Identification and Removal of Man-Made Transients from Geomagnetic Array Time Series: A Wavelet Transform Based Approach

Thomas T. Liu and Antony C. Fraser-Smith
Space, Telecommunications and Radioscience Laboratory
Stanford University, Stanford, California 94305, U.S.A.
e-mail: ttliu@alpha.stanford.edu

Abstract

We address the problem of identifying and removing man-made transients from geomagnetic time series acquired with an array of magnetometers. We model the transients as scaling functions of unknown scales, amplitudes, and delays. We then use an undecimated discrete wavelet transform to identify the transients in the presence of the natural $1/f$ geomagnetic background. The identification criteria incorporate the temporal and spatial characteristics of the transients. Using maximum likelihood estimates of the transient parameters, we subtract the detected scaling functions from the original time series. The efficacy of the method is demonstrated with experimental data.

1. Introduction

This paper describes our efforts to monitor ultra-low frequency (ULF; 0.01 to 10 Hz) geomagnetic fields along active earthquake faults in the San Francisco Bay Area (SFBAY) as part of an ongoing study of the relation between anomalous geomagnetic activity and the occurrence of major earthquakes [9, 11]. Our work is complicated by the presence of electromagnetic interference generated by the Bay Area Rapid Transit System (BART) [7, 9]. The disturbances are due to the magnetic fields generated by train propulsion currents and have amplitudes that are as much as 10 to 100 times greater than the natural background levels. Due to the complexity of the BART system, the exact shapes, amplitudes and times of arrival of these disturbances are quite variable and difficult to predict. Our goal is to identify and remove these man-made disturbances while minimizing the removal of both natural background signals and possible earthquake precursor signals.

We present a method that takes advantage of the temporal and spatial characteristics of the transients. We model the BART transients by scaling functions and show that a detector based upon pattern matching in the undecimated discrete wavelet transform (UDWT) domain is effective at identifying the transients in the presence of the geomagnetic background noise. The identification is further improved by determining if the polarization vector associated with each detection lies within the subset of the array signal space occupied by the BART interference. Removal of the detected transients is accomplished with a hard thresholding operation in the UDWT domain followed by an inverse wavelet transform.

2. Signal Model

The data used in our study are acquired with an array of three stations that span the SFBAY. Each station is equipped with a three-axis (vertical, geomagnetic North-South, and geomagnetic East-West) magnetometer, resulting in a total of 9 data channels. A map of the array is shown in Figure 1. We model the signal on the ith channel as

$$x_i(t) = \sum_j A_{i,j} s_{k_{i,j},l_{i,j}}(t) + v_i(t) + o_i(t) \qquad (1)$$

where $i = 1$ to 9 and $A_{i,j}$, $k_{i,j}$, and $l_{i,j}$ are unknown amplitudes, scales, and delays, respectively. We assume that $A_{i,j} \neq 0$ and $k_{i,j}$ and $l_{i,j}$ are integers. We define $s_{k,l}(t) = 2^{\gamma k/2}\xi_k(t-l)$ where $\xi_k(t) = 2^{-k/2}\xi(2^{-k}t)$ and $\xi(t)$ is the transient signal model. The term $v_i(t)$ represents the natural geomagnetic background which has an approximate $1/f^{\gamma}$ power spectrum [8], while $o_i(t)$ represents other natural signals of interest, such as possible earthquake precursors. In addition to the 9 data channels in the SFBAY, we also use 3 additional channels acquired at a *remote reference* station (henceforth referred to by its station code SAGO) located near Hollister, CA, approximately 100 km to the south of the SFBAY. Because natural ULF geomagnetic fields are highly coherent over this distance [5], the data from the remote station provide an estimate of the natural background signals $v_i(t)$. For the remote channels we

