code Changed

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, ~~geophysical~~ the data are often inverted~~geophysical inversion is required~~. The way in which hydrologic and geophysical data are inverted and integrated can impact the extraction of information from geophysical data (Dam and Christensen 2003; Day-Lewis 2005; Moysey et al. 2005; Linde et al. 2006b; Singha and Gorelick 2006; Singha and Moysey 2006; Hinnell et al. 2010).

The simplest approach is sequential hydrogeophysical inversion (SHI). ~~The simplest approach to incorporating geophysical data is through 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, ~~T~~the inverted geophysical properties are ~~then~~ 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 ~~T~~the geophysical responses are sensitive to some of the same structures and property distributions that the hydrologic data are sensitive to. ~~This means that potentially the hydrologic data can contain information about the geophysical system. Despite that geophysical provides minimal non-invasive measurements of the subsurface characteristics.~~ 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. 2006b; 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 SHI the geophysical models cannot be easily updated to conform to the hydrologic observations.

~~Furthermore, with SHI it is difficult to quantify the uncertainty of groundwater model predictions because different assumptions, smoothing, and inversion approaches may have been used to invert the hydrological and geophysical data (Doherty et al. 2010; Menke 2012).~~

Two alternatives to SHI that extract more information from the data sets are coupled hydrogeophysical inversion (CHI) and joint hydrogeophysical inversion (JHI) (Hinnell et al. 2010). For both alternatives, the hydrologic and geophysical data sets are inverted simultaneously. In CHI, the simulated response of one model (e.g. the hydrologic model) is used as input to constrain the other model (e.g. the geophysical model). (For example, during the inversion a water table simulated by the hydrologic model is used to constrain the depth of a layer boundary of the

Field Code Changed

estimated geophysical model.) CHI has been applied successfully for reducing parameter uncertainty by using ground penetrating radar and electrical resistivity tomography data in hydraulic models (Kowalsky et al. 2005; Hinnell et al. 2010). In JHI, the hydrologic and geophysical models are coupled directly through some of their parameters using assumed relationships among the geophysical and/or hydrologic parameters (Hyndman et al. 1994). For EM data, JHI is typically done using a relationship between hydraulic conductivity and electrical resistivity inspired by Archie´s law (Archie 1942; Revil and Cathles 1999; Purvance and Andricevic 2000; Slater 2007). Application of JHI for simultaneous inversion of hydrologic and geophysical data has been demonstrated by Lochbühler et al. (2013)~~Linde et al. (2006)~~, Herckenrath et al. (2013a) and Vilhelmsen et al. (2014).

Formatted: English (U.S.)

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

1.3 Hydrogeophysical test-bench

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

reduce groundwater model prediction error and uncertainty and how they can best be used for this purpose. This examination is not straight forward because it requires both hydrogeologic and geophysical understanding and competences.

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

1.4 Objectives

The paper has the following objectives. First, it will present the important elements and steps in use of HyTEB. Since HyTEB and its use is interdisciplinary, the presentation and the following case study introduce geophysicists to the methods, challenges, and purposes of groundwater modeling, and groundwater modelers to some of the challenges of using mainly electric and electromagnetic data for groundwater model calibration purposes. Second, HYTEB is used to examine the worth of adding a ground based time-domain electromagnetic data set to a hydrological data set when making a groundwater model for a glacial landscape of a kind that is typical to parts of Northern Europe and North America. It is investigated if the worth of adding the geophysical data depends on the type of groundwater model prediction as well as on whether the geophysical and hydrological data are inverted sequentially or jointly. Section 2 of this paper describes the elements of HYTEB and how they are used, 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 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 ~~response models~~ to generate measurements that would be collected from the reference systems in the absence of noise. We then examine the influence of measurement error and other sources of error on model predictions of interest. By repeating this for different synthetic system realizations (i.e. for different reference systems) and for different data sets it becomes possible to statistically quantify the worth of the various data for improving the predictions of interest. The work flow of HYTEB is shown in 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, 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, ~~hydrological~~ data set, and predictions

Using the same spatial discretization as in step 1, the second step is to define the boundary conditions and the hydraulic and solute transport property values for the generated geological system. The hydraulic and solute transport properties can include, for example, hydraulic conductivity, specific storage, and effective porosity. For categorical deposits (as in 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 simulated for this system are called "reference prediction".

2.3 Step 3 – Generation of reference geophysical system and ~~geophysical~~ data set

The third step is to define the property values of the geophysical system corresponding to the geological realization generated in step 1. Like the hydraulic properties, the geophysical properties can be considered and simulated as categorical random fields. A geophysical property of relevance can, for example, be the electrical resistivity of the spatially variable geological deposits. For some geological systems, it is found or assumed that there is correlation between electrical resistivity and hydraulic conductivity. In this case, the hydraulic and geophysical property fields must be generated to be dependent. Various empirical petrophysical relationships between hydraulic conductivity and electrical resistivity have been proposed (Slater 2007). It is common to use a linear log-log relationship which is given some theoretical support by Purvance and Andricevic (2000). Having defined the property values of the geophysical reference system, the geophysical instrument

responses are simulated to produce a noise-free geophysical data set that can be corrupted by adding random error to represent all sources of measurement error. Ideally a 3D code should be used. ~~3D computation of electrical responses can be efficiently modelled(Günther et al. 2006; Rücker et al. 2006).~~ Codes for 3D computation of TEM responses have ~~also~~ been developed (e.g. Árnason 1999), but the computation is impractical and burdensome. As a practical alternative we suggest to simulate TEM responses by a 1D code, where the 1D geophysical model is created from the reference system~~model~~ 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 ~~a~~ the 1D approach should work~~s~~ well (Auken et al. 2008; Viezzoli et al. 2010a). In HYTEB we use AarhusInv (Auken et al. 2014) to simulate ~~electrical and~~ electromagnetic instrument responses.

