Effect of Using Different Types of Threshold Schemes (in Wavelet Space) on Noise Reduction over GPS Times Series

K. Moghtased-Azar[a, *], M. Gholamnia [b]

[a] Department of Surveying, Faculty of Civil Engineering, 51666-16471, 29 Bahman Boulevard, Tabriz, Iran
moghtased@tabrizu.ac.ir
[b] Department of Surveying, Faculty of Civil Engineering, Zanjan University, Zanjan, Iran
mehgholamnia@gmail.com

KEY WORDS: Time series, Noise, Wavelet analysis, Threshold methods

ABSTRACT:

We applied six types of thresholding techniques in aim to impact of thresholding in denoising of time series, which were penalized threshold, Birgé-Masssart Strategy, SureShrink threshold, universal threshold, minimax threshold and Stein's unbiased risk estimate. In order to compare the effect of them in denoising of noise components (white noise, flicker noise and random walk noise) we have constructed three kinds of stochastic models: the pure white noise model (I), the white plus random walk noise model (II) and the white plus flicker noise model (III). The numerical computations are performed through the analyzing 10 years (Jan 2001 to Jan 2011) of daily GPS solutions which are selected of 264 stations of SOPAC (Scripts orbit and permanent array center). According to results of computations, among the thresholding schemes in denoising of the pure white noise model (I): minimax threshold and Stein's unbiased risk estimate could reduced the distribution of low amplitude of white noise. However, minimax threshold and SureShrink threshold could reduced the distribution of high amplitude of white noise. Birgé-Masssart Strategy and universal threshold could reduced both low and high amplitudes of white noise. In models II and III, all of threshold schemes could reduced both high and low amplitudes of white noise in same level. Whereas for power-law noise (flicker noise and random walk noise) penalized threshold and Stein's unbiased risk estimate led to reduction of low amplitudes and SureShrink threshold and minimax threshold led to reduction of colored noise with high amplitudes. Birgé-Masssart Strategy and universal threshold could reduced both low and high amplitudes of colored noise.

1. INTRODUCTION

The developments of space geodesy (i.e. GPS) allowed the establishment of world geodetic networks observing constellations of satellites permanently. Great numbers of the measurements collected by these systems permit to represent the displacement of the ground stations in terms of coordinate time series. Time series analysis is a quite recent research field in space geodesy used in order to better apprehend the temporal variability of the physical phenomena (deformations of the earth's crust, mass transfers, geodynamic local phenomena, etc).

The most recent studies are interested particularly in the signal noise separation (denoising) of the coordinates time series based on statistical and mathematical tools. In this contribution we are using wavelet - based denoising schemes for time series analysis of permanent GPS stations. The wavelet technique permits to study the signal at different resolutions to better locate the different frequencies.

The wavelet transform decomposes a signal using functions (wavelets) well localized in both physical space (time) and spectral space (frequency), generated from each other by translation and dilation, which is well suited for investigating the temporal evolution of periodic and transient signals. The wavelet analysis has influenced much research field, of which in particular, the applications for the comprehension of the geophysics process.

Appling wavelet transform on the permanent time series of GPS could separate the noise of the signal, in order to provide certain information useful to later geodynamic interpretations. However, due to the large amount of computation time and storage needed to process of wavelet transform, we are using of fast algorithm for computation the wavelet coefficients introduced by Mallat in the context of multi-resolution analysis (MRA). The multi-resolution analysis allows, by successive filtering, producing a series of signals corresponding to an increasingly fine resolution of the signal.

Thereby, signal is separated in two components: one representing the approximation of the signal (represented by its low-frequency) and the other representing its details (represented by its high-frequency). To separate both, we thus need a pair of filters: a low-pass filter to obtain the approximation, and a high-pass filter to estimate its details. In order to not lose information, these two filters must be complementary; the frequencies cut by one must be preserved by the other.

The majority of wavelet algorithms use a decimated discrete decomposition of the signal. This decomposition has the characteristic to be orthogonal and to concentrate information in some great wavelet coefficients. The denoising idea is to conserve only the greatest coefficients and put the others (corresponding to the noise) at zero before reconstruction of the signal (thresholding step). The thresholding step modifies and process all of the discrete detail coefficients at all scale so as to remove noise. We applied six types of thresholding techniques in aim to impact of thresholding in denoising of time series, which were penalized

* Corresponding author.

threshold, Birgé-Masssart Strategy, SureShrink threshold, universal threshold, minimax threshold and Stein's unbiased risk estimate.

