TOWARDS OPTIMAL EXPERIMENTAL DESIGN IN HYDROGEOPHYSICAL STUDIES

J.A. HUISMAN[1], T.P.A. FERRE[2], A. KEMNA[1], H. VEREECKEN[1]

[1] *ICG-IV Agrosphere, Forschungszentrum Jülich, 52425 Jülich, Germany.*

[2] *Department of Hydrology and Water Resources, University of Arizona, USA.*

ABSTRACT

Optimal experimental design can be used to optimize the design of tomographic hydrogeophysical measurement surveys. The aim of this paper is to review and compare three experimental design methods by applying them to two case studies: travel time tomography with ground penetrating radar (TT-GPR) and electrical resistivity tomography (ERT). Both TT-GPR and ERT case studies showed that the method based on a normalized sensitivity can give unwanted measurement designs in terms of model resolution and expected inversion quality. The eigenvalue-based design method performed well in the TT-GPR case, but was too computationally intensive for a meaningful result in the ERT case study. Finally, the method based on a normalized sensitivity weighted with the relative model resolution resulted in good results in the TT-GPR and moderate to good results in the ERT case study.

1. INTRODUCTION

Hydrogeophysics attempts to combine multiple geophysical data sources to better constrain hydrological models of the subsurface. It is generally recognized that such data fusion attempts will be computationally demanding. Sophisticated (hydrogeophysical) inversion methods are one potential way to reduce the computational burden. Methods to reduce the number of measurements by avoiding redundant information (optimal experimental design) could also contribute to such a reduction. Such an optimisation of the experimental design requires that the quality of all possible designs can be quantified and maximized. In this contribution, we first review three methods to quantify the quality of the design. Then we apply these methods to two simple case studies based on travel time tomography with ground penetrating radar and electrical resistivity tomography, respectively. The main aims of these case studies are to compare the performance of the design methods and to gain insights in the feasibility of experimental design in hydrogeophysical studies.

2. REVIEW OF EXPERIMENTAL DESIGN METHODS

There have been a number of studies dealing with optimal experimental design issues in (hydro)geophysics. Most of these studies are based on linearized inverse modelling and, therefore, we introduce some basics of inverse theory before discussing several advanced experimental design methods. Inverse modelling starts with the linearized forward problem, which can be written as

$$\left(d^{obs} - d^{ini}\right) = G\left(m^{true} - m^{ini}\right) \quad (1)$$

where d^{obs} are the observed data corresponding to the unknown true model, m^{true}, and d^{ini} are the simulated data belonging to the initial model estimates, m^{ini} and G is the Jacobian matrix with the sensitivities ($G_{ij} = \partial d_i/\partial m_j$). The least-squares solution of the problem in Eq. 1 is:

$$m^{est} = m^{ini} + \left(G^T G + C_M^{-1}\right)^{-1} G^T \left(d^{obs} - d^{ini}\right) = m^{ini} + G^{-g}\left(d^{obs} - d^{ini}\right) \quad (2)$$

where m^{est} are the estimated model parameters, G^T is the transpose of G, C_M contains the regularization constraints and G^{-g} is the so-called generalized inverse matrix. By using the linear approximations $d^{ini} \approx Gm^{ini}$ and $m^{ini} \approx G^{-g}d^{ini}$, Eqs. 1 and 2 can be rewritten as

$$d^{obs} \approx Gm^{true} \quad \text{and} \quad m^{est} = G^{-g} d^{obs} \quad (3)$$

which can be combined to

$$m^{est} = G^{-g} G m^{true} = R m^{true} \quad (4)$$

where R is the so-called resolution matrix. Conceptually, the resolution matrix is the lens or filter through which the inversion sees the study region (Day-Lewis et al., 2005). High diagonal values in R indicate well identified model parameters and low values indicate unresolved parameters. High diagonal values in R are obtained when the elements of G are large (high sensitivity) and when the rows of G are independent (independent measurements).

