Analysis of geodetic and model simulated data to describe nonstationary moisture fluctuations over southern Africa

by

Joel Ondego Botai

Submitted in partial fulfilment of the requirements for the degree
DOCTOR OF PHILOSOPHY

in the Faculty of Natural & Agricultural Sciences
University of Pretoria
Pretoria
South Africa

May, 2011

Analysis of geodetic and model simulated data to describe nonstationary moisture fluctuations over southern Africa

Author: Botai Ondego Joel
Supervisors: Prof. Ludwig Combrinck[1, 2] and
Prof. Sivakumar Venkataraman[1, 3]
Department: [1]Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria 0002, South Africa
[2]Hartebeesthoek Radio Astronomy Observatory (HartRAO), P.O. Box 443, Krugersdorp 1740, South Africa
[3]National Laser Centre (NLC), Council for Scientific and Industrial Research (CSIR), P. O. Box 395, Pretoria 0001, South Africa; Dept. of Physics, University of Kwa-Zulu Natal, Durban, South Africa.
Degree: Doctor of Philosophy

Abstract

Recent advances in space geodetic techniques such as Very Long Baseline Interferometry, Global Navigation Satellite Services, Satellite Laser Ranging and advanced numerical weather-prediction model simulations, provide huge tropospheric data sets with improved spatial-temporal resolution. These data sets exhibit unique fluctuations that have a spatial-temporal structure which are thought to mimic the complex behaviour of the atmosphere. As a result, the analysis of nonstationary structure in the tropospheric parameters derived from geodetic and numerical model simulations could be used to probe the extent of universality in the dynamics of the atmosphere, with applications in space geodesy. In order to identify the physical causes of variability of tropospheric parameters, parametric and nonparametric data analyses strategies which are investigated and reported in this thesis, are used to inform on the geophysical signals embedded in the data structure. In the first task of this research work, it is shown that the fluctuations of atmospheric water vapour over southern Africa are non-linear and nonstationary. Secondly, the tropospheric data sets are transformed to stationarity and the stochastic behaviour of water vapour fluctuations are assessed by use of an automatic algorithm that estimates the model parameters. By using a data adaptive modelling algorithm, an autoregressive-moving-average model was found to sufficiently characterise the derived stationary water vapour fluctuations. Furthermore, the non-linear and nonstationary properties of tropospheric delay due to water vapour were investigated by use of robust and tractable non-linear approaches such as detrended fluctuation analysis, independent component analysis, wavelet transform and empirical mode decomposition. The use of non-linear approaches to data analysis is objective and tractable because *they allow data to speak for themselves* during analysis and also because of the non-linear components embedded in the atmosphere system. In the thesis, we establish that the non-linear and nonstationary properties in the tropospheric data sets (i.e., tropospheric delay due to water vapour and delay gradients) could be triggered from strongly non-linear stochastic processes that have a local signature (e.g. local immediate topography, weather and associated systems) and/or exogenous. In addition, we explore and report on the presence of scaling properties (and therefore memory) in tropospheric parameters. This self-similar behaviour exhibit spatial-temporal dependence and could be associated with geophysical processes that drive atmosphere dynamics. Satellite Laser Ranging data are very sensitive to atmospheric conditions, which causes a delay of the laser pulse, hence an apparent range increase. A test for non-linearity is applied within specialised software for these data; it is found that the range residuals (i.e., the observed minus computed residuals) are improved when possible non-linearity of the locally measured meteorological parameters as applied to a range delay model are considered.

Preface

Geodetic time series analysis is a necessary procedure of extracting statistical properties and other characteristics of the data and is therefore an important process in modern space geodesy. In general, the analysis involves pre-processing of raw observations from various geodetic techniques, enhancing signals in the raw data, actual analysis (e.g., detection of nonlinearities and nonstationarities, statistical characterisation of the series) and prediction. While different methods are often applied to analyse the geodetic time series, estimating the deterministic (e.g., periodic variations and trend) and stochastic (mostly aperiodic variations) components as well as extracting specific oscillatory modes (which could be linked to geophysical signals) have not received much attention. In this current research work, the stochastic and multiscale properties in tropospheric parameters (hereafter Water Vapour (WV), tropospheric delay and delay gradients) derived from geodetic and numerical weather prediction models are assessed and modelled. The results indicate that WV/tropospheric delay due to WV exhibit self-similar behaviour and that their fluctuations are non-linear and nonstationary.

The layout of this thesis is intended to provide a logical flow of this research endeavour. After the general introduction in Chapter 1, the literature review (Chapter 2) provides an overview of space geodetic techniques, principle operation of Global Navigation Satellite System (GNSS) and Very Long Baseline Interferometry (VLBI) techniques, their applications (e.g., Earth's crustal deformation, plate tectonics, and maintenance of Terrestrial Reference Frames (TRF) as well as atmospheric remote sensing). Current measurements and analysis strategies of tropospheric parameters with application in geodetic analyses are also reviewed in this chapter. In Chapter 3, the sources of data that are studied in this thesis are explained. The spatial-temporal resolution of the geodetic (VLBI and GNSS), Numerical Weather Prediction (NWP) model simulations (e.g. NCEP/NCAR), radiosonde (e.g. the Southern Hemisphere ADditional OZonesondes (SHADOZ) network) and the HALOgen Occultation Experiment (HALOE) satellite data sets are described. The methods used to pre-process these data records are also described briefly.

Chapter 4 examines the stationarity in geodetic WV and adaptively fits a time series ARMA model that describes the stochastic pattern, to the geodetic WV transformed from nonstationary to stationary. Chapter 5 deals with the analysis of WV fluctuations. The SHADOZ radiosonde network is also used to infer the multiscale

structure of WV in low- and mid-tropical Africa. Furthermore, a model for the vertical profile of WV in the southern hemisphere based on the HALOE satellite and the SHADOZ network data is developed. In Chapter 6, firstly the scaling behaviour (testing the underlying memory processes) of tropospheric WV is assessed using wavelets. Secondly, a noise-assisted data analysis methodology is applied to the geodetic tropospheric zenith delay and surface temperature to determine the dominant modes of oscillation in data. Further, WV and surface temperature have been shown to be temporally correlated because the instantaneous phase differences among the associated modes of the Intrinsic Mode Functions (IMFs) derived from the Ensemble Empirical Mode Decomposition (EEMD) of WV and surface temperature have a high degree of synchronisation. Additionally, the benefit of introducing non-linearity and nonstationarity in atmospheric correction to the Satellite Laser Ranging (SLR) range is investigated by introducing a nonlinear function to model the azimuth dependent atmospheric range correction. In Chapter 7, a summary of the findings are presented and recommendations and future research proposed.

List of publications

The following contributions have been published and/or submitted in/to various peer review journals as part of this work or related to it.

1. **Botai O. J.,** W. L. Combrinck, V. Sivakumar and C. J. de W. Rautenbach, (Submitted). Probing nonlinearity in geodetic data by hypothesis testing, Journal of Geodesy.
2. **Botai O. J.,** W. L. Combrinck and V. Sivakumar, (2011). Inferences of α-stable distribution in long-range dependent geodetic data, South African Journal of Geology (Accepted).
3. **Botai O. J.,** W. L. Combrinck and V. Sivakumar, H. Schuh and J. Boehm, (2010). Extracting independent local oscillatory geophysical signals by geodetic tropospheric delay, *In*: IVS 2010 General meeting proceedings, D. Behrend and K. D. Baver (eds). ISBN. NASA/CP-2010-215864.
4. **Botai O. J.,** V. Sivakumar , C. J. de W. Rautenbach, and W. L. Combrinck, (2010). Multiscale organisation of water vapour in the low- and mid- tropical Africa, Advances in Geosciences, 16: 241-251.
5. **Botai O. J.,** W. L. Combrinck and V. Sivakumar, (2009). Assessing the degree of synchronisation between geophysical records using the method of instantaneous phase differences, *In* 11th SAGA biennial technical meeting and exhibition. D. Vogt and S. Fourie (eds). ISBN. 978-0-620-44602-0: 588-593.
6. Sivakumar V., **O. J. Botai**, D. Moema, A. Sharma, C. Bollig and C.J. de W. Rautenbach, (2009). CSIR-NLC mobile LIDAR for atmosphere remote sensing, IEEE, Intern. geosc. Remote sens. symposium, Cape Town. ISBN. 978-1-4244-3394-0.
7. **Botai O. J.,** W. L. Combrinck and C.J. de W. Rautenbach, (2008). Nonstationary tropospheric processes in geodetic precipitable water vapour time series, Michael G. Sideris (ed). Observing our Changing Earth, International Association of Geodesy Symposia, Springer Berlin Heidelberg, 133: 625-630, doi:. 10.1007/978-3-540-85426-5.

Declaration

I, the undersigned declare that the thesis, which I hereby submit for the degree Doctor of Philosophy, Faculty of Natural and Agricultural Sciences at the University of Pretoria, is my own work except where acknowledged. This work has not previously been submitted by me for a degree at this or any other tertiary institution.

_____

Botai Ondego Joel

Acknowledgement

Analysis of geodetic and model-simulated data to describe nonstationary moisture fluctuations over southern Africa is the result of my affiliation with the Space Geodesy programme of the Hartebeesthoek Radio Astronomy Observatory, and the Department of Geography, Geoinformatics and Meteorology, University of Pretoria. I would like to thank those people directly involved with this thesis. My two principal advisors, Prof. Ludwig Combrinck and Prof. Sivakumar Venkataraman, provided superb guidance, which continuously required a thorough mastery and communication of my work. I would also like to thank Prof. C. J. de W. Hannes Rautenbach with whom we started this work. Hannes admitted me to the Meteorology postgraduate degree programme, provided me with an opportunity to pursue an intriguing, fascinating and interdisciplinary research topic and endlessly supported my PhD research cycle.

The assistance of Prof. Dr. Eng. Harald Schuh and Dr. Eng. Johannes Boehm from the Institute of Geodesy and Geophysics, Technical University of Vienna, Austria with regard to granting me a research visit to their institution, mentorship and freely made available high quality geodetic data utilised in this thesis is highly appreciated. I also want to thank the Inkaba yeAfrica project; a German-South African collaborative Earth Science initiative for providing financial support that enabled me to present my work at various international conferences and workshops. Institutions and individual principal researchers whose data sets were used in this thesis are also acknowledged.

I would especially like to thank my family, my father and other siblings for the moral support and holding onto me during the years of absence from the family. Their willingness to make available resources that supported my academic growth throughout my life is the reason why I am completing this degree. Mr. Paul Benedict Mmtoni deserves much credit for editing this work. But, above all, I also want to thank my wife, Mrs M. C. Botai, and all my children, who have been a constant in my life during this long process and whom I owe much for day-to-day love during this work. This piece of work is there because of you.

All these people highlighted here have given me a tremendous sense of self, taught me how to socialise, how to strive and achieve and also how to enjoy myself during happy and sad moments. This work therefore mirrors those I have mentioned here as well as many I have forgotten to list. To all these people, I express my gratitude.

Contents

List of Tables

List of Figures

Acronyms

AAM : Atmospheric Angular Momentum
ACs : Analysis Centres
AIRS : Atmosphere Infrared Sounder
ARMA : Autoregressive Moving Average
BIBER : Bound Influence by Standardised Residuals
CAM : Community Atmosphere Model
CCAM : Cubic Conformal Atmospheric Model
CO_2 : Carbon dioxide
DFA : Detrended Fluctuation Analysis
DORIS : Doppler Orbitography and Radiopositioning Integrated by Satellite
DWT : Discrete Wavelet Transform
ECMWF : European Centre for Medium-range Weather Forecasts
EEMD : Ensemble Empirical Mode Decomposition
EMD : Empirical Mode Decomposition
ENSO : El Nino Southern Oscillation
EOP : Earth Orientation Parameters
FFT : Fast Fourier Transform
FT : Fourier Transform
GGOS : Global Geodetic Observing System
GMF : Global Mapping Functions
GNSS : Global Navigation Satellite Systems
GPS : Global Positioning Satellites
gPWV : Global PWV
HALOE : Halogen Occultation Experiment
HartRAO : Hartebeesthoek Radio Astronomy Observatory
HCB : Highveld and Central Bushveld
HHT : Hilbert-Huang Transform
HT : Hilbert Transform
ICA : Independent Component Analysis
IGG : Institute of Geodesy and Geophysics

IF : Instantaneous Frequency
IR : Infrared
IGG : Institute of Advanced Geophysics and Geodesy
IWV : Integrated Water Vapour
IAG : International Association of Geodesy
ICRF : International Celestial Reference Frame
IGS : International GNSS Service
IMFs : Intrinsic Mode Functions
IMF : Isobaric Mapping Function
ITCZ : Inter-Tropical Convergence Zone
ITRF : International Terrestrial Reference Frame
IVS : International VLBI Service
LRD : Long-Range Dependent
MI : Mutual Information
MODWT : Maximum Overlap Discrete Wavelet Transform
NADA : Noise Assisted Data Analysis
NCAR : National Centre for Atmospheric Research
NCEP : National Centre for Environmental Prediction
NMF : Niell Mapping Function
NO_2 : Nitrogen dioxide
NWP : Numerical Weather Prediction
PCA : Principal Component Analysis
PSC : Polar Stratospheric Clouds
PRN : Pseudorandom noise
PWV : Precipitable Water Vapour
QQ : Quartile-Quartile
QSO : Quasi-Stellar Objects
RINEX : Receiver Independent Exchange
RH : Relative Humidity
RO : Radio Occultation
SAWS : South African Weather Service
SGT : Space Geodetic Techniques
SHADOZ : Southern Hemisphere ADditional OZoneondes

SLR : Satellite Laser Ranging

SOI : Southern Oscillation Index

SO_2 : Sulphur dioxide

SSA : Singular Spectrum Analysis

SS : Self-Similar

SSTs : Sea Surface Temperatures

TD : Tropospheric Delay

TEC : Total Electron Content

US : United States

UV : Ultra Violet

VEOF : Vector Empirical Orthogonal Function

VMF : Vienna Mapping Function

VLBI : Very Long Baseline Interferometry

WT : Wavelet Transform

WV : Water Vapour

WVR : Water Vapour Radiometry

ZTD : Zenith Tropospheric Delay

ZWD : Zenith Wet Delay

1. Introduction

Those among us who are unwilling to expose their ideas to the hazard of refutation do not take part in the scientific game.
Sir Karl Popper (1902-1994

This chapter provides a brief background and overview of space geodetic observations and the associated errors. The focus is on understanding and characterisation of geodetic data from the point of view of appropriately applying robust quantitative analysis strategies of tropospheric water vapour and tropospheric delay due to water vapour time series. The aim, objectives, research questions and scope of the study are also described.

1.1. Background

Measures of physical observables (which change over time) to study natural phenomena is of extraordinary importance for comprehending and characterisation of the underlying physical process. In particular, measurements from space geodetic techniques provide observational data that are used to depict the global picture of the Earth. The geodetic observations have a wide range of scientific applications such as fluctuations of the Earth (including precession, polar motion, nutation etc.,) and all types of height changes and deformation due to mass transfer between the solid Earth, the atmosphere and hydrosphere. However, these measurements are influenced by many geophysical processes such as weather and climate encompassing the atmospheric structure and dynamics, mass fluctuations, large water mass circulations and sea level fluctuations, postglacial rebound and, tide of solid Earth and Oceans; are often embedded with measurement error and are often corrupted by unknown noise sources. All the physical processes and noise introduce biases to the geophysical signal structure that are embedded in the parameters derived from data. It is crucial that these processes are therefore understood in data analysis. Key auxiliary parameters associated with all the primary geodetic target parameters are the tropospheric WV or the tropospheric delay of radio signals induced by WV and meteorological parameters. Geodetic time series often exhibits irregular, nonstationary and wide-band signals due to the complex interaction between different components of the Earth system. In the majority of the applications of studying time series of geodetic data, the underlying structure is often assumed to be

stationary. It is unlikely therefore that analysis techniques developed to reliably display the spatial-temporal variability of geodetic data are robust. Applying non-linear time series techniques to geodetic observations provide new information about the complex dynamics of the atmosphere-Earth system. In order to extract and characterise the relevant geophysical signals embedded in the data, a variety of measures (especially those that are data adaptive) are welcome.

1.2. Significance of the study

The significance of this research is in the realm of atmospheric modelling for space geodetic applications. The concept of non-linear and nonstationary time series analysis in space geodesy is relatively new. Therefore, research which explores data adaptive analysis methods will contribute to the understanding of the signal structure embedded in geodetic data from the point of view of displaying the spatial-temporal features present in the data. The analysis framework reported in this thesis envisage geodetic time series in different frequency bands thereby allowing for individual modes of oscillation to be linked to the different geophysical processes (or systematic variations of technical origin) by use of a variety of different measures that characterise their dynamical characteristics.

The goal of tropospheric geodetic modelling is to mitigate the contribution of tropospheric delays to the delay observable. Estimating the deterministic and stochastic components of the tropospheric delay due to WV using data adaptive techniques contributes to a more realistic robust geodetic parameter estimates. Therefore, the findings from this research have the capacity to impact on the tropospheric modelling strategies taking cognisance of the evolutionary second-order structure and self-similar behaviour in the geodetic data. In addition, this work creates awareness of the non-linearity and nonstationarity problem, as linearity and stationarity is always assumed within current models. A further logical step, which is briefly explored in the current work, will be to incorporate a non-linear component as an estimate and additional tropospheric correction, for instance to the range component of Satellite SLR. This should improve the observed minus computed residuals (as demonstrated herein) and allow for improved precise orbit determination and consequently, an improvement in the estimates of other modelled variables. This is true in general for all the space geodetic techniques, therefore this work paves the way forward for the development and inclusion into analysis software, of advanced

models which can be realised in software algorithms, thereby contributing to global geodetic science and atmospheric/geodetic water vapour modelling.

1.3. Aim and objectives

The overall aim of this study is to apply pragmatic analyses strategies to quantitatively investigate and understand the variability of the geodetic tropospheric WV over southern Africa and other parameters associated to the geodetic delay observable. In order to achieve this goal, the following specific objectives of the research are upheld;

- **a.** To investigate the nature of geodetic WV time series (e.g., linear trends, periodic and non-periodic transients) and determine the appropriate model that describes the variability pattern.
- **b.** To characterise the multiscale behaviour of tropospheric WV in southern Africa.
- **c.** To detect and characterise the non-linear and nonstationary properties of tropospheric geodetic WV and understand the associated geophysical causes.
- **d.** Incorporate non-linearity in atmospheric range correction and therefore evaluate its contribution to the Observed minus Computed (O-C) residuals in geodetic (e.g., SLR) analysis.

In this work, when the term *geodetic* WV is used, it refers to the contribution of atmospheric water vapour to the geodetic delay observable and does not refer to water vapour as a meteorological variable. However when the term water vapour (WV) is used, it refers to meteorological water vapour, which can be determined by meteorological instruments or can be inferred from the geodetic delay models, i.e., by separating the true delay from the total delay, so that the difference is in effect the water vapour contribution.

1.4. Research questions

There have been much theoretical and modelling work done on the methods and validation of geodetic WV derived from different data sources. Statistical methods have been used to analyse the spatial-temporal structure of long-term variations of WV. Additionally, a number of different studies to measure trends in WV have been reported by Nilsson and Elgered, (2008) and references therein. Despite these efforts, this research sought to address the following key questions;

a) What is the appropriate model that could describe a stochastic geodetic WV?
b) What is the scaling behaviour of geodetic WV data records?
c) Are the fluctuations of geodetic WV driven by stationary or nonstationary non-linear tropospheric processes?
d) If non-linearity is inferred, what would be the contribution for incorporating non-linear models to the correction of atmospheric range bias, in geodetic data analysis?