In order to compare the effect of them in denoising of noise components (white noise, flicker noise and random walk noise) we have constructed three kinds of stochastic models: the pure white noise model (I), the white plus random walk noise model (II) and the white plus flicker noise model (III). The numerical computations are performed through the analyzing 10 years (Jan 2001 to Jan 2011) of daily GPS solutions which are selected of 264 stations of SOPAC (Scripts orbit and permanent array center).

2. SPECTRAL ANALYSIS OF GPS TIMES SERIES

2.1 Noise Analysis

The one dimensional stochastic process whose behavior in the time domain is such that its power spectrum has the form

$$P_x(f) = P_0\left(\frac{f}{f_0}\right)^{\kappa} \tag{1}$$

where f is the temporal frequency, P_0 and f_0 are normalizing constants, and κ is the spectral index. Typical spectral index values lie within [−3, 1]; for stationary processes $-1 < \kappa < 1$ and for non-stationary processes $-3 < \kappa < -1$. A smaller spectral index implies a more correlated process and more relative power at lower frequencies.

Special cases within this stochastic process occur at the integer values for κ. Classical white noise has a spectral index of 0, flicker noise has a spectral index of -1, and a random walk *noise* has a spectral index of -2 (Agnew, 1992). The power spectral method can be employed to assess the noise characteristic of GPS time series.

The second way is to use (co)variance component estimation (VCE) methods. The role of the data series covariance matrix is considered to be an important element with respect to the quality criteria of the unknown parameters. Therefore, VCE methods are of great importance.

The noise components of GPS coordinate time-series, i.e. white noise, flicker noise and random walk noise, are usually estimated by the maximum likelihood estimation MLE method which is a well-known estimation principle. According to this theory, the time series of GPS coordinates is composed of white noise, flicker noise, and random walk noise with variances σ_w^2, σ_f^2 and σ_{rw}^2, respectively.

Then, covariance matrix of the time series can then be written as:

$$C_x = \sigma_w^2 I + \sigma_f^2 C_f + \sigma_{rw}^2 C_{rw} \tag{2}$$

where I is the $m \times m$ identity matrix, and C_f and C_{rw} are the cofactor matrices relating to flicker noise and random walk noise, respectively. The structure of C_x matrix is known through matrices I, C_f and C_{rw}, but the contributions through σ_w^2, σ_f^2 and σ_{rw}^2 are unknown (Williams et al., 2004). In global GPS solutions, Williams et al. (2004) showed that a combination of white and flicker noise is appropriate for all three coordinate components (east, north and up components).

2.2 Wavelet Analysis

In wavelet domain, significant information can be extracted simultaneously in time as well as frequency domain due to time-frequency localization property of the wavelets, which makes it suitable to study the non-stationary signals. The scaling of wavelets provides powerful methods to characterize signal structures such as fractal signals, singularities etc. In numerical analysis and functional analysis, a discrete wavelet transform (DWT) is any wavelet transform for which the wavelets are discretely sampled.

As with other wavelet transforms, a key advantage it has over Fourier transforms is temporal resolution: it captures both frequency and location information (location in time). The DWT of a signal x is calculated by passing it through a series of filters. First the samples are passed through a low pass filter with impulse response g resulting in a convolution of the two:

$$y[n] = (x * g)[n] = \sum_{k=-\infty}^{\infty} x[k]\, g[n-k] \tag{3}$$

The signal is also decomposed simultaneously using a high-pass filter h. The outputs giving the detail coefficients (from the high-pass filter) and approximation coefficients (from the low-pass). It is important that the two filters are related to each other and they are known as a quadrature mirror filter. However, since half the frequencies of the signal have now been removed, half the samples can be discarded according to Nyquist's rule. The filter outputs are then sub-sampled by 2 (see Fig.1). This decomposition has halved the time resolution since only half of each filter output characterizes the signal. However, each output has half the frequency band of the input so the frequency resolution has been doubled (Mallat, 1989).

$$\begin{aligned} y_{\text{low}}[n] &= \sum_{k=-\infty}^{\infty} x[k]\, g[2n-k] \\ y_{\text{high}}[n] &= \sum_{k=-\infty}^{\infty} x[k]\, h[2n-k] \end{aligned} \tag{4}$$

This decomosition is repeated to further increase the frequency resolution and the approximation coefficients decomposed with high and low pass filters and then down-sampled. This is represented as a binary tree with nodes representing a sub-space with a different time-frequency localisation. The tree is known as a filter bank.

