

# 1 **Optimized arrays for 2-D resistivity survey lines with a large number of electrodes**

2 Authors : M. H. Loke<sup>1\*</sup>, P. B. Wilkinson<sup>2</sup>, J. E. Chambers<sup>2</sup>, S. S. Uhlemann<sup>2</sup> and J. P. R. Sorensen<sup>3</sup>

3 <sup>1</sup> Geotomo Software, 115 Cangkat Minden Jalan 5, Minden Heights, Gelugor 11700, Penang, Malaysia.

4 email : drmhloke@yahoo.com, Tel : +60 4 6574525, Fax : +60 4 6588437. Corresponding author.

5 <sup>2</sup> British Geological Survey, Environmental Science Centre, Keyworth, Nottingham, United Kingdom  
6 NG12 5GG.

7 email : pbw@bgs.ac.uk, jecha@bgs.ac.uk, suhl@bgs.ac.uk

8 <sup>3</sup> British Geological Survey, Wallingford, Oxfordshire, United Kingdom OX10 8BB.

9 email : jarel@bgs.ac.uk

10

## 11 **Abstract**

12 Previous studies show that optimized arrays generated using the 'Compare R' method have significantly  
13 better resolution than conventional arrays. This method determines the optimum set of arrays by  
14 selecting those that gives the maximum model resolution. The number of possible arrays (the  
15 comprehensive data set) increases with the fourth power of the number of electrodes. The optimization  
16 method faces practical limitations for 2-D survey lines with more than 60 electrodes where the number  
17 of possible arrays exceeds a million. Several techniques are proposed to reduce the calculation time for  
18 such survey lines. A single-precision version of the 'Compare R' algorithm using a new ranking function  
19 reduces the calculation time by two to eight times while providing results similar to the double-precision  
20 version. Recent improvements in computer GPU technology can reduce the calculation time by about  
21 seven times. The calculation time is reduced by half by using the fact that arrays that are symmetrical  
22 about the center of the line produce identical changes in the model resolution values. It is further  
23 reduced by more than thirty times by calculating the Sherman-Morrison update for all the possible two-  
24 electrode combinations, which are then used to calculate the model resolution values for the four-

1 electrode arrays. The calculation time is reduced by more than ten times by using a subset of the  
2 comprehensive data set consisting of only symmetrical arrays. Tests with a synthetic model and field  
3 data set show that optimized arrays derived from this subset produce inversion models with differences  
4 of less than 10% from those derived using the full comprehensive data set. The optimized data sets  
5 produced models that are more accurate than the Wenner-Schlumberger array data sets in all the tests.

6

7 **Keywords** : 2-D; resistivity; fast; array; optimization; resolution

8

## 1 INTRODUCTION

2 In the last twenty-five years there have been major developments in the resistivity method where  
3 it can now provide realistic images in geologically complex areas. Developments in multi-electrode  
4 resistivity meter systems and rapid inversion software have led to the widespread use of two-  
5 dimensional (2-D) and even three-dimensional (3-D) and resistivity surveys in engineering,  
6 environmental and mineral exploration surveys (Auken et al., 2006; Loke et al., 2013). 2-D surveys  
7 have proven to be the most economically efficient method in moderately complex areas, while 3-D  
8 surveys are usually reserved for more complex areas. Most of the early multi-electrode instruments have  
9 up to about 50 independent electrodes, but newer commercial instruments with more than 100 nodes are  
10 becoming more common.

11 There has been significant progress in automatic techniques to find the optimum set of electrode  
12 array configurations (Stummer et al., 2004; Maurer et al., 2010) in recent years. The 'Compare R'  
13 method by Wilkinson et al. (2006) proved to provide arrays with the highest resolution among the  
14 techniques examined by Loke et al. (2010a). The optimized arrays can significantly improve the  
15 resolution obtained with 2-D and 3-D surveys (Wilkinson et al., 2012; Loke et al., 2014c). Previous  
16 research have concentrated on improving the depth resolution for survey lines with a small number (30  
17 to 50) of electrode positions. The 'Compare R' method faces practical computational limitations for  
18 longer survey lines (Blome et al., 2011). Thus the main focus of this paper is on techniques to reduce the  
19 calculation time required to generate the optimized arrays for lines with a large number of electrodes.  
20 Other practical aspects on the use of optimized arrays for field surveys (such as the reduction of  
21 electrode polarization noise and use of data error estimates) can be found in Wilkinson et al. (2012).

22 The following section gives a brief description of the 'Compare R' array optimization method,  
23 followed by computational and numerical techniques devised to reduce the calculation time required to  
24 generate the arrays. Results from a synthetic model and a field survey are then presented.

1

## 2 THEORY

3 The smoothness-constrained least-squares optimization method is frequently used for 2-D  
4 inversion of resistivity data (Ellis and Oldenburg, 1994; Loke et al., 2003). The subsurface model  
5 commonly consists of a large number of rectangular cells where the size and positions of the cells are  
6 fixed but the resistivity is allowed to vary. The least-squares equation that gives the relationship between  
7 the model parameters and the measured data is given by

$$8 \quad (\mathbf{G}^T \mathbf{G} + \lambda \mathbf{C}) \Delta \mathbf{r}_i = \mathbf{G}^T \mathbf{d} - \lambda \mathbf{C} \mathbf{r}_{i-1}. \quad (1)$$

9 The Jacobian matrix  $\mathbf{G}$  contains the sensitivities of the (logarithm of) calculated apparent resistivity  
10 values with respect to the (logarithm of) the model resistivity values  $\mathbf{r}$ .  $\mathbf{C}$  contains the roughness filter  
11 constraint,  $\lambda$  is the damping factor and  $\mathbf{d}$  is the data misfit vector.  $\Delta \mathbf{r}_i$  is the change in the model  
12 parameters. Various modifications have been made to the above equation to incorporate desired  
13 characteristics in the data misfit or model roughness functions, such as a L1-norm criterion for the data  
14 misfit and model roughness filter (Farquharson and Oldenburg, 1998; Loke et al., 2003) and to include  
15 using a data error weighting matrix (Ellis and Oldenburg, 1994). The model resolution matrix  $\mathbf{R}$   
16 (Menke, 1989; Loke et al., 2010a) is given by

$$17 \quad \mathbf{R} = \mathbf{B} \mathbf{A}, \text{ where } \mathbf{A} = \mathbf{G}^T \mathbf{G} \text{ and } \mathbf{B} = (\mathbf{G}^T \mathbf{G} + \lambda \mathbf{C})^{-1}. \quad (2)$$

18 The main diagonal elements of  $\mathbf{R}$  give an estimate of the model cells resolution. The 'Compare R'  
19 method (Wilkinson et al., 2006) attempts to determine the set of array configurations that will maximize  
20 the average resolution value. There are  $E(E-1)(E-2)(E-3)/8$  independent four-electrode configurations  
21 for a system with  $E$  electrodes. Arrays where the current and potential electrodes are interleaved of the  
22 'gamma' (Carpenter and Habberjam, 1956) type as well as those large geometric factors are excluded to  
23 reduce the number of possible configurations (Stummer et al., 2004). The remaining configurations form  
24 the 'comprehensive' data set. A small base data set consisting of the dipole-dipole configurations with

1 the 'a' dipole length of 1 unit electrode spacing (and maximum 'n' value of 6 to 10 depending on the  
 2 maximum geometric factor set) is used as the starting base data set. The change in the model resolution  
 3 for each new array when added to the base set is then calculated. A selected number of the  
 4 configurations that give the largest increase in the model resolution, and have a suitable degree of  
 5 orthogonality to the existing configurations, are then added to the base data set (Wilkinson et al., 2012).  
 6 The model resolution values for the new base data set are then recalculated using equation (2). This is  
 7 repeated until the desired number of optimized arrays is selected.

8 The Sherman-Morrison Rank-1 update is used to calculate the change in the model resolution  
 9 matrix  $\Delta \mathbf{R}_b$  when a new configuration is added to the base set using the following equation (Loke et al.,  
 10 2010b).

$$11 \quad \Delta \mathbf{R}_b = \frac{\mathbf{z}}{1 + \mu} (\mathbf{g}^T - \mathbf{y}^T) \quad (3)$$

$$12 \quad \mathbf{z} = \mathbf{B}_b \mathbf{g}, \mathbf{y} = \mathbf{A}_b \mathbf{z} \text{ and } \mu = \mathbf{g} \cdot \mathbf{z} .$$

13  $\mathbf{A}_b$  and  $\mathbf{B}_b$  are the matrices in equation (2) for the base data set. The vector  $\mathbf{g}$  contains the model  
 14 sensitivity values for the new array. The following function  $F_{CR}$  (Wilkinson et al., 2012) is used to the  
 15 rank the improvement in the resolution for a model with  $m$  cells due to an add-on array.

$$16 \quad F_{CR} = \frac{1}{m} \sum_{j=1}^{j=m} \frac{\Delta R_b(j, j)}{R_c(j, j)}. \quad (4)$$

17 The average relative model resolution given by

$$18 \quad S_r = \frac{1}{m} \sum_{j=1}^{j=m} \frac{R_b(j, j)}{R_c(j, j)}, \quad (5)$$

19 is commonly used to assess the performance of the optimized arrays. In the following sections, a value  
 20 of 0.001 (Wilkinson et al., 2012) is used for the damping factor  $\lambda$  in equation (2) for the model  
 21 resolution calculations. We also use the simple form of the model resolution equation where the identity  
 22 matrix  $\mathbf{I}$  is used in place of  $\mathbf{C}$  (Wilkinson et al., 2006). However, the proposed techniques can be also

1 used for different variations of this equation, such as using a roughness filter matrix (Loke et al., 2010b)  
2 or incorporating a data weighting matrix (Wilkinson et al., 2012).

3

#### 4 **METHODS TO REDUCE THE CALCULATION TIME**

5 A number of numerical and computational techniques that were developed to reduce the  
6 calculation time are described in the section.