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

2.4 Step 4 – Model construction and parameterization~~Make and parameterize models~~

In this step, the synthetic data are used to constrain parameter estimation for a groundwater model of the ~~true~~ 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 "~~true~~ 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~~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, zones, pilot points, or combinations hereof. In the following demonstration we chose pilot points.~~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 ~~real~~ reference groundwater system. States occurring in (i.e. simulated for) the ~~true~~ reference groundwater system are here termed "~~true~~ reference predictions". The objective of model calibration is to make the model predictions as similar as possible to the ~~true~~ reference predictions.

2.5 Step 5 – Calibrate the model(s)

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

~~geophysical models simultaneously, they are subject to the same regularization effects and all of the data are fitted simultaneously by both model types.~~

2.6 Step 6 – Simulate model predictions, then repeat steps 1-6

After successful calibration, the groundwater model is used to make model predictions equivalent to the reference predictions ~~as in~~ of step 2. 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~~ made through~~, steps 1 to 6 needs to be repeated a number of times. Each repetition involves generation of a new realization of the geological system and the corresponding 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 ~~true~~ 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 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). 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. The bedding is represented as a maximally disordered system using ‘‘maximum entropy’’ transition frequencies (Carle 1999).

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

3.2 Reference ~~G~~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.

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; 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~~model~~ output. Independent Gaussian error with zero mean and 0.1 m standard deviation was added to the true head values to produce the head observations. Gaussian error with zero mean and a standard deviation corresponding to 10 % of the true river discharge was added to the discharge to produce the stream flow observation used for model calibration.

3.2.2 Reference ~~P~~predictions

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

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

Prediction types 4 to 7 relate to a new steady-state long after pumping from the well has been stopped. Type 4 is head recovery at the ten locations given in Figure 2 and Table 4. Type 5 is the travel time of a particle flowing with the groundwater from the location ~~that~~ 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 ~~true~~ 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 ground-water 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 ~~G~~geophysical system and data – step 3

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

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

where K is the hydraulic conductivity (m/s), ρ is the electrical resistivity (ohm-m~~), e is random Gaussian noise,~~ 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) was used to compute the resistivity within each cell of the geological system from the corresponding cell hydraulic conductivity.

Using a ~~direct~~ perfect relationship between hydraulic conductivity and resistivity must be characterized as the ideal case because electrical resistivity data can provide maximal information about hydraulic conductivity. When possible, estimation of hydraulic conductivity from electrical resistivity is usually based on a site specific linear log-log relationship (see e.g. Mazáč et al. 1985; Revil and Cathles 1999; Purvance and Andricevic 2000; Slater 2007), which has been found to be a positive relationship in some cases (Urish 1981; Frohlich and Kelly 1985), and a negative

relationship in other cases (Worthington 1975; Heigold et al. 1979; Biella et al. 1983). (A more complicated, or less certain, relationship between electrical resistivity and hydraulic conductivity could also have been chosen for the demonstration; HYTEB is designed to have no such limitation.)

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) 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 according to the noise model suggested by (Auken et al. 2008).~~:~~

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

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

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

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

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

3.4 Model construction and parameterization (Step 4)

Formatted: English (U.S.)

The groundwater model uses the true boundary conditions except that recharge is to be estimated together with hydraulic conductivity. Because the ~~synthetic~~ reference groundwater and geophysical systems ~~are~~ 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). Kriging is used for spatial interpolation (here using the correct correlation lengths) from the pilot points to the model grid. T~~t~~his kind of parametrization creates smooth transition in hydraulic conductivity~~,~~ which may seem problematic to use in the current case where there are ~~not be fully sufficient to resolve the~~ "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 porosity cannot be estimated from the hydrological and geophysical data available here, we always use the ~~true~~ 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)

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

3.5.1 Three calibration methods

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

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

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) 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 during each optimization iteration making ϕ_m equal to a user specified target value (Doherty 2010). In this case, for ϕ_m defined by (4), the target value is set to 2 (indicating that the fitted data residuals correspond to the data noise levels).

Method 2 is joint estimation of groundwater model parameters and geophysical model parameters on the basis of both hydrologic and geophysical data (JHI). The minimized objective function is of the same form as (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_h, n_r and~~ n_{TEM} are ~~the number of head, discharge and~~ of TEM observations, respectively. The first two terms on the right hand side of equation (6) are identical to the terms in (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) 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

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

In (5), $\cancel{k}K_{mf,i}$ is the estimate of the hydraulic conductivity at the ith pilot point of the groundwater model; $\cancel{k}K_{joint,i}$ is also an estimate of hydraulic conductivity, but this estimate is calculated from the estimated electrical resistivity at the same depth and location by using equation (1). In this case, the target value of $\phi_{m,joint}$ is set equal to 3.