It is vital for these questions to be answered to enable the space geodesy research community to fine-tune the existing tropospheric models so as to accommodate the dynamic characteristics and contributions of WV to the geodetic delay observable estimation. Our interest in this research reported herein lies with the usage of time series analysis and the description of the general time series characteristic of WV based on an automatic data-driven model selection process.

1.5. Hypothesis

This study hypothesizes that tropospheric WV and other geodetic parameters exhibit non-linear and nonstationary properties with memory.

1.6. Scope of the study

The only International VLBI Service (IVS) station in Africa (Hartebeesthoek Radio Astronomy Observatory, HartRAO) is located at 25.89^0 S; 27.090 E, South Africa. This is a fiducial geodetic station that has been contributing to the space geodesy community geodetic observations for nearly three decades. The current research covers quantitative analyses of stochastic, non-linear and nonstationary properties of geodetic WV and associated parameter tropospheric delay due to WV (e.g., tropospheric delay and tropospheric delay linear gradients) over and around HartRAO and WV derived from NWP model simulations. In the analysis, the non-parametric time series analysis methodologies such as Detrended Fluctuation Analysis (DFA), Wavelet Transform (WT) and EEMD are used. In order to evaluate the theoretical concepts developed throughout this work within the context of space geodetic techniques, an application was developed for SLR. This approach utilised an enhancement of existing software (Combrinck, 2010) to include non-linear effects not provided for in the current modelling of the additional range delay due to the atmosphere.

2. Tropospheric modelling for space geodetic applications

... by observing the most distant objects of the Universe (Quasars), we can learn things that happen around the corner, here on Earth.
Miguel T. (1999).

A review of the literature on geodetic tropospheric modelling is presented in this Chapter. In particular, the existing literature focussing on the analyses strategies for the derived tropospheric parameters from geodetic measurements and recent developments in tropospheric TD/WV modelling is reviewed.

2.1. Introduction

Over the last three decades, Space Geodetic Techniques (SGT) such as VLBI, GNSS, SLR, Doppler Orbitography and Radiopositioning Integrated by Satellite (DORIS) and Water Vapour Radiometry (WVR) have continued to improve the accuracy, timely and continuous provision of geodetic products, which have benefits to both science and society (Niell, 2005). Some of the unique products from SGT, such as Earth Orientation Parameters (EOP), Earth gravity fields, sea surface and sea level changes, solid surface geometry and kinematics have been provided reliably and consistently. These products form the basis for scientific application in areas such as geodynamics, global change, land management, engineering and navigation. In addition, these products have been used in surveying, global spatial data infrastructure as well as in rural and urban development. Geodetic data have applications in monitoring surface deformation (e.g., GPS, VLBI and SLR), navigation (e.g., GPS), natural hazards like volcanic eruptions and stress levels for earthquake hazard assessment (see for example Nyst *et al*., 2006). A summary of geodetic parameters related to the system Earth reported by Rothacher, (2002) is tabulated in Table 2.1.

Table 2.1. Geodetic parameter groups related to system Earth as reported in Rothacher, (2002).

Parameter type		VLBI	GPS	SLR	DORIS
RS	Quasar positions	x			
	Orbits (Satellites, Moon)		x	x	x
EOP	Nutation	x	x		
	Nutation rates	x	x	x	x
	UT1-UTC	x			
	LOD	x	x	x	x
	Polar motion	x	x	x	x
TRF	Station positions	x	x	x	x
Gravity field	Geocenter		x	x	x
	Low degree		x	x	x
Atmosphere	Troposphere	x	x		x
	Ionosphere	x	x		x

The scientific application of geodetic data has been fundamental in the understanding of the structure and deformations of the Earth's crust (Sansò and Gil, 2006), mantle and core. Geodetic data have also been applied in the analysis of geophysical fluids (Wolfgang and Richter, 2008), the coupling between the free atmosphere and solid earth. For instance, Davies *et al*., (2004), in an analysis of GRACE and GPS data, identified a climate-driven deformation signal of the solid earth. Seasonal variations in gravity fields were positively correlated with climate-driven fluxes of surface water.

Current VLBI community under the auspices of the IVS working groups have redefined the target accuracy goals in their products. For instance, the next generation VLBI system (termed VLBI2010) ought to provide a 1 mm positional and 1 mm per annum velocity accuracies within the ITRF. The VLBI network is depicted in Figure 2.1 where a number of these stations have collocated GNSS, SLR or DORIS instruments, which have applications in

geodesy and atmospheric studies. Continuous measurement of EOP and the rapid generation and distribution of geodetic products (Niell, 2005) are additional targets. Since the atmospheric structure and dynamics influences the accuracy of the estimated geodetic products, these applications require that atmospheric biases in the geodetic products ought to be minimised.

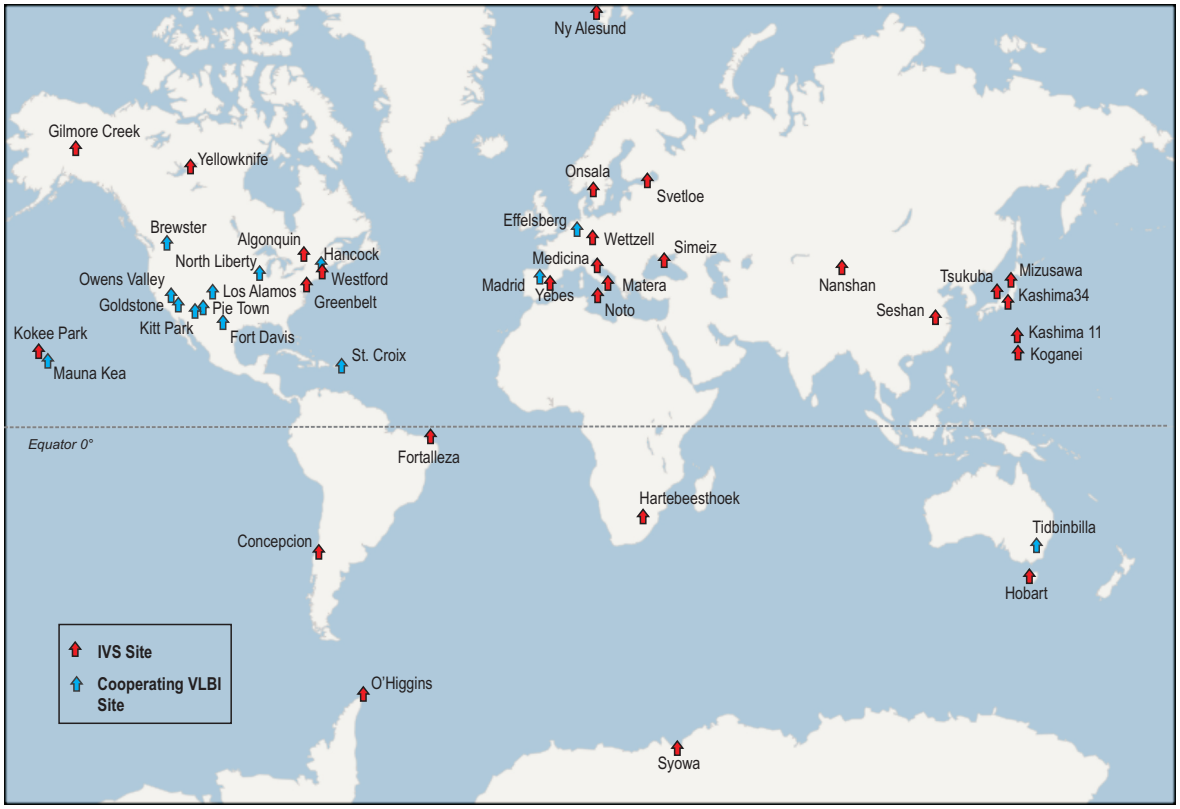

Figure 2.1 The current global VLBI network.

Research that focuses on geodetic tropospheric modelling involves the development of strategies for improving the atmospheric models that reduce random and systematic bias components of the Tropospheric Delay (TD) observable. The TD observable is the extra path length covered by the radio wave due to a delay caused by the change in the direction of propagation of the radio signal in the troposphere (ionosphere, see Ho *et al*., (1997); Hobiger *et al*., (2006)) and troposphere (Saastamoinen, 1972; Haase *et al*., 2003).

Modelling and analyses strategies of the tropospheric contribution to the delay observable ought to be improved if the goal of one millimetre accuracy in the estimation of

EOP, station positions and velocities is to be achieved. To measure the improvement in modelling accuracy, the secular signals associated with crustal deformations are monitored regularly. Other non-secular geophysical deformations (resulting from atmospheric loading, groundwater level seasonal fluctuations etc.) may be seen in the repeatability of baseline length between two stations.

The TD component is the most dominant bias contributor in GNSS and VLBI geodetic measurements (Gradinarsky *et al*., 2000; Behrend *et al*., 2008). As the signal traverses the neutral atmosphere, it experiences a change in the effective path length and direction of propagation. The change in the geometric path of the microwave signal occurs due to spatial-temporal variability of the refractive index. One major contributor to variability of the refractive index is tropospheric WV. In the troposphere, WV induces the highly fluctuating tropospheric delay observable, commonly referred to as the wet delay.

Current tropospheric modelling strategies do not accurately capture actual stochastic and non-homogeneous properties of geodetic WV or TD due to WV. Although a lot of effort has been made to improve the modelling and analysis of the atmospheric contribution to the geodetic TD observable, the envisioned accuracy constraints, as reported by for instance Niell, (2005), require better and in addition regional (rather than just global) modelling techniques that ould correct for the random and systematic components of the geodetic TD observable biases (e.g., troposphere and instrumental). There is need for improved geodetic site specific model analysis strategies of the anisotropic atmosphere and error budgets.

The VLBI2010 project in particular focuses on addressing key constraints, which undermine the attempts to meet mm accuracy geodetic station positions through simulations (Wresnik *et al*., 2008). Research in appropriate geometry, sky coverage, system constraints, tropospheric modelling and software have been reported by Niell, (2005); Boehm *et al*., (2006); Nilsson and Haas, (2008). Notwithstanding all these efforts, the current modelling strategies do not consider possible stand-alone station dependent strategies that might constrain (to some extent) the empirical and systematic biases in geodetic parameters. The next section describes measurements and empirical formulation of tropospheric parameters (e.g., delay gradients, WV) that influence the accuracy of the geodetic TD observable.

2.2. Structure of the atmosphere

Different conventions or physical parameters are used to describe atmosphere layers. In space geodesy, the atmosphere would be classified into troposphere; mesosphere, stratosphere and

ionosphere (see Table 2.2). The troposphere is the neutral part of the atmosphere which includes the troposphere, tropopause and stratosphere as defined in meteorology. The troposphere has a quasi-frequency independent refractive index at microwave signals below 30 GHz. The neutral layer (0 to 15 km height) has a high concentration of neutral particles. This part of the atmosphere is dominated by physical processes which characterise the dynamics of the atmosphere determined by turbulent molecular viscosity. Due to spatial-temporal variability of temperature and pressure, mass re-distribution in the atmosphere causes local and global circulation systems which are affected by Earth rotation, large water bodies and other extra-terrestrial forces. This coupling is often captured in the correlation of EOP and the Atmospheric Angular Momentum (AAM), see for example Zhang *et al*., (2003) and references therein.

Table 2.2. Geodetic view of the atmosphere

Height, km	**0 to 50**	**≥ 50**	**≤ 80**	**≥ 80**
Temperature	Troposphere Stratosphere	Stratopause Mesosphere		Thermosphere Exosphere
Refractivity	Troposphere			
Mixing ratio			Homosphere	Heterosphere
Magneto-electronic structure	Neutral atmosphere	Ionosphere		Ionosphere Plasmasphere Magnetosphere

The troposphere exhibits turbulent processes and therefore the mixing ratio of its constituent's undergoes subtle changes, except for WV which experiences a marked change. However, at high altitudes, > 80 km, the kinetic molecular processes dominate the turbulence processes. Below 80 km, the atmosphere is called the homosphere. The heterosphere is part of the atmosphere medium above 80 km. Furthermore, based on the dynamic processes, the troposphere can be divided into the atmospheric boundary layer (layer adjacent to the Earth's surface with characteristic high Reynolds number, implying high turbulent flow) and the free troposphere.

The ionosphere layer consists of mainly ionised atmospheric constituents. In this layer, the Ultra Violet (UV) and the X-ray radiation produce non-vanishing ionisation density. The ionosphere is conventionally characterised by the Total Electron Content (TEC) and scale height (H). These parameters exhibit spatial-temporal variability caused by

underlying processes. The speed and attenuation of microwave signals traversing the ionosphere and troposphere are affected by the refractive index of the atmosphere. In the neutral atmosphere, the optical and radio frequencies are related to the $n(\boldsymbol{r}, \mathrm{t})$ using the Cauchy Equation given in Equation (1) as reported by Griffiths, 1999):

$$n(\boldsymbol{r}, \mathrm{t}) = 1 + N_a\left(1 + \frac{\mathrm{B}}{\lambda^2}\right); \tag{1}$$

where N_a and B are the refractivity and dispersion coefficient respectively. Following Riepl and Schluter, (2000), B= 1.7× 10^{14} m^2 for the atmosphere. Note that, at optical and radio wavelengths, $\lambda^2 \sim$ B and B < λ or B << λ^2. As a result, the signal propagation in the atmosphere will experience a time delay relative to it's propagation in vacuum due to n < 1.

In Space Geodetic measurements, the microwave signals experience bending due to the gradients, see Equation (2), of the troposphere (Davis, 1992; Davis *et al*., 1993) and this is commonly called the geometric delay. Further, the microwave signals are slowed down and therefore causes an excess delay which is a function of the refractivity given in Equation (2);

$$n(\{\boldsymbol{r}, z\}, \mathrm{t}) = n_0(z) + \vec{\xi}(z) \times \boldsymbol{r} + \cdots; \tag{2}$$

where $n_0(z)=n(r=0, z)$ is the horizontal invariant component of the refractive index and $\boldsymbol{\xi}(z) = \nabla_r n(\{\boldsymbol{r}, z\}, \boldsymbol{r}=0, \mathrm{t})$ is the horizontal gradient. According to Rocken *et al*., (1993), the summation of the geometric and excess delay could be expressed as shown in Equation (3) ;

$$ZTD = \int_l n(\{\boldsymbol{r}, z\}, \mathrm{t})\mathrm{d}r - \tau_G; \tag{3}$$

where $r \in l$ and *ZTD*, l and τ_G are the total tropospheric delay in the zenith direction, the curved ray path and the straight-geometric delay through the atmosphere. Equation (3) can also be expressed in terms of the slowing and bending terms in Equation (4),

$$ZTD = \int_l \left[n\{\boldsymbol{r}, z\}, \mathrm{t}\right]\mathrm{d}r - \left[S - \tau_G\right]; \tag{4}$$

where S is the curved path length along L. The slowing and bending components are given by first and second terms of the right hand side of Equation (4). The bending term ($\left[S - \tau_G\right] \lesssim 10$ mm) is elevation dependent and vanishes if the ray path is in the zenith (in the absence of gradients).

The temperature and density of the tropospheric constituents affect the geodetic *ZTD* observable, which in turn determines its spatial-temporal distribution. The stochastic behaviour of *ZTD* is still one of the limiting factors that restrict the accuracy of space

geodetic techniques as reported by Boehm *et al*., (2006). Since *ZTD* is caused by the dry and wet part of the atmospheric constituents, the delay could be separated into the hydrostatic (referred to as the dry part in the literature as it is caused by the refractivity of the dry gases in the troposphere, but in fact contains the non-dipole component of water vapour refractivity) and wet (non-hydrostatic) components:

$$ZTD(\varepsilon) = \mathrm{Y}_h(\varepsilon)TD_h^z + \mathrm{Y}_w(\varepsilon)TD_w^z; \tag{5}$$

where $\mathrm{Y}_{w,h}$ are the elevation (ε) dependent (and azimuth independent) mapping functions which are determined independently from numerical weather prediction models, see Boehm *et al*., (2006) or radiosonde data (Niell, 2000; Niell, 2001). The predictable ZTD are often estimated from empirical models using unbiased surface meteorological data to mm accuracy (Janes *et al*., 1991). The wet delay component TD_w^z is related to the quantity of WV in the atmosphere and is often used as a passive tracer for atmospheric studies. For instance it can be used to reveal the structure of the atmosphere at all temporal and spatial scales. In the next section, we focus on the characteristics of the troposphere using refractivity, (*N{**r**, z}, t*) and the relationship between N, TD and tropospheric precipitable WV.

2.3. Refractivity, tropospheric delay and precipitable WV

The refractive index, $n(\{\boldsymbol{r}, z\}, \mathrm{t})$ of the troposphere is often expressed in terms of $N(\{\boldsymbol{r}, z\}, \mathrm{t})$ i.e. $N(\{\boldsymbol{r}, z\}, \mathrm{t}) \equiv 10^6 \times n(\{\boldsymbol{r}, z\}, \mathrm{t})$. Expressed as a function of temperature, T, the partial pressure of dry air, p_d and the partial pressure of WV, p_v (e.g., Smith and Weintraub 1953; Thayer, 1974; Bevis *et al*., 1994), the compact form of $N(\{\boldsymbol{r}, z\}, \mathrm{t})$ is given in Equation form as,

$$N = \frac{k_1 \cdot Z_d^{-1} \cdot p_d}{T} + \frac{k_2 \cdot Z_v^{-1} \cdot p_v}{T} + \frac{k_3 \cdot Z_v^{-1} \cdot p_v}{T^2} + \frac{k_4 \cdot p_c}{T^2}; \tag{6}$$

where the symbols denoting the spatial-temporal dependence have been dropped for simplicity. The corrections from the departures of moist air from the ideal gas are given by the inverse compressibility factors, Z_d^{-1} and Z_v^{-1} for air and water respectively. The coefficients $k_i, \forall i = 1,2,3$ and their associated errors are given in Table 2.3 according to Bevis *et al*., (1994). If temperatures T and t are measured in Kelvin and Celsius and using James, (1967), the compressibility factors can be expressed as:

$$Z_d^{-1} = 1 + p_d \left\{ 57.97 \cdot 10^{-8} \cdot \left(\frac{T + 0.52}{T} \right) - 9461 \cdot 10^{-4} \cdot \frac{t}{T^2} \right\},$$
$$Z_v^{-1} = 1 + \frac{1650 \cdot p_v}{T^3} \cdot \left\{ 1 - 0.01317 \cdot t + 1.75 \cdot 10^{-4} \cdot t^2 + 1.44 \cdot 10^{-6} \cdot t^3 \right\}. \quad (7)$$

Table 2.3. Nominal atmospheric refractivity constants and their standard errors.

