Accepted Manuscript

Time dependence of noise characteristics in continuous GPS observations from East Africa

Yelebe Birhanu, Simon Williams, Rebecca Bendick, Shimeles Fisseha

PII: S1464-343X(18)30113-4

DOI: 10.1016/j.jafrearsci.2018.04.015

Reference: AES 3199

To appear in: *Journal of African Earth Sciences*

Received Date: 4 August 2016

Revised Date: 19 April 2018

Accepted Date: 20 April 2018

Time dependence of noise characteristics in continuous GPS observations from East Africa

Yelebe Birhanu[1, 3*], Simon Williams[2], Rebecca Bendick[3], and Shimeles Fisseha[4]

1. School of Earth Sciences, University of Bristol, Wills Memorial Building, Queens Road, Bristol, BS8 1RJ, UK.
2. National Oceanography Centre, Joseph Proudman Building, 6 Brownlow Street Liverpool L3 5DA, UK.
3. Department of Geosciences, University of Montana, Missoula, Montana, 59812, USA.
4. Institute of Geophysics, Space Sciences and Astronomy, Addis Ababa University, Ethiopia.

*Corresponding author: yelebe.birhanu@gmail.com

Abstract

A noise model for the regional continuous GPS (cGPS) timeseries in East Africa (Ethiopia and Eritrea) was computed using the maximum likelihood estimation (MLE) method. Using this method and assigning different noise models for each cGPS site and each component (north, east and vertical) may bias the noise level of the velocity solutions due to the non-uniformity of the length of the timeseries. Within the whole regional network, the length of the timeseries varies from one to seven years. We

compute a preferred regional noise model for the whole data sets using a stacked maximum likelihood values for the different power – law indexes (between -2 and 0 with a time step of 0.1), presuming that there is only one noise model that exists in the regional cGPS timeseries. Therefore, a single power – law index (flicker plus white noise) was assigned for the whole regional network irrespective of the length of the timeseries. This approach is more robust and "realistic" to determine the noise characteristics of the regional GPS network.

Keyword: timeseries, maximum likelihood, noise model, power – law index

Introduction

The noise characteristics of cGPS coordinate timeseries have been studied by various authors [e.g. Agnew, 1992; Mao et al., 1999; Williams et al., 2004; Hackl et al., 2011] in different parts of the world, using different processing methods, from regional networks through to global solutions. It is a well-understood phenomenon that assigning white noise model for the cGPS timeseries will underestimate the noise uncertainty level in the final velocity solution [Mao et al., 1999] since it does not account for the contribution of time varying noises. Work done by these authors [Agnew, 1992; Zhang et al., 1997; Mao et al., 1999; Williams, 2003a; Williams et al., 2004] showed that the cGPS timeseries noise characteristics consists of a combination of white noise and other time-correlated noise models. We used 16 cGPS sites in the northeast part of the East African Rift System (EARS) (Figure 1) and its surroundings, especially Ethiopia and Eritrea, to study the noise characteristics of the whole regional network. The cGPS sites have timeseries that span from early 2007 to 2014 as shown in figure 2. However, the timeseries are non-uniform over that period because of their deployment for various studies and contain discontinuities and gaps in the timeseries caused by various instrumental or logistical problems.

In this study we discuss the noise level of the regional cGPS timeseries using the maximum likelihood estimation (MLE) technique and the time dependence of the noise characteristics as a function of the length of timeseries. We compute a regional noise model for the entire network and compare velocity uncertainties derived from CATS and GLOBK which are based on different noise models.

Methods

All the cGPS data were processed using GAMIT/GLOBK software [Herring et al., 2010]. 15 International GNSS Service (IGS) sites closest to the study area were used as reference sites with one IGS site (ADIS) located within the study area (Figure 1), and included in this study. Using GAMIT, we applied double differencing on each daily phase observation in order to estimate station coordinates, phase ambiguities and satellite state vectors. At every station seven tropospheric delay and two tropospheric gradient parameters per day were estimated [Reilinger et al., 2006]. After applying double differencing to each daily phase observation, the daily solutions (h-files) of the cGPS data were combined with the daily global solutions (H-files) obtained from MIT, using the global kalman filter (GLOBK). The resulting daily timeseries were closely inspected for quality check in order to remove outliers (above two-sigma threshold), and discontinuities caused by antenna changes, receiver changes or earthquakes, together with other time-dependent changes [Reilinger et al., 2006]. Before assigning a Gauss Markov noise model for our cGPS timeseries, the usual routine in GLOBK, we tested different stochastic noise models for each cGPS timeseries. The noise characteristics of the north, east, and up components were computed individually using the CATS GPS coordinate timeseries analysis software [Williams, 2005] that applies maximum likelihood to estimate the noise parameters.

The observed GPS motion at each site is a superposition of secular trend, a periodic component (mostly annual and semi-annual signal), offsets (discontinuities caused by tectonic and non-tectonic processes) and a noise component as shown in equation (1) [Williams, 2003b; Yuan et al., 2007].

$$y_i = a + bt_i + \sum_{j=1}^{l} a_j H(t_j - T_j) + \sum_{n=1}^{m} A_n . \sin(\omega_n t + \varphi_n) + n_i \qquad (1)$$

Where the first term (a) is the site coordinate, the second term (bt_i) is the linear rate (Figure S1), the third term is made up of the Heaviside step function [H (t_j - T_j)] and an offset amplitude (a_j where t_j = T_j) mostly caused by earthquakes, various tectonic processes and other phenomena like antenna or receiver change, and the fourth term consists of periodic components, mainly the annual and semi-annual signals (Figure S2) where $\omega_n = \frac{2\pi}{n}\ rad/year$. The final term is the noise component of the GPS timeseries. In this study we are interested to study the noise term.

From equation (1) above, in order to study the various noise models of the cGPS timeseries the other terms have to be removed from the data. The data has to be detrended, offsets and seasonal variations must also be removed from the data (Figure S3). The correlation in cGPS timeseries may be approximated by a power law process [Agnew, 1992; Mao et al., 1999; Williams, 2003a and b; Williams et al, 2004]:

$$P(f) = P_o\left(\frac{f}{f_o}\right)^{-\alpha} \qquad (2)$$

Where α is the spectral index (white noise has α = 0, flicker noise has α = 1 and random walk noise has a spectral index of 2), Po is a constant, f_0 is a constant frequency and f is the frequency.

We used the method described in CATS [Williams, 2005] in order to compute a "realistic" noise model for the GPS timeseries. The following steps were implemented in order to characterize the noise in the regional network.

1. All the GPS data were detrended and the linear terms in the timeseries were removed using the weighted least squares fit.

2. The detrended data were closely inspected and offsets caused by any tectonic or non-tectonic processes were removed using two-sigma significance level. Tsview software [Herring and McClusky, 2009] was used to inspect and remove the outliers and offsets. Although some of the cGPS sites are located at a closer proximity to the EARS, we did not see any offsets which is caused by the local earthquakes.

3. The annual and semi-annual terms in the data were removed using the fourth term in equation (1). Before removing the annual and semi-annual signals from the data a spectral analysis of the timeseries were computed using the Lomb-Scargle Algorithm [Press et al., 2001] as shown in (Figure S4). This method can help in identifying any dominant periodic signal. We used the R function 'lsp' [R Core Team, 2013] that computes the Lomb-Scargle periodogram [Ruf, 1999; Press et al., 2001]. This algorithm is different from the Fast Fourier Transform because it can handle data sets that are not evenly spaced and does not require data that are a factor of 2 in length (or padded). Since some of our data sets have gaps due to receiver and/or antenna malfunctions or other issues the Lomb-Scargle method is appropriate for these datasets.

4. The residual timeseries (called here the unfiltered timeseries), after the trend, offsets and seasonal signals were removed are then summed on a daily basis to produce mean values that are termed as a Common Mode Error (CME). The CME was then removed from each GPS timeseries to produce a filtered timeseries However since the trends are ultimately derived from the unfiltered series we use these for the MLE estimation of the noise rather than the filtered series.