Method 3 is sequential parameter estimation (SHI) as proposed by Dam and Christensen (2003). First, the geophysical model parameters (electrical resistivities) are estimated on the basis of the geophysical data. ~~Subsequently~~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, ~~T~~the estimated electrical resistivities are used to constrain the subsequent hydrologic inversion, which is carried out as minimization of equation (3~~5~~) where the measurement objective function ϕ_m is defined by equation (2~~4~~) while the preferred difference regularization term is defined by

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

As in (5), $\cancel{k}K_{mf,i}$ is the hydraulic conductivity at the ith pilot point of the groundwater model; $\cancel{k}K_{seq,i}$ 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). In this case the target value of ϕ_m is set equal to 2.

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

3.5.2 Initial parameter values

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

For method 1 (HI), we ran two inversions. In the first run, termed HI-T, we used the ~~true~~ 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 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.

~~In the demonstration example, we examine SHI and JHI approaches, without considering CHI. We made this choice not because we don't see value in examining CHI, nor because of any limitation in~~

~~HYTEB for examining CHI. Rather, for clarity of presentation, we considered SHI and JHI to be more easily comparable. CHI analyses are generally more valuable when considering transient data; essentially, CHI allows the process model to replace smoothing in time when interpreting the geophysical data. Having made the choice not to examine CHI, we could use a realistic case study that did not include transient data.~~

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 ~~that~~ can be calculated by running just one of the models. However, for method 3, the geophysical data were inverted using AarhusInv (Auken et al., 2014).

3.6 Picking 10 realizations

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

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

models that broadly represent the range of model behaviors, including both the range of each prediction and the correlations among predictions.

4 Results

4.1 Model Calibration

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) 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): 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 shows the ~~true~~ reference hydraulic conductivity fields of the uppermost six layers and a representative cross section for one of the 10 chosen system realizations ~~(see section 3.6)~~. It also shows the corresponding estimated hydraulic conductivity fields obtained by six different inversion runs. The figure can thus be used to visually compare the estimated hydraulic conductivity fields and to judge whether they resolve the structures of the ~~true model~~reference system. Figure 6 shows

Field Code Changed

corresponding pilot-point-by-pilot-point scatter plots of ~~true~~ reference versus estimated hydraulic conductivity. Except when noted specifically, the results in Figure 5 and Figure 6 for this realization are typical for all 10 chosen system realizations.

The second and third rows of 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 ~~true~~ reference field but the estimated values vary much less than the ~~true~~ 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 ~~true~~ reference fields. The third row of Figure 5 illustrates that for inversion HI-H, which used homogeneous initial hydraulic conductivity fields, there is almost no structurel similarity between the estimated and ~~true~~ 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 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 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 shows that for low-permeability deposits, the range of estimated log-hydraulic conductivities is twice as large as the ~~true~~ 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 ~~true~~

reference values, and the estimated values tend to be orders of magnitude too small (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) because the resistivity estimated from TEM tends to be too high due to low sensitivity of the TEM data. For the studied system, this shows that resistivities estimated by independent TEM data inversion must be used with caution as estimators of hydraulic conductivity or as regularization means for subsequent hydrological inversion. In this case, the absolute relationship between hydraulic conductivity and ~~true~~ 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 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 shows that estimated versus ~~true~~ reference hydraulic conductivity values plot nicely along the identity line for JHI-T. The resolution of structures (Figure 5) and the quality of the K estimates (Figure 6) deteriorate for JHI-H and JHI-G, both of which use less informative initial parameter values. Figure 5 visually indicates that JHI-G resolves structures better than JHI-H. For sand and gravel deposits 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 ~~true~~ 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 shows five examples of scatter plots of ~~true~~ 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 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), groundwater discharge to the river after pumping has stopped, and recharge area of the pumping well.

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 ~~true~~ 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~~ except in the buried valley near the pumping well~~. Unbiased head prediction is exemplified by the plots in the first column of 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~~However, the scatter around the identity line is larger for HI calibrated models than for JHI calibrated models~~. This indicates that the use of geophysical data in JHI reduces the uncertainty of this head prediction as compared to the HI calibrated models. The scatter plots for the other head predictions are similar to those shown for location 1 with the following exceptions. For head prediction 2 (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 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. The two plots in the top of the second column indicate that head recovery at location 1 tends to be overpredicted by the models calibrated by purely hydrologic inversion (HI-T and HI-H). The third plot in this column ~~(SHI-G)~~ indicates that SHI-G slightly reduces ~~some of~~ the bias seen in the HI-based model~~ prediction may be reduced slightly by using geophysical data in a sequential approach~~. Finally, the last three plots in the second column of 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).

The second plot~~s~~ in the third column of 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 the identity line while the tenth point falls far above the identity line. Generally, the plots indicate a consistent overprediction of head using HI-based inversion. The remaining plots in the third column show that recovery prediction at location 8 turns out to be too large for the models calibrated with geophysical data (JHI) or by using geophysics based regularization (SHI)~~, no matter whether this is done by SHI or by JHI~~. ~~These plots indicate that use of~~ the way ~~the geophysical data~~

~~are used in JHI or SHI introduces further bias in the prediction of head recovery within the buried valley: if a line is visually fitted through the points, the apparent non-zero intercept indicates bias (see section 2.7). The scatter plots for head recovery at locations 9 and 10, also inside the buried valley, are similar to those for location 8.~~

Figure 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. On the contrary, for recovery prediction 8, located within the buried valley, both Figure 7 and Figure 8 show that ~~using~~ including the geophysics~~al data~~ 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.) 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 indicate that discharge to the river without pumping is overpredicted except for the HI-T and JHI-T based models. Further, this is a type of model prediction that is not improved by including geophysical data in the inversion (compare for example the HI-H plot with the JHI-G plot). If anything, the results for the ten realizations indicate that use of geophysical data may bias discharge prediction unless joint inversion is used with initial parameter values being equal (or close) to the reference (true) values (JHI-T). That use of geophysical data is not important to improve this prediction is confirmed by the relative error magnitudes for discharge shown in Figure 8.