Refractivity constant	k_1 **[*Kmb*$^{-1}$]**	k_2 **[*Kmb*$^{-1}$]**	k_3**[*K*^{2}*mb*$^{-1}$]** $\times 10^5$
Smith & Weintraub, (1953)	77.607 ± 0.013	71.60 ± 8.500	3.7470 ± 0.0310
Thayer, (1974)	77.604 ± 0.014	64.79 ± 0.080	3.7760 ± 0.0040
Bevis *et al.*, (1994)	77.600 ± 0.005	70.40 ± 2.200	3.7390 ± 0.0012

From Equation (6), the dry and wet refractivity can also be obtained, which is analogous to the TD_h^z and TD_w^z derived from the geodetic TD observable. The first term on the right hand side of Equation (6) denotes dry refractivity, N_d and is often compared to the TD_h^z. The error in N_d is less than 0.5% (ITU, 2003). The second and last terms of Equation (6) are called the wet refractivity N_w. While N_d is about $0.75N$, N_w is the largest contributor to the variability of tropospheric refractivity. The zenith delay components given in Equation (5) can then be obtained by the integration of the vertical profiles of $N_{w,d}$ (see Equations (8) and (9)) which could be computed from radiosonde and numerical weather prediction simulation data sets.

$$\begin{aligned} TD_h^z &= 10^{-6} \times \int_{\boldsymbol{r}} N_d \mathrm{d}\boldsymbol{r}, \\ &= 10^{-6} \times \int_{\boldsymbol{r}} k_1 R_d \rho \mathrm{d}\boldsymbol{r}; \end{aligned} \quad (8)$$

$$\begin{aligned} TD_w^z &= 10^{-6} \times \int_{\boldsymbol{r}} N_w \mathrm{d}\boldsymbol{r}, \\ &= 10^{-6} \times \int_{\boldsymbol{r}} \left(\frac{k_2' p_v Z_v^{-1}}{T} + \frac{k_3 p_v Z_v^{-1}}{T^2} \right); \end{aligned} \quad (9)$$

where $\boldsymbol{r} = [0 \quad 0 \quad z]$ and $M_d k_2' = M_d k_2 - k_1 M_w = 22.1 \pm 2.2 \ K \ mb^{-1}$ (Bevis *et al.*, 1994). M_w and M_d are the molar weight of wet and dry air respectively. Here, R_d and ρ are the specific gas constant and the mass density of dry air respectively.

Using the hydrostatic equilibrium assumption reported by Houghton, (1986); $dp = -g(z) \times \rho(z) dz$, (here, z denotes the vertical coordinate) and a standard atmosphere profile, TD_h^z can be re-written as:

$$TD_h^z = \frac{2.2768 \times 10^{-3} \pm 5 \times 10^{-7} \times p_s}{1 - 0.00266 \times \cos 2\lambda - 0.00028H}, \tag{10}$$

where λ and H is the latitude and height, in km over the geoid respectively. The bias in TD_h^z due to the surface pressure, p_s hPa, λ and H can be derived using Equation (10) as follows; the delay, surface pressure, latitude and height biases are defined as $\sigma_\tau, \sigma_p, \sigma_H$ respectively such that:

$$\sigma_\tau \propto \sigma_p,$$
$$\sigma_\tau \propto \sigma_\lambda,$$
$$\sigma_\tau \propto \sigma_H.$$

Tropospheric integrated WV (hereafter *IWV*) $\mathrm{kg/m^2}$ of the column above height z_0 can be derived from TD_w^z using mean temperature, T_m of WV in the atmospheric column using Equations (11) and (12) respectively, see for instance Askne and Nordius, (1987) and Bevis *et al*., (1992). [1]

$$T_m = \frac{\int_{z_0}^{z_T} \rho \mathrm{d}z}{\frac{\int_{z_0}^{z_T} \rho \mathrm{d}z}{T}}, \tag{11}$$

$$IWV = \frac{10^6 TD_w^z}{R_v \left\{ -k_1 \epsilon + k_2 + \frac{k_3}{T_m} \right\}}. \tag{12}$$

The *IWV* derived from pressure and mean tropospheric temperature measurements suffer from measurement errors. [2] One way to eliminate propagating these errors is to use TD_w^z in the data assimilation for numerical weather prediction. In geodetic applications, *IWV* is only a

[1] Radiosonde profiles have been used by Bevis *et al*., (1992) and Emardson and Derks, (1998), to derive linear and quadratic relation between surface temperature T_s and T_m

[2] Using Equation (12), $ZWD \propto IWV$ and therefore $\sigma_{p_s} \sim 1$ mbar ~ 0.4 mm IWV.

derived quantity while TD_w^z is more directly used to improve on the accuracy of the delay observable. [3]

2.4. Overview of geodetic VLBI, GNSS and SLR

In the geodetic VLBI technique, extragalactic radio signals (at 2.3 and 8.36 GHz) from Quasars are received simultaneously by a network of terrestrial telescopes, from which very accurate estimates of distances between telescopes and estimates of Earth rotation can be made. Radio signals from Quasars or Quasi-Stellar Objects (QSO) are highly variable and do not suffer from parallax or proper motion since the radio sources are very distant. In fact, they are the furthest known objects. In astronomy, the VLBI measurement principle is based on interferometry where pairs of telescopes from various locations observe/receive and combine simultaneously the signals *in-phase* from the same object to reveal detail structure of the object. For the purposes of geodesy, sources are selected that exhibit little or no structure (even with the high resolution possible with VLBI) so as to use the quasars as point sources, effectively fixed points realising the International Celestial Reference Frame (ICRF). Simultaneous combination of geodetic VLBI signals in real time is not possible; instead, the signals measured at each station are recorded onto some storage media (magnetic tapes, disks) together with a very accurate time-keeping signal from a high-precision atomic clock. The recorded observations and the time signal are later retrieved at various correlators stations at selected data centers.

The time shift in the arrival time of the signal (hereafter the delay) at the relative positions of the VLBI telescopes is used to determine, with high accuracy, the positions and velocities of each station relative to each other and the positions of the radio sources. These measurements have application in the determination and maintenance of the International Terrestrial Reference Frame (ITRF) and ICRF, EOP and atmospheric parameters. In addition, during further processing of the time history of baselines between stations, three dimensional motions of the stations are obtained. Horizontal motion carries information regarding plate tectonics and, in addition, the vertical component could contain information related to geophysical processes such as post-glacial rebound.

Unfortunately, geodetic VLBI measurements contain complex signals that could be translated as *biases*[4]. Some of these errors are as a result of non-synchronised measurements,

[3] For normal atmospheric conditions, it is expected that 10 kg/m^2~ 65 mm of the zenith tropospheric delay.

hardware constraints like:-(i) changes in the hardware or (ii) mechanical response of the hardware to weather; e.g. thermal expansion (Wresnik *et al*., 2005), wind, precipitation and atmospheric delay due to variations in the refractivity (Haase *et al*., 2003; Boehm *et al*., 2006) and geophysical factors such as ocean loading (Ray, 1999; Scherneck, 1991), crustal motion (Haas *et al*., 2000; Haas *et al*., 2003) and atmospheric pressure loading (Petrov and Boy, 2004). To determine geodetic VLBI parameters with very high accuracy, all these signals ought to be modelled and accounted for during processing of the data.

In general, the working principle of GNSS (e.g. GPS, GLONASS, GALILEO and the planned COMPASS system) is similar to VLBI (e.g. radio waves traversing through the ionosphere). However, the major difference is that GNSS uses artificial satellites, while the VLBIs use distant radio sources. In VLBI the main observable is the phase delay, whereas in GNSS it is the range. The American GPS navigation system consists of a constellation of more than 30 satellites orbiting at approximately 20 000 km at 55° inclination. This geometry allows simultaneous visibility of at least four satellites by any receiver on the surface of the Earth (Hofmann-Wellenhoff *et al.* 1997). The Russian GLONASS currently has about 20 satellites in its constellation. The GLONASS satellites orbit at a 64.8° inclination and at an altitude of about 19 000 km.

Radio signals (at different frequencies) from transmitters onboard GNSS satellites are continuously transmitted and received by GNSS receivers. GPS uses different codes; CDMA[5] to separate signals from different satellites, while GLONASS satellites transmit their radio signals at different frequencies, the FDMA[6]. The GPS satellites transmit right-handed polarized radio signals at two frequencies; L_1~ 1.57542 GHz and L_2~ 1.2276 GHz, that are modulated with two Pseudorandom noise, PRN codes (C/A and P(Y)) and with a navigation message. GLONASS transmit right-hand circular polarized radio signals with band specifications of $1.6025625 < L_1 < 1.6155$ and $1.240 < L_2 < 1.260$ GHz as well as the C/A and P carrier band widths of 1.0 and 10.0 MHz respectively.

The travel time or distance between the receiver and the satellite is calculated by comparing the time the signal was transmitted to the time the signal was received (based on

[4] The description of errors in geodetic analysis should be treated with caution. Most of the signals that were initially treated as biases have recently found significant scientific applications. In geodetic VLBI, for astrometric and geodetic applications, the delays other than the geometric delay are regarded as biases.

[5] Code Division Multiple Access

[6] Frequency Division Multiple Access

the receiver clock). The Pseudo distances calculated in this way have ~ 10 m accuracy. Using the carrier phases (measured with an order of accuracy of millimetres), centimetre level of position accuracy could be obtained but for the inherent integer ambiguity (Hofmann-Wellenhoff *et al*., 2001). Just like the measurement errors in VLBI, GNSS error sources such as clock, orbital, receiver noise, multipath, ionospheric and tropospheric errors have to be accounted for to obtain an optimal delay observable. The ionospheric delay is often removed based on the concept of double-differencing[7].

2.4.1. GPS delay observable

In GPS observations, measurements are often carried out using the pseudo-range (or code range) and carrier phase. The primary measurement is the phase measurement, which has applications for high precision positioning. Pseudo-ranges are only accurate to a meter and are therefore considered as ancillary observations to be used for eliminating synchronization, clock, integer ambiguity and cycle slip biases. The basic form of carrier phase observation (also zero difference observation, ρ_z) is given by Equation (13)

$$\rho_z = \rho_0 - \lambda N - \tau_{ion} + \tau_{trop} + \tau_{cl} + \tau_{hw} + \tau_{syn} + \tau_{or} + \tau_{apc} + \tau_{rel} + \zeta, \qquad (13)$$

where ρ_0, λ and N are the geometrical distance from the satellite to the receiver, wavelength of the carrier signal and the ambiguity integer. The ionospheric delay is τ_{ion} and the tropospheric delay is τ_{trop}, with τ_{cl}, τ_{hw}, τ_{syn}, τ_{or} and τ_{apc} the combined receiver and transmitter clock biases, the hardware bias of the receiver, synchronisation error, the receiver/transmitter antenna orientation error and the antenna phase centre offset respectively. A relativistic contribution is denoted by τ_{rel} while ζ describes residual errors.

The satellite transmitter and receiver clock errors in Equation (13) require double-differencing, see Equation (15), for their elimination (Alber *et al*., 2000). The satellite clock error is often eliminated using single-differences, which are formed by differencing the simultaneous one-way measurements from the satellite to two ground receivers given by Equation (14). Furthermore, ionospheric and tropospheric delays are also eliminated (or at

[7] In double-differencing, single differences are first formed by subtracting observation equations from two separate receivers to a single satellite. Taking the difference between these two single differences for a specific receiver pair gives the carrier phase double difference (Alber *et al*., 2000).

least reduced) if the receivers are closely spaced. In double-differencing, the difference between two single-differences is computed and this eliminates the two receiver clock biases.

$$\begin{aligned}\Delta\tau^1_{r_1,r_2} &= \tau^1_{r_2} - \tau^1_{r_1},\\ \Delta\tau^2_{r_1,r_2} &= \tau^2_{r_2} - \tau^2_{r_1},\end{aligned} \tag{14}$$

$$\Delta\tau_{dd} = \Delta\tau^2_{r_1,\,r_2} - \Delta\tau^1_{r_1,\,r_2} \tag{15}$$

Here $\tau^1_{r_1}$ and $\tau^2_{r_1}$ are the observations of satellites 1 and 2 by receiver r_1 and $\tau^1_{r_2}$, and $\tau^2_{r_2}$ the observations of satellites 1 and 2 by receiver r_2, $\Delta\tau_{dd}$ is the double-difference delay observation component. Optimal correction of receiver clock errors is only possible, firstly, if the measurements are taken at the same time or a priori knowledge of antenna position, satellite position, and the pseudo-range measurements are used to constrain the offset of the station clock to within a μs.

For original carrier phases τ_{L_1} and τ_{L_2}, a combined phase measurement with combination factors, κ_1 and κ_2 is given by Equation (16),

$$\tau_{L_{1,\,2}} = \kappa_1\tau_{L_1} + \kappa_2\tau_{L_2} \tag{16}$$

If the carriers phases τ_{L_1} and τ_{L_2} have systematic errors $\delta\tau_{L_1}$ and $\delta\tau_{L_2}$, the combined observation will have a systematic error given by Equation (17);

$$\delta\tau_{L_{1,2}} = \kappa_1\delta\tau_{L_1} + \kappa_2\delta\tau_{L_2} \tag{17}$$

For a dispersive ionosphere, $\delta\tau_{L_2}.f_2^2=\delta\tau_{L_1}.f_1^2$, for which Equation (18) could be written as

$$\delta\tau_{L_{1,2}} = \Gamma_{ion}\delta\tau_{L_1}, \tag{18}$$

Where $f_2^2.\Gamma_{ion}=f_2^2.\kappa_1+k_2.f_1^2$, quantifies the first-order ionospheric contribution in linear combination. Ionospheric contribution would therefore be reduced with the selection of optimal values of κ_1 and κ_2; e.g. in GPS observations, an ionospheric free measurement would be described based on $f_1^2-f_2^2.\kappa_1=f_1^2\sim2.25$ and $f_1^2-f_2^2.\kappa_2=-f_2^2\sim-1.55$. In this case the combined noise level could be given by 2.7322σ. [8]

[8] According to error propagation law, given the σ of independent measurements $\{x, y, z\}$, related as $r = ax + by + cz$, the combined error in r, σ_r is given by $\sigma_r = \sqrt{(a\sigma)^2 + (b\sigma)^2 + (c\sigma)^2}$ or $\sigma_r = \sigma\sqrt{(a^2 + b^2 + c^2)}$.

2.4.2. The geodetic VLBI delay observable

The VLBI observables (Shapiro 1976; Cannon, 1978; Thompson *et al*., 2001), such as the phase delay, delay rate and the group delay, carry positional information of extraterrestrial radio sources and terrestrial telescopes, which have vital geodetic and astrometry applications. The phase delay, though unambiguous, is the most accurate VLBI observable. It is used to determine the quasi-inertial reference frame to sub-milliarcsecond (mas) precision using the method of VLBI differential astrometry. [9]

The physical system of the geodetic VLBI could be represented by the schematic diagram shown in Figure 2.2 (Cannon, 1978). In the Figure 2.2, vector **r** is a unit vector, LO is the local oscillator and RF AMP is the radio frequency amplifier. As depicted in the Figure the signal (with frequency, ν) from a distance radio source arrives at antenna 2 after a time delay, τ_g with respect to antenna 1.

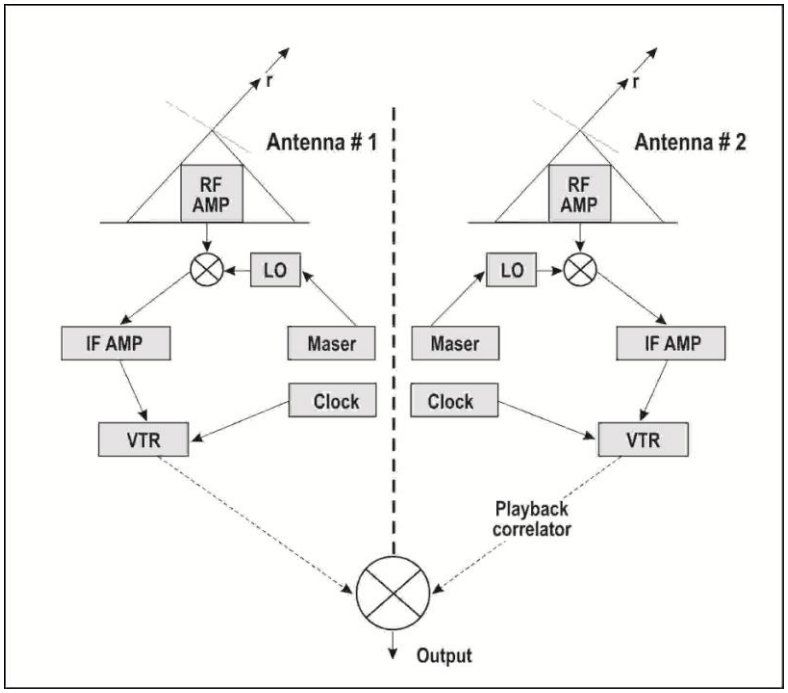

Figure 2.2. Schematic of a long baseline interferometer

[9] In VLBI differential astrometry, the interferometric phases of two radio sources are alternately sampled once every few seconds for a given period of time. The different contributions to the phase delays are removed, thereby isolating the geometric contribution, which is finally modelled using the weighted-least squares algorithm of the geometrical parameter adjustment.

In an ideal case, contributions from the propagation medium and instrumentation are neglected and therefore the geometric delay is given by Equation (19),

$$\tau_g = \frac{\mathbf{B}}{c}\hat{r} \tag{19}$$

where $c, \mathbf{B}$ and $\hat{r}$ are the speed of light, the baseline vector and direction unit vector of the radio source from the centre of the Earth. The interferometer response, $\Lambda(\nu, \boldsymbol{r})$ assuming accurate delay tracking with respect to angular position, $\boldsymbol{r_0}$, is given by:

$$\Lambda(\nu, \boldsymbol{r}) = S_\nu e^{\left\{j2\pi\frac{\nu_0}{c}\mathbf{B}\cdot\boldsymbol{r_0}\right\}} \times e^{\left\{j2\pi\frac{\nu}{c}\mathbf{B}\cdot(\boldsymbol{r}-\boldsymbol{r_0})\right\}}, \tag{20}$$

where S_ν is the total flux density of the source at the observing frequency, ν.

For extended radio sources, the interferometer response will have two components given in Equation (21); a rapidly varying and time dependent response (this is equivalent to the point source response) but centred at $\boldsymbol{r_0}$ at the observing frequency, ν_0, and a slow-varying component called the visibility function, $V(\mathbf{B})$, which describes the amplitude and phase offsets. The visibility function is a complex function that depends on the bandwidth of the radio source, the emission content of the radio sources and the baseline geometry.

$$\Lambda(\nu, \boldsymbol{r}) = S_\nu e^{\left\{j2\pi\frac{\nu_0}{c}\mathbf{B}\cdot\boldsymbol{r_0}\right\}} V(\mathbf{B}). \tag{21}$$

However, for simplicity, Equation (20) could be interpreted as sinusoidal response patterns (also called the fringes), which are formed as a result of the continuous variation of the geometric delay caused by the diurnal motion of the point radio sources, and therefore the interferometer response function could be expressed as:

$$\Lambda(\nu, \boldsymbol{r}) = S_\nu \cos(2\pi\mathbf{B}_\lambda \cdot \hat{\mathbf{r}}), \tag{22}$$

where the interferometer fringe phase (due to the quasi-sinusoidal response) is defined as $\phi_g = 2\pi\nu\tau_g$.[10]

In general, the interferometer phase equation, taking into account the contributions from the source structure and contributions from the propagation medium (e.g. the atmosphere) as reported in Thompson *et al.* (2001), is given by:

$$\phi(t) = \frac{2\pi L}{\lambda}\left\{\cos D\cos\delta\cos\left[A_0 + \omega t - \alpha\right] + \sin D\sin\delta\right\} + \phi^{at}(t) + \phi^{ins}(t) + \phi^{str}(t), \tag{23}$$

[10] Fringe phases are the spatial-temporal patterns of the interferometer response

where $L = |B|$, A_0 and D are the right ascension and declination of the baseline vector, **B**, respectively. The angular velocity of the Earth is denoted as ω rad/sec while α and δ are the right ascension and declination of the radio source respectively. The propagation medium, the instrumentation and source structure contributions to the fringe phase pattern are denoted by ϕ^{at}, ϕ^{ins} and ϕ^{str} respectively. From Equation(23), the geodetic parameters relevant to the thesis reported here are the **B**, A_0, D and ϕ^{at}.

