

## ***Interactive comment on “A framework for testing the use of electric and electromagnetic data to reduce the prediction error of groundwater models” by N. K. Christensen et al.***

**N. K. Christensen et al.**

nkc07@phys.au.dk

Received and published: 25 November 2015

We would like to thank Anonymous Referee #1 for his valuable and relevant comments. Our replies are found below.

Answers to the general comments

1. 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

C5086

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.

1. 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.

2. 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 prevalent 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. Accounting for spatially variable resolution in electrical resistivity tomography through field-scale rock-physics relations By: Singha, Kamini; Moysey, Stephen GEOPHYSICS Volume: 71 Issue: 4 Pages: A25-A28 Pub-

C5087

lished: JUL-AUG 2006

A framework for inferring field-scale rock physics relationships through numerical simulation By: Moysey, S; Singha, K; Knight, R GEOPHYSICAL RESEARCH LETTERS Volume: 32 Issue: 8 Article Number: L08304 Published: APR 19 2005

Effects of spatially variable resolution on field-scale estimates of tracer concentration from electrical inversions using Archie's law By: Singha, Kamini; Gorelick, Steven M. GEOPHYSICS Volume: 71 Issue: 3 Pages: G83-G91 Published: MAY-JUN 2006

Inversion of tracer test data using tomographic constraints By: Linde, N; Finsterle, S; Hubbard, S WATER RESOURCES RESEARCH Volume: 42 Issue: 4 Article Number: W04410 Published: APR 18 2006

2. 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 geophysi-

C5088

cists we talked to at the early stage of this example investigation. Now in hindsight we can all be much cleverer.

3. 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 any 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.

3. 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

C5089

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.

4. 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

C5090

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

4. 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.

5. The title should be changed by removing electric data (only electromagnetic data are shown).

5. We will remove “electric” from the title.

6. 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.

C5091

6. The reviewer is right. We will go through the manuscript and remove confusions between geophysical “model” and “data”.

7. 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 flexible).

7. 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.

8. 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.

8. This is a valid point and we will present the calibration results.

9. 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.

9. We appreciate the reviewer’s opinion on this. We will consider this criticism in detail

C5092

when we have received the remaining reviews.

Answers to the specific comments  
Smaller comments: 9600, line 2: Add “that” before geophysics. The text will be modified.

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.

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

9600, line 11: I would replace “and approaches to correlating” with “used to correlate”. The text will be modified. 9600, line 15: Perhaps state that the bedrock is clay. Personally I expected resistive igneous rocks. The text will be modified.

9600, line 17: These “resistivity estimates” forms a model that here are assumed to be “data” with well-known consequences of this choice.

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?

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.

C5093

9601, line 20: Replace “ramifications” with “impact”. The text will be modified.

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.

9602, line 16: Sensitivity to what? To “small scale heterogeneity”. Isn’t this obvious when reading the entire sentence?

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.

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”.

9603, line 27: Also cite: Accounting for spatially variable resolution in electrical resistivity tomography through field-scale rock-physics relations By: Singha, Kamini; Moysey, Stephen GEOPHYSICS Volume: 71 Issue: 4 Pages: A25-A28 Published: JUL-AUG 2006 A framework for inferring field-scale rock physics relationships through numerical simulation By: Moysey, S; Singha, K; Knight, R GEOPHYSICAL RESEARCH LETTERS Volume: 32 Issue: 8 Article Number: L08304 Published: APR 19 2005 Effects of spatially variable resolution on field-scale estimates of tracer concentration from electrical inversions using Archie’s law By: Singha, Kamini; Gorelick, Steven M. GEO-

C5094

PHYSICS Volume: 71 Issue: 3 Pages: G83-G91 Published: MAY-JUN 2006 Inversion of tracer test data using tomographic constraints By: Linde, N; Finsterle, S; Hubbard, S WATER RESOURCES RESEARCH Volume: 42 Issue: 4 Article Number: W04410 Published: APR 18 2006

We will include the suggested references if we find them relevant.

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).”

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.

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; Piro, Guillaume; Straubhaar, Julien; et al. MATHEMATICAL GEOSCIENCES Volume: 46 Issue: 5 Special Issue: SI Pages: 625-645 Published: JUL 2014 We will read the papers and consider reformulation.

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.

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

C5095

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).

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.

9606, line 2: I would remove “for making experiments”. The text will be modified.

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.

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”.

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

C5096

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 welcome a better wording suggested by the reviewer or the 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.)

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.

9610, title 2.4: Please reconsider the name of the section. The section name will be changed to “model parametrization”

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

C5097

this purist viewpoint.

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".

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.

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 is 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.

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 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.

C5098

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. "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 be noisy to be applied in the further analysis."

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.

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.

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

C5099

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.

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); that 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 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.

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.

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.

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 12, 9599, 2015.

C5100