7

##### 8 **Using the symmetry of 2-D survey lines and models**

9 A modification was introduced by Loke et al. (2010a) whereby whenever an array was added to  
10 the optimized data set, a check was made to ensure the symmetrical counterpart of the array was also  
11 included. This ensures that over a symmetrical structure the optimized data set will also display a  
12 symmetrical anomaly to avoid any bias in the data. Figure 1 shows an example of a symmetrical pair of  
13 arrays for a survey line with 17 electrodes. Assuming the initial base data set and model discretization of  
14 the subsurface into rectangular cells are symmetrical about the center of the line, both arrays will give  
15 the same value for the model resolution change function  $F_{CR}$  when added to the base set. While the array  
16 in Figure 1a will cause a larger change in the model resolution ( $\Delta R_b$ ) values for the model cells under  
17 the left half of the line, and the array in Figure 1b will have a larger effect on the right half, the sum of  
18 the relative changes for all the cells will be the same. Thus it is only necessary to calculate  $F_{CR}$  for one  
19 of the arrays. The time taken by the 'Compare R' method is proportional to the number of arrays in the  
20 comprehensive data set. As an example, if a maximum geometric factor of 4146.9 m is used  
21 (corresponding to a dipole-dipole array with  $a=1$  m and  $n=10$ ), the comprehensive data set for a survey  
22 line with 80 electrodes has 2973047 arrays but slightly less than half (1485564 or 49.97%) are the  
23 symmetrical counterparts. The calculation time is reduced by about half by using the symmetry in the  
24 model resolution values. A small number of arrays that are symmetrical about the centre of the line of

1 electrodes (Figure 1c) do not have a symmetrical counterpart in the comprehensive data set. We note  
2 that 3-D surveys using a rectangular grid of electrodes have a four-way symmetry (Loke et al., 2014c),  
3 so a similar technique can be used to reduce the calculation time for 3-D optimized arrays by nearly  
4 three-quarters.

5 For models with non-symmetrical topography or subsurface resistivity variations the assumption  
6 of symmetrical variations in the model resolution values is not valid. However, from previous field  
7 surveys, it has been found that optimized arrays calculated based on a simple homogeneous model with  
8 a flat surface still performed much better than conventional arrays in areas with significant topography  
9 and large resistivity variations (Stummer et al., 2004; Wilkinson et al., 2012).

10

## 11 **Improvements in GPU technology**

12 Loke et al. (2010b) showed that the Graphics Processing Unit (GPU) in the computer graphics  
13 card can be used as a highly parallel mathematical coprocessor to reduce the calculation time. Figure 2  
14 shows the time taken by the single and double-precision algorithms to generate a data set with 10000  
15 points for a 2-D survey line with 80 electrodes using different GPUs. It includes the GTX285 GPU used  
16 by Loke et al. (2010b) to illustrate the effect of improvements in GPU technology over the years. The  
17 base computer system has an Intel Sandy Bridge-E 3930 hex-core CPU. The calculation times using the  
18 CPU alone are also shown for comparison. The number of arrays in the comprehensive data set is  
19 2973047 while the number of model parameters is 2548. The initial base data set consists of the dipole-  
20 dipole array configurations with  $a = 1$  m and  $n = 1$  to 8. The base data set was increased by 5% after  
21 each iteration of the 'Compare R' method. The time taken by the GPU single-precision version is about  
22 one-half to one-eighth of that taken by the double-precision version. The number of GPU single  
23 precision computational units is 3 to 24 times the number of double precision units (Farber, 2011). Note  
24 that the slowest GPU (GTX285) is still about twice as fast as the CPU in the single-precision test. The

1 GTX285 and GTX560 graphics cards have a PCI-E 2.0 data bus, while the newer GTX670 and GTX780  
 2 have a PCI-E 3.0 data bus that has twice the data transfer rate. The overall effects of improvements in  
 3 GPU technology have reduced the calculation time for the single-precision version by about seven  
 4 times. In the following tests, only the calculation times for the single-precision algorithm using the  
 5 GTX780 GPU are listed. Since the CPU and GPUs follow the IEEE binary floating-point format (Press  
 6 et al., 2007), the numerical results (other than the calculation time) are independent of the type of CPU  
 7 or GPU used.

8 It was noted by Loke *et al.* (2010b) that on tests for lines ranging from 30 to 60 electrodes there  
 9 was a small difference (ranging from 0.31% to 0.94%) in the average relative model resolution values  
 10 (equation (5)) of the results obtained using the single and double precision algorithms. Thus tests were  
 11 conducted to see whether a similar difference occurs for longer survey lines. We found that the  
 12 difference is significantly lower with the current version of the computer program used. As an example,  
 13 the optimized data set with 10000 data points for a line with 80 electrodes (Figure 1) have average  
 14 relative resolution values of 0.6543747 and 0.6534324 for the single and double precision methods  
 15 respectively, i.e. a difference of only about 0.14%. The difference from earlier results (Loke et al.,  
 16 2010b) is probably due to different ranking functions used. The ranking function used in earlier research  
 17 (Wilkinson et al., 2006; Loke et al., 2010b) is given by

$$18 \quad F_{CR} = \frac{1}{m} \sum_{j=1}^{j=m} \frac{R_{b+1}(j, j)}{R_b(j, j)} = \frac{1}{m} \sum_{j=1}^{j=m} 1 + \frac{\Delta R_b(j, j)}{R_b(j, j)}. \quad (6)$$

19 It uses the ratio of the improvement in the model resolution due to the new array compared to the  
 20 resolution of the current base set, whereas the new ranking function proposed by Wilkinson *et al.* (2012)  
 21 uses the resolution of the comprehensive data set (equation (4)) in the denominator. The resolution of  
 22 the base set  $R_b$  changes with each iteration. This probably makes the final result more sensitive to round-  
 23 off errors (and thus the precision of the calculations used). In comparison, the resolution of the



1 comprehensive data set  $R_c$  is constant throughout the optimization process. We repeated the calculations  
 2 using the old ranking function given in Equation (6) for a line with 80 electrodes and obtained average  
 3 relative resolution values of 0.6538870 and 0.6520786 for the single and double precision algorithms.  
 4 This gives a difference of 0.28% which is twice that obtained with the new ranking function. We also  
 5 repeated the tests for lines with different numbers of electrodes. The new ranking function gave  
 6 differences of less than 0.2% between the single and double precision algorithms which are not  
 7 significant in terms of the resolution obtained with the optimized arrays.

8

### 9 **Decomposing the Sherman-Morrison update and cache optimization**

10 The bulk of the calculation time is taken by the Sherman-Morrison update for the  $\mathbf{z}$  and  $\mathbf{y}$  vectors  
 11 in equation (3). We now look into more detail the calculation of these vectors.

$$12 \quad \mathbf{z} = \mathbf{B}_b \mathbf{g} \quad \text{and} \quad \mathbf{y} = \mathbf{A}_b \mathbf{z}$$

13 The  $\mathbf{z}$  vector is formed by a multiplication of an  $m$  by  $m$  matrix  $\mathbf{B}_b$  with the sensitivity vector  $\mathbf{g}$  for a test  
 14 array in the comprehensive data set. The  $\mathbf{y}$  vector is formed by a similar matrix-vector multiplication.  
 15 For a given base data set, the same  $\mathbf{A}_b$  and  $\mathbf{B}_b$  matrices are used for all the test arrays. The potential and  
 16 sensitivity values for any four-electrode array can be calculated from similar values from the four  
 17 relevant current-potential electrode pairs. The sensitivity vector  $\mathbf{g}_{ABMN}$  for a four-electrode array can be  
 18 decomposed into the following vectors.

$$19 \quad \mathbf{g}_{ABMN} = \mathbf{g}_{AM} - \mathbf{g}_{AN} - \mathbf{g}_{BM} + \mathbf{g}_{BN} \quad (7)$$

20  $\mathbf{g}_{AM}$  is the sensitivity vector associated with the two-electrode pair of the  $A$  current electrode and  
 21 the  $M$  potential electrode (Figure 1). Thus the  $\mathbf{z}_{ABMN}$  and  $\mathbf{y}_{ABMN}$  vectors for any four-electrode array can  
 22 be calculated using the following relationships.

$$23 \quad \mathbf{z}_{ABMN} = \mathbf{B}_b \mathbf{g}_{AM} - \mathbf{B}_b \mathbf{g}_{AN} - \mathbf{B}_b \mathbf{g}_{BM} + \mathbf{B}_b \mathbf{g}_{BN} = \mathbf{z}_{AM} - \mathbf{z}_{AN} - \mathbf{z}_{BM} + \mathbf{z}_{BN} \quad (8)$$

$$\mathbf{y}_{ABMN} = \mathbf{A}_b \mathbf{z}_{AM} - \mathbf{A}_b \mathbf{z}_{AN} - \mathbf{A}_b \mathbf{z}_{BM} + \mathbf{A}_b \mathbf{z}_{BN} = \mathbf{y}_{AM} - \mathbf{y}_{AN} - \mathbf{y}_{BM} + \mathbf{y}_{BN} \quad (9)$$

A survey line with  $E$  electrodes has  $E(E-1)/2$  possible two-electrode combinations, i.e. it is proportional  $E^2$ . In comparison, the size of the comprehensive data set ( $N_c$ ) is proportional  $E^4$ . For example, a survey line with 80 electrodes has 3160 possible two-electrode combinations while the comprehensive data set has 2973047 arrays. The calculation time can be greatly reduced by first calculating the  $\mathbf{z}$  and  $\mathbf{y}$  vectors for all the possible unique two-electrode combinations. The results are temporarily stored in the computer memory using two vector arrays ( $\mathbf{z}_{CP}$  and  $\mathbf{y}_{CP}$  where  $\mathbf{C}$  and  $\mathbf{P}$  denotes a general current and electrode position) each with  $E(E-1)/2$  elements. To calculate the  $\mathbf{z}_{ABMN}$  vector for a four-electrode configuration, the pre-calculated four current-potential pairs ( $\mathbf{z}_{AM}$ ,  $\mathbf{z}_{AN}$ ,  $\mathbf{z}_{BM}$ ,  $\mathbf{z}_{BN}$ ) are then added or subtracted. Note that only the vector addition or subtraction operations are needed, as the matrix-vector multiplication operation in equation (3) is avoided by using the pre-calculated two-electrode vector values. A similar operation is used to calculate the  $\mathbf{y}_{ABMN}$  vector. The number of operations required to calculate the vectors is proportional to  $m$ , compared to the Sherman-Morrison update that is proportional to  $m^2$ . As equations (6), (7) and (8) involve the addition or subtraction of only four vectors, computational round-off errors is not a significant problem. The decomposition and direct update methods give the same results. The direct method requires  $N_c m^2$  (or about  $1.9 \times 10^{13}$  for a survey line with 80 electrodes and 2548 model cells) multiply-add operations, while the decomposition method requires  $E \cdot (E-1) m^2 / 2$  (about  $2.1 \times 10^{10}$  for the same line) multiply-add plus  $4N_c m$  (about  $6.9 \times 10^9$ ) add-subtract operations. Note the number of mathematical operations required by the direct method is nearly 1000 times more than the decomposition method.