A variety of experimental design methods have been proposed in the literature: simple forward modelling, direct interpretation of Jacobian sensitivities, methods based on data importance, among others. A thorough discussion of these simpler design methods and their disadvantages is provided by Maurer et al. (2000). In this paper, we only review the advanced design methods of Furman et al. (2004), Stummer et al. (2004) and Curtis et al. (1999).

Furman et al. (2004) presented an algorithm for the optimal selection of electrical resistivity tomography (ERT) measurements (here after referred to as FUR method). The FUR algorithm makes use of the sensitivity of measurements to a series of subsurface perturbations in a homogeneous background. The optimal set of measurement is defined as the set of measurements that maximizes the average sensitivity in the Jacobian matrix:

$$GF = \frac{1}{MN} \sum_{i=1}^{N} \sum_{j=1}^{M} \left| G_{ij} \right| \quad (5)$$

where M is the number of model parameters and N is the number of measurements. When computationally feasible, Furman et al. (2004) also suggested calculating the sensitivity of all possible measurements and normalizing these sensitivities according to

$$G_{ij}^{nor} = \left(\frac{G_j^{\max} - \left| G_{ij} \right|}{G_j^{\max}} \right) \quad (6)$$

where G^{max} is the maximum absolute sensitivity for each model parameter. The solution to the optimization problem is now simply the selection of the N measurements with the lowest mean normalized sensitivity. In the work of Furman et al. (2004), the sensitivity analysis was performed on selected target locations. Although this reduces the computational effort, we chose to evaluate the sensitivity at each cell to allow a better comparison with the other methods.

Stummer et al. (2004) also presented a method to optimize the subsurface information obtained from ERT measurements (here after referred to as STU method). The STU method assumes that there are M model parameters and that the Jacobian matrix of all possible N measurements is defined as G^{all}. From the complete set of N measurements, we select N^{base} measurements and subdivide G^{all} in G^{base} and G^{add}. The corresponding resolution matrices can

be calculated with Eq. 4. In the next step, the measurements in G^{add} are ranked according to the goodness of fit function GF:

$$GF(i) = \sum_{j=1}^{M} \frac{\left|G_{ij}^{add}\right|}{G_j^{sum}} \left(1 - \frac{R_{jj}^{base}}{R_{jj}^{all}}\right) \tag{7}$$

where G^{sum} is a normalization factor based on the average absolute sensitivity for each model parameter calculated from all possible measurements:

$$G_j^{sum} = \frac{1}{N} \sum_{i=1}^{N} \left|G_{ij}^{all}\right| \tag{8}$$

GF favours measurements that have high sensitivities for many model parameters and measurements that add information to parts of the model space not yet optimally resolved in the base measurements as compared to the entire set of possible measurements. After selecting the measurement with the highest GF as a potential candidate to be added to the set of base measurements, the linear independence of the candidate measurement is verified by the inner product LI. If it is assumed that row k has the highest GF, then LI is calculated with

$$LI(i) = \left| \frac{\sum_{j=1}^{M} G_{ij}^{base} G_{kj}^{add}}{\left\| G_i^{base} \right\| \left\| G_k^{add} \right\|} \right| \quad i = 1 \ldots N^{base} \tag{9}$$

If all values of LI are below a given threshold, the candidate measurement is added to base measurements and the next best candidate measurement is evaluated. After adding a number of measurements, the model resolution matrix needs to be recomputed to account for the improved resolution capabilities of the updated set of base measurements. Although the normalization used in Eq. 7 is different from the normalization in Eq. 6, using only the first part of the objective function in Eq. 7 should lead to the same optimal set of measurements for the STU and FUR methods.

