1. An Automated Time Domain Despiking Algorithm for MT data

Karl N. Kappler

1.1 Introduction

A method of automatic time domain spike identification and removal for magnetotelluric array data is outlined. The method is exposed using an example MT array consisting of two sites, each with orthogoanl electric dipoles, and induction coil magnetometers. This process is comprised of three steps. Spikes in the data time are first identified accoridng to an objective criteria; second, a suitable length window enclosing the spike is selected for deletion; and finally, the deleted window is replaced by a timeseries of synthetic, plausible data. The algorithm described performs these three tasks in an automated way, and results in time series of rew data, time series of despiked data, and an array of indices of windows which have been replaced. Identification is done by exploiting the similtaneous nature of geomagnetic variations at sites far (100's of km) away from one another. The selection criteria is controlled by two user defined parameters, identification thresholds, and window length. The spike replacement routine uses short Weiner filters to replace data flagged by the identifcation routine with plausible data. Summary results are presented from application of the method to magnetotelluric stations in Central California. The method sucessfully removes large spikes as well as DC offsets.

1.2 Sites and Instruments

Two ultra-low frequency electromagentic observatories are the sources of the data in the examples. For a detailed description of the sites see [*Kappler* 2004,BSL annula report]. We consider eight channels, four at Parkfield (PKD) and four at Hollister (SAO). At each site are two orthogonal electrodes and two orthogonal induction coils sampling the ambient EM field in the plance of the earth's surface.

1.3 Spike Identification

The science of magnetotellurics (MT) relies upon the recording of tiny variations in the earth's magnetic field. MT data interpretation is predicated on the assumption that the micropulsations of the earth's magnetic field are horizontally polarized, and hence spatially uniform over 100's of km (at least at mid-latitudes). One expects that, to first order, sensors at different sites ought to be strongly correlated. The magnetic fields should be nearly identical (neglecting local noise and instrument malfunctions). The electric field amplitudes and direction, however, will differ by a scale factor which depends on the local conductivity structure, but this relationship will be stationary as evidenced by the stability of TFs [*Eisel and Egbert* 2002]. The strongly correlated behaviour between magnetic channels is a property of the approximate unifirmity of the source processes over the earth [*Egbert* 1989], and the correlation of electric channels is expected from Faraday's Law which describes the voltages induced by the time varying magnetic flux seen in the magneitc channels, for example in Figure 1, which shows orthogonal electric and magnetic fields sampled at the earth's surface at two sites a distance of 120km apart. Note the similarity in field variations between channels of same orientation and field type, as well as identical scaling of the magnetic components, compared to the scale factor difference in the electric channels. This similarity holds at higher frequencies, as shown by repeatedly zooming in on the data in Figures 2, 3, and 4.

The coherence of the fields can be clearly seen to extend from hours down to seconds in period. It is this inherent similarity in the fields which is exploited to identify windows in time when the array is not functioning correctly. Magnetotelluric data is prone ot sudden sharp variations in signal of natural origin observed at a given channel. It is proposed that these varitaions can be differentiated from variations whose origins are local noise phenomena by compariing channels at different sites. A simple statistic which can be harnessed for the purpose of this differentiation is the variance of the data over small time windows. The ratio of the variance in channels which record the same field type at the same orientation should be stationary about some typical value, in the case of magentic fields, this value should be 1. In order to merge the

The choice of window length should be suffciently wide to account for possible intersite timing errors, and FIR filter noise convolved on top of spikes, but sufficiently narrow that moderate sized spikes drive the window variance well above the value it would have without a spike. Experiments have been carried out with window lengths of 256, in the case of both 1 and 40Hz data.

Always leave a blank line after a section, subsection, or subsubsection call. This is not really a problem for latex so much as for the conversion to html.

In the text, please be sure to reference figures using the ref command as illustrated here (Figure **??**). The e12 is the label of the figure and this allows auto-figure numbering. Please note, however, that the labels MUST be unique within the entire document (ie, if more than one person uses e12 as a label, there will be confusion). So please choose a unique label name.