Figure 3 shows the time taken to calculate an optimized data set with 10000 arrays for survey lines with 60 to 200 electrodes using the two different methods of applying the Sherman-Morrison update. The decomposition method is nearly thirty times faster for the longest survey line with 140

1 electrodes where it is still practical to use the direct method. It extends the practical limit of the  
 2 'Compare R' method to survey lines with about 200 electrodes using current PC technology (Figure 3).  
 3 The reduction in the computer calculation time is less than the estimated reduction in the number of  
 4 numerical operations in calculating the Sherman-Morrison update. This is due to the remaining  
 5 operations that are common to both methods, such as the calculation of the model resolution for the new  
 6 base data set after each iteration.

7 In the decomposition method, the time taken by the Sherman-Morrison update for the two-  
 8 electrode pairs is about 1 to 2% of the total computer time while 80 to 90% of the computer time is used  
 9 in reconstituting the  $\mathbf{g}$ ,  $\mathbf{z}$  and  $\mathbf{y}$  vectors for the four-electrode arrays in the comprehensive data set. The  
 10 order at which electrode configurations are processed have a significant effect on the computer time.  
 11 The computer CPU has a fast internal memory cache that stores recently used data fetched from the  
 12 slower main computer memory (the Intel 3930 CPU has a 12 MB cache). It can access data stored in the  
 13 cache several times faster compared to the main memory. The array configurations in the comprehensive  
 14 data set are sorted using the following function

$$15 \quad S = A.(E + 1)^3 + B.(E + 1)^2 + M.(E + 1) + N, \quad (10)$$

16 where the electrodes ( $A, B, M, N$ ) have index numbers ranging from  $I$  to  $E$ . Table 1 shows a partial listing  
 17 of the array configurations from a survey line with 80 electrodes after sorting. Arrays number 36 to 41  
 18 share the same positions for the  $A$ ,  $B$  and  $M$  electrodes, and thus the same values for the  $\mathbf{g}_{AM}$  and  $\mathbf{g}_{BM}$   
 19 (and also for similar  $\mathbf{z}$  and  $\mathbf{y}$ ) vectors. The values used for array 36 reside in the CPU cache that can then  
 20 be used for arrays 37 to 41. Note the  $A$ ,  $B$  and  $N$  electrodes for array 38 are the same as those in array 43,  
 21 so they also share the same values for the vectors involving the  $AN$  and  $BN$  electrodes pairs. The  
 22 calculation time is reduced by 50 to 60% by using a sorted comprehensive data set compared to a  
 23 random sequence.

1           The Intel C++ optimizing compiler (Gerber et al., 2006) was used in this research. The compiler  
2 attempts to generate optimal code for the Intel CPUs by taking advantage of specific CPU characteristics  
3 such as the registers and cache. Nevertheless, sorting the comprehensive data set gave a significant  
4 reduction in the calculation time. This is probably because the CPU cache has a limited size. As an  
5 example, to store all the two-electrode  $\mathbf{z}_{CP}$  and  $\mathbf{y}_{CP}$  vectors for the 80 electrodes survey line example  
6 requires about 61 MB of memory space which is much larger than the CPU cache of 12 MB. The sorting  
7 algorithm ensures that the vector values are reused shortly after they have been loaded into the cache.  
8 This avoids the vector values in the cache from being replaced by other data values before they are  
9 reused. The sorting algorithm also has the advantage that the calculation time is less dependent on the  
10 efficiency of the compiler used which might be different for other compilers or computer systems.

11

## 12 **Modifications to the comprehensive data set**

13           The size of comprehensive data set increases rapidly with the number of electrode positions  
14 (Figure 3). It increases from less than 1 million for a line with 60 electrodes to over 317 million with  
15 260 electrodes. We next examine possible ways to reduce the size of the comprehensive data set without  
16 significantly reducing the model resolution. The arrays in the comprehensive data set can be divided  
17 into those of the alpha and beta types (Figure 4). The first modification examined is when the arrays are  
18 restricted to those of the beta type. This might be a practical consideration if I.P. measurements are also  
19 made. By separating the current and potential dipoles, noise due to EM coupling can be greatly reduced.  
20 For a survey line with 60 electrodes, this modification reduces the comprehensive data set from 931320  
21 to 443685 arrays, i.e. a reduction of slightly over 50%. The second modification considered is to restrict  
22 the beta arrays to those that are symmetrical (Figure 4c) that greatly reduces the number of arrays to  
23 13211. The final case examined is where both the symmetrical alpha (Figure 4d) and beta arrays are  
24 used that gives 29886 possible arrays.

1           We next examine the model resolution sections for the different comprehensive data subsets (that  
2 shows the maximum possible resolution) to study the reduction in the resolution caused by the  
3 restrictions used. The example in Figure 5 has 60 electrodes that are located between the 0 and 59  
4 meters marks on the surface. The model section is extended 6 meters from both ends of the line to  
5 ensure all regions with significant resolution values are included (Loke et al., 2014a). The 0.025 model  
6 resolution contour line for the full comprehensive data set (Figure 5a) reaches to a depth of about 10.3  
7 meters below the center of the survey line. The same contour line for the beta arrays subset reaches to  
8 almost the same depth below the center of the line, but its depth decreases more rapidly towards the ends  
9 of the survey line. The loss in resolution is more clearly illustrated in relative model resolution section  
10 (Figure 6a) which shows the resolution for each model cell (for the beta arrays subset) divided by the  
11 cell resolution for the full comprehensive data set. Below the middle of the line, the relative model  
12 resolution values are fairly high (above 0.8) even towards the bottom of section. However, towards the  
13 sides of the section the relative model resolution values drops to below 0.2.

14           The 0.025 resolution contour line reaches to a depth of about 8 meters (Figure 5c) for the  
15 symmetrical beta arrays subset. The loss of resolution with depth and towards the sides is more clearly  
16 shown in the relative resolution plot (Figure 6b). Below a depth of about 2 meters, the relative resolution  
17 is below 0.725. Towards the sides, the loss of resolution is even more severe than the beta arrays subset.  
18 The subset with both the symmetrical alpha and beta arrays (Figure 3d) has about the same maximum  
19 depth for the 0.025 contour line, but the loss of resolution towards the sides is less rapid compared to  
20 Figure 5c. The better performance of this subset at the sides compared to the beta arrays is better  
21 illustrated in the relative resolution plot (Figure 6c) where it achieves maximum values of above 0.8  
22 beyond the ends of the line. It also shows that the resolution at depth below the center of the line is  
23 slightly better than the symmetrical beta arrays with values of above 0.70 covering a larger region.

1           Thus from the model resolution sections, we expect that the optimized data set using only the  
2 beta arrays will have a maximum depth of investigation similar to the full data set, but the resolution  
3 decreases more rapidly towards the ends of the line. The optimized symmetrical beta arrays will have a  
4 shallower depth of investigation with poorer resolution towards the ends as well. If both the symmetrical  
5 alpha and beta arrays are used, the loss in resolution towards the ends of the line should be reduced.

6           Figure 3 show the number of arrays (and the calculation times to generate an optimized data set  
7 with 10000 arrays) for the full comprehensive data set and the symmetrical arrays subset for survey lines  
8 with 60 to 260 electrodes. Beyond 200 electrodes, it was not practical to calculate the optimized arrays  
9 for the full comprehensive data set. The calculation time is reduced by fifteen times for a line with 200  
10 electrodes by using only the symmetrical arrays. The average resolution of the symmetrical arrays subset  
11 is about 75 to 80% of the full set (Figure 5), although the number of data points is only about 1 to 3% of  
12 the full set.

13           As an illustration of the improvements made using the different techniques, the time taken to  
14 generate 10000 data points for a survey line with 140 electrodes is reduced from 74649 (using the direct  
15 method) to 2572 seconds using the decomposition method. This is further reduced to 409 seconds by  
16 using the symmetrical alpha and beta arrays subset.

17           In the following discussion, the arrays generated using the full comprehensive data set with all  
18 the viable alpha and beta arrays will be referred to as the 'Full' optimized data set. The other data sets  
19 generated using the different comprehensive data subsets will be referred to as the 'Beta', 'Symmetrical  
20 Beta (Sym-Beta)' and 'Symmetrical (Sym)' optimized data sets.

21

## 22 **RESULTS**

23           In this section, we test the quality of the arrays produced by the different data sets using a  
24 synthetic model and a field survey. Since the true resistivity distribution for the synthetic model is

1 known, it provides an objective means to assess the quality of the inversion models from the different  
2 data sets.

3

#### 4 **Synthetic test model**

5 The test model has a two layer medium (500 and 100  $\Omega\text{m}$  with thickness of 40 m for the top  
6 layer) with six low resistivity blocks (50  $\Omega\text{m}$ ) in the upper layer and three high resistivity blocks (1000  
7  $\Omega\text{m}$ ) in the lower layer (Figure 7) below a survey line with 85 electrodes. The first data set consists of all  
8 the Wenner-Schlumberger (W-S) arrays (Loke et al., 2013) where the 'n' factor (ratio of the A-M  
9 spacing to the M-N spacing) are integer values. It has 4992 data points. A maximum geometric factor  
10 20735 m (corresponding to a dipole-dipole array with  $a=5$  m and  $n=10$ ) was set for the W-S and  
11 optimized data sets. Voltage dependent Gaussian random noise with a mean amplitude of 0.5 milliohm  
12 (Zhou and Dahlin, 2003; Press et al., 2007) was added to the data before they were converted to  
13 apparent resistivity values. The data sets have a minimum resistance value of about 6 milliohm. Thus the  
14 data points with the lowest resistance values have a noise level of about 8%. The average noise level is  
15 about 0.5% for the W-S data set, and 2 to 3% for the optimized data sets. The L1-norm method is used  
16 for both the data misfit and model roughness (Farquharson and Oldenburg, 1998; Loke et al., 2003) in  
17 the inversion of the data set. The L-curve method (Farquharson and Oldenburg, 2004; Loke et al.,  
18 2014b) is used to automatically select the appropriate damping factor  $\lambda$  in equation (1).

