ANALYSIS OF GNSS TIME SERIES OBTAINED FROM TURKISH NATIONAL PERMANENT GNSS STATIONS NETWORK-ACTIVE SYSTEM USING HILBERT-HUANG TRANSFORM

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

SONER ÖZDEMİR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
CIVIL ENGINEERING

JUNE 2014

Approval of the thesis:

ANALYSIS OF GNSS TIME SERIES OBTAINED FROM TURKISH NATIONAL PERMANENT GNSS STATIONS NETWORK-ACTIVE SYSTEM USING HILBERT-HUANG TRANSFORM

submitted by **SONER ÖZDEMİR** in partial fulfillment of the requirements for the degree of **Master of Science in Civil Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen _____
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Ahmet Cevdet Yalçıner _____
Head of Department, **Civil Engineering**

Prof. Dr. Mahmut Onur Karslıoğlu _____
Supervisor, **Civil Engineering Dept., METU**

Examining Committee Members:

Prof. Dr. Zuhal Akyürek _____
Civil Engineering Dept., METU

Prof. Dr. Mahmut Onur Karslıoğlu _____
Civil Engineering Dept., METU

Prof. Dr. Gerhard-Wilhelm Weber _____
Institute of Applied Mathematics, METU

Assoc. Prof. Dr. Ali Kılıçoğlu _____
General Command of Mapping – Retired

Assist. Prof. Dr. Metin Nohutcu _____
Geomatics Eng. Dept., Hacettepe University

Date:

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name : SONER ÖZDEMİR

Signature :

ABSTRACT

ANALYSIS OF GNSS TIME SERIES OBTAINED FROM TURKISH NATIONAL PERMANENT GNSS STATIONS NETWORK-ACTIVE SYSTEM USING HILBERT-HUANG TRANSFORM

Özdemir, Soner
M.Sc., Department of Civil Engineering
Supervisor : Prof. Dr. Mahmut Onur Karslıoğlu

June 2014, 126 Pages

In this thesis, time series of position estimates are generated from the data collected at Turkish National Permanent GNSS Stations Network-Active (TNPGN-Active) stations with high-precision GNSS analysis methods, and obtained time series are analysed using Hilbert Huang Transform (HHT) whereby associated problems of this transformation are investigated.

The accuracy of the positional correction parameters sent to the end users on the field from TNPGN-Active system is important when it is considered these coordinates are used in the new cadastral applications. After the generation of time series, problematic sites are detected which can impose adverse effects on coordinate accuracies.

GNSS time series show non-linear and/or non-stationary behaviours due to the underlying physical processes. Adaptive nature of HHT makes it possible to address each time series seperately and sheds light upon the individual characteristics of the time series. With Empirical Mode Decomposition (EMD) method, original GNSS signals are transformed into amplitude/frequency modulated Intrinsic Mode Functions (IMFs) in this study. Together with the Hilbert Spectrum, a more detailed representation of the physical processes is supplied.

Problematic areas in HHT, such as interpolation technique between extrema, end effects, stopping criteria for sifting process, mode mixing, and some coproducts of EMD, such as detrending and denoising are also evaluated.

With this study, HHT is applied to TNPGN-Active time series for the first time. Great efforts are awaiting for physical interpretation of the results obtained in this study.

Keywords : TNPGN-Active, Time series, Hilbert-Huang Transform, Empirical Mode Decomposition, Intrinsic Mode Functions.

ÖZ

TÜRKİYE ULUSAL SABİT GNSS İSTASYONLARI AĞI-AKTİF SİSTEMİNDEN ELDE EDİLEN GNSS ZAMAN SERİLERİNİN HİLBERT HUANG DÖNÜŞÜMÜ İLE ANALİZ EDİLMESİ

Özdemir, Soner
Yüksek Lisans, İnşaat Mühendisliği Bölümü
Tez Yöneticisi : Prof. Dr. Mahmut Onur Karslıoğlu

Haziran 2014, 126 Sayfa

Bu çalışmada, Türkiye Ulusal Sabit GNSS İstasyonları Ağı-Aktif (TUSAGA-Aktif) verileri yüksek hassasiyetli GNSS analiz yöntemleri ile analiz edilerek konum zaman serileri üretilmiş, elde edilen zaman serilerine Hilbert-Huang Dönüşümü (HHD) uygulanmış ve bu dönüşümle ilgili problemler araştırılmıştır.

Yeni kadastral uygulamaların TUSAGA-Aktif sisteminden elde edilen koordinatlara dayalı olarak gerçekleştirildiği göz önüne alındığında, arazideki son kullanıcılara TUSAGA-Aktif sisteminden gönderilen koordinat düzeltme parametrelerinin önemi ortaya çıkmaktadır. Zaman serilerinin oluşturulmasının ardından, koordinat doğruluklarına olumsuz etkileri olabilecek istasyonlar tespit edilmiştir.

GNSS zaman serileri, altında yatan fiziksel süreçler nedeniyle, doğrusal ve/veya durağan olmayan davranışlar sergilemektedir. HHD'nin uyarlanabilir yapısı, her bir zaman serisini ayrı ayrı ele almayı ve kendilerine özgü niteliklerini ortaya çıkarmayı mümkün kılmaktadır. Bu çalışmada, Deneysel Biçim Ayrışımı (DBA) yöntemi ile GNSS zaman serileri, genlik/frekans modülasyonlu Öz Biçim Fonksiyonlarına (ÖBF) ayrıştırılmıştır. Hilbert Spektrumu ile, fiziksel süreçlerin daha detaylı bir sunumu gerçekleştirilmiştir.

HHD'nin, uçdeğerler arasında gerçekleştirilecek interpolasyon yöntemi, sinyal baş ve sonlarında oluşabilecek sallanmalar, ayrıştırmayı sonlandıracak durma kriterinin seçimi, benzer ölçekli sinyallerin aynı ÖBF içinde bulunması gibi problemli alanları ve trend belirleme ve gürültü giderme gibi yan ürünleri de ayrıca değerlendirilmiştir.

Bu çalışma ile ilk kez TUSAGA-Aktif zaman serilerine HHD uygulanmıştır. Sırada, elde edilen sonuçların fiziksel olarak yorumlanabilmesi için gerçekleştirilecek çalışmalar bulunmaktadır.

Anahtar Kelimeler : TUSAGA-Aktif, Zaman serisi, Hilbert-Huang Dönüşümü, Deneysel Biçim Ayrışımı, Öz Biçim Fonksiyonları.

ACKNOWLEDGEMENTS

The author expresses his gratitude to his supervisor Prof. Dr. Mahmut Onur Karslıoğlu for his unprecedented guidance throughout the research.

The assistance of Mr. Ayhan Cingöz is gratefully acknowledged.

The GNSS data used in this thesis are provided by General Command of Mapping.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS

ADSL	Asymmetric Digital Subscriber Line
AM/FM	Amplitude/Frequency Modulated
ANKR	Ankara Station
CA	Coarse Acquisiton
CGNNS	Continuously Operationg GNSS Stations
CMR	Compact Measurement Record
CODE	Center of Orbit Determination in Europe
CUSUM	Cumulative Sum
DCB	Differential Code Biases
DFT	Discrete Fourier Transform
EDGE	Enhanced Data for Global Evolution
EEMD	Ensemble Empirical Mode Decomposition
EMD	Empirical Mode Decomposition
FFT	Fast Fourier Transform
FT	Fourier Transform
GCM	General Command of Mapping
GLR	Generalized Likelihood Ratio
GNSS	Global Navigation Satellite Systems
GPRS	General Packet Radio Service
HHT	Hilbert-Huang Transform

HSA	Hilbert Spectral Analysis
HT	Hilbert Transform
IERS	International Earth Rotation and Reference Systems Service
IFT	Inverse Fourier Transform
IGS	International GNSS Service
IMF	Intrinsic Mode Function
IRLS	Iteratively Rewighted Least Squares
ITRF	International Terrestrial Reference Frame
JPL	Jet Propulsion Laboratory
KLIEP	Kullback-Leibler Importance Estimation Procedure
LTI	Linear Time-Invariant
MIT	Massachusetts Institute of Technology
NRMS	Normalized Root Mean Square
NTRIP	Networked Transport of RTCM via Internet Protocol
P	Protected
PCA	Principal Component Analysis
PCV	Phase Center Variation
PPP	Precise Point Positioning
RINEX	Receiver Independent Exchange Format
RMS	Root Mean Square
RTCM	Radio Technical Commission for Maritime Services
RTK	Real Time Kinematic
SD	Standard Deviation
SNR	Signal-to-Noise Ratio
SOI	Southern Oscillation Index

SOPAC	Scripps Orbit and Permanent Array Center
SR	Stochastic Resonance
STFT	Short-time Fourier Transform
TNPGN	Turkish National Permanent GNSS Stations Network
WRMS	Weighted Root Mean Square
WSD	Weak Signal Detection

CHAPTER 1

INTRODUCTION

GNSS (Global Navigation Satellite Systems) applications have been evolved from their fundamental positioning purposes to wide and diverse practices especially in the last two decades. It is a geodetic technique and serves for the demands of geodetic community comprehensively notwithstanding, it also provides many valuable coproducts for related sciences. For example, for monitoring secular crustal deformation in certain regions, specific GNSS networks have been established while interseismic and post-seismic deformations have been inspected by episodic GNSS observations in Turkey. Velocity field and time-series analyses of GNSS data shed light upon the deformation and seismic hazard areas.

A Real Time Kinematic (RTK) network, Turkish National Permanent GNSS Stations Network-Active (TNPGN-Active), covering all country was established in order to meet intensive cadastral needs (http://www.hgk.msb.gov.tr/english/tnpgn-active.php). While daily GNSS data from TNPGN-Active provide an indispensable tool for investigating tectonic and seismic activities, the accuracy of the coordinate correction parameters sent to the end users on the field is significant when it is considered that the new cadastral applications are started to be based upon these coordinates. Thus, detailed analysis of the coordinate time-series of TNPGN-Active stations and the removal of the unwanted signal which can affect the coordinate and velocity estimates adversely are of utmost importance.

It is known that the accuracy level of the observations is a major constraint on the assessment of the behavior of a physical system, and significant information might be hidden in the noise in many cases. Hence, the necessity of decomposing the

original signal into its smallest constituents is distinct. The main objective of this thesis is to produce TNPGN-Active coordinate time series in a consistent way and to decompose these time series by implementing Hilbert-Huang Transform.

1.1 The Subject and the Scope

Time series are used almost everywhere; in statistics, pattern recognition, signal processing, weather forecasting, earthquake prediction, astronomy, and so on. There are different kinds of classifications for time series analysis methods. They can be divided into frequency-domain methods and time-domain methods. They can be also classified as parametric and non-parametric methods. Further, methods may be divided into univariate and multivariate, or linear and non-linear methods. There are tens of techniques which have been explored intensively, with the primary motivation being forecasting, signal detection and estimation. Recently in Turkey, time series analysis of a data collection has gained significance: daily position solutions of continuously operating GNSS stations.

While GNSS's accurate positioning capability facilitates everyday activities, and meter-level positioning accuracy is sufficient for many individual users, some crucial applications such as foundation of cadastral infrastructure, detection of sea level rise, tectonic monitoring etc. require cm or even sub-cm level positioning accuracy. This necessity entails cutting-edge GNSS processing techniques as well as detailed inspection and careful monitoring of position time series.

Although detecting the motions of the tectonic plates by measuring precise three dimensional positions of geodetic markers is purely a geodetic technique, it has been widely used also by seismologists for decades to assist their inferences about the general tectonic structure of the ground.

Seismology measures seismic motions after an earthquake event. However, geodesy measures aseismic deformations. Thus, episodic GNSS observations have been employed densely for geodynamic studies. For occuring very wide variety of

tectonic phenomena, Anatolia and surrounding regions have always been a natural laboratory for tectonic and geodynamic studies. Specific densified GNSS networks were established for monitoring local and regional secular deformation in certain regions in Turkey. Velocity field solution of episodic long-period GNSS data gives information about the tectonic nature of Anatolia and surrounding regions. Shear strain rates, rigid body rotations and faulting areas obtained from strain analyses generally do well conform to the geological and geophysical evidence of Anatolia. Strain analysis implementation is not possible without velocity field generation. And velocity field estimation solely depends on the sequential position estimates. Accordingly, decomposing the time series into their smallest constituents, detecting and cleaning the erroneous signals are crucial for a correct velocity estimation, and correspondingly correct interpretation of tectonic motions.

CGNNS (Continuously Operationg GNSS Stations) data provide much higher accuracy levels than campaign type episodic GNSS measurements, no matter how big the number of measurement epochs is. CGNSS history dates back to early 1990s in Turkey. 146 CGNSS stations with 80-100 km inter-station distances covering all Turkey were established under TNPGN-Active Project. TNPGN-Active stations are fully operational since the second half of 2008 and temporal resolution has reached 5-year duration which allows a detailed time series analysis.

Using geodetic measurements of surface displacement or strain for geophysical studies require accurate error estimates as well as accurate estimation of the parameters. Generally a long time series of geodetic measurements is required to obtain accurate velocity estimates. Thus, a variety of errors with different timescales may corrupt the data.

GNSS observations' noise is generally assumed to be white (uncorrelated). Yet, studies show that the assumption of purely white noise leads to significantly underestimated site rate uncertainties (Zhang et.al., 1997; Mao et.al., 1999). The effect of white noise can be reduced by frequent measurements and averaging. But it is not a remedy especially for random walk type time-correlated noise. While the

motion of large crustal units is seeked, actually the motion of the monument on the ground surface is measured. Therefore, the spurious motion of the monument is an important noise source.

Since the main purpose of TNPGN-Active system is to provide quick and accurate positional corrections to the users on the field especially to meet cadastral needs rather than to serve as a tectonic monitoring system, majority of the stations were built onto the terraces or roofs of governmental buildings. This makes the importance of detecting monument noise for TNPGN-Active sites quite obvious. Other sources of time correlated noise include reference frame effects, mismodelled antenna phase center effects, mismodelled satellite orbits, and mismodelled atmospheric effects. Due to the gross effect of assumed noise type on the resulting rate uncertainty, classification and quantification of the noise components are important parts of driving crustal motion models from GNSS data.

Unmodelled periodic signals (such as oceanal loading and errors in solid Earth tide models) can impact time series on different scales. For example, solid Earth tide might cause up to 0.4 m variations on station coordinates (Lambeck, 1988). Although the state of the art scientific softwares such as GAMIT-GLOBK (King and Bock, 2005; Herring, 2005) and Bernese (Hugentobler et. al., 2005) are used for GNSS processing, models that are used together with these softwares are general, global models, e.g., global ionosphere models, ocean loading models etc. But, particularly designed local models might give better results in regional analyses. Again, correct time series interpretation can promise to expose proficiency or deficiency of global models in local analyses. On the other hand, it is a general approach to use monthly combined daily solutions for the coordinate and velocity estimation of TNPGN-Active stations. In this approach, a rough analysis is being implemented for outlier detection before combining daily solutions monthly. However, outliers which cannot be detected roughly might have an adverse effect on the quality of coordinate and velocity fields. Therefore, sophisticated algorithms must be applied for the detection of the outliers.

Hilbert-Huang Transform (Huang et.al., 1998) is a relatively new technique which is used for the decomposition of nonstationary and nonlinear signals. The prominent property of Hilbert-Huang Transform is its adaptiveness to initial data, in addition to its locality and completeness. HHT consists of Empirical Mode Decomposition (EMD) part in which the signal is decomposed into intrinsic mode functions and the Hilbert Spectral Analysis (HSA) part in which instantaneous frequency data are obtained. In contrast to the Fourier Transform and Wavelet Transform, Intrinsic Mode Functions (IMF) that are obtained from decomposition are determined by an analyzed sequence alone instead of being set analytically. Basic functions are derived adaptively directly from input data. Each IMF contains higher frequency oscillations than the following IMF components. The simple oscillatory mode generally represented by an IMF might be thought as a counterpart to the simple harmonic function. In HSA part, instantaneous frequency variations of the IMF's are examined in order to find clues about the hidden structures.

The adaptive EMD algorithm is deprived of a definite analytical expression, and extra attention should be given to some specific areas of this tecnique. The cubic spline fitting which is used in generating IMFs has overshoot and undershoot problems; the selected stopping criterion can cause IMFs to have uniform amplitudes or to be unacceptably asymmetric; the end swings can gradually influence inside of time series with the on-going sifting process and distort the results; the decomposition method is not capable of seperating signals which have frequencies too close to each other; over-sampled data are needed to define the instantaneous frequency precisely.

In this thesis, HHT is applied to produced GNSS position time series, problematic areas and some solutions to them are investigated, and detrending the time series with the help of EMD is evaluated.

1.2 Literature Review

Weak signal detection is one of the common tasks in signal processing studies. The meaning of weak signal detection is to obtain beneficial weak signal information under the influence of strong noise with the appropriate detection tools. Weak signal detection (WSD) studies spread over a wide application spectrum. Instances might be seen in RADAR, GNSS, earthquake, experimental physics, etc. Traditional signal detection methods require large signal-to-noise ratios (SNR) and lose their effectivities when the objective signal is weak. There are several modern signal processing methods for weak signal detection, such as correlation detection, wavelet analysis, holospectral analysis, Hilbert Huang transform, independence component analysis etc. Generally, the aim is to constrain or eliminate the noise, and thus to enhance the SNR. Three methods rise out of the ruck.

First one is using chaotic oscillators for detecting weak signal (Birx and Pipenberg, 1992). These methods are based on chaos theory. WSD method based on chaotic oscillators makes use of the feature that the oscillator is immune to noise but sensitive to the weak signal having the same frequency with the period of the system driving force to achieve the detection of the weak signal drowned in noise (Wang et.al., 1998). This method can be combined with traditional methods such as correlation method, cross spectra method as well as Fast Fourier Transform (FFT) in which FFT increases the detection ability and the chaotic oscillator improves frequency precision (Li, 2005). Some types of chaotic resonator might suppress noise excellently and they are also sensitive to the amplitude variation of the forcing component in these attractors. Yet, numerical integration such as Runge-Kutta method is needed to solve a nonlinear differential equation which requires considerable amount of time. Moreover, the behaviour of some chaotic attractors are so complex that the results might be misinterpreted.

The second method is applying a linear difference resonator (Qu and Lin, 1999) to detect the weak characteristic frequency component. Due to the widespread numerical computing of differential equations, their properties have been revealed.

This method is based on difference equations of which properties are much more diversified than differential equations. There is much possibility of selecting proper discrete systems as the resonant detectors. Furthermore, a discrete system is used instead of a continuous system represented via differential equations. Therefore, calculations are faster and it is a suitable method to apply on-line.

The third method is based on stochastic resonance theory which can detect a weak signal in the presence of heavy noise from a very short data record. The signal that is normally too weak to be detected, can be boosted by adding white noise to the signal. Stochastic resonance (SR) term was first proposed in 1981 (Benzi et.al., 1981) as a mechanism for almost periodic occurences of ice ages on Earth during the last 700000 years. Since its birth, SR has been attracting many researchers in electronic, biological, quantum and other areas. Many theoretical proposals are awaiting experimental verification.

SR is a type of nonlinear phenomenon in which partial energy of noise can be converted into the energy of signal, and the transmission of weak input signal through the nonlinear system is enhanced. There are three basic conditions of SR. These are bistable nonlinear system, weak input signals and noises. For linear systems, a signal's quality would be deterioriated when white noise is added. But for nonlinear systems, the white noise added to the input signal can improve the signal in such a way that the output signal has a better quality. That is, the response of a nonlinear system to a weak input signal is optimized by the presence of a randomly fluctuating non-zero level of noise. When a weak signal is not powerful enough to trigger a threshold, added noise can boost the signal enough to exceed the threshold value (Dutta et.al., 2006). Moreover, the added white noise can be filtered out later and the original, previously undetectable signal can be detected.

Although the term "decomposition" formally means seperating a compound process into constituent components, in many areas the term is used in the meaning of a breakdown into certain functions which were not actually present at the formation stage of the initial data. Owing to the artificial and desired structure of these

functions, a deeper analysis of data to identify hidden patterns is allowed. The dominant approach is the decomposition of time series into component basis functions satisfying completeness and orthogonality of the basis. Examples are Fourier methods, wavelets analysis, principal component analysis etc.

The Fourier Transform is very common among well known decomposition and spectral analysis methods. It is an orthogonal transformation that uses fixed harmonic basis functions. Since the transform is always performed in a fixed, priorly set basis of orthogonal functions, it does not depend on the nature of the data. Futhermore, the resulting harmonic components have constant amplitude and frequency values over the entire data span. Thus, changes in the nature of the data over the interval under consideration will not be reflected in the transform results. Due to the assumption of stationary initial data, the obtained results only reflect a certain averaged state of the process. The Fourier spectrum loses track of time location for events and it is not a local description. It is most useful when the underlying process is linear and stationary so that the superposition of sinusoidal solutions make physical sense.

To avoid adverse effects of the non-stationarity of the data, Wavelet Transform might be used instead of Fourier Transform. It is especially used for denoising nonlinear signals widely (Durmaz, 2007; Akyay, 2009). In this method, time series is decomposed into local, time-dilated and time-translated wavelet components. In contrast to Fourier Transform, every component resulting from a wavelet transform has parameters that determine its scale and level over time which solves the problem associated with a possible non-stationarity of the process. It is local and provides uniform temporal resolution for all frequency scales, although resolution is limited by the basic wavelet. It is useful for characterizing gradual frequency changes. The drawback of this method is its non-adaptiveness. Same basic wavelet is used for all data. To cope up with this problem, and to further improve the transform efficiency, different adaptation techniques were introduced to traditional wavelet transforms. Sweldens (1996) proposed a new framework for the wavelets construction called "the lifting scheme". Most of the adaptive wavelet transforms use the lifting scheme.

Typically, wavelets with short wavelengths should be used in the neighbourhood of discontinuities or generally higher frequencies in a signal, while wavelets with longer wavelengths should be used for the signal regions where lower frequencies prevail. Different adaptation algorithms were developed for signal denoising problem with varying success (Tuncer, 2006). Due to the available effective implementation algorithms and well established mathematical techniques, the Fourier Transform and Wavelet Transform have been widely applied in many different areas successfully.

Recently, a new method called Hilbert-Huang Transform is gaining growing popularity for the decomposition of nonstationary and nonlinear signals. Rather than being a theoretical tool, the Hilbert-Huang Transform (HHT) is more like an empirical approach in which a signal is decomposed into so-called intrinsic mode functions (IMF).

The HHT has been increasingly applied in many areas such as biomedical applications, neuroscience, image processing, meteorological and atmospheric applications and so on. Huang et al. (2001) used HHT to develop a spectral representation of earthquake data. Chen et al. (2002) used HHT to determine the dispersion curves of seismic surface waves and compared their results to Fourier-based time-frequency analysis. Salisbury and Wimbush (2002), using Southern Oscillation Index (SOI) data, applied the HHT technique to determine whether the SOI data are sufficiently noise free that useful predictions can be made and whether future El Nino southern oscillation events can be predicted from SOI data.

The process of finding abrupt changes (jumps, steps, shifts) in the mean level of a signal is called step detection. The problem of quick detection of abrupt changes occurs in multiple areas such as genetics, biophysics, geophysics etc. Since the time series is corrupted by noise and the step is small in many cases, the difficulty of the problem is high, as the step may be hidden by the noise. Therefore, weak signal detection and step detection methods are not quite independent from each other.

Step detection methods might be classified into two categories depending on the delay of the detection: real-time detection (Adams and MacKay, 2007; Garnett et.al., 2009; Paquet, 2007) and retrospective detection (Basseville and Nikiforov, 1993; Takeuchi and Yamanishi, 2006; Moskvina and Zhigljavsky, 2003). Real-time detection is needed when immediate responses are required and it becomes a special case of sequential analysis where the sample size is not fixed in advance. By contrast, retrospective detection algorithms are applied to the data long after the possession. In stepwise jump displacement algorithm (Gill, 1970), it is assumed that there are no steps in the signal, and possible candidate steps are introduced one at a time. Each candidate is tested for the minimization of some criteria such as the least squares fit of the estimated signal. In sliding window algorithms, a window slides across the time series and the evidence of a step is tested by, e.g., use of two sample Student's t-test. Global algorithms attempt to find the steps by a kind of optimization procedure like wavelet methods (Mallat and Hwang, 1992) and total variation denoising which uses methods from convex optimization. Hidden Markov Models are also often used where the steps can be modelled as a Markov chain.

Having been studied for decades, some pioneer works demonstrated good change-point detection performance by comparing the probability distributions of time-series samples over past and present intervals. When the two distributions are significantly different, the point is regarded as a change point. Various approaches follow this strategy, including the cumulative sum (CUSUM) (Basseville and Nikiforov, 1993), and Generalized Likelihood Ratio (GLR) (Gustafsson, 1996). In these approaches, the logarithm of the likelihood ratio between two consecutive intervals in time series data is monitored. This statistical framework has been extensively explored in connection with real world applications, such as approaches based on novelty detection (Markou and Singh, 2003), maximum-likelihood estimation (Guralnik and Srivastava, 1999), and online learning of autoregressive models (Yamanishi et.al., 2000). Another group of methods gaining high popularity in recent years is the subspace methods (Moskvina and Zhigljavsky, 2003; Ide and Tsuda, 2007; Kawahara et.al., 2007). One of the major approaches is called subspace identification, which compares the subspaces spanned by the columns of an

extended observability matrix generated by a state-space model with system noise (Kawahara et al., 2007).

Yet, the drawback of these approaches is their reliance on some pre-specified parametric models such as probability density models, autoregressive models, and state-space models, for tracking some specific statistics such as the mean, the variance, and the spectrum which limits their capabilities against real-world scenarios. Hence, some non-parametric approaches have been explored in the community of statistics in which non-parametric density estimation is used for calculating the likelihood ratio (Csörgö and Horvath, 1988; Brodsky and Darkhovsky, 1993). However, non-parametric density estimation is a hard problem especially in high-dimensional problems.

Recently, a new strategy was introduced to overcome this weakness in which the ratio of probability densities, i.e., density ratio, is estimated directly, instead of probability densities themselves. Direct density-ratio estimation is substantially easier than density estimation (Sugiyama et.al., 2012). Direct density ratio estimation has been actively explored in the machine learning community, e.g., kernel mean matching (Gretton et.al., 2009), the logistic-regression method (Bickel et.al., 2007), and the Kullback-Leibler importance estimation procedure (KLIEP) (Sugiyama et.al., 2008). In the context of change-point detection, the last method was reported to outperform competitive approaches (Kawahara and Sugiyama, 2012) such as the one-class support vector machine (Schölkopf et.al., 2001; Desobry et.al., 2005) and singular-spectrum analysis (Moskvina and Zhigljavsky, 2003).

1.3 Motivation and the Purpose of This Study

Turkey is located at the intersection of Eurasian, Anatolian and Arabian tectonic plates. Due to the relative motions of these plates, cm-level displacements can occur in site positions. In addition, earthquakes depending on their magnitudes might cause meter-level displacements. Thus, changes in the coordinates of geodetic sites

have to be monitored continuously, and corresponding velocities of these sites must be generated. For this purpose, episodic measurements are carried out at these sites.

Strain analysis based on GNSS velocity field is an important tool used for understanding the structure of the crustial motions. Since episodic GNSS measurements do not supply high temporal resolution, permanent GNSS stations are established to increase the resolution. Ankara Station (ANKR) is the first permanent GNSS station in Turkey and was established in 1991. Many other permanent stations have been established under Turkish National Permanent GNSS Stations Network (TNPGN) project so far. All TNPGN stations are operating on the ground surface. The primary motivation of TNPGN project is to serve for geodynamical studies. Aktuğ et.al. (2009) can be given as an example of these studies.

Permanent GNSS stations are not used only for tectonic monitoring. Recently, a Real Time Kinematic (RTK) GNSS network has been established under TNPGN-Active project. The main purpose of TNPGN-Active system is to provide RTK corrections to the users on the field. New cadastral applications are implemented based upon the coordinates supplied by TNPGN-Active system. Hence, detailed analysis of the coordinate time-series of TNPGN-Active stations is important to achieve better coordinate and velocity estimates.

To be able to detect periodic signals in the time series and for a correct coordinate and velocity estimation, daily GNSS data collected at TNPGN-Active stations should be processed with consistent strategies. One of the leading purposes of this study is to generate consistent GNSS time series of TNPGN-Active stations through high-precision GNSS analysis techniques and herewith to detect the stations owning undesirable time series. Determining the time series showing high Root Mean Square (RMS) values will help to decide whether the locations of the corresponding stations should be changed.

After the generation of daily time series, for the purpose of outlier detection, a visual inspection of the time series is being implemented in General Command of Mapping

(GCM), who is the responsible organization of calculating the accurate positions of TNPGN-Active stations. But the outliers are evaluated subjectively with this approach. It is aimed in this study to propose a practical outlier detection algorithm enabling a quantiative evaluation of the outliers, which can be used in GCM's future analyses.

The Fourier Transform is the most dominant approach in time series analysis. Since Fourier Transform is always performed in a priorly set basis of sines and cosines functions, it does not depend on the nature of the data. It assumes that the systems are linear time-invariant, and the data are stationary. Thus, it is not an appropriate transformation which can be used in GNSS time series.

Recently, Hilbert-Huang Transform (HHT) is gaining popularity for the decomposition of nonstationary and nonlinear signals. It is an adaptive, local and complete transformation. The primary motivation of this study is to show the applicability of HHT to TNPGN-Active time series.

Since HHT is an empirical approach, there are several problematic areas in this method which need extra attention. With this study, it is intended to evaluate the problematic areas in HHT, such as interpolation technique between extrema, end effects, stopping criteria for sifting process etc., and to find solutions for some of the shortcomings. It is also aimed to investigate the feasibility of HHT in detrending the time series.

1.4 Thesis Outline

This thesis consists of six chapters. In Chapter-1, the scope and the motivation of the study are explained and a literature review is given. Chapter-2 is dedicated to theoretical bases of GNSS data analysis. Outlier detection methods are presented in Chapter-3 very briefly. In Chapter-4, characteristics of Empirical Mode Decomposition are unfolded and the Hilbert Spectral Analysis is elucidated. Chapter-5 forms the main body of this thesis. The features of the data used in this

study are briefed, GNSS data processing software is presented, implementation of the high-precision GNSS analysis techniques are explained, obtained time series are illustrated, a practical outlier detection algorithm is presented, application of EMD and HSA to TNPGN-Active time series is explained, and problematic areas of HHT are investigated. A discussion-style concluding remarks are given in Chapter-6 together with the issues which will be addressed in future studies.

CHAPTER 2

HIGH-PRECISION GNSS DATA ANALYSIS

2.1 GNSS Data Processing Software

GNSS data processing softwares are results of extensive geodetic research generally carried out by universities or research laboratories. Such softwares typically include; accurate observation models, orbit integration, detecting cycle slips and outliers, estimation of station coordinates, orbits and tropospheric bias, ambiguity resolution algorithms, and estimation of transformation parameters between reference frames.

There are many packages developed for scientific purposes, but three high-precision sofwares are widely used by researchers:

- GAMIT-GLOBK software, Massachusetts Institute of Technology (MIT), USA (King and Bock, 2005)
- BERNESE software, Astronomical Institute, University of Bern, Switzerland (Rothacher et al., 1990)
- GIPSY-OASIS II software, Jet Propulsion Laboratory (JPL), California Institute of Technology, USA (Webb and Zumberge, 1993)

2.2 Temporal Resolution

Annual signals can significantly bias estimation of site velocities, especially when the GNSS time series is shorter than 2.5 years (Blewitt and Lavallee, 2002). Below 2.5 years, velocity estimates become unstable due to correlated parameters and even the simultaneous estimation of annual and semiannual sinusoidal signals together

with the site velocity and initial position does not always improve velocity estimates. On the other hand, beyond 4.5 years, velocity bias becomes negligible rapidly and estimation of annual and semiannual signals may not be necessary. In Figure 2.1, velocity bias from an annual sinusoidal signal versus data span can be seen.

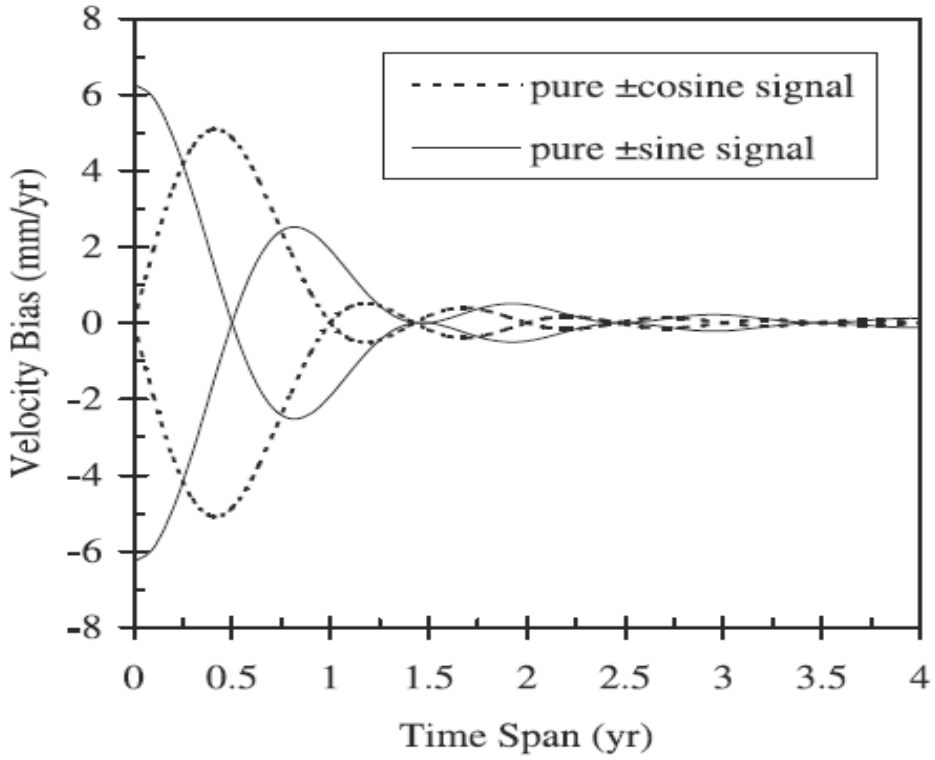

Figure 2.1 Velocity bias from an annual sinusoidal signal versus data span for cosine and sine signals (Blewitt and Lavallee, 2002).

2.3 GPS Observables

Basic principle of GPS positioning is to determine positions by measuring distances between receivers and the satellites in view whose coordinates are known. Since GPS receivers cannot measure these distances directly, these observations are called "pseudoranges". Time difference between the receiver's local clock and satellite's atomic clock is multiplied by the speed of light to obtain pseudorange in meters.

300-m-wavelength CA (coarse acquisiton) code and 30-m-wavelength P (protected) code measurements are used in the calculation of pseudoranges by performing cross-correlation between the incoming signal and the replica signal generated in the receiver. This way, required time shift is computed to align codes. Pseudoranges are not used alone in geodetic surveys due to their lack of precision. In this thesis, they are used to synchronize receiver clocks, resolve ambiguities and repair cycle slips in phase observations.

High-precision geodetic applications require sophisticated GPS receivers that can measure carrier phase observable in addition to the pseudorange observable. All TNPGN-Active stations are equipped with such receivers. This observable is the difference between the phase of the incoming carrier wave and the phase of a signal generated by the receiver internally. This difference is multiplied by the wavelength of the carrier wave and a biased distance to the satellite is obtained. Actually this is a kind of pseudorange which is much more precise than the coded pseudoranges. Yet, the disadvantage of this observable, in addition to the receiver clock bias, is that the integer number of wavelengths is not known. It is essential to resolve this bias by ambiguity resolution techniques.

Apart from the fact that carrier phase measurement is much more precise (about 100 times) than the pseudorange measurement, the only difference between them is the integer-wavelength ambiguity. Because of the similarity of models, the pseudorange data can be used together with the carrier phase data to correct for the initial ambiguity (Blewitt, 1989) and for subsequent cycle slips (Blewitt, 1990).

Satellite clock bias is eliminated by differencing the phases or pseudoranges of signals received simultaneously between a single satellite and two ground stations. This is called single-differencing. It also reduces the tropospheric and ionospheric refraction effects if the stations are close to each other (less than 30 km). However, the receiver clock bias is still unknown. Forming a double-difference by differencing the between-station differences also between satellites cancels receiver clock bias and there is no need to estimate it as a parameter. Double-differencing is based upon

two factors: at any time (1) clock bias at a specific satellite is same for all stations, (2) clock bias at a specific station is same for all satellites being tracked. Later on, phase ambiguities are resolved and an unambiguous measure of doubly differenced range is obtained.

There are three different approaches for estimating positions in the literature. They all give results with acceptable geodetic quality and results typically agree well. These are:

- Precise Point Positioning (PPP) of single stations.
- Relative positioning of networks by clock estimation.
- Relative positioning of networks by double-differenced data.

All approaches need precise orbit data while PPP also needs precise clock data.

When an interruption of the signal happens, it causes the phase exhibit a discontinuity of an integer number of cycles which is called "cycle-slip". In the presence of cycle slips, initial processing of phase data might be performed by triple-differencing in which differences of doubly differenced phase between epochs are used. It is a useful method to obtain preliminary estimate of station or orbital parameters. However, with triple-differencing, the degradation in precision is substantial. Thus, for precise applications, triple-difference data are inappropriate.

Delay induced by ionosphere is a major error source in single-frequency GPS measurement. But, the ionospheric refraction bias can be reduced to a millimeter or less by constructing a combined ionosphere-free phase observable from L1 (1575.42 Mhz) and L2 (1227.6 Mhz) data. This is called LC or sometimes L3 combination. The equation for a linear combination is:

$$\Phi = \alpha_1 \Phi_1 + \alpha_2 \Phi_2 \,, \qquad (2.1)$$

where α_1 and α_2 are arbitrary linear coefficients. For LC combination, these coefficients are chosen such a way that the ionospherical influence vanishes:

$$\alpha_1 = \frac{f_1^2}{f_1^2 - f_2^2} \approx 2.546 \text{ and } \alpha_2 = \frac{-f_1 f_2}{f_1^2 - f_2^2} \approx -1.984 \quad (2.2)$$

with f_1 and f_2 frequencies of L1 and L2 signals. So, the phase of LC combination is (measured in units of L1 wavelengths):

$$\Phi_{LC} = 2.546\Phi_{L1} - 1.984\Phi_{L2} \quad (2.3)$$

(Bock et.al.,1986). However, LC combinations have approximately three times the noise of the L1 observations. Thus, treating L1 and L2 as two independent observables is a better approach than forming linear combinations for short baselines (less than a few kilometers) since ionospheric errors largely cancel.

It is also useful to examine several combinations of L1 and L2 residuals. In LC, single-cycle slips in L1 appear as jumps of 2.546 cycles while in L2 they appear as jumps of 1.984 cycles. When single-cycle slips happen in both L1 and L2, they appear as jumps of 0.562 cycles in LC. Another combination called "LG" is also useful:

$$\Phi_{LG} = \Phi_{L2} - 0.779\Phi_{L1} \, . \quad (2.4)$$

LG can be used to measure ionospheric variations, since all geometrical and non-dispersive delays are eliminated.