5. The CME is useful in explaining how spatially coherent the timeseries are. We use MLE estimation to compare different candidate noise models for each series individually. We also use a method to produce a combined regional estimate of the noise in the timeseries.

6. Finally, Principal Component Analysis (PCA) was applied as an alternate to the CME approach to the three individual components of the unfiltered timeseries in order to gauge the usefulness of each technique.

Results

In order to ascertain which noise model is the most realistic for the regional datasets, we look at the sum of the natural logarithm of the ML values for each noise model and each component for the whole data set. After that, the summed noise models were differenced from the ML estimate of the white noise model (the white noise model is used as a null hypothesis), and values greater than the white noise ML indicate that this model is more likely. The top rows of bar graphs in Figure 3 show the difference between the seven stochastic models (flicker noise (F), random walk noise (R), power-law (P), Gauss Markov noise (G), flicker plus white noise (F+W), random walk plus white noise (R+W) and Gauss Markov plus white noise (G+W)) with the null hypothesis (white noise (W)). All the models that include time correlated noise and white noise have larger collective MLs. In the plot the time-correlated noise (R and G) models have values generally smaller than the null hypothesis. In the second row bar plots (Figure 3) instead of taking white noise as the null hypothesis, flicker plus white noise was used as a null hypothesis, and similar computation has been done as before. The result shows that power law, first order Markov noise plus white noise model have almost zero MLE differences with flicker plus white noise, except the north component of the first order G+W noise model which has a higher MLE (Figure 3). The other noise models show a negative value with the null hypothesis (F+W) and therefore less likely to be the noise model candidates

Unlike the above procedure i.e. computing the ML values for the known noise models (flicker, random walk, power-law, etc.) the spectral index of the cGPS timeseries was estimated from each components of the whole regional network. Dmetrieva et al., [2015] used a combination of Kalman filter and MLE to estimate a single noise model for a network of 15 sites in Eastern North America, instead of assigning an individual noise model for each site and component separately. We use a similar approach here and solve for the spectral index that maximizes the likelihood of the whole regional network. If we assume that the individual timeseries are not spatially correlated (or at least that the time and spatial correlations are orthogonal to one another) then the maximum likelihood of the network wide estimation of the spectral index would be the sum of the individual site MLs. In order to estimate a regional noise model for the network the following procedures were implemented. First the ML values for a set of fixed spectral indices for each site and each component (north, east, and up) were computed for the unfiltered timeseries. The computed ML values for each spectral index (between -2 and 0 and time steps of 0.1) were summed in order to compute the ML for the whole network. Figure 6 shows the network-wide ML as a function of spectral index for each component. Instead of using a maximization algorithm to find the spectral index that gives the largest network-wide ML, which could potentially be CPU intensive, we took the three largest

ML values for each component (based on the fixed spectral index values) and fitted a polynomial of degree 2 to those values. Using only the three largest values ensures we can find a maximum to the polynomial. We then solve the polynomial to find the spectral index which maximizes the fit. In Figure 4, the computed spectral index for each cGPS site and each component are compared against the length of the timeseries. The plot shows that, as the length of the timeseries increases, the estimated spectral index tends towards -1 (flicker noise) and when the length of the timeseries is short the estimated spectral index tends to zero (white noise). This is primarily a result of lack of data, biasing the results [Santamaria-Gomez et al., 2011].

Principal Component Analysis (PCA) [Björnsson and Venegas, 1997] attempts to split time series up into several significant modes that are orthogonal to each other. The first mode is often very similar to the CME unless it has been prior to the PCA. The first three Eigenvectors for each component are plotted in Figure 5. In all components, there is a distinct change in the values for sites that started in 2012 onwards. The Eigenvectors are a lot more consistent with more variation in the older sites. This probably reflects that many of the major modes are reflecting the dual subspaces of the dataset, that is those sites containing data mainly prior to 2012 and those sites with data after 2012. In terms of the amount of variance explained the first mode accounts for 68%, 57% and 43% for the north, east and vertical components respectively. The second and third modes explain around 19% and 9%.

For the whole network approach the result showed that the three components have a power-law dependence with estimated spectral index between (-1 < spectral index < -

0.85) for the north, east and up components. We then assign the index that maximizes the fit as the preferred noise model for the whole network. This computation is preferred over assigning different noise models for each component, and each site, as this uses the entire cGPS data sets as compared to using the individual timeseries.

Discussion

The preferred noise model was computed based on the above two methods: using the difference of the natural logarithm MLE values and the stacked MLE values and fitting a polynomial of degree 2 that maximizes the spectral index. The regional noise analysis that uses the stacked MLE values is a more appropriate way to assign a "realistic" noise model for the regional timeseries computation. For the entire cGPS network, we expected to have one type of noise in the timeseries and not different power-law noises for each site and each component. Since our data sets have different timeseries length, assigning a different noise model for each component of the timeseries may bias the noise estimates (Figure 4). The regional noise model analysis uses the whole data sets and assign a single spectral index for the entire network. This regional analysis is a more robust way of assigning a noise model.

Conclusion

The GLOBK realistic sigma estimation, that uses first order Gauss Markov noise, shows coherent sigma estimates with the CATS estimates, that we assign flicker plus white noise as a preferred noise model. As shown in Figure S5 GLOBK and CATS velocity noise analysis are most correlated when the timeseries are longer and least correlated when the timeseries are shorter. Shorter timeseries bias the noise estimates (Figure 4) and affect the velocity error estimates. In order to assign the optimal noise model for the entire cGPS network, for a non-uniform length of a cGPS timeseries, care has to be taken. Therefore, in this study, we have assigned flicker plus white noise model as a preferred noise model for the whole regional network regardless of the length of the timeseries.

Acknowledgements

This work is supported by NSF EAR-1119209 and NERC funding through RiftVolc NE/L013932/1. We would like to thank Institute of Geophysics, Space Sciences and Astronomy, Addis Ababa University and Eritrean Institute of Technology. Data are archived in the UNAVCO archive.

References

Agnew, D. (1992), The time domain behavior of power law noises, Geophys. Res. Lett., 19, 333–336.

Björnsson, H. and S. A. Venegas (1997), A Manual for EOF and SVD Analyses of Climate Data. Department of Atmospheric and Oceanic Sciences and Centre for Climate and Global Change Research, McGill University.

Dmitrieva, K., P. Segal, and C. DeMets, (2015), Network-based estimation of time-dependent noise in GPS position timeseries Journal of Geodesy.

Hackl, M., R. Malservisi, U. Hugentobler, and R. Wonnacott, (2011), Estimation of velocity uncertainties from GPS timeseries: Examples from the analysis of the South African TrigNet network", J. Geophys. Res., vol. 116, no. B11.

Herring, T., R. King, and S. McClusky (2010), Introduction to GAMIT/ GLOBK Release 10.4, Mass. Inst. of Technol., Cambridge.

Herring T., and McClusky S., (2009), Timeseries viewing, editing and interface to GLOBK Vers. 2.02, MIT.

Mao, A., C. G. A. Harrison, and T. H. Dixon (1999), Noise in GPS coordinate timeseries, J. Geophys. Res., 104(B2), 2797–2816.

Press, W. H., Teukolsy, S. A., Vetterling, W. T., and Flannery, B. P., (2002), Spectral Analysis of Unevenly Sampled Data. *Numerical Recipes in C++ (2nd edition)*, Cambridge University Press.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org

Ruf, T., (1999) The Lomb-Scargle Periodogram in Biological Rhythm Research: Analysis of In- complete and Unequally Spaced Time-Series. Biological Rhythm Research 30: 178–201.

Santamaria- Gómez, A., M. Bouin, X. Collilieux, G. Wöppelmann (2011), Correlated errors in GPS position timeseries: Implications for velocity estimates, J. Geophys. Res., 116, B01405, doi:10.1029/2010JB007701.