19 The W-S model (Figure 8a) correctly shows the general two-layer structure and resolves the  
20 shallow low resistivity blocks, but does not resolve the three deeper high resistivity blocks. The  
21 inversion model for 'Full' optimized data set with the same number of data points resolves all blocks  
22 (Figure 8b) although the data misfit (2.6%) is significantly higher than the W-S data set (0.5%). This is  
23 due to the higher average geometric factor (13824 m) of the optimized data set compared to the W-S  
24 data set (2343 m). This is consistent with earlier results obtained by Loke et al. (2010a). The models for

1 the 'Beta' (Figure 8c) and 'Symmetrical Beta' (Figure 8d) optimized data sets manage to resolve the  
2 central high resistivity block, but fail to clearly resolve the left and right high resistivity blocks. As  
3 shown earlier in the model resolution sections, these arrays have poorer resolution towards the ends of  
4 the line. The 'Symmetrical' optimized data set model (Figure 8e) shows the correct shapes of the three  
5 high resistivity blocks. It also has a lower data misfit of 1.9% compared to the 'Full' optimized data set  
6 due to a lower average geometric factor of 10416 m.

7 To compare the accuracy of the models, we calculate the mean squared misfits between the true ( $r_t$ )  
8 and the calculated ( $r_c$ ) model values using the following equation.

$$9 \quad \delta_m = \frac{1}{n} \left[ \sum_{i=1}^n [\log(r_t(i)) - \log(r_c(i))]^2 \right]^{0.5} \quad (10)$$

10 The W-S model has the largest misfit (Table 2) while the 'Full' optimized data set model has the lowest,  
11 which agrees with a visual inspection of the models. The misfit for the 'Symmetrical' optimized data set  
12 model is only about 2.5% higher than the 'Full' optimized data set model, although the arrays are derived  
13 from a data subset that is only 2% of the full comprehensive data set. The models for 'Beta' and  
14 'Symmetrical Beta' optimized arrays data sets do have lower model misfits than the W-S model, but  
15 significantly higher misfits compared to the 'Full' and 'Symmetrical' data sets. This is mainly due to the  
16 poorer resolution of the left and right high resistivity blocks achieved by these data sets.

17

## 18 **Field data set**

19 The field data were acquired on a well characterised wetland site (Chambers et al., 2014), located  
20 on the River Lambourn at Boxford, Berkshire, UK. The survey target was a relict channel (~0.5 – 2 m  
21 thick) comprising peat and alluvium, occupying a topographic low on the surface of fluvial sands and  
22 gravels, which are several meters thick and overlie chalk bedrock.



1           The ERT line consisted of 100 electrodes at 0.5 m spacing perpendicular to the long axis of the  
2 channel. The survey line intersected the location of a borehole, and peat depths were probed at 1 m  
3 intervals along the line and along parallel lines 1 m to either side. The average standard deviation of the  
4 three depths at each 1 m interval along the parallel lines was small (6.3 percent). The resistivity data  
5 were measured using a GeoLog-2000 GeoTom instrument.

6           To complete the field surveys in reasonable time, the Wenner-Schlumberger (“W-S”) survey was  
7 limited to symmetric configurations where the ratio of the current-potential electrode separation to the  
8 potential dipole length,  $n$ , was an odd integer. Since electrical noise at the site was very low (Chambers  
9 et al., 2014), all possible combinations of dipole length and odd integer  $n$ -level were used. This gave  
10 4242 measurements with a maximum geometric factor of 3849 m and a maximum median depth-of-  
11 investigation of 10 m. These limits were applied to select optimized surveys designs from the Full, Beta,  
12 Symmetrical Beta (“Sym-Beta”), and Symmetrical (“Sym”) comprehensive sets. Reciprocals of these  
13 surveys were also constructed so that data quality could be assessed (Labrecque et al., 1996). The  
14 arrangement of the measurements in the command sequences was reordered to minimize electrode  
15 polarization effects (Wilkinson et al., 2012). The minimum separation between electrodes being used for  
16 current injection followed by potential measurement was 126 configurations for W-S, 279 for Full, and  
17 209 for Sym. The measurement rate was ~55 measurements per minute, so these corresponded to 2.3,  
18 5.1 and 3.8 minutes respectively, which has proved to be more than sufficient for electrode polarization  
19 to decay in previous field surveys (Wilkinson et al., 2012). The Beta and Sym-Beta surveys were  
20 arranged so that electrode polarization would not occur, as with a standard dipole-dipole survey.

21           The apparent resistivity for each configuration was taken to be the mean of the normal and  
22 reciprocal measurements and its error estimate (“reciprocal error”) was the percentage standard error in  
23 that mean. In the first survey (Full), a large proportion of the data involving electrodes at 28, 32.5, 36,  
24 37, 37.5 and 38 m along the line had high reciprocal errors (>5 percent). The contact resistance estimates

1 for these electrodes were still settling down when this survey was undertaken, being some 20-25 percent  
2 higher than for all subsequent surveys. To provide a better comparison between Full and the other  
3 optimized surveys, measurements with reciprocal errors >5 percent in the original Full survey were  
4 replaced by the same measurements with reciprocal errors <5 percent from either the Sym survey (for  
5 preference since this was measured closest in time to the Full survey) or the Beta survey. This new data  
6 set is referred to as the "Full (reconstructed)" survey ("Full(r)") and is used in the remainder of this  
7 analysis in place of the original Full data. The distributions of reciprocal error in each of the surveys are  
8 shown in Table 3. The data quality is generally excellent with at least 97.6 percent of every survey  
9 having reciprocal errors <5 percent (Table 3).

10 The inversions used an L1-norm for both data misfit and model roughness and the L-curve  
11 method was used to select the damping factor. The inverted models and mean absolute misfit errors are  
12 shown in Figure 9. The misfits are consistent with the levels of data noise as characterized by the  
13 reciprocal error distributions. All the inverted models (Figure 9) exhibit a conductive surface layer  
14 between ~0.5 – 2 m thick corresponding to the peat, overlying a thicker resistive layer of sand & gravel.  
15 Below this is a region of intermediate resistivity corresponding to the chalk bedrock. In the optimized  
16 images, but not the W-S image, the resistivity of the chalk increases with depth below about 82 m AOD,  
17 which agrees with a previous larger scale 3D resistivity survey at the site. Intrusive investigations have  
18 confirmed that the top of the chalk is typically weathered to a putty-like consistency, which has a higher  
19 clay content and lower resistivity than the deeper unweathered chalk (Chambers et al., 2014). The white  
20 dots show the locations of the peat / sand & gravel interface determined by the depth probe. The  
21 borehole is 9 m deep and is situated at a distance 23.3 m along the line. It shows the elevations of the  
22 peat / sand & gravel and the sand & gravel / chalk interfaces.

23 To compare the images with the available ground truth, *k*-means clustering was performed on the  
24 logarithms of the cell resistivities to segregate the images into clusters representing lithological units

1 (Ward et al., 2014). Initially two clusters were used to identify the transition between the conductive  
2 peat and the resistive sand & gravel. The resistivity value representing this interface was taken to be the  
3 mean boundary resistivity between the clusters across all the inverted models. The mean  
4  $\log_{10}(\text{resistivity})$  boundary value was found to be 1.728 with a standard deviation of 0.007,  
5 corresponding to 53.5  $\Omega\text{m}$  with a standard deviation of 0.8  $\Omega\text{m}$ . To identify the sand & gravel / chalk  
6 interface, three clusters had to be used. This gave a mean  $\log_{10}(\text{resistivity})$  boundary value of 2.039 with  
7 a standard deviation of 0.013 (109.4  $\Omega\text{m}$  with a standard deviation of 3.2  $\Omega\text{m}$ ). The interface contours at  
8 53.5  $\Omega\text{m}$  and 109.4  $\Omega\text{m}$  are shown as black lines in Figure 9. The *k*-means boundaries are compared to  
9 the borehole interfaces in Table 4. For the optimized surveys, the differences between the borehole and  
10 ERT-derived elevations for both interfaces are less than half those obtained with the standard W-S  
11 survey. The plots on the right of Figure 9 show the variation of resistivity with elevation at the borehole  
12 location. They demonstrate that the optimized surveys more clearly define both the upper (peat) and  
13 lower (chalk) boundaries of the sand & gravel layer. This improvement is considerably more  
14 pronounced for the chalk interface, which is characterized by steeper resistivity gradients in the  
15 optimized images than in the standard image. This is because the chalk interface is deeper, and hence in  
16 a region of lower sensitivity, where the advantages of optimized survey designs are greater. For the peat  
17 interface, comparisons with the intrusive probe data are shown in Table 5 in terms of mean, mean  
18 absolute, and root-mean-squared differences. By all these measures, the shallow peat interface  
19 boundaries determined from the optimized survey images also correspond more closely to the intrusive  
20 data than the boundary from the standard W-S survey image.

21

## 22 **CONCLUSION**

23 The single-precision version of the 'Compare R' algorithm using a new ranking function  
24 (Wilkinson et al., 2012) reduced the calculation time by a factor of two to eight times compared to the

1 double-precision version while providing almost identical results. Recent improvements in GPU  
2 computer technology further reduces the calculation time by about seven times. The calculation time is  
3 reduced by half by taking advantage of symmetry in the model resolution values for symmetrical pairs  
4 of arrays. The calculation time is reduced by more than thirty times by calculating the Sherman-  
5 Morrison update for all the possible two-electrode combinations which are then used to reconstruct the  
6 model resolution update for the four-electrode arrays in the comprehensive data set. The techniques  
7 developed in this research have reduced the calculation time by more than two orders of magnitude, and  
8 pushed the practical limit of the 'Compare R' method to about 200 electrodes with current (2014)  
9 microcomputer technology.

