Dear authors and Editor,

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

1. It is still confusing if the manuscript is about HYTEB or a comparison study of sequential and joint inversion. As pointed out by the Editor in the early stage of the review process, the methodology (behind HYTEB) described here is common to a lot of studies considering geophysical data or hydrogeological modeling (i.e. considering synthetic benchmarks to assess a new technique). A technical paper (probably in a different journal) about HYTEB would be interesting if it was presenting an end-product and demonstrated with various examples (different geophysical data and modeling tools, different inversion procedures, etc.). For such a paper, a much more thorough description of the way HYTEB works would be needed. Is HYTEB very flexible if one wants to use another forward simulator or inversion packages? How can specific petrophysical relationships be introduced, relating other parameters (for example porosity and resistivity)? How easily can the user define his own objective function? What about transport (e.g. time-lapse geophysical data)? What are the limitations (stochastic inversion, use of McMC, multiple-point geostatistics, etc.)?
   On a scientific basis, we are more interested in the outcomes of the comparison study. I suggest to limit the reference to HYTEB to the methodology section and avoid further references in the text (stating that the methodology has been implemented in a software called HYTEB, available at …). Many sentences referring to author choices are followed by comments in parenthesis saying that it is not limited by HYTEB. This should be avoided, since the paper is actually not about HYTEB.
2. Considering the first comment, I think that the authors should focus on their comparison of sequential and joint inversion. They should propose a more thorough literature review and discussion of both sequential and joint inversion approach to highlight the novelty of their approach (is it the comparison between the techniques, the level of heterogeneity considered, the scale, the use of EM data? Given the cited references, I would suggest to insist on the heterogeneity and the number of parameters to optimize).
   As an example, the paper of Lochbuhler et al. (2013) describes a structure-coupled inversion in terms of penalizing cross-gradient between geophysical and hydraulic parameters as traditionally done in joint geophysical inversion, avoiding the need for a petrophysical relationship, which is quite different of the proposed approach.
   On another level, the authors do not discuss stochastic inversion (e.g. Irving and Singha, 2010, which is also a joint inversion with transport data) or the possibility to use a Bayesian technique in the sequential inversion which inherently introduce some uncertainty in the petrophysical relationship (Dubreuil-Boisclair et al. 2011, Ruggeri et al., 2013, 2014, for an application to sequential hydrogeological inversion, see Hermans et al., 2015).

3. The importance of the choice of the petrophysical relationship has already been raised previously. I agree with the authors that this is a simplification that can be accepted for a synthetic case (direct link between hydraulic conductivity and resistivity) to simulate a realistic geophysical data set. However, even if such a relationship is assumed to build the data and perform the joint inversion approach (where the updated hydraulic conductivity are transformed in resistivity to simulate the geophysical data), it is clearly not valid the other way around in the sequential inversion. The limited resolution of geophysical method and inversion will degrade the relationship. This will not necessarily degrade the results. Indeed, if we assume that the inversion of geophysical data smoothed the resistivity distribution, then the range of variation of resistivity in the inverted model is small and high/low hydraulic conductivity will not be represented in the hydrogeological model. This should be taken into account in the sequential approach (which will also influence the JHI-G results)
4. The authors seem to use some estimates of resistivity to deduce the hydraulic conductivity to constrain the joint inversion approach. Clarification is needed, since the basic idea of joint inversion is to avoid geophysical inversions so that there is no error occurring from the "bad" recovery of geophysical parameter after inversion. Adding such a constraint would therefore re-introduce some error due to the limited resolution of geophysics and limit the improvement brought by the method.