The difference between L1 and L2 phases with the L2 phase scaled to L1 wavelength is called "widelane" and used to fix cycle slips. But, it is affected by fluctuations in the ionospheric delay. On the other hand, Melbourne-Wübbena linear combination (sometimes called L6) is free of ionospheric, tropospheric and geometric effects:

$$MW - WL = \Phi_{L2} - \Phi_{L1} + (P1 + P2)(f1 - f2)/(f1 + f2)\ . \quad (2.5)$$

The term P denotes the pseudoranges.

2.4 Orbit Models

GPS positioning is based on the principle of determining position by measuring distances to points of known positions, i.e., to satellites. Thus, exploiting an accurate model of the satellite's motion is significant. Gravitational (Sun, Moon, Earth's gravity field) and non-gravitational (e.g. solar radiation pressure) forces perturb the motion of the GPS satellites and causes the orbits to deviate from a Keplerian ellipse in inertial space.

Satellite motions are described in a space-fixed inertial (in fact quasi-inertial; changes in the motion behaviour of the Earth around the Sun causes the system to be quasi-inertial one, because of the constant gravitational effects of the Sun, Moon and other celestial bodies it is not possible to establish an inertial frame at all, neither on the earth nor in the space) system while positions of ground points are determined in an Earth-fixed non-inertial system. Furthermore, GPS measurements are carried out in a local horizon system.

There are three main temporal changes in the space-fixed system, namely precession, nutation and polar motion. Precession and nutation results from the gravitational forces applied by the Sun, Moon and the other planets acting on the bulge of the earth. Consequently rotation axis of the Earth is not fixed in space. It moves on a cone around the normal of the ecliptic plane. In polar motion, on the contrary of the precession and nutation, Earth rotation axis move with respect to the Earth surface. In an approximate sense, precession and nutation are caused by external torques applied to Earth and polar motion happens due to the mass and angular momentum redistributions and the elasticity of Earth. It should be also considered that the rotational rate of the Earth is decreasing due to the energy dissipation in the space.

2.5 Parameter Estimation

2.5.1 Least-Squares Estimation

To solve the parameters with the help of the measurements, a mathematical relationship must be defined between the measurements and the parameters. In most cases, there exists no direct relationship between the unknown parameters and the observations, that is, the unknown parameters are not measured directly. Such a model is given by the following linear equation:

$$\boldsymbol{y_{nx1}} + \boldsymbol{e_{nx1}} = \boldsymbol{X_{nxu}\beta_{nx1}} \,. \tag{2.6}$$

The stochastic part of this model is given by:

$$\boldsymbol{D(y)} = \sigma^2 \boldsymbol{P}^{-1} \tag{2.7}$$

where $\boldsymbol{D(y)}$ is the covariance matrix of the observations, $\boldsymbol{P}$ is the positive definite data weight matrix and σ^2 is the variance of unit weight. In the adjustment theory, $\boldsymbol{P}^{-1}$ is also called cofactor matrix of the observations. Equation (2.6), together with Equation (2.7), is called Gauss-Markoff Model. For GNSS positioning, where a non-linear relationship between the observations and the parameters exists, Equation (2.6) can be written in the following form after the linearization:

$$\boldsymbol{\Delta y_{nx1}} + \boldsymbol{e_{nx1}} = \boldsymbol{X_{nxu}\Delta\beta_{nx1}} \,. \tag{2.8}$$

Here, the reduced observations, in other words *prefit residuals*, are listed in the $\boldsymbol{\Delta y}$ vector, which has dimensions *nx1*, where *n* is the number of linearly independent double differenced data. The configuration matrix $\boldsymbol{X}$ has dimensions *nxu* where *u* is the number of parameters, and the parameter corrections are contained in the $\boldsymbol{\Delta\beta}$ vector, which has dimensions *nx1*. The random observation errors are represented by the $\boldsymbol{e}$ vector, which has the same dimensionality as $\boldsymbol{\Delta y}$. The configuration or coefficient matrix $\boldsymbol{X}$ is formed by taking the partial derivative of the function

according to the model parameters. Using the least squares estimation method, corrections to the model parameters are calculated as follows (details can be found in Koch, (1999)):

$$\boldsymbol{\Delta\hat{\beta}} = (\boldsymbol{X^T P X})^{-1}\boldsymbol{X^T P \Delta y} \,. \tag{2.9}$$

Later on, approximate values of the parameters, $\boldsymbol{\beta}_0$, can be updated and the adjusted parameters are obtained:

$$\boldsymbol{\hat{\beta}} = \boldsymbol{\beta}_0 + \Delta\boldsymbol{\hat{\beta}} \,. \tag{2.10}$$

Then, corrected observations are computed:

$$\boldsymbol{\hat{y}} = \boldsymbol{X\hat{\beta}} \,. \tag{2.11}$$

The estimated or *postfit residuals* are obtained by subtracting corrected observations from original observations:

$$\boldsymbol{\hat{e}} = \boldsymbol{y} - \boldsymbol{\hat{y}} \,. \tag{2.12}$$

The conventional measure of goodness-of-fit is the estimated variance scale factor ($\hat{\sigma}_0^2$) and calculated by using estimated residuals and usually normalized by dividing by the "degrees of freedom":

$$\hat{\sigma}_0^2 = \frac{\boldsymbol{\hat{e}^T P \hat{e}}}{n-u} \,. \tag{2.13}$$

The ideal value of $\hat{\sigma}_0^2$ for properly weighted, independent random observations is 1.0. Indeed, we assume that we know the correct observation covariances beforehand. If this assumption is true, $\boldsymbol{\hat{e}^T P \hat{e}}$ random variable would be χ^2 (chi-square) variable and the value of this variable would be equal to the degrees of freedom (*n*-*u*) of the adjustment. If $\hat{\sigma}_0^2$ > 1, either there are rough errors in the

observations or variances of observations are overestimated. If $\hat{\sigma}_0^2 < 1$, it means that variances of observations are actually better than we assume. The covariance matrix for the estimated parameters is given by:

$$\boldsymbol{D}(\widehat{\boldsymbol{\beta}}) = \hat{\sigma}_0^2(\boldsymbol{X^T P X})^{-1} . \quad (2.14)$$

With white noise, these formal uncertainties are realistic and will be better as the number of the observations (sampling interval is significant) increases.

However, the uncertainties cannot be treated with white noise statistics in a GPS analysis. Because time-correlations dominate phase observations. When sky plots generated by GAMIT are examined, it is seen that the visible noise in the phase residuals due to multipath and troposheric fluctuations is typically correlated over 15-30 minutes. The receiver can measure the phase with a precision better than a mm, but unmodelled errors such as multipath cause post-fit residuals show several mm standard deviations. There are also errors with longer correlation times absorbed into the parameter adjustments that cannot be seen in the residuals (Mao et.al., 1999).

2.5.2 The Kalman Filter

The basic Kalman filter technique was invented by Kalman (1960). For a more detailed treatment of Kalman filters, Brown and Hwang (1997), Grewal and Andrews (2000) can be consulted.

The Kalman filter uses knowledge of the deterministic and statistical properties of the system parameters and the measurements. In addition to the parameter estimates, it also maintains a set of uncertainties in its estimates and a measure of the correlations between the errors in the estimates of the different parameters (Groves, 2008).

For the purpose of reference frame definition and time series generation, produced daily GAMIT solutions of TNPGN-Active data (the resulting least squares adjustment vectors and corresponding variance-covariance matrices) were combined with the global solutions produced by SOPAC (Scripps Orbit and Permanent Array Center) analysis center by using GLOBK in this thesis. GLOBK uses Kalman filter method. In this section, utilization of Kalman filter in data combination is explained based on (Dong et.al., 1998).

Let $\boldsymbol{\delta I_k = I_k - I_{0k}}$ be the vector of linearized quasi-observations at time t_k, where $\boldsymbol{I_{0k}}$ represents the a priori values and $\boldsymbol{I_k}$ the estimates of the quasi-observations from the analysis of the original observations. The observation equation is

$$\boldsymbol{\delta I_k = A_k \delta x_k + \varepsilon_k} \tag{2.15}$$

in which $\boldsymbol{x_k}$ are the parameters to be estimated, $\boldsymbol{A_k}$ is the design matrix of partial derivatives, and $\boldsymbol{\varepsilon_k}$ is a zero-mean, white-noise process with covariance $\boldsymbol{P_k}$. The parameter state is represented by

$$\boldsymbol{\delta x_{k+1} = S_k \delta x_k + q_k} \ , \tag{2.16}$$

where $\boldsymbol{S_k}$ is the state-transition matrix representing the dynamic evolution of the parameters and $\boldsymbol{q_k}$ is a Markov stochastic process with covariance $\boldsymbol{Q_k}$. To estimate the parameters at time t_{k+1}, one first propagates the estimates and their covariance,

$$\boldsymbol{\delta \hat{x}_{k+1|k} = S_k \delta \hat{x}_k} \ , \tag{2.17a}$$

$$\boldsymbol{C_{k+1|k} = S_k C_k S_k^T + Q_k} \ , \tag{2.17b}$$

then updates the estimates using the current observations,

$$\boldsymbol{\delta \hat{x}_{k+1} = \delta \hat{x}_{k+1|k} + K_{k+1}(\delta I_{k+1} - A_{k+1} \delta \hat{x}_{k+1|k})} \ , \tag{2.18a}$$

$$\boldsymbol{C_{k+1} = C_{k+1|k} - K_{k+1} A_{k+1} C_{k+1|k}} \quad , \tag{2.18b}$$

where the "Kalman gain" matrix is,

$$\boldsymbol{K_{k+1} = C_{k+1|k} A_{k+1}^{T} (P_{k+1} + A_{k+1} C_{k+1|k} A_{k+1}^{T})^{-1}} \quad . \tag{2.19}$$

In these equations, $k+1 \mid k$ denotes the prediction at t_{k+1} using data through t_k. For a detailed explanation of this method, see Herring et.al. (1990), Dong et.al. (1998).

2.6 Reference Frame Definition

Defining a reference frame is necessary to obtain meaningful site coordinates, and thereby crustal motion. The main idea is to align the estimated site positions to a set of well defined locations that have physical significance. There are two options in GLOBK to impose constraints on the solution. In the first option, with *finite* constraints, realistic a priori uncertainties are assigned to the coordinates of one or more stations. This approach is most useful when the study area is very local or there are only one or two reference sites are available. Because, constraining aditional sites can distort the network, although provides redundancy.

In the second approach, frame definition with *generalized* constraints (Dong et.al., 1998), reference sites with good a priori values are chosen as many as possible. Residuals of these sites are minimized while estimating translation, rotation and scale parameters. Since all reference coordinates are free to adjust, outliers can be easily detected and removed. Network can translate and rotate, but not distort. This approach works best with strong redundancy.

CHAPTER 3

OUTLIER DETECTION AND IRLS FOR GNSS TIME SERIES

The concept of an outlier was defined by Hawkins (1980) as follows:

"An outlier is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism."

The determination of the level of this "deviation" in long-term GNSS daily position time series entails elaborate assessment. Time series of real world phenomena always provide us a limited time-window for analysis. The observation (the daily position estimate in this case) that we accept as an outlier, might be a part of a longer-period signal for which we have not reached the necessary time resolution yet to comprehend. Likewise, an outlier detected towards the end of a time series might just be the beginning of an incident, like e.g., of an earthquake.

On the other hand, the uncertainty of the position estimate is as decisive as the deviation from other position estimates when labelling a position estimate as an outlier. While higher uncertainty rates are anticipated from the estimates which deviate from the trend of the other estimates, this is not always the case. Thus, if the uncertainty rate of a deviated estimate is low, one should think twice to condemn it as an outlier.

The basic outlier models can be categorized as; extreme value analysis, probabilistic and statistical models, linear models (linear regression, principal component analysis etc.), proximity-based models (distance-based, density-based etc.), information theoretic models, and high-dimensional outlier detection. Details about these models

can be found in, e.g., Aggarwal (2013). In this section, only *Z-value test* is explained briefly which is a test used in extreme value analysis.

The level of outlierness of a data point is generally indicated by a score which gives an idea about the outlier tendency of that point. One of the common and simple model used in outlier analysis is *Z-value test* (Aggarwal, 2013). The *Z-value* for a data point X_i is denoted by Z_i, and is defined as follows:

$$Z_i = \frac{|X_i - \mu|}{\sigma}, \tag{3.1}$$

where μ is the mean and σ is the standard deviation of the observations. The *Z-value* test computes the number of standard deviations by which the data varies from the mean. It is implicitly assumed that the data is modeled from a normal distribution. An ubiquitous rule is to use $Z_i \geq 3$ as a detection tool for the outlier.

However, it cannot be assumed that the daily position estimates are normally distributed in a GNSS time-series. Nevertheless, the data can be modeled into lower dimensional embedded subspaces with the use of linear correlations (Rousseeuw and Leroy, 2003).

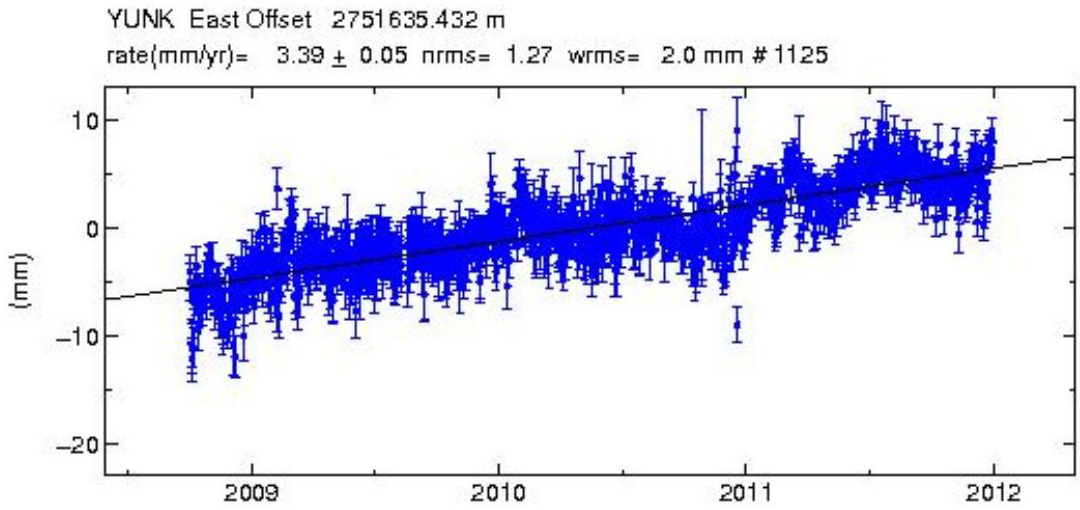

Figure 3.1 The East component of the time series of YUNK station.

In Figure 3.1, the east component of the time series of YUNK station is given. It is apparent that the data are aligned along a 1-dimensional line in a 2-dimensional space. The optimal line which intersects these points can be obtained by using regression analysis. This line typically is obtained by a least squares fit. Once the optimal line is obtained, extreme value analysis can be applied on the deviations from this line in order to determine the outliers.

In linear regression to a line, a model of the form $y_i = \beta_0 + x\beta_i$ (polynomial of degree 1) is seeked that best fits a set of data points. The unknown parameters are the polynomial coefficients. This polynomial model can be designed as a special Gauss-Markoff model (Koch, 1999) for parameter estimation purposes.

When the number of the observations (measurements) is bigger than the number of the parameters (unknowns) in a full rank system, there is not only one solution that satisfies all equations. Yet, a particular approximate solution may be found that minimizes the discrepancies between the observations and the model predictions.

Least squares solution minimizes the L_2 – *norm* of the residual. It is statistically the most likely solution if observational errors are independent and normally distributed. This minimizing procedure can be represented analytically as follows:

$$\min \sum_{i=1}^{m} \frac{(y_i - f(\beta)_i)^2}{{\sigma_i}^2}, \tag{3.2}$$

where $\boldsymbol{y}$ is the data vector, $\boldsymbol{\beta}$ is the vector of model parameters and f is the function that relates the parameters to the data.

But if there is a big outlier, least squares solution (L_2 – *norm*) can mislead to the wrong direction. When an outlier occurs, least squares solution takes into account the relative contribution of that observation to the total likelihood. In such cases, the L_1 – *norm* solution might be useful.

Instead of minimizing the L_2 – *norm*, L_1 – *norm* solution minimizes:

$$\sum_{i=1}^{m} \frac{|y_i - f(\beta)_i|}{\sigma_i} \ . \tag{3.3}$$

The *L1-norm* solution is more robust than the *L2-norm* solution, because it does not square the residuals. Minimizing the $|y_i - f(\beta)_i|$ can be accomplished by IRLS (Iteratively Rewighted Least Squares) method (Scales et.al., 1988). It solves a sequence of weighted least squares problems whose solutions converge to a *L1-norm* minimizing the solution:

$$\boldsymbol{X^T R X \beta = X^T R y} \ , \tag{3.4}$$

where $\boldsymbol{X}$ is the configuration matrix, and $\boldsymbol{R}$ is a diagonal weighting matrix corresponding to the $\boldsymbol{P}$ matrix in Gauss-Markoff model (see Section 2.5.1), with $R_{i,i} = 1/|r_i|$.

This is a nonlinear system of equations which cannot be solved directly. A simple iterative scheme can be used to find the appropriate weights. The IRLS algorithm begins with the least squares solution $\boldsymbol{\beta^0 = \beta L_2}$**.** Then the initial residual vector $\boldsymbol{r^0 = y - X\beta^0}$ is calculated. Later, $\boldsymbol{X^T R X \beta = X^T R y}$ is solved to obtain a new model $\boldsymbol{\beta_1}$ and residual vector $\boldsymbol{r_1}$. The process is repeated until the model and residual vector stabilize.

CHAPTER 4

HILBERT-HUANG TRANSFORM

4.1 Motivation

Currently, Fourier spectral analysis is the most dominant approach for data analysis due to its simplicity and versatility. Signals are decomposed into a series of sines and cosines in this approach. However, there are several inherent assumptions and restrictions of the Fourier analysis. It assumes that the systems are linear time-invariant (LTI), and the data are stationary. *Linear* systems ensure the *scaling* and the *superposition* conditions, i.e., input $ax_1(t) + bx_2(t)$ produces output $ay_1(t) + by_2(t)$ where *a* and *b* are real scalars. In *time-invariant* systems, the output does not depend on when the input is applied. The basis functions of the transforms (sines and cosines in Fourier case) for a LTI system are always complex exponentials. Considering that the sinusoids are a sum of complex exponentials, a sinusoid input to the system produces a sinusoid output. Although the output can be of different amplitude and phase, it is not possible to produce frequency components that are not present in the input.

A *stationary* signal possesses constant statistics (such as mean and variance) across each time segments. For a discrete time signal, Fourier Transform (FT) computes a coefficient for each data point and each non-zero Fourier coefficient shows the existence of an uniform harmonic component globally (from t=-∞ to t=+∞) rather than locally that is present just in the time interval of the signal. Thus, FT assumes a stationary signal with constant statistics across broad time segments.

Yet, rarely linear and/or stationary are real world signals. Even if a process is stationary, the observed data span is almost always shorter than the longest time scale of the process. Consequently, the process appears non-stationary.

Linear and stationary assumptions of Fourier spectral analysis necessiate many additional and spurious harmonics to simulate non-stationary and/or non-linear data. This causes energy spreading over a wide frequency range. Hence, energy-frequency distribution of such data may be misleading.

In order to lessen the disadvantages (especially lack of timing information) of global FT, Short-time Fourier Transform (STFT) – so-called spectrogram – can be used in which the data are assumed piecewise stationary. FT is applied in a limited time window which slides along the time axis to get a time-frequency distribution. However, it is not known whether the adopted window size overlaps stationary time scales. In addition, there is a trade-off between the resolution in time and in frequency. A narrow window enhances frequency resolution, while necessiates longer time series. The wavelet analysis (an adjustable window Fourier spectral analysis), the Wigner-Ville distribution (Fourier transform of the central covariance function) and Principal Component Analysis are other widely used methods for data analysis. They all try to modify the global representation of Fourier analysis, but none of them completely could succeed it.

In Hilbert-Huang Transform (HHT) method, the data are first preprocessed by the empirical mode decomposition (EMD) method to produce intrinsic mode functions. These functions form the bases (which are complete, local and adaptive) in which the data are expanded. Since they are derived from the data, they are a posteriori (not like predefined a priori sines and cosines). Later on, Hilbert Transform (HT) is applied to the intrinsic mode functions to obtain energy-frequency-time distribution (Hilbert spectrum). Instantaneous frequency and energy are needed to construct Hilbert spectrum rather than global frequency and energy which are used to construct Fourier spectrum.

EMD method needs discussions concerning particularly the problems with respect to the interpolation technique, stopping criteria and end effects in context of the sifting process.

4.2 Empirical Mode Decomposition (EMD) Method

To construct Hilbert spectrum, instantaneous frequency and energy are needed. Although there are other definitions of instantaneous frequency, the commonly accepted definition is "the derivative of the phase of the analytic signal":

$$w = \frac{d\theta(t)}{dt}, \tag{4.1}$$

where w is the instantaneous frequency and θ is the phase angle. How to make the data analytical and to obtain intantaneous properties will be explained in Section 4.4.1.

In Fourier analysis, the frequency is defined for sine and cosine functions globally with constant amplitude. A truncated sinusoid requires additional frequency components. Hence, at least one full oscillation of a sine or cosine wave is needed to define the frequency. However, GNSS time series are non-stationary and the frequency changes intermittently. This situation dictates a local study of the frequency content.

For any function to have a meaningful instantaneous frequency, the real part of its Fourier transform has to have only positive frequency (Gabor, 1946; Bedrosian, 1963; Boashash 1992). In this context, examination of a simple $x(t) = \alpha + sint$ function leaded Huang et.al. (1998) to the definition of a class of functions, that is, Intrinsic Mode Functions (IMFs). The Fourier spectrum has a DC term due to the α term. If $\alpha > 1$, the phase function and the instantaneous frequency (which are computed with the help of Hilbert Transform) have meaningless negative values. Thus, the function should be restricted to be symetric locally with respect to the zero

mean level (α = 0) to be able to define the instantaneous frequency. This restriction set off the foundation of a method to decompose the data into components, namely IMFs, for which the instantaneous frequency can be defined.

An intrinsic mode function (IMF) is a function that satisfies two conditions: (1) in the whole data set, the number of extrema and the number of zero crossings must either be equal or differ at most by one; and (2) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero (Huang et.al., 1998).

According to the statements above, it can be concluded that an IMF involves only one mode of oscillation and can be both amplitude and frequency modulated. The second condition is obtained by removing the dissymetry between the upper and the lower envelopes iteratively which is called "sifting process". This process transforms the original signal into an AM signal. As the instantaneous frequency can change from time to time, each IMF is an amplitude/frequency modulated (AM/FM) signal. During sifting, low frequency components are removed and the highest frequency is left. Consequently, the IMFs appear in a high to low frequency order.

4.3 Sifting Process

Given a signal $x(t)$, the IMFs are found by an iterative procedure which can be summarized as follows (Huang et.al., 1998);

1. Find all the local maxima and minima in $x(t)$,
2. Interpolate between minima and between maxima using cubic spline ending up lower envelope $e_{min}(t)$ and upper envelope $e_{max}(t)$ respectively.
3. Compute the mean of the envelopes: $m(t) = (e_{min}(t) + e_{max}(t))/2$.
4. Subtract *m(t)* from the signal to obtain the detail: $x(t) = x(t) - m(t)$.
5. Repeat step (1) to (4), until $x(t)$ can be considered as zero mean according to some stopping criterion (sifting). Once this is achieved, the detail $x(t)$ is referred to as an IMF ($i(t)$).

6. Remove the obtained IMF from the signal: $x(t) = x(t) - i(t)$ and return to step (1) if $x(t)$ has more than one extremum (neither a constant nor a trend).

The number of extrema is decreased through subsequent IMFs and the whole decomposition is expected to be completed with a finite number of IMFs.

The sifting process can be stopped either when the residue becomes a monotonic function or a function containing only one internal extremum from which no more IMF can be extracted. The final residue should be the trend, if the data contains a trend. Finally the EMD of the data $x(t)$ can be expressed mathematically as:

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) , \quad (4.2)$$

where c_i is the ith IMF and r_n is the residue.

4.4 Detrending Signals with EMD

As can be followed through the EMD figures in this thesis, a zero reference is not needed to apply EMD to an arbitrary data. The sifting process generates zero references for IMFs. The Hilbert spectrum has an advantage that it is possible not to include the obtained residue (last IMF component) in its calculation. On the other hand, situation is not so simple for the ubiquitous Fourier Transform, if it is applied to the data without zero reference. To avoid the large DC term in data with non-zero mean, the mean values should be removed first. It is common to remove the best straight-line fit linear trend from the data before applying Fourier Transform (FT). The reason of detrending the data is to make the time series stationary, which is a pre-condition to apply FT to a time series. Evidently, it is controversial how well a straight line can represent the real trend of a non-stationary process. On the other hand, non-stationarity has not always to do with the trend of the data.

One of the important properties of EMD algorithm is its completeness. That is, according to Equation (4.2), the original data can be reconstructed perfectly by

adding all IMF components to each other successively. This property makes detrending the data possible with the help of EMD process.

For detrending the data, the method proposed in Flandrin et.al. (2004) was used in the thesis. If the analyzed signal *x(t)* consists of a trend and a fluctuating process *y(t)*, the trend is expected to be captured by the lower frequency components together with the residue. Thus, detrending *x(t)* amounts to computing the partial, fine-to-coarse, reconstruction

$$\hat{y}_D(t) = \sum_{k=1}^{D} d_k(t) \,, \tag{4.3}$$

where *D* is the IMF index prior to contamination by the trend. Each of the IMFs $\{d_k(t); k=1,\ldots,D\}$ being zero mean, a rule of thumb for choosing *D* is to observe the evolution of the standardized empirical mean of $\hat{y}_d(t)$ as a function of a test order *d*, and to identify for which $d = D$ it departs significantly from zero (Flandrin et.al., 2004). *D* is accepted as the change point, and IMF components which have higher indices than *D* form the trend of the data together.

4.5 Problems in HHT

Since HHT is an empirical approach, specific parts of this approach requires further attention. These parts are explained in this section.

Spline fitting between the extrema points should be improved. After first round of sifting, there can be positive minima and negative maxima. This situation implies riding waves, and renders the obtained signal still not an IMF. Ideally it should be an IMF, but an important drawback of cubic spline interpolation takes the stage: overshoots and undershoots. They can generate new extrema, and shift or exaggrate existing ones. The original signal might cross the envelope. This affects the IMF estimation. Overall, cubic splines are recommended to be preferred in Huang et.al. (1998). Many prominent studies also favor cubic spline interpolation. Other types of interpolation (linear or polynomial) tend to increase the required number of sifting

iterations and to over-decompose signals by spreading out their components over adjacent modes (Rilling et.al., 2003).

Recently, several studies emerged which propose to apply constrained cubic spline in EMD instead of natural cubic spline to eliminate overshoots in the upper and lower envelopes. An example can be found in, Kokes and Ba (2011). The constrained cubic spline was proposed in Kruger (2002) for chemical engineering applications.

Traditionally, the natural cubic spline function is constructed based on the criterion that the second order derivative is the same for both functions on either side of a point. The principle behind the proposed cubic spline is to prevent overshooting by sacrificing smoothness. This is achieved by eliminating the requirement for equal second order derivatives at every point and replacing it with specified first order derivatives. How to calculate the specified first order derivatives for intermediate and end points can be found in Kruger (2002).

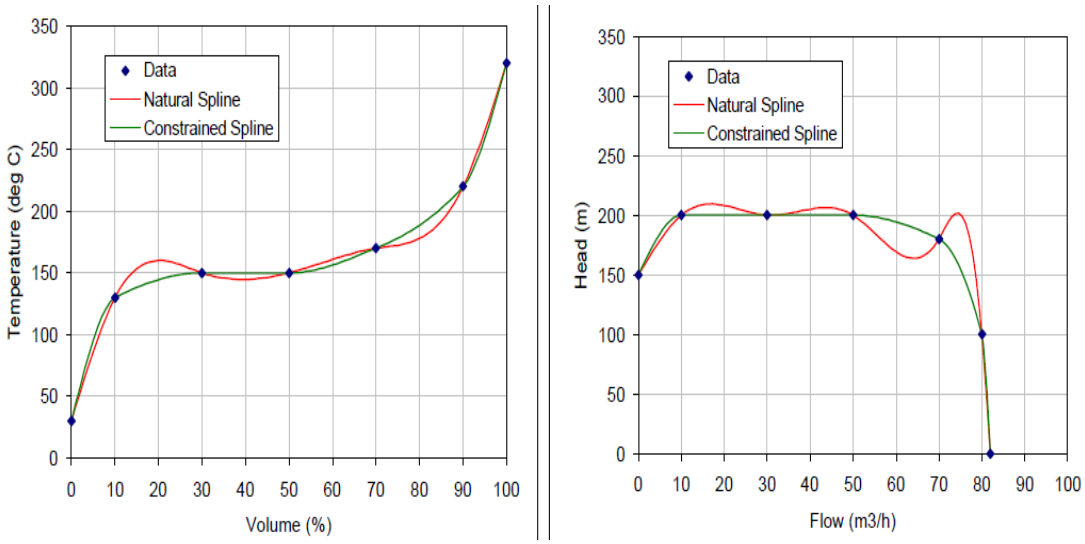

Figure 4.1 Comparison of natural and constrained cubic spline interpolations (Kruger, 2002).

In Figure 4.1, the illustration of the comparison between the natural and constrained cubic spline interpolations by using some data cooked-up ahead of time is given (Kruger, 2002). It is seen that the constrained cubic spline overcomes the overshoots and undershoots.

A tedious situation occurs near the ends of the time series when forming the upper and lower envelopes with cubic spline function. Several border conditions can be used. For example, the second order derivative of the splines at the end points can be assumed zero, or the continuity of the third order derivatives at the second and penultimate points can be forced. The latter results in the first two and last two intervals having the same spline function. However, the farther the first (last) maximum (minimum) point from the end point is, the more likely the end swings become large. The end swings can gradually influence inside of time series with the on-going sifting process and distort the results, especially in low-frequency components. For longer (oversampled) TNPGN-Active daily time-series, it is a good choice to throw away the first and the last segments which can be affected by distortions in each sifting. But this choice is almost non-existing for monthly combined time series with ~30 times less data points. Obviously, the known first derivatives at the first and last points (known as clamped end conditions) would decrease the level of the swings.

In Huang (1998), extending the original series by using characteristic wave method is mentioned. Rilling et.al. (2003) proposed to mirrorize the extrema closest to the edge. Chiew et.al. (2005) used the average of the two closest maxima (minima) for the upper and lower envelopes. Datig and Schlurmann (2004) proposed a method based on the slopes of maxima and minima near both ends of the time series.

For the analysis of TNPGN-Active time series, the first and the last samples were considered as maxima and minima simultaneously. This forces all the IMFs to be zero at the ends. Seeing that the essential scales could be extracted from the data successfully, endeavor for extending the signal properly was kept for a future study.

Sifting eliminates the riding waves and makes the wave profiles more symmetric which are essential conditions to obtain meaningful instantaneous frequency. With this aspect, it would be desirable to sift as many times as needed. However, it also smoothes the uneven amplitudes. This causes a decline in the range of variations in the IMF, and removes physically meaningful amplitude fluctuations. Hence, the number of the siftings should be limited to preserve the natural amplitude variations. To obtain physically meaningful IMF components without constant amplitude frequency modulations, determination of a stopping criterion for sifting process is needed.

In Huang et.al. (1998), the size of the standard deviation, *SD*, computed from the two consecutive sifting results was limited:

$$SD = \sum_{t=0}^{T} \left[\frac{\left| (h_{1(k-1)}(t) - h_{1k}(t) \right|^2}{h_{1(k-1)}^2(t)} \right], \tag{4.4}$$

where *h* stands for corresponding sifting result. It was offered to set this *SD* value between 0.2 and 0.3. However, how small the *SD* value should be is arguable. In addition, it is not related to the definition of the IMF. Same numbers of zero-crossings and extrema are not guaranteed.

According to another criterion proposed by Huang et.al. (1999), the sifting is stopped when the numbers of zero crossings and extrema are equal or differ at most by one for *S* consecutive times. But how to select the *S* number is ambiguous. In Huang et.al. (2003), the effects of selecting different *S* numbers were investigated statistically. Typically a value of $3 \leq S \leq 5$ has proved successful as the default stopping criterion. It is not recommended to choose too high a *S* number due to the adverse effects of over-sifting.

Fixing the sifting number to a pre-determined value was proposed as another stoppage criterion in Wu and Huang (2004, 2005, 2009). In Wu and Huang (2010) the filtering properties of EMD were studied as its sifting number changes. Based on

the numerical results of the decompositions using EMD of a delta function and white noise, as the (pre-assigned and fixed) sifting number is changed from a small number to infinity, the EMD corresponds to filter banks with a filtering ratio that changes accordingly from 2 (dyadic) to 1; the filter window does not narrow accordingly, as the sifting number increases (Wu and Huang, 2010). It was proposed to select 10 as a fixed sifting number if it is desired to have EMD similar to an adaptive dyatic filter bank.

Recently, another stopping criterion was proposed by Rato et.al. (2008). A resolution factor is defined by the ratio between the energy of the signal at the beginning of the sifting, and the energy of the mean of the envelopes:

$$R = \frac{E_x}{E_m}, \tag{4.5}$$

where E_x is the energy of the signal at the beginning of the sifting and E_m is the energy of the mean of the corresponding envelopes. If this ratio grows above the allowed resolution, sifting stops.

Contrary to the stopping criterion based on iteration numbers, resolution factor allows to set the number of IMFs. Traditional EMD algorithm generally results in close to but no more than $log_2 N$ IMF components for a time series with length *N*. But a predetermined sifting number can result different number of IMFs for different signals. Choosing a different sifting number for stopping criterion may not change the number of the computed IMFs. However, reducing the resolution factor reduces the number of the computed IMFs. This way, different signals can be decomposed into the same number of IMFs by designating a particular resolution factor for each signal for comparison purposes. It should be noted that with this approach, it is not guaranteed that the number of the zero crossings and the number of the extrema are equal or differ at most by one.

Finally, before proceeding to the next part, a major drawback of the EMD should be mentioned: mode mixing. It is a consequence of signal intermittency and defined as

1- An IMF consisting of signals of widely disparate scales, or 2- A signal of a similar scale residing in different IMF components. It can cause an IMF to be physically meaningless. To solve this problem, a noise-assisted data analysis method, the Ensemble EMD (EEMD) was developed by Wu and Huang (2009). It defines the true IMF components as the mean of an ensemble of trials, each consisting of the signal plus a white noise of finite amplitude. The effects of the added white noise series cancel each other in the ensemle average, the only persistent part is the signal. EEMD method is not in the scope of this thesis, but it is reserved as a future work. The final time-frequency projection which will be obtained after Hilbert Transform has a capability to rectify the mixed modes to some degree. Thus, the physical meaning of the decomposition should be evaluated in the Hilbert spectrum.

4.6 Hilbert Transform (HT) and Hilbert Spectrum

4.6.1 Hilbert Transform

To be able to compute the instantaneous frequency according to Equation (4.1), the Hilbert Transform should be applied to each IMF. According to the Euler's formula, for any real number *z*,

$$e^{jz} = \cos(z) + j\sin(z) , \tag{4.6}$$

where e is the base of the natural logarithm, j is the imaginary unit, with the argument z given in radians. This formula is used to represent the harmonic waveforms in the complex notation:

$$e^{jwt} = \cos(wt) + j\sin(wt) . \tag{4.7}$$

David Hilbert (1862-1943) showed that the function *sin(wt)* is the Hilbert Transform of *cos(wt)*. Hilbert Transform *Y(t)* of a function *X(t)* is defined for all *t* by,

$$Y(t) = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{X(\tau)}{t-\tau} d\tau \, . \tag{4.8}$$

It is normally not possible to calculate the Hilbert Transform because of the situation $\tau = t$. However, the Cauchy principal value P expands the class of functions for which the integral in Equation (4.8) exists. HT can be computed either in time domain or frequency domain. In essence, Equation (4.8) defines the HT as the convolution of *X(t)* with the Hilbert transformer $1/\pi t$ in time domain. So, this operation is filtering *X(t)* by a filter $1/\pi t$.

The convolution operation in time domain is equivalent to a multiplication in the frequency domain. Thus, we need to transform *X(t)* and the operator $1/\pi t$ into the frequency domain by using Fourier Transform (FT). Therefore in the frequency domain we have,

$$y(t) = x(t) \cdot (-j \cdot sgn(t)) \, , \tag{4.9}$$

where $y(t)$ and $x(t)$ are the FT of $Y(t)$ and $X(t)$, respectively, and $sgn(t)$ is the sign function where $sgn(t) = 1$ if $t > 0$ and $sgn(t) = -1$ if $t < 0$.

Having now $y(t)$, we can compute $Y(t)$ just by applying the Inverse Fourier Transform (IFT) of $y(t)$. This way, we have finally computed the HT ($Y(t)$) by just using $X(t)$.

$X(t)$ and $Y(t)$ form the complex conjugate pair, so we can build an analytic signal, $Z(t)$, as

$$Z(t) = X(t) + jY(t) = a(t)e^{j\theta(t)} \, , \tag{4.10}$$

where instantaneous amplitude $a(t)$ and instantaneous phase $\theta(t)$ are computed as follows,

$$a(t) = \sqrt{X^2(t) + Y^2(t)}\,,\ \theta(t) = \arctan\left(Y(t)/X(t)\right). \tag{4.11}$$

Now, instantaneous frequencies can be computed according to Equation (4.1). After performing the HT on each IMF, the data can be expressed as follows:

$$X(t) = \sum_{i=1}^{n} a_i(t) e^{j \int w_i(t) dt}\,. \tag{4.12}$$

4.6.2 Hilbert Spectrum

The residue is not included in the above expression. In Equation (4.12), the amplitude a_i and the instantaneous frequency w_i are variables and they change as functions of time. They are not constants as in Fourier expansion. Hence, it is possible to represent the amplitude and the instantaneous frequency as functions of time in a three-dimensional plot, in which the amplitude can be contoured on the frequency-time plane. This representation (frequency-time distribution of the amplitude) is called Hilbert spectrum *H(w,t)*. Instead of amplitude, amplitude squared can be also used, which is a more common way to display energy density.

Before proceeding to the next part, an important issue should be clarified which cannot be swept under the rug. As explained above, HT of a signal is calculated with the help of Fourier Transform and Inverse Fourier Transform. Then, restrictions on the calculation of FT are also valid on the calculation of HT.

FT assumes that the data are equally sampled. Sampling-rate might be faster or slower, but it is not possible to apply FT on an unevenly sampled data. However, in many cases, GNSS time series contain gaps with varying extent due to the malfunctions of GNSS stations and detected and cleaned outliers. Therefore, it is a must to fill up the gaps before applying FT and HT. There are many interpolation

techniques which can serve for this purpose. Finding the most appropriate interpolation technique for GNSS time series is outside the scope of this thesis. Performance of any interpolation technique decreases as the data gap extends. Keeping in mind the necessity of a sophisticated interpolation technique, linear interpolation was used to fill up the gaps in TNPGN-Active time series in this thesis. Another approach could be to cut and paste all data side to side, as if there is no gap. Then it would be a cumbersome operation to find the correct time location of a physical event which would be probably detected in the Hilbert spectrum.

