

Restoration of magnetometry data using inverse-data methods

1 Introduction

Magnetometry has become one of the most popular techniques for the geophysical prospection of archaeological sites. Modern instruments are reliable, easily portable, convenient in use, and reasonably inexpensive. The data are logged automatically and can readily be transferred to a small computer for subsequent processing. Anyone who has received proper training in the operation of field magnetometers can expect to survey a considerable area of ground each day and, provided that conditions are favourable, be able to present the results in an archaeologically meaningful form.

A typical opportunity for magnetometry occurs when ancient ditches or pits have been cut into inert soil, but have subsequently been filled in with material which is magnetically active. The locations of the features can then be detected as induced magnetic anomalies relative to the earth's main field. The response of the magnetometer to such an anomaly is somewhat complicated, presenting a positive lobe along the southern side together with a negative shadow towards the north (Linington 1964). The relative sizes of the positive and negative lobes depend on the mode of operation of the magnetometer (gradiometer or single mobile detector), on the survey's location on the surface of the planet, and on the depth of the archaeological features below the modern ground surface.

Some care has to be taken in interpreting the results of a magnetometer survey, because the actual position of the anomaly corresponds neither to the positive lobe nor to the negative lobe, but is close to the junction between them. This may not be important when the survey reveals only a limited number of well spaced features, but it may cause crucial difficulties if there are many features, overlapping each other at different depths.

The data logged by the magnetometer provide, in effect, a digital image of the site. Such an image differs from more familiar electronic images, such as those derived from conventional photography, only in that it contains both positive and negative readings, whereas in most cases only positive intensities are permitted. Many different methods of image restoration have been developed over the years, and the majority are still valid when applied to images containing a mixture of positive and negative values.

Standard methods of image restoration include spatial filtering, for smoothing and edge enhancement, Fourier transform methods, and construction of the inverse response function. In recent years a number of alternative methods have been developed, based upon statistical estimation techniques, such as the EM algorithm which is discussed in this paper. Such techniques model the distribution of the survey data, on the basis of the known magnetometer response. The magnetic intensity is then estimated from the data, taking into account any reasonable prejudice about the nature of the anomalies. These techniques are known collectively as 'inverse-data methods'.

The aim of our project is to apply suitable inverse-data methods to various types of archaeological magnetic data, and to appraise their success in comparison with standard techniques of image restoration. Some preliminary results are presented in this paper.

2 Outline of this project

Recognising that the analysis of full-scale archaeological field data presented formidable problems, because of both the size of the problem and the complexity of the response function involved, it was decided that the project should be developed in three distinct stages, each progressing towards the ultimate objective.

Stage I

The analysis of digitised measurements of magnetic susceptibility over the length of earth cores from archaeological sites. These measurements arise from a project undertaken by the Department of Archaeological Sciences at the University of Bradford with the aim of determining the location and depth of the magnetically active regions of a site. Earth cores, extracted from several locations over the site, are passed through a detector coil which allows the susceptibility to be measured continuously along the core; the readings are recorded digitally.

Since the detector coil is sensitive to the susceptibility over a considerable length of the core, its effective response function to the susceptibility of any point in the core has very long tails (fig. 1a). In consequence, the curve of the continuous measurements shows very broad, smooth peaks,

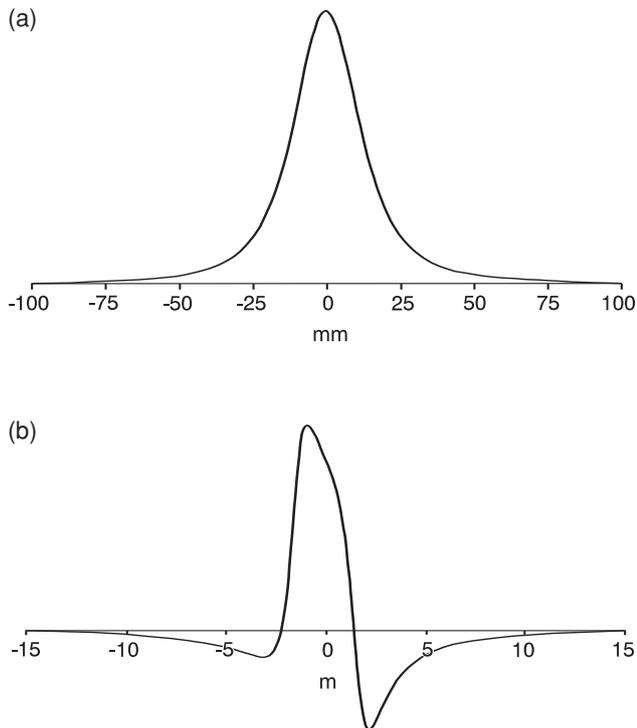


Figure 1. Theoretical response functions for (a) measurements of core susceptibility and (b) line magnetometer measurements.

and responses from different regions of the core may overlap very strongly. Our aim is to account for the broad spread of the response function and to allow the measured susceptibility to be attributed to sharply defined regions of the core, representing distinct epochs in the development of the site.