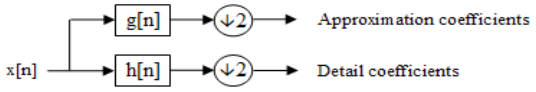

Figure 1. Block diagram of filter analysis, where operator ↓ denotes to the sub-sampling operator

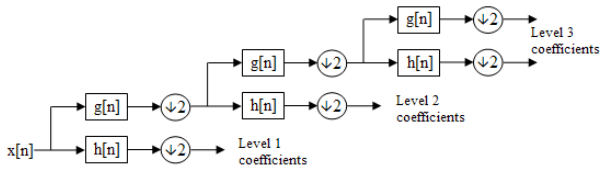

Figure 2. A three level filter bank.

2.2.1 Denoising: Assume that the observed noisy signal (y) is composed of true signal x and Gaussian white noise ε centred independent and identically distributed of variance σ^2, such as $\varepsilon \overset{iid}{\sim} N(0,\sigma^2)$. The general way to denoised is to find $\hat{x}$ such that it minimizes the mean square error of $\hat{x}$ by:

$$\mathrm{MSE}(\hat{x}) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x})^2 \tag{5}$$

Donoho and Johnstone (Donoho and Johnstone, 1994; Donoho, 1995) developed a methodology called waveShrink for estimating x. There are two commonly used shrinkage function: the hard and soft shrinkage functions:

$$T_\lambda^{\mathrm{Hard}}(d) = \begin{cases} d & \text{if } |d| > \lambda \\ 0 & \text{if } |d| \le \lambda \end{cases} \tag{6}$$

$$T_\lambda^{\mathrm{Soft}}(d) = \begin{cases} d-\lambda & \text{if } d \ge \lambda \\ d+\lambda & \text{if } d \le -\lambda \\ 0 & \text{if } |d| \le \lambda \end{cases}$$

where d is the detailed coefficient and λ is threshold value ($\lambda > 0$). Determining threshold values λ is the key issue in waveShrink denoising. In the following subsection we briefly discuss five standard methods for selecting threshold rules.

2.3.1.1 Universal Threshold: This thresholding strategy comes from Donoho-Johnstone (Donoho and Johnstone, 1993) in which is of the following form:

$$\lambda^U = \sqrt{2\log n}\ \hat{\sigma} \tag{7}$$

where n is the signal length and $\hat{\sigma}$ is the estimated noise standard deviation. A robust estimator for the estimating $\hat{\sigma}$ could be comes from

$$\hat{\sigma} = \frac{\mathrm{median}(|d_1(k)| - \mathrm{median}(|d_1(k)|)}{0.6745};\ k = 1,2,\ldots,n/2 \tag{8}$$

where d_1 is the finest level of detail coefficients.

2.3.1.2 SureShrink Threshold: Donoho and Johnstone (1994) developed a technique of selecting a threshold by minimizing Stein's unbiased estimator of risk:

$$\lambda_j^S = \arg\min_{\lambda>0}\left[\mathrm{SURE}^S(\lambda, d_j)\right] \tag{9}$$

where $\mathrm{SURE}^S(\lambda, d_j)$ is the Stein's unbiased estimator of risk of threshold function and d_j represents detail coefficients at level j of decomposed signal. For instance, application of Eq. (10) to soft threshold function ($T_\lambda^{\mathrm{Soft}}$) gives:

$$\mathrm{SURE}^S(\lambda, d_j) = N_j - 2\sum_{k=1}^{N_j}\mathbf{1}\left(|d_k| \le \lambda\right) + \sum_{k=1}^{N_j}\min\left(|d_k|, \lambda\right)^2 \tag{10}$$

where $\mathbf{1}$ denotes the indicator function and N_j is number of coefficients in level j of decomposed signal. In above equation, we assumed that $\sigma = 1$ otherwise it could be estimated using Eq. (8) and detail coefficients could be normalized using it.