4.6.3 Marginal Hilbert Spectrum

The Marginal Hilbert Spectrum can also be defined by using Hilbert spectrum amplitudes as follows:

$$h(w) = \int_0^T H(w,t)dt \ . \tag{4.13}$$

The marginal spectrum offers a measure of total amplitude contribution from each frequency value.

CHAPTER 5

APPLICATIONS

5.1 GNSS Data Analysis

5.1.1 Introduction

In this part of the thesis, GNSS observations collected at TNPGN-Active stations have been processed and coordinate time series of all TNPGN-Active stations have been produced. TNPGN-Active is a Real Time Kinematic (RTK) GNSS network established under a Turkish National Scientific and Technological Council project. It consists of 146 permanent GNSS stations covering whole country with 80-100 km inter-station distances. Locations of TNPGN-Active sites can be seen in Figure 5.1.

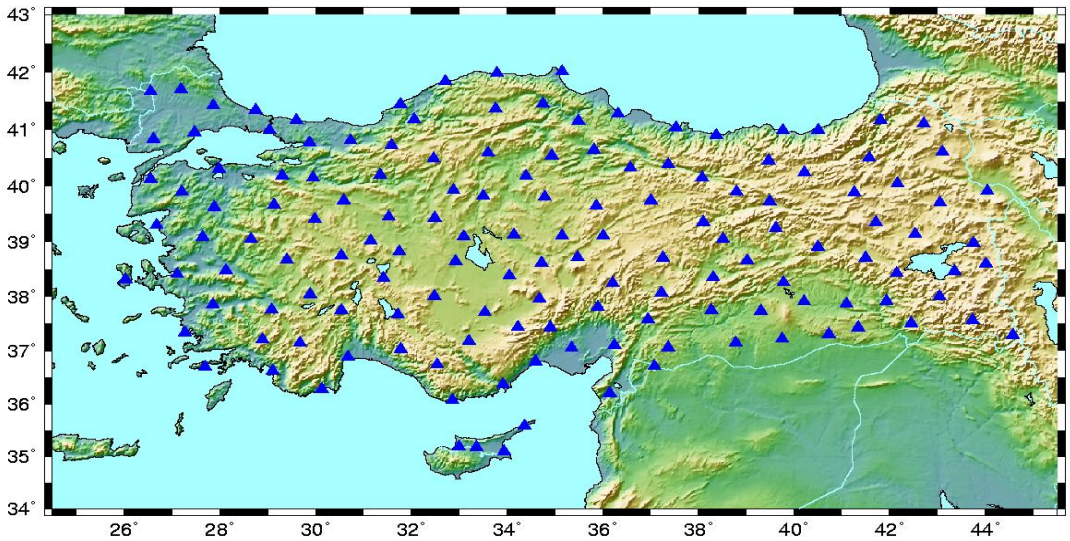

Figure 5.1 TNPGN-Active Stations (represented with blue triangles).

The main purpose of TNPGN-Active System is to provide RTK corrections and herewith to enable users on the field with cm-level positional accuracies. Correction parameters are calculated at the control centers in General Command of Mapping and in General Directorate of Land Registry and Cadastre. Observations collected at the stations are transferred to the control centers through ADSL (Asymmetric Digital Subscriber Line) and GPRS/EDGE (General Packet Radio Service/Enhanced Data for Global Evolution) while corrections are sent to the users via GPRS and NTRIP (Networked Transport of RTCM via Internet Protocol) in the recent RTCM (The Radio Technical Commission for Maritime Services) and CMR+ (Compact Measurement Record) formats.

Although all stations are equipped with a GNSS (GPS+GLONASS+GALILEO) receiver, only GPS observations are included in all calculation procedures.

While TNPGN-Active is serving mapping community by providing accurate coordinate information in couple minutes, daily GNSS data also provides an indispensable tool for earth science community. No matter in which study area TNPGN-Active based information will be used, detailed analysis of the coordinate time-series of TNPGN-Active stations and the removal of the unwanted signal which can affect the coordinate and velocity estimates adversely are of utmost importance.

TNPGN-Active system is operational since the second half of 2008. Though stations were established in piecemeal fashion, almost all stations reached 5-year temporal resolution which allows implementing an appropriate time series inspection. In Section 2.2, relationship between the temporal resolution of the GNSS time series and velocity bias is explained. In Table 5.1, NRMS (Normalized Root Mean Square) and WRMS (Weighted Root Mean Square) values of north, east and up coordinate components of some selected TNPGN-Active stations which were obtained after the analysis of data between 2009 and 2012 can be seen. NRMS and WRMS are two statistical measures of goodness of fit. NRMS reflects some combination of measurement errors and model errors and it is simply the square root of the χ^2

divided by degrees of freedom. How to interpret the NRMS values is given in Section 2.5.1. The explicit formula of NRMS is,

$$NRMS = \left[\frac{1}{N-1}\sum_{i=1}^{N}\frac{e_i^2}{\sigma_i^2}\right]^{1/2} \tag{5.1}$$

(Spinler et.al., 2010), where N is the number of the observations, e_i is the ith residual (observed minus computed observations), and σ_i is the uncertainty associated with the ith observation. WRMS is a measure of repeatibility, or precision, of the measurements. It is calculated as,

$$WRMS = \left[\frac{\frac{1}{N-1}\sum_{i=1}^{N}\frac{e_i^2}{\sigma_i^2}}{\sum_{i=1}^{N}\frac{1}{\sigma_i^2}}\right]^{1/2} . \tag{5.2}$$

In Table 5.1, RMS values of TNPGN stations of which data spread over a longer period can also be seen. TNPGN is a static continuos network established and is being operated by General Command of Mapping. TNPGN-Active stations are in red color while TNPGN stations in black.

Table 5.1 RMS values of selected TNPGN-Active (red) and TNPGN Sites (black).

Station	WRMS (mm)			NRMS (mm)			Duration (year)
	N	E	U	N	E	U	
SINP	1.7	2.2	5.8	1.0	1.4	1.0	3.42
TRBN	2.0	2.1	6.9	0.9	1.0	0.9	3.38
ZONG	2.0	1.8	5.8	1.1	1.1	1.0	3.36
CESM	2.1	1.8	5.0	1.1	0.9	0.7	3.17
ANTL	1.8	1.6	5.2	1.1	1.0	0.9	3.14
SILF	2.1	2.0	5.4	1.2	1.1	0.9	3.14
MIHA	1.9	2.2	6.3	0.9	1.0	0.9	8.64
ESME	1.9	2.0	5.9	0.8	0.9	0.7	8.51
AKHI	2.1	2.3	6.9	0.9	1.0	0.8	8.27
KNYA	2.0	1.8	6.2	1.2	1.1	1.1	8.28
ANKR	2.7	2.4	7.4	2.0	1.6	1.7	13.00

To make comparison between the permanent type and episodic (campaign) type GNSS analysis results, some tide gauge stations which are close to the selected TNPGN-Active stations (shown in Table 5.1) and have campaign type GNSS data were selected. These tide gauge stations are also operated by General Command of Mapping. For consistency, tide gauge stations data were processed in the same manner with TNPGN-Active stations. RMS values of tide gauge stations, duration and numbers of epochs can be seen in Table 5.2.

When RMS values in Table 5.2 are examined, it is observed that the precision obtained from campaign type analysis is lower than the precision of permanent type analysis no matter how big the number of epochs is. On the other hand, it is apparent in Table 5.1 that RMS values of TNPGN-Active stations with a duration 3-3.5 years are more or less in the same level of the RMS values of TNPGN stations which have data over 8 years.

Table 5.2 RMS values of selected tide gauge stations.

Station	WRMS (mm)			NRMS (mm)			Duration(yr) – #Epoch
	N	E	U	N	E	U	
SNOP	4.9	6.3	11.7	1.6	2.0	1.2	5.94 - 7
TRBZ	5.7	3.9	36.9	1.8	1.1	3.6	8.02 - 8
AMSR	3.8	4.5	13.7	1.5	1.7	1.7	7.06 - 7
MENT	9.4	7.4	12.5	2.8	2.0	1.7	18.83 - 14
ANTG	7.1	7.3	15.6	1.1	2.0	1.7	16.72 - 13
ERDM	3.8	3.4	6.2	1.9	1.5	1.0	8.12 - 8

Prominent scientific GNSS data processing softwares is listed in Section 2.1. GAMIT-GLOBK Version 10.5 software has been used in this study. GAMIT-GLOBK is a GPS analysis package developed for estimating station coordinates and velocities, representation of post-seismic deformation, atmospheric delays, satellite orbits and Earth orientation parameters.

GAMIT is used to process phase data to estimate the positions of stations and satellite orbits, atmospheric zenith delays and Earth orientation parameters while

GLOBK, which is a Kalman filter, is used to combine various geodetic solutions. Estimates and covariance matrices of parameters produced by GAMIT are the input data for GLOBK. The software uses C-shell scripts which invoke the Fortran or C programs, and it is designed to run under any UNIX operating system supporting X-Windows.

5.1.2 Daily GNSS Data Processing

In the analysis of TNPGN-Active data, singly and doubly differenced phases were used in a relative positioning sense and correlations stemmed from differencing process were considered (see Schaffrin and Bock (1988), for a detailed discussion of this algorithm). Triple differences were used only for editing purpose and not used in parameter estimation. Whenever a cycle slip that cannot be repaired detected, an extra bias parameter to be estimated added. However, relative positioning of networks by clock estimation algorithm mentioned in Section 2.3 was also used to compute one-way (undifferenced) carrier phase residuals for editing and display in the analysis. For this, phase offsets due to the station and satellite clocks at each epoch were estimated.

Various combinations of phase and pseudorange observations explained in Section 2.3 were used in the analysis for resolving the phase ambiguities as well as for fixing cycle slips. If ambiguity parameters can be fixed in the model, the ambiguities are not estimated, and theoretically a more precise solution can be generated by using the same data. This reduces the covariance matrix and lowers the expected errors in the station coordinates.

To determine L1 and L2 ambiguities, first L1-L2 cycles and then L1 cycles were resolved by using Melbourne-Wübenna widelanes with pseudoranges. Double-differenced phase ambiguities which could not be resolved were adjusted. Whenever the precise pseudoranges are available, it is possible to resolve widelane ambiguities for very long baselines (up to thousands of kilometers) no matter how the ionospheric conditions are. If precise pseudoranges are not available, the phase

observations must be used alone with a constraint on the ionosphere to resolve the widelane ambiguities (Feigl et.al., 1993). Since precise pseudoranges are available in TNPGN-Active system, no ionospheric constraints were given in the analysis.

Satellite-dependent differential code biases (DCBs) computed from tracking data by the Center of Orbit Determination in Europe (CODE), must be used in the analysis when different receiver types are included. Since all TNPGN-Active sites are equipped with same type of receivers, these DCB values were not included in the analysis.

In the analysis, the most accurate orbit products (SP3 final) produced by International GNSS Service were used. Latency of this product is 12-18 days and the accuracy of the orbits is ~2.5 cm. Offsets between the center of mass of a satellite and the phase-center of a transmitting antenna for each satellite were also taken into account. For yaw-axis attitude of satellites (must be considered due to maneuvers a satellite executes to keep its solar panel facing Sun), model of Bar-Sever (Bar-Sever, 1995) was applied.

Precession, nutation, Earth rotation and polar motion were included in the analysis to perform a reliable transformation between the Space-fixed and Earth-fixed systems. In addition, other factors affecting the position of the ground station such as luni-solar solid-body tides, loading of the crust by ocean tides and the atmosphere corotating with the Earth were incorporated into the model.

Variations in the phase centers of the satellite and ground station antennas may cause change in the phase and pseudorange observations. To avoid the adverse effects of this situation, nadir angle-dependent “absolute” phase center variations (PCVs) were applied for satellite antenna phase center corrections while “absolute” elevation and azimuth-dependent PCVs and L1/L2 offsets were applied for ground antenna phase center offsets and corrections.

Zenith delay parameters were estimated by using Saastamoinen (1972) model for dry and wet components. Niell Mapping Function (Niell, 1996) was used to extend delays to other elevation angles.

Doubly-differenced phase observations were used to estimate the relative positions of the stations, orbital and Earth rotation parameters, zenith delays, and phase ambiguities by incorporating a weighted least squares algorithm as explained in Section 2.5.1. It is always good to have good a priori values of the unknowns (station coordinates in our case) for processing. For this purpose, a preliminary solution was produced to generate and examine pre-fit residuals, and a robust editing of the data was performed. Errors up to 10 m were allowed and coordinates were estimated with accuracy better than one meter. Then, the entire solution was repeated with the updated coordinates. Certainly, another consideration was the linearity of the least squares adjustment since the mathematical model between the observations and unknowns was non-linear. Later on, a second solution was produced to obtain the final estimates in GAMIT. However, not to bias the combination performed by GLOBK, daily solutions generated by GAMIT were loosely-constrained. Naturally, several intermediate solutions were generated by GAMIT to be able to resolve phase ambiguities with constraints, but they were loosened before final solution.

For final solution of the station coordinates, GLOBK was used which takes the loosely-constrained daily GAMIT solution files as input. Data editing and linearity requirements must be met in the GAMIT part. ITRF-2008 (International Terrestrial Reference Frame-2008- Altamimi et.al (2011)) coordinates of the fiducial sites were supplied in GLOBK part. GLOBK uses a Kalman filter (see Section 2.5.2) which operates on error covariance matrices rather than normal equations and hence requires the specification of a non-infinite a priori constraint for each parameter estimated (Herring et.al., 1990).

In GAMIT, generated uncertainties (which passed to GLOBK) are determined by the a priori error assigned to the phase observations and by the sampling interval. In

the preliminary solution of TNPGN-Active data, an uncertainty of 10 mm were assigned to each one-way L1-phase. In a LC phase, the assigned uncertainty becomes 32 mm. The mean RMS of one-way LC residuals is typically ~6-9 mm. RMS values bigger than 15 mm give rise to thought of intensive multipath affect around the station. RMS values of one-way LC residuals were examined for all stations included in the analysis. RMS statistics for all stations in a chosen day (10th day of 2010) are shown in Figure 5.2.

At the end of this investigation, it was concluded that there is not any multipath problem for TNPGN-Active stations.

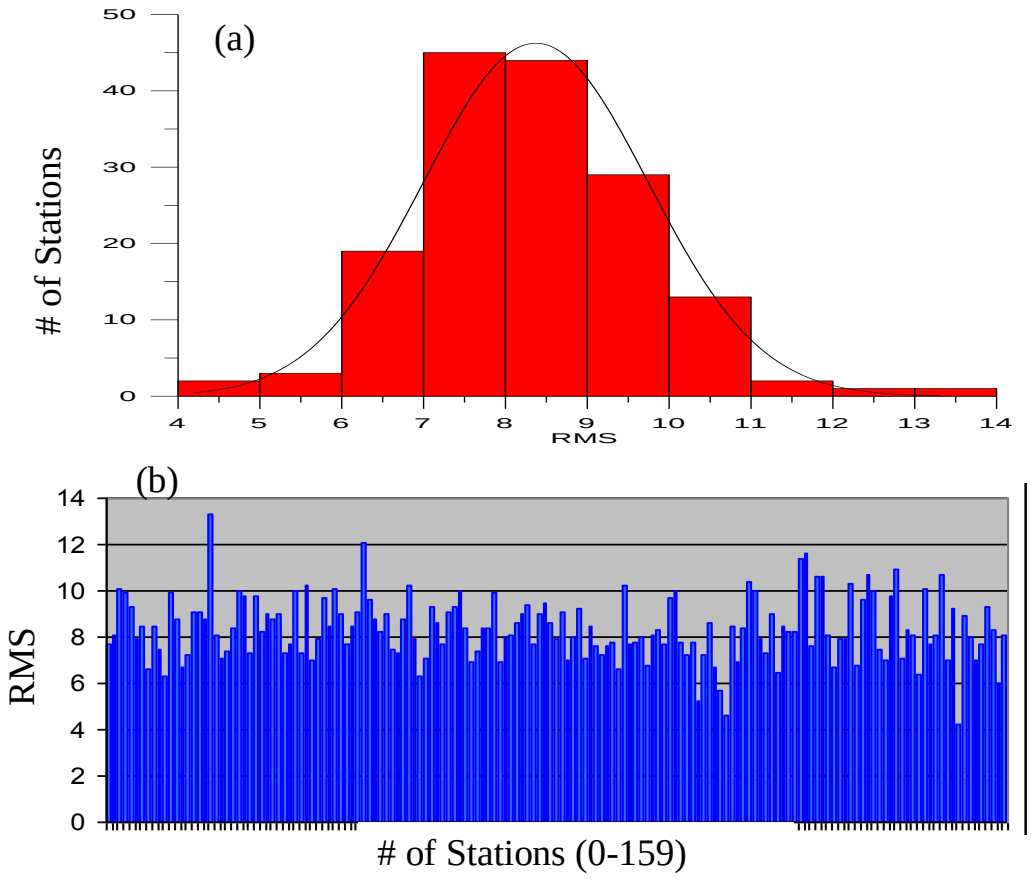

Figure 5.2 (a) Histogram of RMS values of one-way LC residuals in 10.01.2010. (b) RMS values of one-way LC residuals for all stations in 10.01.2010.

In the second solution of GAMIT, first, observations were rewighted using a constant and elevation-dependent term computed from the actual one-way LC phase residuals. In order to generate realistic coordinate uncertainties, inflated values of the a priori phase error were used and not rescaled by the NRMS. Actually, rescaling by the NRMS, i.e., making $\hat{\sigma}_0^2$ equal to 1.0, and compensating later for the unrealistically low coordinate uncertainties would be another valid approach. In GLOBK, it is always possible to rescale all covariances of GAMIT solutions and to add white or random-walk noise to the variances of individual stations.

For frame definition, *generalized* constraints was used in the analysis. In Figure 5.3, the numbers of the IGS stations used in reference frame definition for days of 2010 can be seen.

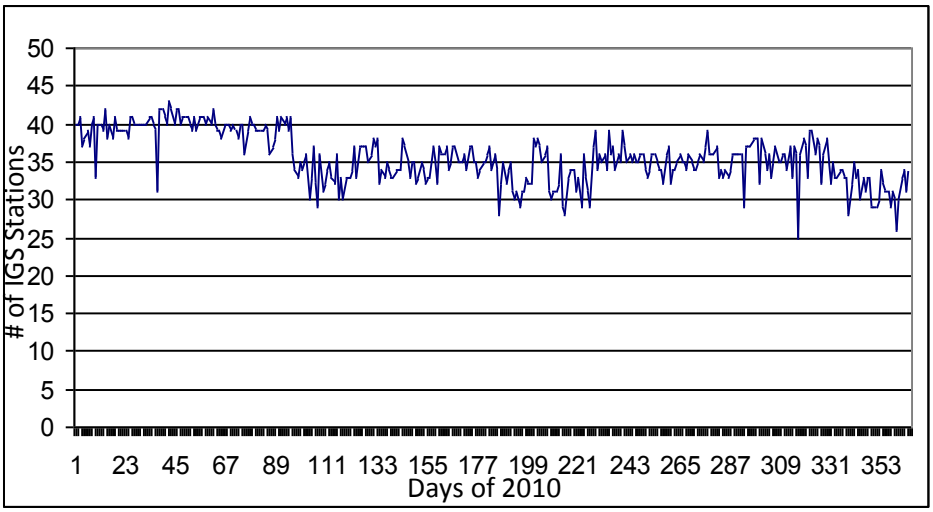

Figure 5.3 IGS stations numbers used in reference frame definition for days of 2010.

IGS sites used in reference frame definition were selected from IGS reference frame sites shown in Figure 5.4. On the other hand, the data of 16 IGS stations (as long as they are operational) shown in Figure 5.5 were also processed together with TNPGN-Active data on a daily basis.

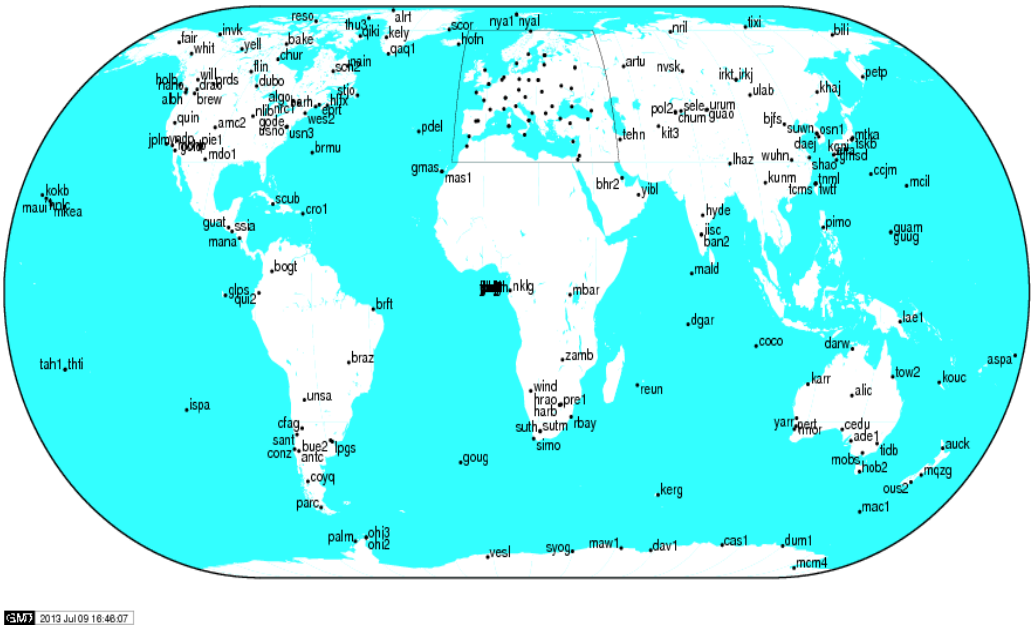

Figure 5.4 IGS (IGb08) Reference Frame Network

(http://igs.org/network/refframe.html)

These stations alone could have been used in reference frame definition by using 7-Parameter Helmert Transformation as long as their good a priori coordinates in ITRF2008 system were known. But daily GAMIT solutions were not used alone in the analysis, instead they were combined with the global solutions produced by SOPAC (Scripps Orbit and Permanent Array Center) analysis center. Thus, those high quality 16 IGS stations served as "tie" sites between local solutions and global solutions. Indeed, they do not need to be in the frame-realization list.

Combining TNPGN-Active local solutions with SOPAC global solutions gives access to over 100 sites which can be used for frame definition without having to include them in the local analysis. Since vertical coordinates have lower precision than horizontal coordinates in GPS analysis, their weights were decreased in the stabilization.

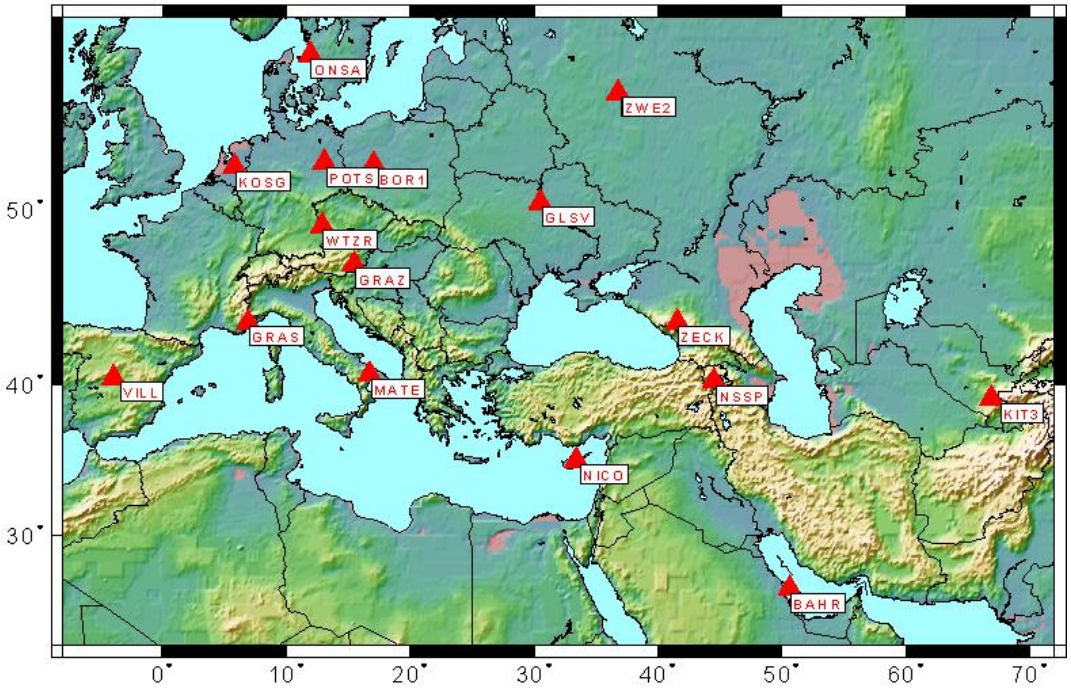

Figure 5.5 IGS stations processed together with TNPGN-Active data.

To be able to accept a stabilization reliable, obtained RMS values should be around 2-3 mm, especially no more than 5 mm. In Figure 5.6, RMS values obtained in reference frame definition for the same days in Figure 5.3 are given.

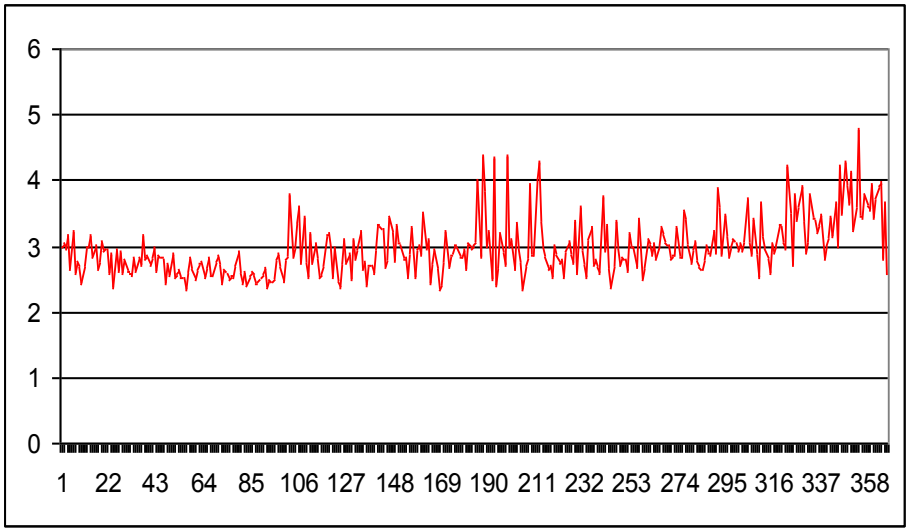

Figure 5.6 Obtained RMS values in reference frame definition for days of 2010.

5.1.3 Processing Strategy Summary

Mostly, MIT (Massachussetts Institute of Technology) analysis strategy was followed in this thesis (ftp://igscb.jpl.nasa.gov/pub/center/analysis/mit.acn). The processing strategy summary is given in Table 5.3.

Table 5.3 Processing Strategy Summary

PROCESSING STRATEGY SUMMARY	
Software used	GAMIT v. 10.5, GLOBK v. 5.12
GNSS system(s)	GPS
Effective date for data analysis	2008 and afterwards
Data processed	146 TNPGN-Active stations + 16 IGS stations (when available)
MEASUREMENT MODELS	
Basic observables	Doubly differenced, ionosphere-free combination of L1 and L2 carrier phases. Pseudoranges are used only to obtain receiver clock offsets and in ambiguity resolution Melbourne-Wuebbena widelane method. Non-redundant double differences are used (Schaffrin and Bock, 1988).
	Elevation angle cutoff: 10 degrees Sampling rate:30 seconds for cleaning; 2 minutes in least-squares analysis.
Satellite antenna -center of mass offsets	SV-specific z-offsets & block-specific x- & y-offsets (from manufacturers) from file igs08_wwww.atx.
Satellite antenna phase center corrections	Block-specific nadir angle-dependent "absolute" PCVs applied from file igs08_wwww.atx; no azimuth-dependent corrections applied.
GPS attitude model	GPS satellite yaw attitude model: applied (Bar-Sever, 1996) based on nominal yaw rates
Ground antenna phase center offsets & corrections	"Absolute" elevation- & azimuth-dependent (when available) PCVs & L1/L2 offsets from ARP applied from file igs08_wwww.atx.
Antenna radome calibrations	Calibration applied if given in file igs08_wwww.atx; otherwise radome effect neglected.
Marker -> antenna ARP eccentricity	dN,dE,dU eccentricities from site logs applied to compute station marker coordinates
Troposphere a priori model	Met data input: latitude, longitude, height, DOY climate model from (Boehm et al., 2007); rel. Humidity set to 50% for all sites
	Zenith delay: (Saastamoinen, 1972) "dry" + "wet" using synthesized input met data
	Mapping function: GMF (Boehm et al., 2006) for dry & wet zenith delays individually
	Horiz. Grad. Model: no a priori gradient model is used
Ionosphere	1st order effect: accounted for by dual-frequency observations in linear combination
	2nd order effect: no corrections applied
	Other effects: no corrections applied

Table 5.3 (cont'd) Processing Strategy Summary

Tidal Displacements	Solid Earth tide: IERS 2003 (IERS, 2004).
	Permanent tide: zero-frequency contribution left in tide model, NOT in site coordinates
	Solid Earth pole tide: IERS 2003; mean pole removed by linear trend.
	Oceanic pole tide: no model is applied
	Ocean tide loading: IERS Conventions 2003 using site-dependent amps & phase for 11 main tides from Bos & Scherneck website for FES2004 model.
	Ocean tide geocenter: site-dependent coeffs corrected for center of mass motion of whole Earth.
Non-tidal loadings	Atmospheric pressure: not applied
	Ocean bottom pressure: not applied
	Surface hydrology: not applied
	Other effects: none applied
Earth orientation Variations	Ocean tidal: diurnal/semidiurnal variations in x,y, & UT1 applied according to IERS 2003.
	High-frequency nutation: prograde diurnal polar motion corrections (IERS 2003) applied using IERS routine.
REFERENCE FRAMES	
Time argument	GPS time as given by observation epochs, which is offset by only a fixed constant (approx.) from TT/TDT
Inertial frame	Geocentric; mean equator and equinox of 2000 Jan 1.5 (J2000.0)
Terrestrial frame	ITRF2008 reference frame realized through the set of up to 50 station coordinates and velocities given in the IGS internal realization IGS08.snx (aligned to ITRF2008). Reference sites may be removed from the realization if the standard deviation of their position estimates deviates too much from the median sigma of the remaining reference sites or if their position estimate deviates by more than 4-sigma from the apriori value. Conditions are applied iteratively.
Interconnection	Precession: IAU 1976 Precession Theory
	Nutation: IAU 2000A Nutation Theory
	A priori EOPs: polar motion & UT1 interpolated from IERS Bulletin A, updated weekly, with the restoration of subdaily EOP variations using IERS models.
ORBIT MODELS	
Geopotential (static)	EGM96 degree and order 9.
	GM=398600.4415 km**3/sec**2 (for TT/TDT time argument)
	AE = 6378136.3 m
Tidal variations ingeopotential	Solid Earth tides: C20,C21,S21,C22, and S22 as in IERS (1992).
	Ocean tides: none
	Solid Earth pole tide: None applied in orbit models
	Oceanic pole tide: no model applied
Third-body forces	Sun & Moon as point masses
	GM_Sun 132712440000.0000 km**3/sec**2 GM_Moon 4902.7989 km**3/sec**2
Solar radiation pressure model	A priori: nominal block-dependent constant direct acceleration;
	Earth shadow model: umbra & penumbra included
	Earth albedo: not applied
	Moon shadow: umbra & penumbra included
	Satellite attitude: model of Bar-Sever (1995) applied

Table 5.3 (cont'd) Processing Strategy Summary

Relativistic effects	Dynamical correction: not applied.
	Gravitational time delay: IERS 2003.
Numerical integration	Adams-Moulton fixed-step, 11-pt predictor-corrector with Nordsieck variable-step starting procedure (Ash, 1972)
	Integration step-size: 75 s; tabular interval: 900 s
	Starter procedure: Runge-Kutta Formulation; initial conditions taken from prior orbit solution at 12:00
	Arc length: 24 hours (00:00:00 – 23:59:30 GPS time)
ESTIMATED PARAMETERS (& APRIORI VALUES & CONSTRAINTS)	
Adjustment method	Weighted least squares to generate loosely constrained covariance matrices and solutions that are passed to a Kalman filter for network combinations.
Data span	24 hours used for each daily analysis (00:00:00 – 23:59:30 GPS time)
Station coordinates	All station coordinates are adjusted, relative to the a priori values from IGS08.snx; a no-net-rotation condition is applied wrt the IGS08 frame using up to 50 reference frame stations; apriori sigmas for all stations are 10 m for each component.
Satellite clocks	Estimated using one-way phase data aligned with pseudorange.
	Sp3,clock files: Estimated values included 30-sec sampling for clock files.
Receiver clocks	Estimated during clock estimation. Stations clocks except the reference clock station are decimated to 300 seconds.
Orbits	sp3 files: orbits transformed to crust-fixed (rotating) frame accounting for geocenter motions due to ocean tides and for subdaily tidal EOP variations
Troposphere	Zenith delay: residual delays are adjusted for each station assuming mostly dominated by "wet" component; parameterized by piecewise linear, continuous model with 2-hr intervals
	Mapping function: GMF (Boehm et al., 2006) wet function used to estimate zenith delay residuals
	Zenith delay epochs: each even-integer hour
	Gradients: two N-S & two E-W gradient parameter per day for each station, with linear variation during the day; 30mm at 10-deg elevation 1-sigma constraint is applied at all stations.
Ionospheric correction	Not estimated.
Ambiguity	Real-valued double-differenced phase cycle ambiguities adjusted except when they can be resolved confidently in which case they are fixed using the Melbourne-Webana widelane to resolve L1-L2 cycles and then estimation to resolve L1 and L2 cycles.

5.1.4 Evaluation of Daily Processing Results

Three main criteria were considered for assessing the quality of the daily solutions. These are whether:

- All data were included in the processing.
- The data fit the model appropriately.
- The uncertainties are satisfactory.

These three requirements were evaluated for all TNPGN-Active daily solutions day by day. In Figure 5.7, the summary file from 19th day of 2011 of the 2nd group of TNPGN-Active stations can be seen. To speed up the process, TNPGN-Active sites were grouped into 4 groups with selected common points included in all groups.

```
Processing directory: /Users/process/analiz/2011_002
Disk Usage:  168379  Free  441523 Mbyte. Used 28%

Number of stations used 52 Total xfiles 52
Sites excluded by xsite command

Postfit RMS rms, total and by satellite
RMS IT Site All 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
RMS 22 ALL   7.1 77 79 82 78 64 64 60 58 78 65 59 61 85 80 79 80 69 81 63 65 66 65 65 61 74 63 71 66 80 79 64
Best and Worst two sites:
RMS 22 HINI 4.3 8 0 8 4 0 3 5 0 5 3 0 3 0 0 6 5 0 5 4 0 0 3 2 0 3 0 3 0 0 3 2
RMS 22 TOKA 5.4 7 6 5 6 4 5 3 4 6 4 3 4 7 7 6 6 4 6 4 5 5 5 4 4 7 4 6 5 6 6 4
RMS 22 RDIY 8.8 12 13 9 8 9 6 6 6 9 8 7 9 10 8 10 9 7 11 8 7 8 9 7 7 11 8 9 8 8 8 7
RMS 22 ARTV 9.8 10 12 12 10 8 10 8 9 11 10 8 10 10 11 9 11 9 11 8 9 10 9 9 9 10 8 9 8 11 10 8

Double difference statistics
Prefit nrms:  0.63779E+00   Postfit nrms: 0.18620E+00
Prefit nrms:  0.63594E+00   Postfit nrms: 0.18875E+00
Prefit nrms:  0.63779E+00   Postfit nrms: 0.18450E+00
Prefit nrms:  0.63594E+00   Postfit nrms: 0.18656E+00
Number of double differences: 167063

Phase ambiguities (Total  WL-fixed  NL-fixed): 1496 1337 954  AUTCLN (Total Expected)  1496 1496
Phase ambiguities WL fixed  89.4% NL fixed  63.8%

Processing 2011 002 GPS week 1617 0 Using node: tusagas-Mac-Pro.local Finished at: 11_03_27_00:58:35
```

Figure 5.7 The summary file of a daily solution (19th day of 2011) from TNPGN-Active analysis.

In Figure 5.7 (and in all solution summaries) it was checked whether the "Total xfiles" is equal to the number of RINEX (Receiver Independent Exchange Format) files available. In addition, "the number of stations used" should be equal to the total x-files (GAMIT ASCII observation files). If not, it means that some x-files were created from RINEX files but not included since the data duration is below some specified threshold. In the analysis of TNPGN-Active data, this threshold was specified as 300 which equals to about 3 hours of tracking.

The lines beginning with "RMS" show one-way RMS residuals by satellite and station. In the second RMS line, the first value is an overall RMS in mm. The remaining values on that line shows the scatter by satellite in tenths of mm, and can be used to detect a bad orbital model for one satellite. The last four RMS lines report the stations which have the lowest (HINI and TOKA in Figure 5.7) and the highest (RDIY and ARTV in Figure 5.7) scatter respectively. Typically, values of 3-5 mm for best sites and values of 7-9 mm for worst sites are expected while values between 10 and 15 mm show high but acceptable level of noise (Herring et.al., 2008).

In TNPGN-Active analysis, observations were reweighted according to the one-way phase residuals computed in the "preliminary solution", before using them in the "final solution". Thus, stations with 10-15 mm RMS values in the one-way phase residuals do generally show 2-3 times higher position uncertainties than those sites with better one-way phase residuals RMS values. RMS values higher than 15 mm may indicate a tracking problem in the receiver, high multipath, bad weather conditions, poor a priori coordinates, or short data span. Whenever encountered such high RMS values, phase versus elevation plots and sky plots of the related stations were examined in order to understand the source of the noise. In Figure 5.8, LC phase residual versus satellite elevation angles graphic for Edirne station is shown. In this figure, phase residuals show a distribution with mean values around 0 mm for all elevation angles.

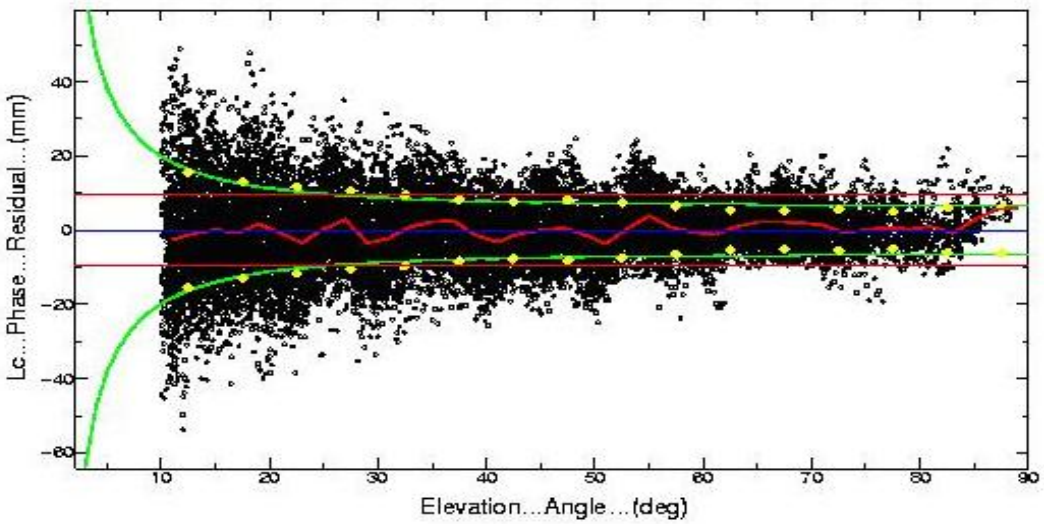

Figure 5.8 LC phase residuals versus satellite elevation angles for Edirne station.

