Denoising of coherent noises to detect transient behavior in geophysical data

Ashutosh Chamoli, Department of Earth Science, IIT Roorkee, Uttarakhand, India-247667
Nikhil Singh, Department of Earth Science, IIT Roorkee, Uttarakhand, India-247667*

Email: *martinlehora@gmail.com(presenting author)

Keywords

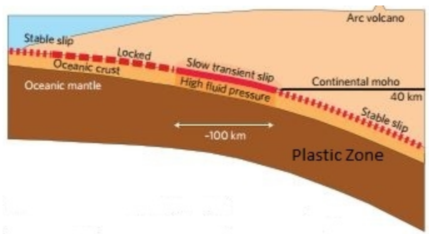
Denoising; Coherent noise; Time series; Transient behavior

Summary

Coherent noises exhibit the similar behavior as the signal in the time series. The problem is seen particularly for transient signals in geophysical time series. In this present study a hybrid approach based on Wavelet theory and Principal Component Analysis(PCA) theory is proposed for denoising the coherent noise in geophysical time series. The application is shown for a GPS time series to enhance slow slip earthquake events of Cascadia region in north west of North America. Exploitation of wavelet based denoising techniques such as soft or hard thresholding and then PCA has a pronounced effect on denoising of time series.

Introduction

The removal of coherent noise from a geophysical data in a continuously operating network is a challenging problem. The problem can affect any geophysical recording network such as Global Positioning System(GPS) time series, seismological etc. GPS stations in a continuously operating network measure surface motion and produce time series of position changes in each coordinate direction(north, east and vertical usually) relative to a reference frame(Ji & Herring 2013). Coherent noises are temporally correlated noise in phase with the signal and in case of small amplitude tectonic processes such as Silent Slip or Slow Slip Events(SSE) these noise waves arrive at the same time as the signals and thus, difficult to separate out from the noise present in the raw data such as identification of transients in GPS time series. SSE involve transient aseismic slip(measurable surface displacement along a fault in the absence of notable earthquake) across a fault at rates intermediate between plate boundary slip rates and those required to generate seismic waves(Hamling et. al. 2015). Figure 1 depicts different zones of subduction plate interface with variation in depth and pressure. More than 40 km depth rocks are heated and softened and the plates slide past each other producing stable slip unlike near the surface there is a lot of friction between the moving plates building up tension in the locked zone. SSE occurs in the transition zone as a result of slow transition slip where the rocks are partly stuck and partly melted producing slow rather than



fast release of energy, though still enough to deform the surface.

Figure 1: Diagram showing different zones of a subduction plate boundary where oceanic plate descends beneath the continental plate as locked, transient at a depth of 40 km and plastic zone.(Peng et. al., 2010)

Apart from the presence of coherent noise in the GPS time series it is also time consuming to inspect a large GPS time series without priori information of the transient signals. As well as misidentification of noise as a signal and vice-versa will always exist with a non-zero probability.
The methodology addresses above mentioned problems associated with detection of transient signals in GPS time series and it incorporates denoising properties of wavelet and PCA theory. The wavelet and PCA based approach attempts good resolution in time , frequency and space domain. PCA enhances the SNR in space domain by considering the spatial coherence of transient signals.

Method

The present methodology involves denoising behavior of wavelet transform and PCA theory. Wavelets are functions which result in values different than zero in a relatively short interval. Conditions for a real function ψ of real variables u to be a wavelet are:-

$$\int_{-\infty}^{+\infty} \psi(u)du = 0 \qquad \text{----(1)}$$

and

$$\int_{-\infty}^{+\infty} \psi^2(u)du = 1 \qquad \text{----(2)}$$

Wavelet transform of a function f(t) can be obtained using the integral transform :

$$Wf(a,t) = \int_{-\infty}^{+\infty} f(u)\cdot\overline{\;}_{a,t}(u)du \quad \lambda>0 \qquad \text{----(3)}$$

where

$$\psi_{\lambda,t}(u)=\frac{1}{\sqrt{\lambda}}\,\psi\left[\frac{u-t}{\lambda}\right] \qquad \text{----(4)}$$

represents a family of functions called wavelets defined in equation 1 and 2. Here λ is the scale parameter, t is the location parameter and $\overline{\psi_{\lambda,t}}$ is the complex conjugate of $\psi_{\lambda,t}$. Variation in λ has effects on wavelet function such as for λ>1(dilation) and λ<1(contradiction). Changing t has effects of analyzing the function f(t) around different points at t. Wavelet transform offer a powerful tool for exploratory data analysis (EDA) through which we can get a preliminary indication about the characteristics of the data and the nature of further analysis it calls for(Kumar and Foufoula-Georgiou, 1997). It gives us a basic idea about nonstationarity and the dominant scales of variations present in the series. PCA is a statistical procedure which allows us to identify the principal direction in which data varies.