However, this procedure has some drawbacks in situations of extreme sparsity of wavelet coefficients. To avoid this drawback, Donoho & Johnstone (1995) considered a hybrid scheme of the SureShrink threshold by the following heuristic idea: if the set of empirical wavelet coefficients is judged to be sparsely represented, then the hybrid scheme defaults to the level wise universal threshold

$$\lambda_j^{HS} = \begin{cases} \lambda_j^U & \text{if } \frac{1}{N_j}\sum_{j=1}^{N_j}\left(\left(\frac{d_j}{\hat{\sigma}}\right)^2 - 1\right) \le \frac{(\log_2 N_j)^{3/2}}{\sqrt{N_j}} \\ \lambda_j^S & \text{otherwise} \end{cases} \tag{11}$$

Otherwise the SURE criterion is used to select a threshold value.

2.3.1.3 Minimax Threshold: minimax threshold is one of the commonly used thresholds. The minimax threshold is defined as threshold λ which minimizes the expression

$$\inf_{\lambda}\sup_{\theta}\left\{\frac{R_\lambda(\theta)}{n^{-1} + \min(\theta^2, 1)}\right\} \tag{12}$$

where $R_\lambda(\theta) = E\left(\delta_\lambda(d) - \theta\right)^2,\ d \sim N(\theta, 1)$.

2.3.1.4 Penalized Threshold: This is level wise threshold method which is provided by Birge and Massart (2001). In this method the detail coefficients could be sort in descending order then according the following equation threshold λ could be computed:

$$\lambda = \arg_t^{\min}\left[-\sum_{k=1}^{t} d_k{}^2 + 2\sigma^2 t\left(\alpha + \ln\frac{n}{t}\right)\right]; t = 1\ldots n \qquad (13)$$

where $\alpha > 1$ is the sparsity parameter.

2.3.1.4 Birge and Massart Strategy Threshold: This is level wise threshold method which is also provided by Birge and Massart (1997). It uses level-dependent thresholds obtained by the following wavelet coefficients selection rule. Let j be the decomposition level, m be the length of coarsest approximation coefficients over 2 and $\alpha > 1$. The numbers j, m and α define the strategy: at level $j+1$ (and coarser levels), everything is kept. For level i from 1 to j, the n_j larger coefficients in absolute value are kept with:

$$n_j = \frac{m}{(j+2-i)^{\alpha}} \qquad (14)$$

2.3.2 Wavelet Reconstruction:

This part denotes how the decomposed components can be assembled back into the original signal without loss of information. This process is called reconstruction, or synthesis. The mathematical manipulation that effects synthesis is called the inverse discrete wavelet transform (IDWT). Where wavelet analysis involves filtering and down-sampling, the wavelet reconstruction process consists of up-sampling and filtering.

The reconstruction filters are designed in such a way to cancel out the effects of aliasing introduced in the wavelet decomposition phase. The reconstruction filters together with the low and high pass decomposition filters, forms a system known as quadrature mirror filters. For a multilevel analysis, the reconstruction process can itself be iterated producing successive approximations at finer resolutions and finally synthesizing the original signal.

3. NUMERICAL ANALYSIS

GPS data are collected from Scripps Orbit and Permanent Array Center (SOPAC), which include archive high-precision GPS data particularly for the monitoring of earthquake hazards, tectonic plate. Given positions by SOPAC are provided in ITRF2000, and include both horizontal and vertical velocities and their accuracies. All the chosen stations (264 permanent GPS stations) have individual and continuous solutions up to 10 years, between January 2001 and January 2011. Fig. 3 illustrates the sites of SOPAC across Western United States, Western Canada and Alaska.

In order to compare the effect of threshold methods in denoising of noise components (white noise, flicker noise and random walk noise) we have constructed three kinds of stochastic models: the pure white noise model (I), the white plus random walk noise model (II) and the white plus flicker noise model (III).

To graph the results, we used of angle histogram plot which is a polar plot showing the distribution of values grouped according to their numeric range. This type of graph shows the distribution of theta in 20 angle bins or less. The radial angle, expressed in radians, determines the angle of each bin from the origin. The length of each bin reflects the number of elements in theta that fall within a group, which ranges from 0 to the greatest number of elements deposited in any one bin (see Fig. 4).