10 For surveys that involve a larger number of electrodes, the calculation time can be further  
11 reduced by another order of magnitude by using a subset consisting of the symmetrical arrays from the  
12 full comprehensive data set. Although the number of arrays in the symmetrical subset is only about 1 to  
13 3% of the full set, the average model resolution is about 75 to 80% of the full comprehensive data set.  
14 Optimized arrays generated from the symmetrical subset are also less sensitive to noise due to a smaller  
15 average geometric factor value. Tests with a synthetic model show that the misfit between the calculated  
16 and true resistivity values derived for the "Symmetrical" optimized arrays is less than 3% higher than the  
17 model from the 'Full' optimized data set. The combined effect of the techniques described in this paper  
18 now makes it practical to use the 'Compare R' array optimization techniques for surveys lines with more  
19 than 200 electrode positions.

20 A field test of the optimized survey designs was undertaken at a wetland site with considerable  
21 ground truth in the form of high-resolution intrusive probe data and a borehole. The results showed that  
22 the images from the optimized data sets located the lithological interfaces with quantitatively greater  
23 accuracy than the standard Wenner-Schlumberger array. This was the case for both the shallow interface  
24 determined by the probe data and the deeper interface from the borehole log. Compared to the standard

1 image, the variations in resistivity in the deeper regions of the optimized images were also qualitatively  
2 in better agreement with observations from previous, larger scale, surveys and other intrusive  
3 investigations.

4 For long 2-D survey lines the best compromise at present seems to be the use of the symmetrical  
5 arrays subset of the comprehensive data set, although this is at the expense of a small reduction in the  
6 model resolution. We are presently investigating the optimum balance between increasing the resolution  
7 while minimizing the calculation time by allowing a limited number of non-symmetrical arrays. We are  
8 also looking methods to adapt the techniques for multi-channel survey instruments (Wilkinson et al.,  
9 2012). The use of a space varying weighting function to generate arrays that are optimized for resolving  
10 structures in specified areas is also being studied. Another area of current investigation is the use of  
11 optimized pole-dipole arrays to obtain a deeper depth of investigation (Bloome et al., 2011). As 3-D  
12 effects can be a significant factor in some areas, research is being conducted to adapt these techniques to  
13 optimized arrays for 3-D surveys (Loke et al., 2014c). We are also studying the possible improvement in  
14 the performance of the optimized arrays by incorporating known topography and non-homogeneous  
15 subsurface resistivity distributions.

16

## 17 **ACKNOWLEDGEMENTS**

18 This paper is published with the permission of the Executive Director of the British Geological Survey  
19 (NERC). We would like to thank two reviewers for their insightful comments that have helped to  
20 improve the paper.

1   **REFERENCES**

- 2   Auken, E., Pellerin, L., Christensen, N.B. and Sørensen, K. 2006. A survey of current trends in near-  
3       surface electrical and electromagnetic methods. *Geophysics* 71, G249-G260.
- 4   Blome, M., Maurer, H. and Greenhalgh, S., 2011, Geoelectric experimental design — Efficient  
5       acquisition and exploitation of complete pole-bipole data sets. *Geophysics* 76, F15-F26.
- 6   Carpenter, E.W. and Habberjam, G.M., 1956. A tri-potential method of resistivity prospecting.  
7       *Geophysics* 11, 455-469.
- 8   Chambers, J.E., Wilkinson, P.B., Uhlemann, S., Sorensen, J.P.R., Roberts, C., Newell, A.J., Ward, W.,  
9       Binley, A., Williams, P.J., Goody, D.C., Old, G. and Bai, L., 2014. Derivation of lowland  
10      riparian wetland deposit architecture using geophysical image analysis and interface detection.  
11      *Water Resources Research*, 50, 5886–5905.
- 12  Ellis, R.G. and Oldenburg, D.W., 1994. Applied geophysical inversion. *Geophysical Journal*  
13      *International* 116, 5-11.
- 14  Farber, R., 2011. *CUDA Application Design and Development*. Elsevier, Amsterdam.
- 15  Farquharson, C.G. and Oldenburg, D.W. 1998. Nonlinear inversion using general measures of data  
16      misfit and model structure. *Geophysical Journal International* 134, 213-227.
- 17  Farquharson, C.G. and Oldenburg, D.W., 2004. A comparison of automatic techniques for estimating  
18      the regularization parameter in non-linear inverse problems. *Geophysical Journal International*  
19      156, 411-425.
- 20  Gerber, R., Bik, A. J.C., Smith, K.B. and Tian, X. 2006. *The Software Optimization Cookbook* (second  
21      edition). Intel Press, Intel Corporation, Hillsboro, Oregon, USA.
- 22  LaBrecque, D.J., Miletto, M., Daily, W., Ramirex, A. and Owen, E., 1996. The effects of noise on  
23      Occam’s inversion of resistivity tomography data. *Geophysics* 61, 538–548.

1 Loke M.H., Acworth, I. and Dahlin, T., 2003. A comparison of smooth and blocky inversion methods in  
2 2D electrical imaging surveys. *Exploration Geophysics* 34, 182-187.

3 Loke, M.H., Wilkinson, P. and Chambers, J., 2010a. Fast computation of optimized electrode arrays for  
4 2D resistivity surveys. *Computers & Geosciences* 36, 1414-1426.

5 Loke, M.H., Wilkinson, P. and Chambers, J., 2010b. Parallel computation of optimized arrays for 2-D  
6 electrical imaging. *Geophysical Journal International* 183, 1202-1315.

7 Loke, M.H., Chambers, J.E., Rucker, D. F., Kuras, O. and Wilkinson, P. B., 2013. Recent developments  
8 in the direct-current geoelectrical imaging method. *Journal of Applied Geophysics* 95, 135-156.

9 Loke, M.H., Wilkinson, P., Chambers, J. and Strutt, M., 2014a. Optimized arrays for 2-D cross-borehole  
10 electrical tomography surveys. *Geophysical Prospecting* 62, 172-189.

11 Loke, M.H., Dahlin, T. and Rucker, D.F., 2014b. Smoothness-constrained time-lapse inversion of data  
12 from 3-D resistivity surveys. *Near Surface Geophysics* 12, 5-24.

13 Loke, M.H., Wilkinson, P.B., Uhlemann, S.S., Chambers, J.E. and Oxby, L. S., 2014c. Computation of  
14 optimized arrays for 3-D electrical imaging surveys. *Geophysical Journal International* 199,  
15 1751-1764.

16 Maurer, H., Curtis, A. and Boerner, D.E., 2010. Recent advances in optimized geophysical survey  
17 design. *Geophysics* 75, A177–A194.

18 Menke, W., 1989. *Geophysical data analysis: Discrete inverse theory (Revised edition)*. Academic Press  
19 Inc., San Diego.

20 Press, W.H. , Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P, 2007. *Numerical Recipes in C (Third*  
21 *Edition)*. Cambridge University Press, Cambridge, UK.

22 Stummer, P., Maurer, H. and Green, A., 2004. Experimental design: Electrical resistivity data sets that  
23 provide optimum subsurface information. *Geophysics* 69, 120-129.

- 1 Ward, W.O.C, Wilkinson, P.B., Chambers, J.E., Oxby, L.S. and Bai, L., 2014. Distribution-based fuzzy  
2 clustering of electrical resistivity tomography images for interface detection. *Geophysical*  
3 *Journal International* 197, 310-321.
- 4 Wilkinson, P.B., Meldrum P.I., Chambers J.E., Kuras, O. and Ogilvy R.D., 2006. Improved strategies  
5 for the automatic selection of optimized sets of electrical resistivity tomography measurement  
6 configurations. *Geophysical Journal International* 167, 1119-1126.
- 7 Wilkinson, P.B., Loke M.H., Meldrum, P.I., Chambers, J.E, Kuras, O., Gunn, D.A. and Ogilvy, R.D.,  
8 2012. Practical aspects of applied optimised survey design for Electrical Resistivity  
9 Tomography. *Geophysical Journal International* 189, 428-440.
- 10 Zhou, B. and Dahlin, T, 2003. Properties and effects of measurement errors on 2D resistivity imaging  
11 surveying. *Near Surface Geophysics* 1, 105-117.
- 12



- 1 Table 1. Example partial list of electrode numbers of arrays from the comprehensive data set for a
- 2 survey line with 85 electrodes.

Array number	A	B	M	N
36	1	9	2	3
37	1	9	2	4
38	1	9	2	5
39	1	9	2	6
40	1	9	2	7
41	1	9	2	8
42	1	9	3	4
43	1	9	3	5
44	1	9	3	6
45	1	9	3	7
46	1	9	3	8
47	1	9	4	5
48	1	9	4	6
49	1	9	4	7
50	1	9	4	8

3

4

1 Table 2. RMS model misfits for the inversion models obtained using the different data sets for the  
 2 synthetic test model.