The third optimal experimental design method is based on the work of Curtis (1999), here after referred to as CUR method. The starting point of the CUR method is the realization that instable inverse solutions occur when the eigenvectors of the matrix G^TG in Eq. 2 have some extremely small eigenvalues. The aim of the CUR method is therefore to maximize the eigenvalues from G^TG. In the survey design context, positive eigenvalues correspond to independent pieces of information obtainable from the survey, and zero eigenvalues correspond to information that is unobtainable. For a more detailed introduction of the CUR method, the reader is referred to Curtis (2004). A range of eigenvalue positivity measures have been conceived (Curtis, 1999). In this paper, we focus on the following measure:

$$\Theta = \sum_{i=1}^{M} \frac{\lambda_i}{\lambda_1} \tag{10}$$

where λ are the eigenvalues sorted in order of decreasing magnitude (λ_1 highest, λ_M lowest). This measure was selected because the number of computations required to obtain this measure is an order of magnitude lower than that of the other measures presented by Curtis (1999). Curtis (1999) optimized the eigenvalue measure with a genetic algorithm. To make the results of this method more comparable with the other two methods, we used a modified approach. Starting from a set of base measurements, the eigenvalue measure is calculated for the set of base measurements plus each candidate measurement. In the next step, the candidate

measurement with the highest eigenvalue measure is added to the set of base measurements and the procedure is repeated. To reduce the computation time, several measurements can be added to the set of base measurements before recalculating the eigenvalue measure for all remaining candidate measurements. It should be noted that sequential addition does not necessarily lead to the most optimal set of measurements, but this simplification was required to make the design method computationally feasible.

3. MATERIALS AND METHODS

3.1 Travel time tomography with ground penetrating radar (TT-GPR)

In TT-GPR, travel times of EM waves are measured between transmitter and receiver antennas located at many different positions within two boreholes. The first arrival times of all measurements can be used to reconstruct a (tomographic) two-dimensional image of radar velocity by discretizing the area between the boreholes and inverting for the velocity of each cell by minimizing the difference between measured and modelled arrival times. In our TT-GPR case study, we assume two vertical boreholes with a depth of 20 m and a separation of 10 m. The area between the boreholes is discretized in cells of 0.25 m, which results in a total of 3200 model parameters in the inversion. We used the commonly used straight-ray approximation as the forward model. This approximation reduces the inversion to a linear problem and makes the sensitivity and resolution matrix independent of subsurface heterogeneities. This approximation does not account for EM wave refraction and EM wave frequency effects on the ray traces. We used damping to regularize the inverse problem. To investigate experimental design issues of TT-GPR, we assumed that there were 40 equally spaced transmitter and receiver locations in each borehole starting at 0.125 m and ending at 19.625 m with steps of 0.50 m. This results in a total of 1600 potential measurements (the 'complete' dataset). The experimental design methods were then used to test whether there exists a subset of the complete dataset that results in an approximately equally good inversion. This was tested with the average relative model resolution (RMR):

$$RMR = \frac{1}{M}\sum_{j=1}^{M}\left(\frac{R_{jj}^{base}}{R_{jj}^{all}}\right) \tag{11}$$

which equals 1 if all model parameters are as well resolved in the subset as in the complete dataset. The model resolution matrix was updated after adding ten measurements in the STU method, and the eigenvalue measures were recalculated after adding 50 measurements to the set of base measurements in the CUR method. The initial set of base measurements for the STU and CUR method consisted of the 10 and 50 most sensitive measurements identified in the FUR method, respectively. In the STU method, the initial value of the LI threshold defined in Eq. 9 was set to 0.33 and this value linearly increased to 1.00 as the size of the set of base measurements approached the size of the complete dataset.