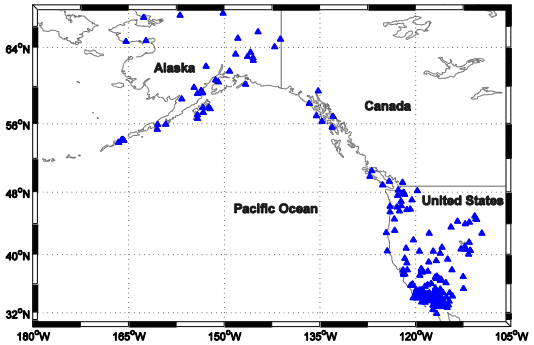

Figure 3. Location of selected sites in the study area.

First row in the illustration is related to the distributions of white noise in Up, North and East components, respectively. The second row is related to estimated white noise after denoising by penalized threshold, the third row shows the estimated white noise after denoising by Birge and Massart strategy threshold, the fourth row shows the estimated white noise after denoising by hybrid scheme of the SureShrink threshold, the fifth row is illustrated the estimated white noise after denoising by universal threshold, the sixth row shows the estimated white noise after denoising by minimax threshold and the seventh row is related to the estimated white noise after denoising by Stein unbiased risk estimates.

Fig. 4 shows the high level of white noise (in model I) in vertical component compared to horizontal components. Comparison of different type of threshold methods shows that: all methods could reduce the distribution of white noise in horizontal components in a same level. Universal threshold and Birge and Massart strategy threshold could reduce both low and high amplitudes of white noise in vertical components. However, minimax threshold and hybrid scheme of the SureShrink threshold could decrease only large amplitudes. Distribution of white noise after applying penalized threshold shows the distribution of low amplitudes is reduced and distribution of high amplitudes is increased.

Figs. 5 and 6 show the distribution of white noise in model II and model III, respectively. It can be seen by comparison of Figs, 4 and 5 and 6 that there is no significant difference between the effects of threshold procedures in denoising of horizontal components. In model (I), distribution of white noise in raw vertical component data for both of low and high amplitudes are nearly equal. In model (II), distribution of white noise in raw vertical component data for low amplitudes is less than high amplitudes. In model (III), distribution of white noise in raw vertical component data for high amplitudes is less than low amplitudes.

Fig. 7 illustrates the estimated random walk noise in model (II). Comparison of types of threshold procedures shows that the level of random walk noise (in both low and high amplitudes) with

universal threshold and Birge and Massart strategy threshol
has reduced significantly.

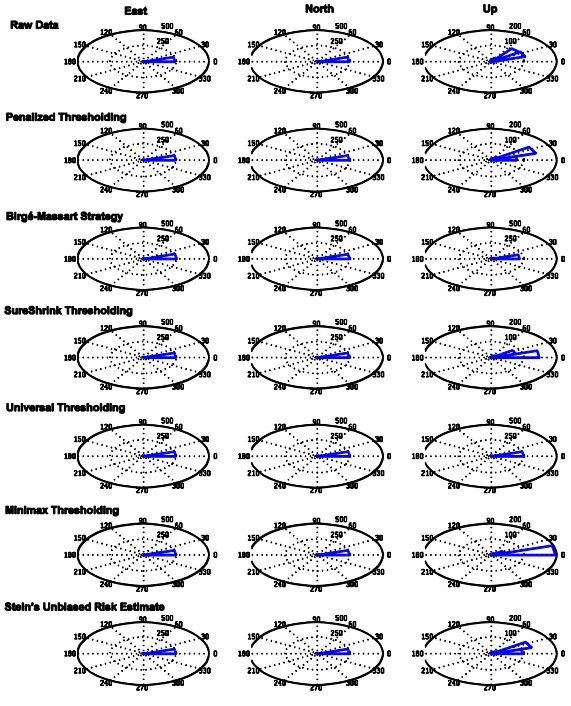

Figure 4. Effects of different threshold methods in denoising of white noise (model I).