4.3.4 Recharge area and other particle tracking predictions

The plots in the fifth column of 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~~s~~ 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 for the prediction of average age of groundwater pumped from the well and for the prediction of particle travel time. The explanation for these poor predictive performances is similar to that just given for the prediction of the well's recharge area. Figure 7 and 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 ~~Summary~~ 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 reduces to the question: what type and abundance of geophysical (and other) data should be gathered, and how can they be used optimally? This can be clarified by doing controlled experiments. For large spatial scales, these questions can best (only?) be done by synthesizing both the actual hydrogeological setting and the alternative data gathering, doing modeling experiments using these data, and comparing modeling results with the known "synthetic reality" for the alternative choices.~~ This paper presents a newly developed framework 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 a **HY**drogeophysical **TE**st-**B**ench (HYTEB). HYTEB allows for ~~sophisticated~~ treatment of hydrologic and geophysical data and inversion approaches. It can be used to examine the combined use of hydrologic and geophysical data, including model

parameterization, inversion, and the use of multiple geophysical or other data types. It can also be used to discover potential errors that can be introduced through petrophysical models used to correlate ~~and approaches to correlating~~ 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. ~~It can be used at any point in an investigation, with all existing data integrated into the analysis together with synthetic data. It is intentionally flexible to allow for simple or treatments of geophysical response, hydrologic processes, and inversion approaches. This should allow HYTEB to contribute to studies ranging from practical groundwater studies to large scale research efforts.~~

~~The advantage~~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. ~~The bedrock has low electrical resistivity in significant contrast to the higher resistivity of the glacial deposits. The resistivity of the glacial deposits varies with the grain size and clay content. TEM data is often collected (now by airborne systems) to map this type of environment because it is ideal for mapping the depth to the top of the Tertiary bedrock, including the depth and location of buried valleys (shown by Auken et al., 2008). The bedrock structure represented in groundwater models for this type of environment is therefore often constructed primarily on the basis of TEM data while the groundwater models are calibrated on the basis of hydrologic data only. Here we use~~ HYTEB 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).

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

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. ~~It was found that the estimation results depend on the type of regularization used as well as on the choice of initial parameter values.; neither~~ ~~Neither~~ the ~~hydrologic inversion (HI, only using hydrologic data)HI nor joint or sequential hydrogeophysical inversion (JHI and SHI)the JHI~~ the ~~objective function could be minimized uniquely. SHI or JHI with regularization based on the use of TEM data produced estimated hydraulic conductivity fields that bear more resemblance to the true fields than when using HI with Tikhonov regularization. However, even for the studied case, for which there is a perfect and known relationship between hydraulic conductivity and electrical resistivity, the hydraulic conductivity field estimated by the inversion methods used here depends on the choice of initial parameter values. Not surprisingly it is found best to use true parameter values as initial values, but the true values will, of course, never be known. However, HYTEB could be used to dig deeper into this finding. That is, how well does the initial field have to reflect the true field? Are there specific components of the field that need to be captured in the initial estimate?~~

For HI, the estimated hydraulic conductivity field turns out to be very smooth in the top layers and almost homogeneous in the deeper layers, which is expected for this type of (Tikhonov) regularization. For SHI and JHI, the estimated hydraulic conductivity field resolves ~~some~~ 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 ~~from~~ for the ~~TEM data~~geophysical models either in the first step of the SHI scheme or during the JHI scheme can turn out to be very erroneous ~~either too small or too large~~ when the sensitivity of the TEM data with respect to resistivity is low. For the studied system, this shows that resistivities estimated by SHI or JHI must be used with caution as estimators of hydraulic conductivity or as regularization means for subsequent hydrological inversion. In this case, the use of the absolute relationship between hydraulic conductivity and ~~true~~ electrical resistivity led to an over-reliance on the use of inferred resistivities to populate the model's hydraulic conductivity field.~~. That is, much of the lack of value of the geophysical data~~ model ~~arose from a mistaken faith in the power of the petrophysical model relationship in combination with geophysical data of low sensitivity, thereby propagating geophysical estimation errors into the hydrologic model parameters.~~ In other words, even when there is a correlation between electrical resistivity and hydraulic conductivity, reliance on the relationship can lead to errors. This is exactly the kind of insight that can be gained from the use of HYTEB before collecting geophysical or other data that would be difficult or impossible to infer without this integrated platform.

With respect to reducing model prediction error, it depends on the type of prediction whether it has value to include ~~geophysical data~~ 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 ~~by~~ for models calibrated ~~using both hydrologic data and geophysical data models (~~by SHI or JHI~~)~~ as compared to models calibrated by ~~only using hydrologic data (~~HI~~)~~.