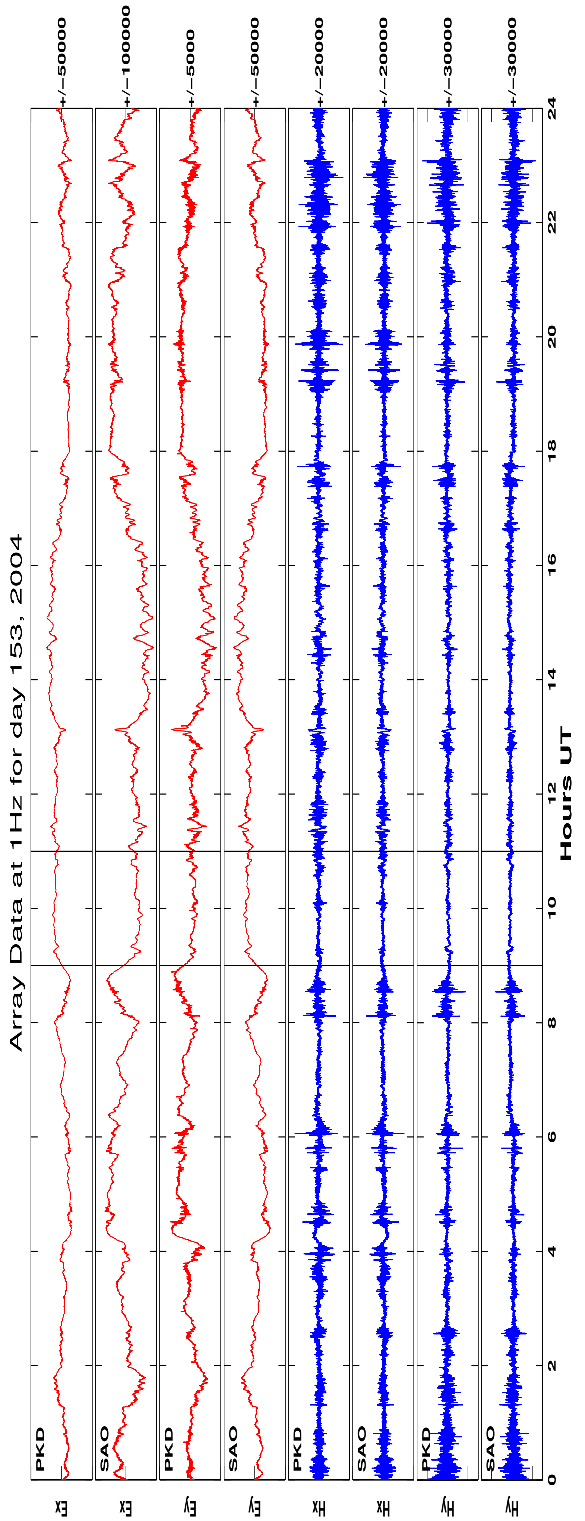

Figure 1: A plot of mean-subtracted array data for a full day in 2004. Electric fields are shown in red and magnetic fields in blue. Plots alternate between PKD and SAO at each field Polarity. Y values are in counts with axis limits shown to the left. The vertical lines mark the domain boundaries of Figure 2.

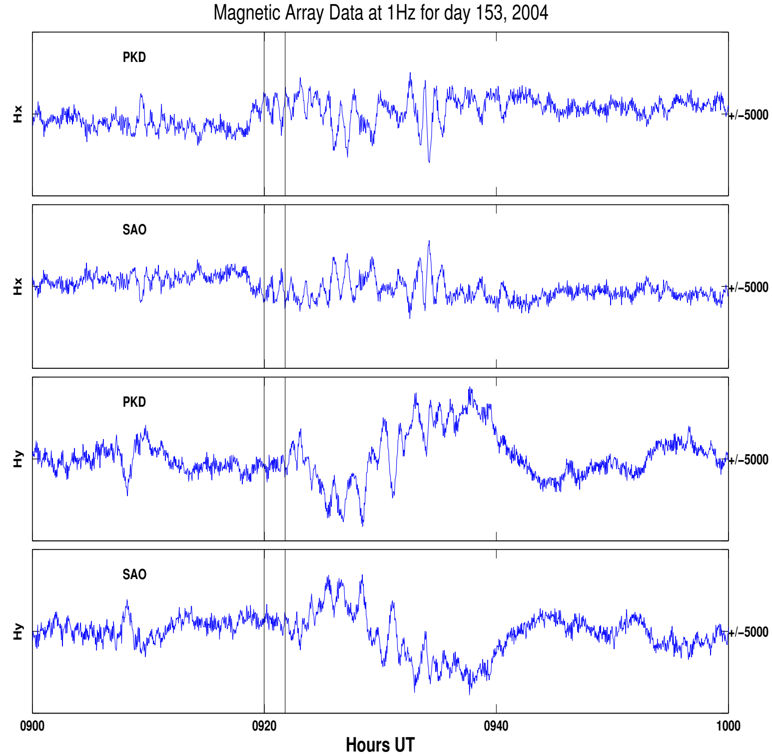

Figure 2: Mean-subtracted magnetic array data for one hour of the day shown in Figure 1. Plots alternate between PKD and SAO at each field polarity. The vertical lines mark the domain boundaries of Figure 3.