0-7803-5148-7/98/$10.00©1998 IEEE

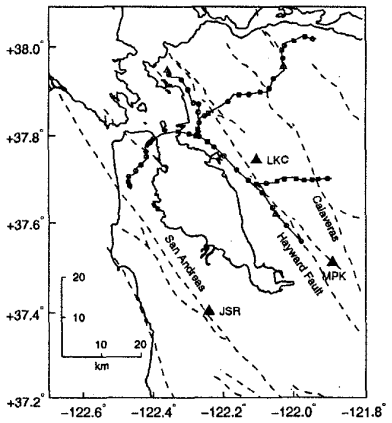

Figure 1. Map of field sites JSR, LKC, and MPK. Dashed lines are earthquake faults. Solid lines and open symbols are BART tracks and stations, respectively

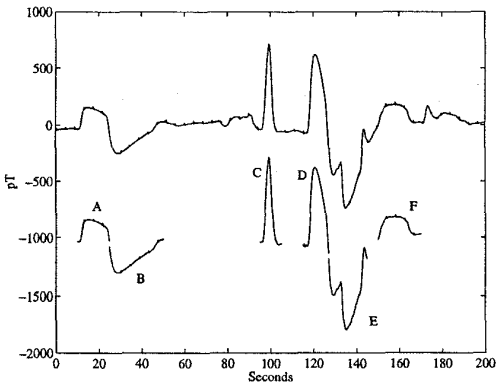

Figure 2. An example of BART interference. Significant transients are re-plotted below the original time series. The temporal widths for transients A through F are 3.1, 4.4, 0.9, 1.8, 3.6 and 3.5 seconds, respectively.

assume that $x_i(t) = v_i(t), i = 10$ to 12, i.e. BART interference and earthquake precursor signals are not present.

The signal model in (1) approximates the BART interference as the superposition of scaled and delayed versions of a single transient $\xi(t)$. The approximate nature of the signal model can be observed in Figure 2 where we provide an example of BART interference measured during a period when the natural background signals $v_i(t)$ are negligible and $o_i(t)$ is assumed to be zero. Although the transients that compose the interference have varying shapes, they can be roughly characterized as non-zero mean transients with signal energies concentrated in the main lobes. Our signal model serves as a first order approximation if we choose a $\xi(t)$ with these general characteristics. In Figure 3 we show that the Coiflet parameter 2 (henceforth denoted as C_2) scaling functions [3] and Gaussian functions serve as good models for the transient waveforms. For the remainder of this paper we choose $\xi(t)$ to be a C_2 scaling function

We also note that the temporal widths of the observed transients lie within a limited range. Using the square root of the variance of the squared modulus as defined in [1] as a measure of the temporal width σ_t of a transient, we find that the temporal widths range in value from 0.8 to 6 seconds. The time between the start and stop of a transient is approximately $5\sigma_t$. With these observations, we limit the scale parameter k to lie within a range $\mathcal{K} = \{6, 7, 8, 9\}$ in order to obtain C_2 scaling functions with the desired temporal widths.

3. Single Channel Case

In this section we consider the problem of detecting and removing transients from a single channel of the array. The detection criteria depend solely on the temporal characteristics of the transients.

3.1. Transient Detection

In order to introduce our method, we begin by examining the detection of a single transient of unknown amplitude, scale, and delay in the presence of $1/f$ noise, i.e. $x(t) = As_{k,l}(t) + v(t)$. In this case we have the standard choice between two hypotheses

$$\begin{aligned} H_0: \quad x(t) &= v(t) \\ H_1: \quad x(t) &= As_{k,l}(t) + v(t) \end{aligned}$$

where for now we assume that $A > 0$. A standard scheme for detection with unknown parameters is the generalized likelihood ratio test (GLRT) [12]. It has the form: choose H_1 if the likelihood ratio $r(x(t)) > r_1$ where

$$r(x(t)) = \frac{\max_{A,k,l} P(x(t)|A, k, l, H_1)}{P(x(t)|H_0)}$$

and r_1 is a threshold value chosen to achieve a desired probability of false alarm, and choose H_0 otherwise. Detection in $1/f$ noise is facilitated by the observation that the discrete wavelet transform (DWT) acts as an approximate whitening transform for $1/f$ processes [13]. Specifically, if $v(t)$ has a power spectrum $\frac{\sigma_v^2}{f^\gamma}$ then the DWT coefficients $V_{m,n} = \int v(t) 2^{-m/2} \psi(2^{-m}t - n)$ are approximately uncorrelated, and the variance at each analysis scale m is $\sigma_m^2 = \sigma_w^2 2^{\gamma m}$ where $\sigma_w^2 = \kappa \sigma_v^2$ and κ is a constant that depends on γ and the orthonormal wavelet $\psi(t)$. In [10] we use this observation to form a GLRT detector based upon