Since the restoration of susceptibility values is a line problem, it requires relatively little computing power. Furthermore the response function does not possess the positive and negative lobes which normally appear in magnetometry work. Consequently this should provide a fairly straightforward problem on which to test our techniques.

Stage II

Gradiometer measurements along a line transecting linear features. Rather than progress straight to full survey over an extended area of ground, we first look at data from a survey along a line transecting some linear feature, such as a long straight ditch. Here the ground is modelled as a collection of magnetised prisms of rectangular cross-section, and infinite in length in the direction of the feature. The response function has positive and negative lobes, and may vary with location on the planet, the depth of the feature,

and the strike (or transection) angle (fig. 1b). Nevertheless the problem is more manageable than full area field survey at Stage III.

Stage III

Gradiometer measurements from area surveys. This type of problem usually involves a large data set and a more complex response function than either of the earlier stages. Hence it is only to be tackled when we have gained considerable experience of the methodology during the earlier stages.

3 Techniques tested during Stage I, using susceptibility data

Two forms of susceptibility data were used for testing restoration methods. The first form was entirely simulated, by convolving a known pattern of susceptibility with a suitable response function, and finally adding some Gaussian noise. The second form came from actual measurements, but made on blocks of material of known susceptibility fabricated to resemble earth cores, rather than on true earth cores. The advantage of using 'phantom' data of this type is that the expected answer is known, and therefore the accuracy of the results may be judged.

All the techniques worked much as we would have predicted with the simulated data. The straight Fourier method and the direct calculation of the inverse response function tended to give fragmented results in the presence of noise, for reasons discussed below, but the other methods worked well enough. We therefore decided to concentrate on the 'phantom' data, since these should give a fairer indication of how the various techniques would perform in practice.

3.1 FOURIER TRANSFORM METHODS

It is possible to calculate the (discrete) Fourier transform of the response function, to find its exact inverse, which can then be used to calculate the restored susceptibility. In fact the results are extremely disappointing and have little resemblance to the 'known' pattern of susceptibility. The reasons for this poor performance are well known. The response function used here is very smooth and has long tails; consequently its Fourier transform has very small amplitude in the high-frequency components. Hence the inverse of the Fourier transform has very large amplitude in those components. Furthermore, since the observed data represent a convolution which includes the smooth response function, they should have very small high-frequency components; any significant component in those frequencies is almost certainly associated with noise. One effect of dividing by the transformed response function is to exaggerate components arising from noise, which often gives rise to unsatisfactory results.

One technique to counteract the exaggeration of noisy components is the use of the Wiener filter (Gonzalez/Wood 1992). The basis of this method is an analysis of the frequency spectrum resulting from the noise, but in practice it is often interpreted as the simple addition of a small positive constant Φ to the denominator of each component of the inverse of the transformed response function. The constant Φ prevents division by near-zero, but it is difficult to provide a prescriptive formula for its optimum value; a suitable value is usually found by subjective trial and error.

3.2 CONSTRUCTION OF INVERSE RESPONSE FUNCTION

It is possible to construct the inverse response from a set of simultaneous equations derived from a precisely constrained problem (Tsokas *et al.* 1991). The results are effectively equivalent to the straight Fourier method, and have similar deficiencies. Better results may be obtained by setting up an over-constrained problem, which is solved by minimising a sum of squares. This takes account of the presence of noise in the data and leads to results which are qualitatively similar to those from the application of the Wiener filter in Fourier transforms. Singular value decomposition provides a more stable and more controlled approach to the over-constrained problem, but the results are not markedly improved.

3.3 MAXIMUM LIKELIHOOD ESTIMATION

Our initial trial of inverse-data methods was based on the well established Metropolis-Hastings algorithm (Hastings 1970; Metropolis *et al.* 1953). The calculations proved to be extremely slow and the final results did not show any significant improvement over those discussed above. These conclusions are not entirely surprising since the algorithm was devised in order to solve non-linear problems, whereas the problems of fitting the magnetic data are linear ones. As a result of these observations we abandoned the Metropolis-Hastings method in favour of an alternative statistical technique which has proved to be successful in other imaging applications.

