The Cryosphere Discuss., 5, C2085–C2093, 2012
www.the-cryosphere-discuss.net/5/C2085/2012/

The Cryosphere Discussions

Interactive comment on "Optimisation of quasi-3D electrical resistivity imaging – application and inversion for investigating heterogeneous mountain permafrost" *by* D. Schwindt and C. Kneisel

A. Kemna (Referee)

kemna@geo.uni-bonn.de

Received and published: 14 March 2012

This paper describes a numerical study on the resolving power of 3D electrical resistivity tomography (ERT) in permafrost studies for different data acquisition strategies and measurement configurations. In particular imaging results based on data sets collected along parallel acquisition lines are compared with those obtained for additional perpendicular lines. A considerable amount of scenarios are being considered this way, based on a 3D model which exhibits heterogeneity as representative of a typical permafrost site. The objective is not only the qualitative comparison of the imaging results but

even the quantitative comparison of inverted resistivity values, as it is of importance if for instance ice contents are to be estimated from inverted resistivity values using petrophysical relationships. Given the increasing use of ERT in permafrost studies, the manuscript represents a relevant and timely contribution.

However, the methods and tools used in the paper to achieve the objectives (assessment of images, optimization procedure and stopping criterion in the inversion) are not state of the art in the field of quantitative ERT. Important aspects and approaches in quantitative ERT, which are well known and documented in the literature, are ignored:

(1) On the one hand this holds for the different methods available for image appraisal, i.e. the assessment of the imaging results. Here a number of approaches based on the analysis of cumulated sensitivity, resolution matrix or model parameter uncertainty are at hand (which all account for the varying resolution of ERT in different regions of the model); however, the authors do not adopt any of them for a more differentiated analysis of their simulation results. References in this context include for example: Ramirez et al., 1995, J. Environ. Eng. Geophysics; Oldenburg & Li, 1999, Geophysics; Alumbaugh & Newman, 2000, Geophysics; Friedel, 2003, Geophys. J. Int.; Day-Lewis, Singha & Binley, 2005, J. Geophys. Res.; Nguyen et al., 2009, Near Surface Geophysics.

(2) The other issue which is not properly accounted for in the manuscript is the influence of data error, its description in the inversion, and – closely related – the applied criterion to stop the inversion iterations on the imaging result. Instead of rigorously computing the smoothest image which fits the data to a predefined degree (depending on the data error level), the final model is chosen in a rather arbitrary way. Importantly, chosen images do not yet explain the data within the assumed errors – as revealed by Figure 6 –, which suggests problems of the used code to model/invert with the required accuracy. Also for the field data no information/analysis regarding the data errors is included. However, this is essential to avoid over-fitting and possible misinterpretation. Useful references addressing the issue of data errors in ERT and their description in

the inversion scheme are for example: LaBrecque et al., 1996, Geophysics; Slater et al., 2000, J. Appl. Geophysics; Koestel et al., 2008, Water Resources Res.; Krautblatter et al., 2010, J. Geophys. Res.

There are also some problems with the used terminology. For instance, the term "reliability" (e.g. in the Abstract) is not appropriate in the given context because it lacks proper definition. When is an inverted value reliable, when not? The relevant quantity here is "uncertainty", which can be computed for each resistivity value of the inverted model (e.g. Alumbaugh & Newman, 2000, Geophysics). I also disagree with the use of the term "quasi-3D" (see list of specific comments below for more details on this).

A final general remark is on the title: I don't think that "optimization" is justified here. Optimization implies a systematic procedure where by some means one approach is preferred over a range of other approaches. What is done in the manuscript is rather a comparison and assessment of different approaches, but not an optimization. There are actually papers around which address the data acquisition strategy from an optimization point of view, including: Furman et al., 2004, Vadose Zone J.; Stummer et al., 2004, Geophysics; Furman et al., 2007, Geophysics.

The above problems (1) and (2) are of fundamental nature given the scope and objectives of the paper and therefore put the subsequent discussions and conclusions at question. These issues will have to be properly discussed and accounted for in the manuscript, involving considerable revision, before it can be published.

List of specific comments:

1) P 3384, lines 15-16: This statement only makes sense if the characteristic length scale of the heterogeneity (size of anomalies, correlation length of structures with statistical heterogeneity etc.) is related to the electrode separation.

2) P 3385, line 1: What is meant with "small-scale"? m, dm . . .?

3) P 3385, line 16: Dahlin and Loke (1997) is not a valid reference here (because

they only apply 2D inversion). Other references in addition to Loke and Barker (1996) should be cited, in particular earlier work such as Park and Van (1991, Geophysics), Sasaki (1994, Geophysics), Ellis and Oldenburg (1994, Geophys. J. Int.). See e.g. references in Günther et al. (2006, Geophys. J. Int.)

4) P 3385, lines 19-20: Put references in chronological order.

