FR TG P04

Processing of Electromagnetic Arrays with Missing Data - Source Field Structure and Source Free Transfer Functions

M.Yu. Smirnov* (Luleå University of Technology) & G.D. Egbert (Oregon State University)

SUMMARY

New approach for multivariate analysis of EM data MsDEMPCA (Missing data EM Principal Components Analysis) is applied to several large arrays: the Baltic Electromagnetic Array Research (BEAR), Electromagnetic Mini Array (EMMA), Magnetotellurics in Scandes (MASCA), IMAGE observatory data and various EarthScope MT arrays, with the goal of demonstrating how the MV approach can clarify signal and noise characteristics, improve estimates of TFs, and, perhaps most importantly, quantify biases in these estimates.

Introduction

With advances in electromagnetic (EM) geophysical instrumentation large synoptic digital EM array datasets are increasingly common, begging the development of multivariate (MV) analysis methods that can fully exploit the simultaneous character of these data. However, with many stations there will inevitably be gaps and/or heavy noise in some channels at any given time, so practical MV array methods must allow for missing data, and temporal and spatial heterogeneity of noise. We developed and validated such a scheme (MsDEMPCA, for Missing data EM Principal Components Analysis). Here we focus on applications of MsDEMPCA to several large arrays: the Baltic Electromagnetic Array Research (BEAR), Electromagnetic Mini Array (EMMA), Magnetotellurics in Scandes (MASCA), IMAGE observatory data and various EarthScope MT arrays, with the goal of demonstrating how the MV approach can clarify signal and noise characteristics, improve estimates of TFs, and, perhaps most importantly, quantify biases in these estimates. This in turn has an implication for mineral exploration, when natural source magnetotelluric method is used at high latitudes (Finnascandia, Canada).

Method

As discussed in Smirnov and Egbert (2012) one can apply MV statistical methods to EM array data having missing data to reduce bias, improve signal- to-noise ratios, and provide better control over source effects and coherent noise contamination in estimates of EM transfer functions (TFs). Furthermore, within an array framework one is not limited to the classical uniform source magnetotelluric (MT) impedance and geomagnetic vertical field TFs, but can also include inter-station horizontal magnetic TFs which can be used to map anomalous induced currents, as well as hybrid impedance tensors between electric and magnetic fields from any pair of stations. Within the framework of the MV analysis several different approaches can be taken to estimation of these TFs.

(1) The simplest approach, described in Egbert and Booker (1989), is to assume the two largest amplitude spatial modes correspond to plane wave external source, and use these to compute TFs for all components relative to any two reference channels.

(2) Considering that all modes may contain a mixture of plane wave and more spatially complex sources, one can seek linear combinations of modes for which the horizontal magnetic components most closely approximate idealized uniform linearly polarized N-S and E-W sources, and estimate TFs from these mixtures of modes.

(3) The full cross-product matrix can be computed from the robustly cleaned data projected onto the dominant spatial modes, and the usual formulae can be applied to compute weighted least squares estimates of TFs from these robust cross- products.

(4) The estimated PCA modes can be used to combine data from all sites (or from a subset) into a pair of idealized reference channels, which can be used in a conventional robust remote reference (RR) scheme, as in Egbert (1997).

In fact, all of these TF estimates can be viewed as variants on a RR scheme, with the reference signal defined as a particular linear combination of data channels Egbert (1997). As with the remote reference technique, the estimates are not biased by incoherent noise in the input channels. However, RR estimates of TFs are degraded by noise in all of h , e , and r , and while combining data from many sites may reduce noise in the reference fields r , these additional sites do nothing to reduce noise in the local electric and magnetic channels h , e . As a result only rather limited reductions in estimation errors can be achieved by using multiple sites in the processing. Improved characterization of signal and coherent noise, which can sometimes suggest data selection criteria that can lead to significant improvements in TF estimates, are generally of greater value. Another potentially significant advantage of a MV approach is in quantifying possible bias errors due to external source complications, or coherent cultural noise. These deviations from idealized plane wave sources show

up as additional coherent spatial modes, at levels well above the incoherent noise. In some cases (e.g., estimating the MT impedance in the presence of large scale external source gradients) all of these modes, and hence any mixture, result in the same TF. In others (e.g., MT impedance with data contaminated by strong cultural noise; horizontal magnetic TFs in the presence of persistent source gradients) different mixtures of modes result in large variations of TFs. By sampling a range of plausible linear combinations of the leading significant PCA modes the magnitude of possible source biases due to coherent noise can be quantified. As we will show, this bias range can be quite large, even when random error estimates based on the level of incoherent noise are small. The other extension of the method as was previously discussed by Egbert (2002) is estimation of C functions from only magnetic fields. We have successfully implemented this approach to estimate C function for MASCA array. (Figure 2.)

Figure 1 The first six spatial modes of MaSca and IMAGE arrays for the period of 224 s. Real part of horizontal magnetic field is shown. The first two modes represent mostly plain wave part, whilst the other higher order modes can be described by gradients. There is no perfect separation of different parts of the signal, meaning that for instance the first two modes clearly contain gradient part in the signal.

Figure 2 C function estimation for MASCA array compared to Schmucker's GDS estimations from european observatorie and resulting 1D model.

Conclusions

The MASCA array (Figure 1), which has a certain amount of missing data, and is located at high latitudes, where source fields are spatially complex, results in estimates of impedances which are rather stable (independent of approach, with relatively small estimated source bias) and compare well to usual multi remote reference results. This result supports the conclusion that ordinary RR impedances are not seriously affected by source field distortions, as already observed from previous analyses of BEAR array data. The situation with magnetic transfer functions is quite different, as persistent source gradients (Figure 1) now result in a very strong dependence on the TF estimation approach, and large bounds on possible source biases. Approach (2) sorting out modes based on spatial structure produces the most reasonable results for both inter-station and vertical field TFs, resulting in relatively clean maps of hypothetical events for this large array. The EarthScope arrays provide examples of very short spatial wavelength PC source contamination, and also of large-scale coherent noise, and provide examples where estimated biases even in impedances are much larger for some periods than the statistical error bars.

References

- Egbert, G.D. [1997] Robust multiple-station magnetotelluric data processing. *Geophys. J. Int.*, **130**, 475-496.
- Egbert, G.D. [2002] Processing and interpretation of electromagnetic induction array data. *Surveys in Geophysics*, **23**, 207-249.
- Egbert, G.D. and Booker, J.R. [1989] Multivariate analysis of geomagnetic array data: 1. The response space. *J. Geophys. Res.*, **94**, 14227-14247.
- Smirnov, M.Yu. & Egbert, G.D. [2012] Robust principal component analysis of electromagnetic arrays with missing data. *Geophys. J. Int.*, **190**, 1423-1438.