3.4 THE EM ALGORITHM

This algorithm was published by Dempster *et al.* (1977) as a summary of various earlier methods, one of the best known of which is the Lucy-Richardson method (Lucy 1974; Richardson 1972). We offer a brief description of the algorithm here, with the intention of publishing more of the mathematical detail elsewhere.

Suppose that the susceptibility profile along the core is divided into m discrete elements, that x_j is the ‘true’ susceptibility of element j , and that \hat{x}_j is some estimate of x_j . Suppose also that data are observed at n locations, and that y_i is the observed value at location i , whereas μ_i is the

expected value when the ‘truth’ is convoluted with the response function.

Then

$$\mu_i = \sum_{j=1}^m h_{ij}x_j \quad \text{and} \quad y_i = \sum_{j=1}^m z_{ij}$$

where h_{ij} is the response function coupling location i to element j , and z_{ij} is the contribution to observation i from element j . The values z_{ij} may be envisaged as ‘unobservable’ data whose expected values are $h_{ij}x_j$. The introduction of such ‘missing’ or ‘unobservable’ data is an essential requirement of the EM algorithm.

The algorithm defines two separate steps:

E step: (Expectation), where z_{ij} is estimated by its conditional expectation, given the data:

$$z_{ij} = E [z_{ij} | y_i] = h_{ij}x_j + \frac{1}{m} (y_i - \mu_i)$$

M step: (Maximisation), where the value \hat{x}_j is found to maximise the log-likelihood, or minimise the error sum of squares, assuming that the z_{ij} are observed data.

The E and M steps may be combined to give a revised estimate of x_j :

$$\hat{x}_j^{new} = \hat{x}_j^{old} + \sum_{i=1}^n (y_i - \mu_i)h_{ij} \Big/ m \sum_{i=1}^n h_{ij}^2$$

This equation provides an iterative process where the estimate of each ‘truth’ element is decoupled from the estimates of the other elements.

Since this simple implementation of the EM algorithm is based on minimising an error sum of squares, the results are essentially similar to those from the Wiener filter or the other least squares methods. In consequence there is a choice of methods leading to similar results:

EITHER The normal equations are set up for a least squares calculation, or the equivalent Fourier transforms are used, both of which involve very large arrays, so that the calculation is *memory intensive*;

OR The EM algorithm may be used as described above, which results in a very slowly convergent iterative process, and hence is *processor intensive*.

The true advantage of the EM algorithm only becomes apparent when the expected results are influenced by pre-existing concepts of their pattern.

3.5 THE EM ALGORITHM WITH PENALISED LIKELIHOOD

There are likely to be many different solutions in the region close to the optimum defined by a maximum-likelihood procedure. The solution defined as strictly optimal is unlikely to conform to our prejudices, so preferred solutions are selected by introducing a penalty which favours restorations felt to be more appropriate to the problem.

In the case of the susceptibility, restorations are expected to show features with clearly defined boundaries, and a smooth variation of intensity elsewhere. The EM algorithm can be modified to take account of such a penalty, but strictly leads to a set of non-linear simultaneous equations. A linear approximation may be obtained, however, by replacing the value \hat{x}_j^{new} in the penalty term by the value \hat{x}_j^{old} obtained from the last iteration step. This is known as the OSL approximation (one step late), and has been shown to be valid provided that convergence to the required solution is reasonably slow (Green 1990).

The E step remains the same as in the previous subsection, but a penalised likelihood is introduced into the M step. On combining the two steps, the OSL approximation gives the following iterative formula for the estimated susceptibility:

$$\hat{x}_j^{new} = \hat{x}_j^{old} + \frac{1}{m \sum_{i=1}^n h_{ij}^2} \left[\sum_{i=1}^n (y_i - \mu_i) h_{ij} - \beta \sigma^2 \frac{\partial \phi}{\partial x_{ij}} \Big|_{x_j^{old}} \right]$$

The function ϕ , differentiated in the right-hand term, defines the nature of the penalty and is often referred to as the *potential function*; apart from this term, the equation is identical to the pure EM algorithm. The value σ^2 is the assumed variance of noise in the data and the coefficient β defines the strength of the penalty. The larger the value of $\beta \sigma^2$, the more likely is the restoration to conform to our prejudices, at the expense of goodness of fit to the data.

One significant advantage of the EM algorithm with the OSL approximation is that it is possible to introduce a penalty without greatly increasing the computational expense. This is not the case with other methods known to us.

4 Results from Stage I

The calculations described in the last section were applied to two sets of *phantom* susceptibility data (fig. 2). The observed data are shown as a solid line, and the underlying 'truth' as a dashed line; the 'truth' line is repeated in subsequent figures.