When the stress situation is changed from that of the hydrologic calibration data, then all models make biased predictions of head change. Use of geophysical data or models (with JHI or SHI ~~or JHI~~) reduces error and bias of head prediction at shallow depth but not in the deep part of the buried valley near the pumping well (where the stress field ~~was changed~~ change the most). Analyzing the prediction results by the method described by Doherty and Christensen (2011) indicates that ~~the geophysical data~~ 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 ~~data~~ better resolves the shallow variations in glacial deposits' resistivity than the variations inside the deep buried valley.

For model prediction of change of discharge to the stream, there is no improvement in using ~~the TEM data~~ 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 this data, but the estimated areal recharge turned out to be two thirds or less of the actual areal recharge. The predicted recharge area of the pumping well and the predicted age of the pumped water therefore turn out to be much too large. So another important insight from the HYTEB analysis is that recharge should be parameterized and estimated in a different way than it was done in the demonstration example. 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.

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

(2014) : 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

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

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 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, Binley A, Tryggvason A, Pedersen LB, Revil A (2006a) : Improved hydrogeophysical characterization using joint inversion of cross-hole electrical resistance and ground-penetrating radar traveltime data. Water Resour Res 42 (12):W12404. doi: 10.1029/2006WR005131

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

Lochbühler T, Doetsch J, Brauchler R, Linde N (2013) : Structure-coupled joint inversion of geophysical and hydrological data. GEOPHYSICS 78:ID1–ID14. doi: 10.1190/geo2012-0460.1

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.

Menke W (2012) Geophysical Data Analysis: Discrete Inverse Theory, Third Edition: MATLAB

Edition. Elsevier, Academic Press, Boston, USA.

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

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

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

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

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

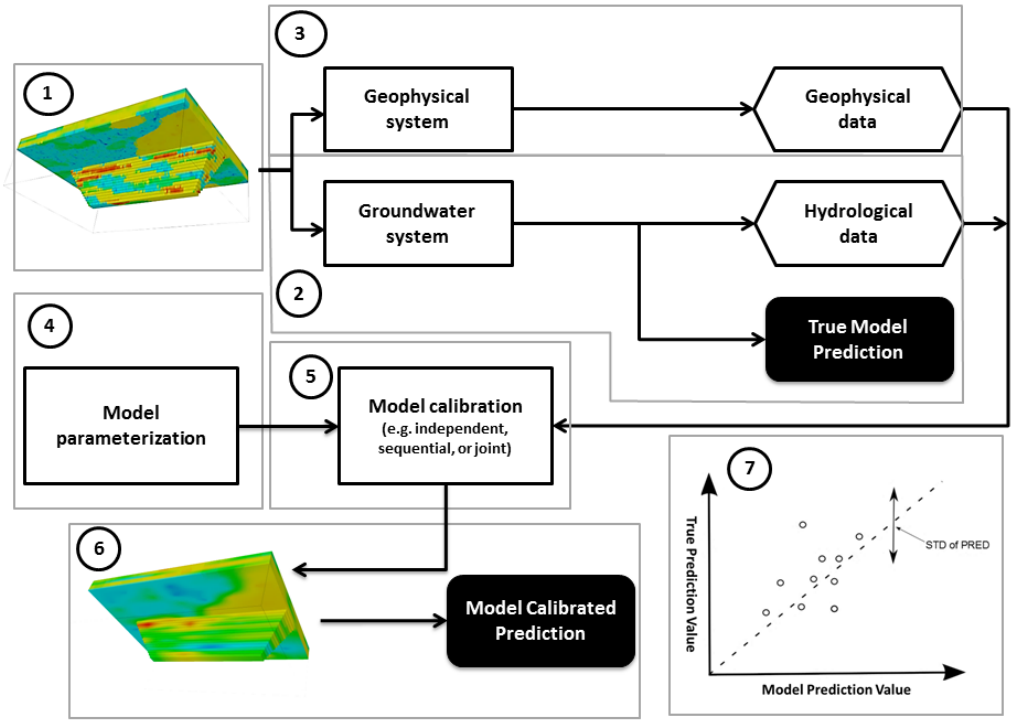

Figure 1: Workflow of the HYTEB. Each numbered dashed box marks a major step in the work flow. In parts 1 and 5 the red, yellow, blue and green colors indicate different categories (types) of geological deposits; color variation within each category (in part 6) indicates variation in hydraulic conductivity