The phase delay τ_ϕ, is the ratio of the observed fringe phase (with phase ambiguities), which is an integral number of 2π's and the reference angular frequency is given by Equation (24).

$$\tau_\phi = \frac{\phi(t) + 2\pi n}{\omega}; \quad \omega = 2\pi\nu. \tag{24}$$

The group delay τ_g is the derivative of the fringe phase with respect to angular frequency and is given by Equation (25),

$$\begin{aligned} \tau_g \mathrm{d}\omega &= \mathrm{d}\phi, \\ &= \frac{1}{2\pi}\frac{\mathrm{d}\phi}{2\nu}. \end{aligned} \tag{25}$$

The time derivative of the phase delay is given by:

$$\begin{aligned} \dot{\tau}_\phi &= \frac{\mathrm{d}\tau_\phi}{\mathrm{d}t}, \\ &= \frac{1}{2\pi\nu}\frac{\mathrm{d}\phi}{\mathrm{d}t}. \end{aligned} \tag{26}$$

The uncertainties in the VLBI observations are dependent on temperature of the antenna (T_{ant}) and of the receiver (T_s), the observing frequency and band width ($\Delta\nu$) of the radio signal. According to Thompson *et al.* (2001), the uncertainty in the phase delay σ_ϕ, group delay σ_g, and the phase delay rate, $\sigma_{\dot{\tau}_\phi}$ are given by Equations (27), (28) and (29) respectively.

$$\sigma_{\tau_\phi} = \frac{T_s}{2\pi T_{ant}\sqrt{2}}\left\{\frac{N}{\nu^2 \Delta\nu \Delta t}\right\}. \tag{27}$$

$$\sigma_{\tau_g} = \frac{T_s}{2\pi T_{ant}\sqrt{2}}\left\{\frac{1}{\Delta\nu_{rms}^2 \Delta\nu\Delta t}\right\}. \quad (28)$$

$$\sigma_{\tau_\phi} = \frac{T_s}{\pi T_{ant}}\left[\sqrt{\frac{3}{2}}\right]\left\{\frac{1}{\sqrt{\Delta\nu\Delta t}}\right\}. \quad (29)$$

Here, Δt and N are the single source scan period and sampling frequency over Δt. The root mean square band width is $\nu_{rms} \sim \frac{\Delta\nu}{12\nu}$. Using Equations (27) and (28), the relationship between phase delay and group delay standard deviation could be written as:

$$\frac{\sigma_{\tau_\phi}}{\sigma_{\tau_g}} \sim \frac{\Delta\nu}{12\nu}\sqrt{N}. \quad (30)$$

2.4.3. The geodetic SLR delay observable

Geodetic VLBI and GNSS systems described above operate in the radio wavelengths. The SLR is the only SGT that operates in the optical region with a good ranging accuracy of 1 to 2 cm (Combrinck, 2010). Some of the applications of SLR data include measurement of small scale geodetic station position variations that arise from geophysical processes, the contribution to the development of gravity field models, and the establishment and maintenance of ITRF.

The SLR observable is the Time-of-Flight (ToF) of a laser pulse between the SLR station and target satellite that can be translated to the range to the target satellite; this is often corrected for system delay. The range to the satellite is used to derive other parameters such as EOP, station position, gravity coefficients, etc. The range Equation (31), often used in SLR processing, takes into account atmospheric effects (Δa), Centre-of-Mass (ΔCoM), correction of the satellite, station range bias (ΔR_b) and a relativistic correction (ΔGR).

$$NPR_i = \left\{\frac{c \times NP_{TOF_i}}{10^{12}}\right\}\left\{2 - \Delta a_i + \Delta CoM_i - \Delta R_{b_i} - \Delta GR_i - \Delta\varepsilon_i\right\}^{-1} \quad (31)$$

Here, NPR and $\Delta\varepsilon$ are the normal point range or the observed range and the correction for unknown systematic and random errors respectively. The systematic errors (which are mainly from tropospheric influence) in the SLR observations are contained in Equation (31) and therefore critically influence the absolute ranging accuracy.

SLR operates at optical wavelengths where the atmosphere is dispersive. Therefore, during the processing of SLR data, a correction for the additional delay due to the atmosphere is required. Mendes and Pavlis (2004) reported closed-form expressions in Equation (32) and suitable for calculating the additional zenith SLR ranges due to troposphere;

$$\begin{aligned} \tau^z_{atm} &= \tau^z_h + \tau^z_{nh} \\ \tau^z_h &= 0.002416579 \times \frac{f_h(\lambda)}{f_s(\phi, H)} \times P_s \\ f_h(\lambda) &= \left[k_1^* \frac{k_0 + \lambda^{-2}}{k_0 + \lambda^{-2}} + k_3^* \frac{k_2 + \lambda^{-2}}{k_2 + \lambda^{-2}} \right] \times 9.9995995 \times 10^{-3} \\ f_s(\phi, H) &= 1 - 0.00266 \cos 2\phi - 0.00000028H \end{aligned} \tag{32}$$

$$\begin{aligned} \tau^z_{nh} &= 10^{-4} \frac{e_s}{f_s(\phi, H)} \left[5.316 f_{nh}(\lambda) - 3.759 f_h(\lambda) \right] \\ e_s &= 10^{-2} RH \left[6.1078 \times \left(\frac{7.5t}{237.3 + t} \right)^{10} \right] \\ f_{nh}(\lambda) &= 0.003101 \left(\omega_0 + 3\omega_1 \lambda^{-2} + 5\omega_2 \lambda^{-4} + 7\omega_3 \lambda^{-6} \right) \end{aligned} \tag{33}$$

Here, P_s and $f_h(\lambda)$ are the surface barometric pressure and the hydrostatic dispersion equation respectively. The geodetic latitude and height are denoted by ϕ and H respectively while k_0=238.0185 μm^{-2}, k_2=57.562 μm^{-2}; k_1^* and k_3^* are 19990.975 μm^{-2} and 579.55174 μm^{-2} respectively. The WV pressure (e_s) at the surface is calculated from Relative Humidity (RH). Further, the dispersion component of the non-hydrostatic is denoted by $f_{nh}(\lambda)$ and $\omega_{0,1,2\&3}$ are given as 295.235, 2.6422 μm^2, -0.032380 μm^4 and 0.004028 μm^6.

The atmospheric model of Mendes and Pavlis (2004) adopted by the Analysis Working Group of the International Laser Ranging Service (ILRS) is currently used in the SLR data satellite analysis software package developed at HartRAO (Combrinck, 2010). In order to demonstrate the bias contribution of the troposphere to the range measurements in SLR, the tropospheric zenith delay derived from LAser GEOdynamics Satellite-1 (LAGEOS I) satellite data is plotted as a function of elevation (left panel) and azimuth (right panel) in Figure 2.3. Figure 2.3 illustrates a band of corrections derived from laser wavelengths 423 nm, 532 nm and 846 nm and a range of ground level measurements of relative humidity (%), pressure and temperature. In this example, the elevation angle of 15°; where the delays were more than 8 m, was set as the cut-off angle. As depicted in the right panel, the azimuth angle

only influences the geometry of the satellite orientation as determined by their orbital parameters.

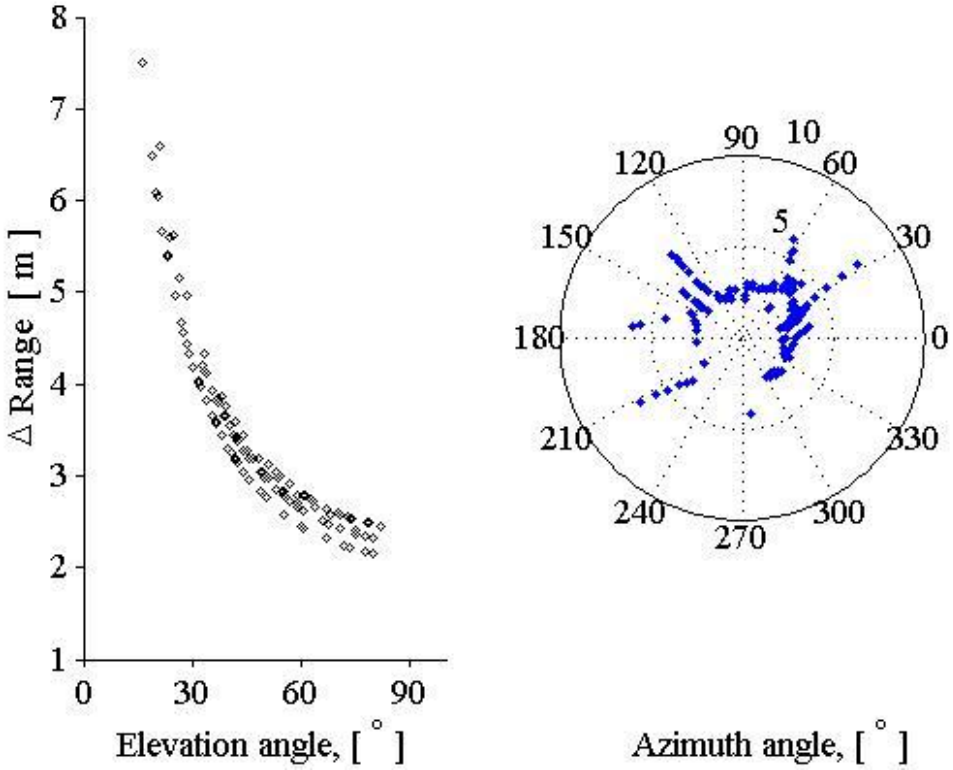

Figure 2.3. Mendes and Pavlis (2004) model of the absolute change in the SLR range measurement due to atmospheric delays

2.5. Derived tropospheric parameters in geodetic analyses

The presence of the dynamic troposphere along the lines of sight from each antenna to the radio source will affect the propagation of the radio signal due to variations of the refractive index of the traversed medium. This geodetic radio signal interaction with troposphere induces a bias in the geometric delay measurement. In addition, the rate of change of phase delay over short time spans will also be affected. Therefore, geodetic parameters such as baseline length and orientation, position of radio signal receivers and transmitters as well as clock offsets; which are estimated from the group delay and phase delay rate observables would be biased due to the fluctuating troposphere.

Tropospheric delay parameters are often estimated in most GNSS and intercontinental VLBI data reduction. In the analysis strategies employed by various geodetic analysis groups, only the spatial-temporal average-troposphere-parameter component (e.g., *ZTD*) is estimated

for each station over the measurement period. As a result, the TD_w^z fluctuations around these averages are the dominant tropospheric errors, which in turn map into biases in geodetic parameters. To improve the accuracies in the geodetic observables, the deviations from the spatial-temporal TD_w^z average ought to be identified and quantified so as to understand the character and effect of WV fluctuations on the estimated geodetic parameters.

Biases due to the modelling of tropospheric delays of microwave signals from satellites or radio sources due to the neutral atmosphere lower the accuracy of the delay observable in GPS and VLBI analyses. The parameterisations often used to compute the total delay, TD is given by Equation (34) according to Davis *et al*., (1985).

$$ZTD(\vartheta,\varepsilon) = \mathrm{Y}_h(\varepsilon)\{TD_h^z\} + \mathrm{Y}_w(\varepsilon)\{TD_w^z\} + \mathrm{Y}_g(\varepsilon)\times\cot(\varepsilon)\{G_n\times\cos(\vartheta) + G_e\times\sin(\vartheta)\} \quad (34)$$

Here Y_g is either the wet or hydrostatic gradient mapping function, G_n and G_e are the North and East gradients respectively. The tropospheric mapping functions and gradients form a set of derived parameters that influence proper modelling of the tropospheric delay. As a result, the mapping functions and gradients affect the accuracy of the geodetic parameters such as station coordinates and velocities computed from space geodetic techniques (Petrov *et al*., 2009). In particular, biases in the station height component are directly related to tropospheric biases using a rule-of-thumb reported in Niell *et al*., (2001) and Boehm and Schuh (2004). Both of these derived parameters aid in characterising the azimuthally symmetric component of the tropospheric delay.

Due to the Earth's surface (and therefore troposphere) asymmetry, the second-order terms of the refractivity, $\boldsymbol{\xi}(z) = \nabla_{\boldsymbol{r}} n(\{\boldsymbol{r}, z\}|_{t,\boldsymbol{r}=0})$, given by Equation (2) often emerge. Though their contribution to the delay observable is arguably minimal, correcting them would certainly play an important role in meeting the goals of 1 mm and 0.1 mm/year accuracy of station positions and velocities respectively. These goals are described in the Global Geodetic Observing System (GGOS) of the International Association of Geodesy (IAG).

Mapping functions

Many mapping functions have been suggested (Niell 2000; Niell 2001; Boehm *et al*., 2006) in many geodetic tropospheric modelling works. Among them, the Niell Mapping Function (NMF) is the most common mapping function used in many geodetic software packages.

Mapping functions $Y_h(\varepsilon)$ and $Y_w(\varepsilon)$ are computed by fitting the coefficients a, b and c in Equation (35) to the standard atmospheres, to in-situ radiosonde measurements or to NWP models (Niell, 2001; Boehm and Schuh, 2007).

$$Y_{h,w}(\varepsilon) = \frac{1+\dfrac{a}{1+\dfrac{b}{1+c}}}{\sin(\varepsilon)+\dfrac{a}{\sin(\varepsilon)+\dfrac{b}{\sin(\varepsilon)+c}}}. \qquad (35)$$

The NMF are provided at five latitude bands and are quasi-symmetric (they have a seasonal dependency with inherent half-year phase shift) with an annual signal. Tropospheric delays are usually separated into hydrostatic and wet components; computed as a product of the delay in the zenith direction and the corresponding hydrostatic and wet mapping functions. Later, mapping functions based on the NWP models, the Isobaric Mapping Functions (IMF) were developed (Niell, 2000). The hydrostatic IMF ($Y_h(\varepsilon)$) uses a height of 200 hPa pressure level while the wet IMF ($Y_w(\varepsilon)$) is based on a coarse ray-trace at 3.3° elevation through the NWP pressure levels.

Boehm and Schuh, (2007) reported on the new Vienna Mapping Function (VMF) which are dependent only on the elevation angle, ε, assuming a symmetric atmosphere around the stations. The values of the *b* and *c* coefficients of Equation (35) are obtained from the IMF and NMF for the hydrostatic and wet components respectively. Updated VMFs (hereafter VMF1) were developed based on new values of *b* and *c* coefficients of the Y_h. The coefficients *c* were derived from ray tracing and fitted to a function of latitude and day of year to remove systematic errors. The systematic station height improvement of up to ~10 mm (which is equivalent to 2 mm improvement in the station height) is obtained on application of VMF1 (Boehm *et al*., 2006; Niell, 2006). Therefore, using VMF1 improves the precision of geodetic parameters. Global Mapping Functions (GMF) that are comparable to the VMF1 have also been developed based on the global ECMWF numerical weather model data sets. The coefficients *b* and *c* are obtained from the spherical harmonics expansion of the VMF1 and then mapped onto a global grid.

Tropospheric gradient

To consider the four dimensional structure of the tropospheric parameters, G_n and G_e are used (Boehm and Schuh, 2007). The gradient components G_n and G_e are mapped by $\mathrm{Y}_g(\vartheta)$ to the slant direction of the observation. In this way, the tropospheric delay and gradient asymmetries are accounted for. Current geodetic analysis approaches model G_n and G_e using the assumption that the atmosphere is driven by stationary processes and two gradients per station which are the North-South and East-West components. These components are used to describe the tilting of the zenith as described by the mapping functions.

The gradient components G_n and G_e are estimated using two methods. These are using priori hydrostatic gradients that have no temporal dependence or by determining the time dependent gradients from a 200 hPa tilted pressure level (see for instance in Boehm and Schuh, (2007) and others therein). Notice that $G_\alpha = \{G_n, G_e\}$ could also be determined from NWP as the vertical integral of $\xi(z)$ weighted with height (hereafter linear horizontal gradients (Boehm and Schuh, 2007; Davis *et al.*, 1993), see Equation (36)). The linear horizontal gradients are computed based on the assumption that the vertical refractivity gradient $\xi(z)$ is constant over some finite distance around a geodetic station.

$$G_\vartheta = 10^{-6} \times \int_{z=0}^{\infty} \xi(z) \times z \times \mathrm{d}z. \qquad (36)$$

Figure 2.4 illustrates that the hydrostatic linear horizontal gradients (North-South and East-West) have very minimal fluctuations of > 0.5 mm. However, the North-South linear horizontal gradient contains fluctuations of ~ 1.5 mm. Similarly, the East-West linear horizontal gradient exhibits amplitude of ~1.0 mm.

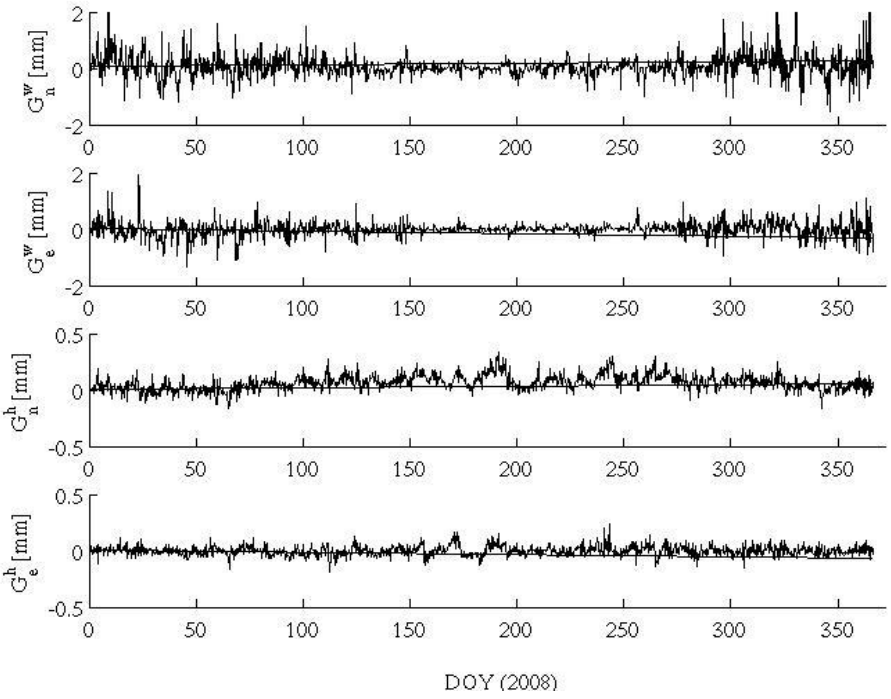

Figure 2.4. The linear horizontal gradients at HartRAO estimated from ECMWF data.

2.6. Recent developments in modelling TD and WV

Ray-tracing has been used to derive mapping functions and also assess the theoretical models used in many geodetic applications. Over a long-term period, the theoretical assessments of these models exhibit reasonable accuracy for both the TD_w^z and TD_h^z assuming accurate surface meteorological measurements. There is however larger scatter in the geodetic path delay and position estimates in the short-term atmospheric fluctuations. A number of factors could be attributed to this scatter. Firstly, though the mm-level accuracy in the delay observable is achievable with precise meteorological sensors, surface measurements are often not representative of the vertical profiles through the whole troposphere; for instance in the boundary layer, the humidity and temperature reveal strong evolution during the day and are stable in the free atmosphere (Stull, 1994). Secondly, empirical meteorological models yield better results with standard atmosphere parameters instead of with measured meteorological parameters (Wang *et al*., 2008). Furthermore, the presence of horizontal gradients in the refractivity field induces the errors when mapping the delay observable onto the line-of-sight. To address this gap, parametric estimation and external correction strategies have been suggested.