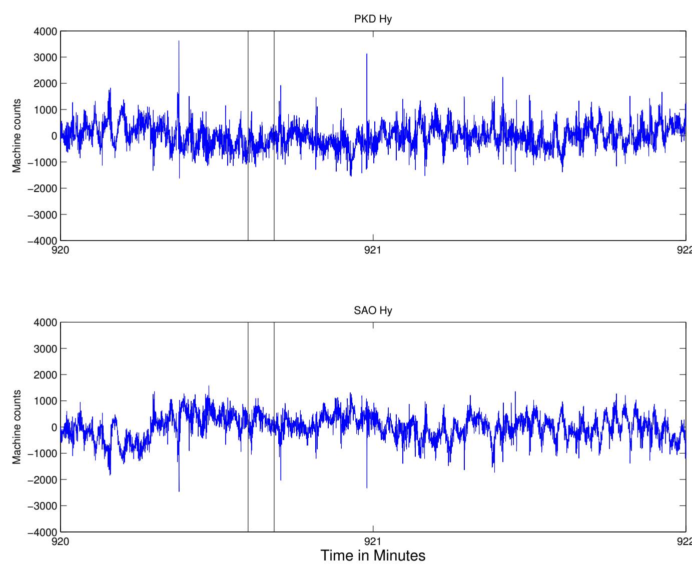

Figure 3: Mean-subtracted time series of y-polarity magnetic channels for two minutes within the hour shown in Figure 2. The vertical lines mark the domain boundaries of Figure 4

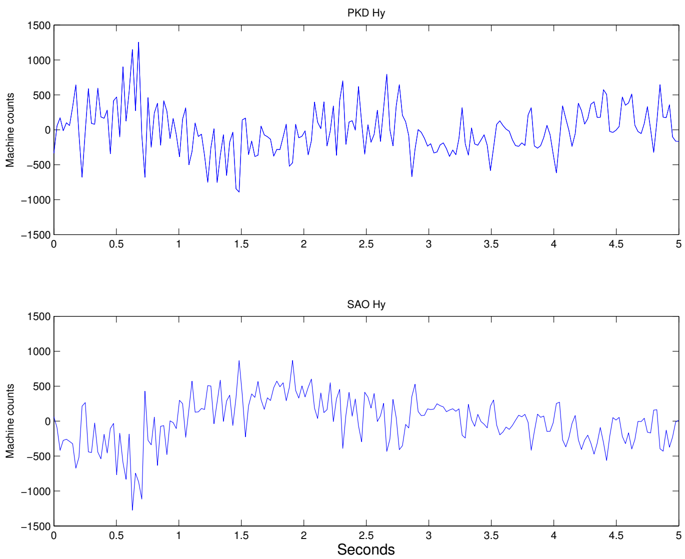

Figure 4: Mean-subtracted time series of y-polarity magnetic channels for 5 seconds within the window shown in Figure 3.

PLEASE MAKE SURE THAT THE LINES ON YOUR FIGURES ARE SUFFICIENTLY THICK. This is often a problem for reproduction.

PLEASE be sparing in your use of color. Advances in technology have not really extended to color printing and this is still very expensive. Please make every effort to produce your figure in black and white - or in colors that print adequately in black and white.

PLEASE make sure that your figures are a reasonable size for the space. If there is too much white space, use gv to check the bounding box. The first two parameters are the x and y values for the lower left corner of the figure, and the two following, the x and y values for the upper right corner. You can edit the postscript file and make changes to the Bounding Box. You can also specify the bounding box in the epsfig call.

When you reference other work, please do so in the following style (*Author et al.*, 1999). PLEASE be consistent. And please conform to some of the following usages: local magnitude M_L and moment magnitude M_w. To indicate the symbol "approximately", use $\sim$. To use the symbol for degrees, use $^\circ$.

Some commonly used names and how to represent the accent marks: Roland Bürgmann, Charles Mégnin.

Other usage issues: FTP, Web, Internet - all capitalized as shown here.

If you are making reference to a URL, please do so in the following style, `http://www.xxx.yyy.com/`. This ensures consistency across the report.

PLEASE check your spelling before submitting the contribution. This is not particularly easy. The Unix utility spell has problems with the LaTex commands. However, it is ok for a quick check. Dave Dolenc has suggested using "ispell -t", which does handle TEX commands and offers changes. I've tried it and think it works well.

As a general recommendation, I suggest short or brief section and subsection names.

For 1-column figures, I strongly recommend that you place them at the top of the page. It makes a cleaner break with the text. To do that, replace the [htbp] with [t!].