3.2 Electrical resistivity tomography (ERT)

Cross-borehole ERT is used to estimate the electrical conductivity distribution in the subsurface by establishing an electrical gradient between two source electrodes and measuring the resultant potential distribution at two other receiving electrodes and repeating this procedure for many different electrode combinations (Binley and Kemna, 2005). For p electrodes, there exists a total of $N=p\times(p-1)\times(p-2)\times(p-3)/8$ nonreciprocal four-point electrode

configurations. Despite a large flexibility on the acquisition side, most 2D imaging applications have used standard electrode configurations (either Wenner or dipole-dipole arrays). However, Stummer et al. (2004) showed that non-standard electrode configurations can lead to much better imaging results with the same amount of arrays. Similar conclusions were drawn by Furman et al. (2004). Our ERT case study is based on an actual field set-up used to monitor water fluxes in a forest stand. In this 10 by 10 meter field setup, 36 boreholes are installed on a 2 by 2 meter grid. Each borehole is 2.84 meter deep and contains 16 electrodes with an increasing spacing with depth. This set-up has more than 10 billion potential arrays. To make this set-up suitable for experimental design methods, we only consider two central boreholes in a single 2D plane (see Fig. 3). The actual measurement set-up is based on a dipole-dipole scheme and consists of 106 arrays, whereas the 'complete dataset' contains a total of 107880 possible arrays. The 2D plane extends from x=-1.0 m to x=12.0 m and from y=0.0 m to y=-4.0 m and is discretized in 56 cells in x-direction and 35 cells in the y-direction. CRMOD (Kemna, 2000) was used to calculate the sensitivity matrix based on all candidate arrays for a homogeneous resistivity model. Regularization was based on a first-derivative filter (roughness) and a field-average value for the regularization parameter (average value of many regularization parameters obtained at the end of actual inversions). As with the TT-GPR case study, the quality of the optimal experimental design methods was evaluated with the relative model resolution given in Eq. 11. The settings of the STU method were the same as for the TT-GPR case study, except for the LI threshold, which was fixed at 0.90. For the CUR method, we started with the 10 most sensitive points from the FUR method, added 10 measurements in the first iteration and 20 measurements in the following iterations.

4. RESULTS AND DISCUSSION

Fig. 1a and 2a present the resolution plots for TT-GPR and ERT for the complete dataset, respectively. As was already shown by Day-Lewis (2005), TT-GPR measurements provide most information in the centre between the boreholes, whereas the cells near the surface and the bottom of the boreholes and along the boreholes are much less resolved. For ERT measurements, the resolution is high near the boreholes and reduces strongly away from the boreholes. This complementary distribution of information content for TT-GPR and ERT certainly is one of the motivations for current joint inversion efforts.

Figs. 1b to 1d present the RMR for subsets of 400 'optimal' TT-GPR measurements obtained with the three experimental design methods. The FUR method emphasizes measurements with a high normalized sensitivity. Due to the nature of TT-GPR measurements, a single measurement can have the highest sensitivity in a large number of cells. In the simplified linearization used in this study, this sensitivity only depends on the length of the raypath in each cell. Therefore, the FUR method preferentially selects high-angle measurements and the resulting RMR is heterogeneously distributed in the 2D plane. In terms of inversion results, this heterogeneous RMR distribution implies that velocity variations cannot be detected in considerable parts of the model space (zero model resolution).

The distribution of the relative RMR for the STU and the CUR method are similar in the TT-GPR case study. The RMR is high in areas with a low absolute resolution, and relatively evenly distributed in the other areas of the 2D plane. In the densely sampled region of y=-5 to -15 meter, the RMR is 0.1817 for the STU method and 0.1692 for the CUR method.