Arguably, the meteorological models have not been able to predict the geodetic delay observable with sufficient accuracy. As a result, the parameter estimation approach has been developed which involves a post acquisition processing approach that uses geodetic measurements to constrain the parameters of a meteorological model for determining the geodetic TD observable (Tralli, 1988; Bock and Doerflinger, 2001). In this approach, the path delay is corrected for *a priori* with a meteorological model using *ZTD* and mapping functions. This methodology has inherent theoretical accuracy limitations due to clock offsets, orbit and multipath offsets (e.g., in GNSS) in the delay observable.

Deterministic parameter estimation by least squares adjustment and stochastic estimation based on Kalman filtering (Pacione and Vespe, 2003; Jin and Park, 2005) are the two common strategies used for the parameter estimation of the TD. In deterministic parameter estimation, the tropospheric delay is modelled either as a series of piecewise independent constant terms or correlated parameter which are closely related to a random model. This approach is computationally demanding in large geodetic networks or over long observing periods.

In the Kalman filtering, arbitrary values of the TD with high turn-around time can be estimated. However, this method suffers from increased multipath and tropospheric mis-modelling of the gradients and small-scale in-homogeneities due to low elevation angle observations. Nevertheless, parameter estimation has improved the estimation of the TD observable; e.g., a few parts in $\sim 10^8$; improvement in the VLBI baseline repeatabilities (Tralli, 1988). Recently, improvement in parameter estimation is based on the turbulent atmosphere model (Nilsson and Haas, 2008). In this model, the *ZTD* is simulated through a turbulent atmosphere. The delays are simulated to vary both as a function of direction of observation and time. This approach yields more realistic delays as compared to those simulated from random walk processes.

For high accuracy geodetic positioning, external correction strategies are used. External correction strategies rely on the use of independent techniques for measurements of the TD_h^z and TD_w^z (Bock and Doerflinger, 2001). In this strategy, the wet path delays are retrieved by remote sensing the troposphere, in order to correct the geodetic delay observable *a priori*. The TD_h^z is modelled by empirical meteorological models and evaluated from either surface meteorological measurements or standard atmosphere data. Thereafter, this

component is mapped onto the path of the geodetic radio signal using a mapping function. The wet path delay is derived from integrated water vapour content that is remotely sensed. The WV in the zenith direction is expressed as either the *IWV* or Precipitable Water Vapour (*PWV*) and are formulated in Equations (37):

$$\begin{aligned} IWV &= \int \rho_v \mathrm{d}z \quad \left[\frac{\mathrm{g}}{\mathrm{cm}^2}\right]; \\ PWV &= \frac{IWV}{\rho_l} \quad [\mathrm{cm}]. \end{aligned} \tag{37}$$

Here, ρ_l is the density of liquid air. WV radiometers and infrared solar hygrometers were the early instruments developed mainly for meteorologists and astronomers respectively for remote sensing *PWV*. Ground microwave radiometers have been extensively developed for correcting tropospheric path delay in geodetic analysis (Nilsson *et al.*, 2006). These radiometers are however limited to sensing the wet part of the delay only. Ground-based WVRs measure the sky brightness temperatures at a given frequency, ν, see Equation (38).

$$T_{b,\nu} = T_{b0,\nu} e^{\tau(\nu,\infty)} + \int_S T(s)\alpha(\nu,s) e^{-\tau(\nu,s)} \mathrm{d}s, \tag{38}$$

where S is the straight path from the radiometer to the top of the atmosphere. The frequency, the cosmic background temperature and total attenuation coefficient (due to WV, oxygen and liquid water) are denoted by $\nu, T_{b0,\nu} \sim 2.8\mathrm{K}$ and $\alpha(\nu,s)$ respectively. Whereupon $\tau(\nu,s) = \int_0^s \alpha(\nu,s')\mathrm{d}s'$ is the opacity between location s and the radiometer. The total atmospheric attenuation along path S is denoted by $e^{\tau(\infty,\nu)}$.

In the estimation of WV, an observational frequency where the WV content influences the brightness temperature; in most cases one of the water absorption lines is often used. Further, contribution by other gases and liquid water to the brightness temperature ought to be accounted for to obtain accurate WV. Accounting for the cloud liquid water is not an easy task. Nevertheless, liquid water contribution can be removed by combining two measurements at two frequencies (22.225 GHz and 30 GHz WV line) since the contribution to the absorption coefficient $\alpha \propto \nu^2; \forall r \ll f^{-2}$.

The relation between brightness temperature and WV is fairly non-linear due to the $e^{-\tau(\infty,\nu)}$. Using a linearised brightness temperature $(T_{b,\nu'})$ or introducing a new parameter expressed in terms of the brightness temperature solves the problem of separating water

vapour from brightness temperature. Here, $\tau(\infty,\nu)$ is often derived from Equation (39) using the effective temperature T_e :

$$\tau(\infty,\nu) = -ln\frac{T_e - T_{b,\nu}}{T_e - T_{b0,\nu}}; \tag{39}$$

where $T_e(\nu)\times 1-e^{-\tau(\infty,1)} = \int_S T(s)\alpha(\nu,s)e^{-\tau(\nu,s)}\mathrm{d}s$. An approximate model expressing T_e as a function of surface temperature and elevation angle of the path S can be used thereafter and the amount of WV in the atmosphere would be retrieved from $\tau(\infty,\nu)$. Some of the biases in the estimated WV could therefore be as a result of error propagation from the T_e bias. Similarly, by defining $T_{b,\nu'}$ and using the opacities, the linearised effective temperature T_e could be obtained based on Equations (40):

$$\begin{aligned} T_{b,\nu'} &= T_{b0,\nu}\{1-\tau(\infty,\nu)\} + \int_S T(s)\alpha(\nu,s)e^{-\tau(\nu,s)}\mathrm{d}s, \\ T_{b,\nu'} &= T_{b0,\nu} + \{T_{e'}(\nu) - T_{b0,\nu}\}\tau(\infty,\nu), \\ T_{e'} &= \frac{\int_S T(s)\alpha(\nu,s)e^{-\tau(\nu,s)}\mathrm{d}s}{\tau(\infty,\nu)}. \end{aligned} \tag{40}$$

The relationship between opacity and WV has been reported by Deuber *et al*., (2005). As can be deduced, surface temperature could therefore be used to derive $T_{e'}(\nu)$. Biases in $T_{b,\nu'}$ due to T_e are cancelled out by the errors in $T_{e'}(\nu)$ due to the correlation between $T_{e'}(\nu)$ and $T_{b,\nu'}$ (Jarlemark, 1997). As reported by Jarlemark, (1997), the model that is used to convert the linearised brightness temperature to TD_w^z is given by Equation (41):

$$TD_w^z = c_b\left\{\left[\frac{\nu_2}{\nu_1}\right]^2 T_{b(\nu_1')} - T_{b(\nu_2')} - T_{bO_2})\right\}. \tag{41}$$

Here, T_{bO_2} is the cosmic microwave background radiation and oxygen contribution to the linearised brightness temperatures. Contribution of the cloud liquid water to the linearised brightness temperature is eliminated in the two frequencies by double differencing during the analysis process. The conversion factors c_b are determined in two possible ways. Firstly, simultaneous measurements of brightness temperatures with the radiometer are used together with TD_w^z from either GNSS or radiosondes measurements. Secondly, TD_w^z and brightness temperatures could be simulated from radiosonde data and models of the attenuation coefficients and thereafter use these values to determine c_b. As a result, the brightness

temperature and TD_w^z biases emanate from the systematic errors in radiosonde measurements. These errors are minimal owing to the fact that they cancel out when estimating c_b and T_{bO_2}. However, models that are used to compute $\alpha(\nu, s)$ from radiosonde measurements may introduce systematic biases in the derived algorithm parameters.

Space-borne microwave and Infrared (IR) radiometers have also been used for measuring WV and water profiles; these provide accurate tropospheric vertical WV structures. In addition, Raman LIDARS have also been used for external correction of both the wet and hydrostatic path delay by measuring the inelastic backscatter of WV incident on the laser pulse over large scales (Tarniewicz *et al*., 2002). The Raman LIDARS have been particularly useful in sensing the lower atmosphere where gradients are highly pronounced. In general, active and passive remote sensing techniques (on different platforms) have been used to obtain n-D fields of WV each employing different retrieval methods (e.g., GPS occultation (Ao, 2007) and rapid WV retrieval using Raman and differential absorption LIDAR (Dinoev *et al*., 2006). Some of the global monitoring campaigns of WV have been carried out on space-borne platforms such as Moderate Resolution Imaging Spectroradiometer (MODIS), CHAllenging Minisatellite Payload (CHAMP)- Constellation Observing System for Meteorology, Ionosphere & Climate (COSMIC) and Upper Atmosphere Research Satellite (UARS) (Banks *et al*., 1978) while others are terrestrial based (e.g., SHADOZ). Retrieval of WV over regional scales and short time scales based on dedicated campaigns has recently increased.

All the efforts (this includes the development of measurement systems and retrieval techniques) to monitor tropospheric WV take cognisance of the difficulties in modelling the spatial temporal variability and evolution of WV. In particular, the fluctuations of WV exhibit complex modes, each associated with different (coupled) physical processes that act as a feedback system in Earth's climate system (useful in climate modelling and meteorology) and imposes the accuracy limitation to the geodetic delay observable (vital to the space geodesy research community).

2.7. Analysis strategies for TD/WV in space geodesy

Tropospheric WV rarely attains a permanent hydrostatic equilibrium but continuously changes into or from ice, liquid and WV (which dominates). It is primarily in the vapour phase that water is globally transported into the air. Though WV constitutes 1 part per million

of water on Earth, it transports and redistributes huge amount of moisture and energy (latent heat). The two week life-span of WV in the atmosphere is characterised by vertical and horizontal transports, mixing, condensation, precipitation and evaporation. WV is therefore a key element in climate of the Earth and the hydrological cycle. It is the most variable amongst the major components of the Earth and a vital element in numerical weather prediction as reported in Cucurull, (2000).

Treuhaft and Lanyi, (1987) used a statistical model to quantify the effect of the dynamic wet atmosphere on radio interferometry measurements. The statistical model employed had two major assumptions: a) that a simplified Kolomogorov theory could be used to approximate the spatial structure of the refractivity fluctuations and, b) the temporal fluctuations are caused by spatial patterns driven by wind. In addition, the model assumed that the WV spatial structure and the wind vector were independent of atmosphere height up to some predefined effective height. The structure function given by Equation (42) was then used to describe the spatial characteristics of the wet troposphere.

$$\mathrm{D}_{\mathrm{wv}}(\boldsymbol{r},\Delta\boldsymbol{r}) = \left\langle \left[\mathrm{wv}(\boldsymbol{r}+\Delta\boldsymbol{r}) - \mathrm{wv}(\boldsymbol{r}) \right]^2 \right\rangle \tag{42}$$

In Equation (42), $\mathrm{D}_{\mathrm{wv}}(.) \Leftrightarrow \mathrm{D}_{\mathrm{wv}}(|\Delta\boldsymbol{r}|)$ and therefore can be written in the form

$$\begin{aligned} \mathrm{D}_{\mathrm{wv}}(\Delta\boldsymbol{r}) &= \left\langle \left[\mathrm{wv}(\boldsymbol{r}+\Delta\boldsymbol{r}) - \mathrm{wv}(\boldsymbol{r}) \right]^2 \right\rangle \\ &= \mathrm{C}^2 \Delta\boldsymbol{r}^{\alpha} \end{aligned} \tag{43}$$

where the spatial in-homogeneity of WV is characterised by C. For geodetic VLBI, the general expression for the spatial structure function of the slant delay, $\tau_{\theta,\phi}$ (here, $\{\theta;\phi\}$ are elevation and azimuth respectively) between two VLBI antennas separated by baseline vector $\boldsymbol{b}$ for an atmosphere of effective height, h is given by Equation (44), see Treuhaft and Lanyi, (1987) for further details.

$$\begin{aligned} \mathrm{D}^{\boldsymbol{b}}_{\tau_{\theta,\phi}} &= \left(\frac{1}{\sin\theta} \right)^2 \int_0^h \int_0^h dz dz' \times \left\{ \mathrm{D}_f \left(\left[b^2 + 2(z-z')b\cot\theta\cos\phi + A^2 \right]^{0.5} \right) - \mathrm{D}_f A \right\} \\ A &= \left(\frac{z-z'}{\sin\phi} \right) \end{aligned} \tag{44}$$

where $\mathrm{D}_{wv} \in \mathrm{D}_f$. Additionally, based on the *frozen* spatial structure assumption, the temporal structure function given by Equation (45) could be derived by setting $b = vt$, where v is the wind speed at time t.

$$\mathrm{D}^{\Delta t}_{\tau_{\theta,\,\phi}} = \left\langle \left[\tau_{\theta,\,\phi}\left(\boldsymbol{r},\ t+\Delta t\right) - \tau_{\theta,\,\phi}\left(\boldsymbol{r},\ t\right) \right]^2 \right\rangle \tag{45}$$

Using Equations (44) and (45), the differential fluctuations between two geodetic stations and the per-station dependent temporal fluctuations can be captured based on the equality constraint given by Equation (46).

$$\mathrm{D}^{\Delta t} = \mathrm{D}^{b=v\Delta t} \tag{46}$$

In the above description of WV fluctuations, the Kolmogorov turbulence model is suited for representing the local spectrum of WV fluctuations. Jarlemark *et al.*, (1998) used the structure function given by Equation (47) to describe changes over time of the zenith total delay depending on the time lag, t_τ. The linear dependence between $\mathrm{D}_{\tau^{ztd}}$ and τ was suggested as a special case of random walk process.

$$\mathrm{D}_{\tau^{ztd}}(t_\tau) \overset{\mathrm{det}}{=} \left\langle \left[\tau^{ztd}_{t+t_\tau} - \tau^{ztd}_{t} \right] \right\rangle \tag{47}$$

This representation has several complicating properties such as nonstationarity and the passage of fronts. As a result, it is difficult to reconcile the changes associated with the passage of synoptic scale systems with the apparent observed Kolmogorov behaviour. In Hogg *et al.*, (1981), the determinations of WV from most geodetic techniques are modelled in terms of atmospheric turbulence. As a result, WV could be viewed as a passive tracer that is blown turbulently and the methods used to analyze the WV observations are therefore statistical. The spatial characteristics of WV fluctuations based on the inherent statistical properties could be probed directly using a network of instruments such as GPS network. In addition, Zhang *et al.*, (2003) surveyed satellite and *in situ* observations and reported that the probability distribution functions of the troposphere WV in the tropics was predominantly bimodal due to the spatial-temporal gradients components.

Bevis *et al.*, (1992) reported that the geodetic TD^z_w could be estimated through two approaches. Firstly, a simple estimation method where the geodetic WV could be kept constant for a given time interval and its value obtained as part of the overall least-squares inversion. In this approach, the geodetic WV was assumed constant for a time period ranging from 1 to 24-hours. This deterministic approach implies that the WV is constrained to some value and its space-time derivative kept over some bounds. Secondly, Bevis *et al.*, (1992) used an estimation method where the analysis of geodetic WV utilises the statistical properties of the spatio-temporal variability of geodetic WV. In this approach, the fluctuation of geodetic WV is assumed to be driven by a stochastic process. This implies that process

parameters could be estimated using a Kalman filter or other related class of optimal filters based on the state-space time domain formulation.

It worthy to note that, geodetic WV and/or alternative geodetic troposphere quantities estimated via stochastic filtering would require a specific class of stochastic processes to be selected to capture the inherent fluctuations. This option would be based on *a priori* knowledge of the underlying process. However, the current practice of selecting the representative class of stochastic processes involves the visual inspection of the power spectrum. Documented evidence on the early efforts to model the variability of TD_w^z/WV using a random walk process or a first-order Gauss-Markov process was reported by Herring *et al*., (1990) and other references therein. As reported in the literature, the variation of TD_w^z is space-time constrained to 1-20 mm/year using a specific stochastic process noise model.

Stoev and Elgered, (2005) used realisations of random walk stochastic process to characterise the spatial-temporal variation of TD using a ground network of GPS receivers in Europe with a time interval of 1 to 3-hours. In their study, using monthly TD data between 1997 and 1998, a standard deviation < 50 mm without a clearly visible seasonal component was reported. To capture the excursions present in the data, temporal structure functions were computed. It is vital to remark that using random walk processes to model the nature of TD above a given geodetic site is a first order approximation that is useful for geodetic inter-technique comparisons. The absorbing barriers model (see Grimmett and Stirzaker, 2001) was introduced as an extension to the random walk paradigm to capture local TD fluctuations associated with the passage of atmospheric fronts, thunderstorms and other local weather systems.

Recently, the work reported by Boehm *et al*., (2007) used the turbulence strategy reported by Nilsson *et al*., (2007) to estimate TD_w^z via simulations. In the simulations, the asymmetry in TD_w^z variations is taken into account by the covariance information between all observations at each station. As a result, a time series of equivalent TD_w^z are derived that includes the elevation and azimuth dependency as opposed to the random walk or Gauss-Markov simulations. The turbulent framework reported hinged on the following factors; a) initial zenith wet delay ($TD_{w,0}^z$), b) the wind speed and direction, $\{v;\theta\}$, c) the structure constant parameter (C_n), d) the troposphere effective height (h) and e) the height increment (Δh). A typical simulation scenario of TD_w^z and clock biases based at HartRAO, as described

by Boehm *et al*., (2007) is given in Figure 2.5. The parameterisations in the simulation of the 24-hour equivalent TD_w^z variability illustrated in Figure 2.5 were as follows: a) $\tau_0^{zwd} = 128$, b) $C_n = 2.4\times10^{-7}\ \mathrm{m}^{-1/3}$, c) $\{h;\Delta h\} = \{1000;100\}$ m and d) $\{v;\ \theta\} = \{12\ \mathrm{ms}^{-1};\ 180^{\circ}\}$. In the simulations of integrated and random walk clock excitations, the Allan standard deviation, $\sigma_{asd}\ (\sim 2\times10^{-15})$ set at 50 minutes was used. The Allan variance (defined as the average fractional deviation stability) is used here to characterise the fluctuations of the noise contribution from the geodetic system clock offsets.

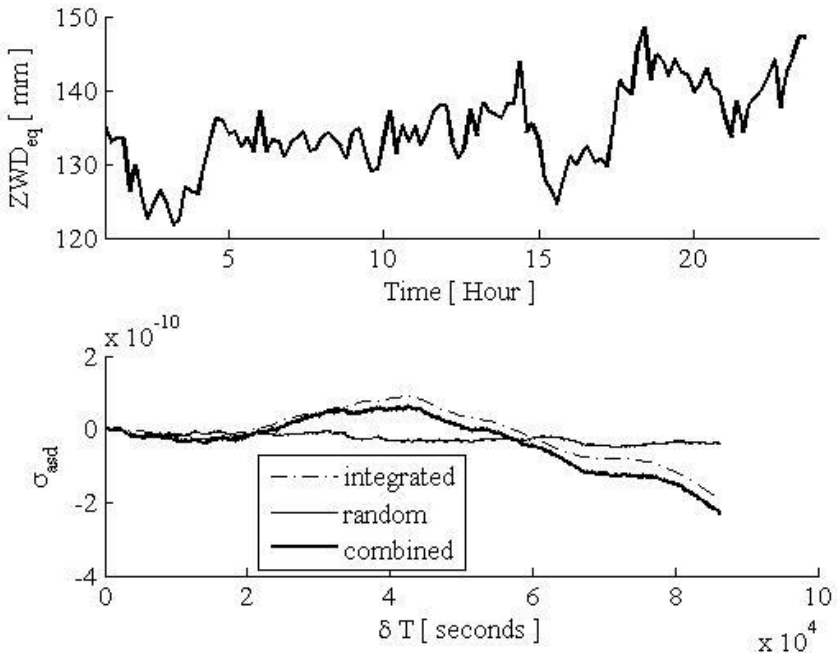

Figure 2.5. Simulation of equivalent zenith wet delay (top panel) and clock (bottom) variability using the random walk process.