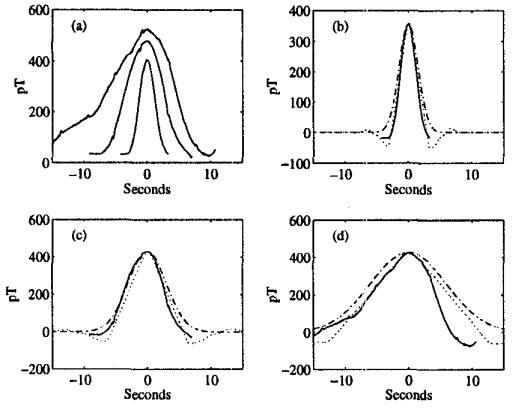

Figure 3. Panel (a): BART transients with temporal widths of 5.8, 2.9 and 0.8 seconds. Waveforms of C_2 scaling functions (dotted lines) and Gaussian functions (dash-dotted lines) are superimposed in panels (b)-(d).

pattern matching in the undecimated discrete wavelet transform (UDWT) domain. The detection statistic has the form

$$\begin{aligned}\lambda(\tilde{\mathbf{X}}) &= \max_{\{k\in\mathcal{K},l\}} \lambda(\tilde{\mathbf{X}},k,l) \\ &= \max_{\{k\in\mathcal{K},l\}} \sum_{m,n} 2^{-\gamma m}\tilde{X}_{m,2^m n+J_k+l}\tilde{S}^{\{k,0\}}_{m,2^m n+J_k}\end{aligned}$$

where $\tilde{X}_{m,n} = \int x(t)2^{-m/2}\psi(2^{-m}(t-n))dt$ is the UDWT of $x(t)$ and $\tilde{S}^{\{k,l\}}_{m,n} = \int s_{k,l}(t)2^{m/2}\psi(2^{-m}(t-n))dt$ is the UDWT of $s_{k,l}(t)$. The parameter J_k is an integer that depends on k, $\xi(t)$ and $\psi(t)$, while $\tilde{\mathbf{X}} = \left\{\tilde{X}_{m,n},\ m,n\in\mathcal{Z}\right\}$ is the vector of UDWT observations.

For the case of multiple transients we return to the signal model stated in (1). We partition each time series into a sequence of overlapping sections of size T. Within each section we find the maximum of $|\lambda(\tilde{\mathbf{X}},k,l)|$. We use the absolute value since we no longer restrict the unknown amplitude to be non-negative. We can obtain maximum likelihood (ML) estimates of A, k and l from the magnitude, scale and location of the maximum value of $|\lambda(\tilde{\mathbf{X}},k,l)|$. In practice, however, we have found that the quality of the estimates can be quite poor due to the presence of signal model mismatch errors and deviations of the actual noise process from idealized $1/f$ noise. In [10] we show that the local maxima of the UDWT provide good estimates of the delays l which maximize $|\lambda(\tilde{\mathbf{X}},k,l)|$. Applying this observation to the geomagnetic data set, we have found that delay estimates based on the local maxima of the UDWT are more robust with respect model mismatch errors. As a result, we use a modified detection statistic

$$\lambda(\tilde{\mathbf{X}}) = \max_{k\in\mathcal{K}} |\lambda(\tilde{\mathbf{X}},k,\hat{l})|$$