5) P 3385, lines 26-29: Of course there's many other works on 3D ERT. It would make sense here to cite a review chapter such as "Daily, W., Ramirez, A., Binley, A., and LaBrecque, D., 2005. Electrical resistance tomography – theory and practice, in Butler, D.K., Ed., Near-Surface Geophysics, Investigations in Geophysics, 13: Society of Exploration Geophysicists, Tulsa, 525-550" and/or "Binley, A., and Kemna, A., 2005. DC resistivity and induced polarization methods, in Rubin, Y., and Hubbard, S.S., Eds., Hydrogeophysics: Springer, 129-156"

6) P 3386, lines 3-23: I strongly disagree with the discussion in this paragraph and the proposed convention for the use of "quasi-3D". Actually the terminology is not at all unclear or inconsistently used among ERT developers: "2D" or "3D" refers to the parameterization in the underlying inversion, i.e., whether the resistivity distribution is allowed to vary in two or three dimensions, respectively. Of course, the resolution of model parameters in 2D or 3D requires appropriate data acquisition. So for instance nobody would expect to fully resolve a 3D image from a data set collected along a single linear array of surface electrodes. However, it is certainly not necessary – as long as isotropy is assumed (but that's beyond the scope of the paper) – to collect data using crossing (i.e. parallel and perpendicular, and perhaps diagonal too) surface lines. Parallel lines can be indeed sufficient because of the 3D measurement volume of a collinear 4-electrode array (i.e. the associated current and potential field lines do also penetrate the region outside the vertical plane defined by the collinear electrode array). The advantage of additional perpendicular acquisition lines may be the general improvement of the signal-to-noise ratio in the whole data set (by having more measurements) and perhaps a more homogeneous coverage of the subsurface (in terms

of cumulated sensitivity). “Quasi-3D” refers to building a 3D image from independent 2D inversions, and not to running a 3D inversion of 3D data collected along several parallel acquisition lines. To avoid further confusion among ERT users, the paragraph (and actually the whole manuscript) needs to be revised in this respect. I also suggest differentiating between the inversion process and the data acquisition strategy here.

7) P 3386, line 27: “This results in . . .” should be rephrased because it suggests a strict implication which is not at all the case.

8) P 3387, lines 11-12: on “to optimize . . . by forward modelling”: I think it is rather an investigation than an optimization; moreover it is not only forward modelling but also the inversion of simulated data, so something like “investigate . . . in numerical simulations” would be more appropriate.

9) P 3387, lines 16-17: “imaging” instead of “illustrating”; verb missing with respect to “resistivity values” (or positioning of “(1)”, “(2)” not correct).

10) P 3387, line 23: Note that the stopping criterion of the inversion is directly related to the issue of data errors.

11) P 3388, lines 3-5: This statement should be rephrased and, if at all, only key references on this general topic should be given (only Scales and Snieder (2000) qualifies in this respect; however, it’s better to refer to standard text books on the topic). The statement suggests that there are several studies dealing with geophysical forward modelling and inversion in recent years, while actually geophysics mainly CONSISTS of geophysical modelling and inversion and hence there’s thousands of studies in the last 50 years (or more) on this. I suggest focussing on ERT only, where still a huge amount of literature exists.

12) P 3388, lines 9-15: Again, these are all fundamental aspects of inversion theory in general and resistivity inversion in particular, which is not appropriately reflected by the selected references Hilbich et al. (2009) and Fortier et al. (2008). The “back-

and-forth interpretation procedure” by Fortier et al. is what geophysicists do for ages, i.e. setting up a model, simulating data and comparing inversion results for different inversion settings, data error characteristics etc.

13) P 3389, lines 20-25: Please provide reference with an explanation of the “Wenner-Schlumberger” configuration; should be “dipole-dipole”; however, I’m not sure what the authors actually want to say here (“more sensitive to horizontal changes . . . and therefore a good tool for mapping vertical structures” is contradictory); actually the dipole-dipole array is likewise sensitive to resistivity changes in both horizontal and vertical directions

14) P 3389, line 27: I understand that “grid size” here only refers to the spatial extension of the grid (which is normally related to the electrode layout and thus electrode spacing and number of electrodes); however, grid size may also refer to the size of an individual grid cell or the number of total grid points (both being independent of electrode locations) – make clear what is meant here

15) P 3389, line 27 – P3390, line 4: I don’t really understand the interdependencies between electrode spacing, number of electrodes etc. and grid specifications. If I understand it correctly, then electrode spacing is directly linked to the grid increments? If so, then the corresponding simulations differ systematically, with (unknown) effect on inversion behaviour, resolution of the final results etc. Why don’t you use the same grid for all simulations? Then any grid effects on the imaging results could be ruled out. What is the idea of “roll-along” here (which actually is a data acquisition concept)? Why don’t you just set up the grid as you need it?