A stochastic mathematical model of the combined solution of TD was formulated by Heinkelmann *et al*., (2007). This methodology was based on the assumption of zero correlation between the solutions of the individual VLBI analysis centres (ACs); implying that the solutions of the individual ACs are independent. However, the presence of inhomogeneous structure of the standard deviations of the ACs imply that weighting of individual observations among the ACs would be unrealistic and therefore the standard deviations ought to be ignored in further analysis. As an alternative to the stochastic model, a functional approach was proposed. In this functional form, the trend and seasonal components of TD_w^z time series is incorporated into the model. The combined IVS time series

of TD given by Equation (48) are determined for each station separately by a weighting mean of hourly delay values of the ACs using the relative weighting factors derived from the variance component estimation given by Equation (48)

$$w_k = \frac{N \times \frac{1}{\hat{\sigma}_k^2}}{\sum_{k=1}^{N} \frac{1}{\sigma_k^2}} \qquad (48)$$

$$ZTD_c = \sum_{k=1}^{N} \frac{w_k}{\sum w} \times ZTD_k \Big|_{N>2} .$$

Here, k and N are the individual AC and total number of ACs respectively, which contribute troposphere delay solutions, to the combined solution. The weighting scheme follows the rigorous Bounded Influence by Standardised Residual (BIBER) estimator reported by Wicki, (2001). The BIBER estimator reported (Heinkelmann *et al.*, 2007) neglects the standard deviation computed by each AC.

3. Data and methodology

I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot express it in numbers, your knowledge is of meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever matter may be.
- Lord Kelvin, W. T., 1866.

Chapter 3 outlines the research method, data used, their sources and method of analysis. The schematic framework of the research is discussed. The Chapter focuses on the geodetic and model simulated data sub-sampled at the geodetic VLBI stations and the NWP grid cells over the SHADOZ network. In addition, the non-parametric techniques used in data analyses are also described.

3.1 Introduction

In this thesis, investigation of tropospheric delays due to geodetic WV and WV fluctuations over the Southern Africa region by geodetic and simulation data are reported. The stochastic behaviour of local WV time series is investigated by use of Auto-Regressive Moving Average (ARMA). In addition, the multi-scale variability and scaling behaviour of WV is studied in the time-frequency domain (wavelets) as well as using a data adaptive (noise assisted) methodology (i.e. EMD methods). All these methods take into account the inherent nonlinearity and nonstationary characteristics based on the local time scales of the data. This chapter describes the sources and different types of data that were used in the present research work. In addition, methods used to pre-process these data records are briefly described. In the analysis section, a general and brief description of the mutual information concept, often used in information theory, is discussed and its linkage to the correlation paradigm is presented. Further, for the purpose of studying the scaling behaviour in the WV fluctuations a general description of the wavelet transform, DFA and HHT techniques are also presented. Specific applications of each of these methods, which have been presented in

various international conferences, peer reviewed and published are presented in the subsequent chapters.

3.2. Research methodology

In order to investigate the nature of WV fluctuations over Southern Africa, this research was undertaken from three important viewpoints as depicted in Figure 3.1. Firstly, geodetic data (VLBI and GPS ZTD and delay gradients) at the HartRAO fiducial geodetic station were used to compute a long time series of geodetic WV. Troposphere gradients, VMF and WV derived from ECMWF data were used to investigate the nature of stochastic processes in the time series. Thereafter, the parameters of the ARMA model that characterise the stationarity of WV were adaptively estimated from geodetic tropospheric delay time series.

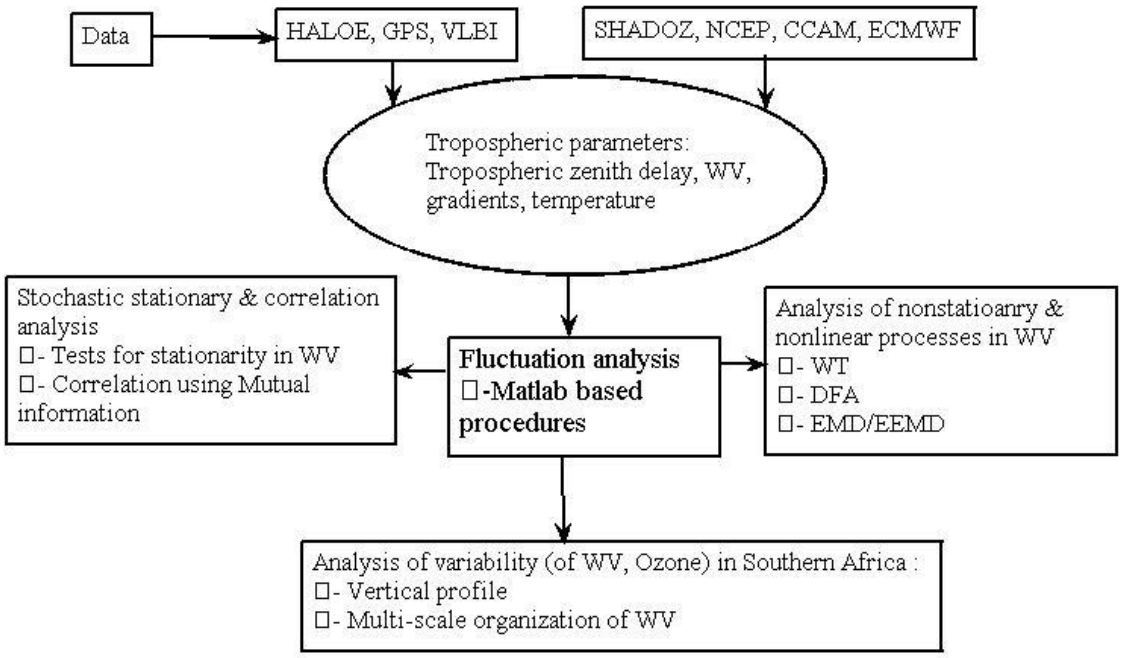

Figure 3.1. Flow diagramme of the research framework.

Secondly, the multi-scale structure of WV in the low- and mid-tropical Africa is investigated by use of *in situ* radiosonde observations of the SHADOZ station network comprising of Ascension, Irene (South Africa), Reunion (Reunion) and Nairobi (Kenya) and the numerical model simulations for the period from 1998 to 2006. Thereafter, the vertical model of

tropospheric WV over southern Africa is developed by use of radiosonde and NCEP/NCAR reanalysis data. Finally, the scaling properties of WV fluctuations were identified and measured using wavelet-based and DFA (DFA is described by Chen *et al*., 2002) approaches. Wavelet analyses could be viewed as a microscope and telescope. This is due to the high-frequency/low-time resolution in the low-frequency part and low-frequency/high-time resolution in high-frequency part characteristic of this technique. In addition, it uses a predefined wavelet basis; the mother wavelet. This implies that the wavelet analysis results are limited by the mother wavelet. Nevertheless, this nonparametric estimate of scaling behaviour is computationally efficient (e.g., the discrete wavelet transformation) and is robust because it has low variance and negligible bias. Further, the Hilbert-Huang transforms, HHT (Huang *et al*., (1998)) and the EEMD reported by Zhaohua and Huang, (2009) were used to adaptively analyse the nonlinear and nonstationary processes in WV. The HHT is built on the assumption that any data set consists of different, simple and intrinsic modes of oscillations (ranging from low to high frequency) that are derived from the observations objectively (adaptively). As a result, this methodology is suited for presenting the WV distributions (derived from observations) in time-energy-frequency distributions.

3.3. Data

Troposphere parameters (N, WV, ZTD and delay gradients) that were analysed and presented in this thesis were derived from geodetic, radiosonde, other space-borne measurements and NWP model simulations. For clarity, Figure 3.1 depicts the data, processing and analysis methods that have been used to study the fluctuations of troposphere parameters (ZTD and WV).

Geodetic data

The central theme in geodetic processing is to derive the delay observable which has position information of the geodetic receiver and the source of the radio signal. For geodetic VLBI, the delay observable also has the structure information of the radio source. To derive this information with high accuracy, the troposphere contribution to the delay observable must be removed. This thesis addresses *an inverse problem*:

a) The results of actual geodetic observations are used to assess and compute the inherent properties of the fluctuating troposphere parameters that characterise the tropospheric structure and dynamics.

b) It also addresses the effect of the atmosphere on the geodetic observations, through the use of actual observations and numerical simulations of meteorological parameters.

The IVS for Geodesy and Astronomy provides tropospheric products such as zenith total delay and zenith wet delay, $\tau_{ztd/zwd}^{atm}$ for all IVS-R1 and IVS-R4 sessions since January 2002 (Schuh and Boehm 2003). All available VLBI observations are processed by the IVS ACs with three main analysis software packages, OCCAM (maintained by the Institute of Applied Astronomy, Russia), CALC/SOLVE (maintained by NASA Goddard Space Flight Centre, GSFC) and Steel Breeze (maintained by Main Astronomical Observatory-MAO, the National Academy of Sciences, Ukraine). The corresponding products such as $\tau_{ztd/zwd}^{atm}$ are transferred to the IGG (Institute of Geodesy and Geophysics, Vienna University of Technology, Austria) for comparison and combination. The motivation for combining the tropospheric parameters is to average out the systematic differences in $\tau_{ztd/zwd}^{atm}$ arising from the use of the different analysis software packages using different parameterisation and models, such as the thresholds of outlier detection, or elevation cut-off angles. For further details, please refer to Schuh and Boehm (2003) and Heinkelmann *et al*., (2007). The combined long time-series of $\tau_{ztd/zwd}^{atm}$ is determined from all geodetic VLBI sessions and can conveniently be obtained from all IVS data centres (see, ftp://cddis.gsfc.nasa.gov/vlbi/ivsproducts/trop).

The geodetic delay and other derived parameters such as troposphere gradients, WV, mean atmospheric temperature and VMF derived from the ECMWF were obtained from IGG. The data is archived at http://mars.hg.tuwien.ac.at. The archive consists of files which contain a record of the global geodetic VLBI, GPS and DORIS station names. The temporal resolution for troposphere parameters archived is six hours corresponding to the NWP model simulations. Since our concern is to assess the local and regional fluctuations of troposphere WV, we study WV (and those parameters that influence WV) variability over a geodetic station; HartRAO-South Africa (see Figure 3.2).

Figure 3.2. Fiducial geodetic site: Hartebeesthoek Radio Astronomy Observatory.

In addition to geodetic VLBI data, troposphere delays derived from GNSS observations were also analysed. The International GNSS Service (IGS) which was formally established in 1993 by the IAG, began routine operations in January 1994 (Beutler *et al.*, 1999). The IGS operates a global network of more than 350 permanent GPS tracking stations, each equipped with a GPS receiver that provides raw GPS tracking data in a Receiver Independent Exchange (RINEX) data format. All available near-real time global IGS observation data are transmitted to the global IGS data centres from where a combined tropospheric product (e.g., weekly files containing $\tau_{ztd/zwd}^{atm}$ in a 2-h time interval from the IGS tracking stations and archived for instance at ftp://cddis.gsfc.nasa.gov/gps/products/trop_new) can be downloaded. Other data sets used in VLBI and GPS processing are presented in Table 3.1.

Table 3.1. Data products and sources used in VLBI analysis.

Geodetic data	Description	Source
Ephemerides & Earth orientation parameters	• Current series: EOP 0504 (Standish, 1990)	http://hpiers.obspm.fr/eop-c
Atmospheric loading coefficients	• Tidal and non-tidal atmospheric tides (Petrov and Boy, 2004)	- http://www.ecgs.lu/atm (tidal S_1/S_2 sine and cosine components)
Thermal deformation	• Thermal deformation of VLBI antennas (Wresnik *et al*., 2005)	- http://mars.hg.tuwien.ac.at~vlbi/thermal
Ocean loading	• Ocean tide model GOT00.2 based on the global ocean tide model from TOPEX/ Poseidon Altimetry/ GOT99.2 (Ray, 1999)	Source, H.G. Scherneck

Numerical prediction model simulations, satellite and Ozonesonde data

NWP model simulated pressure, temperature, and specific and relative humidity fields derived from the reanalysis project of the NCEP / NCAR (Kalnay *et al*., 1996) in the United States (US)); which is the NCEP/NCAR data set were also used to compute WV. These data sets were obtained from NASA's website at http://www.cdc.noaa.gov. In addition, surface temperature measurements based on the automatic weather stations over the HCB were provided by the South Africa Weather Service (SAWS).

The vertical profile of WV model for southern Africa reported in this thesis was derived from Ozonesonde data sets based on the SHADOZ network (i.e. Nairobi - Kenya, Malindi - Kenya, Irene – South Africa, Reunion - Reunion and Ascension–Ascension Islands stations); the data is archived at http://croc.gsfc.nasa.gov/shadoz/. Refer to Thompson *et al*., (2003) for a detailed and technical description of the data sets. In addition, data from the HALOE on board the upper atmosphere research satellite were from HALOE server.

3.4. Data pre-processing

Both Geodetic data and those derived from model simulations utilised in the present research exhibit inherent systematic biases. For instance, geodetic data sets are often acquired by use of geodetic instruments that have measurement biases. In addition, bias contribution from external environment such as changing weather systems is a known problem among the space geodesy community. Furthermore, the difference in the analysis schemes employed by different software packages is known to bias geodetic parameters such as geodetic WV. On the other hand, model simulated data sets are often constrained by the inherent parameterization schemes.

While the analysis strategies employed in this research are robust and some-worth data adaptive, the data sets considered were pre-processed before analysis. A static transformation function was applied to most of the data sets analysed in the thesis in order to ensure a symmetric frequency distribution and also to obtain a more Gaussian-like shape. This transformation is particularly important in assessing the stationary processes in the geodetic data. In addition, the periodic cycles and polynomial trends were estimated and subtracted from the original data sets as a method of disintegrating the time series into inherent components from which the stochastic characteristics of each component would be deduced. In particular, the data sets were transformed using Box-Cox transformation (Box and Cox, 1964) while second order statistics were adaptively subtracted from the data using the Wessel *et al*, (2000) adaptive filtering methodology.

3.4.1 Box-Cox transformation

The Box-Cox transformation is a non-linear static transformation function which converts non-normal datasets to a set of data which approximates a Gaussian distribution. Though the Box-Cox transformation is a family of power transformation, in this thesis, a geodetic data record $\{Y_j\}_{j=1,2,\cdots,N}$ for which $Y_j > 0 \quad \forall j \in \{1,\cdots,N\}$ is Box-Cox transformed by Equation(49),

$$Y'(\lambda) = \begin{cases} \dfrac{\left(Y^{\lambda} - 1\right)}{\lambda}, & \lambda \neq 0 \\ \log(Y), & \lambda = 0 \end{cases} \tag{49}$$

The power parameter, λ is often selected based on the maximising logarithm of the likelihood function given by Equation (50),

$$f(Y,\lambda) = \frac{N}{2}\ln\left[\sum_{j=1}^{N}\frac{(Y_j(\lambda)-\bar{Y}(\lambda))^2}{N}\right] + (\lambda-1)\sum_{j=1}^{N}\ln(Y_j) \tag{50}$$

where geometric mean, $\bar{Y}(\lambda) = \frac{1}{N}\sum_{j=1}^{N} Y_j(\lambda)$.

3.4.2. Estimation of periodic cycles and adaptive filtering

Geodetic tropospheric data exhibit periodic components. A periodic component can be obtained by estimating the mean and the variance for a particular time span over some reference epoch, T_{ref} in the cycle. If the geodetic data record has p cycles of length N, according to Hipel and Mcleod (1994), the mean can be calculated from Equation (51);

$$\bar{\mu}_{T_{ref}} = \frac{1}{p}\sum_{k=1}^{p} Y_{T_{ref},j} \quad T_{ref} = 1,2,\cdots,N \tag{51}$$

where T_{ref} is the reference time epoch and k indexes the successful cycles. Note that $N=365$ for daily measurements over one year cycle. The variance is given by Equation (52);

$$\sigma^2_{T_{ref}} = \frac{1}{p-1}\sum_{k=1}^{p}\left(Y_{T_{ref},k} - \bar{\mu}_{T_{ref}}\right)^2, \quad T_{ref} = 1,2,\cdots,N \tag{52}$$

The normalised anomalies time series can be calculated by Equation (53)

$$Y'_{T_{ref},k} = \frac{Y_{T_{ref},k} - \bar{Y}_{T_{ref}}}{\sigma_{T_{ref}}}. \tag{53}$$

The main objective in the current analysis of geodetic data is to investigate the characteristics of the fluctuations of the WV and the nature of the underlying processes that drive this variability. However, some amount of noise is always expected to be embedded in the geodetic data records. Analysis of such data records in the presence of noise often fail to give the required accurate spatial-temporal structures of interest to the space geodesy community.

It is therefore necessary to exclude the artifacts, systematic and manual errors by use of a robust *cleaning* tool. A robust platform for denoising the data which is used in this thesis is based on Wessel *et al.*, (2000). The advantage of using this methodology stems from the fact that the filter coefficients are spontaneously adapted in the event of the sudden changes in the time series.

Apart from the ordinary gap filling of data with missing data records, adaptive filtering proceeds via two important steps; the adaptive filtering and adaptive control procedures. In the adaptive filtering procedure, the adaptive second order properties such as the mean, μ_a^n and standard deviation, σ_a^n given in Equations (54) are computed from a reconstructed time series Y^{*k} (wherein the obvious errors such as gaps due to missing values) have been removed or filled.

$$\begin{aligned} \mu_a^k &= \mu_a(k-1) - q(\mu_a^{k-1} - Y^{*k} - 1) \qquad \text{k=1, 2, ... n} \\ \sigma_a^k &= \sqrt{\mu_a^k - \lambda_a^k} \end{aligned} \tag{54}$$

where $q \in \{0,1\}$ is the controlling coefficient and the adaptive second moment $\lambda_a^k = \lambda_a^{k-1} - q\left(\lambda_a^{k-1} - Y^{*k-1} \times Y^{*k-1}\right)$. Outliers are often identified using a filter constraint imposed on the raw data. The data point is an outlier if;

$$\begin{aligned} \left|Y^k - Y^{k-1}\right| &> \frac{\gamma Y^{k-1}}{100} + q_f \times \bar{\sigma}_a \\ \left|Y^k - Y^v\right| &> \frac{\gamma Y^v}{100} + q_f \times \bar{\sigma}_a \end{aligned} \tag{55}$$

where γ is the proportionality constant (~ 0.1), $q_f \times \bar{\sigma}_a$ is the generalised 3σ sigma rule. The last valid observation is denoted by Y^v while $\bar{\sigma}_a$ is the average of σ_a^n. A random number generated from Equation (56) is used for gap filling (replace all those values recognised as outliers).

$$\hat{\lambda} = \left\{ \mu_a^k - \frac{1}{2}\sigma_a^k, \mu_a^k + \frac{1}{2}\sigma_a^k \right\} \tag{56}$$

This gap filling procedure is used to avoid the false decreased variability that is often noticed after the adaptive filtering phase. In the adaptive control procedure, a percentage time series is build from the adaptively filtered time series. Thereafter, a new adaptive mean and standard deviation of the reconstructed percentage time series, $\left\{Y_k^{\%}\right\}$; $\forall k = 1, 2, \cdots, n$ of the adaptive filter and binomial-filtered series are calculated. Then, a constraint is imposed on the binomial-filtered series such that an outlier data point is detected using the following inequality:

$$\left|Y_a^{\%} - \mu_a^k\right| > q_{f_1} \times \sigma_a^k + \sigma_0. \tag{57}$$

Here, q_{f_1} and σ_0 are the filter coefficient and parameter that accounts for basic variability respectively. Equation (57) is introduced to dampen filtering errors due to minimal variability in the geodetic time series.