Williams, S. D. P. (2003a), The effect of coloured noise on the uncertainties of rates estimated from geodetic timeseries, J. Geod., 76, 483–494.

Williams, S. D. P. (2003b), Offsets in Global Positioning System timeseries, J. Geophys. Res., 108(B6), 2310, doi:10.1029/2002JB002156.

Williams, S. D. P., Y. Bock, P. Fang, P. Jamason, R. M. Nikolaidis, L. Prawirodirdjo, M. Miller, and D. J. Johnson (2004), Error analysis of continuous GPS position timeseries, J. Geophys. Res., 109, B03412, doi:10.1029/2003JB002741.

Williams, S.D.P. (2005), Create and Analyze Time Series: CATS software V3.1.1.

Yuan L., X. Ding, W. Chen, S. Kwok, and R. Chan, (2007), Statistic analysis of daily position timeseries from the Hong Kong local dense GPS network, international Global Navigation Satellite Systems Society IGNSS Symposium.

Zhang, J., Y. Bock, H. Johnson, P. Fang, S. Williams, J. Genrich, S. Wdowinski, and J. Behr (1997), Southern California Permanent GPS Geodetic Array: Error analysis of daily position estimates and site velocities, J. Geophys. Res., 102, 18,035–18,055.

Figure captions

Figure 1. Location of the cGPS sites overlain on the DEM map of the study area. Each triangle shows the location of the GPS sites with site names next to each triangle.

Figure 2. The detrended cGPS displacement timeseries for each site and for the north, east and vertical components respectively. The y-values are scaled for plotting purpose only.

Figure 3. Histogram of the log maximum likelihood values for the various noise models. The labels in the horizontal axis, W- white noise, F – flicker noise, R- random walk noise, P – power-law noise, G – first order Gauss Markov noise, F+W – flicker plus white noise, R+W – random walk plus white noise, and G+W – first order Gauss Markov

plus white noise model. The first row, three bar graphs are for the north, east and vertical components respectively. The first row three bar plots show white noise (W) as a null hypothesis and compared against the mean of the difference of the natural logarithm ML of white noise with other noise models. In the second row, the bar plots (for the north, east and vertical), the similar procedure was done and in this case flicker plus white noise (F+W) was taken as a null hypothesis.

Figure 4. Time dependence of the estimated spectral index of the cGPS timeseries for the north south, east west and up components. The estimated spectral index is computed for the different duration of the timeseries (from one to seven years).

Figure 5. The Principal Component Analysis (PCA) for the north, east and up components. The three colors (purple, blue and red) indicates the first, second and third principal component (pc) respectively for the north, east and vertical components. The cGPS site names are labeled below the horizontal axis.

Figure 6. The first three plots show, the stacked ML values for a selected spectral index between -2 and 0 with an interval of 0.1. In the second three plots the three maximum-stacked MLE values are selected and the associated polynomial fit of degree 2 was fitted to the values that maximizes the spectral index. The broken red vertical lines indicate the maximum of the fit (-0.89, -0.92 and -0.98) for the north, east and vertical respectively.

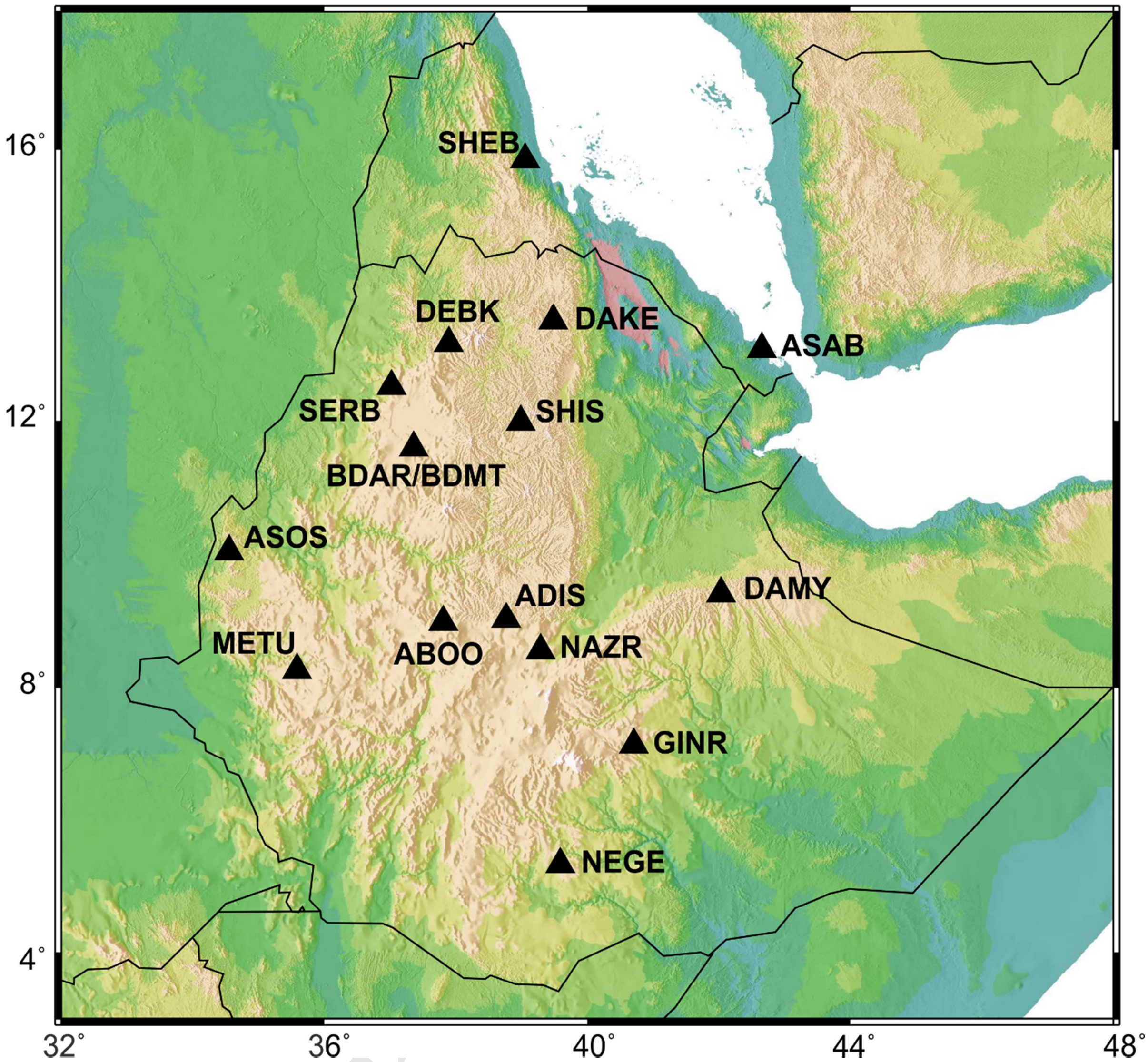
Figure S1 GPS time series of ADIS station. In this pot the linear trend, annual and semiannual terms are not removed. The dotted red line is the weighted least square fit for the north, east and up components. The north and east components show significant linear trend.

Figure S2 The linear trend of the time series were removed and the red line is the annual and semiannual fit. The up component shows significant annual and semiannual component.

FigureS3 Residual GPS time series where the linear, annual and semiannual terms were removed from the time series.

Figure S4 Lomb – Scargle periodogram of the cGPS timeseries for north, east and up components with the corresponding site names on the left bottom corners of the plots. These are the selected periodogram that have longer timeseries and based on the geographic distribution. The blue plots are for the north, purple for the east and red for vertical components. The broken black lines in the plots show the annual frequency. The label in the bottom right corner indicates, slope of north component (slope_n), slope of east component (slope_e) and slope of up/vertical component (slope_u).