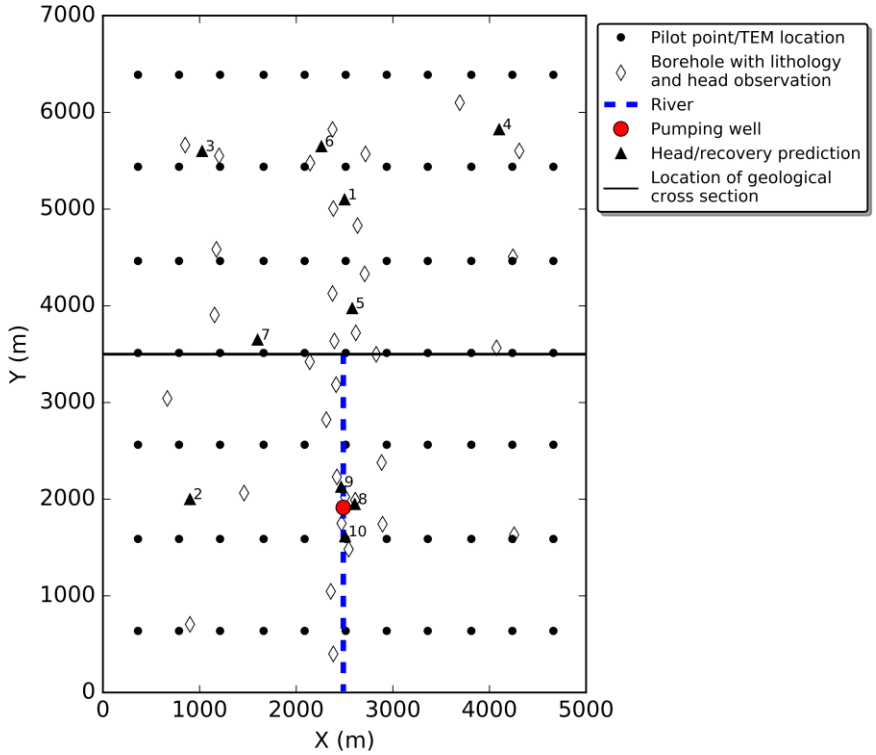

Figure 2: A map of locations of boreholes, a pumping well, geophysical data, pilot points, predictions of interest and location of a geological cross-section. (The positions of the pilot points and geophysical measurements are coincident.)

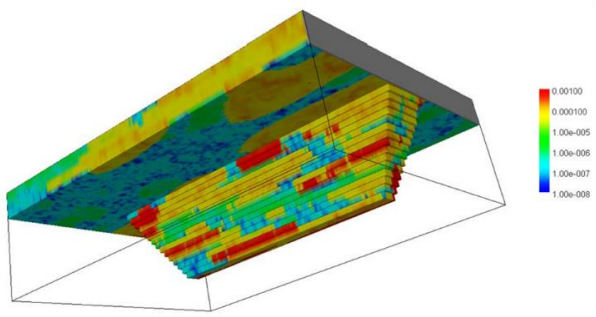

Figure 3. Hydraulic conductivity field for one of the model realizations. (Red shades are for gravel, yellow for sand, green for silt, and cyan/blue for clay.)

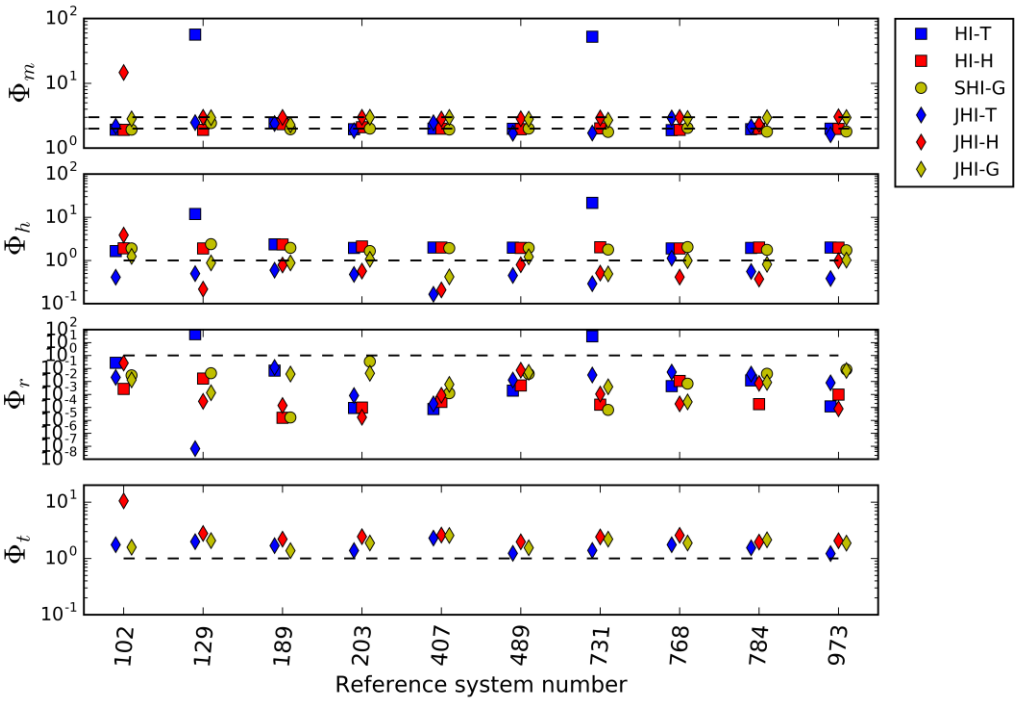

Figure 4: Measurement objective function value obtained for the various groundwater model calibration cases and for the 10 different system realizations. The two dashed lines in the top plot indicate~~s~~ the target value for the various model calibrations:~~.~~ ~~T~~the upper dashed line is the target 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.