3.5. Data analysis strategies

The only links we have with the unexplained reality are the data and therefore the only way of investigating the underlying processes of any given phenomena is through data analysis, refer for example Lin *et al.*, (2009). Geodetic tropospheric parameter time series, such as tropospheric delay (and delay gradients), WV, tropospheric mean temperature and pressure consist of complex components which are manifestations of non-linear processes. The dynamics of the troposphere often evolve as a complex system with various spatio-temporal correlation scales that are either discrete (e.g., precipitation) or continuous (e.g., tele-connection patterns). These correlations often embed different components with, perhaps a wealth of unique statistical information about the interactions among the inherent tropospheric constituents: the geophysical signals. Traditional methods of determining characteristic time-frequency scales (e.g., Fourier and Principal Component Analysis) for each component involve decomposing the time series into component basis functions that satisfy two conditions; completeness of the basis and orthogonality. In terms of Fourier analysis, a given time series '$Y(t)$' is decomposed into global sinusoidal components of fixed amplitude a_j given by Equations (58) ,

$$Y(t) = \sum_{j=0}^{n} a_j e^{i\omega_j t};$$
$$a_j = \frac{1}{2\pi} \int_t Y(t) e^{-i\omega_j t} dt \qquad (58)$$

Equations (58) imply that the spectral amplitudes, a_j represent the energy contributed by a sinusoidal basis with frequency ω_j that spans the whole time series. The Fourier representation is most useful when the underlying geophysical process which causes variability in the time series is linear and therefore the superposition of the sinusoidal signals would make physical sense. As alluded to earlier, a_j remains time invariant thus $Y(t)$ is fairly constant. However, most of the geodetic time series do not meet this stationarity condition (they are non-uniform, non-linear and nonstationary). This would mean that the time series exhibits a broad spectral energy. In order to reconstruct the time series, global (e.g., harmonic) sinusoids are often required. Fourier transforms do not provide local features and therefore not suited for local description of the embedded dynamical structure of the observations.

3.5.1. Detrended fluctuation analysis

In this thesis, the presence or absence of random walk-type behaviour in troposphere WV is assessed using the DFA. The DFA methodology has been proven useful in revealing the extent of long-range correlations in diverse time series (e.g., Talkner and Weber, 2000; Király and Jánosi, 2005; Qian *et al*., 2008; Peña *et al*., 2009; Rybski and Bunde, 2009 and Varotsos *et al*., 2009). The DFA method is used to analyse WV fluctuations and also provide characteristics of the correlated stochastic components as well as effectively filtering out slow trends. The DFA approach handles nonstationary trends and also amplifies the intrinsic correlation structure of WV fluctuations of different time scales for analysis. The most important advantage of DFA over conventional methods such as autocorrelation and spectral analysis is that it has provision for the detection of intrinsic self-similarity that is embedded in the nonstationary WV. In the following, the general procedure of the DFA methodology is presented.

- Step 1: A fluctuating WV time series $Y_t \quad \forall t = 1, 2 \cdots T$ is integrated to determine the profile:

$$\chi_t = \sum_{t=1}^{i} (Y_t - \bar{Y}) \tag{59}$$

In Equation (59);

$$\bar{Y} = \frac{1}{T} \sum_{i=1}^{T} Y_t \tag{60}$$

- Step 2: χ_t is segmented into $K = \text{int}\left|\tau^{-1} T\right|$ non-overlapping time intervals, η_k of equal size τ where $k = 1, \cdots, K$. The above procedure is repeated from the other side of the series (from t=T, T-1, ..., T-(T-1)) in order to include all parts of the profile. This yields $2K$ segments.
- Step 3: For each of $2K$ segments, a local trend is calculated and a polynomial function of the form χ'_k is determined by the least-squares fit to the series. Thereafter, the variance is calculated using Equation (61)

$$F^2(\tau, k) = \frac{1}{\tau} \sum_{i=1}^{\tau} \left\{ \chi\left[(k-1) + i \right] - \chi'_k(i) \right\}^2; \tag{61}$$

for each segment, $k = 1, \cdots, K$ and

$$F^2(\tau, k) = \frac{1}{\tau} \sum_{i=1}^{\tau} \left\{ \chi\left[T - (k-K)\tau + i \right] - \chi'_k(i) \right\}^2, \tag{62}$$

for $k = K+1, \cdots, 2K$.

- Step 4: An m[th] order fluctuation is calculated by averaging each scale over all segments using Equation (63)

$$F_m(\tau) = \left\{ \frac{1}{2K} \sum_{k=1}^{2K} \left\{ F^2(\tau, k) \right\}^{m/2} \right\}^{\frac{1}{m}}. \tag{63}$$

In this report, m=2. Steps 2, 3 and 4 are repeated for several time scales in order to assess the dependence of $F(:)$ on the time scales.

- Step 5: The scaling behaviour of the WV fluctuations is then determined by analysing the log-log plots of $F(:)$ versus τ. Note that a power law relationship between F_m^τ and τ indicates the scaling with an exponent υ given by;

$$F_m^\tau \sim \tau^\upsilon. \tag{64}$$

Here, ν is a self-similarity parameter that represents the long-range power-law correlation in the data record. It is worth noting that if WV exhibits self-similar behaviour with $\nu > 0$ the fluctuations would grow with the window size in a power law way. This implies that the fluctuations on large observation windows exponentially grow faster than those with small windows. This would mean that WV fluctuations are unbounded. If $\upsilon = 0.5,$ the fluctuations are uncorrelated and are expected to be driven by processes that are a random walk and WV exhibit a Gaussian distribution; however, if $\upsilon < 0.5,$ the fluctuations are anti-correlated and for $\upsilon > 0.5,$ the signal is correlated. Processes exhibiting this behaviour have a power-law autocorrelation function expressed as;

$$C_\gamma = \left(Y_t Y_{t+\gamma}\right) \sim \gamma^{-\alpha}. \tag{65}$$

Here, $0 < \alpha < 1$. According to Talker and Weber, (2000), the relationship between the correlation exponents could be given by;

$$\frac{\alpha}{2} = 1 - \gamma \tag{66}.$$

3.5.2. Wavelet transform

The WT has been introduced and developed to study a large class of phenomena such as image processing, data compression, chaos, fractals, etc (Whitcher, 1998). Mallat, (1989) proposed a concept of multi-resolution analysis for constructing an orthonormal wavelet basis and further illustrated the wavelet multiresolution characteristic from the space aspect. As a

result, the works demonstrated the functions of wavelet theory in the frequency analysis of various data signals. Though recently developed, wavelets analyses techniques provide a powerful and insightful representation of the structure in data appropriate to both linear and nonlinear systems. The basic functions of the WT are related to the property of spatial-temporal-frequency localisation, contrary to what happens with trigonometric functions. The WT works as a mathematical microscope on a specific part of a signal to extract local structures and singularities. This makes the wavelets ideal for handling non-stationary and transient signals, as well as fractal-type structures.

Let $L^2(\mathbb{R})$ denote the two dimensional space of all square integral functions, $\varphi(t)$ with finite energy. If $\varphi(\mathrm{t}) \in L^2(\mathbb{R})$ is a fixed function, then the $\varphi(t)$ is said to be a wavelet if and only if its Fourier Transform (FT), $\hat{\varphi}(w)$ satisfies the permitted admissibility condition (also called complete reconstruction condition) given by Equation (67),

$$C_\varphi = \int_0^\infty \frac{\left|\hat{\varphi}(\omega)\right|^2}{|\omega|} d\omega, \\ < \infty. \tag{67}$$

Here, $\varphi(t)$ is the mother or basic wavelet. Equation (67) implies that the wavelet value is centred on the mean (see Equation(68))

$$\int_{-\infty}^{\infty} \varphi(t)dt = \hat{\varphi}(t), \\ = 0, \tag{68}$$

and therefore is oscillatory (some sort of a wave) as described by Daubechies, (1992), Mallat, (1999) and Qian, (2002).

If the flex (also called the dilation) and translation transform is applied to the mother wavelet $\varphi(t)$, then $\varphi(t)$ can be decomposed into some wavelet series $\phi_{(a,b)}$ defined such that;

$$\varphi_{(a,b)}(t) = \mid a \mid^{-0.5} \varphi\left(\frac{t-b}{a}\right). \tag{69}$$

Here, $b \in \mathbb{R}$ is the translation parameter and $a \in \mathbb{R}^{+}(a \neq 0)$ is the flex/dilation or scale parameter (this is the scaling in frequency range). The normalisation factor $a^{0.5}$ ensures that $\phi_{(a,b)}$ has the same energy along all the scales. Given that tropospheric WV data sets are represented by finite number of observations or measurements, the orthogonal (discrete) wavelets associated with orthonormal bases of $L^2(\mathbb{R})$ are often appropriately used for their analysis. Therefore, WT is performed only on a discrete grid of the tropospheric WV over some dilation and translation. This implies that a and b parameters take only integral values, where in general terms, the expansion of the WV time series, $Y(t)$ can be expressed by Equation (70)

$$\mathrm{Y(t)}=\sum_{n}\sum_{m} Y_n^m \varphi_{m,n}(t) \tag{70}$$

From Equation (70) the orthonormal wavelet basis functions are related according to;

$$\varphi_{m,n}(t) = 2^{\frac{m}{2}} \varphi\left(2^m t - n\right), \tag{71}$$

where m and n are the dilation and translation indices respectively. Equation (71) is derived from equation (69) when $a = 2^{-m}$ and $b = 2^m \times n$. At any particular wavelet level m, the contribution of a time series could be given by Equation (72),

$$Y_m(t) = \sum_{n} Y_n^m \varphi_{m,n}(t) \tag{72}$$

The significance of Equation (72) is that it provides temporal behaviour of the time series within different scales as well their contribution to the total energy WV time series. As discussed in Qian, (2002), the wavelet function $\varphi(t)$ is related to scaling function $\phi(t)$ and scaling coefficients a_n^m.

For a given wavelet basis to be as representative as possible, some degree of regularity is often desired. This condition is met by wavelets that exhibit n vanishing moments;

$$\int_{-\infty}^{\infty} Y^k \varphi_n(y)dx = 0 \;;\; \forall k = 0,1,\cdots,n-1, \quad \text{and} \tag{73}$$

$$\int_{-\infty}^{\infty} Y^k \varphi_n(y)dx \neq 0; \;\; \forall k = n. \tag{74}$$

Equations (73) and (74) imply that a wavelet with n vanishing moments is orthogonal to polynomials up to order $n-1$. Note that the admissibility condition imposes the condition that a wavelet ought to have at least one vanishing moment. In general, a wavelet transform of $Y(t)$ with a wavelet $\varphi_n(y)$ and n vanishing moments is simply a smoothed version of the n^{th} derivative of $Y(t)$ on various scales.

Here, we have employed the Haar wavelet as the analysing signal where a set of non-continuous (and therefore non-differentiable) functions whose mother wavelet takes the form of:

$$\varphi(t) = \begin{cases} 1 & 0 \le t \le 0.5 \\ -1 & 0.5 \le t \le 1 \\ 0 & otherwise \end{cases} \tag{75}$$

with a scaling function $\phi(t)$ described as,

$$\phi(t) = \begin{cases} 1 & 0 \le t \le 1 \\ 0 & otherwise \end{cases} \tag{76}.$$

Important features of the Haar wavelet include:

a) the basis functions are often expressed by linear combinations: $\phi(2^0 t), \phi(2^1 t), \phi(2^2 t), \ldots, \phi(2^k t), \ldots$ and their shifted functions,

b) the constant functions, $\varphi(2^0 t), \varphi(2^1 t), \varphi(2^2 t), \ldots, \varphi(2^k t), \ldots$ and their shifted function form are used for approximations,

c) they exhibit the orthogonality, $\int_{-\infty}^{\infty} 2^m \varphi(2^m t - n)\varphi(2^{m_1} t - n_1)\, dt = \delta_{m,m_1}\delta_{n,n_1}$, where $\delta_{i,j}$ is the Kronecker delta and

d) the wavelet and scaling functions are related as shown in equations (77) ;

$$\begin{aligned} \varphi(t) &= \varphi(2t) + \varphi(2t-1), \\ \phi(t) &= \varphi(2t) - \varphi(2t-1). \end{aligned} \tag{77}$$

The Haar wavelet transform cross-multiplies a function against the Haar mother wavelet with various shifts and stretches which are derived from the Haar matrix. See Equation (78) for a 2 by 2 Haar matrix sample;

$$H_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{78}$$

The Haar WT therefore aids the sampling processing in which rows of the transform matrix act as samples of the finest resolution (this is the basis for multi-resolution analysis in wavelet methodology).

In the analysis of scaling behaviour in WV, nonparametric estimators (e.g. DFA described above) were considered in this thesis. These estimators are based on fitting a power-law on the n^{th} order moment of the data values themselves or of their variations as a function of some scale/lag parameter. The approach has however two presuppositions for scaling processes. For long memory processes:-

a) a statistically sufficient evidence that the relevant points on the curve do indeed represent a straight is required, and

b) that the line's slope is such that $0.5 < H < 1$, where H is the Hurst parameter.

If WV data (of length N) is assumed to be stationary, then a simple sample estimator of the mean, μ_Y of a second order process Y_t is a reasonable choice. However, as $N \to \infty$, μ_Y follows a normal distribution with;

$$\begin{aligned} E[Y] &\sim \mu_Y \\ \text{var}[Y] &\sim \frac{\sigma_Y^2}{N} \end{aligned} \tag{79}$$

If Y_t exhibits self-similar behaviour, the sample mean is asymptotically and normally distributed with μ_Y but the variance is expressed according to Beran, (1994) by Equation (80),

$$\sigma_{LRD}^2 = \frac{2C_r N^{\alpha}}{(1+\alpha)\alpha} \frac{1}{N} \tag{80}$$

where $\alpha \in [0,1]$ and $C_r \in [0,\infty]$ parameters describe self-similar behaviour. Equation (80) implies that the variance of the sample mean decreases with sample size at a slower rate than the classical one with the ratio $\sigma_Y^2 : \sigma_{LRD}^2 \to \infty$ quickly with N. Clearly, computing any confidence interval of μ_Y would be biased. A robust approach that is capable of handling self-similar behaviour in WV time series could therefore be required. The wavelet estimator is one of the methods chosen in this thesis. Using the wavelet based approach; scaling properties in the data sets could be detected, identified and quantified. This is because, the WT often utilises an analysing *a priori* mother wavelet such as the Haar basic wavelet; which has inherent scale invariant properties. These semi-parametric estimators are computationally efficient and allow data analysis of arbitrary length. In addition, the estimators can also detect and isolate the deterministic components (trends) in the data.

There are many classes of scaling processes (Beran, 1994). In this thesis, it is desirable to distinguish between self-similarity and long range dependent processes. Self-similar (SS) processes (e.g. Fractional Brownian motion) are stochastic processes that are invariant in distribution under suitable scaling of time and space. A stochastic process $(Y_t, t \geq 0)$ is SS with exponent $H \in \mathbb{R}$ of SS if and only if all $c > 0$,

$$\left(Y_{ct}, t \geq 0\right) \overset{d}{=} c^H \left(Y_t, t \geq 0\right) \tag{81}$$

where $\overset{d}{=}$ indicates an equality in the statistical and/or distribution sense. For Gaussian processes with finite variance (these processes exhibit stationary increments), the following properties hold:

1. If $H < 0$, then $Y_t = 0 \ \ \forall t \geq 0$,
2. If $H = 0$ and $(Y_t, t \geq 0)$ is continuous probability, then $\forall t \geq 0$, $P(Y_t = Y_0) = 1$, which implies that $H > 0$ for this particular SS processes.
3. If for some $0 \leq \gamma \leq 1$, $E\left[Y_{t=1}\right]^{\gamma} < \alpha$, then $0 < H < 1$.

Some processes exhibit inbuilt memory which is dependent upon widely separated values that are significant even across large time shifts. Such stochastic processes are referred to as Long-Range Dependent (LRD) and their autocorrelations decay to zero slowly such that their

sum does not converge. Processes with long memory (or LRD) are stationary processes and contain spectral density that satisfies,

$$\rho_{\nu} \sim c_f |\nu|^{-\alpha} \quad \text{as } \nu \to 0 \tag{82}$$

where $\alpha = 0 < \alpha < 1$ (this describes the quantitative nature of the scaling) and $c_f > 0$ (is the measure of the size of the LRD and has the dimension of variance). Equation (82) implies that the auto-covariance, $r_{\tau} = E\left[Y_t Y_{t+\tau}\right]$ satisfies,

$$r_{\tau} \sim c_r \tau^{\alpha-1} \text{ as } \tau \to 0 \tag{83}$$

where $c_r = c_f 2\Gamma(1-\alpha)\sin\left(2^{-1}\pi\alpha\right)$, and Γ is the Gamma function (Beran, 1994). Equations (82) and (83) imply that the covariances, r_{τ} decays slowly. Increments of finite variances of SS processes have LRD as long as $0.5 < H < 1$, where H and α are related through

$$\alpha = 2H - 1 \tag{84}$$

Based on the wavelet analysis framework, the wavelet coefficients $d_{j,k}$ represent the difference between the aggregated time series by factors 2^{j-1} and 2^j for a fixed scale j. In this regard, the underlying assumption is that $d_{j,k}$ are short-range correlated. Given that $d_{j,k}$ are the wavelet coefficients at octave j, and if the mother wavelet has M vanishing moments and that its Fourier transform is M differentiable at the origin and m_j is the number of wavelet coefficients available at octave j, then $d_{j,k}$ is second order stationary. Furthermore, $E\left[d_{j,k}\right]$ can be estimated as reported by Abry *et al.*, (2000) by;

$$\mu_j = \frac{1}{m_j}\sum_{k=1}^{m_j} d_{j,k}^2 \tag{85}$$

In addition, the estimator of the log $E\left[d_{j,k}\right]$ is,

$$s_j = \log \mu_j - \frac{1}{m_j \log 2} \tag{86}$$

where the last term of right-hand side cancels the bias contributed from the nonlinear component of log2. As a result, a plot of j versus s_j yields the log-scale diagramme as described by Abry and Veitch (1998).

The log-scale diagramme of the coefficients of the WT was used to analyse WV fluctuations and investigate the presence of two most important self-similar behaviours; the LRD and SS. The wavelet-based estimator of the LRD and SS is based on the discrete wavelet transform, DWT. The analyses of the WV's LRD/SS and other derived parameters are based on the following procedure. Firstly, the data is discretely pre-filtered to eliminate outliers in the WV sequence. Thereafter, the DWT of the pre-filtered WV data series is computed and then the squares of the coefficients of WT are averaged. A linear regression on the log of the averaged coefficients of the WT (plotted on the y-axis) and the log of the scale (plotted on the x-axis) is fitted. In this regard, the log-scale diagramme was used to:

a) select the scale range where scaling is observed, and

b) estimate the scaling properties in the coefficients of the WV.