Wenner- Schlumberger array	Optimized data sets			
	Full	Beta	Symmetrical Beta	Symmetrical
0.0146	0.0120	0.0138	0.0139	0.0123

3

4

5 Table 3. Percentage of data set with reciprocal error less than the given level.

Field Data Set	Reciprocal Error Level (%)					
	0.1	0.2	0.5	1.0	2.0	5.0
W-S	53.72	79.75	96.25	99.69	99.91	99.95
Full	23.71	41.95	78.18	92.15	93.28	93.42
Full(r)	24.94	44.14	82.02	96.51	97.74	97.95
Beta	24.06	42.42	77.94	95.88	97.76	98.14
Sym-Beta	31.13	51.36	84.00	95.92	97.24	97.60
Sym	30.64	50.81	86.85	97.93	98.87	99.01

6

7

1 Table 4. Differences between borehole interfaces and *k*-means boundaries.

Field Data Set	Peat / sand & gravel elevation difference (m)	Sand & gravel / chalk elevation difference (m)
W-S	0.24	-1.08
Full(r)	0.10	0.45
Beta	0.06	0.23
Sym-Beta	0.11	0.42
Sym	0.05	-0.07

2

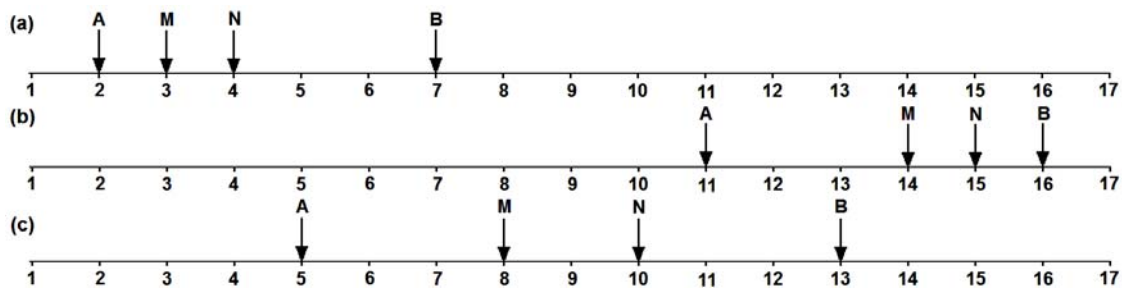
3

4 Table 5. Differences between peat probe interface and *k*-means boundary.

Field Data Set	Mean Difference (m)	Mean Absolute Difference (m)	Root Mean Squared Difference (m)
W-S	0.022	0.098	0.129
Full(r)	0.001	0.078	0.107
Beta	0.005	0.066	0.093
Sym-Beta	0.006	0.069	0.093
Sym	0.005	0.071	0.093

5

6



1

2

Figure 1. (a) Example array of an alpha type array and it's (b) symmetrical counterpart for a survey line

3

with 17 electrodes. (c) Example of an array that is symmetrical about the center of the line and does not

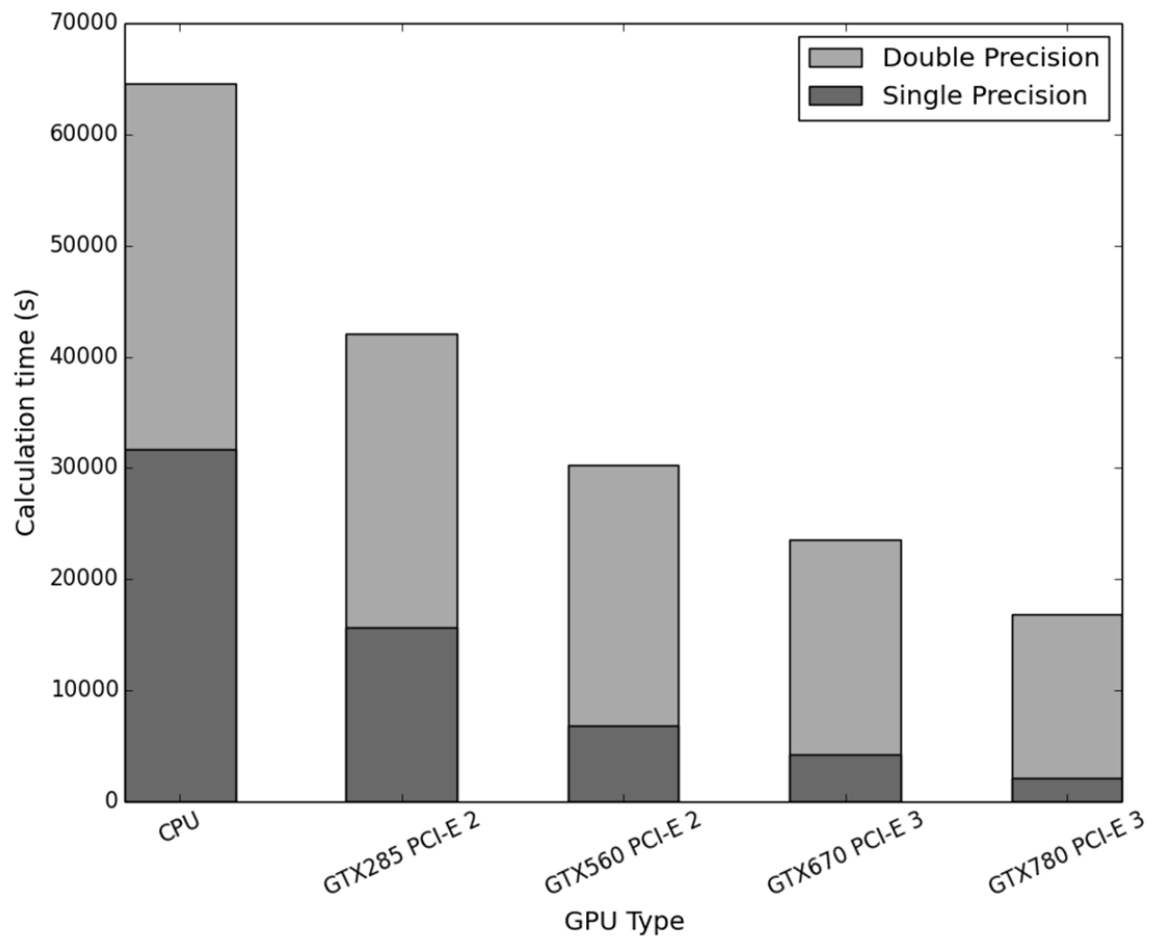
4

have a symmetrical counterpart. A and B are the current electrodes, while M and N are the potential

5

electrodes.

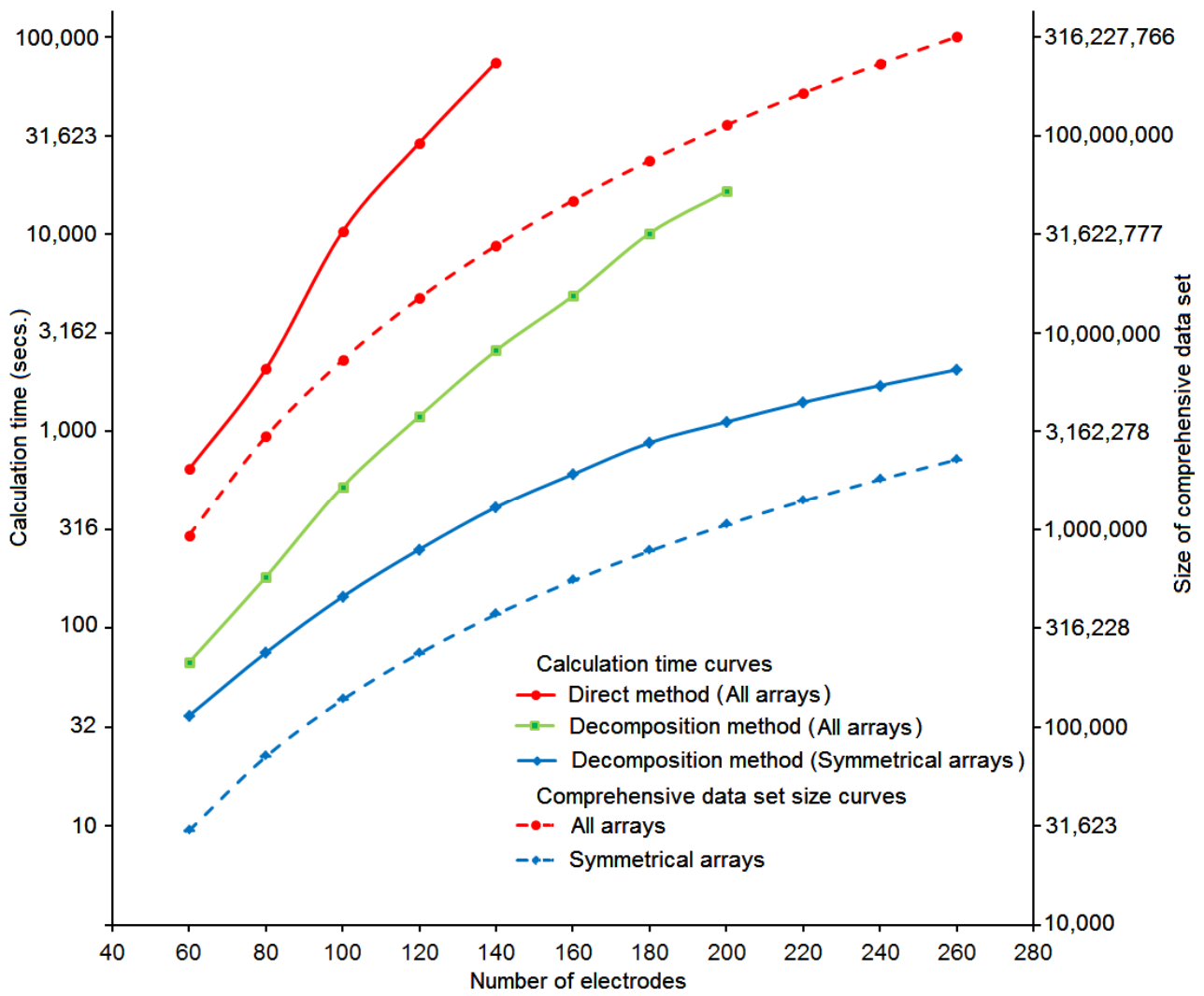
6



1

2 Figure 2. The times taken to generate 10000 data points for a 2-D survey line with 80 electrodes using  
 3 different graphics cards. The single- and double-precision versions gave average relative model  
 4 resolution values of 0.6543747 and 0.6534324, and average resolution values of 0.1216988 and  
 5 0.1218544 respectively.

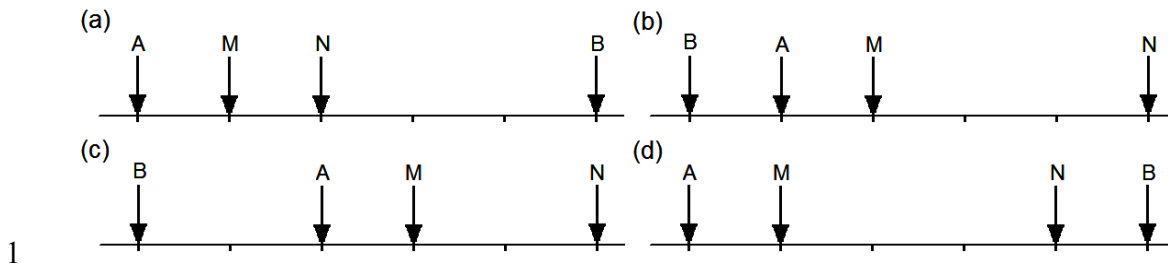
6



1

2 Figure 3. Change of the size of the comprehensive data set and calculation time to generate 10000 data  
 3 points for a 2-D survey line with 60 to 260 electrodes. The calculation times are given for the direct and  
 4 decomposition methods of applying the Sherman-Morrison update. The comprehensive data set size for  
 5 the full data set with all viable alpha and beta arrays, and a subset consisting of the symmetrical arrays  
 6 are shown. Note a logarithmic scale is used for the calculation time and comprehensive data set size.