The Wiener filter was first applied to both sets of data (fig. 3); the solid line represents the answer returned by the calculation. When distinct blocks of susceptibility are well separated their locations are predicted quite well, but their shapes are entirely wrong, since they are quite smooth and contain no sharp edges. There is no meaningful information to enable us to separate the three adjacent blocks in the upper diagram. The side-lobes visible at the edges of the main peaks are a characteristic feature of Fourier analysis. The methods based on maximum-likelihood procedures, including the simple EM algorithm, gave such similar results to the Wiener filter, that we have not illustrated them here.

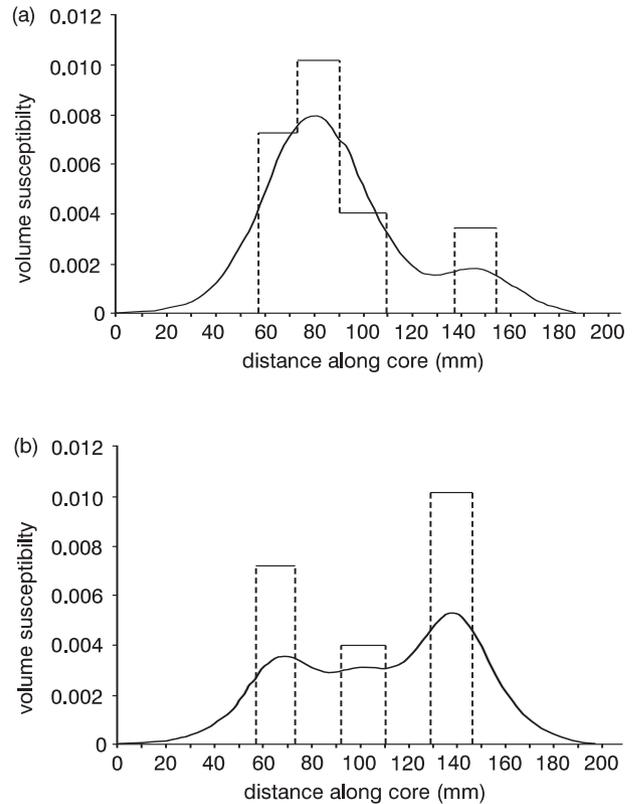


Figure 2. Two sets of 'phantom' susceptibility data (solid lines), measured from synthetic cores constructed from material of known susceptibility (dotted lines).

Using the EM-OSL algorithm for penalised likelihood estimation, we experimented with several different types of potential, of which two were found to give good results. Following the suggestion of Besag (1989), our first choice for the potential function ϕ was the absolute difference between the values of x_j and its neighbouring elements, so that the penalty was the sum of such differences. A suitable value for the constant $\beta \sigma^2$ was determined by experimental investigation; the final choice was made on the basis of a subjective balance between goodness of fit and the anticipated form of the results (fig. 4).

The separated blocks are more clearly defined than they were with the Fourier and maximum likelihood methods, but there is some filling of the intervals between them. The triple block is not resolved and has the appearance of a broad single block. All the blocks have sloping sides rather than the sharply defined vertical edges shown in the 'truth', indicating that they are made up of a succession of small steps rather than one large step.

In an attempt to eliminate this last problem, we introduced a second potential ϕ which incorporates a cut

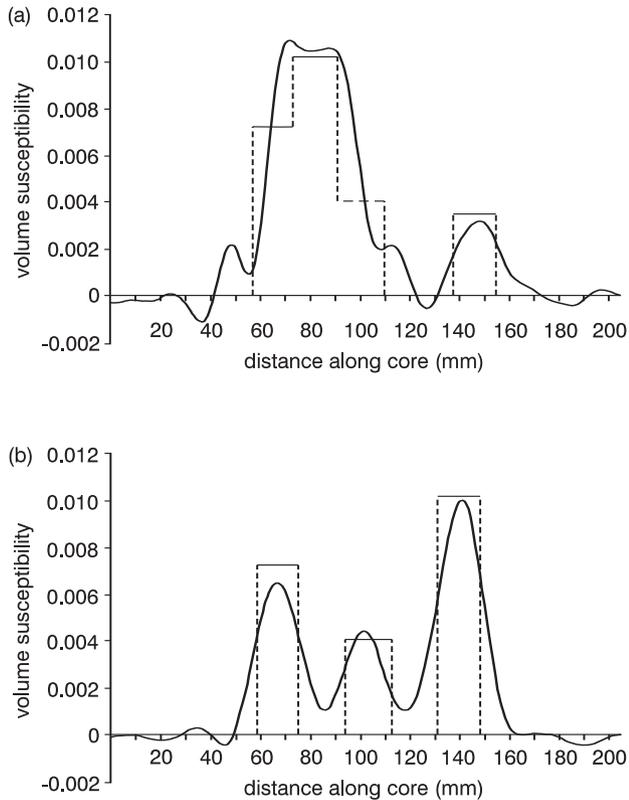


Figure 3. Restored susceptibility (solid lines) from the Fourier method with Wiener filter applied to the data of figure 2.