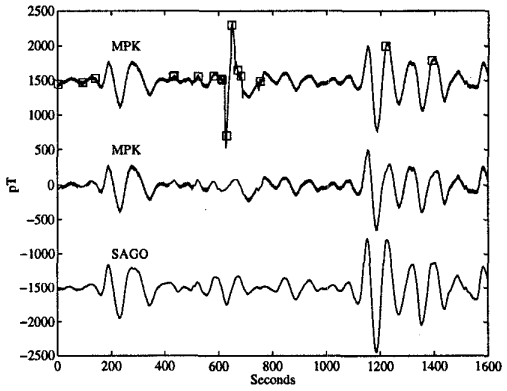

Figure 4. An isolated set of BART transients in the presence of high levels of natural activity. The top curve shows the original time series at MPK with estimated transient locations marked with squares. The middle curve shows the MPK time series after transient removal. Remote reference data from SAGO are shown in the bottom curve.

where $\hat{l} = \arg\max_n |\tilde{X}_{m,n}|$ for $m-1\in\mathcal{K}$.

To complete the detection process, we compare $\lambda(\tilde{\mathbf{X}})$ to a threshold value that depends on the noise parameter γ, the noise variance σ_w^2, the ML scale estimate $\hat{k}$, and A_{min}, the minimum amplitude that we wish to detect. For the results in this paper, we assume an A_{min} of 150 pT. In addition, the threshold is chosen to obtain a desired trade-off between the probability of false alarm (PFA) and the probability of detection (PD). In general, we set the threshold to obtain a PFA of 0.05 or less and a PD of 0.90 or greater. Estimates of γ and σ_w^2 are obtained from discrete wavelet transforms of the remote reference data. When averaged over many days, the natural geomagnetic background exhibits a $1/f$ power spectrum [8]. We have found, however, that the actual spectrum cannot, in general, be modeled as a pure $1/f$ process for periods on the order of one hour. As an approximation, we determine the parameters σ_w^2 and γ that best fit the observed spectrum. We then increase the value of σ_w^2 to obtain a spectrum that upper bounds the observed spectrum. The resultant detector design is conservative since the assumed noise levels are higher than the estimated actual levels.

In Figures 4 and 5 we provide examples of the detection method applied to time series from the North-South channel at station MPK. Remote reference data from SAGO are also presented to demonstrate the natural background activity. Figure 4 shows an isolated set of transients in the presence of elevated background signals, while Figure 5 demonstrates a more typical example in which there are numerous transients. We find that in both cases the detection scheme performs reasonably well at identifying the BART

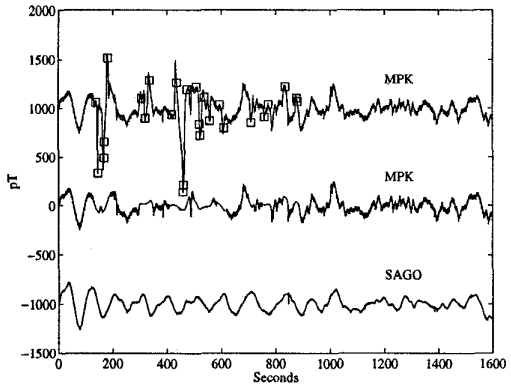

Figure 5. A more typical case in which there are many BART transients. The format is explained in Figure 4.

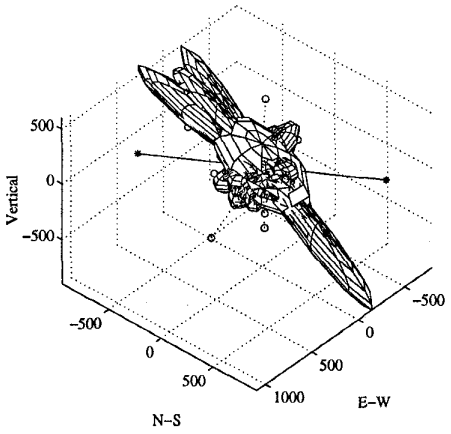

Figure 6. Polarization vectors of signals at MPK. The collection of ellipsoids define the BART transient subset at MPK. Also shown are the polarization vectors for a hypothetical earthquake related signal (stars) and for some typical natural signals (circles).

transients while discriminating against the natural geomagnetic signals.