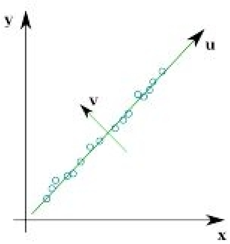


Figure 2: Dimensionality reduction using PCA.(IDAPI Lecture 15)

In the above figure PCA demonstrates the way of reducing the dimensionality of data set Linearly related data variables in X-Y axis system has been shown. The principal direction in which data varies is in U-V axis system where U-axis is the primary direction of variation and V-axis is the secondary direction of variation. V-coordinates of data are very close to zero and can be assumed that it is non zero because of noise in the data. Thus in the U-V axis system we can represent the data set by only one variable U discarding V. And in this way we reduced the dimensionality of the data set by 1. There are some examples where this property of PCA has been widely used for dimensionality reduction such as multichannel data processing(Aminghafari et. al. 2006) and detection of transient signals(Ji and Herring 2013).

In practice discrete wavelet transform is used for transform of discrete signals to discrete coefficients in wavelet domain. Low pass(LP) and High pass(HP) filters are used to get the approximation(A) and detail(D) respectively. Since approximation coefficients are low frequency terms therefore mostly it contains important part of signals and less affected by noise unlike detail coefficients. Denoising of signal is performed by comparing the detail coefficients with a threshold in order to consider whether it constitutes a desirable part of signal or not is known as thresholding. Different thresholding strategies exist such as soft, hard or universal thresholding. Soft thresholding is used for smoothing of the signal whereas hard thresholding gives better edge preservation. Universal thresholding is a fast, easy and automatic thresholding and universal threshold is estimated on the basis of data behavior.

PCA requires the concept of eigenvalues and eigenvectors for the calculation of principal components(PCs). The PCs can be calculated using the below equation:

$$X=ZA \qquad \text{----(5)}$$

where Z be a data matrix of n*p and A is a p*p orthonormal eigenvector matrix can be obtained from sample covariance matrix $S=Z^{T}Z/(n-1)$ of p*p using equation $S=ALA^{T}$ and L is a p*p diagonal eigenvalue matrix(Ji and Herring 2013). Above discussed properties of wavelet and PCA has been formulated and shown in figure 3.

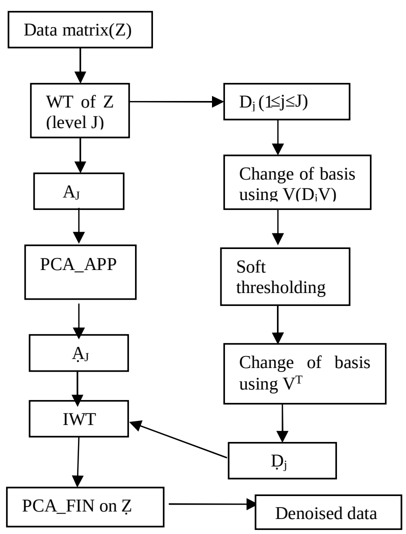


Figure 3:Flow chart for denoising procedure. WT stands for wavelet transform; D stands for detail; A stands for approximation; Ḍ stands for denoised detail; Ạ stands for approximation after PCA; PCA_APP stands for PCA on approximation; IWT stands for Inverse wavelet transform; Ẓ stands for denoised after IWT; PCA_FIN stands for final PCA on Ẓ .

More precisely, the general steps for proposed denoising method is as follows:

- Step 1: Perform the wavelet transform on each column of GPS time series at level J .
- Step 2: Define $\sum_e$ the estimator of the noise covariance matrix and then compute V such that $\sum_e = VLV^T$ where $L=diag(\lambda_{i,}\ 1\le i\le no$ of components of a GPS station). Apply to each detail coefficient(D) after change of basis (namely D_jV , $1\le j\le J$) the soft thresholding on each column of D_jV.
- Step 3: Perform the PCA on the approximation coefficient(A) and select appropriate number of principal components.
- Step 4: Reconstruct a denoised matrix of GPS data from the simplified detail and approximation coefficient by changing of basis using V^T and inverting the wavelet transform.
- Step 5:Perform a final PCA of the denoised matrix obtained at step 4 and select principal components.