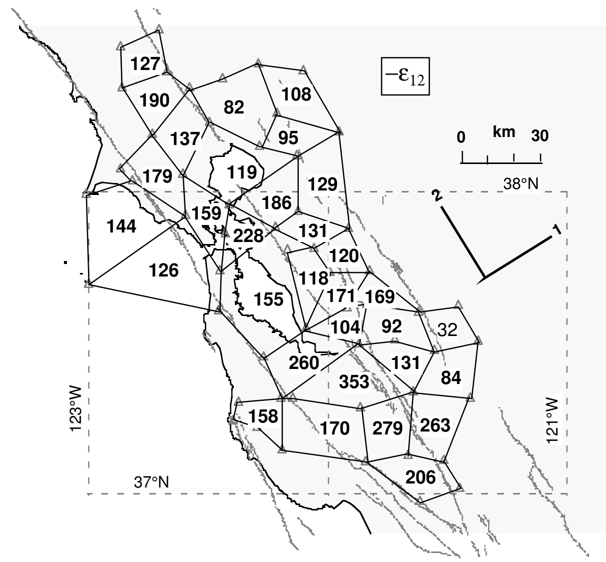

Figure 5: Enlightening figure caption

1.4 New subtopic

More brilliant exposition. Most research contributions should be limited to 2 pages. Here is another reference to a Figure 6. PLEASE make sure that you change the labels of the figures to some name of your own. Otherwise there will be problems with uniqueness.

1.5 Acknowledgements

Please don't change the spelling of Acknowledgements in the subsection header. This is the spelling we use.

Special thanks to everyone. PLEASE acknowledge your funding agency, ideally with the grant number. We send the report to a number of agencies and they always like to see themselves acknowledged!

1.6 References

Wald, D., D.V. Helmberger, and H.T. Heaton, Rupture model of the 1989 Loma Prieta earthquake from the inversion of strong-motion and broadband teleseismic data, *Bull. Seism. Soc. Am., 81,* 1540, 1991.

Bréger, L., B. Romanowicz and L. Vinnik, Test of tomographic models of D" using differential travel time data, *Geophys. Res. Lett., 25,* 5-8, 1998.

Li, X.D. and B. Romanowicz, Comparison of global waveform inversions with and without considering cross-branch modal coupling, *Geophys. J. Int., 121,* 695-709, 1995.

Li, X.D. and B. Romanowicz, Global mantle shear-velocity model developed using nonlinear asymptotic coupling theory, *Geophys. J. R. Astr. Soc., 101*, 22,245-22,272, 1996.

Nataf, H.-C. and Y. Ricard, 3SMAC: an a priori tomographic model of the upper mantle based on geophysical modeling, *Phys. Earth and Plan. Int., 95* 101-122, 1996.

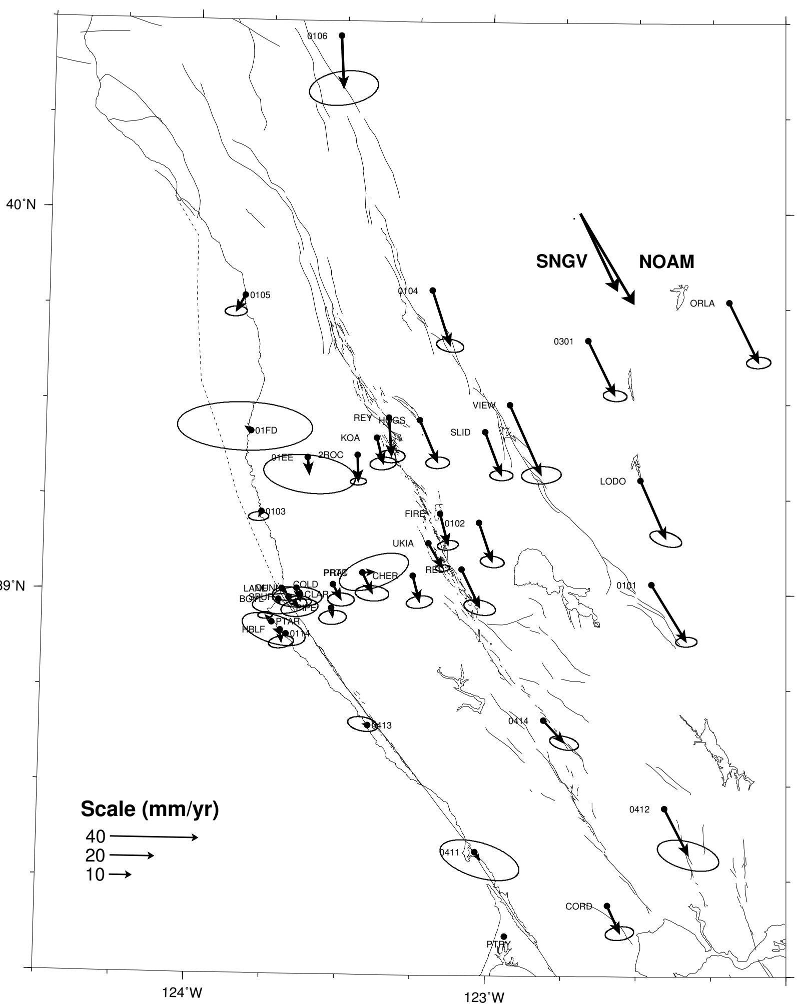

Figure 6: Another figure caption