Specific comments

1. P2.L25-29. The authors describe here one way to tackle the hydrogeological problem, which is the deterministic approach. They ignore the stochastic approach to solve the problem and propose predictions with uncertainty estimate (see Zhou et al., 2014).
2. P3.L20. See also Xu and Valocchi for Bayesian inference to manage structural error.
3. P3. The prediction can also be uncertain because (iv) the non-unicity of the solution. Many models can explain the data, but could produce different predictions (v) Error in the prior distribution of parameter (for example in this case the pilot-point method will generate a smooth model, whereas the truth is much more heterogeneous and can only be represented by a prior corresponding to the one used to generate the true model). To calibrate the model while conserving the prior, methods were developed such as the gradual deformation (Roggero and Hu, 1998) or probability perturbation method (Caers, 2003).
4. P3.L21-L26. You don't include isotopic or chemical data, why mentioning them? HYTEB does not seem to be able to integrate such analysis.
5. P5.L3. See major comment 2. With probabilistic relationship, uncertainty related to inversion and petrophysics is included and the proposed hydrological model is in accordance with the data.
6. P5.L4, "extract more information" is not clear, especially given the outcomes of this study. More information on the hydrogeological parameter? More accurate information because the regularization step is avoided?
7. P5.L13. Note that the assumed relationship can be considered uncertain (see Irving and Singha, 2010).
8. P5.L20-25. You don't answer all of those questions in the paper.
9. P5.L12-23. I would not call HYTEB a framework, since the described framework has been applied in many studies in hydrogeophyiscs. It is a software or a platform.

10. P6.L28. Delete “mainly electric and”, only electromagnetic is shown.
11. P7.L7. HYTEB instead of HYTEM.
12. P7L28. Are T-Progs and Blocksis integrated to HYTEB? What about other metods such as multiple-point geostatistics (SNESIM, Direct sampling)? Can they also be used in a stochastic inversion framework? Why is this step necessary, can we directly simulate a continuous field with sgsim (step 2)?
13. P8.L28. If it is not the case, why would one use geophysical data to calibrate a hydrogeological model?
14. P9.L5. Random error is not able to represent all sources of error measurement such as systematic error linked to the measuring device. Remove “to represent all sources of error”.
15. P10.L13-14. Zones and pilot-point are two types of parameterizations, but they are not any type. Does HYTEB works only with deterministic optimization approaches (where the number of parameter must be relatively small?)
16. P14.L11. Your framework does not assess uncertainty in the prediction, it calculates one prediction for the calibrated model. So the actual influence of the geophysical data on the uncertainty of the estimate cannot be estimated. Uncertainty estimates would require a stochastic approach.
17. P15.L19. See my major comment 3. The relationship between resistivity after inversion and hydraulic conductivity should be checked as well to verify that.
18. P15.L24-26. Delete this kind of statement (see major comment 1), here and later.
19. P16.L15. Better give the noise level than refer to another study.
20. P16.L25. What is the true correlation length? Given the categorical nature of the true model, the variogram is probably not able to capture the heterogeneity and the pilot-point method not well suited for inversion.
21. P16.L30. Your results show the contrary.
22. P17.L7-8. Delete what is in parenthesis.
23. P17L19. Delete, Traditionally, it suggests that scientists rely only on what was done in the past and do not improve their methods.
24. P17L12. It is not clear how mu is optimized, line search method?
25. P17L14. Do you mean equation 2 (not 4)?
26. P19L5L8. This is not clear to me. Why in a joint inversion would you have two different estimates of K? I expect that a K value is obtained trough the inversion process, then, based on this K field, you can calculate the outputs of the model (heads, discharge, and simulated geophysical data through the petrophysical relationship). There is no need to estimate the hydraulic conductivity from the resistivity, it is the other way around. Actually, the whole point of joint inversion is to avoid the regularization of geophysical data (and therefore smoothing) to estimate the resistivity. See major concern 4. Please clarify.
27. P20L16. Does it verify equation 1?
28. P21L15-17. You don’t study the effect of conceptual models, boundary conditions, etc. Why would you use such characteristic to group them?
29. P22L16. Of course, the minimization problem is dependent on the starting model, this is a well-known fact. Also, the procedure does not ensure that you find a global optimum.