  The summary of the above steps is that the method for denoising considers thresholding strategy involving change of basis for the detail coefficients after wavelet transform and then PCA is performed for the selection of convenient number of components for the approximation coefficients. Performing PCA after wavelet transform has a pronounced affect on denoising as it leads to the dimensionality reduction which serves as noise reduction. (Aminghafari et. al. 2006)

Examples

Testing of method has been done on both synthetic as well as real example. Synthetic case is generated by tanh analysis where as the real example is of Cascadia Subduction Zone(CSZ), North western region of North America.

1. Synthetic Data

In the present paper we test the method on the synthetic GPS time series of duration 12 years using hyperbolic tangent function analysis (HTFA), which is an efficient method and applied to GPS deformation studies in northern Basin and Range(Chamoli et. al. 2014) and Guerrero, Mexico(Larson et. al. 2004). Anomalous displacements during slow slip events are estimated by fitting the GPS coordinate time series with a function of the form:

$$\vec{x}(t) = \vec{x_0} + \vec{v}t + \sum_{i=1}^{n} \frac{\vec{U_i}}{2}\left[\tanh\left(\frac{t-T_{oi}}{\tau_i}\right) - 1\right] + \vec{x_1}(t) \quad .....(6)$$

where $\vec{x}(t)$ represents three-component GPS site coordinates at time t, $\vec{x_0}(t)$ represents coordinates at reference time, $\vec{v}$ represents background or "steady state" velocity, $\overrightarrow{U_i}$ represents anomalous displacement during the ith of n slow slip events, T_{0i} is the median time of the ith event, τ_i scales the period over which the event occurred, $\vec{x_1}(t)$ is displacement due to non-tectonic phenomena which can be calculated from the formula :

$$\overrightarrow{x_1}(t) = p1\cos 2\pi t + p1\sin 2\pi t + p2\cos 4\pi t + p2\sin 4\pi t \quad ---(7)$$

where p1 is the annual seasonal component and p2 is the semi-annual seasonal component. (Chamoli et. al. 2014). Table 1 shows the parameters used to generate the GPS time series shown in figure 4.

Table 1: Values of parameters involved in the generation of synthetic GPS time series from HTFA.

Stations	p1	p2	SNR	τ	U	i	RMS error
1	0.7	0.7	2	0.2	-1,1,-1	3	0.1676
2	0.9	0.9	2	0.2	-1,1,-1	3	0.1965
3	0.1	0.1	2	0.2	-1,1,-1	3	0.1499
4	0.7	0.7	2	0.2	-1.5,1.5,-1.5	3	0.1369

Figure 4 shows the synthetic GPS time series generated from tanh method using equation 6 and 7 and values from the table 1 .

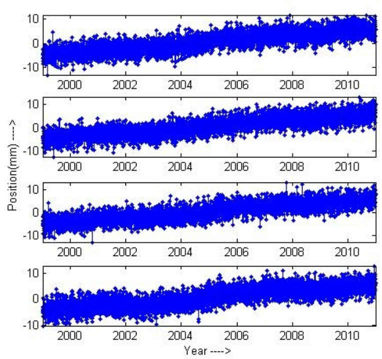


Figure 4: Synthetic GPS time series using HTFA (Station 1,2,3,4 from top to bottom).

Formatted: Left, Indent: Left: 0", First line: 0"

GPS time series in Figure 5 shows the denoised one where blue lines represent the denoised GPS time series and red lines represent the ideal transient behavior modeled using hyperbolic tangent function analysis(HTFA) and the slast column of table 1 gives root mean square(RMS) error which is basically the sum of square of difference between the denoised GPS time series and the modeled one. The RMS error shows the efficiency of the proposed method.