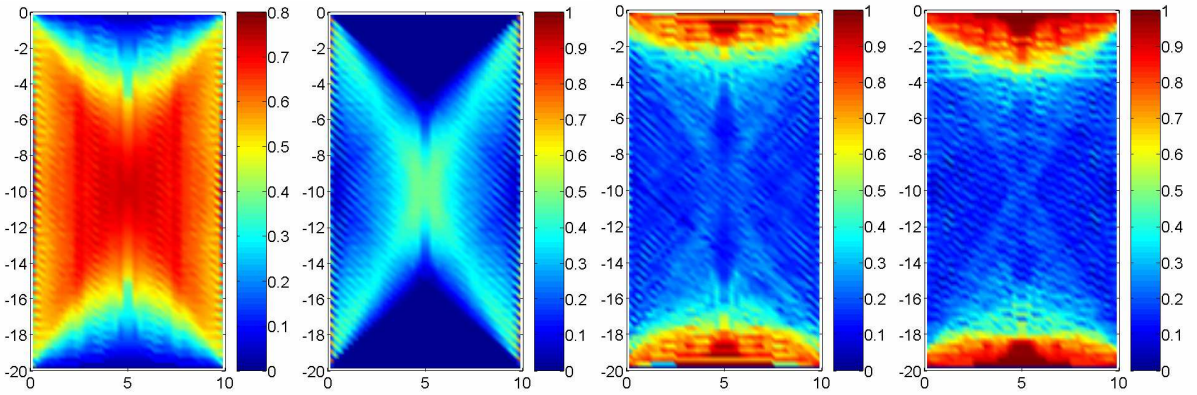

FIGURE 1. (from left to right). Resolution plot of the complete TT-GPR dataset and three relative resolution plots for the FUR, STU and CUR methods based on 400 measurements.

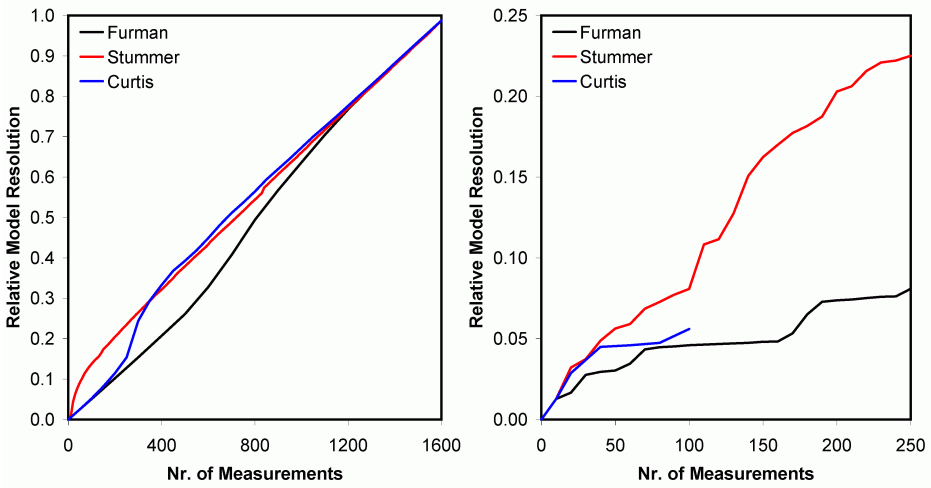

FIGURE 2. Relative model resolution as a function of the number of measurements in the survey for TT-GPR (left) and ERT (right)