30. P23L13-L16. This shows that the hydraulic conductivity field has little influence on the steady-state results. The influence of the boundary conditions and pumping rate are probably preponderant.
31. P23L18-L20. This is expected given the smooth model recovered by the pilot-point method. Basically, Figures 5 and 6 show the same results since the spatial distribution is unequivocally determined by the value at the pilot-points and the variogram. What about the comparison of hydraulic conductivity at other locations? This would highlight the limitations of the pilot-point method to recover the true conductivity field.
32. P24L12-L14. A petrophysical relationship accounting for the limitations of EM inversion (for example by comparing co-located measurements after inversion) could reduce this and make a more realistic synthetic benchmark. See my major comment 3.
33. P24L22-L23. Not sure about that. According to the cross-section, JHI-H performs better.
34. P25L1-L2. Figure 7. Isn't it possible to show the results for all ten realizations at all locations with some color code, as you did for figure 6?
35. P25L21-L26. Can you see a particular reason for that? If this can be observed for the 10 models, then it is not related to the geological heterogeneity. Points 8, 9 and 10 are very close to the pumping well what suggests an influence of the pumping, maybe related to grid discretization?
36. P26L7. I think head recovery at location 1 are underpredicted, as they fall below the identity line and the model prediction is on the Y-axis. Please verify.
37. P26L21. Replace above by below?
38. P26L31. "Explanation for this predictive degradation is given above". Where? It is not clear.
39. P27L9-L13. I would remove, this is highly speculative. The results seem very similar for all models and JHI-G and SHI-G are as good as the other, or even slightly better. This also shows that the mean error is a poor choice here, because the overprediction of 8 points counterbalances the huge underprediction of the two other points for HI-T, leading to the smallest error. I would rephrase the whole interpretation of the discharge prediction to be coherent with figure 7.
40. P27L30-P28L4. This behavior is expected. Using the pilot-point method as calibrating method, smoothing is introduced in the distribution of K. If it may be sufficient to explain steady-state head and discharge, it affects much more flow paths and therefore transport (and in this case recharge, as it is linked to hydraulic conductivity). To get reliable transport estimate, a more geologically realistic distribution of K should be obtained, i.e. one which is consistent with the prior.
41. P28L8. The framework is not new, the software might be.
42. P29L15. "…erroneous when the sensitivity of the TEM data with respect to resistivity is low". In this case, geophysical data should be disregarded. See Beaujean et al. (2014) for an example of model filtering for sequential hydrogeological inversion using geophysical data.
43. P29L18-23. This is actually common sense, basically what you suggest is to always perform synthetic modeling before acquiring data, which is a common practice in research. Also, the fact that the use of the chosen petrophysical relationship will lead to error in using inverted resistivities is straightforward and does not require a full synthetic case.

Proposed References

Irving and Singha, 2010. Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities, Water
Resour. Res., 46, W11514,

Dubreuil-Boisclair et al. 2011. Heterogeneous aquifer characterization from ground-penetrating radar tomography and borehole hydrogeophysical data using nonlinear Bayesian simulations, Geophysics, 76(4), J13-J25.

Ruggeri et al., 2013. Regional-scale integration of multiresolution hydrological and geophysical data using a two-step Bayesian sequential simulation approach, Geophysical Journal International, 194, 1, 289-303.

Ruggeri et al., 2014. Integration of hydrological and geophysical data beyond the local scale: Application of Bayesian sequential simulation to field data from the Saint-Lambert-de-Lauzon site, Québec, Canada. Journal of hydrology, 514, 271-280.

Hermans et al. 2015. Uncertainty in training image-based inversion of hydraulic head data constrained to ERT data: Workflow and case study. Water Resources Research, 51, 5332-5352.

Zhou et al., 2014. Inverse methods in hydrogeology: Evolution and recent trends, Adv. Water Resour., 63, 22–37.

Xu and Valocchi, 2015. A Bayesian approach to improved calibration and prediction of groundwater models with structural error, Water Resour. Res., 51, 9290–9311.

Roggero and Hu, 1998. Gradual deformation of continuous geostatistical models for history matching, paper presented at Proceedings Society of Petroleum Engineers 49004, Annual Technical Conference, Soc. of Pet. Eng., New Orleans, La.

Caers, 2003. Geostatistical history matching under training-image based geological model constraints, SPE J., 8(3), 218–226.

Beaujean et al. 2014. Calibration of seawater intrusion models: Inverse parameter estimation using surface electrical resistivity tomography and borehole data, Water Resources Research, 50, 6828-6849.