16) P 3390, lines 11-15: Here I am a bit confused: I understood before that data along linear survey lines are modelled for a 3D resistivity distribution. Could you please specify which modelling code (and not only the type of modelling approach) was used? And you should explicitly state that the code is 3D to avoid confusion about the dimensionality in the modelling.

17) P 3390, lines 15-16: 3% is generally very optimistic for field measurements (actually even the numerical modelling error can easily amount to more than 3%), definitely for surveys in permafrost environments. Therefore this statement should be rephrased to avoid the impression that 3% is a typical value. Did you use different noise ensembles for the different data sets?

18) P 3390, line 21: on "robust inversion": this term refers to the use of the L1 norm to measure data misfit and/or model roughness in the objective function (which is being minimized). Please specify for which term the L1 norm was used. Following the argumentation in the text, it should only be applied to the measure of model roughness, not to the measure of data misfit. Is this correct?

19) P 3390, lines 22-24: How do you know that the inversion has indeed computed the smoothest possible model that explains the data within the given uncertainty (3%) always after 5 iterations (if at all)? For a non-linear inversion problem the inversion behaviour can vary strongly, particularly depending on the data errors (e.g. the used noise ensemble in the present case) and the resistivity distribution itself, exhibiting either slow or fast convergence. For quantitative ERT, the inversion process has to be completed, i.e. the smoothest model has to be computed that provides a data-error weighted RMS value of 1. With the applied approach you might arbitrarily over- or under-fit the data and thus produce inconsistent images. This is an absolutely crucial point if afterwards inverted resistivity values are interpreted quantitatively.

20) P 3390, line 26: What does "1% accuracy" refer to? Is there any validation for this statement?

21) P 3391, line 5: "dipole-dipole" (check throughout the whole manuscript)

22) P 3391, lines 10-11: Using the 3rd iteration is totally arbitrary. On that basis a fair assessment of the resolving power of ERT is not possible. See comment #19 above.

23) On the section "Reproduction of default resistivity values": Based on the arbitrariness in the selection of the final model (see comment #19), the whole paragraph is pointless in my opinion (and to me the arbitrariness explains some of the inconsistencies observed regarding over-/underestimation). All data sets need to be fitted exactly to the level of data noise. Different noise levels could then be considered to investigate the influence of the data noise on the recovery of specific resistivity values.

24) P 3395, lines 15-22: The cited statement by Hauck and Vonder Mühll supports a very important misconception, again related to comment #19 above: The number of iterations is by no means a well-defined criterion to decide upon whether the inversion has converged, i.e. the optimization problem is solved. Whatever the number of iterations is (which will normally depend on the iterative scheme used to solve the non-linear inverse problem, the accuracy in the computation of the Jacobian matrix if a gradient-based scheme is used, etc.), the mathematical task is to find the smoothest model which yields a well-defined data misfit (depending on the a-priori data noise level) (note that the task is not to minimize the data misfit! – which would never prevent from over-fitting the data, with the obvious problem mentioned in the last sentence of the citation by Hauck and Vonder Mühll). Normally this is achieved by, once a data-error weighted RMS value of 1 is reached, increasing again the value of the regularization parameter to indeed find the smoothest model. If a code is not doing this, then it is unclear which solution is actually computed and thus it can hardly be used for a quantitative ERT study!

25) P 3395, line 23: On Figure 6: What would be interesting to see in addition to the RMS curve (as a function of iteration) are the curves of the model roughness and the regularization parameter. The model roughness curve will indicate whether there is already any convergence here (note that seeking the smoothest model implies convergence of the model roughness curve). However, Figure 6 reveals another problem, which I presume is a general problem of the whole study: The RMS value at the 5th iteration is approx. 12%! This means that the inversion is not able to explain the data to the desired degree, in turn indicating that the numerical modelling errors of the used

code are much larger than the assumed data noise level of 3%. This raises the very critical question on the point of investigating the quantitative imaging capability of ERT with a tool that obviously has not the sufficient accuracy. To me the only solution here is to add noise to the data which is larger in amplitude than these modelling errors. Then of course all results would suffer from this relatively large noise level and the different comparisons might become pointless.

26) P 3398, line 16: Res2Dinv? How was a 3D inversion performed with Res2Dinv?

27) P 3398, lines 17-18: On which basis were the inversion parameters adjusted? Sounds again arbitrary. Why didn't you use the settings of the numerical study?

28) P 3398, line 26: see comment #20

29) P 3399: What is missing here, but absolutely necessary, is the analysis of the data error. Whatever approach is used, you have to come up with some estimate of the data error in order to avoid over-fitting the data in the inversion and thus avoid misinterpretation and invalid conclusions. Given the number of severe concerns outlined so far, I decided to not provide a detailed review of the Discussion and Conclusions sections at this point. All results are too much biased by the problems pointed out above, and many of the issues reappear here.

Interactive comment on The Cryosphere Discuss., 5, 3383, 2011.