Figure S5 Comparison of the north-south and east-west CATS velocity uncertainties (purple color), which uses flicker plus white noise as a preferred noise model, in this study, and the GLOBK velocity uncertainties (blue color) based on Gauss Markov noise model. The cGPS site names are labeled below the horizontal axis and the velocity uncertainties in the vertical axis.



16°
12°
8°
4°
32°
36°
40°
44°
48°
SHEB
DEBK
DAKE
ASAB
SERB
SHIS
BDAR/BDMT
ASOS
ADIS
DAMY
METU
ABOO
NAZR
GINR
NEGE

Figure 1

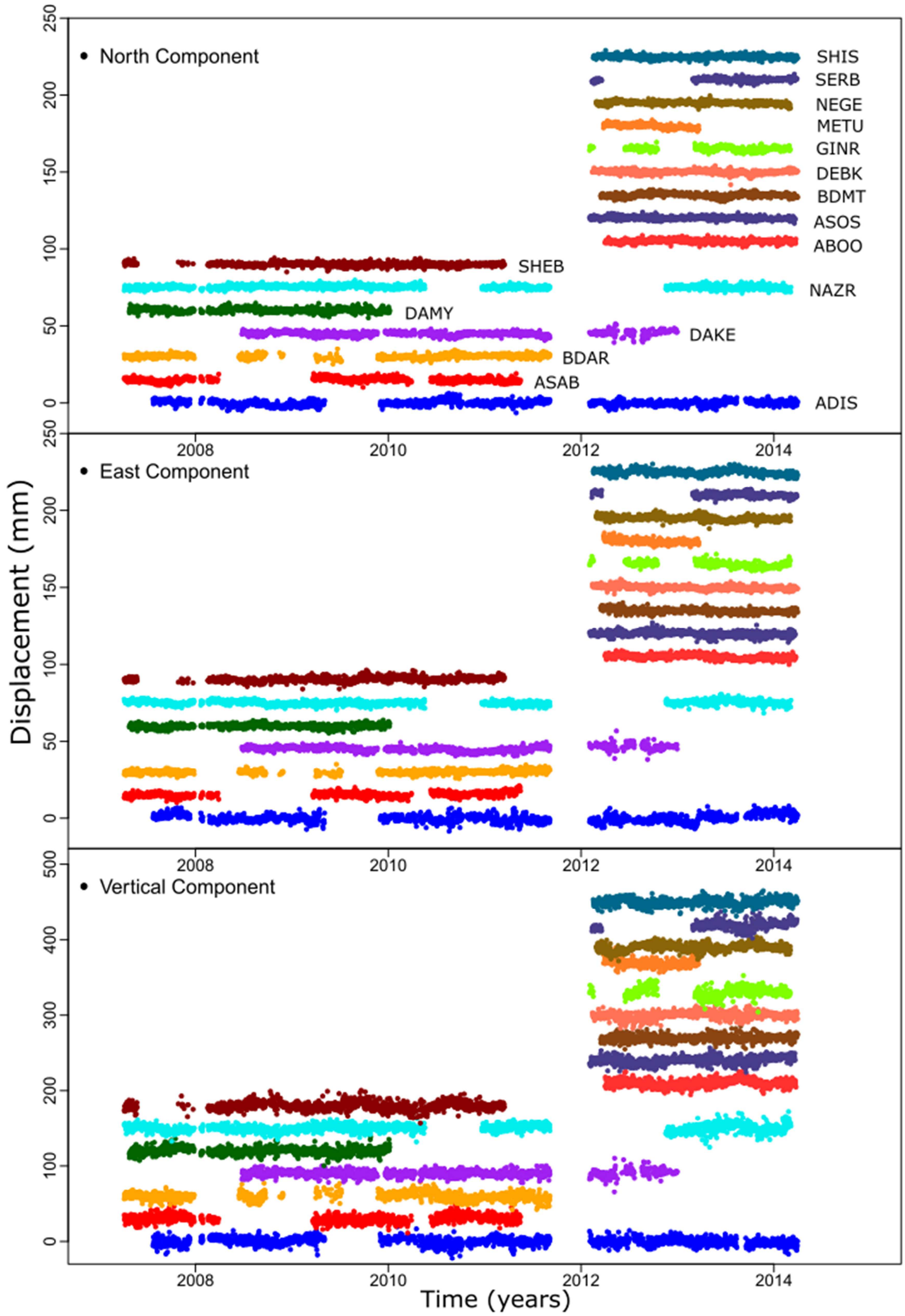

North Component
East Component
Vertical Component
SHIS
SERB
NEGE
METU
GINR
DEBK
BDMT
ASOS
ABOO
SHEB
NAZR
DAMY
DAKE
BDAR
ASAB
ADIS
Displacement (mm)
Time (years)
2008
2010
2012
2014
0
50
100
150
200
250
300
400
500