off, so that a single large step is penalised less heavily than an equivalent series of small steps. In this case, it was necessary to find suitable values not only for the constant $\beta\sigma^2$ discussed above, but also for the parameter defining the cut-off. The blocks are now very sharply defined, with nearly vertical sides (fig. 5). The positions and widths of the blocks are largely coincident with the ‘truth’, but there is still some filling-in of the intervals between the blocks, and smaller blocks seem to appear on the edges of the main blocks. The triple block is now partially resolved (fig. 5a), but has not been fully accounted for along its left-hand edge.

Of the various methods described above, it is clear that the EM-OSL algorithm seems to produce the most realistic results, particularly when used with the potential function ϕ which incorporates a cut-off. The general agreement between the results and the assumed truth is excellent, apart from the small blocks on the side of the main blocks and the filling-in of the intervals between blocks. Since the data used here are actual measurements, the precise mathematical nature of the response function is uncertain to an extent.

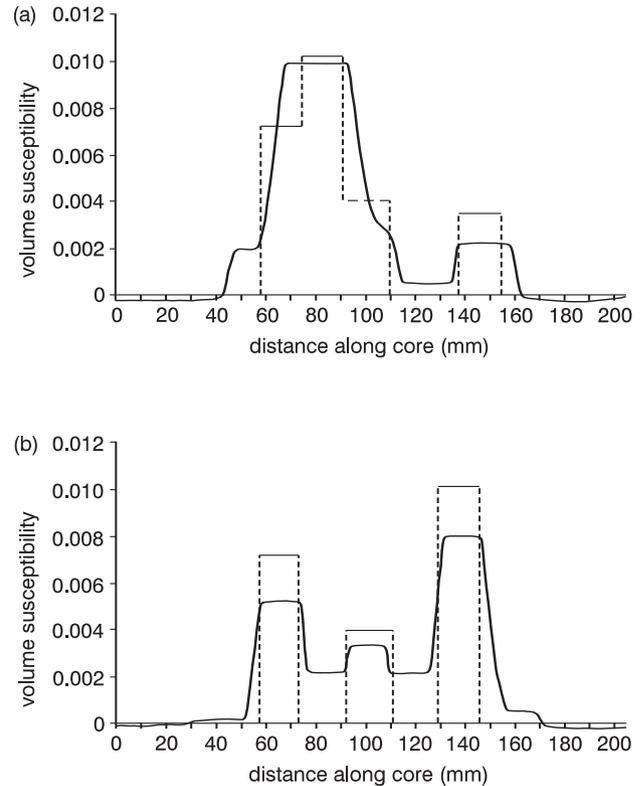


Figure 4. Restored susceptibility (solid lines) from the EM-OSL algorithm applied to the data of figure 2; the potential function is the absolute difference between neighbouring elements.

It is possible that the small additional blocks may arise from minor errors in the response function that was used in the modelling.

5 Results from Stage II

Following the successful application of the EM-OSL algorithm at Stage I, using the cut-off penalty function, it was also tested at Stage II. Because of difficulty in locating suitable field data, we decided to work entirely with simulated data, creating various models in which the magnetised features were prisms of infinite length and rectangular cross section, each located at the same depth below the soil surface.

A typical simulation from our tests comprised a large prism of low magnetic intensity juxtaposed with a smaller prism of higher intensity. This magnetic distribution was convolved with the response function (appropriate to the depth below the soil surface, the location of the model on the planet’s surface, and the strike angle between the line of survey and the line of the feature) and a reasonable measure of Gaussian noise was added to give the simulated data