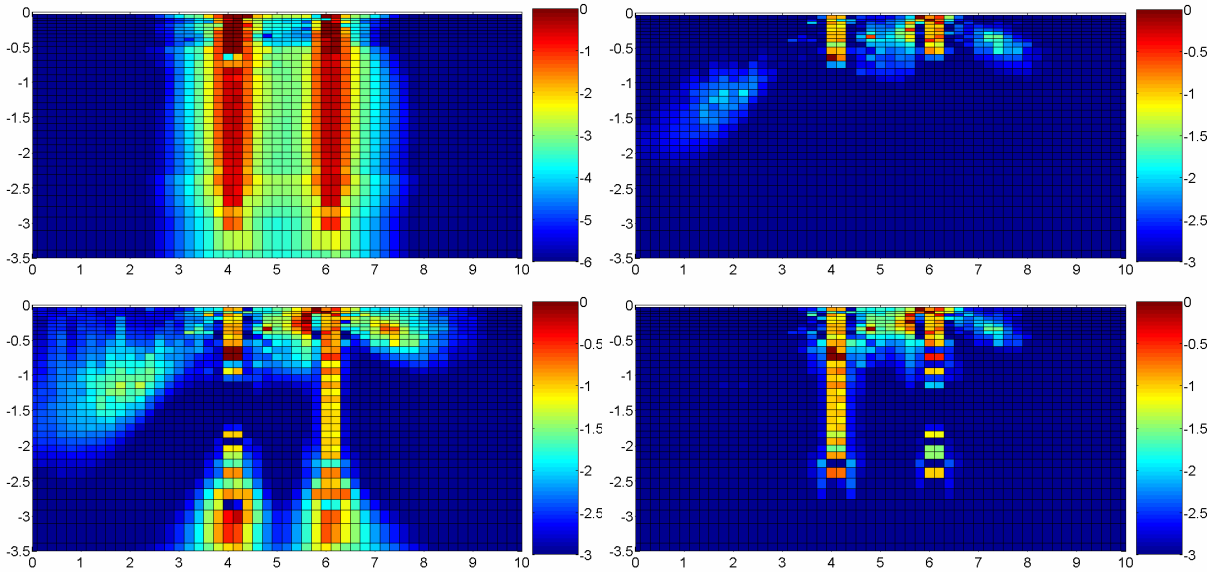

FIGURE 3. (from upper left to bottom right). Log resolution plot of the complete ERT dataset and three log(RMR) plots for the FUR, STU and CUR methods based on 106 measurements.

The areas of high RMR can be explained by the low measurement density: the inclusion of only a few measurements can increase the RMR from 0 to 1. The homogeneous distribution of the RMR in the remaining parts of the model space was expected for the STU method. The weighting of the sensitivity with the relative model resolution in Eq. 7 leads to the inclusion of measurements that add information to unresolved parts of the model space. The difference between the FUR and the STU method indicates that the objective function of the STU method is dominated by the model resolution part and not by the sensitivity part. A repeat run of the STU method without consideration of the LI criterion resulted in a very similar result for the RMR, which indicates that this criterion did not play an essential role in the TT-GPR case study.

The homogeneous distribution of the RMR was not expected for the CUR method. The objective function for this method can be increased by increasing the sum of the eigenvalues, while not increasing the highest eigenvalue. Since eigenvalues and model resolution are both measures of how well model parameters are resolved, it is tempting to interpret this objective function as 'increasing the model resolution without increasing the highest model resolution' (i.e. preferential selection of measurements that provide information in unresolved parts of the model space). This seems exactly what happened in the CUR method. This is an interesting result because in contrast to the STU method, the CUR method does not require the calculation of the model resolution matrix for the 'complete data set'.

Fig. 2a shows the average RMR as a function of the number of measurements included in the survey. Although the spatial distribution of the RMR was quite different for three design methods, they result in a similar development of the average RMR with an increasing number of measurements. The increase in RMR is approximately linear and indicates that the use of 25% of the measurements will lead to a RMR of ~0.25 and the use of 50% of the measurements to a RMR of ~0.50. In other words, every additional measurement adds approximately the same amount of information. Finally, it should be noted that the optimized designs are not necessarily practical in the field. Since TT-GPR measurements are made manually, it might be required to accept a lower RMR value in exchange for a feasible acquisition scheme in practice.

Figs. 3b – 3d show the spatial distribution of the RMR for the 'optimal' sets of 106 ERT measurements for all three design methods. This small number of measurements was dictated by acquisition constraints related to the actual field set-up. As in the TT-GPR case study, the FUR method results in a set of measurements that only samples a limited part of the model space. Fig. 3b shows that the measurements selected by the FUR method are all located near the surface. The selected arrays all have one current and potential electrode in each borehole and are very similar. Electrodes deeper in the borehole are not selected because of the increased electrode separation with depth.