However, in Figure 5.9, a problematic phase residuals versus elevation angles graphic is displayed. In this Figure, the presence of a bias in the phase residuals is obvious.

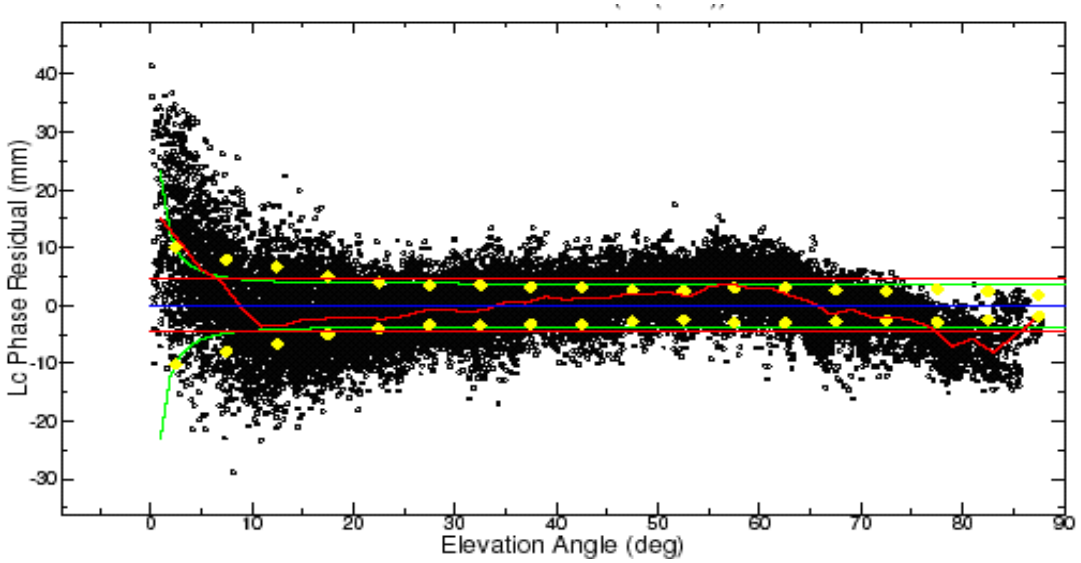

Figure 5.9 A problematic LC phase residuals versus satellite elevation angles graphic (Herring, 2012).

The reason of this problem is possibly the employment of a wrong receiver antenna information during the process. It causes the wrong antenna phase center variations definition.

Sky plots (satellite visibility graphics for 24 hours) can also give information about the source of the noise. Hence, sky plots for all TNPGN-Active stations were examined carefully. En example sky plot for Aydın station can be seen in Figure 5.10.

Generally, tropospheric fluctuations cause high residuals temporary. However, constant residuals give rise to thought of multipath affect. By examining the sky plots (which exhibit high residuals) for an appropriate time span, it was tried to detect the characteristics of the noise. For Aydın station, sky plots were inspected for 6 days as can be seen in Figure 5.11.

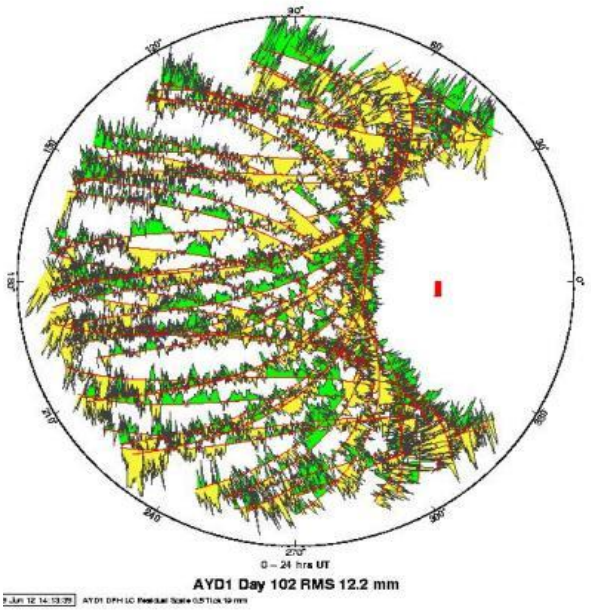

Figure 5.10. Sky plot of Aydın station for 102nd day of 2009. Positive residuals are shown in green color while negative ones in yellow.

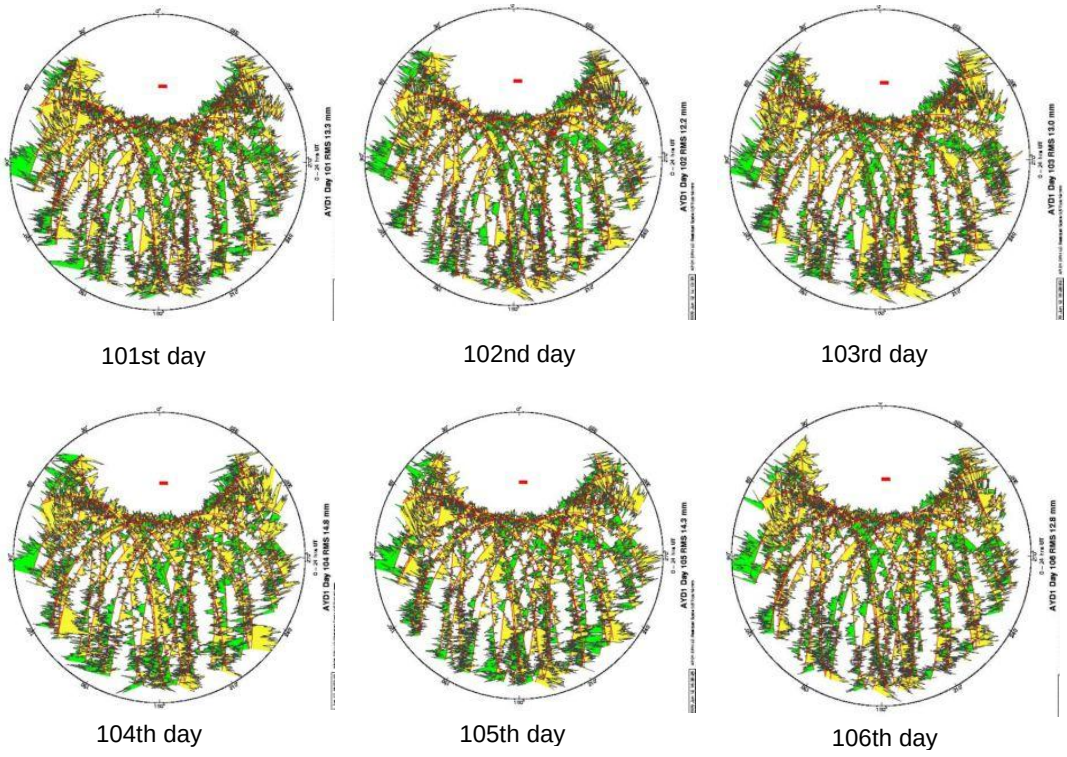

Figure 5.11. Sky plots of Aydın station for 101st-106th days of 2009.

It can be seen from Figure 5.11 that the noise has a constant behaviour in Aydın station. Yet, it was detected that the uncertainy rate in position estimates for Aydın station was still at an acceptable level.

It is also possible to find out the presence of geographic obstacles that interrupt the signals between the receiver and the satellites. Negligible obstacles were detected in the sky plots of couple TNPGN-Active stations. One of them is Çatak station and the sky plot is given in Figure 5.12. In this figure, the impeding affect of a small hill can be seen at the right hand side of the sky plot.

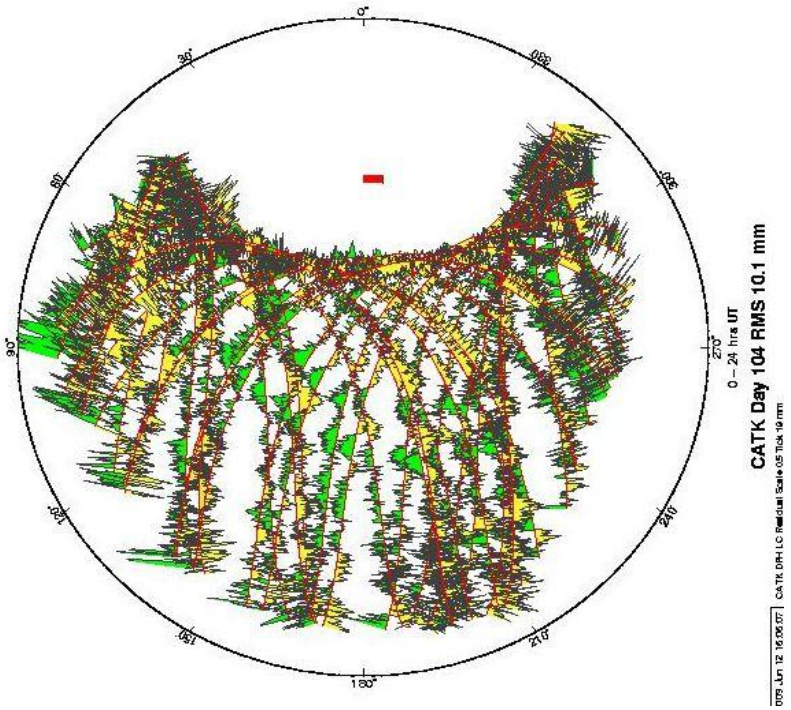

Figure 5.12 Sky plot of Çatak station.

In Figure 5.7, if both of the worst sites had high RMS, all other sites were also examined. Because they could have also high RMS values but did not make the top two, and this could give hint about an overall problem for that day.

Poor a priori coordinates or poor receiver performance were also controlled with the help of Figure 5.7. In such situations, one of the best two sites has 0 RMS, since all of the data from these sites are removed due to the detection of too many possible cycle slips. To determine the reason, clock and range noise statistics were investigated.

In Figure 5.13, a short part from clock and range noise statistics file for 10.01.2010 is displayed. Typically, maximum values of 1-2 m are expected for pseudorange RMS values. If the value is greater than 5 m, most likely it is due to a bad site coordinate.

```
AUTCLN  SUMMARY FILE: Version 3.33I

Clock and Range noise statistics at iteration  5
Site/PRN    Allan SD@100  #     Range rms    #
            sec  (ppb)            (mm)
ANKR        5.117287   2863        1002.1  24034 TPS
ANRK        0.035457   2879        1002.2  24073 TRM
AYVL        0.035513   2879         846.5  23712 TRM
BALK        0.033540   2879         809.0  23827 TRM
BAN1        0.035662   2879         876.5  23772 TRM
BILE        0.034256   2879         890.2  23839 TRM
BOLU        0.035060   2879         906.5  23808 TRM
BURS        0.033965   2879        1291.0  23818 TRM
CANA        0.035162   2663         827.2  22212 TRM
CANK        0.037754   2879        1030.2  23084 TRM
CMLD        0.035140   2879         949.2  23958 TRM
EDIR        0.034187   2866         836.4  23547 TRM
EMIR        0.036821   2872         942.9  23803 TRM
ESKS        0.035204   2879         914.5  23956 TRM
GLSV        0.039735   2879         940.9  24263 NOV
GRAS        0.031711   2879         646.5  23664 TRM
GRAZ        0.029131   2877         491.7  23313 LEI
HARC        0.032854   2879         791.8  23877 TRM
HEND        0.039291   2879        1035.7  22851 TRM
HYMN        0.033138   2844         884.9  23777 TRM
INE1        0.033841   2879        1095.9  23103 TRM
IPSA        0.032585   2879         735.9  23621 TRM
```

Figure 5.13 A part from the clock and range noise statistics file for 10.01.2010.

At the "double differences statistics part" of Figure 5.7, the first four lines starting with "Prefit NRMS" show the normalized RMS respectively for:

- The constrained solution with ambiguities free,
- The constrained solution with ambiguities resolved,
- The loose solution with ambiguities free,
- The loose solution with ambiguities resolved.

If the constraints imposed on the coordinates or orbits in the GAMIT solution would be too tight, the constrained NRMS values would be significantly larger than the loose NRMS. Around 0.2 mm postfit NRMS values were obtained for all daily solutions in TNPGN-Active analysis.

At the last part of Figure 5.7, the number of ambiguities estimated (1496) and the number of wide-lane (1337) and narrow-lane (954) ambiguities resolved are shown (with the ambiguity resolution ratio in the following line). If precise pseudoranges are available (as in TNPGN-Active), WL ambiguity resolution ratio should be normally over 80% for any size of network. Lower levels may indicate short sessions, noisy pseudoranges, or not using diferential code biases (DCB) values. NL ambiguity resolution ratio depends on the session length, size and configuration of the network, quality of the orbits and a priori coordinates, and atmospheric conditions (Herring et.al., 2008). Resolution of more than 50% NL ambiguities is expected. For all TNPGN-Active daily solutions, ambiguity resolution ratio for WL ambiguities was over 80% and for NL ambiguities over 50%.

5.2 The Time Series

Produced daily GAMIT solutions of TNPGN-Active data were combined with SOPAC global solution files by using GLOBK for the purpose of reference frame definition and time series generation. Strategies for the combination of quasi-observations (GAMIT solutions) to estimate station positions and velocities are described most completely in Dong et.al. (1998). The basic algorithms and a description of Kalman Filtering are given in Herring et.al. (1990). It is expected to get uncertainties and repeatibilities at the level of 1-2 mm and 3-5 mm for horizontal coordinates and heights respectively, with 24 hour sessions and a robust reference frame definition.

In Figure 5.14, the histograms of WRMS values for three components of the 282 GNSS stations (stations in SOPAC files included) are given. TNPGN-Active stations which have WRMS values over 5 mm for east and north components, and over 10 mm for up component are given in Table 5.4. In Figure 5.15, time series of the stations of which WRMS values are given in Table 5.4 are illustrated. Time series of other stations are not presented since they occupy very spacious space.

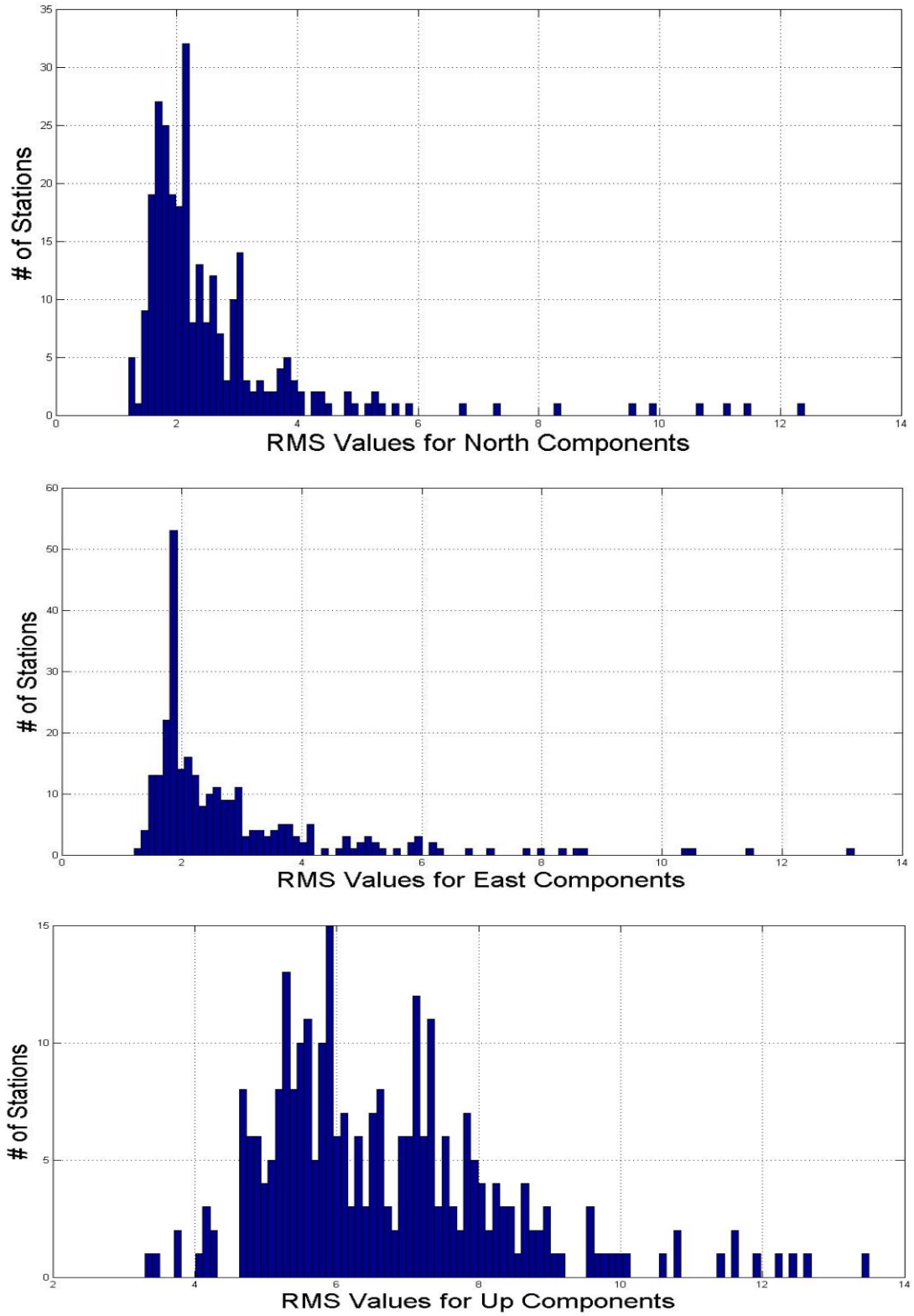

Figure 5.14 Histograms of WRMS values for north (top), east (middle), and up (bottom) components of all GNSS stations included in the analysis.

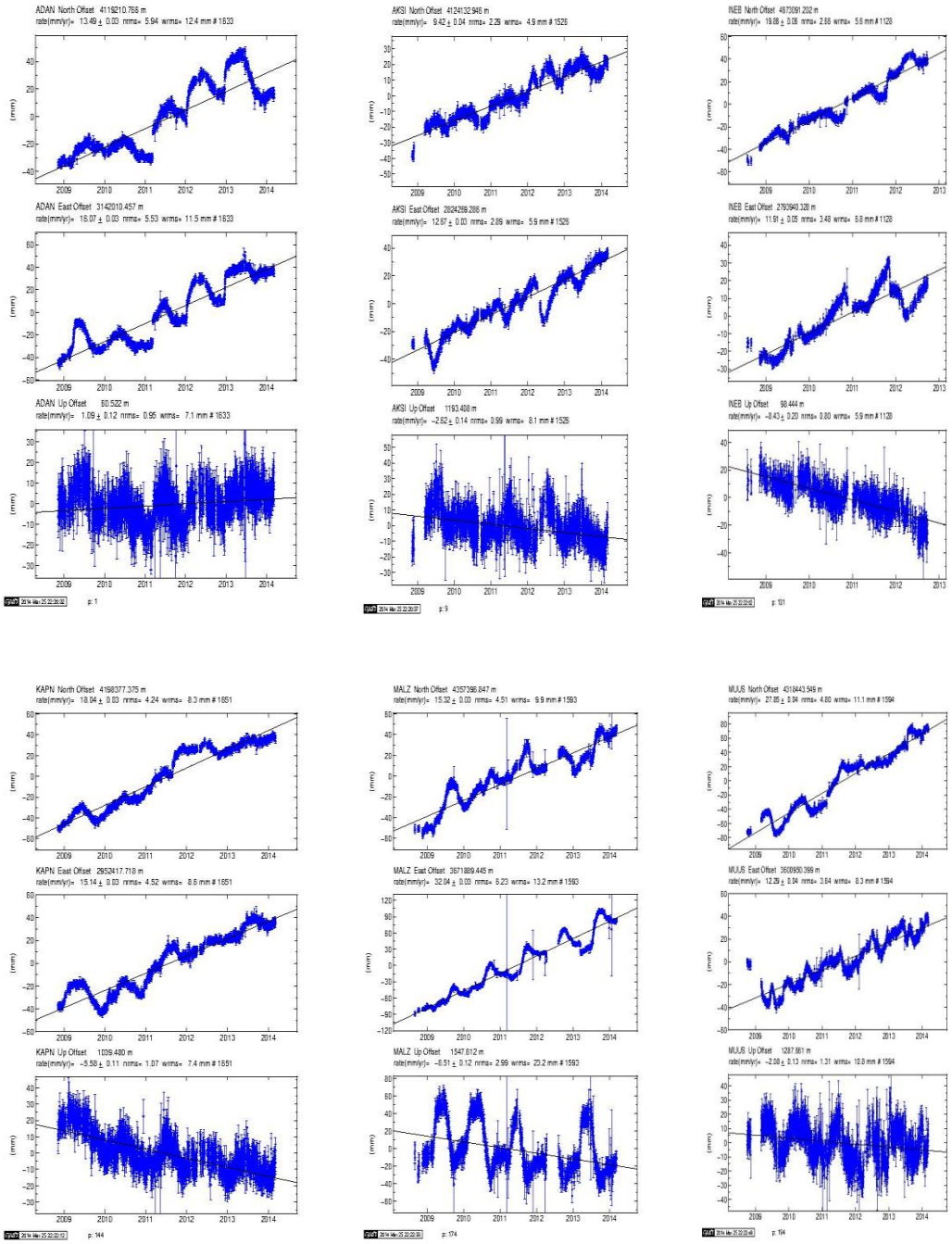

Figure 5.15 North-east-up (from top to bottom) time series of TNPGN-Active stations which exhibits high WRMS values: ADAN (top-left), AKSI (top-middle), INEB (top-right), KAPN (bottom-left), MALZ (bottom-middle), MUUS (bottom-right).

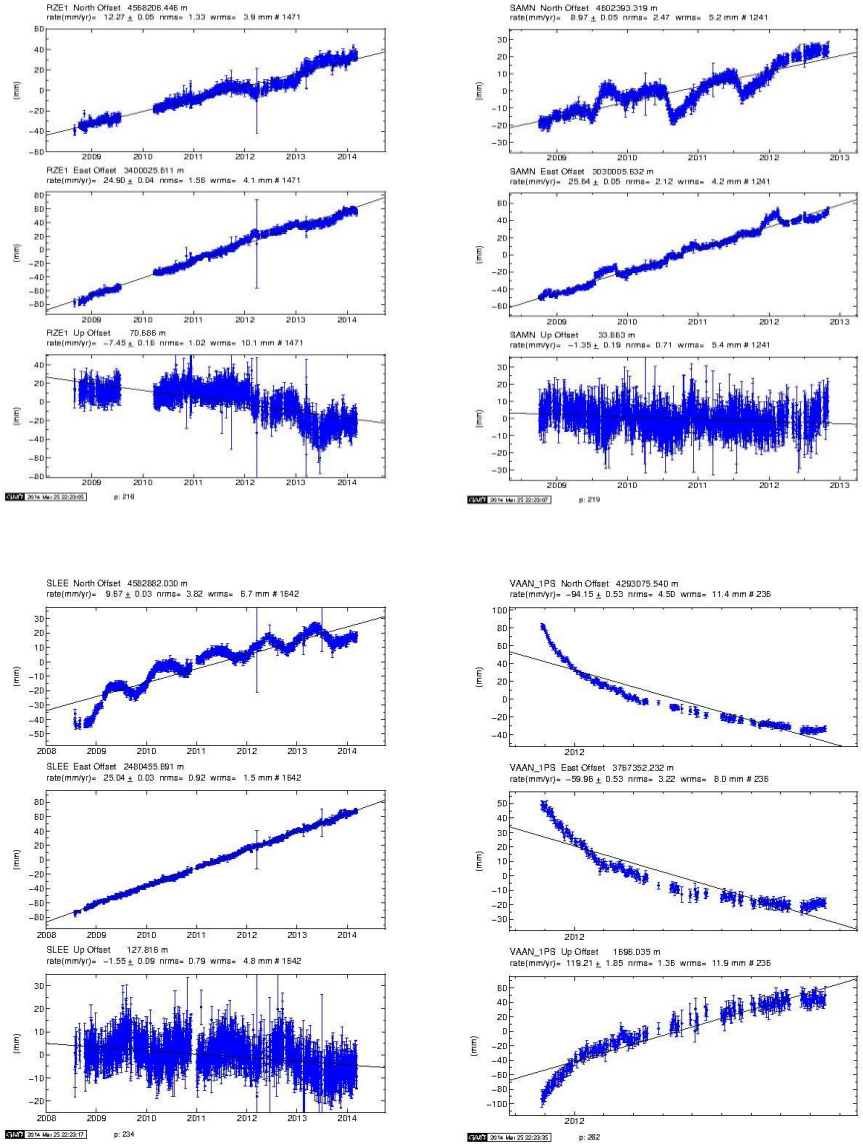

Figure 5.15 (cont'd) North-east-up (from top to bottom) time series of TNPGN-Active stations which exhibits high WRMS values: RZE1 (top-left), SAMN (top-right), SLEE (bottom-left), VAAN_1PS (bottom-right).

Table 5.4 TNPGN-Active Stations with Remarkable WRMS Values

Name	*WRMS Values (mm)*		
	North	*East*	*Up*
ADAN	12.4	11.5	7.1
AKSI	4.9	5.9	8.1
INEB	5.6	6.8	5.9
KAPN	8.3	8.6	7.4
MALZ	9.9	13.2	23.2
MUUS	11.1	8.3	10.8
RZE1	3.9	4.1	10.1
SAMN	5.2	4.2	5.4
SLEE	6.7	1.5	4.8
VAAN_1PS	11.4	8.	11.9

Even without knowing the WRMS values for these stations, visual inspection of the time series would tell us that something is wrong with these stations. VAAN_1PS is the new moniker of the original VAAN station, which had been named after 23rd October 2011 earthquake. The effect of the on-going post-seismic deformation is obvious. ADAN, AKSI, KAPN, RZE1, SLEE and VAAN_1PS stations are functioning on the roofs of buildings while INEB, MALZ, MUUS and SAMN stations are on the ground. Spurious motions of the monuments or buildings, and ground-water level changes come forward as possible reasons. In any case, these stations call for careful investigations.

As mentioned in Section 5.1.2, obtained post RMS values should be between 1 and 5 mm in reference frame definition. These values were lower than 5 mm for all TNPGN-Active combinations (in Figure 5.6, these values for year 2010 were given). In Figure 5.16, a part of a position summary for the last monthly combination of year 2012, which shows the adjustments and uncertainties of sites in north, east, and up is given. In this figure, starred sites are the ones that were used in the reference frame definition. The stabilization statistics at the end of the position summary are shown for each component (north east up) rather than a combination of the three as given in Figure 5.6.

SUMMARY POSITION ESTIMATES FROM GLOBK Ver 5.20I

Long. (deg)	Lat. (deg)	dE adj. (mm)	dN adj. (mm)	dE +- (mm)	dN +- (mm)	RHO	dH adj. (mm)	dH +- (mm)	SITE
27.08182	38.39481	22.74	11.79	0.39	0.41	-0.017	2.52	1.47	IZMI_GPS
26.71743	38.42659	29.97	9.33	1.34	1.36	-0.043	-21.50	7.56	MNTS_GSI
26.68618	39.31144	44.57	33.32	0.30	0.32	-0.020	-12.94	1.14	AYVL_GPS
26.55101	41.67671	93.13	78.05	0.31	0.34	-0.019	12.67	1.16	EDIR_GPS
26.41434	40.11121	24.41	39.18	0.31	0.33	-0.014	6.71	1.14	CANA_GPS
26.37978	40.91753	103.38	83.54	0.30	0.32	-0.017	13.81	1.12	IPSA_GPS
26.37257	38.30382	32.02	-0.18	0.39	0.41	-0.013	0.64	1.46	CESM_GPS
23.39473	42.55610	-3.41	0.23	0.39	0.42	-0.022	-8.88	1.25	SOFI_GPS
17.07346	52.27696	1.79	-2.24	0.16	0.19	-0.016	-6.94	0.58	BOR1_GPS*
16.70446	40.64913	1.60	2.27	0.15	0.14	-0.030	-3.80	0.50	MATE_GPS*
15.49348	47.06713	-1.15	1.78	0.16	0.17	-0.002	-13.70	0.55	GRAZ_GPS*
14.98979	36.87585	-4.81	-2.07	0.35	0.31	-0.019	-3.45	1.03	NOT1_GPS
13.06609	52.37930	-0.70	1.39	0.14	0.16	-0.038	-1.17	0.49	POTS_GPS*
12.87891	49.14420	2.08	0.16	0.17	0.19	-0.016	-10.86	0.62	WTZR_GPS*
11.92552	57.39530	0.27	-0.11	0.13	0.16	-0.107	-3.35	0.46	ONSA_GPS*
6.92058	43.75474	6.14	0.54	0.18	0.18	0.035	-5.74	0.61	GRAS_GPS*
5.80965	52.17843	11.57	6.14	0.29	0.34	-0.036	3.57	1.05	KOSG_GPS
0.33627	50.86732	-0.63	5.72	0.32	0.35	0.021	-20.25	1.00	HERS_GPS*

POS STATISTICS: For 39 RefSites WRMS ENU 3.20 2.76 7.75 mm NRMS ENU 11.83 9.77 8.80 L1

Figure 5.16 A part of a position summary for the last monthly combination of year 2012.

Later on, an outlier detection algorithm proposed in Section 5.3 was applied on each time series. Detected outliers were eliminated and daily solutions were combined loosely to generate monthly combinations which are used in the velocity estimation. As an example, the daily time series of GEME station and its correponding monthly combinations are displayed in Figure 5.17.

This approach (monthly combinations) shortens the processing time in velocity estimation and allows to generate statistics for long-term repeatibility without mixing in the short-term scatter. But it also serves for an implicit but significant purpose especially in this thesis. It excludes the possible problems due to the lack of days or weeks of data that can arise while forming the envelopes in Hilbert-Huang Transform. Lack of data leads to make interpolation to fill up the gaps in the signal which may betray the original behaviour of the signal.

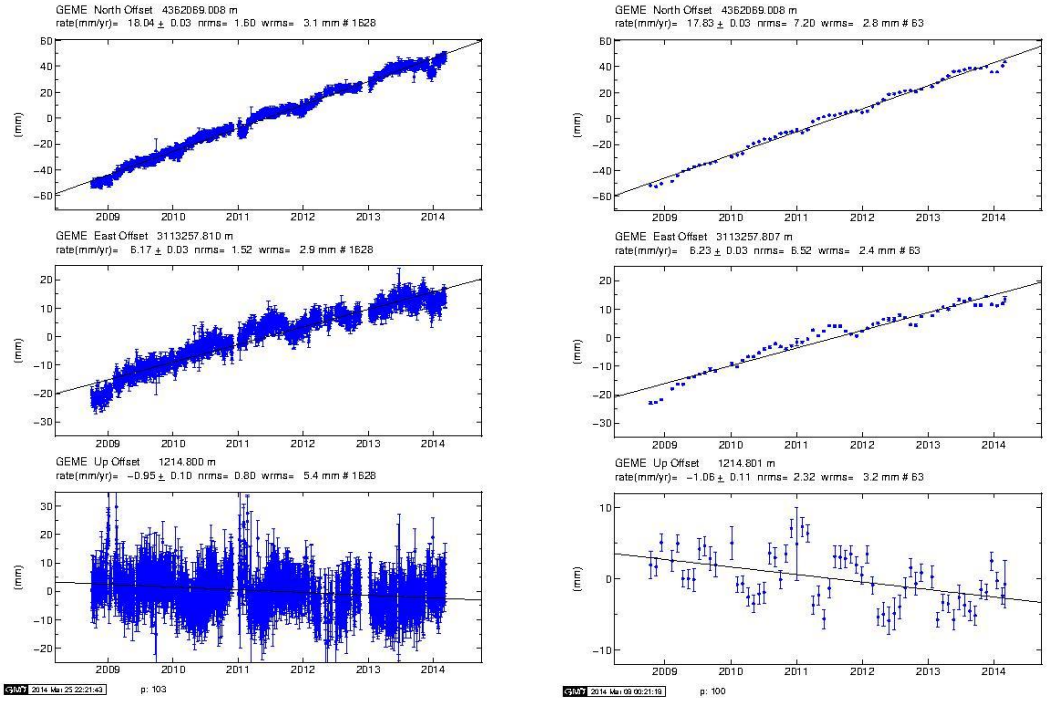

Figure 5.17 The daily time series of GEME station (left) and its correponding monthly combinations (right).

5.3 Outlier Detection

The *Z-value test* explained in Section 3 is a common and simple model used in outlier analysis. However, it cannot be assumed that the daily position estimates are normally distributed in a GNSS time-series. Thus, *Z-value* test cannot be used directly even as a heuristic in TNPGN-Active coordinate time-series.

In Figure 5.18, the east component of the time series of YUNK station until the end of 2011 is given. Normally, it is expected from an outlier detection test to catch the position estimate shown in red circle in Figure 5.18 as an outlier. *Z-value* test results are shown in Figure 5.19. Position estimates that are more than 3σ away from the mean value are shown in red circles.

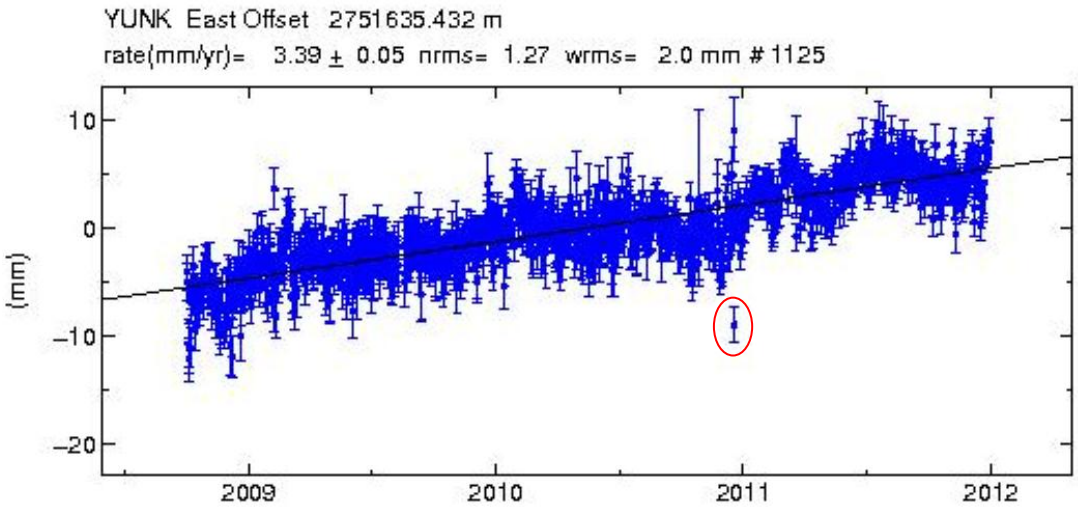

Figure 5.18 The East component of the time series of YUNK station.

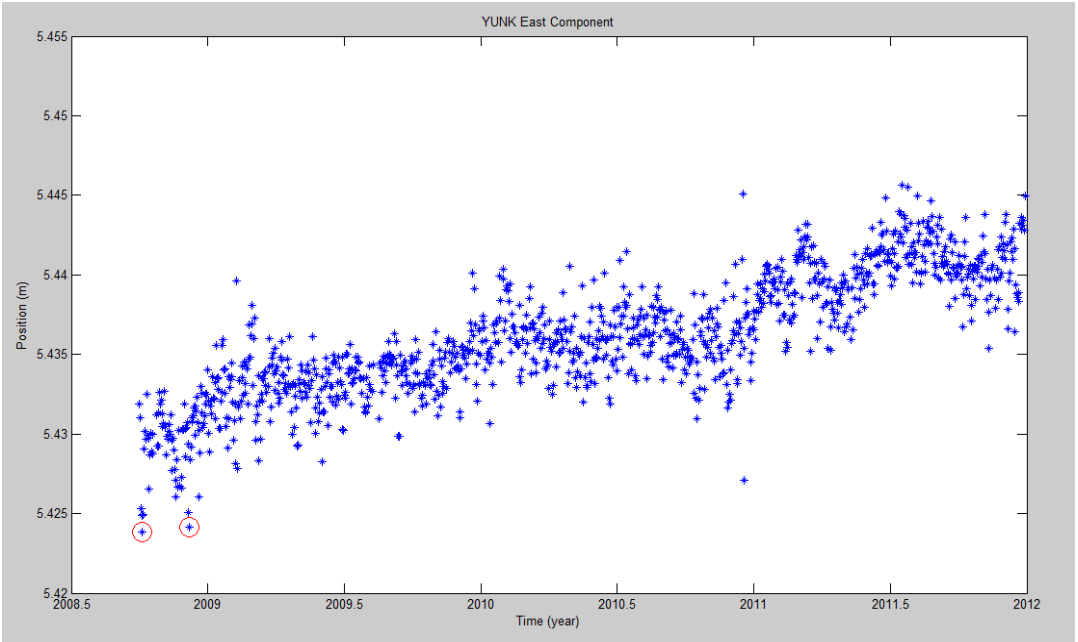

Figure 5.19 Z-value test results for YUNK East component. Outliers (> 3σ) are indicated with red circles.

As can be seen in Figure 5.19, *Z-value* test could not catch the expected estimate as an outlier. Instead, it detected two estimates close to the beginning of the time-series as outliers, although they are not. These two estimates are detected only because of

their relative positions, rather than their extreme positions. Thus, extreme value analysis cannot be applied to the position time series directly.

In Chapter 3, least squares fit and IRLS fit are explained as two methods to determine the optimal line which passes through the data points. Once the optimal line is obtained, extreme value analysis can be applied on the deviations from this line in order to find the outliers. In Figure 5.20, time-series of the north component of MURA station is displayed. The earthquake effect can be seen through the end of the year 2011. Although these estimates are not outliers (interseismic and postseismic velocities are estimated seperately), they can help to illustrate how the least squares fit swerves due to the extreme values. Blue line shows the least squares fit while green one IRLS fit. It is obvious that the IRLS robust fit is less affected from outliers than least squares fit. Thus, instead of deviations from least squares fit, deviations from IRLS fit were used to apply extreme value analysis in order to determine the outliers in the thesis. Indeed, it is not reasonable to use least squares fit to detect outliers while outliers have a gross effect on the calculation of the least squares fit.