Figure 2

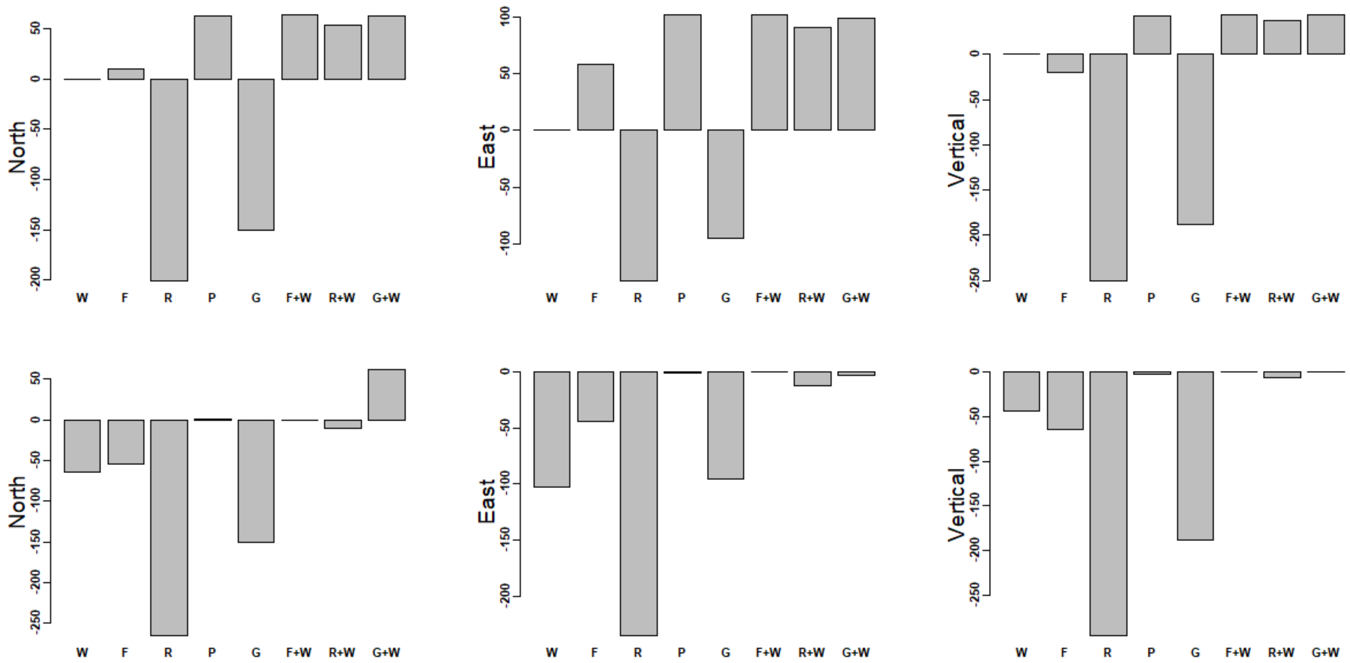


Figure 3

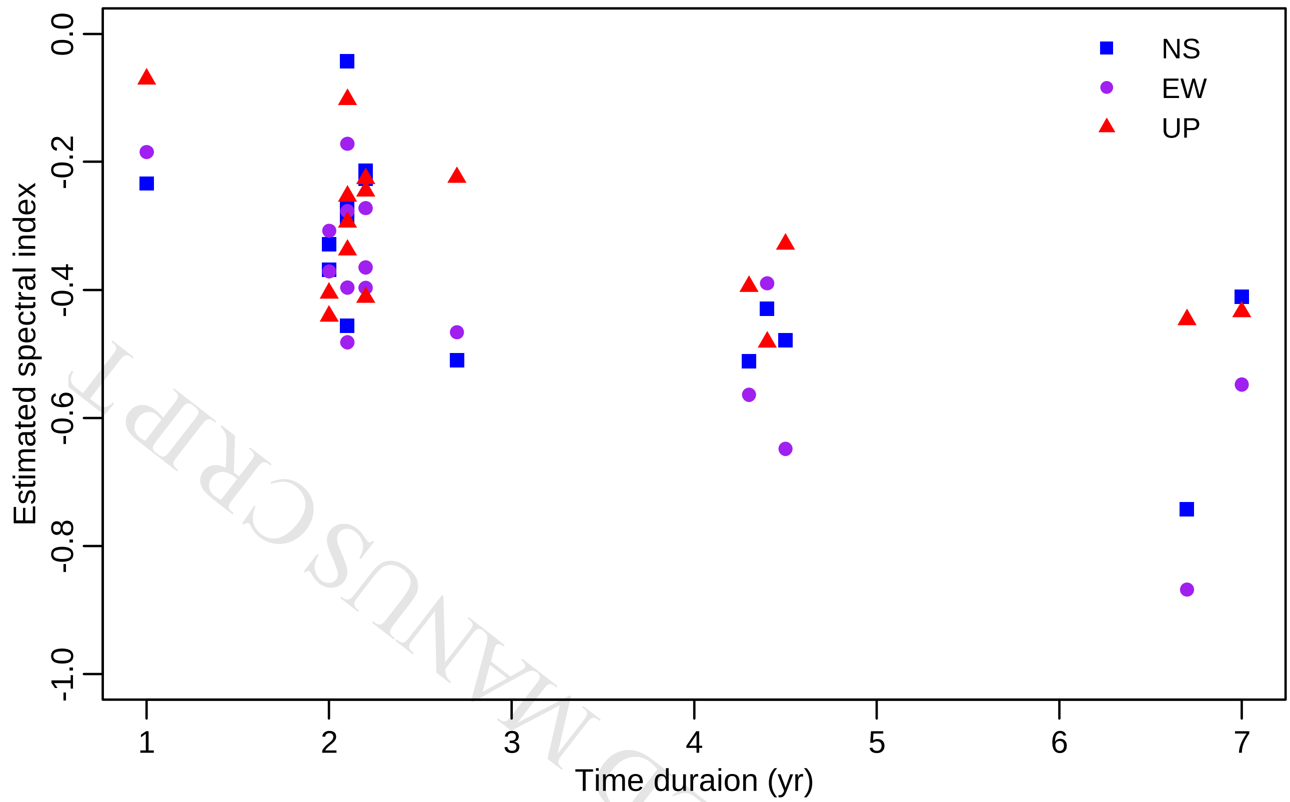

Estimated spectral index
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
1
2
3
4
5
6
7
Time duraion (yr)
NS
EW
UP