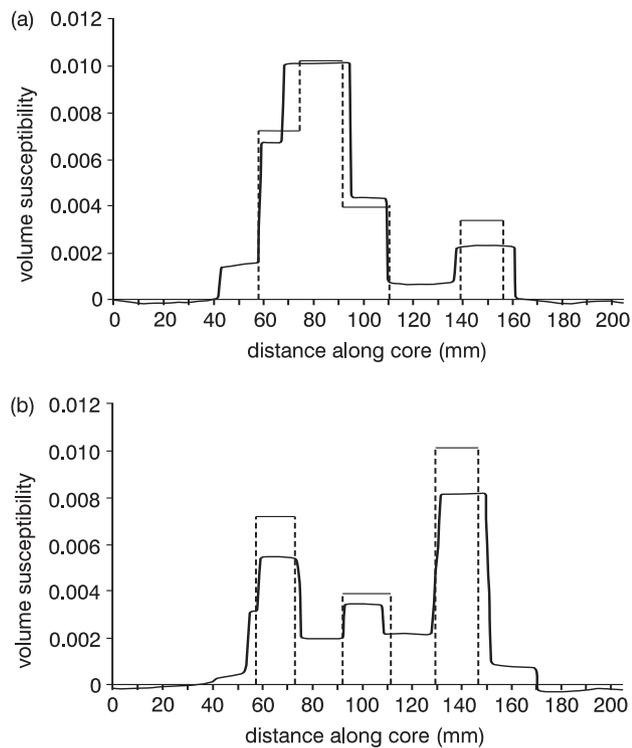


Figure 5. Restored susceptibility (solid lines) from the EM-OSL algorithm applied to the data of figure 2; the potential function incorporates a cut-off, penalising a single large step less than an equivalent series of small steps.

(fig. 6a). The result of applying the EM-OSL algorithm to the data can be seen to be in remarkable agreement with the original model (fig. 6b).

Working with simulated data, we were confident that the response function used in the restoration was precisely the same as the one used to create the data. This might be an unrealistic situation in practice, given the wide variation in form of the magnetometer response function. We repeated the restoration of the same data, but deliberately using inappropriate response functions, first a response function which assumed that the magnetic features were above their true level (fig. 6c), and then one which assumed the features were below their true level (fig. 6d). It can be seen that the location and general shape of the simulated feature are recovered reasonably well, but nowhere near as accurately as with the correct search depth. There is also a fair amount of spurious background activity.

In order to test the noise model, whose specification is somewhat problematic for magnetometry data, we repeated the whole simulation, doubling the magnitude of the noise. The signal from the feature is now substantially hidden by the noise (fig. 7a). The general shape of the feature is still recovered by the EM-OSL algorithm, but the details are not

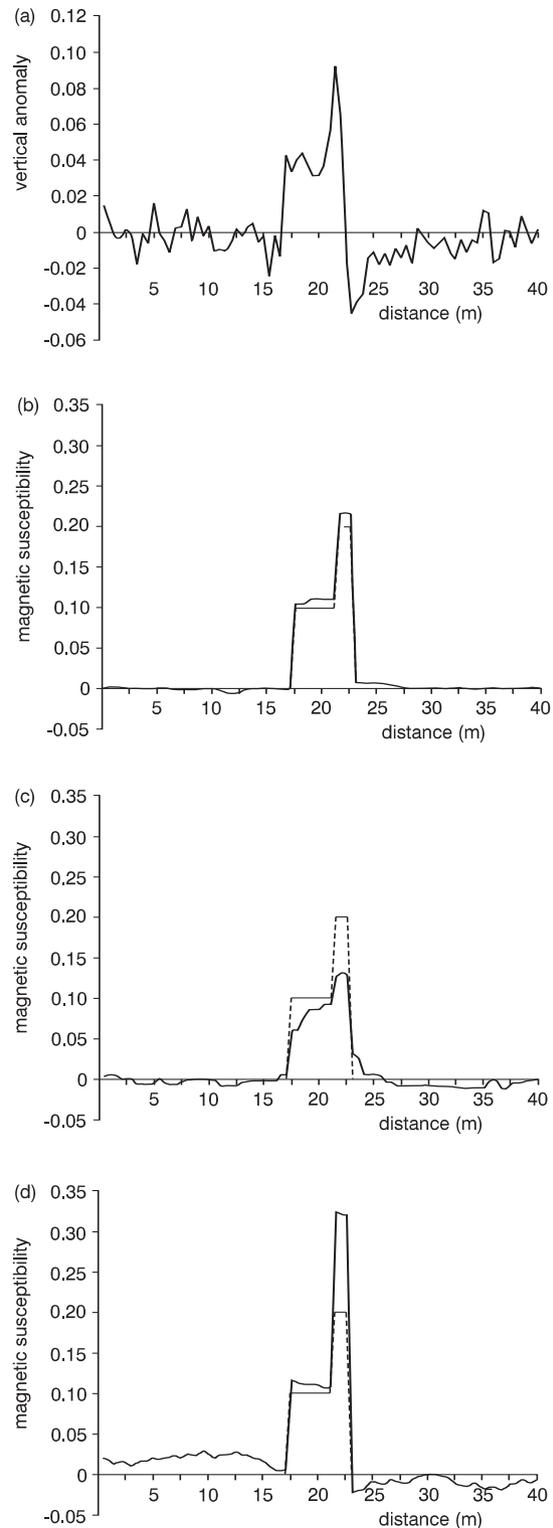


Figure 6. (a) Simulated line magnetometry data restored with EM-OSL algorithm using response functions that assume (b) the correct depth of the feature, (c) too shallow a depth for the feature, and (d) too great a depth for the feature.

as accurate (fig. 7b); there is a fair amount of spurious activity in the background. The results from the searches at the incorrect levels still give some indication of the location of the feature, but are generally inaccurate in other respects (figs 7c, 7d).