3.2. Transient Removal

To remove the detected transients we first form estimates of the transient waveforms. In the neighborhood of each detection, we apply a hard thresholding [4] operation $H(\cdot)$ to the UDWT coefficients $\tilde{X}_{m,n}$ such that

$$H\left(\tilde{X}_{m,n}\right) = \begin{cases} \tilde{X}_{m,n}, & |\tilde{X}_{m,n}| > \sigma_m \\ 0, & |\tilde{X}_{m,n}| \leq \sigma_m \end{cases}$$

where σ_m is the standard deviation of the background noise at scale m. We next use an inverse translation invariant wavelet transform [2] to obtain an estimate of the transient signal from the thresholded coefficients. We then subtract this estimate from the original time series. Examples of the transient removal process are provided in Figures 4, 5, and 8.

4. Extension to Multiple Channels

In the single channel case, the detector may sometimes incorrectly identify natural signals as BART transients. We can reduce the number of false alarms in the detection process by using the polarization vectors of the transients to further distinguish BART related signals from naturally occurring signals.

Despite the complexity and size of the BART system, the interference as observed across the array occupies a relatively well defined subset of the total array signal space. For each observed BART transient we define a polarization vector $\vec{T}$ as the 9-vector composed of the peak amplitudes of the transient as observed across the array. We find that $\vec{T} \in S$ where $S \subset \mathcal{R}^9$ is the well bounded subset. This observation is further confirmed by models of the generation and propagation of the BART magnetic fields. We have found it convenient to define the subset S as the union of three disjoint subsets S_1, S_2 and S_3, where each S_l can be considered as a description of the three-dimensional polarization of the interference at the lth station. While the resultant S is not the smallest possible subset that describes the interference, it has the advantage of being easy to visualize. At each station we construct the subset S_l as the union of overlapping ellipsoids.

In Figure 6 we present an example of the subset S_2 that describes the polarization at station MPK. We also plot polarization vectors of observed natural signals and of a hypothetical earthquake related source. The source is a horizontal electric dipole located at depth of 10 km along the southern portion of the Hayward Fault with a dipole moment of 5×10^6 A-m, a value consistent with previous observations [6]. We observe that both the natural and hypothetical signals are readily distinguished from the BART subset.

We use the subset S to decide if a transient detected on channel i was produced by BART. To do this we first associate detections from different channels with the same underlying transient when the absolute difference in the estimated delays $|\hat{l}_{i,j} - \hat{l}_{i',j'}|$ is less than a specified width δ. We then construct the nine-dimensional polarization vector $\vec{P}$ from the amplitudes of the waveforms evaluated at the estimated delays. If we find that $\vec{P} \in S$, we consider all the associated detections to be BART related.

To demonstrate the above approach we add transient signals with the polarization of the hypothetical earthquake

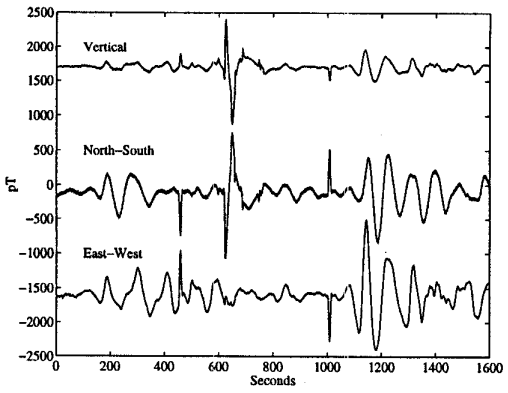

Figure 7. MPK time series with hypothetical precursor transient signals added at 450 seconds and 1000 seconds.