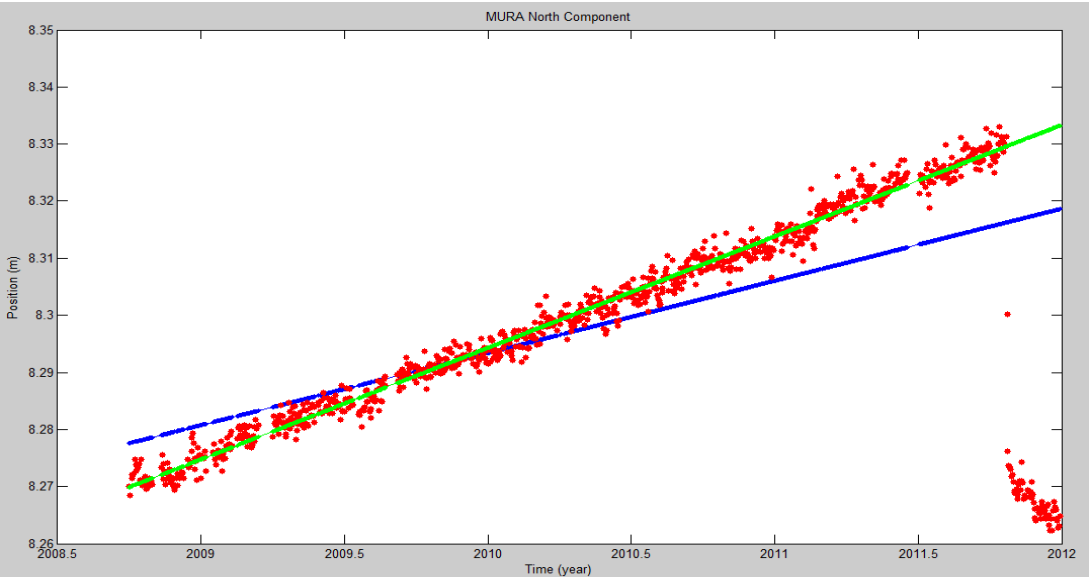

Figure 5.20 Time series of MURA North component. Position estimates are shown in red points, least squares fit in blue line, IRLS fit in green line.

After fitting the position estimates by IRLS, a *Z-value* test was applied to the residuals (deviations from the line) of each position estimate. This is called Z_1 value in the thesis.

Another important issue that should be considered in outlier detection of position estimates is the uncertainties of the estimates. Uncertainty rate shows the reliability level of the particular position estimate. It is not a good idea to eliminate a position estimate which deviates from the optimal line remarkably but possesses a very low uncertainty rate. Therefore, uncertainties were also somewhat included in the outlier analysis of TNPGN-Active time-series. For that purpose, a *Z-value* test was applied for uncertainties directly for there is not any apparent reason that they are not normally distributed. This is called Z_2 value in the thesis.

If a position estimate has a value of Z_1 lower than 2, it was accepted "normal" no matter how big the corresponding Z_2 value is. If a position estimate has a value of Z_1 between 2 and 3, corresponding Z_2 value was checked. If it is lower than 3, the position estimate again was accepted "normal".

A position estimate was accepted "outlier" if it has a value of Z_1 greater than 3, no matter how small the corresponding Z_2 value is. If a position estimate has a value of Z_1 between 2 and 3 but has a value of Z_2 greater than 3, then again it is accepted as an "outlier".

In Figure 5.21, the time series of YOZT east component is seen. After performing the outlier analysis algorithm described above, detected outliers can be seen as red stars in Figure 5.22. After cleaning the outliers, obtained time-series is displayed in Figure 5.23.

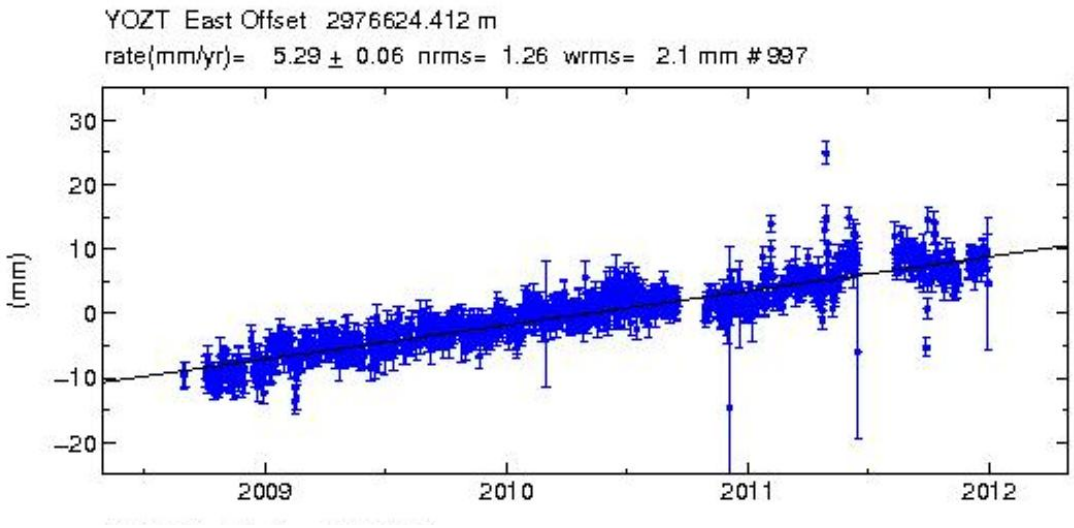

Figure 5.21 Time series of East component of YOZT station. Blue points are position estimates. Blue vertical lines intersect points are corresponding uncertainties.

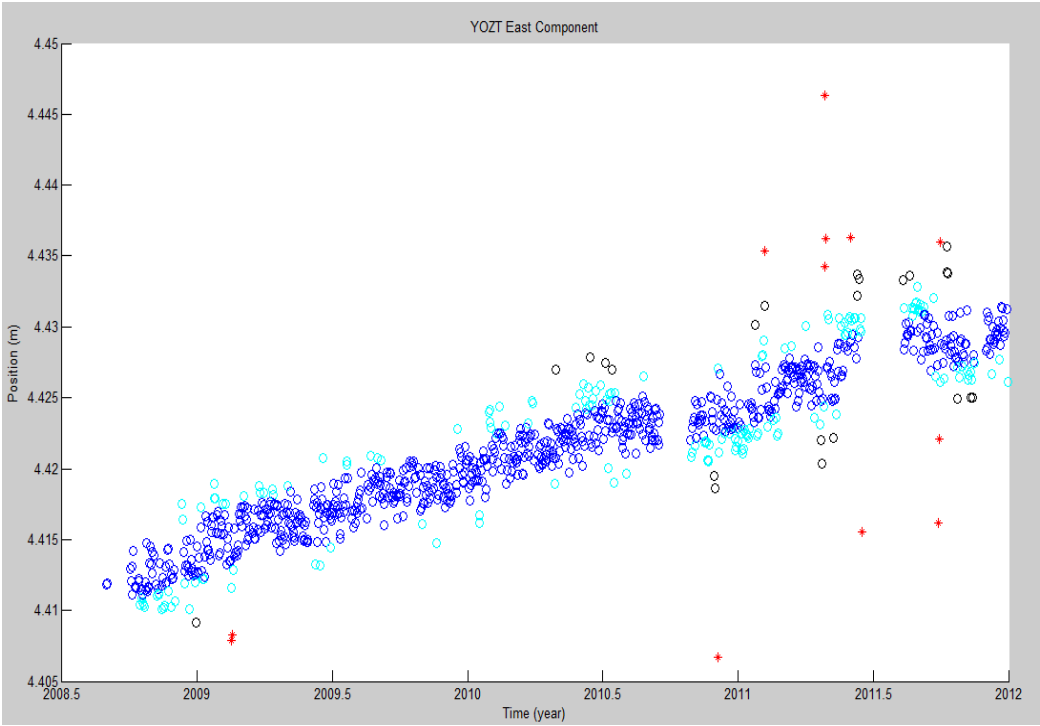

Figure 5.22 Outlier analysis result for YOZT East component. Blue circles: Z_1<=1; Cyan circles: 1< Z_1<= 2; Black circles: 2< Z_1<=3 and Z_2<=3; Red stars: outliers

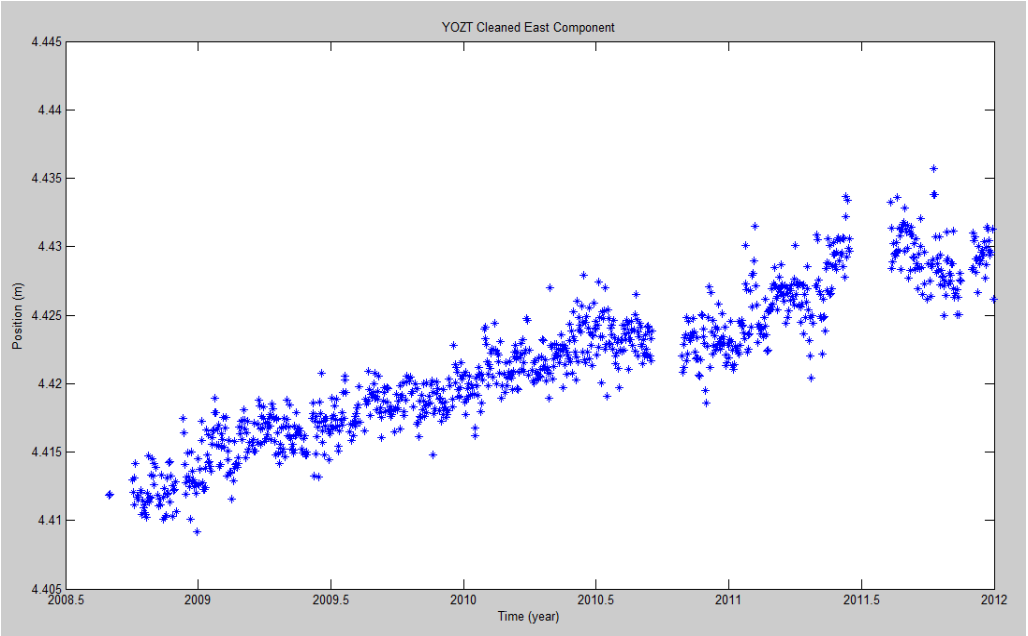

Figure 5.23 Cleaned YOZT East component after outlier detection.

The concept of line fitting which was implemented in this thesis and Principal Component Analysis (PCA) is similar (Jolliffe, 2002). But PCA uses the covariance matrix to describe how the points are spread around the mean. A new coordinate system is seeked in which the data can be represented without correlation, in other words, the covariance matrix in the new coordinate system is diagonal. On the other hand, some proximity-based methods such as k-nearest neighbor can be effectively used in outlier detection. Such methods model outliers as points which are isolated from the remaining data. Although it is worth to search PCA and k-nearest neighbor methods for outlier detection in GNSS time series, it is not expected to have significantly different results from these methods.

5.4 Hilbert-Huang Transform

In Chapter 4, theoretical bases of the Hilbert-Huang Transform and corresponding problematic areas are explained. In following sections, EMD and Hilbert Spectrum results are illustrated, and some solutions to the problematic areas of HHT are

proposed. An application of detrending by using EMD, and the effect of detrending in Fast Fourier Transform are also presented.

In this thesis, EMD method was applied to AFYN monthly combined east offsets to illustrate sifting process; to TNCE monthly combined up and east offsets to compare the natural cubic spline and constrained cubic spline interpolations; to BAYB monthly combined up offsets, KLIS daily north offsets and ISTN daily north offsets to show the obtained IMFs by using different stopping criteria; to daily north offsets of all TNGPN-Active stations to show the differences between the EMD results obtained by using *S* number and *SD* in a tabular form, to ISTN daily north offsets for completeness test and detrending; to KAPN daily north offsets to present the effect of detrending in Fast Fourier Transform; to CAVD daily east offsets to illustrate Hilbert Spectrum and Marginal Hilbert Spectrum and to MALZ daily up offsets to present the effect of linear interpolation in Hilbert Spectrum.

5.4.1 Sifting

The IMF's were found by an iterative procedure as explained in Section 4.3. In Figure 5.24, the sifting process is illustrated with the help of monthly combined east offset values in AFYN station.

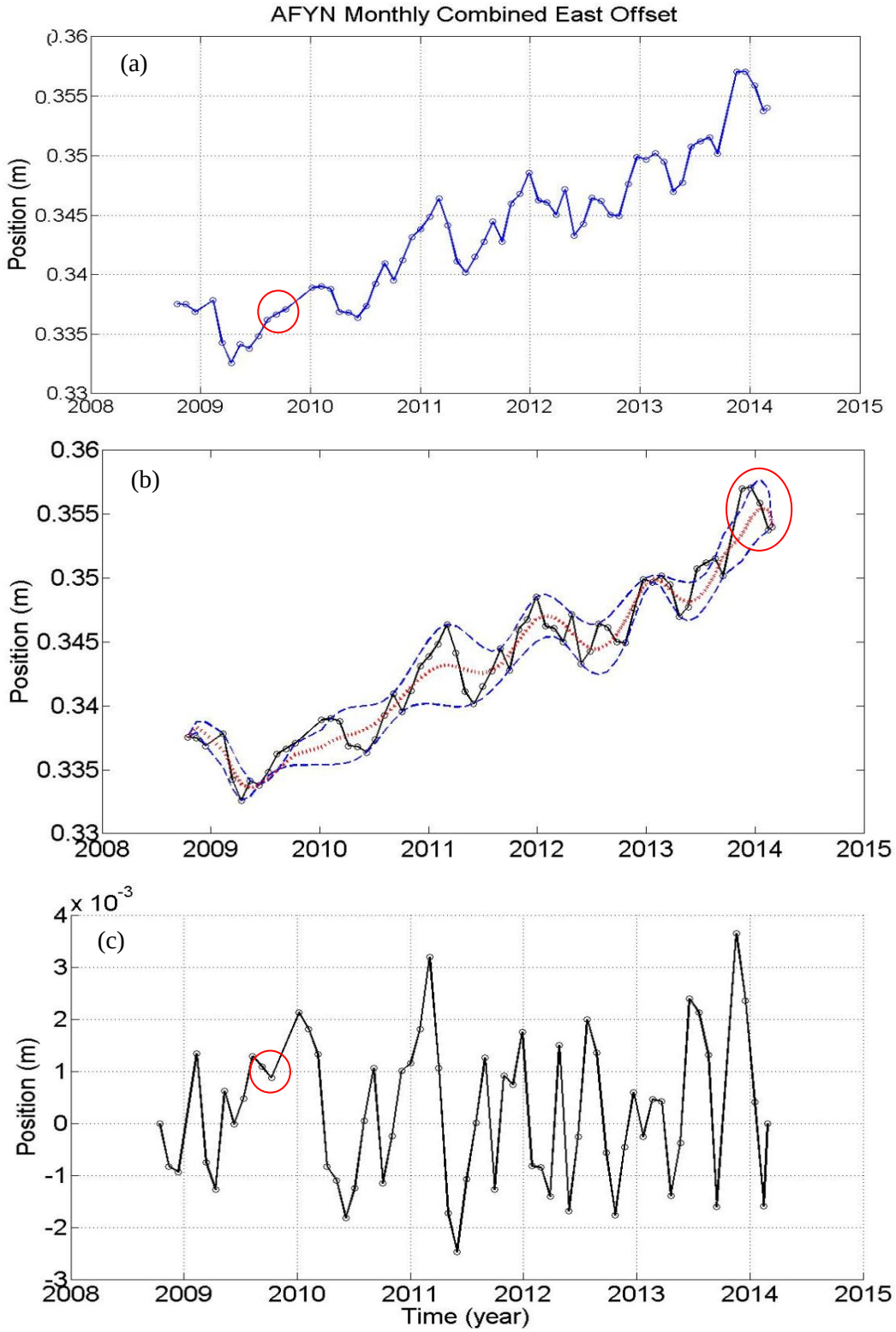

Figure 5.24 Sifting process: the original AFYN station monthly combined east offset data (a); the data in solid line, envelopes in dashed lines and the mean in dotted line (b); the difference between the data and m_1 (c).

5.4.2 Interpolation

In Figure 5.24 (c), there can be seen positive minima. In Figure 5.24 (a), a gentle hump on a slope through the end of 2009 can be seen which is highlighted by a red circle. After sifting, this hump became a local minimum at the same time location in Figure 5.24 (c). An overshoot example can be seen around 2014 in Figure 5.24 (b) which is also highlighted by a red circle. It is seen that the signal crosses the envelope. This affects the IMF estimation. These are the adverse effects of the cubic spline interpolation. To alleviate these effects, the constrained cubic spline interpolation (see Section 4.5) was tried to apply to TNPGN-Active time series in this thesis.

However, results were not as impressive as in Figure 4.1, after the application of constrained cubic spline interpolation to TNPGN-Active data. In Figure 5.25, interpolations between maxima of combined up offsets of TNCE station for a short duration by using constrained and natural cubic splines are given. It can be seen that it is the constrained cubic spline interpolation which causes overshoots rather than natural cubic spline. On the other hand, interpolations between maxima of combined east offsets of TNCE station for a short duration by using constrained and natural cubic splines are given in Figure 5.26. In this figure, clearly, constrained cubic spline fits the data better than natural cubic spline. The reason is obvious: to be able to obtain a better fit with constrained cubic spline than with natural cubic spline, the data should be monotonically increasing or decreasing. Otherwise, natural cubic spline is a better choice. Indeed, when the monotonically increasing data used in Figure 4.1 are changed slightly, interpolation results change substantially. This scenario is displayed in Figure 5.27.

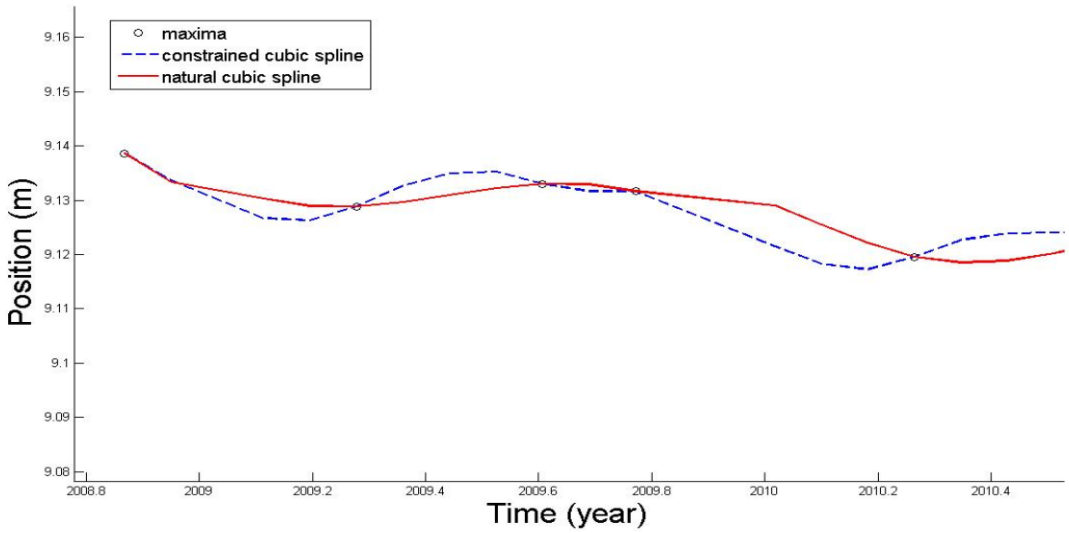

Figure 5.25 Interpolation between maxima of combined up offsets of TNCE station for a short duration by using constrained and natural cubic splines. Constrained cubic spline is in blue dashed line and natural cubic spline is in red solid line.

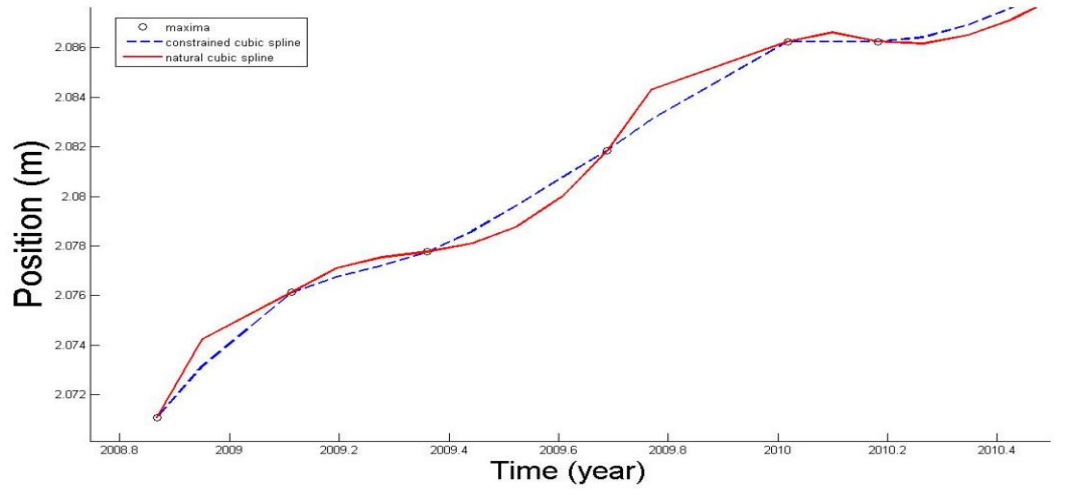

Figure 5.26 Interpolation between maxima of combined east offsets of TNCE station for a short duration by using constrained and natural cubic splines. Constrained cubic spline is in blue dashed line and natural cubic spline is in red solid line.

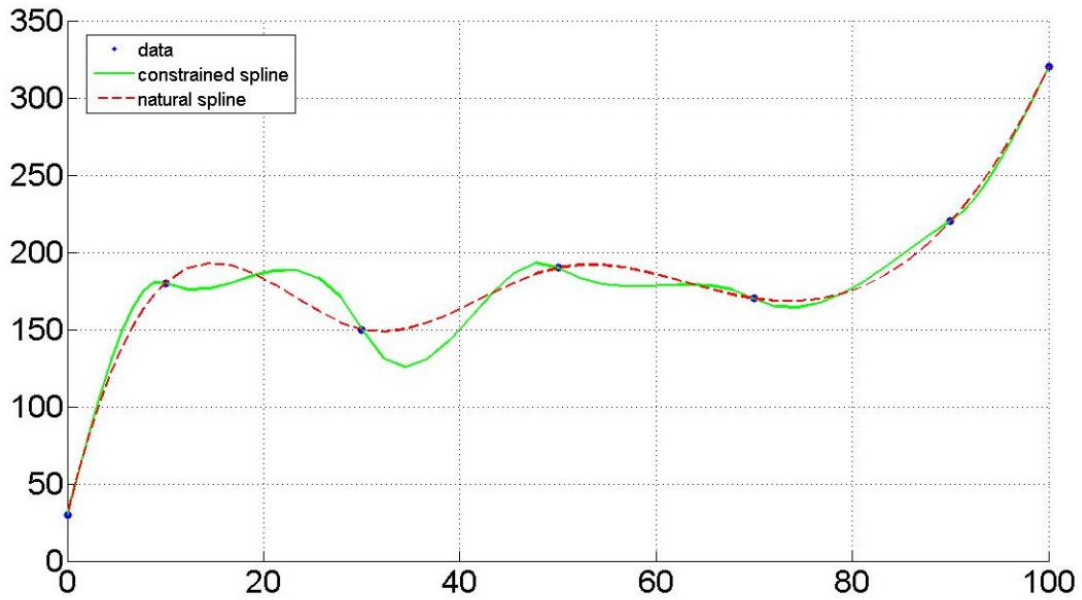

Figure 5.27 Comparison between constrained and natural splines by using the data used in Figure 4.1. Monotonicity of the data was impaired by changing the y-coordinates of the second and fourth points.

Due to the unsatisfactory results explained above, the idea of using constrained cubic spline was dropped and the traditional natural cubic spline was used instead in the analysis of TNPGN-Active data. It is not common to encounter monotonically increasing or decreasing GNSS position time series especially when the daily solutions are used. However, it is worth to investigate the utilization of the constrained cubic spline whenever possible especially in monthly combined time series. Yet, it needs detailed and time consuming exploration particularly when the number of the existing time series is huge. In TNPGN-Active case, this number is three (north-east-up) times the number of the stations (146) which equals to 438.

5.4.3 Stopping Criterion

In Figure 5.24, how a local hump near an inflection point became an extremum after the first round of sifting was displayed. Thus the obtained detail function (which is a proto-mode function) should be processed further until attaining the true IMF. This

repeated process is called “sifting”. In Figure 5.28, the effects of sifting process on AFYN monthly combined data are given. After two siftings, the result is improved but still asymmetric and there are local maxima below the zero line. After four siftings all the local maxima are positive and all the local minima are negative.

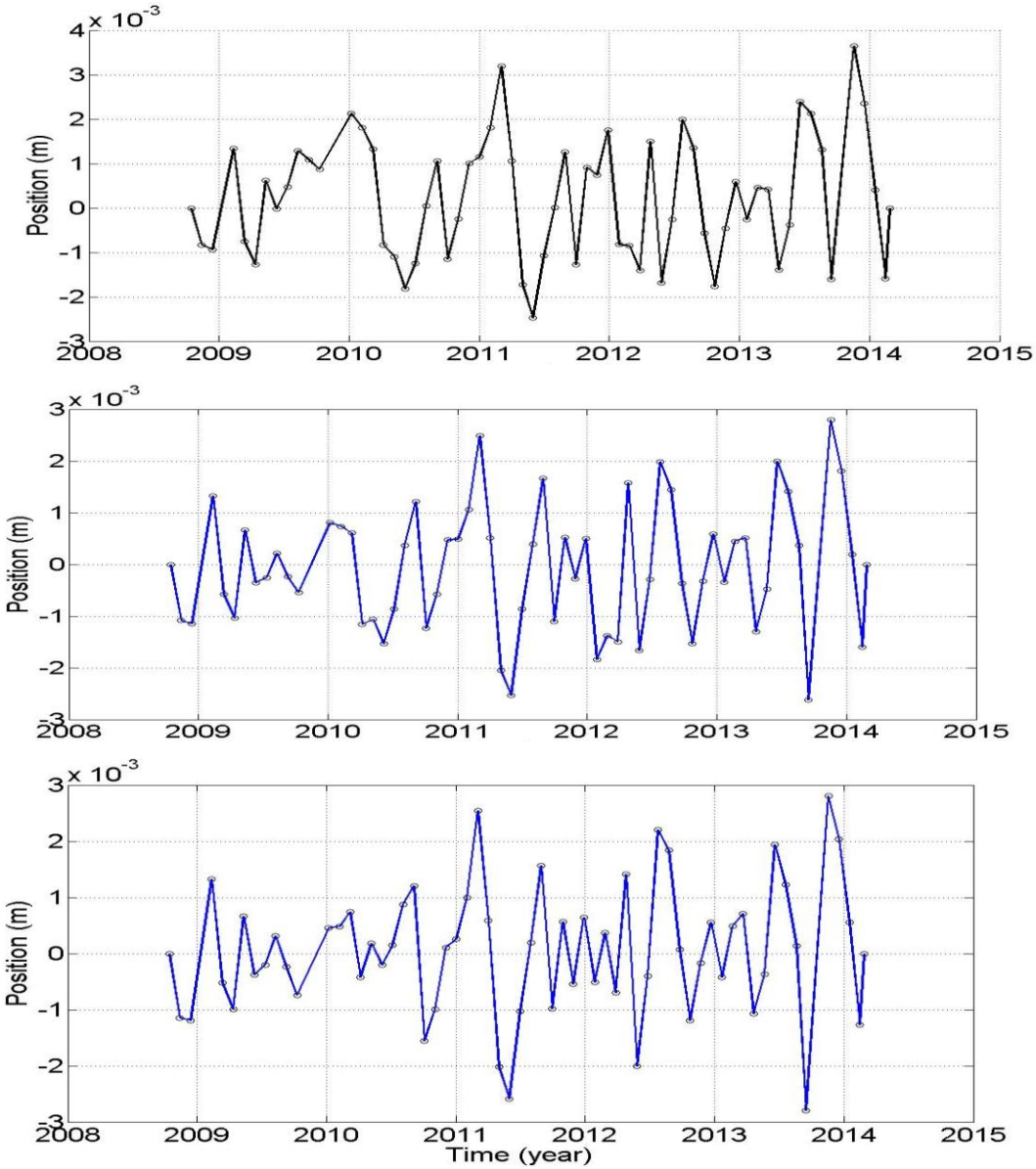

Figure 5.28 The effects of sifting process on AFYN monthly combined data. (Top) after first round of sifting, (middle) after two siftings, (bottom) after four siftings.

In Section 4.5, different stopping criteria proposed in the literature are presented. When the number of samples is small (just like in monthly combined TNPGN-Active time series), choosing a small *S* number ($3 \leq S \leq 5$) is generally giving good results. In the second and third columns of Table 5.5, numbers of computed IMFs and the corresponding *S* numbers used as a stopping criterion are given. For each time series, *S* number was chosen as 3 at the beginning. If the numbers of zero crossings and extrema are equal or differ at most by one for each IMF, calculated IMFs were accepted. If not, *S* number was incremented until they are equal or differ at most by one. In the fourth column of Table 5.5, the numbers of IMFs created by using standard deviations (*SD*) as stopping criterion are given. In the fifth column, the numbers of executed siftings for each IMF are given. It is clear that, using *S* number instead of *SD* is decreasing the numbers of siftings dramatically, and in this manner preserving the natural amplitude variations better than its competitor. Besides, it never betrays the definition of the IMF: the number of extrema and the number of zero crossings are equal or differ at most by one. This is not a must for *SD* case. However, it was observed during the analysis of monthly combined TNPGN-Active data (less than one hundred data points presently) that using *SD* as stopping criterion disobeys the aforementioned rule very rarely. In Figure 5.29, IMFs calculated from monthly combined up offsets of BAYB station using different stopping criteria are given. They are slightly different. But it is seen in Table 5.5 that the numbers of computed IMFs are not always the same.

Table 5.5 Different EMD results obtained by using *S* number and *SD* for selected monthly combined north offsets.

Station	IMF results by using *S* number		IMF results by using standard deviations (*SD*)	
	# of IMF	S number	# of IMF	# of iteration times for each IMF component, using standard deviations *SD*.
ADAN	4	6	4	11, 11, 4, 0
ADIY	2	3	2	6, 0
AFYN	3	3	3	5, 6, 0
AKDG	2	5	2	4, 0
AKHI	3	3	3	4, 7, 0

Table 5.5 (cont'd) Different EMD results obtained by using *S* number and *SD* for selected monthly combined north offsets.

AMAS	2	3	2	8, 0
AKHR	3	3	3	6, 5, 0
AKSI	4	5	4	14, 5, 4, 0
AKSR	2	3	2	5, 0
ALBH	3	4	4	9, 12, 7, 0
ANMU	3	3	3	6, 4, 0
ANRK	3	3	3	5, 9, 0
ANTA	3	3	4	9, 14, 7, 0
ANTL	3	3	3	4, 7, 0
ARTV	3	5	3	6, 9, 0
AVCT	3	5	3	14, 6, 0
AYD1	3	3	4	5, 7, 5, 0
AYVL	4	3	4	4, 3, 4, 0

Disparity is more distinct in TNPGN-Active daily time series analyses results. In Table 5.6, EMD results obtained by using *S* number and SD for all north offsets of TNPGN-Active stations are given. Station names end with lowercase "a" represent the time series after an earthquake event and station names end with lowercase "p" represent the time series before an earthquake event for they are treated as different points in GNSS analysis. Almost all time series have more than a thousand data points.

The numbers of calculated IMFs are higher than those of monthly time series. It is seen that using *SD* results in tremendous iteration numbers while *S* number gives reasonable iteration numbers. However, keeping *S* small (3-6) is not possible any more if it is desired to keep the number of extrema and the number of zero crossings equal or differ at most by one.

The pulse-quickening point is that using *SD* does not always keep the number of extrema and the number of zero crossings equal or differ at most by one. For example, in Table 5.7, it is seen that the number of zero crossings and the number of extrema differ more than by one for the first and sixth IMF components for KLIS daily north offsets.

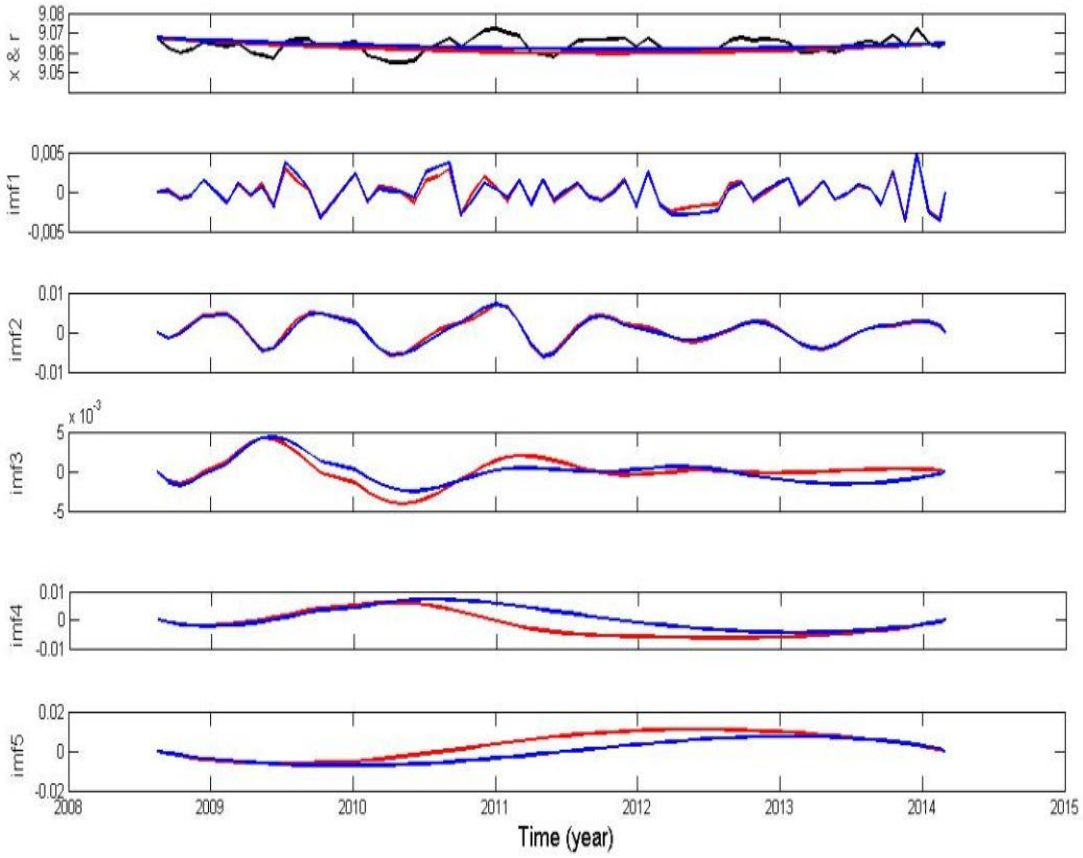

Figure 5.29 IMFs calculated from monthly combined up offsets of BAYB station using different stopping criteria. IMFs obtained by using standard deviations (SD) are shown in red color. Iteration numbers for each IMF (except residue) are 9, 5, 18, 6 and 3, respectively. IMFs obtained by choosing 4 as *S* number are shown in blue color. Original signal and the calculated residues are displayed in the top figure.

In Figure 5.30 the original KLIS north offsets are displayed. In Figure 5.31, IMFs calculated by using *SD* and by using *S* number are displayed respectively. Though, this is one of the rare situations that the *S* number (26) is bigger than each iteration numbers obtained from using *SD* (see Table 5.6).

Table 5.6 Different EMD results obtained by using *S* number and *SD* for all daily north offsets of TNPGN-Active stations.

Station	IMF results by using *S* number		IMF results by using standard deviations (*SD*)	
	# of IMF	*S* number	# of IMF	# of iteration times for each IMF component, using standard deviations *SD*.
ADAN	9	10	10	21, 29, 29, 47, 56, 44, 24, 54, 10, 0
ADIY	7	12	7	19, 19, 23, 17, 25, 23, 0
AFYN	9	14	10	28, 21, 17, 20, 19, 18, 53, 34, 48, 0
AGRDa	8	11	8	16, 18, 21, 18, 26, 9, 4, 0
AGRDp	8	21	8	12, 15, 17, 18, 26, 14, 21, 0
AKDG	7	20	7	22, 25, 24, 15, 23, 32, 0
AKHR	9	12	10	32, 38, 22, 26, 22, 28, 32, 30, 27, 0
AKSI	9	10	10	27, 19, 20, 27, 26, 32, 25, 35, 6, 0
AKSR	8	29	8	21, 25, 22, 21, 35, 23, 17, 0
AMAS	7	13	7	20, 24, 20, 14, 16, 22, 0
ANMU	7	14	8	20, 22, 18, 18, 17, 14, 18, 0
ANRK	8	16	8	24, 20, 16, 15, 40, 26, 24, 0
ANTE	7	18	7	20, 22, 29, 18, 17, 20, 0
ANTL	8	12	9	23, 17, 18, 19, 14, 18, 13, 11, 0
ARDH	8	10	8	24, 26, 29, 17, 20, 27, 34, 0
ARPK	7	12	7	24, 20, 23, 16, 15, 24, 0
ARTV	8	8	8	25, 28, 26, 15, 13, 18, 13, 0
AYD1	11	17	12	19, 22, 23, 17, 18, 21, 11, 105, 31, 4, 4, 0
AYVL	11	17	11	30, 18, 15, 15, 36, 13, 21, 47, 13, 5, 0
BALK	9	17	9	24, 20, 18, 17, 14, 19, 29, 10, 0
BAND	9	17	10	22, 30, 22, 15, 34, 16, 21, 26, 7, 0
BASKa	6	7	6	19, 13, 11, 8, 13, 0
BASKp	6	8	7	16, 15, 22, 13, 12, 25, 0
BAYB	7	11	8	26, 20, 23, 22, 12, 16, 31, 0
BEYS	10	13	10	20, 25, 19, 17, 26, 24, 53, 7, 7, 0
BILE	8	8	8	25, 20, 21, 17, 19, 30, 32, 0
BINGa	7	16	8	26, 27, 20, 14, 13, 18, 15, 0
BINGp	7	15	6	18, 11, 12, 12, 8, 0
BOGZ	9	16	12	30, 30, 26, 17, 37, 92, 129, 45, 53, 12, 10, 0
BOLU	8	19	8	22, 29, 20, 26, 22, 12, 16, 0
BOYT	8	14	8	23, 21, 16, 18, 32, 18, 86, 0
BTMN	7	25	7	31, 25, 20, 22, 20, 16, 0
BURS	9	58	9	25, 26, 27, 13, 18, 19, 14, 20, 0
CANA	9	14	10	29, 20, 17, 21, 17, 39, 13, 17, 37, 0
CANK	7	12	7	22, 19, 15, 17, 16, 23, 0
CATKa	7	10	9	15, 14, 16, 17, 26, 25, 55, 11, 0
CATKp	8	14	8	20, 18, 19, 21, 10, 65, 20, 0
CAVD	10	6	10	16, 18, 13, 19, 21, 9, 58, 11, 10, 0
CESM	8	9	9	22, 18, 14, 19, 10, 22, 41, 83, 0
CIHA	7	13	8	20, 23, 27, 13, 14, 18, 17, 0
CMLD	8	18	9	25, 32, 22, 15, 28, 14, 22, 19, 0
CORU	8	15	8	20, 27, 18, 18, 18, 26, 21, 0
DATCa	9	8	9	13, 11, 14, 14, 7, 28, 6, 6, 0
DATCp	10	39	9	26, 18, 15, 16, 27, 60, 27, 40, 0
DEIR	9	19	9	24, 20, 30, 15, 13, 11, 6, 14, 0

Table 5.6 (cont'd) Different EMD results obtained by using *S* number and *SD* for all daily north offsets of TNPGN-Active stations.