6 Questions to be answered

We have shown that the EM-OSL algorithm, maximising the penalised likelihood, produces good results at both Stage I and Stage II, with the ‘phantom’ susceptibility data and the simulated data for linear traverses. The results suggest that it is now worthwhile to set up the more complicated calculations at Stage III, so that the algorithm may be applied to data from magnetometer surveys over areas of land. Before moving to Stage III, however, a number of questions should be answered.

- How should the parameter β (or the product $\beta\sigma^2$) and the cut-off parameter in the penalty function be chosen? We have experimented with various combinations of values until finding a set which appeared to give a near optimal restoration. Although the results are relatively insensitive to the choice of these parameters, it is clear that a more objective approach is desirable.
- How can it be ensured that the response function is appropriate to the physical conditions of the survey? We have shown that the choice of response function for the simulated magnetometry data makes a considerable difference to the quality of the results. This is a critical consideration when moving into 3-dimensional modelling at Stage III, where the response functions are more complicated than those at Stage II.
- What noise model is appropriate to the data? We have experimented with simple Gaussian models at Stage II, in order to see how the amount of noise affects the quality of the results. The choice of noise model for actual magnetometer survey is problematic, since the signal from features close to the surface is often regarded as noise.
- Is it possible to predict the vertical depth of restored features as well as horizontal location? The response function from magnetised features differs with depth, as is clearly shown by our experiments in attempting to restore data using the wrong depth. The question is whether it is possible for the EM-OSL algorithm to detect the difference between the response functions sufficiently clearly to attribute a feature to the correct depth. Our experiments suggest that it may be possible, but the results are far from conclusive; further experiments are needed at Stage II, before any depth analysis is tried at Stage III.

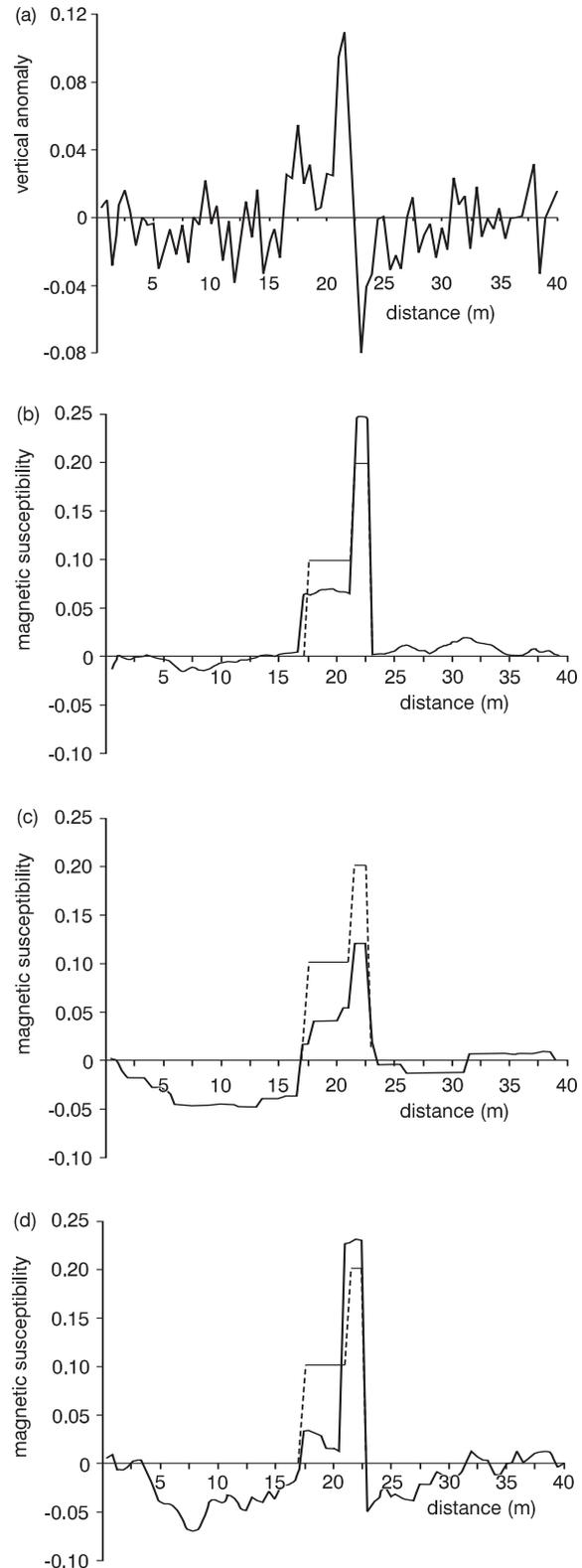


Figure 7. Similar to figure 6, except that noise of doubled intensity has been incorporated into the simulated data.