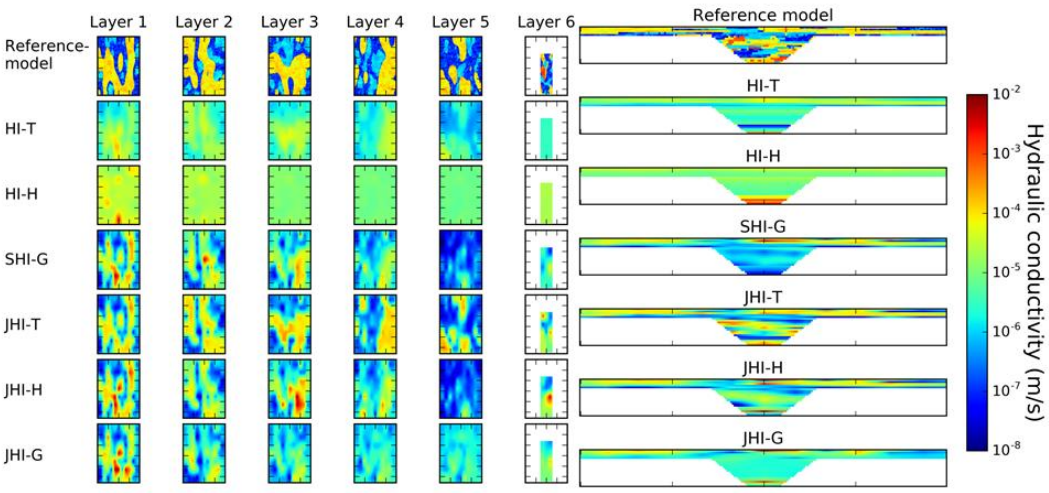

Figure 5: ~~True~~ Reference and estimated hydraulic conductivity fields for model realization number 189: a) shows the fields for layers 1 to 6 ; b) shows the field along an east-west cross section in the middle of the domain.

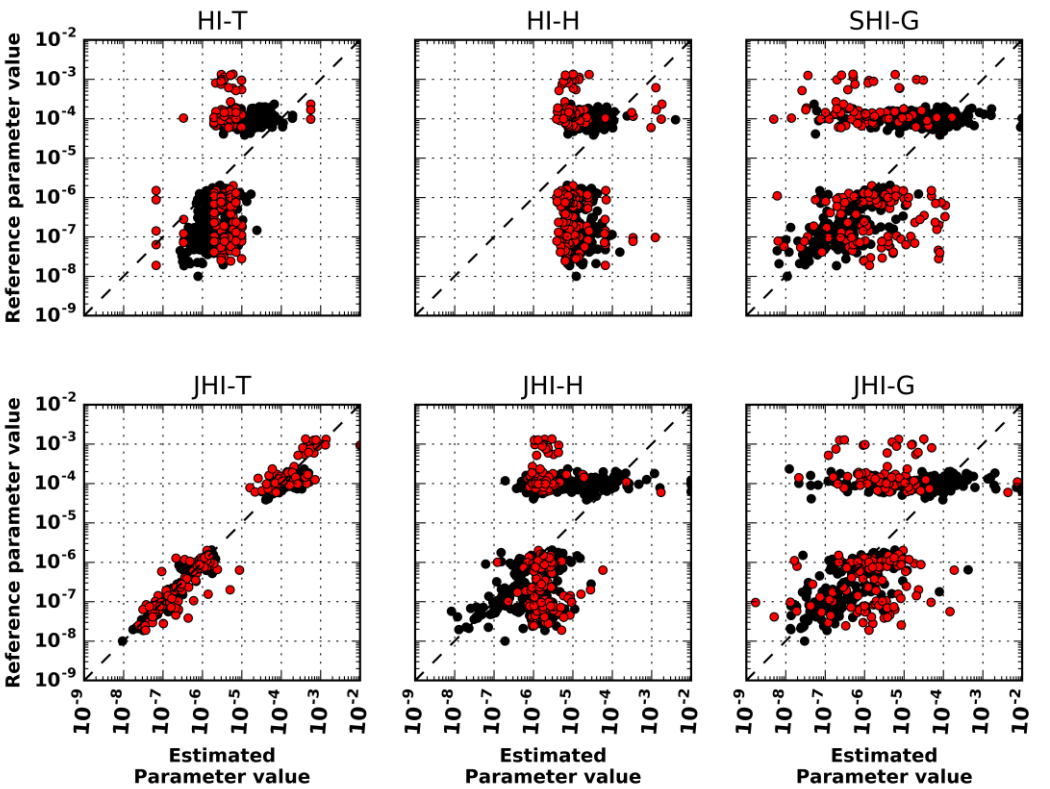

Figure 6. Pilot-point-by-pilot-point scatter plot of ~~true~~ reference versus estimated hydraulic conductivity for the six inversion runs. Black dots are estimated parameter values from the capping part of the model, while the red dots are estimated parameter values within the buried valley.