DIDI	9	16	9	21, 21, 19, 14, 12, 25, 64, 23, 0
DENI	12	45	13	46, 42, 32, 29, 39, 21, 36, 39, 33, 52, 10, 6, 0
DINA	7	5	9	19, 25, 23, 12, 21, 25, 27, 11, 0
DIPK	6	6	6	13, 14, 14, 14, 12, 0
DIVR	7	15	7	29, 28, 18, 29, 18, 13, 0
DIYBa	7	13	7	23, 24, 14, 20, 16, 17, 0
DIYBp	6	5	6	11, 12, 11, 13, 9, 0
EDIR	8	19	8	24, 22, 20, 12, 21, 19, 30, 0
EKIZ	7	8	8	25, 34, 20, 28, 19, 14, 63, 0
ELAZ	7	13	7	21, 17, 17, 19, 15, 20, 0
EMIR	9	10	9	20, 26, 23, 14, 20, 15, 27, 34, 0
ERGN	9	46	8	28, 30, 23, 25, 21, 29, 28, 0
ERZI	8	37	8	19, 18, 18, 15, 18, 12, 17, 0
ERZR	10	20	10	26, 35, 25, 25, 29, 32, 173, 142, 9, 0
ESKS	8	11	8	18, 19, 30, 13, 18, 15, 13, 0
FASA	8	22	8	34, 21, 19, 20, 16, 35, 15, 0
FEEK	8	20	8	26, 26, 22, 20, 21, 15, 18, 0
FETH	13	29	13	24, 29, 20, 23, 22, 36, 17, 24, 56, 28, 5, 5, 0
FINI	8	6	10	28, 23, 24, 26, 25, 21, 29, 34, 47, 0
GEME	8	19	8	21, 17, 19, 17, 21, 21, 34, 0
GIRS	8	13	8	25, 21, 27, 21, 23, 17, 32, 0
GUMU	10	22	11	20, 24, 22, 13, 29, 19, 70, 102, 16, 14, 0
GURU	8	42	7	35, 30, 20, 20, 14, 14, 0
GYUR	8	14	8	20, 15, 14, 21, 18, 20, 28, 0
HAKKa	9	8	11	17, 12, 15, 26, 37, 45, 35, 13, 13, 7, 0
HAKKp	8	10	10	23, 14, 28, 13, 15, 15, 24, 34, 6, 0
HALP	8	15	9	41, 42, 34, 29, 20, 27, 33, 12, 0
HARC	9	12	10	20, 31, 19, 18, 18, 20, 27, 30, 36, 0
HATA	8	21	8	27, 38, 25, 18, 23, 16, 17, 0
HEND	10	27	9	29, 20, 17, 29, 34, 50, 170, 11, 0
HINI	8	37	8	30, 32, 26, 37, 22, 19, 29, 0
HORS	10	18	10	22, 22, 25, 14, 21, 28, 56, 61, 63, 0
HYMN	9	27	10	22, 20, 16, 19, 20, 16, 39, 168, 58, 0
IGIRa	10	10	9	21, 14, 20, 13, 21, 36, 21, 5, 0
IGIRp	9	7	7	15, 19, 13, 25, 25, 10, 0
INEB	9	16	9	18, 17, 15, 22, 31, 17, 47, 32, 0
IPSA	9	13	8	18, 16, 18, 19, 19, 27, 16, 0
ISPT	10	23	9	29, 25, 15, 19, 48, 13, 63, 12, 0
ISTN	9	12	10	20, 17, 26, 44, 66, 15, 40, 173, 30, 0
IZMI	14	43	13	31, 34, 30, 16, 19, 17, 63, 52, 52, 133, 39, 6, 0
IZMT	9	17	9	22, 18, 20, 16, 18, 35, 27, 32, 0
KAMN	9	20	9	21, 17, 21, 19, 20, 65, 44, 125, 0
KAPN	9	13	9	23, 24, 21, 17, 25, 43, 28, 11, 0
KARB	8	30	8	18, 21, 20, 15, 35, 14, 21, 0
KAYS	9	21	10	23, 25, 23, 25, 21, 56, 102, 38, 29, 0
KIKA	9	16	9	25, 21, 23, 17, 15, 19, 21, 7, 0
KIRL	7	9	8	20, 26, 19, 15, 16, 21, 17, 0
KIRS	10	16	11	24, 19, 29, 16, 24, 19, 78, 139, 39, 12, 0
KKAL	7	8	9	25, 19, 18, 15, 24, 27, 22, 16, 0
KLIS	7	26	7	24, 20, 25, 19, 19, 13, 0
KLUU	7	11	9	21, 24, 17, 18, 19, 27, 19, 8, 0
KNYA	9	23	11	24, 30, 25, 30, 30, 37, 39, 32, 69, 12, 0
KRBK	9	19	8	22, 25, 17, 17, 26, 16, 23, 0

Table 5.6 (cont'd) Different EMD results obtained by using *S* number and *SD* for all daily north offsets of TNPGN-Active stations.

KRS1a	9	6	11	14, 25, 13, 19, 31, 39, 38, 14, 8, 6, 0
KRS1p	6	11	7	20, 18, 14, 11, 18, 11, 0
KSTM	8	24	9	25, 22, 29, 14, 25, 37, 10, 7, 0
KURU	7	15	8	23, 22, 25, 14, 15, 18, 21, 0
KUTA	11	43	11	28, 30, 23, 24, 21, 21, 44, 42, 81, 40, 0
LEFK	7	12	7	18, 29, 14, 12, 12, 17, 0
MALY	7	15	8	21, 17, 20, 19, 26, 23, 12, 0
MALZ	9	10	11	28, 28, 23, 26, 29, 77, 47, 132, 31, 4, 0
MARA	7	9	8	22, 24, 21, 18, 26, 20, 20, 0
MARD	10	36	9	20, 24, 25, 20, 31, 30, 34, 22, 0
MGOS	7	9	9	17, 17, 18, 21, 28, 23, 62, 31, 0
MDIY	7	12	7	25, 29, 22, 23, 22, 12, 0
MRSI	8	23	8	25, 20, 17, 20, 17, 14, 14, 0
MUGL	9	20	9	27, 24, 19, 17, 25, 13, 19, 5, 0
MURAa	9	8	10	15, 14, 14, 21, 16, 43, 69, 11, 7, 0
MURAp	7	14	7	18, 17, 16, 14, 15, 13, 0
MUUS	9	17	9	28, 34, 28, 26, 33, 36, 30, 14, 0
NAHA	10	13	10	19, 22, 29, 24, 26, 37, 77, 69, 10, 0
NEVS	9	50	8	26, 22, 20, 22, 19, 27, 46, 0
NIGD	8	19	7	24, 23, 18, 18, 32, 18, 0
ONIY	7	18	7	21, 20, 22, 19, 17, 17, 0
OZALa	6	6	6	14, 20, 13, 14, 14, 0
OZALp	6	8	7	16, 17, 14, 14, 8, 13, 0
POZA	8	29	8	28, 28, 40, 16, 16, 19, 30, 0
RDIY	12	53	7	31, 27, 35, 19, 26, 20, 0
RHIY	8	18	8	25, 26, 25, 22, 30, 32, 21, 0
RZE1	10	28	10	22, 38, 17, 25, 23, 25, 53, 87, 12, 0
SALH	11	23	11	30, 33, 25, 17, 22, 22, 27, 15, 9, 4, 0
SAMN	9	16	10	22, 27, 17, 17, 23, 41, 43, 39, 9, 0
SARVa	9	17	9	20, 21, 14, 14, 30, 17, 24, 5, 0
SARVp	9	3	9	16, 20, 23, 14, 17, 23, 31, 10, 0
SARY	8	8	7	31, 17, 18, 22, 14, 16, 0
SEMD	9	23	9	25, 29, 20, 22, 15, 78, 36, 29, 0
SIHI	9	20	10	24, 18, 16, 18, 36, 24, 30, 57, 12, 0
SILF	8	12	9	25, 22, 19, 18, 19, 22, 50, 17, 0
SINP	8	13	8	22, 20, 21, 23, 31, 13, 14, 0
SIRN	9	26	8	23, 19, 25, 19, 34, 35, 29, 0
SIRT	7	8	9	28, 22, 30, 21, 21, 33, 35, 31, 0
SIVE	7	14	7	24, 27, 22, 23, 16, 13, 0
SLEE	8	17	10	20, 17, 29, 19, 52, 34, 14, 58, 8, 0
SSEH	7	16	7	32, 23, 22, 17, 18, 16, 0
SUNL	7	12	8	30, 25, 22, 15, 22, 22, 53, 0
SURF	7	15	7	25, 19, 18, 20, 14, 27, 0
TEKR	9	22	9	25, 19, 19, 14, 12, 50, 42, 19, 0
TNCE	8	15	9	22, 19, 19, 26, 22, 73, 27, 37, 0
TOKA	10	21	9	24, 21, 21, 20, 27, 31, 107, 35, 0
TRBN	7	16	8	29, 20, 18, 26, 27, 22, 23, 0
TUFA	8	16	8	22, 21, 22, 18, 19, 14, 42, 0
TVANa	7	8	7	17, 16, 10, 16, 17, 12, 0
TVANp	7	11	7	33, 17, 20, 25, 10, 20, 0
UDER	7	14	8	24, 25, 21, 16, 13, 31, 31, 0
USAK	11	19	12	21, 20, 26, 18, 45, 75, 89, 37, 48, 11, 8, 0
VAANa	4	4	4	8, 8, 9, 0

Table 5.6 (cont'd) Different EMD results obtained by using *S* number and *SD* for all daily north offsets of TNPGN-Active stations.

VAANp	7	8	7	19, 16, 18, 16, 16, 7, 0
VIRA	7	14	7	23, 22, 18, 18, 17, 24, 0
YENC	12	21	11	29, 32, 24, 18, 25, 23, 82, 44, 45, 6, 0
YOZT	9	12	9	24, 27, 22, 18, 32, 40, 49, 41, 0
YUNK	9	9	10	22, 28, 20, 19, 20, 25, 88, 15, 9, 0
ZONG	8	19	9	20, 16, 20, 34, 27, 18, 50, 46, 0

Table 5.7 Comparison of zero crossings and extrema for KLIS north offsets.

No.	IMFs calculated by using SD		IMFs calculated by using *S* number	
	# of zero crossings	# of extrema	# of zero crossings	# of extrema
1	1173	1176	1175	1176
2	597	598	609	610
3	318	319	328	329
4	176	177	192	193
5	97	98	100	100
6	43	46	47	48
Res.	0	1	0	1

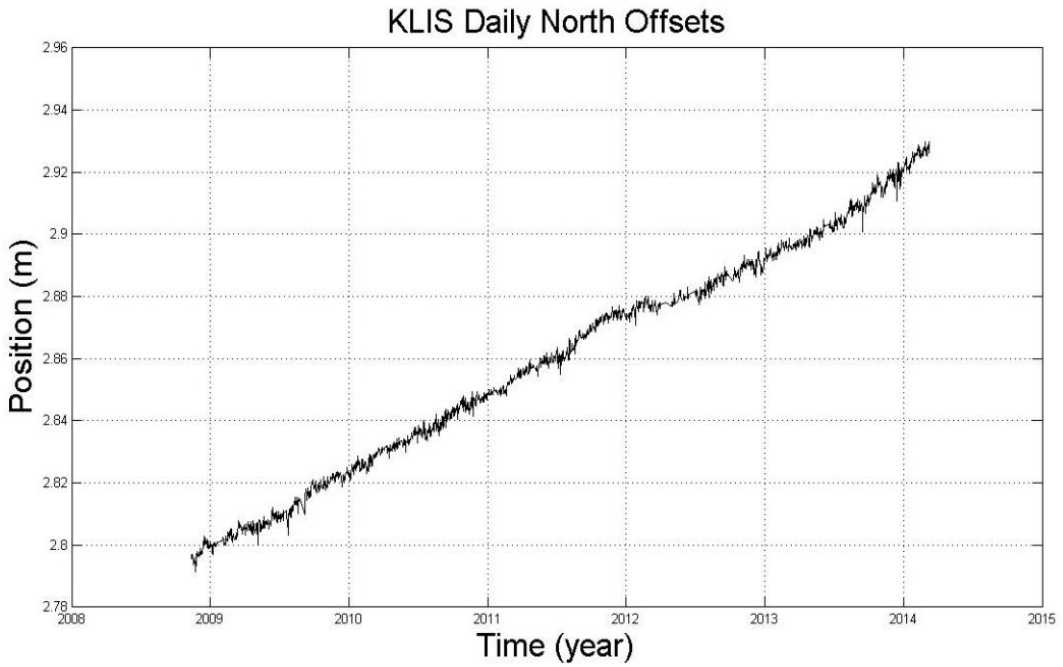

Figure 5.30 Daily north offsets of KLIS station.

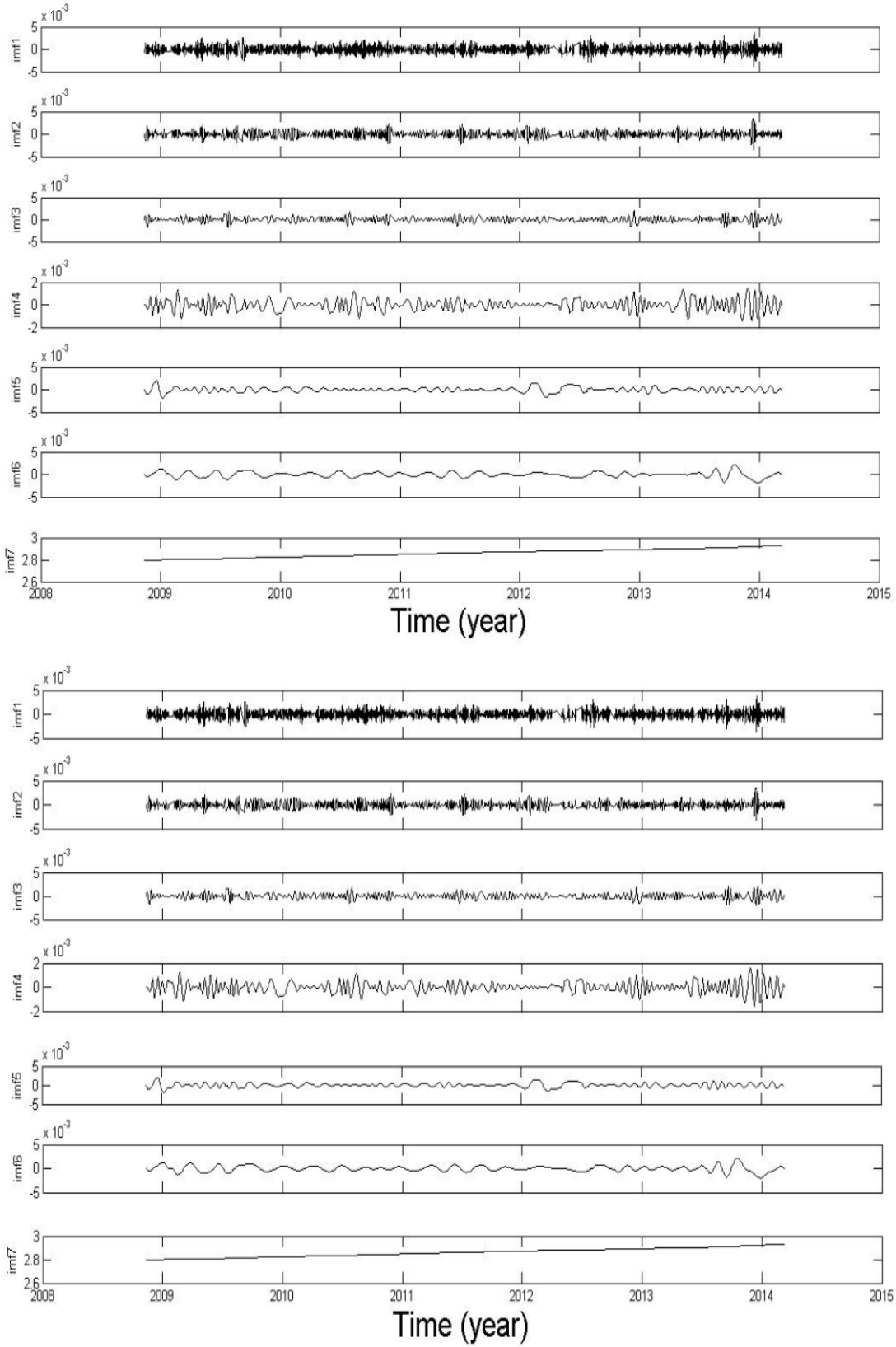

Figure 5.31 The resulting EMD components from KLIS data: (top) IMFs with stopping criterion *SD*; (bottom) IMFs with stopping criterion *S*.

The differences between the IMFs calculated by using *SD* and *S* are almost indiscernable for KLIS station for the iteration numbers are so close to each other (see Table 5.6). Nevertheless, the fact that *SD* disobeys the basic IMF rule cannot be ignored. Further, results are not always similar as in KLIS case. In Figure 5.32, the resulting EMD components from daily north offsets of ISTN station by using *SD* are shown. There are ten IMFs.

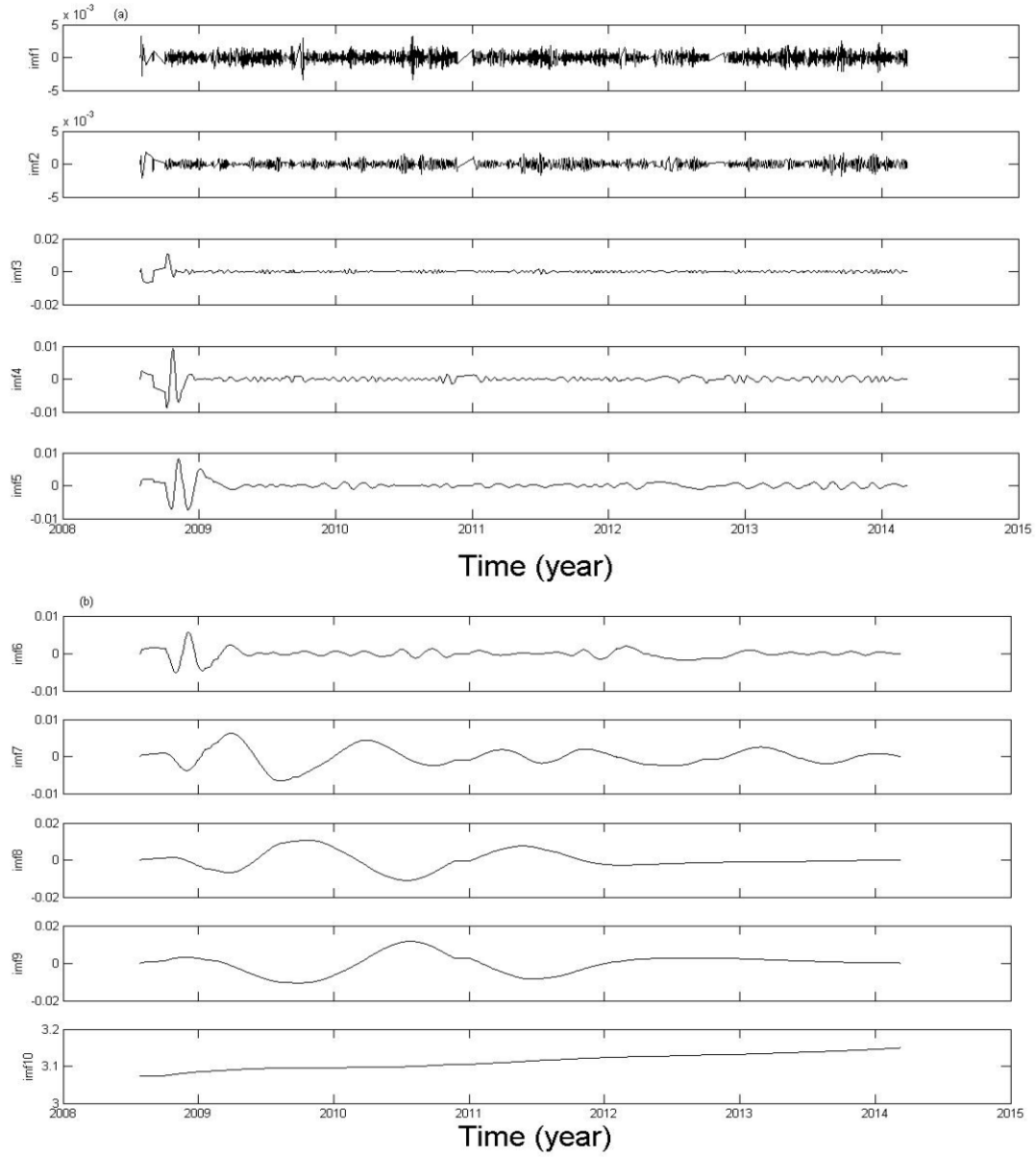

Figure 5.32 The resulting EMD components from ISTN data by using *SD* stopping criterion: (a) EMD components 1-5; (b) EMD components 6-10.

In Figure 5.33 EMD components from the same ISTN data by using *S* number are given. This time there are nine IMFs. Obviously, results are not so similar to those in Figure 5.32.

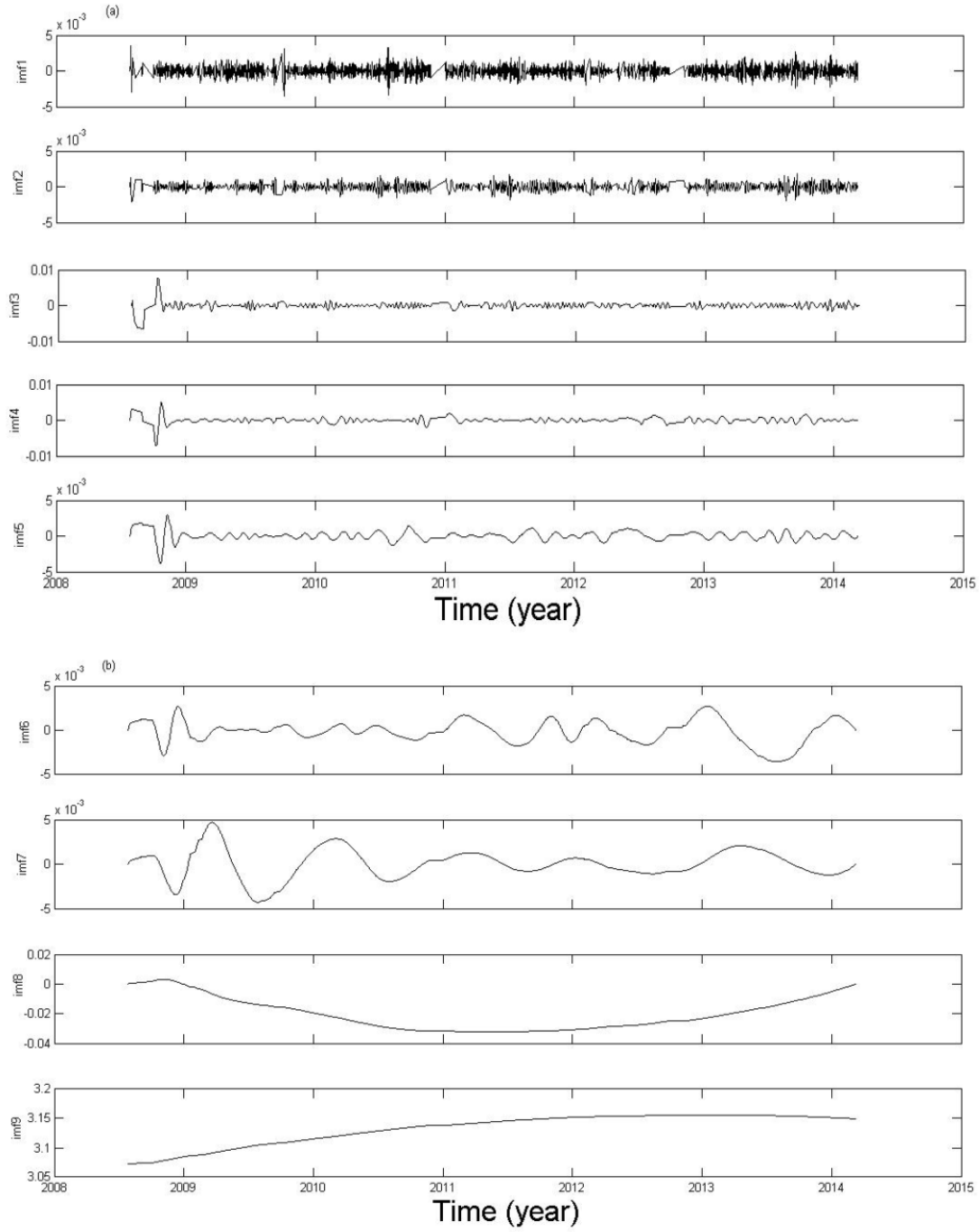

Figure 5.33 The resulting EMD components from ISTN data by using *S* (12) stopping criterion: (a) EMD components 1-5; (b) EMD components 6-9.

On the other hand, illustration of the residues (which are expected to be a constant or trend) obtained by using *SD* and *S* as stopping criteria is highly intriguing. In Figure 5.34, residues delivered by the above explained approaches are plotted on the original ISTN daily north offset data.

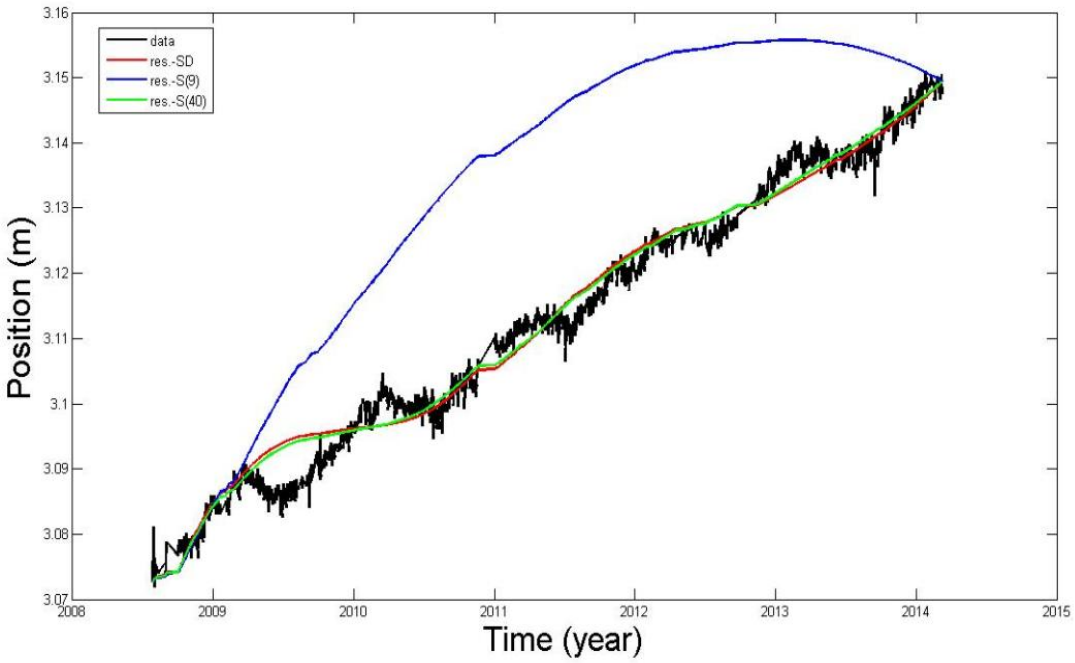

Figure 5.34 Residues obtained from different approaches for ISTN station; blue line: *S*=12, green line: *S*=40, red line: standard deviation (*SD*).

It is seen that choosing *S* number as 12 (which is keeping the number of the zero crossings and the number of extrema equal or differ at most by one for each IMF) yields a residue (line in blue color) somewhat different from the other residues. Thus we would hardly name it as the trend of the time series. Obviously, it is easier to accept the residue (line in red color) obtained from *SD* approach as the trend of the process. Yet, increasing the *S* number and choosing for example as 40 (line in green color) pulls down the residue to the level of the residue of *SD* approach, albeit, it causes over-sifting.

Actually, detailed inspection of Figure 5.34 reveals that it is not the residue which represents the trend, but it also reveals that summing up the last couple IMFs and the residue would produce the trend of the process. How to find the trend is explained in Section 4.4.

After the examination of TNPGN-Active results, it was seen that to select 10 as a fixed sifting number generally gives good results. But resolution factor approach was also applied to TNPGN-Active time series. According to the characteristics of the data, sometimes *R* was chosen as 40 dB and sometimes 50 dB. It was observed that using *R* as the stopping criterion generally gives quite good results with small iteration numbers (most of the time even less than 10 times). For integrity, EMD results for the same ISTN data by using R = 40dB and R = 50dB are given in Figure 5.35 and Figure 5.36 respectively.

Contrary to the stopping criterion based on iteration numbers, resolution factor allows to set the number of IMFs. Traditional EMD algorithm generally results in close to but no more than $log_2 N$ IMF components for a time series with length *N*. At the moment, TNPGN-Active time series usually include approximately 2000 data points. $log_2 2000$ equals to almost 11. Indeed, it can be seen in Table 5.6 that the number of the computed IMFs are almost always less than 11.

After the inspection of various stopping criteria in this thesis, it was conceived that the proposed stopping criteria in the literature are devoid of strong theoretical bases. Rather, they were created based on the researchers' experience during their research. More or less they disregard the rules of IMF construction. They secure the symmetry of upper and lower envelopes only in an approximate sense and they do not always ensure the equality of the zero crossings and extrema numbers. Maybe it is impossible to define a perfect (in every sense) stopping criterion. Nonetheless, specific stopping criteria can be designed according to the needs and the characteristics of the phenomena. As explained, to keep the EMD temporally local, the sifting number should be fixed to a low number. Thus, instead of using standard deviations approach, the sifting number was fixed to 10, through this study.

Whenever the EMD results give rise to suspicion, resolution factor was used which also gives satisfactory results.

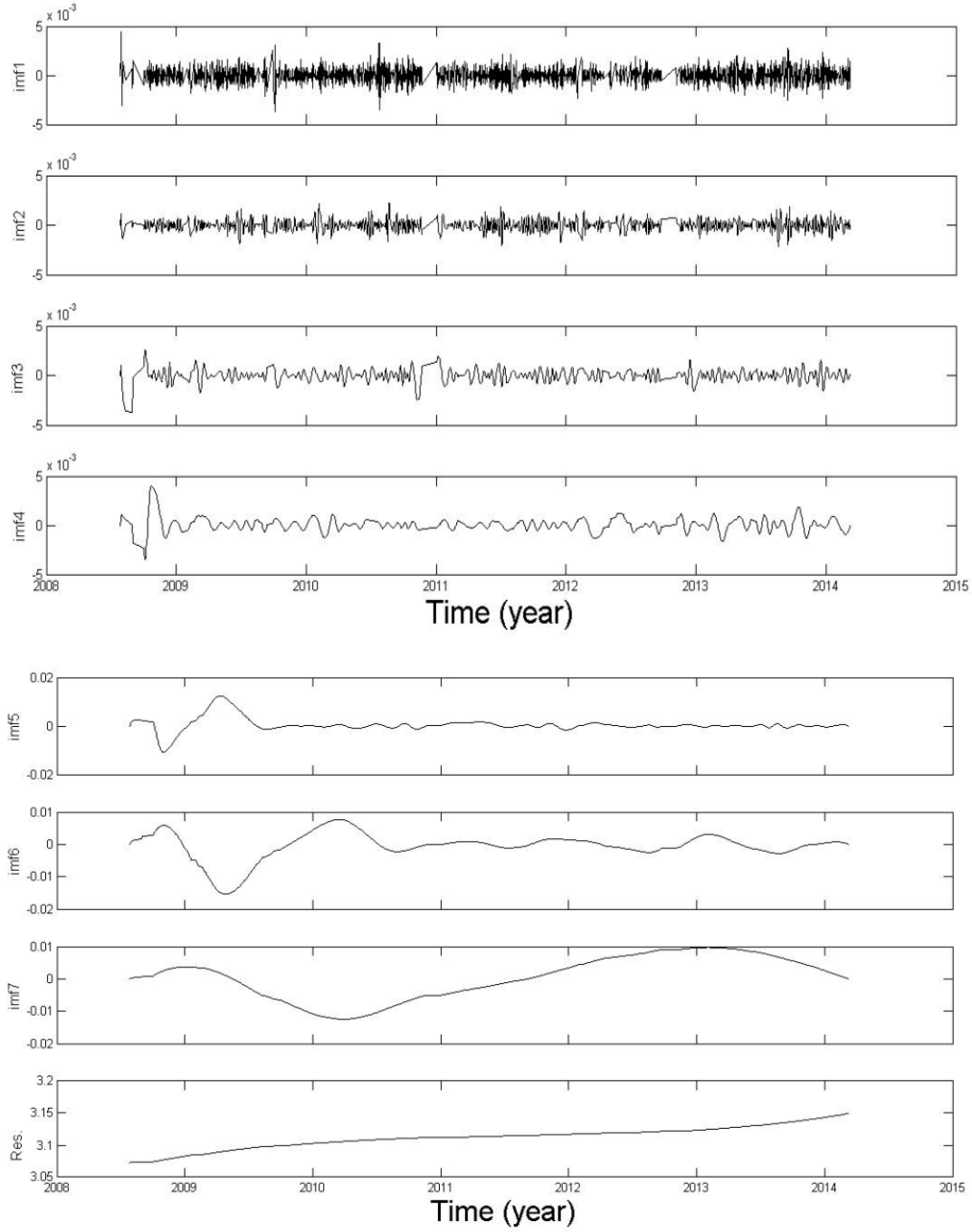

Figure 5.35 EMD results for ISTN daily north offsets by using R=40dB as stopping criterion. (Top) IMF components 1-4; (Bottom) IMF components 5-7 and the residue.

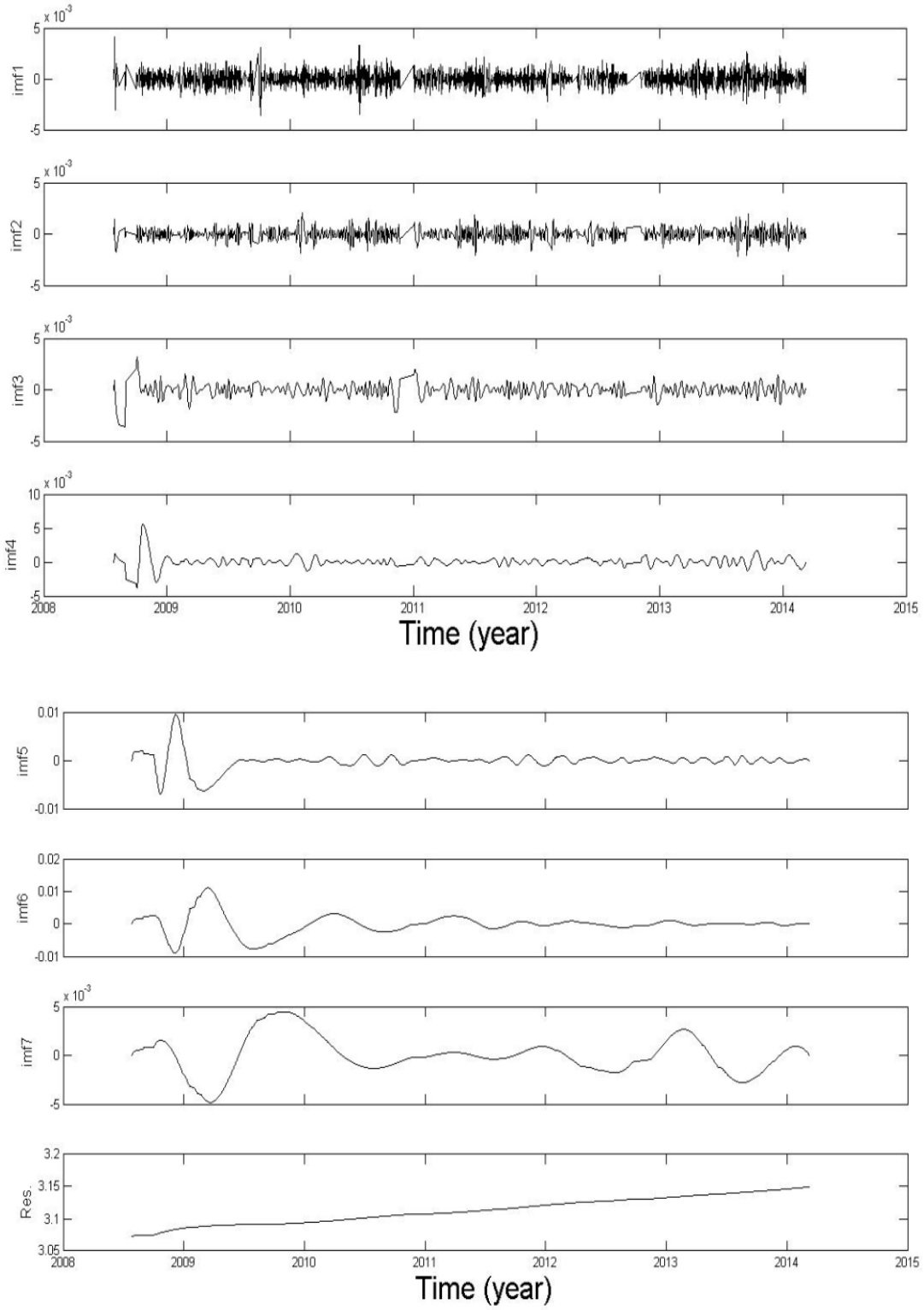

Figure 5.36 EMD results for ISTN daily north offsets by using R=50dB as stopping criterion. (Top) IMF components 1-4; (Bottom) IMF components 5-7 and the residue.

As can be seen in Figure 5.35, ISTN data was decomposed to eight components together with the residue. In comparison with the traditional Fourier Transform, this is an appreciable number. It is also seen that the data was decomposed into locally non-overlapping time scale components.

5.4.4 Completeness Test

For completeness test, EMD results for ISTN station illustrated in Figure 5.33 are used. In Figure 5.37, the reconstruction of tha data from the IMF components starting from the longest to shortest periods is given. By successively adding the higher frequency IMFs, the data could be reconstructed almost perfectly.