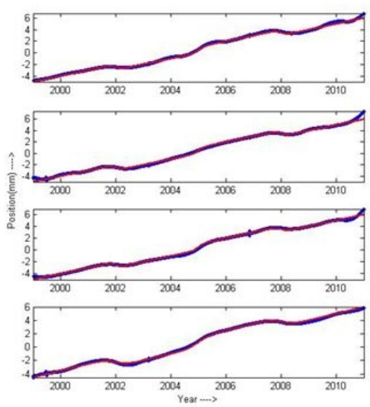


Figure 5: Denoised synthetic GPS time series.

2. Real Data

The method is tested for the ALBH(48° 23' 23.211" -123° 29' 14.892")GPS site which form a part of the PANGA (Pacific Northwest Geodetic Array) in the Cascadia region

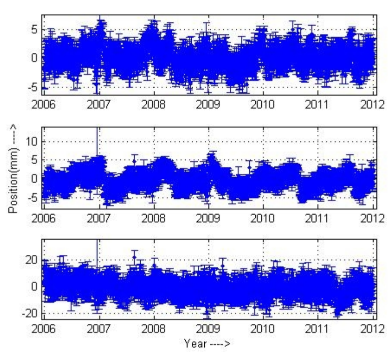


Figure 6: GPS time series of ALBH station(north, east, vertical from top to bottom) from PANGA.

where the oceanic Juan de Fuca plate descends beneath the North American Plate. GPS times series of duration 6 years is used here for testing of the method. Figure 6 represents GPS time series of ALBH station whereas Figure 7 represents denoised one showing transient behavior. Events marked by green lines in figure 8 are showing prominent signatures for quick identification after application of the proposed methodology in figure 7.

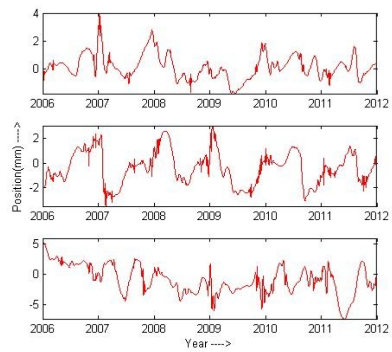


Figure 7: Denoised GPS time series of ALBH station(north, east, vertical from top to bottom) in Pacific Northwest Geodetic Array.

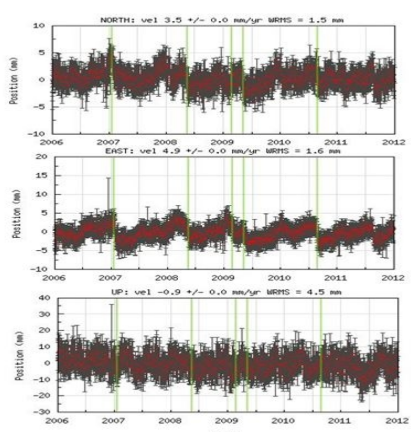


Figure 8: GPS time series of ALBH station from PANGA with green lines indicating Slow earthquakes(www.geodesy.cwu.edu)

Conclusions

A proposed methodology is proposed to denoise coherent noises from geophysical time series. The application on GPS time series has substantially improved the resolution in identification of transient signals related to Slow Slip Events(SSE). The scope of application is in network based recording and multichannel data processing such as GPS time series, multichannel seismic processing.

References

Aminghafari, M., N. Cheze, and J.M. Poggi(2005), Multivariate denoising using wavelets and principal component analysis, Elsevier, 2381-2398, doi:10.1016/j.csda.2004.12.010

Chamoli, A., A.R. Lowry, and T.N. Jeppson(2014), Implications of transient deformations in northern Basin and Range, western United States, J. Geophys. Res. Solid Earth, 119, 4393-4413, doi:10.1002/2013BJ010605

Intelligent Data Analysis and Probabilistic Inference(IDAPI) Lecture 15:DOC493

Ji, H.K., and T.A. Herring (2013), A method for detecting transient signals in GPS position time series : smoothing and principal component analysis, Geophys. J. Int., 1-16,doi:10.1093/gji/ggt03

Larson, K.M., A.R. Lowry, V. Kostoglodov, W. Hutton, O. Sanchez, K. Hudnut, and G. Suarez(2004), Crustal deformation measurements in Guerrero, Mexico, J. Geophys. Res., 109, B04409, doi:10.1029/2003JB002843

Peng, Z., and J. Gomberg (2010), An integrated perspective of the continuum between earthquakes and slow slip phenomena, Nature Geoscience, Vol 3, 599-607, doi:10.1038/NGEO940