The results of the CUR method are presented in Fig. 3d. The RMR is distributed better throughout the model space, but the computational effort to require this plot was excessive compared to the other two methods. In the CUR method, 10 measurements were accepted in the first iteration and 20 measurements in each subsequent iteration without updating the objective function. The effect of this infrequent updating can clearly be recognized because similar types of candidate measurements are accepted within each iteration. However, more frequent updating increases the computational burden beyond an acceptable limit. The results of the STU method are presented in Fig. 3c. Visually, this method performed best in the ERT case study and this is confirmed by the average RMR as a function of the number of

measurements shown in Fig. 2b. As was already noticed in the TT-GPR case study, the STU method tends to favour adding measurements that provide information in areas with a low absolute resolution. Therefore, it is expected that an analysis focussed on the area between the boreholes will further improve the relevancy of the STU design method. Finally, Fig. 2b also shows that experimental design is worthwhile in the case of ERT. In the case of the STU method, only 250 measurements (0.23%) are required to explain 22% of the model resolution of the complete dataset. Approximately 1000 measurements (~0.92%) are required with the FUR method (not shown). It should be noted that the RMR of the expert-based set of 106 measurements was only 0.0139, albeit well distributed in the model space.

5. CONCLUSIONS AND OUTLOOK

We have compared three experimental design methods in two case studies. Based on the evaluation of the average model resolution and the spatial distribution of the model resolution, it was concluded that the method of Stummer et al. (2004), which combines a normalized sensitivity and a relative model resolution measure, performed best. The method of Furman et al. (2004), which is only based on a normalized sensitivity measure, left large parts of the model space unsampled. The method of Curtis (1999) based on eigenvalue spectra performed well in the TT-GPR case study, but was too computationally intensive in the ERT case study.

There are two important issues that were not addressed in this study. The first issue is the dependence of the optimal measurement design on the subsurface distribution of the variable of interest. Without prior information on the site conditions, one strategy would be to first acquire a limited initial dataset and then use a series of "data acquisition – data inversion – update experimental design" cycles to improve the quality of the measurements. The second issue is the dependence of the data signal-to-noise ratio on the measurement type (i.e. higher relative error for high angle GPR rays or ERT measurements with a high geometry factor). This issue can be investigated by including realistic error models in the calculation of the resolution matrix.

REFERENCES

Binley, A. and A. Kemna (2005), DC resistivity and induced polarization methods, in *Hydrogeophysics, Water Sci. Technol. Libr.* Vol. 50, edited by Y. Rubin and S.S. Hubbard, Springer New York, 129-156.

Curtis, A. (1999), Optimal experiment design: cross-borehole tomographic examples, *Geophysical Journal International, 136*, 637-650.

Curtis, A. (2004), Theory of model-based geophysical survey and experimental design. Part A – linear problems. *The leading edge, 23(10)*, 997-1004.

Day-Lewis, F.D., K. Singha and A.M. Binley (2005), Applying petrophysical models to radar travel time and electrical resistivity tomograms: resolution-dependent limitations, *Journal of Geophysical Research, 110*, B08206, doi:10.1029/2004JB003569.

Furman, A., T.P.A Ferre and A.W. Warrick (2004), Optimization of ERT surveys for monitoring transient hydrological events using perturbation sensitivity and genetic algorithms, *Vadose Zone Journal, 3*, 1230-1239.

Kemna, A. (2000), *Tomographic inversion of complex resistivity – Theory and application*. PhD-thesis Bochum University, Germany.

Maurer, H., D.E. Boerner and A. Curtis (2000), Design strategies for electromagnetic geophysical surveys, *Inverse Problems, 16*, 1097-1117.

Stummer, P., H. Maurer and A.G. Green (2004), Experimental design: electrical resistivity data sets that provide optimum subsurface information, *Geophysics, 69(1)*, 120-139.