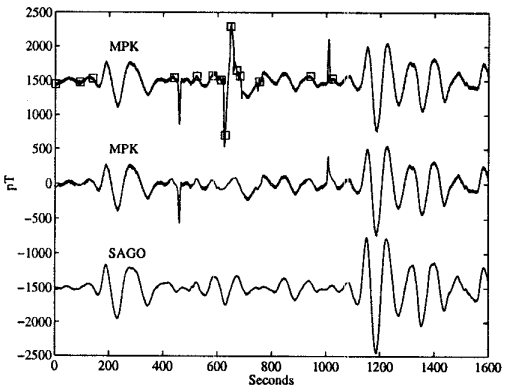

Figure 8. The results of applying the detection method with spatial constraints. The North-South channel data are shown for MPK and SAGO.

source to time series data from MPK. In Figure 7 we show the resultant waveforms for the 3 channels at MPK. In Figure 8 we show the outcome of the detection method as applied to the North-South channel. As a result of incorporating the spatial constraints, the method correctly determines that the hypothetical transient signals are not BART related. In addition, comparison with Figure 4 shows that the method is also better at distinguishing natural background signals from BART generated signals.

5. Conclusion

We have presented a wavelet transform based approach for the identification and removal of transients generated by the BART system. By incorporating the temporal and spatial characteristics of the interference, the technique detects and removes the transients while largely preserving natural signals of interest. Our results show that the monitoring of natural geomagnetic signals is possible even in regions of severe man-made interference.

6. Acknowledgements

Partial support for this work was provided by the U.S. Geological Survey through Grants no. 1434-HQ-97-GR-03124 and 1434-HQ-96-GR-02715. Additional funding for the first author was provided by the generous support of a Gerald J. Lieberman Fellowship. We thank D. Karakelian for use of the data set from station JSR and the Northern California Earthquake Data Center and the Seismological Laboratory at UC Berkeley for use of data from the remote reference station SAGO.

References

[1] R. Bracewell. *The Fourier Transform and Its Applications.* McGraw-Hill, 1986.

[2] R. Coifman and D. Donoho. Translation-invariant denoising. In A. Antoniadis and G. Oppenheim, editors, *Wavelets and Statistics.* Springer-Verlag, 1995.

[3] I. Daubechies. *Ten Lectures on Wavelets.* SIAM, 1992.

[4] D. Donoho and I. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

[5] G. Egbert and J. Booker. Multivariate analysis of geomagnetic array data 1. The response space. *J. Geophys. Res.*, 94(B10):14227–14247, Oct. 10, 1989.

[6] A. Fraser-Smith, A. Bernardi, R. Helliwell, P. McGill, and O. Villard, Jr. Analysis of low-frequency-electromagnetic-field measurements near the epicenter. In *The Loma Prieta, California earthquake of October 17, 1989 – preseismic observations*, USGS Professional Paper 1550-C, pages 17–26. U.S. Gov. Printing Office, 1993.

[7] A. Fraser-Smith and D. Coates. Large amplitude ULF electromagnetic fields from BART. *Radio Science*, 13(4):661–668, July-August 1978.

[8] L. Lanzerotti, C. Maclennan, and A. Fraser-Smith. Background magnetic spectra: 10^{-5} to 10^{5} Hz. *Geophys. Res. Letters*, 17(10):1593–1596, Sept. 1990.

[9] T. Liu and A. Fraser-Smith. Hayward fault earthquake prediction project: ULF magnetic field measurements. Technical Report WO8035-02, Electric Power Research Institute, 1997. 57 pages.

[10] T. Liu and A. Fraser-Smith. An undecimated wavelet transform based detector for transients in $1/f$ noise. Submitted to ICASSP-99, 1998.

[11] S. Park, M. Johnston, T. Madden, F. Morgan, and H. Morrison. Precursors to earthquakes in the ULF band: A review of observations and mechanisms. *Rev. Geophys.*, 31(2):117–132, May 1993.

[12] A. Whalen. *Detection of Signals in Noise.* Academic Press, 1971.

[13] G. Wornell. *Signal Processing with Fractals: A Wavelet-Based Approach.* Prentice Hall, 1996.