Figure 4

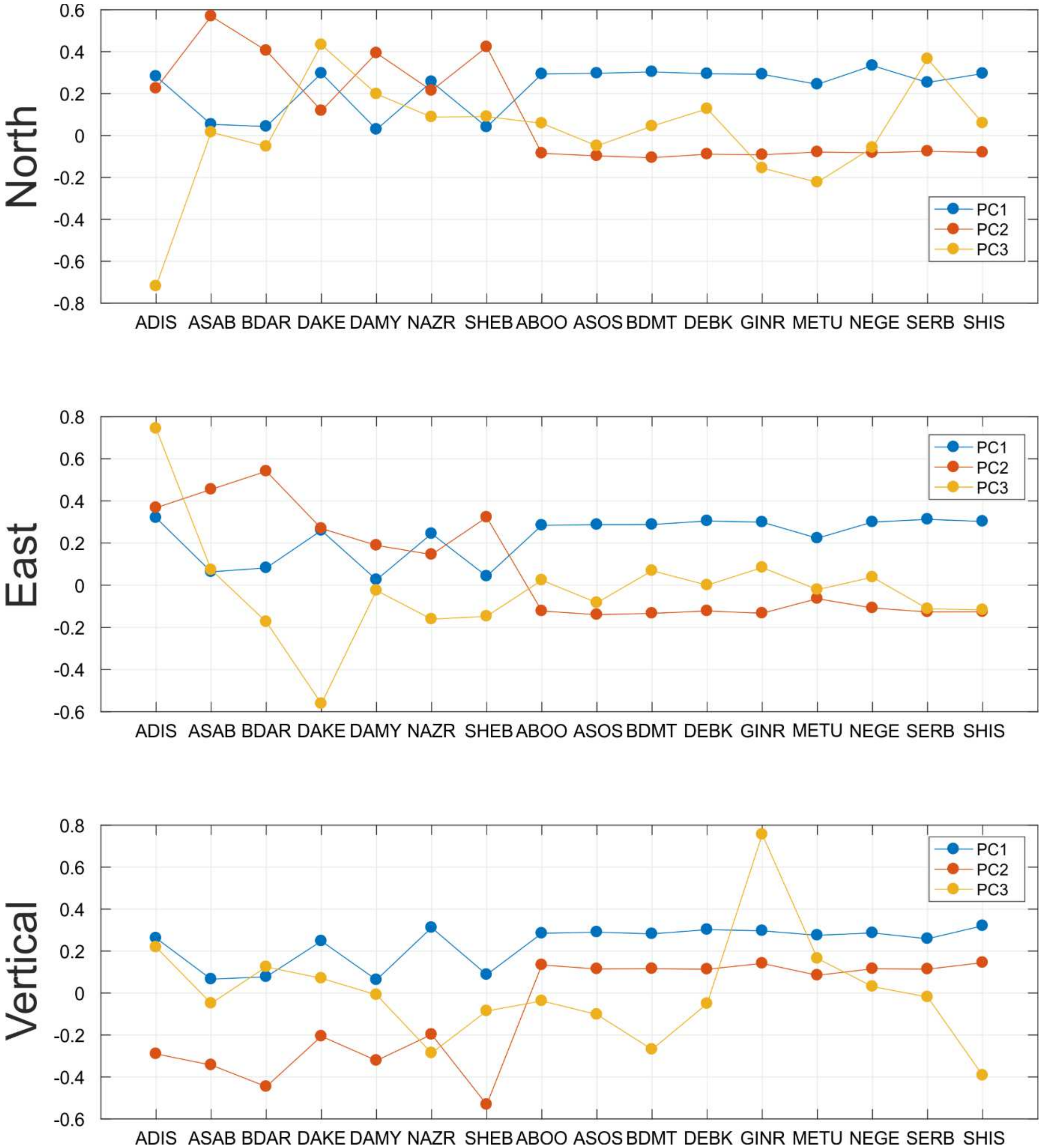

North
East
Vertical
PC1
PC2
PC3
ADIS ASAB BDAR DAKE DAMY NAZR SHEB ABOO ASOS BDMT DEBK GINR METU NEGE SERB SHIS

Figure 5

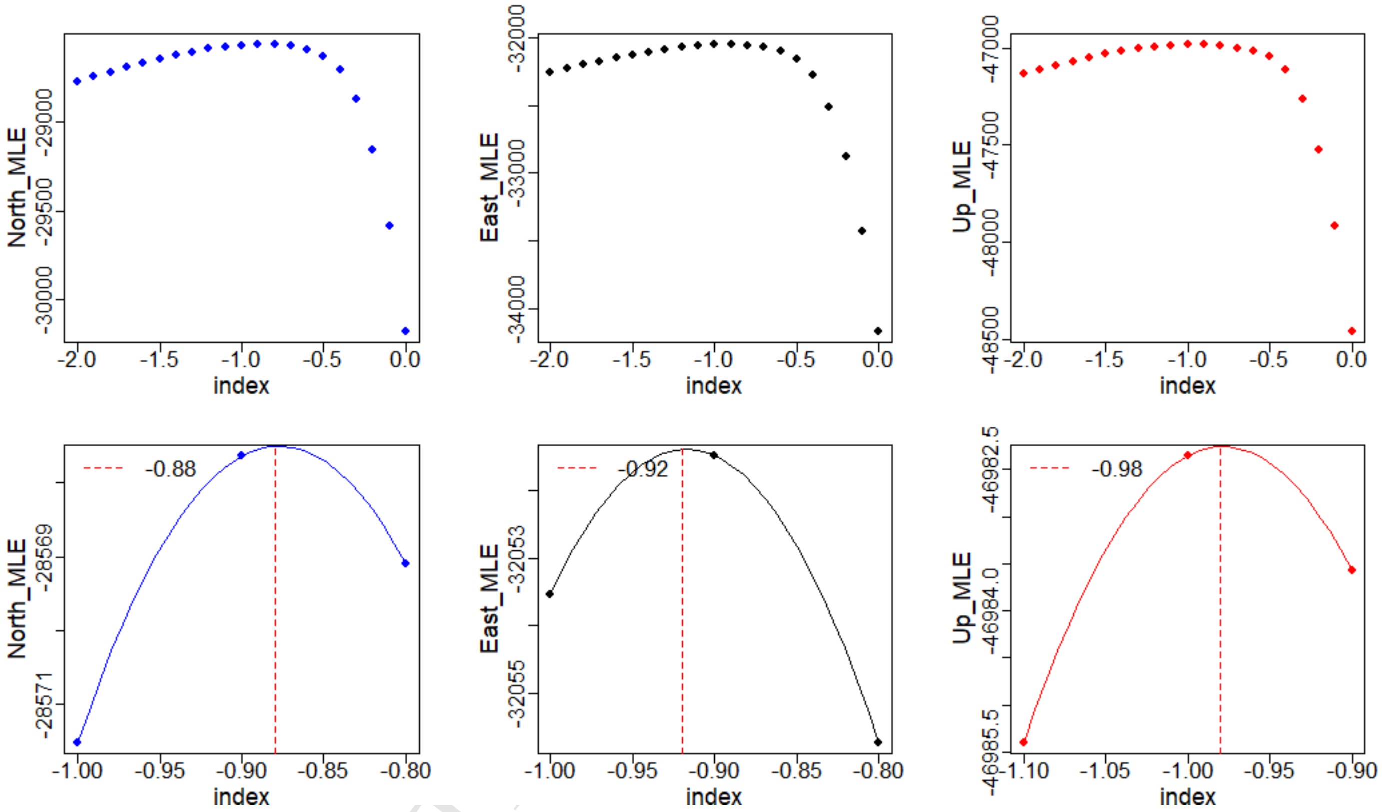


Figure 6

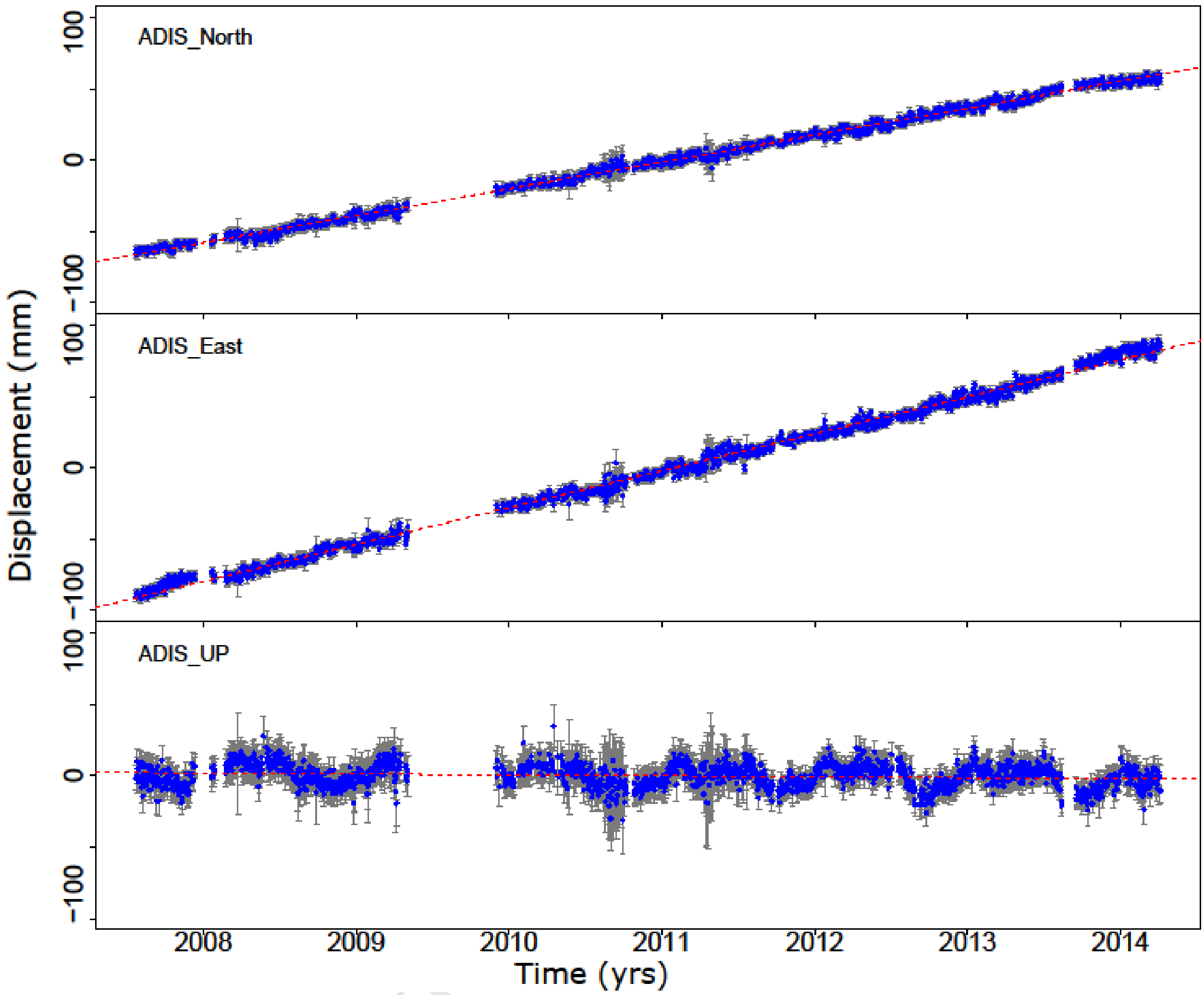

ADIS_North
ADIS_East
ADIS_UP
Displacement (mm)
100
0
−100
2008
2009
2010
2011
2012
2013
2014
Time (yrs)