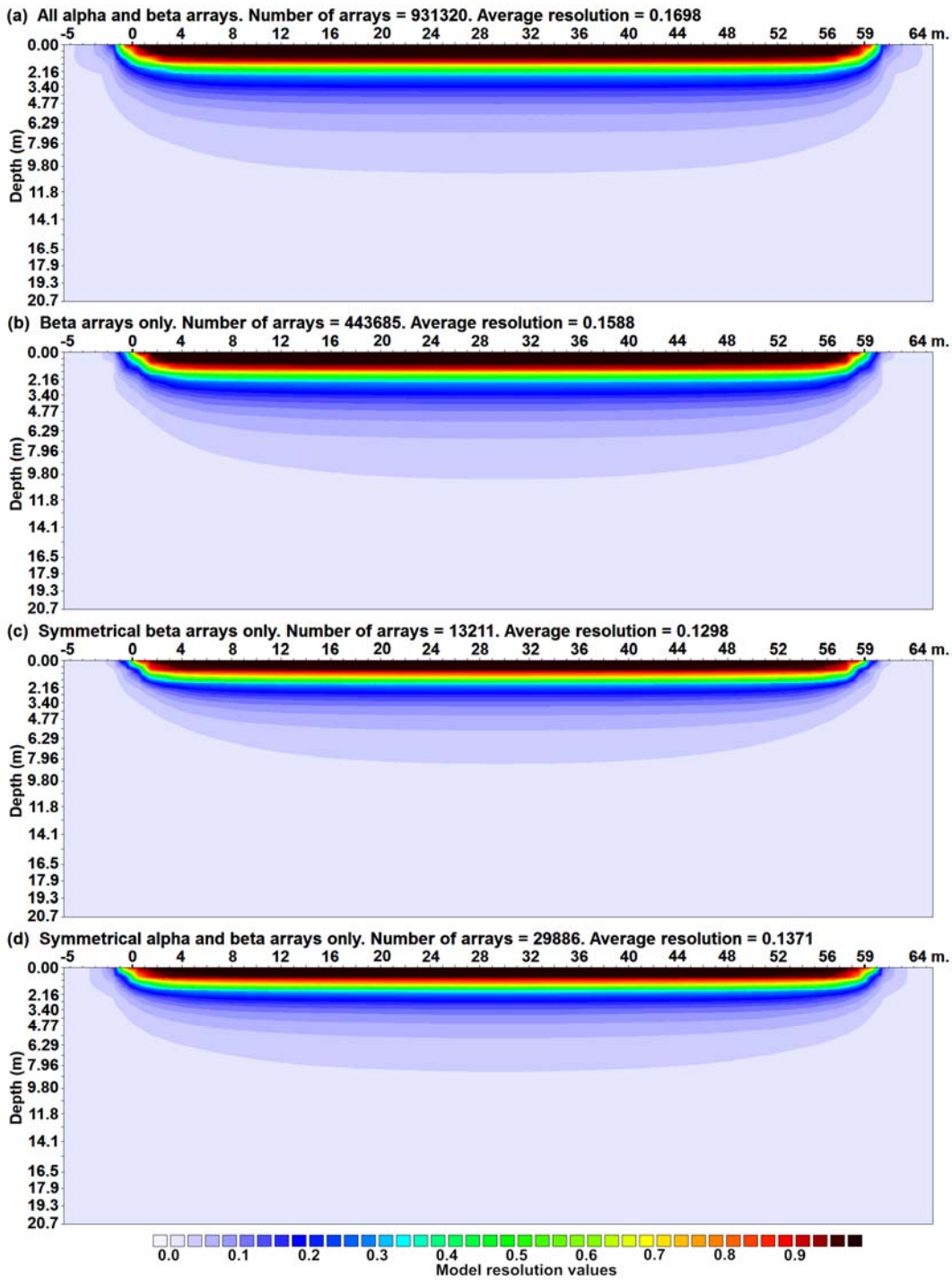
7



1  
2 Figure 4. Arrangement of electrodes for arrays of the (a) alpha, (b) beta and (c) symmetrical beta and (d)  
3 symmetrical alpha types.

4

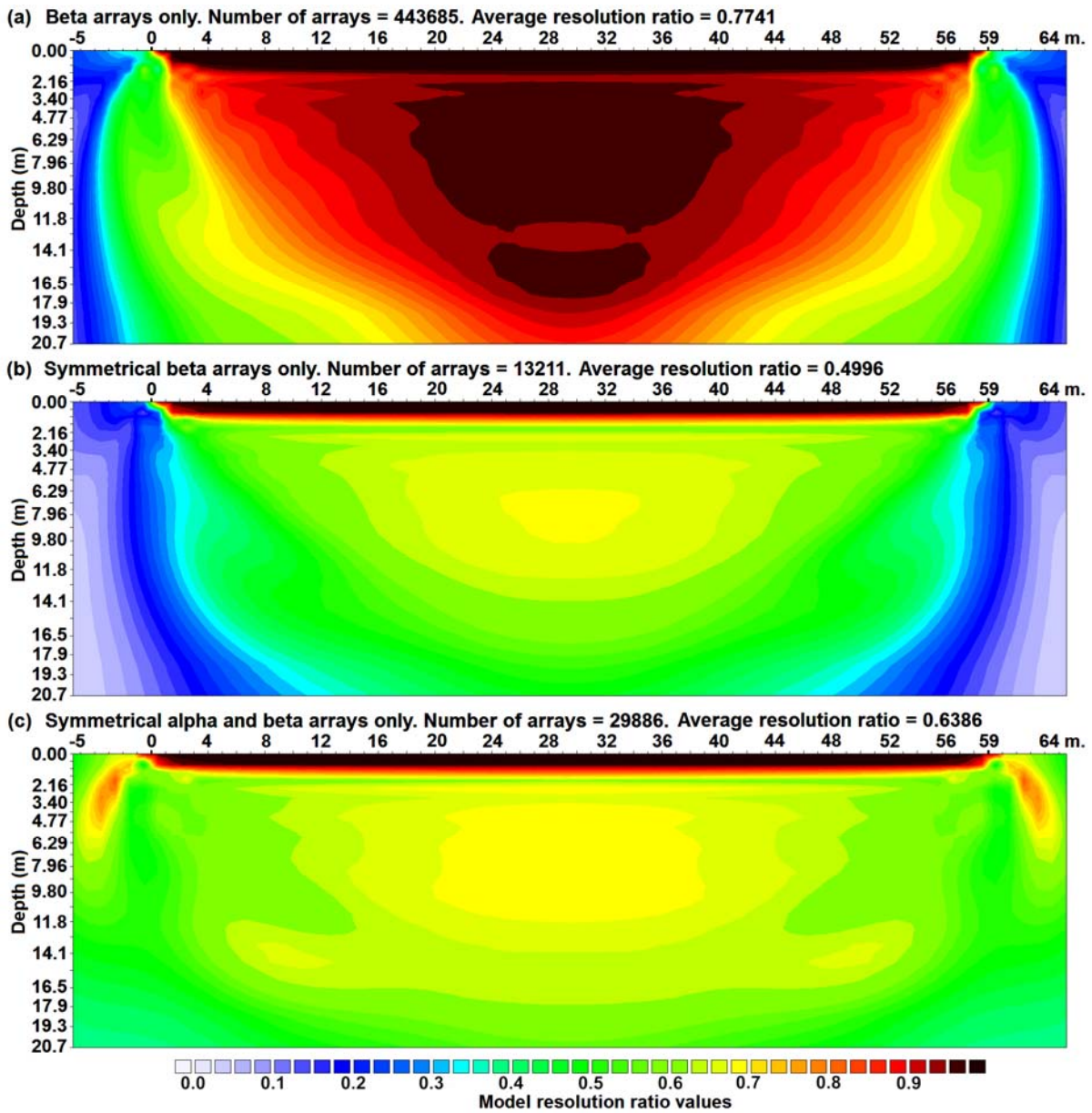
5



1

2 Figure 5. Model resolution sections for the comprehensive data sets for the (a) ‘Full’ data set with all the  
 3 viable alpha and beta arrays, (b) the ‘Beta’ arrays, (c) the ‘Symmetrical Beta’ arrays and (d) the  
 4 ‘Symmetrical’ arrays. The electrodes are located between the 0 and 59 meters marks on the surface of  
 5 the model sections.