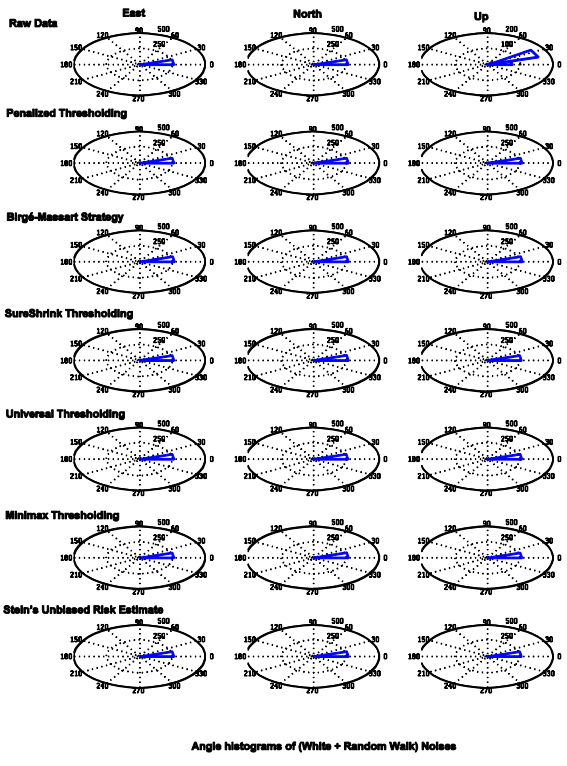

Figure 5. Effects of different threshold methods in denoising of white noise (model II).

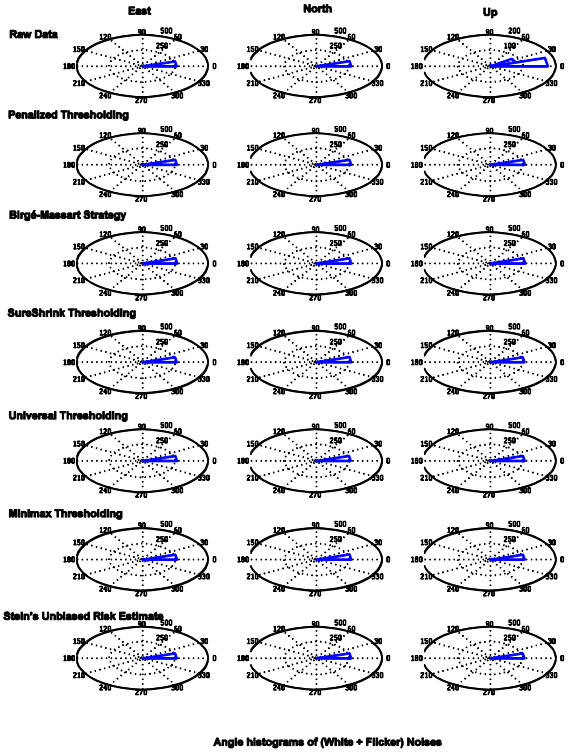

Figure 6. Effects of different threshold methods in denoising of white noise (model III).

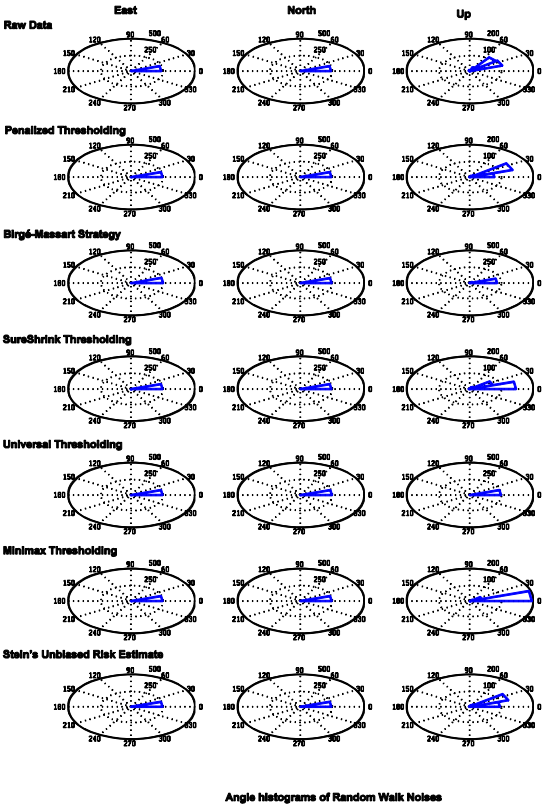

Figure 7. Effects of different threshold methods in denoising of random walk noise (model II).