Figure S1

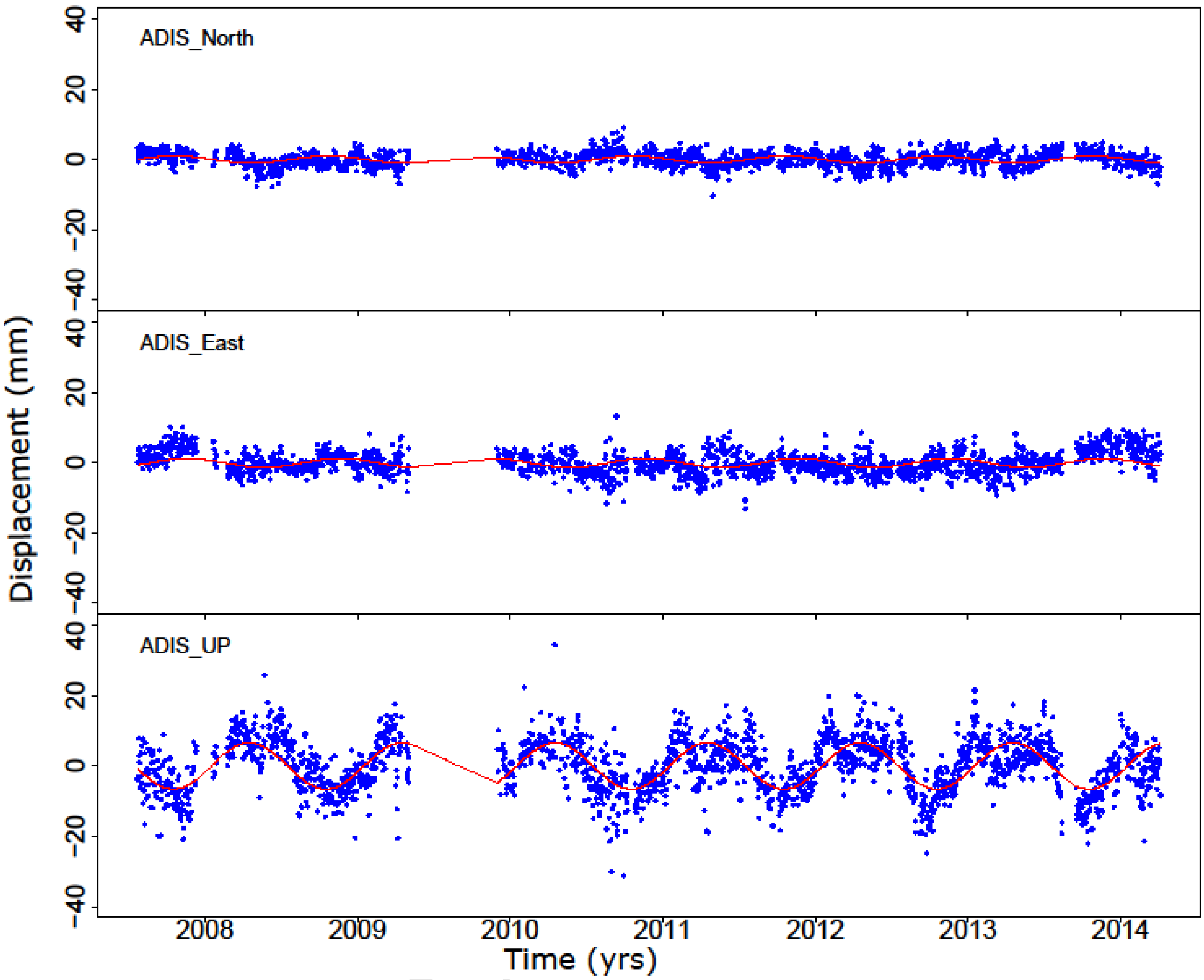

ADIS_North
ADIS_East
ADIS_UP
Displacement (mm)
40
20
0
-20
-40
2008
2009
2010
2011
2012
2013
2014
Time (yrs)

Figure S2

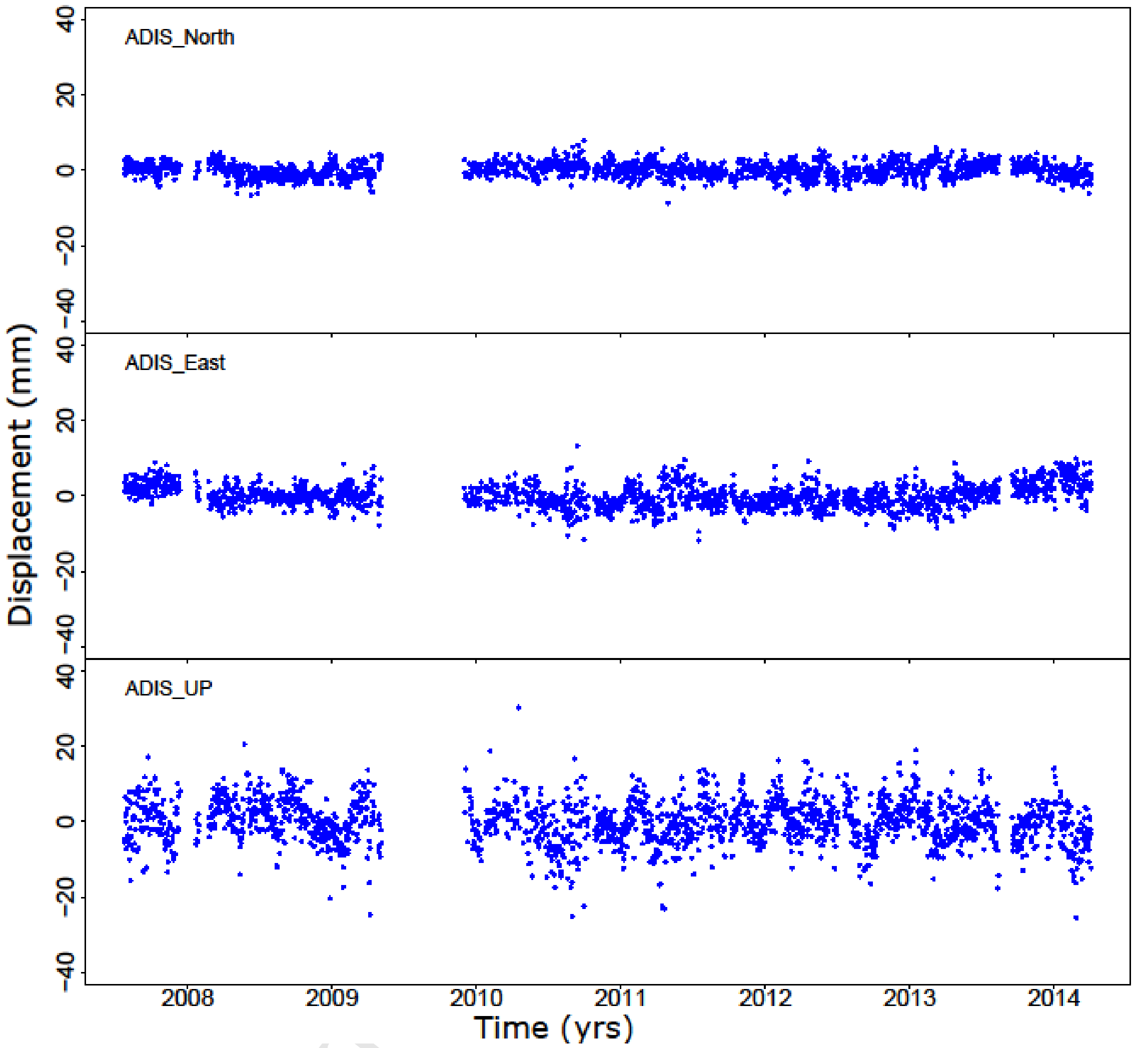
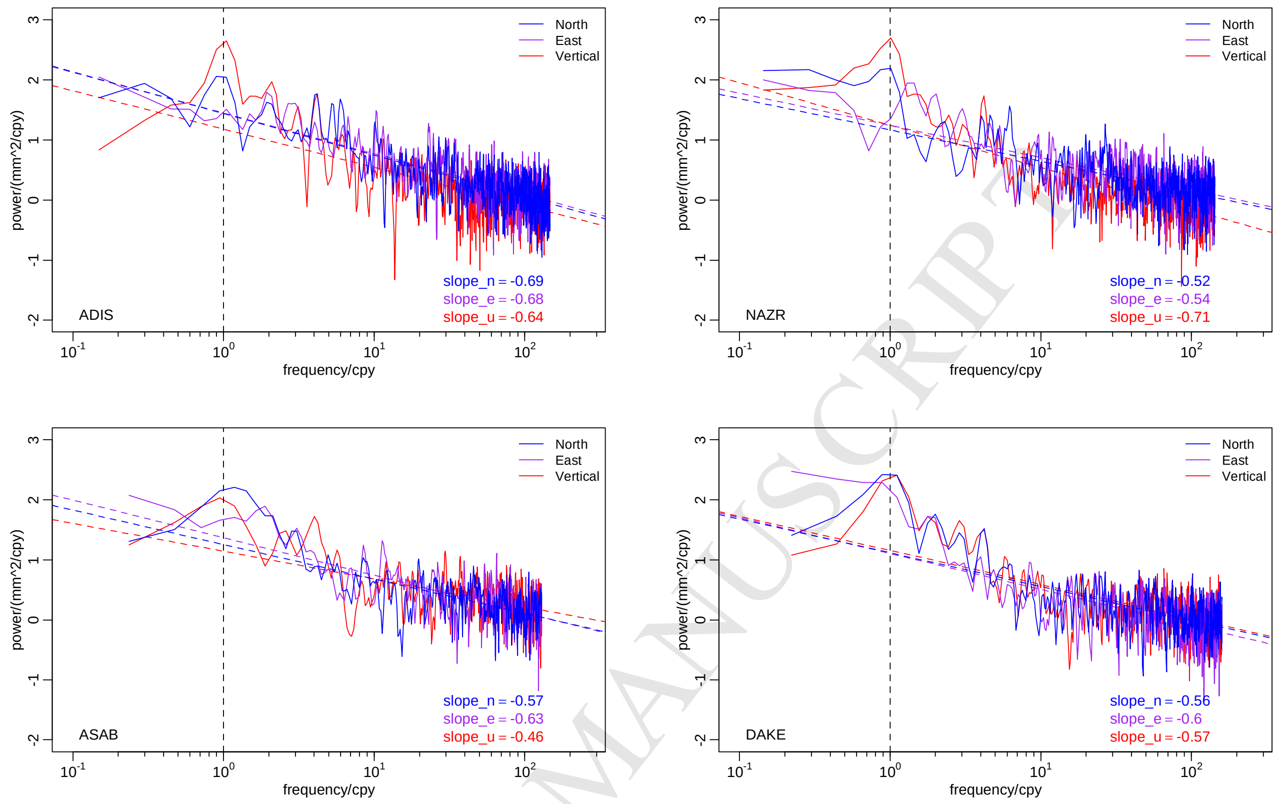