- e. Can a large data set be divided into manageable 'chunks' for calculation? Although each iterative step of the EM-OSL algorithm is calculated quite swiftly, it may still become burdensome if m and n (the number of model elements and the number of observations) are both large. Since a data set at Stage III may contain several hundred thousand readings, it would be more efficient to work with small subsets. It would then be necessary to ensure that the results at the edges of each portion matched correctly with those of neighbouring subsets.

7 Prospects for Stage III

The good progress through Stages I and II of the project has encouraged us to move on to Stage III as rapidly as possible. It is at Stage III that the project will become widely useful to archaeological geophysicists, allowing access to plenty of field data on which to test our mathematical methods. One important aspect of the usefulness of the techniques is their likely computational cost, which we now consider. With around 80 items of line data, the EM-OSL algorithm requires about 2000 iterations to converge to its final answer, taking about two minutes of processor time on a Sun 4 workstation. It is likely that similar computation

times would be achieved on personal computers equipped with the current Pentium processors.

Extrapolating to larger data sets for Stage III, we expect the computational time to be roughly proportional to the size of the data set, although allowance must be made for the more complicated response functions of the 3-dimensional model. A typical field data set of 400 readings over a square grid might take 10 minutes to process on a fast personal computer. If this estimate proves to be reasonable, then it should be possible to process data as rapidly as it can be produced from the field survey.

We conclude that the EM-OSL algorithm is capable of providing the basis for a practical method to restore magnetometry data. We are confident that satisfactory answers to the questions of the previous section will be found, allowing a useful implementation of the technique in general archaeological field survey.

Acknowledgements

We gratefully acknowledge the advice and technical cooperation of Dr Arnold Aspinall and Dr Armin Schmidt, both of the Department of Archaeological Sciences, University of Bradford; in particular, we thank them for providing the susceptibility data.

references

- | | | |
|--|------|--|
| Besag, J. | 1989 | Towards Bayesian image analysis, <i>Journal of Applied Statistics</i> 16, 395-407. |
| Dempster, A.P.
N.M. Laird
D.B. Rubin | 1977 | Maximum likelihood from incomplete data via the EM algorithm, <i>Journal of the Royal Statistical Society Series B</i> 39, 1-38. |
| Gonzalez, R.C.
R.E. Wood | 1992 | <i>Digital Image Processing</i> . Reading, Massachusetts: Addison-Wesley. |
| Green, P.J. | 1990 | Bayesian reconstruction from emission tomography data using a modified EM algorithm, <i>Institute of Electrical and Electronic Engineers Transactions on Medical Imaging</i> 9, 84-93. |
| Hastings, W.K. | 1970 | Monte Carlo sampling methods using Markov chains and their applications, <i>Biometrika</i> 57, 97-109. |
| Linington, R.E. | 1964 | The use of simplified anomalies in magnetic surveying, <i>Archaeometry</i> 7, 3-13. |
| Lucy, L.B. | 1974 | An iterative technique for the rectification of observed distributions, <i>The Astronomical Journal</i> 79, 745-765. |

- Metropolis, N.
A.W. Rosenbluth
M.N. Rosenbluth
A.H. Teller
E. Teller
- 1953 Equations of state calculations by fast computing machines, *Journal of Chemical Physics* 21, 1087-1092.
- Richardson, W.H.
- 1972 Bayesian-based iterative method of image restoration, *Journal of the Optical Society of America* 62, 55-59.
- Tsokas, G.N.
C.B. Papazachos
M.Z. Loucoyannakis
O. Karousova
- 1991 Geophysical data from archaeological sites: inversion filters based on the vertical-sided finite prism model, *Archaeometry* 33, 215-230.

Gayle T. Allum
Department of Statistics
University of Leeds
Leeds LS2 9JT
and
Department of Mathematics
University of Bradford
Bradford BD7 1DP
United Kingdom
e-mail: x11032@bradford.ac.uk

Robert G. Aykroyd
Department of Statistics
University of Leeds
Leeds LS2 9JT
United Kingdom
e-mail: robert@amsta.leeds.ac.uk

John G.B. Haigh
Department of Mathematics
University of Bradford
Bradford BD7 1DP
United Kingdom