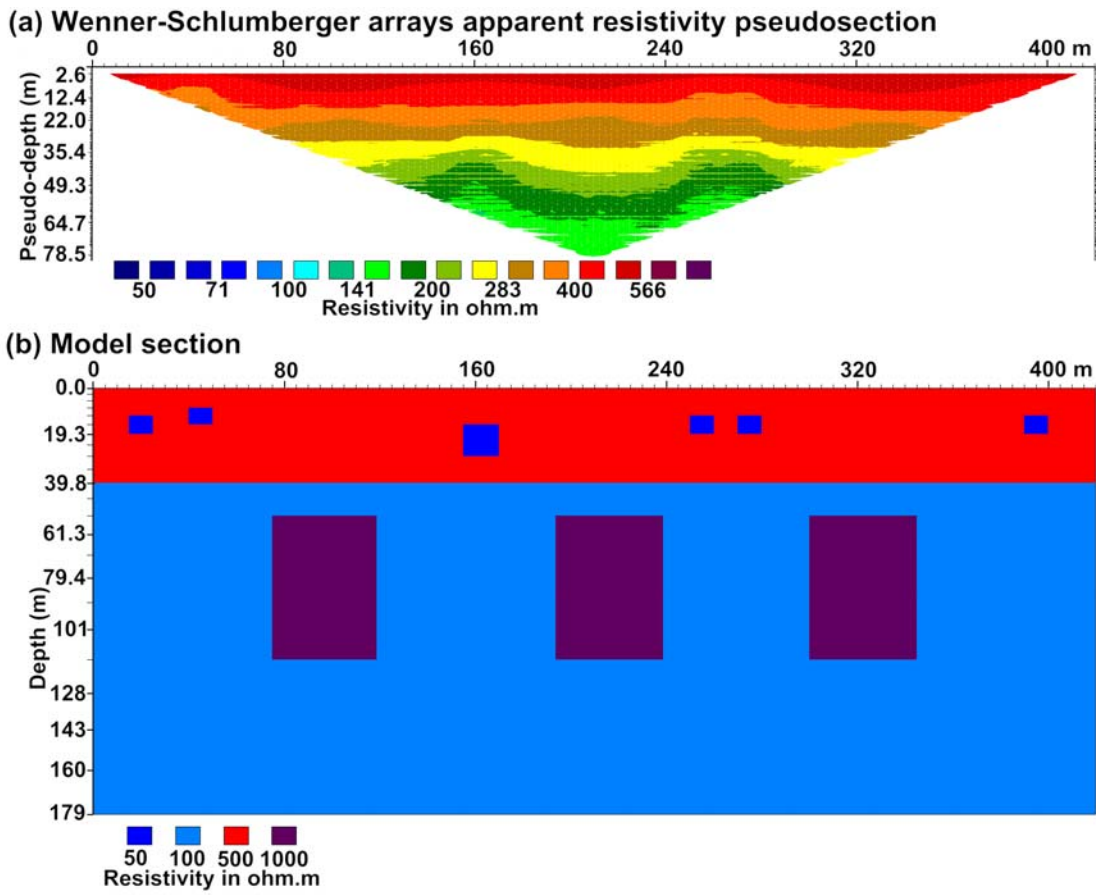




1

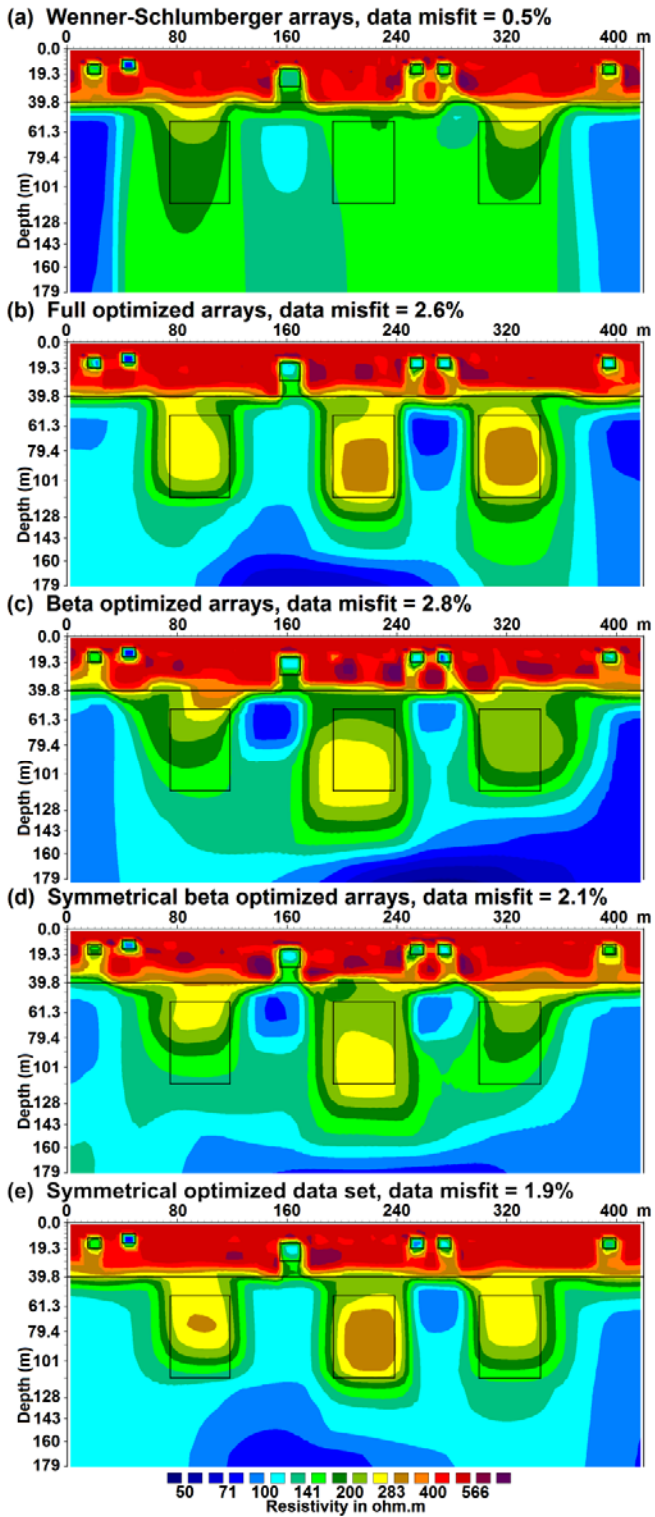
2 Figure 6. Relative model resolution sections for the (a) the ‘Beta’ arrays, (b) the ‘Symmetrical Beta’  
 3 arrays and (c) the ‘Symmetrical’ arrays comprehensive data subsets.

4



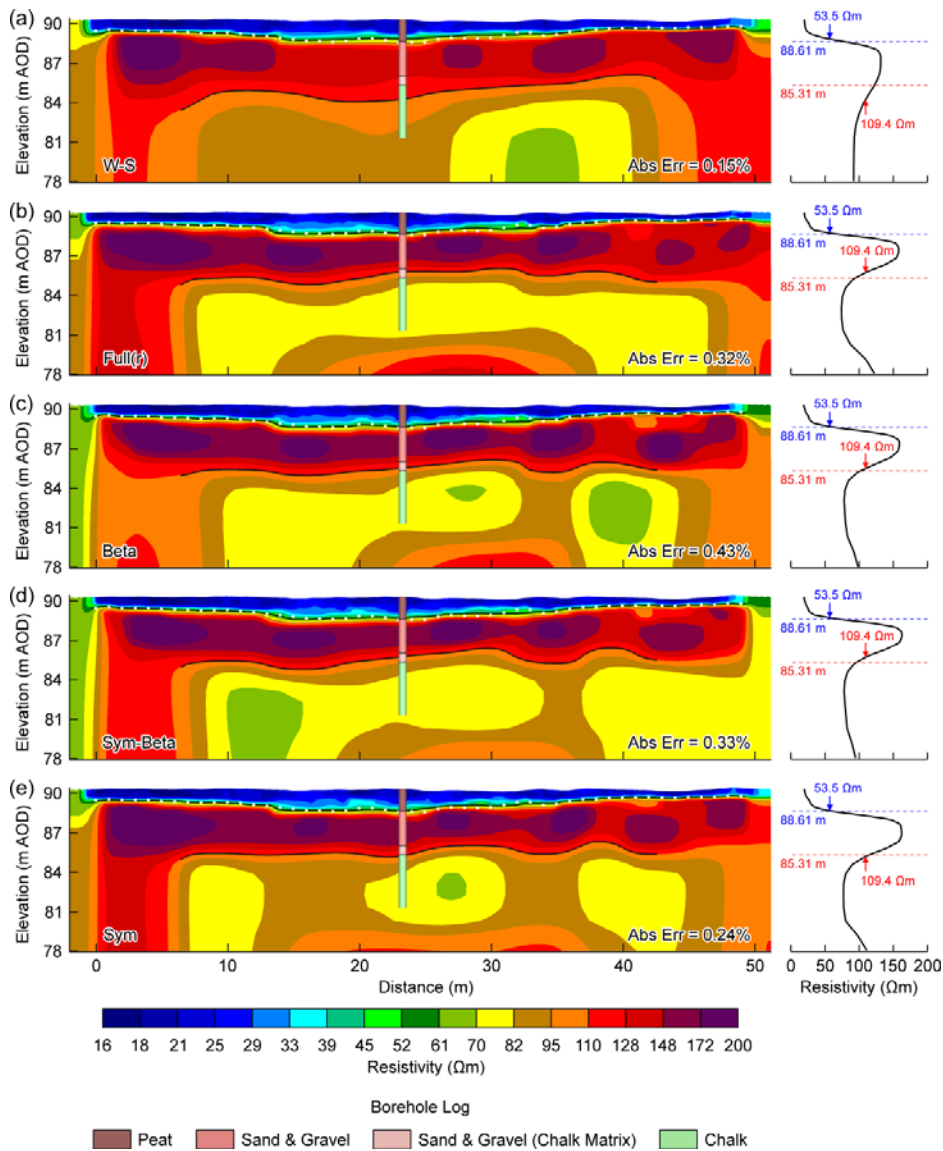
1  
 2 Figure 7. (a) Apparent resistivity pseudosection for the Wenner-Schlumberger array with 4992 data  
 3 points. (b) Synthetic test model consisting of two layer medium with low resistivity blocks in the first  
 4 layer and high resistivity blocks in the second layer below a survey line with 85 electrodes and unit  
 5 electrode spacing of 5 m.

6  
 7



1

2 Figure 8. Results for synthetic test model using different data sets. Inversion models for (a) Wenner-  
 3 Schlumberger data set, the (b) 'Full', (c) 'Beta', (d) 'Symmetrical Beta' and (e) 'Symmetrical' optimized  
 4 arrays.



1  
 2 Figure 9. Inversions of field data from (a) Wenner-Schlumberger, (b) Full, (c) Beta, (d) Symmetric Beta,  
 3 (e) Symmetric survey designs. The mean absolute misfit errors are shown inset. The white dots show the  
 4 intrusively determined locations of the peat / sand & gravel interface. The black lines show iso-  
 5 resistivity contours representing the lithological boundaries determined from *k*-means clustering. The  
 6 plots on the right show the variation of resistivity with elevation at the borehole location. The dashed  
 7 lines show the elevations of the peat / sand & gravel interface (blue) and sand & gravel / chalk interface  
 8 (red). The arrows indicate the resistivities and elevations of the corresponding interfaces determined  
 9 from the iso-resistivity contours.