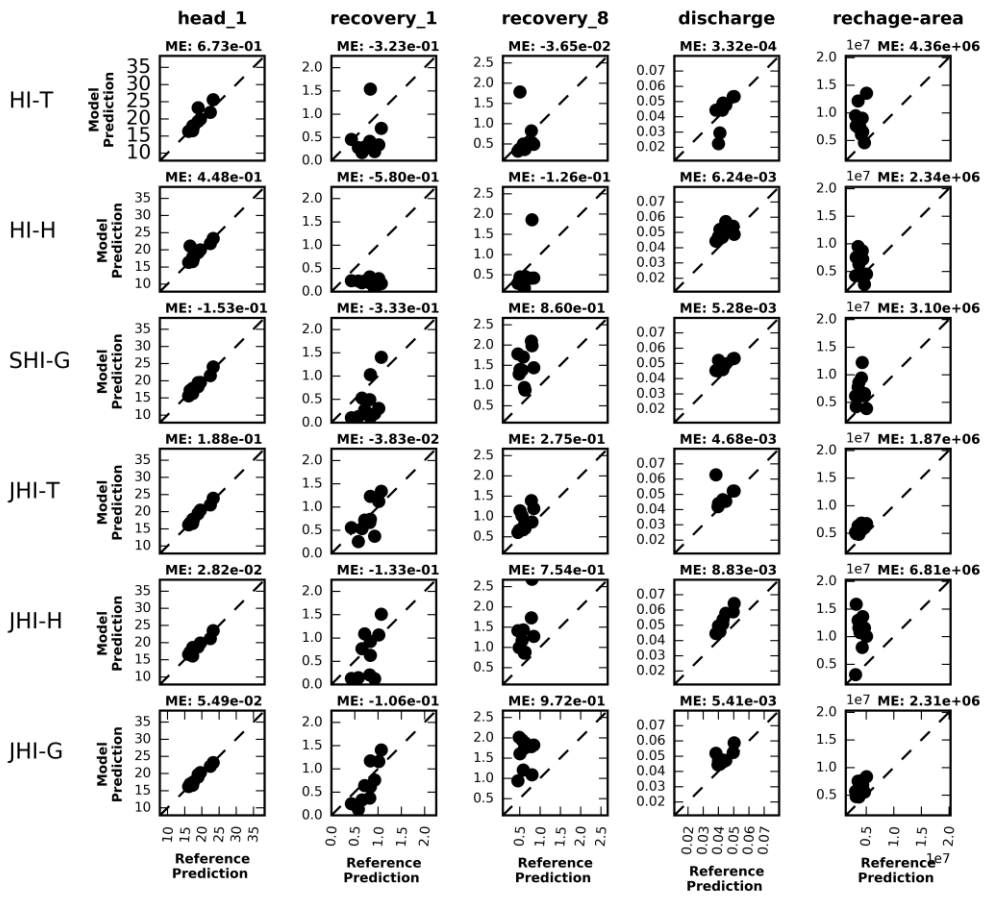

Figure 7: Scatter plots of calibrated model prediction versus ~~the true model~~ reference prediction ~~using results from the~~ for five predictions (explained in body text). Each plot shows results from ten ~~geological~~ system realizations. ~~The plots in the first column is for groundwater discharge to the river after pumping has stopped, the second column is for head recovery in the capping layer at location 1, and the third column is for head recovery within the buried at location 8 (Figure 2). ME quantifies the mean prediction error calculated on basis of the ten realizations.~~

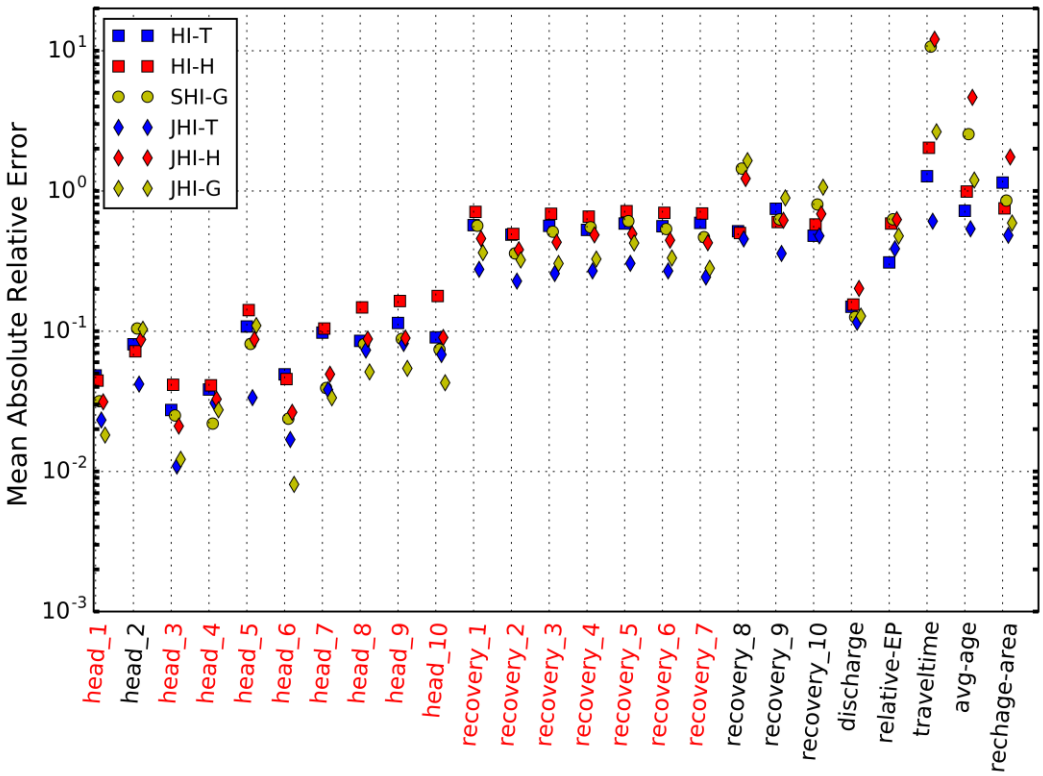

Figure 8: Mean absolute relative prediction error calculated from the ten geological realization results. The symbol type indicates the inversion approach and the symbol color indicates the initial parameter values used when calibrating the groundwater model. Red labels at x-axis highlight prediction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