ADIS_North
ADIS_East
ADIS_UP
Displacement (mm)
40
20
0
-20
-40
2008
2009
2010
2011
2012
2013
2014
Time (yrs)

Figure S3



North
East
Vertical
ADIS
slope_n = -0.69
slope_e = -0.68
slope_u = -0.64
NAZR
slope_n = -0.52
slope_e = -0.54
slope_u = -0.71
ASAB
slope_n = -0.57
slope_e = -0.63
slope_u = -0.46
DAKE
slope_n = -0.56
slope_e = -0.6
slope_u = -0.57
power/(mm^2/cpy)
frequency/cpy

Figure S4

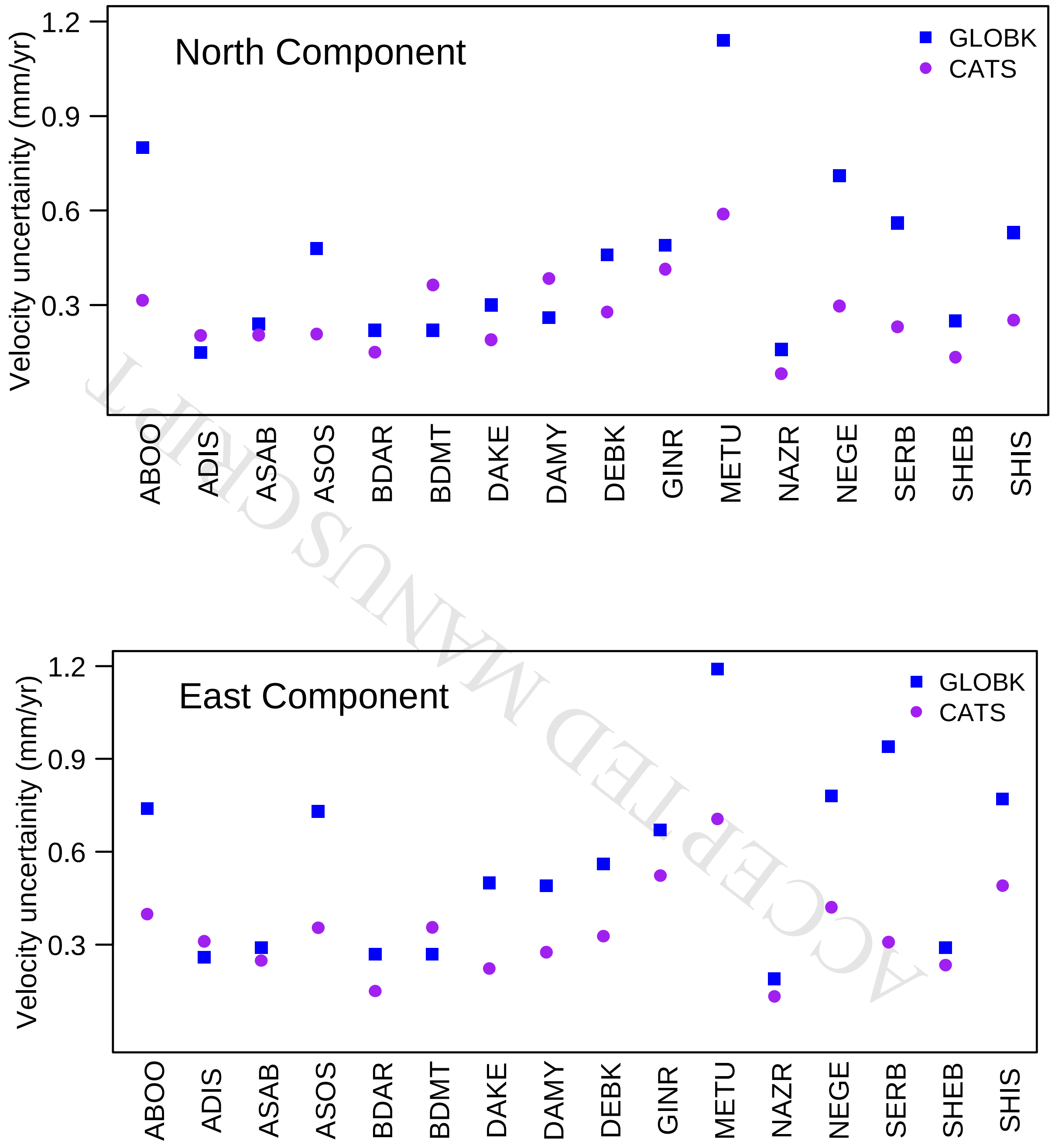

North Component
GLOBK
CATS
Velocity uncertainity (mm/yr)
0.3
0.6
0.9
1.2
ABOO
ADIS
ASAB
ASOS
BDAR
BDMT
DAKE
DAMY
DEBK
GINR
METU
NAZR
NEGE
SERB
SHEB
SHIS
East Component
GLOBK
CATS
Velocity uncertainity (mm/yr)
0.3
0.6
0.9
1.2
ABOO
ADIS
ASAB
ASOS
BDAR
BDMT
DAKE
DAMY
DEBK
GINR
METU
NAZR
NEGE
SERB
SHEB
SHIS

Figure S5

- 16 continuous GPS sites in Ethiopia and Eritrea have been used to compute the noise characteristics of the GPS velocity uncertainties.
- The network wide regional noise analysis shows flicker plus white noise model is the "robust" noise model for the regional GPS velocity uncertainties.
- The uncertainties of the GPS velocity estimates are biased due to the length of the GPS timeseries (we have used 1 to 7 years of GPS data).