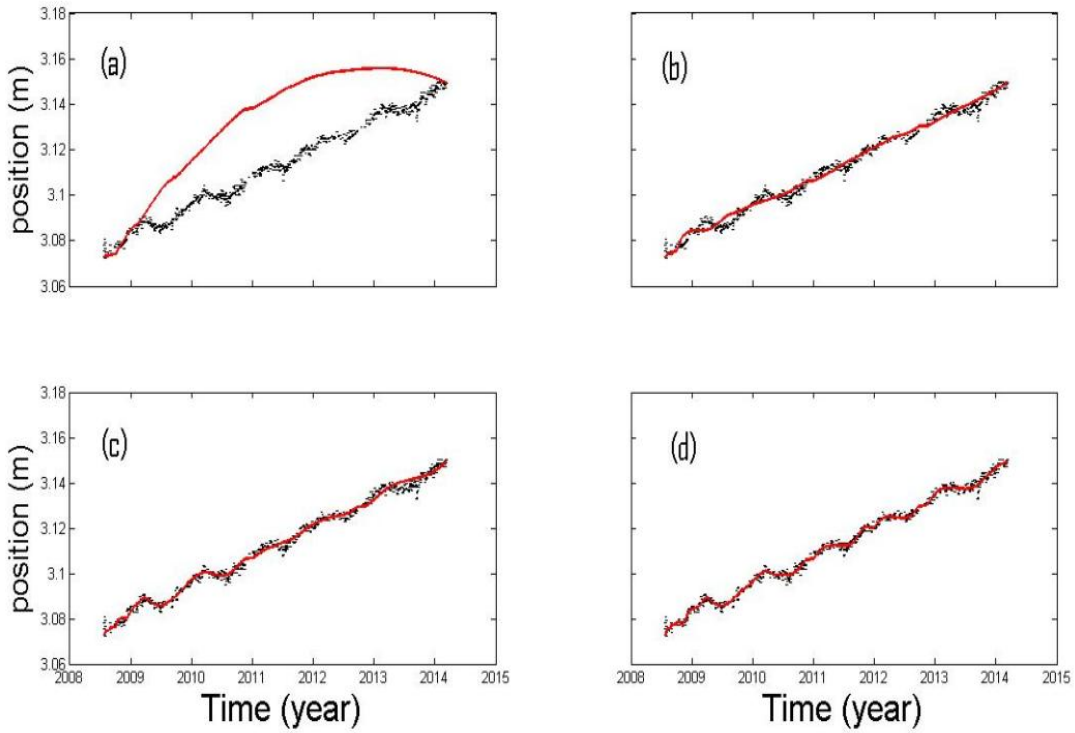

Figure 5.37 Reconstruction of the original data from the IMF components. (a) Data (in the dotted black line) and the residue (in the solid red line). (b) Data and the sum of residue and eighth component. (c) Data and the sum of residue-eighth-seventh components. (d) Data and the sum of residue-eighth-seventh-sixth components.

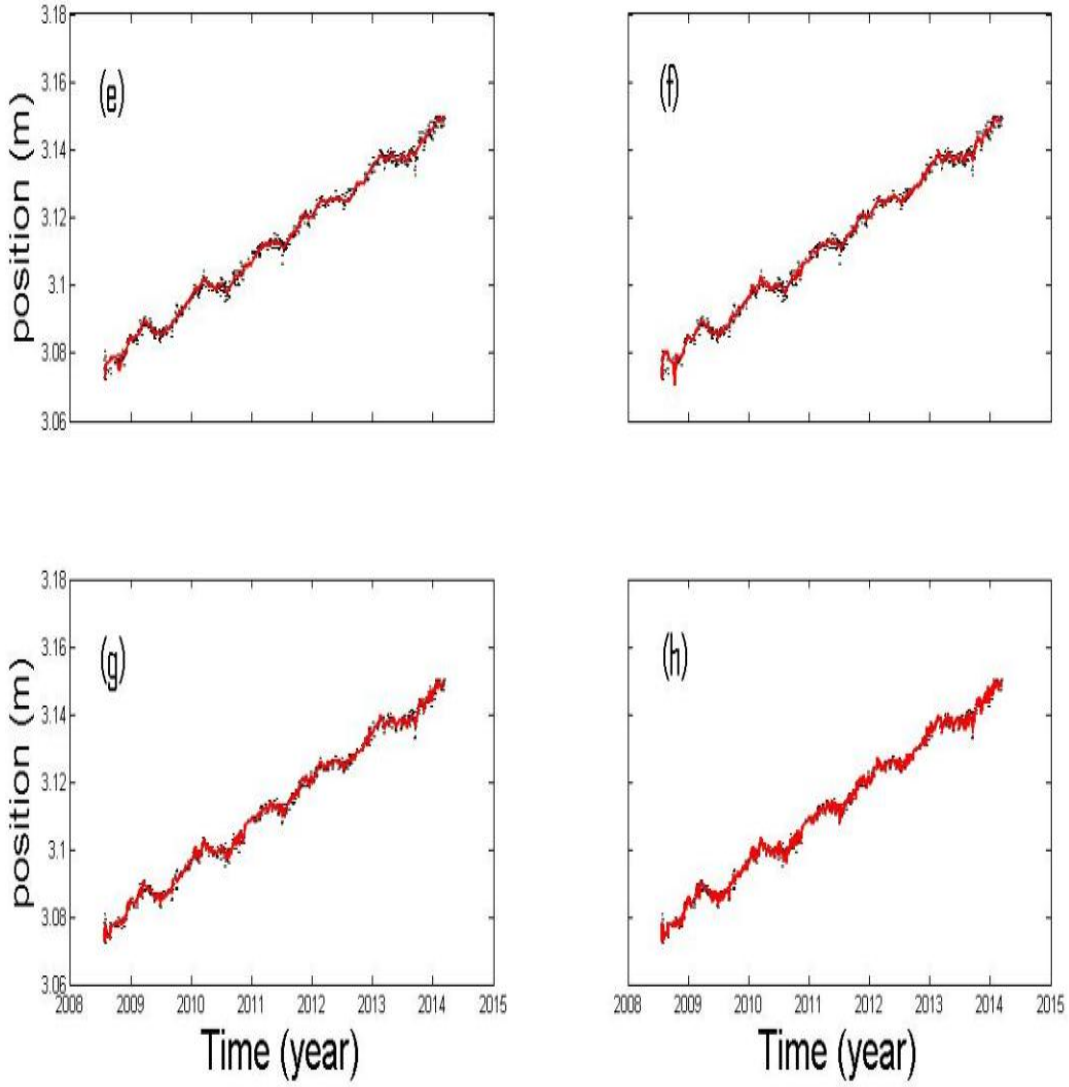

Figure 5.37 (cont'd) (e) Data (in the dotted black line) and the sum of residue-eighth-seventh-sixth-fifth components. (f) Data and the sum of residue-eighth-seventh-sixth-fifth-fourth components. (g) Data and the sum of residue-eighth-seventh-sixth-fifth-fourth-third components. (h) Data and the sum of residue-eighth-seventh-sixth-fifth-fourth-third-second components.

It can be seen that the sum of the IMF components until the third component (in Figure 5.37g) essentially covers all the energy containing the plaits. First two components (having highest frequencies) add little more energy. Most probably, the highest frequency component represents the noise.

In Figure 5.38, the final reconstruction of the data from the IMFs and the differences between the original data and the reconstructed data are displayed. The maximum amplitude of the differences is 1.33×10^{-15}, the roundoff error from the precision of the computer.

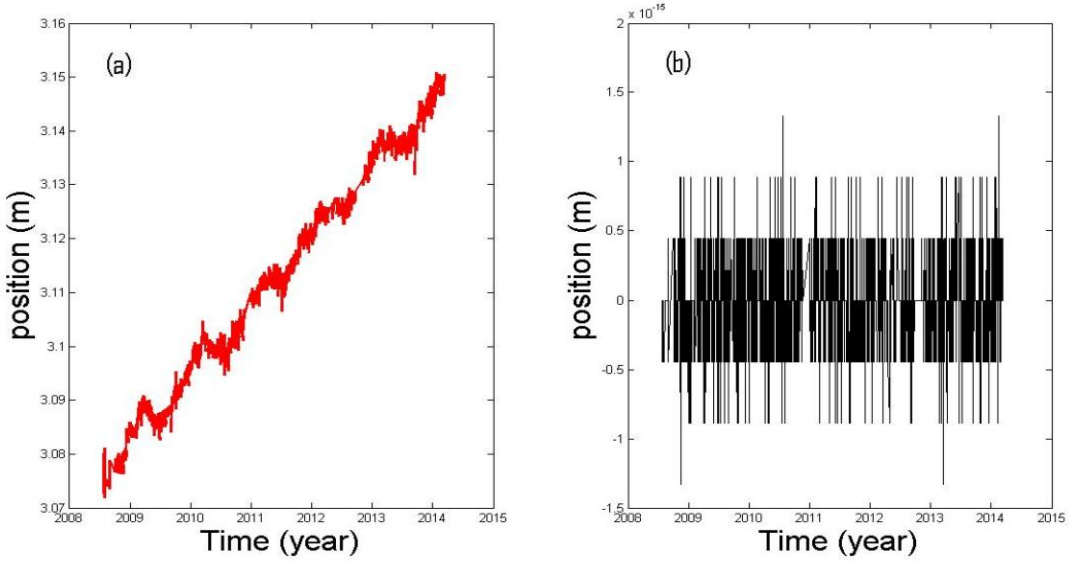

Figure 5.38 (a) The final reconstruction of the data from the IMFs. (b) The differences between the original data and the reconstructed one.

5.4.5 Detrending with EMD and Its Effect on Fast Fourier Transform

For detrending the data, the method proposed in Flandrin et.al. (2004) was used in the thesis (see Section 4.4). In Figure 5.39, the standardized empirical means of the computed IMFs for ISTN station are shown. It can be seen that after the fifth IMF, empirical mean significantly departs from zero. Hence, the IMF components from sixth to nine form the trend of the data. Accordingly, the obtained trend for ISTN station can be seen in Figure 5.40.

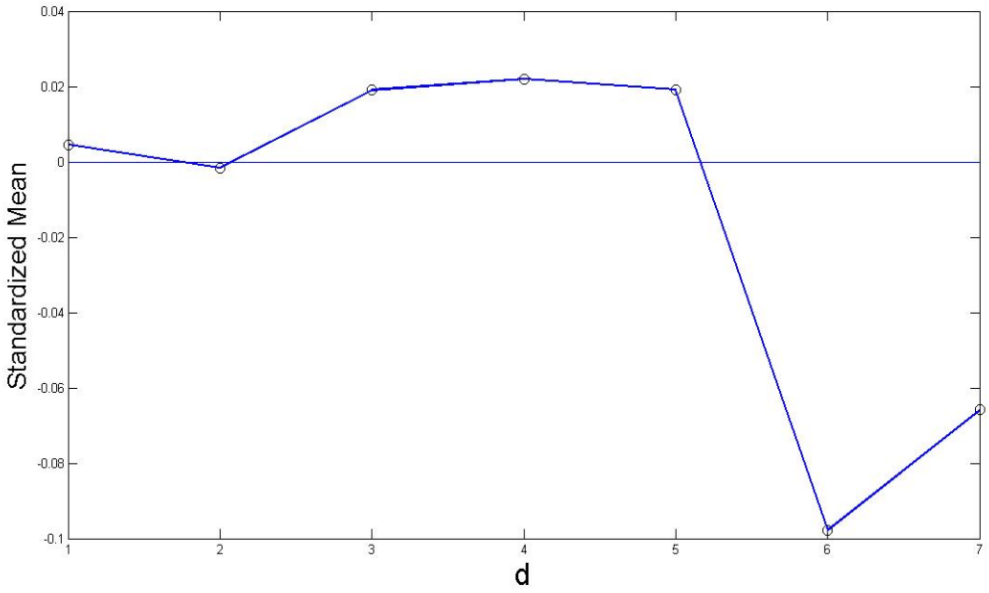

Figure 5.39 Standardized empirical mean of the fine-to-coarse EMD reconstruction for ISTN station. *D = 5* is the change point.

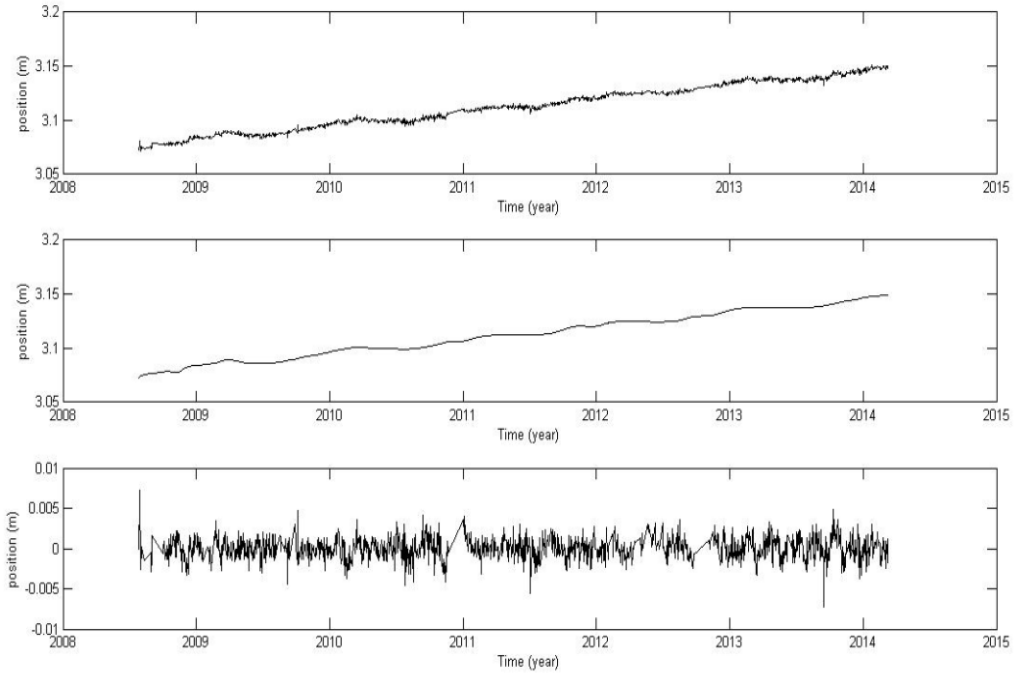

Figure 5.40 Detrending results for ISTN station. (Top) original signal, (middle) estimated trend obtained from the partial reconstruction with IMFs 6 to 8 and the residue, (bottom) detrended signal obtained from the partial reconstruction with IMFs 1 to 5.

Now, the effect of detrending the signal by EMD prior to Fast Fourier Transform (FFT) processing can be shown in comparison with removing a linear trend from the signal. For this purpose, daily north offsets of KAPN station were used. In Figure 5.41, KAPN north offsets, detrended signal by using EMD, and detrended signal by using linear trend can be seen. Even without moving further, it is seen how successfully the KAPN signal was detrended by EMD in this figure. In Figure 5.42, single-sided amplitude spectrum of KAPN signal after applying Discrete Fourier Transform (DFT) is illustrated.

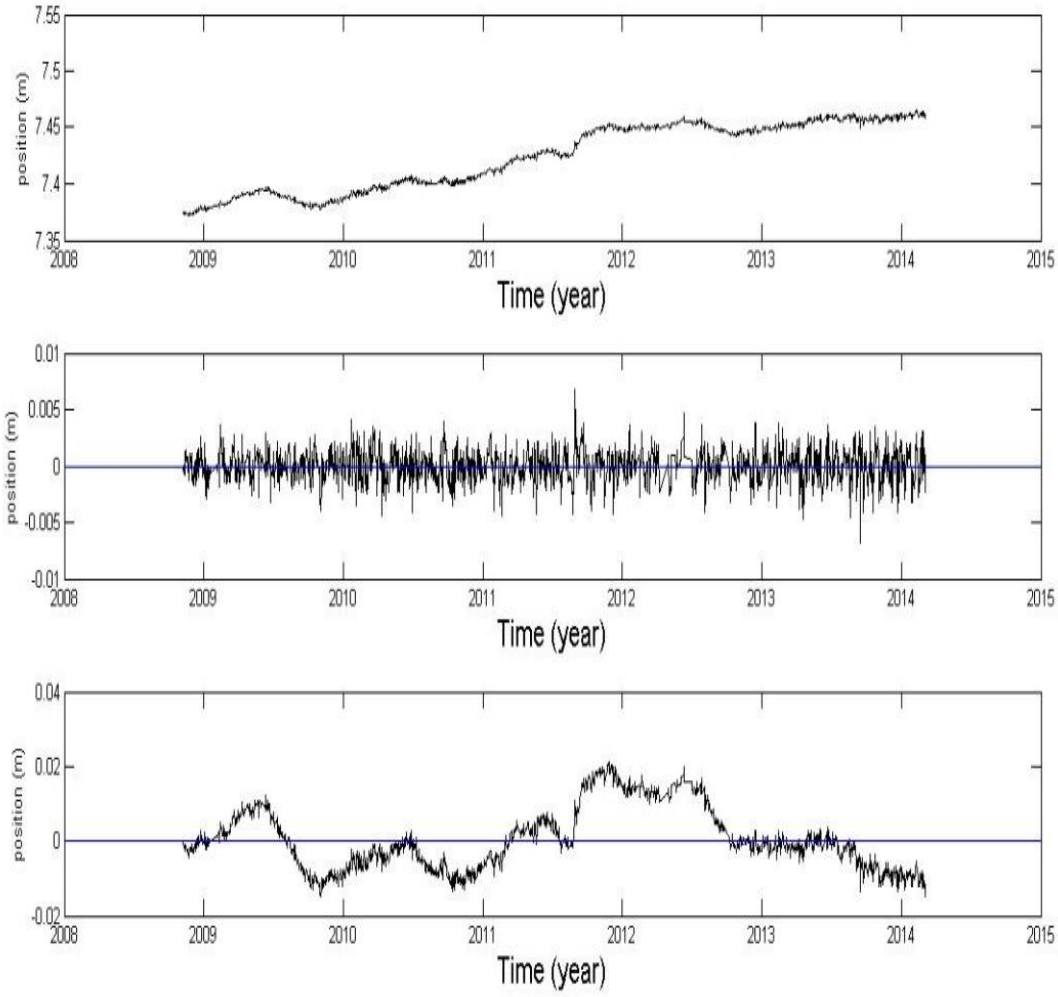

Figure 5.41 (Top) KAPN daily north offsets, (Middle) Detrended KAPN signal by using EMD, (Bottom) Detrended KAPN signal by fitting a linear trend.

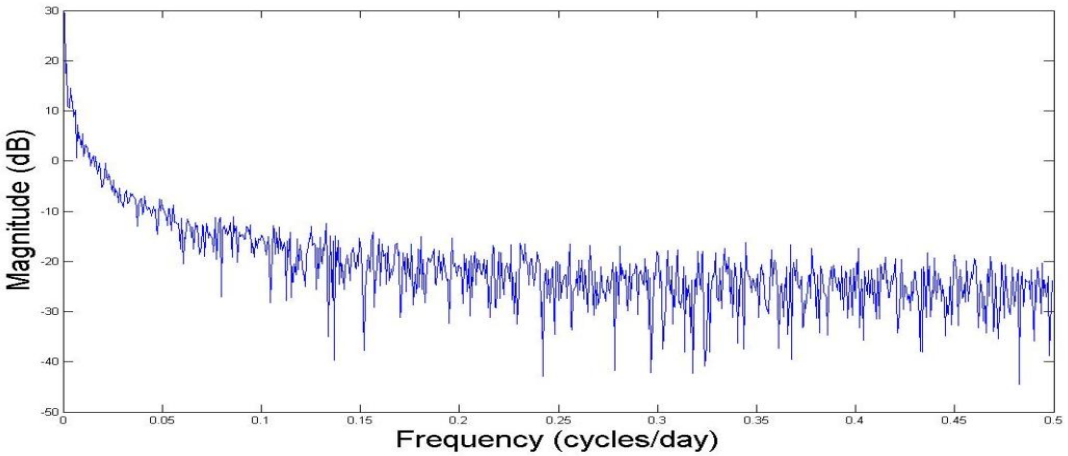

Figure 5.42 Single-sided amplitude spectrum of KAPN north offsets. The magnitude is plotted in logarithmic scale. The DC component was removed for better visualization.

In Figure 5.43, magnitude responses of detrended signals (with EMD and linear fit) are given.

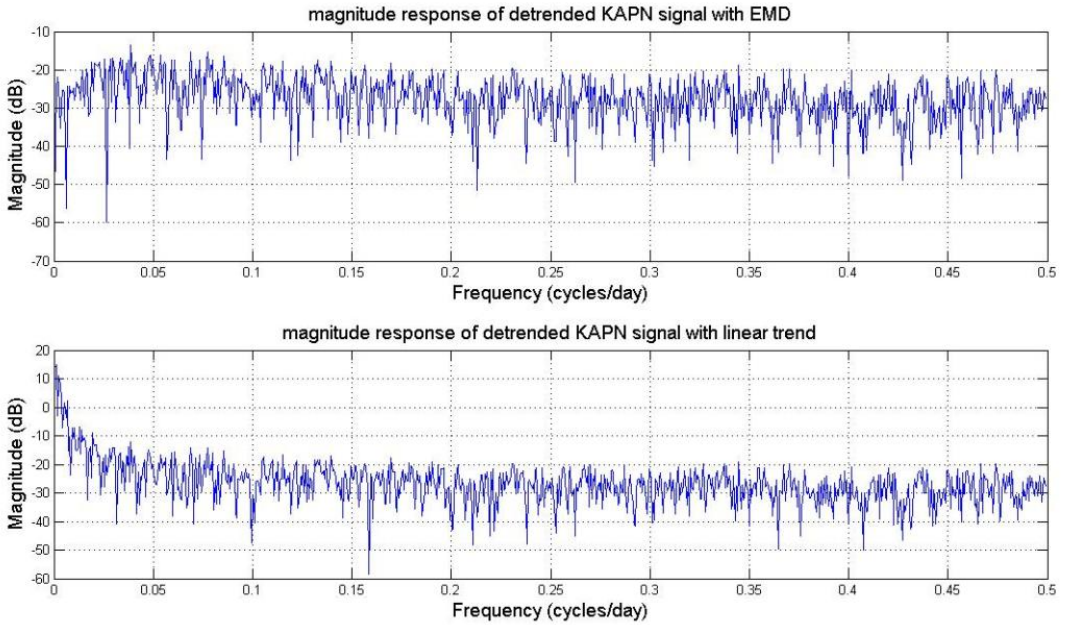

Figure 5.43 (Top) Magnitude response of detrended KAPN signal with EMD (Bottom) Magnitude response of detrended KAPN signal with linear fit (The DC component was removed from linear fit for better visualization).

The norm of the differences between the magnitude responses of detrended KAPN signal with EMD and detrended KAPN signal with linear fit is 18.342, which is a considerable amount.

5.4.6 Hilbert Spectrum

In Figure 5.44, the Hilbert spectrum for CAVD daily east offsets is given in the color map format. The scale for amplitudes is 20*log($H^2(w,t)$). Frequency variations of each IMF are visible. A more continuous frequency variation can be presented by selecting a lower frequency resolution. Such a presentation is given in Figure 5.45 with 50 frequency cells. In Figure 5.44, the number of the frequency cells is chosen as 286. The optimal frequency resolution in the Hilbert Spectrum can be computed as follows:

$$N = \frac{T}{n\Delta t} \ , \tag{5.3}$$

where N is the maximum number of frequency cells, T is the total data length, Δt is the sampling rate, and n is the minimum number of Δt needed to define the frequency accurately (Huang, 1998). For CAVD data, T=1430 days and Δt is 1 day. When we consider that the absolute minimum number of data points for a whole sine wave is five, N can be calculated as 286. Although there is no need to have a whole sine wave to define its frequency, many points are needed within any part of the wave to make the derivative stable. Thus, around five adjacent cell values have been averaged as in Figure 5.45.

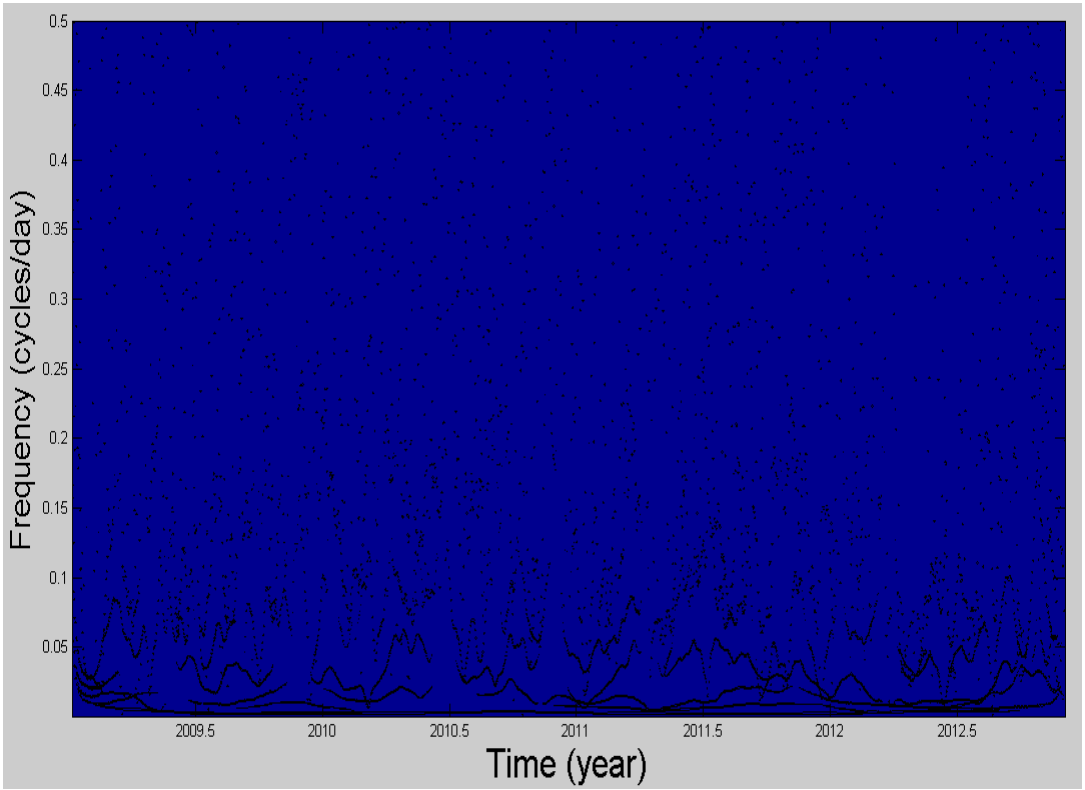

Figure 5.44 The Hilbert spectrum for CAVD daily east offsets with 286 frequency cells.

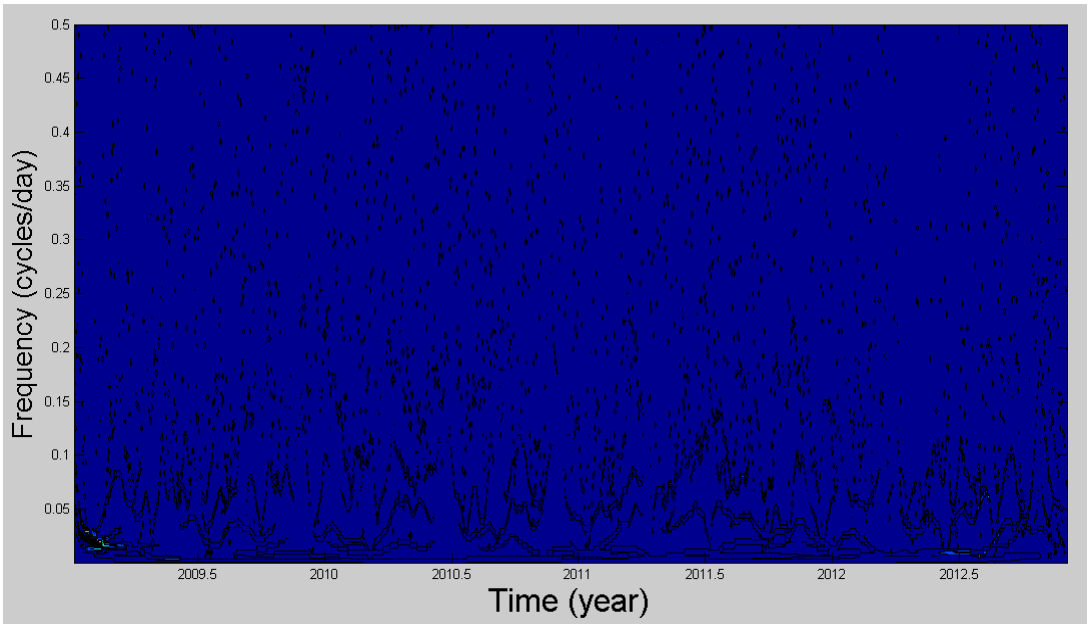

Figure 5.45 The Hilbert spectrum for CAVD daily east offsets with 50 frequency cells.

Individual instantaneous frequency and amplitude distributions for each IMF of CAVD data are shown in Figure 5.46.

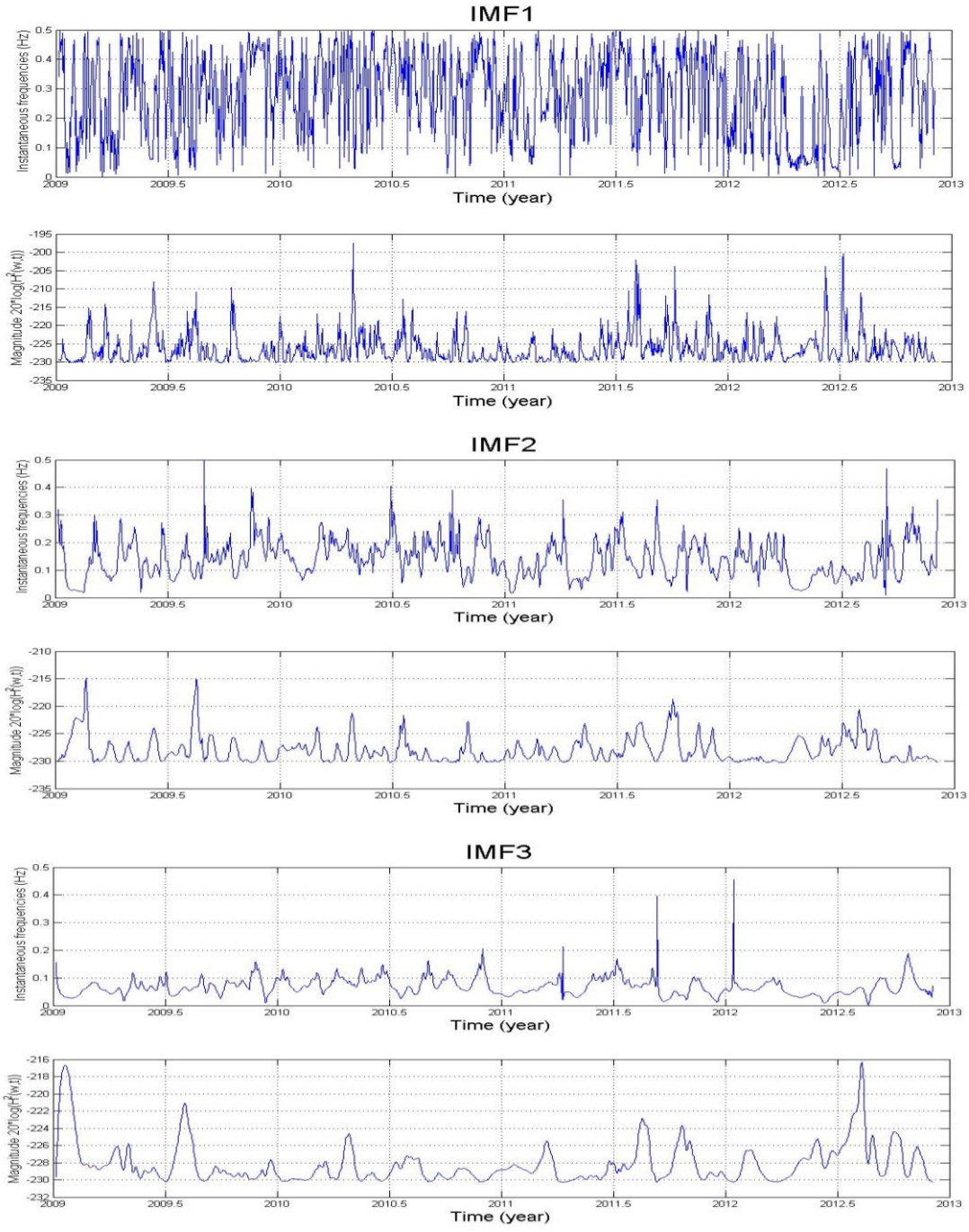

Figure 5.46 Individual instantaneous frequency and amplitude distributions for IMF components of CAVD data (IMF1-IMF2-IMF3).

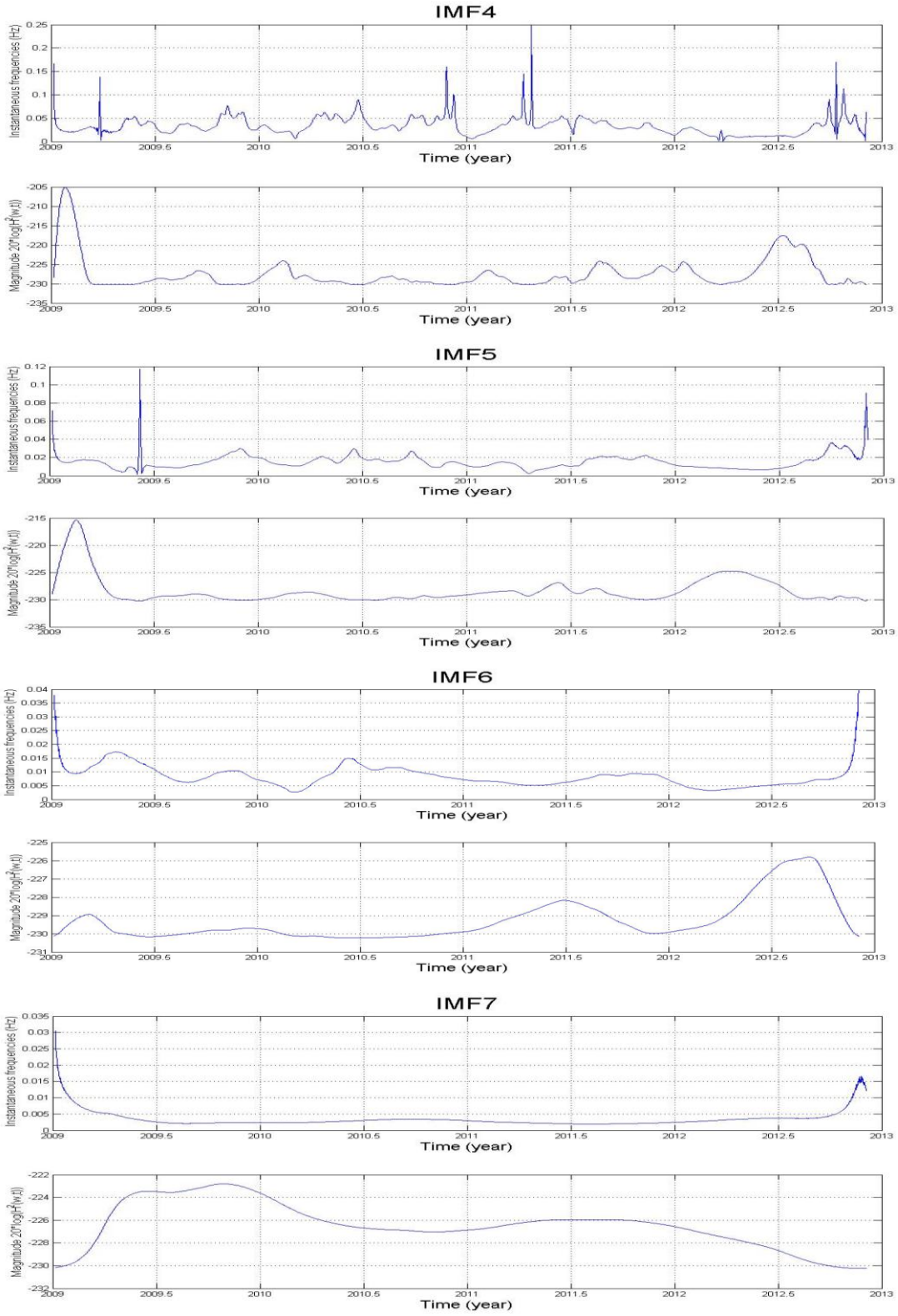

Figure 5.46 (cont'd) Individual instantaneous frequency and amplitude distributions for IMF components of CAVD data (IMF4-IMF5-IMF6-IMF7).

The effect of filling up the data gaps artificially was investigated through the evaluation of MALZ up component time series. Data gaps were eliminated by using linear interpolation. Striking parts of the time series are highlighted by black circles in Figure 5.47. Corresponding Hilbert Spectrum is given in Figure 5.48. The effect of linear interpolation is clear between year 2012 and 2013. Contradictory to the nature of the data, linear interpolation causes unvarying frequency behaviour. Its effect can be seen better in the higher frequency components of EMD, which are displayed in Figure 5.49.

In Figure 5.50, the sixth IMF component is plotted on the linearly detrended original data. The sixth IMF gives a sensation that it could be used to represent the general behaviour of the MALZ up component, thereby to predict future events. But it is seen that the beginning part of the IMF does not fit the data well. It is thought that the linear interpolation highlighted in Figure 5.47 causes this misfit. Thus, that part of the data has been cut, and then EMD process reemployed. Obtained sixth IMF component is displayed in Figure 5.51. This time, the sixth IMF component fits the data better. Utilization of sixth IMF component for the prediction of future events in MALZ time series will be studied in future.

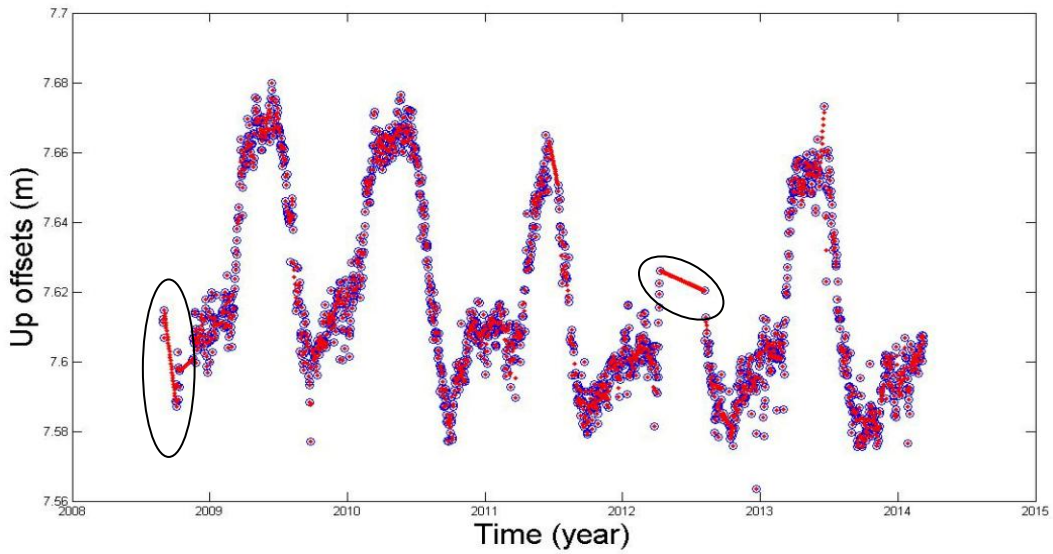

Figure 5.47 Linearly interpolated MALZ up offsets. The effects of linear interpolation are highlighted with black circles.