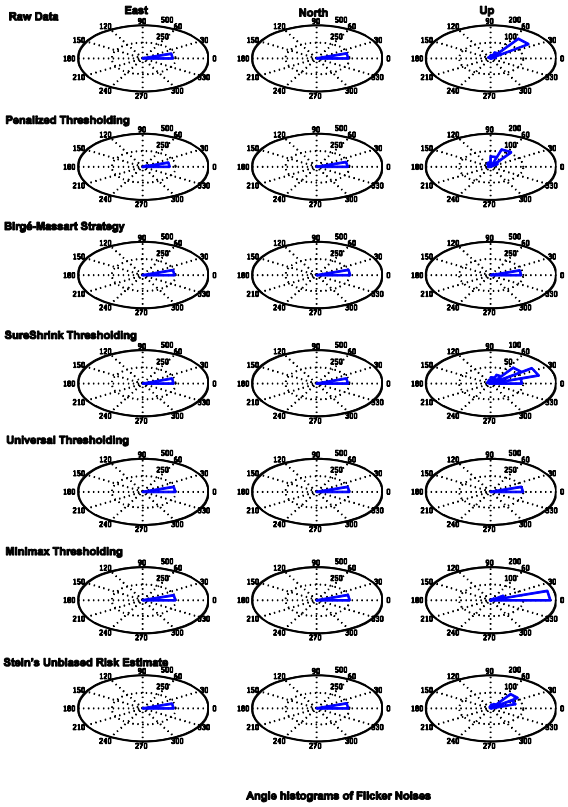

Figure 8. Effects of different threshold methods in denoising of flicker noise (model III).

Fig. 8 illustrates the estimated flicker noise in model (III). As it can be seen in this illustration, the level of flicker noise (in both low and high amplitudes) with universal threshold and Birge and Massart strategy threshold has reduced significantly. However, the other methods could not reduce it significantly.

4. CONCLUSIONS

In this paper we applied six types of thresholding techniques in aim to impact of thresholding in denoising of time series, which were penalized threshold, Birgé-Masssart Strategy, SureShrink threshold, universal threshold, minimax threshold and Stein's unbiased risk estimate. In order to compare the effect of them in denoising of noise components (white noise, flicker noise and random walk noise) we have constructed three kinds of stochastic models: the pure white noise model (I), the white plus random walk noise model (II) and the white plus flicker noise model (III). The results are:

1. In model (I), all methods could reduce the distribution of white noise in horizontal components in a same level. Universal threshold and Birge and Massart strategy threshold could reduce both low and high amplitudes of white noise in vertical components. Minimax threshold and hybrid scheme of the SureShrink threshold could decrease only large amplitudes. Distribution of white noise after applying penalized threshold shows the distribution of low amplitudes is reduced and distribution of high amplitudes is increased.

2. In model (I), distribution of white noise in raw vertical component data for both of low and high amplitudes are nearly equal. In model (II), distribution of white noise in raw vertical component data for low amplitudes is less than high amplitudes. In model (III), distribution of white noise in raw vertical component data for high amplitudes is less than low amplitudes.

3. In models (II) and (III), level of random walk noise and flicker noise (in both low and high amplitudes) with universal threshold and Birge and Massart strategy threshold has reduced significantly. However, the other methods could not reduce it significantly.

REFERENCES

Agnew, D., 1992. The time domain behavior of power law noises. Geophysical Research Letters, 19: 333–336.

Birge´, L., Massart, P., 1997. *From model selection to adaptive estimation*. In Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics, 55 - 88, Springer-Verlag, New York.

Birge´, L., Massart, P., 2001. *Gaussian model selection*. Journal of the European Mathematical Society. 3 (3), 203–268.

Donoho, D., 1993. *Nonlinear Wavelet Methods for Recovery of Signals, Densities, and Spectra from Indirect and Noisy Data*. Proceedings of Symposia in Applied Mathematics, American Mathematical Society.

Donoho, D.L., Johnstone, I.M., 1994. Ideal de-noising in an orthonormal basis chosen from a library of bases .*Comptes Rendus Acad. Sci., Ser. I*,1317-1322,319.

Donoho, D.L., 1995. De-noising by soft-thresholding .*IEEE Transactions on information theory.*, 613-627, (3) 41.

Mallat, S.G., 1999. *A Wavelet Tour of Signal Processing*, Academic Press, 1999, ISBN 012466606X.

Nikolaidis, R., 2002. *Observation of Geodetic and Seismic Deformation with the Global Positioning System* .San Diego: Phd Thesis,University of Califonia.

Williams, SDP., Bock, Y., Fang, P., Jamason, P., Nikolaidis, R.M., Prawirodirdjo, L., Miller, M., Johnson, D.J., 2004. Error analysis of continuous GPS position time series. Journal of Geophysical Research, 109.