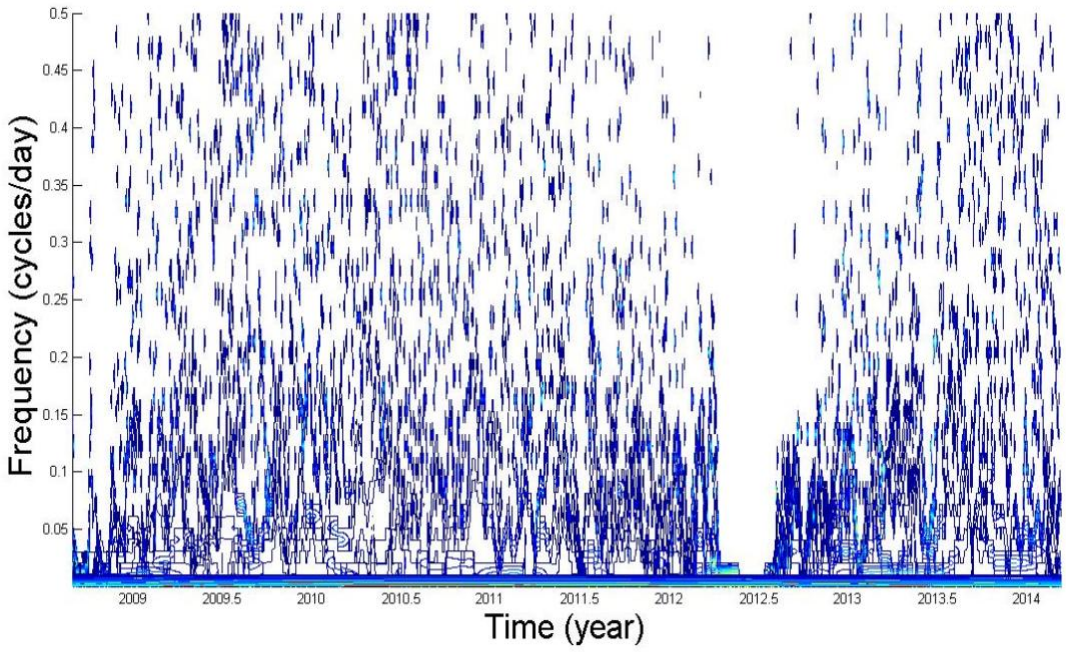

Figure 5.48 Hilbert Spectrum of linearly interpolated MALZ up offsets.

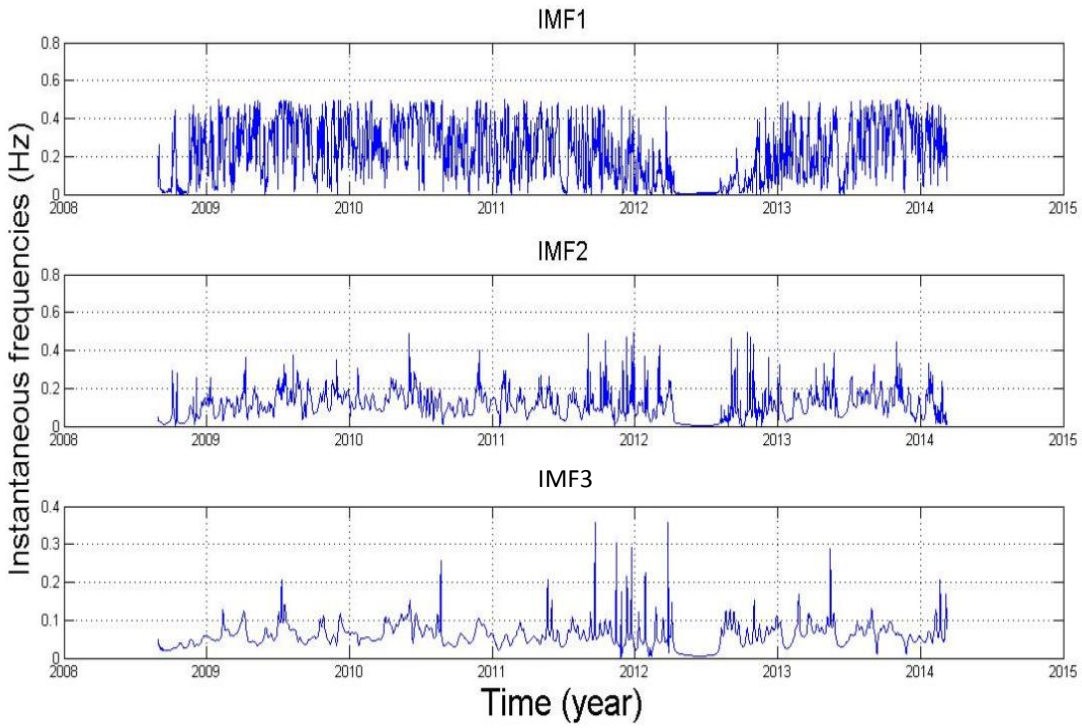

Figure 5.49 Instantaneous frequencies of IMF1, IMF2 and IMF3 components of MALZ up offsets.

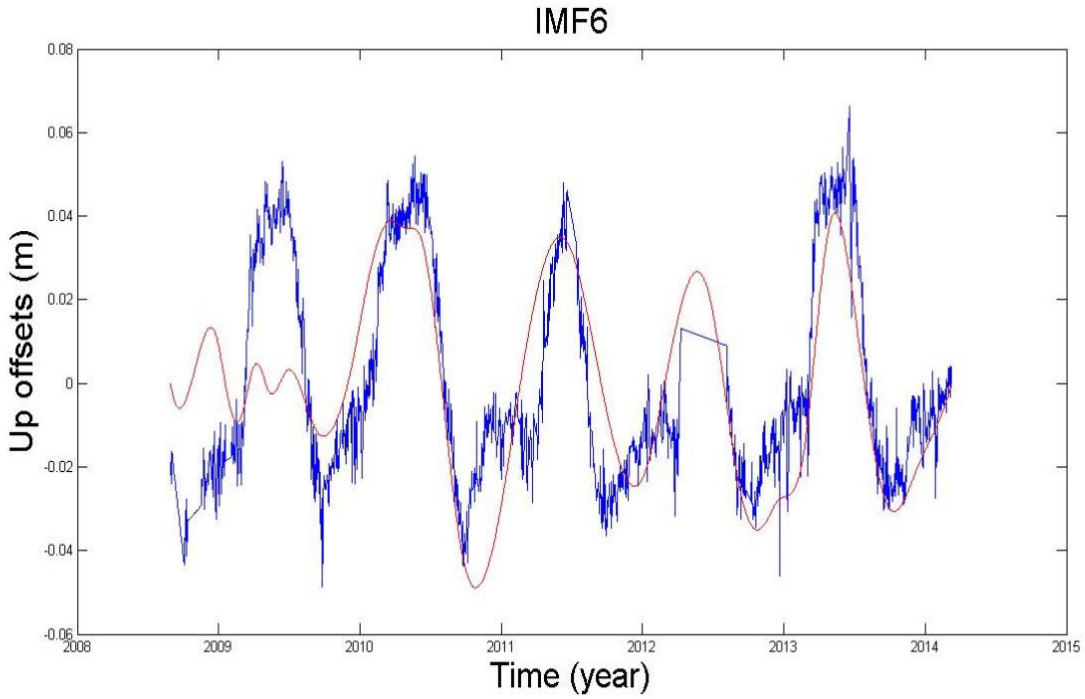

Figure 5.50 IMF6 component of MALZ up offsets before truncation.

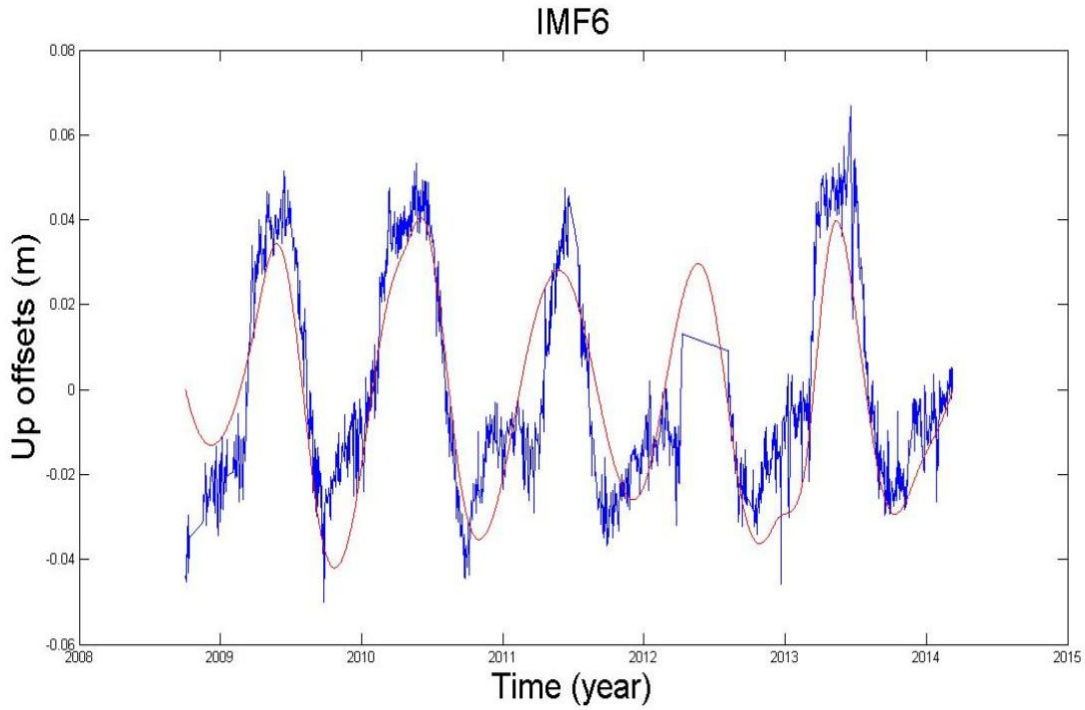

Figure 5.51 IMF6 component of MALZ up offsets after truncation.

5.4.7 Marginal Hilbert Spectrum

In Figure 5.52, the marginal Hilbert spectrum of CAVD data is given in blue solid line, while dotted red line shows the Fourier spectrum of the same data. Although they are presented together, frequency in Hilbert spectrum has a different meaning from the frequency in Fourier spectrum. It means that, in the whole time span of the data, there is a higher likelihood for such a wave to have appeared locally.

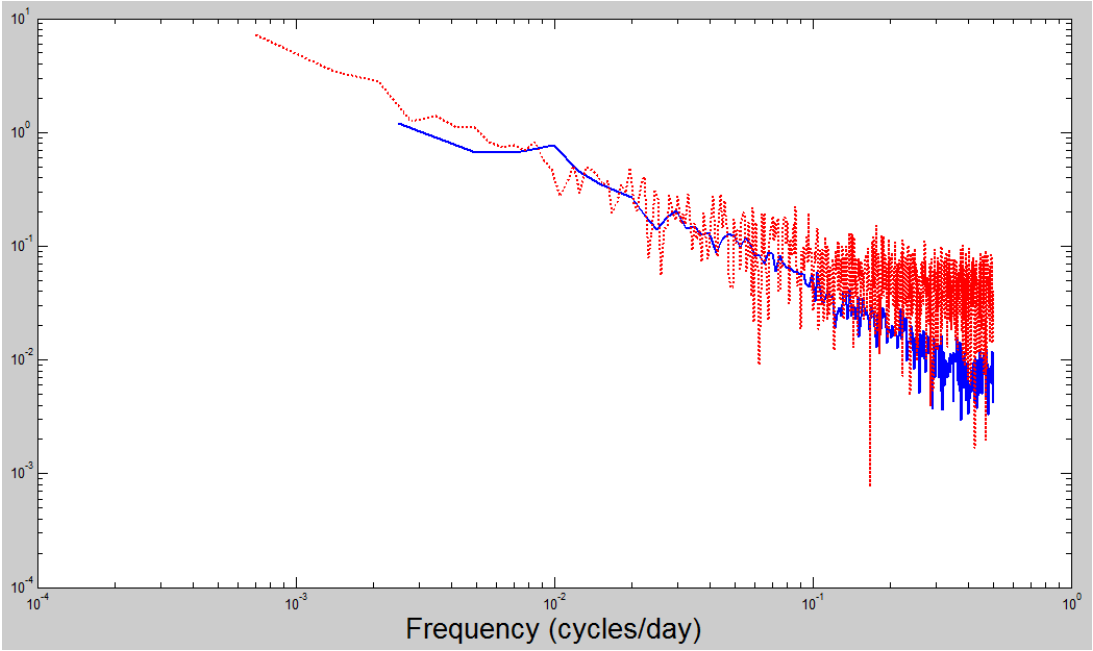

Figure 5.52 Marginal Hilbert spectrum (blue solid line) and Fourier Spectrum (red dotted line) of CAVD east data.

It can be seen in Figure 5.52 that the Fourier spectrum is dominated by the DC term.

CHAPTER 6

CONCLUSIONS

TNPGN-Active system is gaining increasing popularity in geodetic community, owing to its quick and accurate position information providing property. Inevitably, it will also pervade other practices in which accurate position information is needed. The accuracy of the position information which is sent to the users on the field, solely depends on the accuracy of the base stations' positions. Hence, it is important to analyse TNPGN-Active static data carefully.

GNSS observations collected at TNPGN-Active stations have been processed and coordinate time series of these stations have been produced in this thesis. Atmospheric pressure loading corrections (Tregoning and van Dam, 2005), and second and third order ionospheric effects (Bassiri and Hajj, 1993) were not included in the analysis, for it is not expected that having these effects included will make a considerable contribution to the position estimates. Majority of the TNPGN-Active stations are built onto the terraces or roofs of buildings. Thus, the spurious motion of the monument is a bigger error source than these effects. After the evaluation of the pros and cons (analysis of the whole data from the beginning), inclusion of these effects can be assessed in future.

The time series of TNPGN-Active stations reached 5-year temporal resolution and the velocity bias due to the annual signals became negligible. After the examination of the WRMS values of the daily time series, ADAN, AKSI, KAPN, RZE1, SLEE, INEB, MALZ, MUUS and SAMN stations were detected as the sites showing remarkable WRMS values. ADAN, AKSI, KAPN, RZE1, SLEE stations are functioning on the roofs of the buildings. Environmental effects, motion behaviour

of the buildings on which these stations are installed, station equipments etc. should be investigated cautiously. If the sources which cause the high RMS values cannot be found, the locations of these stations should be changed. Post-seismic effect was detected in VAAN_1PS station which should be considered in velocity estimation.

After the generation of daily time series, for the purpose of outlier detection, a visual inspection of the time series is being implemented in General Command of Mapping (GCM), who is the responsible organization of calculating the accurate positions of TNPGN-Active stations. A practical outlier detection algorithm enabling a quantiative evaluation of the outliers is proposed in this thesis, which can be used in GCM's future analyses. In addition to the extreme value analysis with the incorporation of Iteratively Rewighted Least Squares fit of the data, the uncertainty rates of the daily solutions were also considered.

After cleaning the outliers, daily solutions were combined and monthly combined time series were created. Using daily or monthly solutions for velocity estimation is a contentious issue. Both have advantages and disadvantages. Monthly combinations shorthens the processing time (though it is not a problem any more with high-capacity computers) in velocity estimation and allows to generate statistics for long-term repeatibility without mixing in the short-term scatter. By averaging, it also inhibits the contamination originating from unreliable daily solutions. Albeit, with the incorporation of a robust outlier detection algorithm, the effects of bothersome daily solutions could be decreased thoroughly.

One of the most attractive properties of Hilbert Huang Transform (HHT) is its ability to detect the smallest scales in the data. Thus, daily time series, in comparison with the monthly time series, probably would give more clues about the underlying physical processes, especially about the transient ones. Furthermore, in daily time series, the risk of loosing useful data at the ends can be taken, and some windowing operation can be implemented to alleviate the end effects. In monthly time series with ~30 times less data points, it is almost impossible at the moment to make windowing at the ends.

Another difficulty that should be faced with in monthly time series is the extrema locations. Although GNSS time series are discrete signals, they are originated from continuous processes. Quite possibly, the real extrema are somewhere between consecutive data points in monthly time series. This is not such a significant problem in over-sampled daily time series.

On the other hand, since it is rare that a data gap happens for more than one month in TNPGN-Active stations, monthly time series exclude the possible problems that can arise while forming the envelopes in HHT due to the lack of days or weeks of data. Lack of data leads to make interpolation to fill up the gaps in the signal which may betray the original behaviour of the signal. The effect of interpolation in Hilbert Spectrum is illustrated in Chapter 5. In conclusion, it is proposed in this thesis to use daily time series, together with a robust outlier detection algorithm and appropriate adaptive interpolation techniques in velocity estimation as well as in Empirical Mode Decomposition and Hilbert Spectral Analysis, instead of monthly time series.

In this thesis, TNPGN-Active time series were first preprocessed by the Empirical Mode Decomposition method (EMD), and Intrinsic Mode Functions (IMF) were produced. Later on, Hilbert Transform was applied to IMFs and energy-frequency-time distributions were obtained. HHT is an appropriate tool to decompose non-linear, non-stationary GNSS position time series. It is an adaptive and relatively new technique in which there are areas that require further improvements.

The cubic spline fitting has overshoot and undershoot problems. To alleviate these problems, the constrained cubic spline fitting was applied to TNPGN-Active time series. For some monthly time series, the constrained cubic spline fitting was superior to the natural cubic spline. But for most of the TNPGN-Active time series, the constrained cubic spline caused more severe overshoots and undershoots than the natural cubic spline did. It was understood that, the constrained cubic spline fitting is giving better results just in monotonically increasing or decreasing time series. However, it is worth to investigate the utilization of the constrained cubic

spline whenever possible especially in monthly combined time series. Due to the unsatisfactory results, the idea of using constrained cubic spline was dropped and the traditional natural cubic spline was used instead in the analysis of TNPGN-Active data.

A tedious situation occurs near the ends of the time series when forming the upper and lower envelopes with cubic spline function. The end swings can gradually influence inside of time series with the on-going sifting process and distort the results, especially in low-frequency components. For longer (oversampled) TNPGN-Active daily time-series, it is a good choice to throw away the first and the last segments which can be affected by distortions in each sifting. For the analysis of TNPGN-Active time series, the first and the last samples were considered as maxima and minima simultaneously. This forces all the IMFs to be zero at the ends. Seeing that the essential scales could be extracted from the data successfully, endeavor for extending the signal properly (for the known first derivatives at the first and last points would decrease the level of the swings) was kept for a future study.

Sifting in EMD process, eliminates the riding waves and makes the wave profiles more symmetric which are essential conditions to obtain meaningful instantaneous frequency. With this aspect, it would be desirable to sift as many times as needed. However, it also smoothes the uneven amplitudes. This causes a decline in the range of variations in the IMF, and removes physically meaningful amplitude fluctuations. Hence, the number of the siftings should be limited to preserve the natural amplitude variations.

To obtain physically meaningful IMF components without constant amplitude frequency modulations, determination of a stopping criterion for sifting process is needed. In this thesis, four different approaches were considered for stopping criterion. First, the size of the standard deviation, *SD*, computed from the two consecutive sifting results was limited. Second, the sifting was stopped when the numbers of zero crossings and extrema are equal or differ at most by one for *S*

consecutive times. Third, the sifting number was fixed to 10. Fourth, a resolution factor was defined by the ratio between the energy of the signal at the beginning of the sifting, and the energy of the mean of the envelopes. Pros and cons of these stopping criteria were explained in Section 4.5 in detail.

After the inspection of various stopping criteria in this thesis, it was conceived that the proposed stopping criteria in the literature are devoid of strong theoretical bases. Rather, they were created based on the researchers' experience during their research. More or less they disregard the rules of IMF construction. They secure the symmetry of upper and lower envelopes only in an approximate sense and they do not always ensure the equality of the zero crossings and extrema numbers. Maybe it is impossible to define a perfect (in every sense) stopping criterion. Nonetheless, specific stopping criterion can be designed according to the needs and the characteristics of the phenomena. To keep the EMD temporally local, the sifting number should be fixed to a low number. Thus, instead of using standard deviations approach, the sifting number was fixed to 10, through this study. Whenever the EMD results give rise to suspicion, resolution factor was used which also gives satisfactory results.

Credibility of the decomposition was ensured via completeness test. The maximum amplitude of the differences between the original data and the reconstructed data is 1.33×10^{-15}, the roundoff error from the precision of the computer.

A zero reference is not needed to apply EMD to an arbitrary data. The sifting process generates zero references for IMFs. The Hilbert spectrum has an advantage that it is possible not to include the obtained residue (last IMF component) in its calculation. On the other hand, situation is not so simple for the ubiquitous Fourier Transform, if it is applied to the data without zero reference. To avoid the large DC term in data with non-zero mean, the mean values should be removed first. It is common to remove the best straight-line fit linear trend from the data before applying Fourier Transform. Evidently, it is controversial how well a straight line can represent the real trend of a non-stationary process. Hence, it is proposed to

detrend the data by using EMD prior to Fourier analysis. How to detrend the data, and the effect of detrending in Fast Fourier Transform were explained in Chapter 5.

Similar to detrending the data, it might also be possible to denoise the data by using EMD with an elaborate evaluation of the high-frequency IMF components. Denoising the data with the help of EMD will be studied in a future work.

A major drawback of the EMD is mode mixing. It can cause an IMF to be physically meaningless. To solve this problem, a noise-assisted data analysis method, the Ensemble EMD (EEMD) will be investigated. It defines the true IMF components as the mean of an ensemble of trials, each consisting of the signal plus a white noise of finite amplitude. The effects of the added white noise series cancel each other in the ensemle average, the only persistent part is the signal. However, the noise in GNSS position time series cannot be assumed purely white. Therefore, the effectiveness of EEMD on GNSS time series is arguable.

After EMD process, instantaneous frequencies were calculated for each IMF and corresponding Hilbert Spectra were formed. With the concept of instantaneous frequency, it became possible to define the interwave and intrawave frequency modulations of the data.

It should be noted that, physical soundness of the individual IMFs are not guaranteed with this method. Interpreting the physical meaning of the IMFs for GNSS time series needs experience and careful assessment. In any case, first consideration must be the clear seperation of the scales. With this study, HHT was applied on TNPGN-Active time series for the first time. Great efforts are awaiting for physical interpretation of the results obtained in this study.

REFERENCES

Adams, R.P. and Mackay, D.J.C., (2007), Bayesian Online Changepoint Detection, Technical Report, University of Cambridge, Cambridge UK.

Aggarwal, C.C., (2013), Outlier Analysis, Springer Publishing Company, Incorporated.

Aktuğ, B., Nocquet, J.M., Cingöz, A., Parsons, B., Ercan, Y., England, P.C., Lenk, O., Gürdal, M.A., Kılıçoğlu, A., Akdeniz, H., Tekgül, A., (2009), Deformation of Western Turkey from a Combination of Permanent and Campaign GPS Data: Limits to Block-like Behaviour, Journal of Geophysical Research, 114, B10404.

Akyay, T. And Karslıoğlu, M.O., (2009), Wavelet-based Outlier Detection and Denoising of Airborne Laser Scanning Data, 4th International Conference on Recent Advances in Space Technologies-RAST2009, Istanbul, 11-13 June.

Altamimi, Z., Collilieux X., Metivier L., (2011), ITRF2008: An Improved Solution of the International Terrestrial Reference Frame, Journal of Geodesy, Volume 85, Issue 8, 457-473.

Ash, M.E., (1972), Determination of Earth Satellite Orbits, Tech. Note 1972-5, Lincoln Laboratory, MIT.

Bar-Sever, Y., (1996), A New Model For GPS Yaw Attitude, J. Geodesy, 70, 714-723.

Basseville, M. and Nikiforov, I.V., (1993), Detection of Abrupt Changes: Theory and Application, Upper Saddle River, NJ, USA: Printice-Hall, Inc.

Bassiri, S. And Hajj, G.A., (1993), Higher-order Ionospheric Effects on the Global Positioning System Observables and Means of Modeling Them, Man. Geodaetica, 18, 280-289.

Bedrosian, E., (1963), A Product Theorem for Hilbert Transform, Proc. IEEE 51, 868-869.

Benzi R., Stuera, A., Vulpiani, A., (1981), The Mechanism of Stochastic Resonance, J. Phys. A, 14(5), 453-457.

Bickel, S., Brückner, M., Scheffer, T., (2007), Discriminative Learning for Differing Training and Test Distributions, Proceedings of the 24th International Conference on Machine Learning, 81-88.

Birx, D.L. and Pipenberg, S.J., (1992), Chaotic Oscillators and Complex Mapping Feed Forward Networks for Signal Detection in Noisy Environments, International Joint Conference on Neural Networks, Vol. 2, pp. 881-888.

Blewitt, G. And Lavallee, D., (2002), Effect of Annual Signals on Geodetic Velocity, Journal of Geophysical Research, 107(B7), 10.1029/2001JB000570.

Blewitt, G., (1989), Carrier Phase Ambiguity Resolution for the Global Positioning System Applied to Geodetic Baselines up to 2000km, Journal of Geophysical Research 94(B8), 10187-10283.

Blewitt, G., (1990), An Automatic Editing Algorithm for GPS Data, Geophysical Research Letters 17(3), 199-202.

Boashash, B., (1992), Estimating and Interpreting the Instantaneous Frequency of a signal. I. Fundamentals, Proc. IEEE 80, 520-538.

Bock, Y., Gourevitch, S.A., Counselman, C.C.III, King, R.W., Abbot, R.I., (1986), Interferometric Analysis of GPS Phase Observation, Man. Geodaetica, 11, 282-288.

Boehm, J., Niell, A.E., Tregoning, P., Schuh, H., (2006), Global Mapping Function (GMF): A New Empirical Mapping Function Based on Numerical Weather Model Data, Geophys. Res. Lett., 33, L07304, doi: 10.1029/2005GL025545.

Boehm, J., Heinkelmann, R., Schuh, H., (2007), Short Note: A Global Model of Pressure and Temperature for Geodetic Applications, J. Geod., doi:10.1007/s00190-007-0135-3.

Brodsky, B., and Darkhovsky, B., (1993), Nonparametric Methods in Changepoint Problems, Dordrecht, Netherlands: Kluwer Academic Publishers.

Brown, R.G., and Hwang, P.Y.C., (1997), Introduction to Random Signals and Applied Kalman Filtering, 3rd ed., New York: Wiley.

Chen, C.H., Li, C.P., Teng, T.L., (2002), Surface Wave Dispersion Measurements Using Hilbert-Huang Transform, Terrestrial Atmospheric and Oceanic Sciences, Vol. 13, 71-84.

Chiew, F.H.S., Peel, M.C., Amirthanathan, G.E., Pegram, G.G.S., (2005), Identification of Oscillations in Historical Global Streamflow Data Using Empirical Mode Decomposition, Seventh IAHS Scientific Assembly at Foz do Iguscu, Brazil, pp. 53-62.

Csörgö, M. and Horvath, L., (1988), 20 Nonparametric Methods for Changepoint Problems, Handbook of Statistics, Vol.7, 403-425, Amsterdam, Netherlands:Elsevier.

Datig, M. and Schlurmann, T., (2004), Performance and Limitations of the Hilbert-Huang Transformation (HHT) with an Application to Irregular Water Waves, Ocean Engineering 31, 1783-1834.

Desobry, F., Davy, M., Doncarli, C., (2005), An On-line Kernel Change Detection Algorithm, IEEE Transactions on Signal Processing, 53(8), 2961-2974.

Dong, D.N., Herring, T.A., King, R.W., (1998), Estimating Regional Deformation from a Combination of Space and Terrestrial Geodetic Data, J. Geodesy, 72, 200-214.

Durmaz, M., (2007), One-dimensional Real-time Signal Denoising Using Wavelet-based Kalman Filtering, (M.Sc.), Middle East Technical University, Geodetic and Geographic Information Technologies.

Dutta, R., Das, A., Stocks, N.G., Morgan, D., (2006), Stochastic Resonance-based Electronic Nose: A Novel Way to Classify Bacteria, Sensors and Actuators B: Chemical, 115(1), pp. 17-27.

Feigl, K.L., and 14 co-authors (1993), Space Geodetic Measurement of Crustal Deformation in Central and Southern California, 1984-1992, J. Geophys. Res., 98, 21677-21712.

Flandrin, P., Gonçalves, P., Rilling, G., (2004), Detrending and Denoising with Empirical Mode Decompositions, Proc. EUSIPCO, Vienna, Austria, pp. 1581-1584.

Gabor, D., (1946), Theory of Communication, Proc. IEE 93, 429-457.

Garnett, R., Osborne, M., Roberts, S., (2009), Sequential Bayesian Prediction in the Presence of Changepoints, Proceedings of the 26th International Conference on Machine Learning, 345-352, Montreal, QC.

Gill, D., (1970), Application of a Statistical Zonation Method to Reservoir Evaluation and Digitized Log Analysis, American Association of Petroleum Geologists Bulletin, 54:719-729.

Gretton, A., Smola, A., Huang, J., Schmittfull, M., Borgwardt, K., Schölkopf, B., (2009), Covariate Shift by Kernel Mean Matching, Dataset Shift in Machine Learning, 131-160, Cambridge, MA, USA: MIT Press (Chapter 8).

Grewal, M.S., and Andrews, A.P., (2000), Kalman Filtering: Theory and Practice, 2nd ed., New York: Wiley.

Groves, P.D., (2008), Principles of GNSS, Inertial and Multisensor Integrated Navigation Systems, Artech House Inc, Boston, London.

Guralnik, V. And Srivastava, J., (1999), Event Detection from Time-series Data, Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 33-42.

Gustafsson, F., (1996), The Marginalized Likelihood Ratio Test for Detecting Abrupt Changes, IEEE Transactions on Automatic Control, 41(1), 66-78.

Hawkins, D., (1980), Identification of Outliers, Chapman and Hall.

Herring, T.A., Davis, J.L., Shapiro, I.I., (1990), Geodesy by Radio Interferometry: the application of Kalman Filtering to the Analysis of Very Long Baseline Interferometry Data, J. Geophys. Res., 95, 12561-12581.

Herring, T.A., (2005), GLOBK Global Kalman Filter VLBI and GPS Analysis Program, Version 10.1, Mass. Inst. Of Technol., Cambridge.

Herring, T.A., King, R.W., McClusky, S.C., (2008), Introduction to GAMIT/GLOBK Release 10.3, Department of Earth, Atmospheric, and Planetary Sciences, MIT.

Herring, T.A., (2012), GAMIT/GLOBK Workshop Notes, Montserrat.

Huang, N.E., Chern, C.C., Huang, K., Salvino, L.W., Long, S.R., Fan, K.L., (2001), A New Spectral Representation of Earthquake Data: Hilbert Spectral Analysis of Station TCU129, Chi-Chi Taiwan, 21 September 1999, Bulletin of the Seismological Society of America, 91:1310-1338.

Huag, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., Liu, H.H., (1998), The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis, Proceedings of the Royal Society of London A 454 (1971): 903–995.

Huang, N.E., Shen, Z., Long, S.R., (1999), A New View of Nonlinear Water Waves: the Hilbert Spectrum, Annual Reviews of Fluid Mechanics 31, 417-457.

Huang N.E., Wu, M.C., Long, S.R., Shen, S.S.P., Qu, W., Gloersen, P., Fan, K.L., (2003), A Confidence Limit for Empirical Mode Decomposition and Hilbert Spectral Analysis, Proc. Roy. Soc. London 459, 2317-2345.

Hugentobler, U., Dach, R., Meindl, M., Fridez, P., (2005), Bernese GPS Software, Version 5.0, Astronomical Institute, University of Berne.

Ide, T. And Tsuda, K., (2007), Changepoint Detection Using Krylov Subspace Learning, Proceedings of the SIAM International Conference on Data Mining, 515-520.

IERS Conventions 2003, (2004), D.D. McCarthy and G. Petit (editors), IERS Technical Note 32, Frankfurt am Main: Verlag des Bundesamts fuer Kartographie und Geodaesie.

Jolliffe, I., (2002), Principal Component Analysis, Springer.

Kalman, R.E., (1960), A New Approach to Linear Filtering and Prediction Problems, ASME Transactions, Series D: Journal of Basic Engineering, Vol. 82, pp. 35-45.

Kawahara, Y., Yairi, T., Machida, K., (2007), Changepoint Detection in Time-series Data Based on Subspace Identification, Proceedings of the 7th IEEE International Conference on Data Mining, 559-564.

Kawahara, Y. And Sugiyama, M., (2012), Sequential Changepoint Detection Based on Direct Density Ratio Estimation, Statistical Analysis and Data Mining, 5(2), 114-127.

King, R.W. and Bock, Y., (2005), Documentation for the GAMIT GPS Processing Software, Release 10.2, Mass. Inst. Of Technol., Cambridge.

Koch, K.-R., (1999), Parameter Estimation and Hypothesis Testing in Linear Models, 2nd ed., Springer-Verlag, Berlin, Heidelberg.

Kokes, J., Ba, N.N., (2011), Using Constrained Cubic Spline Instead of Natural Cubic Spline to Eliminate Overshoot and Undershoot in HHT, Annals of Faculty Engineering Hunedoara-International Journal of Engineering, Tome IX (Year 2011), Fascicule 3, (ISSN 1584-2673).

Kruger, C.J.C., (2002), Constrained Cubic Spline Interpolation for Chemical Engineering Applications, http://www.korf.co.uk/spline.pdf, (last accessed on 20.06.2014).

Lambeck, K., (1988), The Slow Deformations of the Earth, Goephysical Geodesy, xii+718 pp. Oxford: Clerandon Press.

Li, C., (2005), Study of Weak Signal Detection Based on Second FFT and Chaotic Oscillator, Nature and Science, 3(2), 59-64.

Mallat, S. And Hwang, W.L., (1992), Singularity Detection and Processing with Wavelets, IEEE Transactions on Information Theory, 38:617-643.

Mao, A., Harrison, C.G.A., Dixon, T.H., (1999), Noise in GPS Coordinate Time-series, J. Geophys. Res., 104, 2797-2896.

Markou, M. and Singh, S., (2003), Novelty Detection: A Review, part 1: Statistical Approaches, Signal Process, 83(12), 2481-2497.

MIT Analysis Strategy, ftp://igscb.jpl.nasa.gov/pub/center/analysis/mit.acn, (last accessed on 23.06.2014).

Moskvina, V. And Zhigljavsky, A., (2003), An Algorithm Based on Singular Spectrum Analysis for Changepoint Detection, Communication in Statistics: Simulation and Computation, 32(2), 319-352.

Niell, A., (1996), Global Mapping Functions for the Atmosphere Delay at Radio Wavelengths, Journal of Geophysical Research, 101, pp. 3227-3246.

Paquet, U., (2007), Empirical Bayesian Changepoint detection, Graphical Models, 1995, 1-20.

Qu, L. And Jing, L., (1999), A Difference Resonator for Detecting Weak Signals, Measurement, Vol. 26, pp .69-77.

Rato, R.T., Ortigueira, M.D., Batista, A.G., (2008), On the HHT, its Problems, and Some Solutions, Mechanical Systems and Signal Processing 22, 1374-1394.

Rilling, G., Flandrin, P., Gonçalves, P., (2003), On Empirical Mode Decomposition and its Algorithms, in: IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP-03, Grado (I).

Rothacher, M., Beutler, G., Gurtner, W., Schildknecht, T., Wild, U., (1990), BERNESE GPS Software Version 3.2. Printing Office, University of Berne, Switzerland.

Rousseeuw, P. And Leroy, A., (2003), Robust Regression and Outlier Detection, Wiley.

Saastamoinen, J., (1972), Atmospheric Correction for the Troposphere and Stratosphere in Radio Ranging of Satellites, in the use of artificial satellites for Geodesy, Geophys. Monogr. Ser. 15, AGU, Washington, D.C., pp.247-251.

Salisbury, J.I. and Wimbush, M., (2002), Using Modern Time-series Analysis Techniques to Predict ENSO Events from the SOI Time-series, Nonlinear Processes in Geophysics, 9:341-345.

Scales, J.A., Gersztenkorn, A., Tretiel, S., (1988), Fast lp Solution of large, sparse, linear systems: Application to Seismic Travel Time Tomography, J. Computational Physics, 75(2):314-333.

Schaffrin, B. And Bock, Y., (1988), A Unified Scheme for Processing GPS Phase Observations, Bull. Geodesique, 62, 142-160.

Schölkpf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C., (2001), Estimating the Support of a High-dimensional Distribution, Neural Computation, 13(7), 1443-1471.

Spinler, J.C., Bennett, R.A., Anderson, M.L., McGill, S.F., Hreinsdottir, S., McCallister, A., (2010), Present-day Strain Accumulation and Slip Rates Associated with Southern San Andreas and Eastern California Shear Zone Faults, Journal of Geophysical Research, Vol. 115, B11407.

Sugiyama, M., Suzuki, T., Kanamori, T., (2012), Density Ratio Estimation in Machine Learning, Cambridge, UK: Cambridge University Press.

Sugiyama, M., Suzuki, T., Nakajima, S., Kashima, H., Von Buenau, P., Kawanabe, M., (2008), Direct Importance Estimation for Covariate Shift Adaptation, Annals of the Institute Statistical Mathematics, 60(4), 699-746.

Sweldens, W., (1996), The Lifting Scheme: A custom-design construction of biorthogonal wavelets, Appl. Comput. Harmon. Anal., 3(2):186-200.

Takeuchi, J. and Yamanishi, K., (2006), A Unifying Framework for Detecting Outliers and Changepoints from Non-stationary Time-series Data, IEEE Transactions on Knowledge and Data Engineering, 18(4), 482-492.

TNPGN-Active system, http://www.hgk.msb.gov.tr/english/tnpgn-active.php, (last accessed on 24.06.2014).

Tregoning, P. and Van Dam, T., (2005), Atmospheric Pressure Loading Corrections Applied to GPS Data at the Observation Level, Geophys. Res. Letters, Vol. 32, L22310.

Wang, G.Y., Chen, D.J., Lin, J.Y., Chen, X., (1998), The Statistical Characteristics of Weak Signal Detection Based on Duffing Oscillator, Acta Electronica Sinica 26(10), 38–44.

Webb, F.H. and Zumberge, J.F., (1993), An Introduction to GIPSY-OASIS II. JPL Publication D-11088. Pasadena, CA, Jet Propulsion Laboratory.

Wu, Z. and Huang, N.E., (2004), A Study of the Characteristics of White Noise Using the Empirical Mode Decomposition Method, Proc. Roy. Soc. London, 460A: 1597-1611.

Wu, Z. and Huang, N.E., (2005), Statistical Significant Test of Intrinsic Mode Functions, in Hilbert-Huang Transform: Introduction and Applications, pp. 125-148, Eds. N. E. Huang and S. S. P. Shen, World Scientific, Singapore, 311.

Wu, Z. and Huang, N.E., (2009), Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method, Adv. Adapt. Data Anal., 1: 1-41.

Wu, Z. and Huang, N.E., (2010), On the Filtering Properties of the Empirical Mode Decomposition, Adv. Adapt. Data Anal., Vol. 2, No. 4, 397-414.

Yamanishi, K., Takeuchi, J., Williams, G., Milne, P., (2000), On-line Unsupervised Outlier Detection Using Finite Mixtures with Discounting Learning Algorithms, Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 320-324.

Zhang, J., Bock, Y., Johnson, H., Fang, P., Williams, S., Genrich, J., (1997), Wdowinski S., Behr J., Southern California Permanent GPS Geodetic Array: Error Analysis of Daily Position Estimates and Site Velocities, J. Geophys. Res., 102, 18,035-18,055.