It is assumed that, a scaling phenomena could occur over a range of scales, $j = \{j_1, j_2\}$ and therefore for LRD processes, j_2 is infinite but j_1 is where the LRD begins (this value has to be selected). However, for SS processes, $j = \infty$ as j_2 remain infinite (Abry *et al*., 1999).

3.5.3. Hilbert-Huang transform

Geodetic data collection, pre-processing, analysis and visualisation of the inherent signal structure by use of DFA and WT methodologies often assume that the underlying processes are weakly stationary. Ideally, stationarity in geodetic data and tropospheric WV fluctuations cannot be guaranteed. In order to accommodate the inherent nonlinear and nonstationary properties of WV sequence, the reported research utilised an objective and flexible method that could describe the oscillatory events in WV fluctuations whose associated time-frequency characteristics evolve over time called the HHT. The HHT approach is able to deal with WV fluctuations in the multiple resolutions and therefore distinguishes different processes driving variability.

The gist of the HHT is the EMD whose basic concept involves the empirical identification of oscillatory modes in the data by means of the local extrema. The decomposition is based on the assumption that:-

a) the data must have at least two extrema,

b) there exists a characteristic time-scale defined by the time interval between consecutive extremes; and

c) the presence of an inflection point (no extreme) requires interpolation in order to obtain the extrema.

The EMD approach assumes that the target data set consists of different, simple and intrinsic modes of oscillation that need not be sinusoidal (e.g. slowly varying amplitude and phase), called *IMF*. Each *IMF* ought to satisfy two criteria (to resemble the generalised Fourier decomposition); a) an *IMF* may only have one zero between successful extrema; and b) an *IMF* ought to have zero local mean.

The EMD adaptive process is a recursive 'sifting' algorithm described by e.g., Huang *et al*., (1998) and Pegram *et al*., (2008). Given a time series $\{Y_t, t \geq 0\}$, the recursive 'sifting' procedure can be summarized as follows:

1. Take the input signal Y_{t-1} to decompose, where Y_0 is the original time series;

   1.1. identify the local extrema of Y_{t-1}

   1.2. construct the upper/lower envelope ($\Omega_{u,t}$ / $\Omega_{l,t}$ by interpolation

   1.3. approximate the local average envelope by $\mu_{\Omega_t} = 0.5\left(\Omega_{u,t} + \Omega_{l,t}\right)$

   1.4. extract the detail $d_{t,1} = Y_{t-1} - \mu_{\Omega_t}$

   1.5. If $d_{t,j}$ is an *IMF*, decompose Y_{t-1} into an *IMF* i.e. $IMF_t = d_{t,j}$ and the residual $Y_t = Y_{t-1} - IMF_t$. Otherwise repeat steps 1.1 through 1.5.

2. If Y_t has an implicit oscillation, set Y_t as an input signal and repeat from step1.

If the IMF_s are added together with the residual trend, the original signal is often recovered without any distortions or loss of information as shown in Equation (87),

$$Y_t = \sum_j \left(IMF_j\right) + Y_{res} \tag{87}$$

A key advantage of EMD is that *IMF*$_s$ can be transformed from the temporal-space to time-frequency space by applying the Hilbert Transform (HT) to each *IMF* component determined by Equation (88)

$$\begin{aligned} Y_t^H &= H\left[Y_t\right] \\ &= \frac{1}{\pi} pv \int_{-\infty}^{\infty} \frac{Y_\tau}{t-\tau} d\tau \end{aligned} \tag{88}$$

where pv is the Cauchy principal value or principal value of the singular integral. Note that Y_t and Y_t^H form a complex conjugate pair. Based on HT, the analytic signal is defined by,

$$\begin{aligned} z_t &= Y_t + iY_t^H \\ &= A_t e^{i\theta_t} \end{aligned} \tag{89}$$

where the instantaneous amplitude and phase are given by $A_t = \sqrt{Y_t^2 + Y_t^{H2}}$ and $\theta_t = A_t \tan\left(Y_t^{-1} Y_t^H\right)$ respectively. From Equation (89), the instantaneous frequency (which is also a function of time) of each *IMF* can be defined as

$$\omega_t = \frac{d\theta_t}{dt} \tag{90}$$

This implies that the HT of the n^{th} *IMF* components of Y_t can be written as:

$$Y_{n,t}^H = \sum_{j=1}^{n} A_{j,t} e^{i\int \omega_t^j dt}, \tag{91}$$

where $A_{j,t}$ is the amplitude of the analytic signal associated with j^{th} IMF. It is worth mentioning that the $\{A_{j,t}, \theta_{j,t}\}$can be projected on the time-frequency-energy (=$|A_{j,t}|^2$): forming the Hilbert-Huang spectrum. This spectrum has the same information as in the continuous WT reported in Torrence and Compo, (1998).

4. Modelling the stochastic properties of WV time series

A mathematician is a device for turning coffee into theorems.

- Paul Erdős

In this Chapter, the stochastic properties of WV are investigated. Firstly, results reveal that the underlying process driving fluctuations in the monthly averaged tropospheric WV derived from geodetic VLBI measurements spanning 1999-2008 is non-linear and nonstationary. Furthermore, an ARMA(10, 9) could generally model the monthly averaged tropospheric WV (transformed to stationary) over the period 1999-2008.

4.1. Introduction

A time series represents a path (also called a realisation) of a stochastic process or a sequence of data-points measured at successive time intervals. The collection of variables indexed according to the order in which they are obtained in time, forms the basis for the statistical description of the data which might have inherent spatial-temporal fluctuations (Shumway and Stoffer, 2006). The purpose of time series analysis is to develop mathematical models that provide robust descriptions about the observables; and as a result obtain an understanding of the mechanism that generated the data records (e.g., the nature and structure of the underlying forces). Analysis of time series could also involve fitting models to the data for purposes of future predictions of the phenomena in question. In geodetic applications, tropospheric characteristics can be presented as a vector time series that hold geophysical information which is of interest to the space geodesy community. The tropospheric state at each moment in time can be modelled as a mapping function denoted as $\Re^3 \rightarrow \Re^N$; which assigns an n-dimensional (nD) vector of real tropospheric parameters such as temperature, pressure, refractivity, WV and tropospheric gradients to every point of the 3D space, herein referred to as a time series or data sequence.

The problems encountered in time series modelling and prediction dates back to the pioneering work of Yule in 1972 (Yule, 1972). Until 1970's, most of the research work on time series analysis concentrated on the use of parametric methods to describe underlying process in the observed data. In these parametric approaches, simple linear models are fitted to the data; see for example the text-book by Brockwell and Davis, (1996). Parametric

approaches such as the maximum likelihood estimate (MLE) are commonly used when there is sufficient prior knowledge that the model in the data set has a parametric form with unknown parameters i.e., that the model is from some parametric family-set $\{\mathbb{Z}_\theta, \theta \in \Theta\}$, where $\mathbb{Z}_\theta$ is a known parametric form with unknown parameter θ to be estimated. Though the parametric approaches have established appealing mathematical properties over time, they often impose unsound rigid structure upon the underlying process. To study nonlinear time series, nonparametric models such as the Multivariate Adaptive Regression Splines, (MARS) (Chen *et al*., 1997) and the EEMD (Zhaohua and Huang, 2009) that do not impose any structural assumptions have been developed to model underlying processes. Nonparametric models are often formulated based on the principle of *"letting data speak for themselves."* Nonparametric models are therefore useful when little information or when flexibility about the underlying model is required.

One aspect of investigating a time series involves finding appropriate models for the time series. Generally, for a given model, the central theme is to estimate the unknown quantities of the model based on discrete observations. The process often involves model identification, fitting and model diagnostics. Given some data, there are often an infinite number of models or hypothesis that fit the data equally well. As a result, there is no reason to prefer one model over another. Therefore, one is forced to make assumptions that lead to an inductive bias. In model selection, the model parameters are selected such that a model of optimal complexity for a given (finite) data is created. Such models are said to have the correct inductive bias. Additional details on model section can be found in the text-book by Burhan and Anderson (2002).

Time series models have been central in the study of some behaviour of a process or metric over a period of time. The application of time series models are manifold, the applications range from geophysical problems such as daily weather forecasts, electricity (Taylor, 2008) and astronomy (Subba and Priestly, 1997). In decisions that involve some element of uncertainty of future values, time series models been found to be robust methods of forecasting. Additionally, time series models can be used to understand the structure in the data and predict the future trends and patterns in the data. Bates, (1994) analysed the adjusted time series of global WV derived from satellite infra-red observations and reported the linkage between upper troposphere WV time series seasonal component and the monsoon circulations. In addition, the inter-annual variability of WV was found to be correlated to the

ENSO warm-cold events. Based on the notion that all of the power in the integrated WV variations is contained in a synoptic-scale motion of air-mass, Davis, (2001) investigated the statistics of integrated WV time series using GPS data in order to assess the large-scale weather systems. It was found that the power spectral density could be a robust estimator of the integrated WV than the structure function is over long time scales.

Kruger, (2006) examined the spatial-temporal variations of trends in daily extreme precipitation indices for 138 rainfall stations for the period 1910 to 2004 in South Africa and reported of some certain areas where significant changes in certain characteristics of precipitation amid the lack of real evidence of overall changes in precipitation over the past century. Documented studies on the spatial-temporal characteristics of WV assume that the variability of WV is driven by stationary processes. Despite all these areas of application as mentioned above, assessment of the stochastic and self-similar properties of tropospheric WV has remained unexplored. In addition, there is no literature known to the author that has reported on the model of WV variability.

To characterise the stochastic behaviour of WV fluctuations, a general Auto-Regressive Moving-Average (ARMA) time series model will be considered. In the analysis of WV fluctuations, an automatic algorithm could be used to estimate the appropriate model parameters such as the model order from the tropospheric WV and the estimated model could be used to investigate the nature of the underlying process that drives the variability of tropospheric WV.

4.2. Basic concepts of time series analysis

If a random variable Y varies with time, then a simple time series, expressed as $Y_t=\{ y_t, \forall t=1, 2, ...\}$ (here $\forall t$ denotes for all integer valued time index) assumes that the measured data points are realizations of random processes (defined in the next section as) y_t that comprises of four components as shown in Equation(92) (Trömel and Schönmise, 2006);

$$Y_t=T_t+C_t+S_t+R_t. \qquad (92)$$

Equation (92) is similar to the model proposed by Li *et al.*, (2000) to investigate the presence of secular tectonic deformation fields and to distinguish between tectonically active and inactive regions in central Japan using GPS data. The additive model-components, T_t, C_t, and S_t in Equation (92) refer to the trend, non-random long term and short (seasonal) periodic

components respectively. The random variable R_t accounts for all the deviations from the ideal non-stochastic components. In many time series analysis strategies, the assumption that $E(R_t)$ exist or can also achieved by modifying one or more of the non-random components. Furthermore, if the underlying process is dominated by the growth component T_t only then:

$$\begin{aligned} E(Y_t) &= T_t, \\ &= f(t). \end{aligned} \tag{93}$$

Though the function f(t) is known, it is dependent on the unknown elements of the parameter space $\{\beta_1, \beta_2, \cdots, \beta_q\}$ such that,

$$f(t) \equiv f(t; \beta_1, \beta_2, \cdots \beta_q). \tag{94}$$

Using the ideal realisations, y_t the parameters, $\{\beta_1, \beta_2, \cdots \beta_q\}$ can be determined based on the least squares estimate, i.e.,

$$\sum_t \left(y_t - f(t; \beta_1, \beta_2, \cdots \beta_q) \right)^2 = \min_{\beta_1, \beta_2, \cdots \beta_q} \sum_t \left(y_t - f(t; \beta_1, \beta_2, \cdots \beta_q) \right)^2. \tag{95}$$

A feasible solution from the numerical problem given in Equation (95) entails;

$$y_t = f\left(t; \beta_1, \beta_2, \cdots \beta_q\right), \tag{96}$$

as the predictand of $y_t \; \forall t=1, 2, ...$; where t=1 is the current time. Therefore, the residuals in the realisations i.e., $\sim y_t - y_t$, possess the goodness of fit information of the model to the data.

4.3. Random variables

Over the probability space $\{\Omega, \mathcal{P}\}$, a discrete random variable Y is a function that maps a space of events Ω to the real axis $Y : \mathcal{P} \to \mathbb{R}$ by $\aleph \to Y(\aleph)$ where $\aleph \in \Omega$ is a particular elementary event. As a result, the probabilities $\mathcal{P}_Y(\aleph)$ map each $\aleph$ result between $0 \le \aleph \le 1$. A particular $\aleph$ with a probability $\mathcal{P}_Y(\aleph)$ is obtained based on the realisation of Y. The relative frequency $\mathcal{P} = \frac{N(\aleph)}{N}$ is the ratio between realisations of $\aleph$ and the total number of realisations N.

On the continuous space, $\mathcal{R}$, the definition of the random variable Y has the corresponding probability density function, $\rho_Y(y)$ with $y \in \mathcal{R}$. The probability distribution is therefore the integrated probability density function given in Equation (97)

$$P_Y(\lambda) = \int_{-\infty}^{\lambda} dy \rho_Y(y). \tag{97}$$

A value of $y < \lambda$ results from a realisation of Y with a probability $\mathcal{P}\{y|_y < \lambda\}$ based on the probability distribution function given by equation(98) where a typical example is one that exhibits a Gauss distribution with a density function given by Equation(99);

$$\mathcal{P}\left\{y|_y < \lambda\right\} = P_y(\lambda) \tag{98}$$

$$\begin{aligned} \rho_Y(y) &= N\left(\mu, \sigma^2\right) \\ &= \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{y-\mu}{2\sigma^2}} \end{aligned} \tag{99}$$

4.5. Stochastic processes

A stochastic process Y_t is a time dependent random variable with some dependency structure. A random process is often described by a probability density function $\rho(y_t)$ where, only the probability distributions are computed as compared to the deterministic case where a given outcome is determined by initial conditions and equations. A stochastic process Y_t is denoted as an indexed collection of random variable with $i = t_1$ indices specifying some time ordering in the discrete or continuous space. Two data records observed between t_1 and t_2 might exhibit linear dependency that could be described by an auto-covariance function given by;

$$Cov(t_1, t_2) = \left\langle \left(Y_1 - \langle Y_1 \rangle\right)\left(Y_2 - \langle Y_2 \rangle\right) \right\rangle.$$

If the joint probability distribution is time invariant, the stochastic process is strictly stationary for any moment of time $\left\{t_i, t_j\right\}$. A second-order (weakly) stationary process exhibits a constant mean and the auto-correlation function depends only on the lag.

One way to describe a stochastic process is to specify the joint probability distribution of $Y_{t_1},...,Y_{t_n}$ for any set of times $t_1,...,t_n$ and any value n. However, a simpler, more useful way of describing a stochastic process is to obtain the moments of the process which are the mean, variance and the auto-correlation function. Stochastic processes can be characterized by stationary, auto-covariance function a spectrum and ergodicity properties. Stationary processes play an important role in the analysis of time series. However many observed time series are nonstationary in nature. Therefore the stationary property (which means that $\rho(y_{t_i}, y_{t_j}, ...)$ is invariant) in real data is strictly rare. Instead, a first order stationary process exhibiting a time independent mean is more common. If a process also has time independent

variance and its covariance function is a function of time differences only, then the process is described as weakly stationary or second order stationary.

$$\begin{aligned} \left\langle Y_{t_i} \right\rangle &\equiv \mu \\ &= C_{\mu}, \\ \mathrm{Var}(Y_i) &\equiv \sigma^2 \\ &= C_{\sigma}, \\ \mathrm{Cov}(\tau) &= \mathrm{Cov}(Y_i(t_i), Y_j(t_j)). \end{aligned} \tag{100}$$

Here $C_{\mu,\,\sigma}$ are constants. The auto-covariance functions of a stationary process could be;

$$\mathrm{Cov}(\tau) \equiv \left\langle (Y_{t_1+\Delta t} - \left\langle Y \right\rangle)(Y_{t_1} - \left\langle Y \right\rangle) \right\rangle, \tag{101}$$

This auto-covariance function can be normalized to the variance to obtain the auto-correlation function, $C(t) = \left[\mathrm{Cov}(\tau_0)\right]^{-1} \mathrm{Cov}(\tau)$. Furthermore the spectrum $S(\omega) = \mathrm{Cov}(\tau)$ is frequency-domain equivalent of the auto-covariance. In order to represent a stochastic process, multiple realisations are often averaged. This gives rise to the power spectral density given by,

$$S = \lim_{T\to\infty} E\left[\frac{G_T(\omega)^2}{2T}\right], \tag{102}$$

where $G_T(\omega) = (2\pi)^{-0.5} \int_{-T}^{T} Y_T(t) e^{-i\omega t} dt$ is the Fourier integral of the stationary process $Y(t)$. The Fourier transform of the auto-covariance function is related to the spectral density as follows;

$$\mathcal{S}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} COV(\tau) e^{-i\omega t} d\tau. \tag{103}$$

Equation (103) implies that the spectral density and the auto-covariance function describe the linear dynamic properties of a process equivalently. Sometimes real data may exhibit a temporal mean and auto-covariance function that converges to a value equivalent to the average of a set of realisations (say for example, ensemble average). This property is referred to as the ergodicity of the underlying process.

A stochastic process is called short-range correlated or Short-Range Dependent (SRD) if the auto-covariance function, $\rho_{SRD}(\tau)$ is summable and decays exponentially,

$$\sum_{\tau=-\infty}^{\infty} \rho_{SRD}(\tau) = \mathrm{const} < \infty. \tag{104}$$

Not all data records exhibit SRD, instead they posses the long-range dependence which can be described by an algebraic decay of the auto-covariance function; $\rho_{LRD}(\tau) \propto \tau^{\gamma}$. Such type

of processes are called long-memory processes or long-range correlated and their decay exhibits a diverging sum as shown in Equation (105)

$$\sum_{\tau=-\infty}^{\infty} \rho_{\mathrm{LRD}}(\tau) = \infty. \tag{105}$$

It is worth mentioning that long-memory processes are non-periodic stochastic processes. Dependence over infinite time lags due to the deterministic behaviour in the data record does not influence the long-range correlated datasets.

4.6. Geodetic parameters time series

Geodetic time series is derived from a sequence of geodetic data sets observed over a period of time and arranged according to observation time. There are numerous reasons to record and analyze the geodetic time series data. For instance, geodetic time series analysis could be vital in understanding the structure of the processes that generate the data as well as aid in the prediction of future values. The characteristic property of a time series is that, the data records are not independently generated; they exhibit a time dependent dispersion and are often governed by trend and may also exhibit periodicity. In geodetic applications, there are two steps of geodetic time series analysis. The first is identifying the nature of the process represented by the sequence of observations and, secondly prediction of the future values. Both of these goals require that a pattern of time series data is observed and described. Thereafter, the internal pattern (which could be the autocorrelation, trend or seasonal components) could then be interpreted and integrated to formulate models in geodetic time series that could be vital for tropospheric modelling which improves the accuracy of the geodetic delay observable.

In general, analysis of the geodetic time series reported in this thesis, focuses on the estimation and extraction of the deterministic (e.g. trend and seasonal) components in the geodetic (e.g. WV) data, see the first three terms in Equation(92). The components in geodetic WV time series could be used to determine the best model representing their variability. In the analysis of geodetic data, it is assumed that the bias term (which is contained in last term in Equation (92)) turns out to be a stationary random process. Thus, the theory of random processes can be used to find a satisfactory (and probabilistic) model for the bias term, analyze the properties of bias term and use it in conjunction with the deterministic components to predict the observed geodetic WV series. Alternatively, analysis of the observed geodetic WV series could be approached from Box and Jenkins paradigm

(Box and Jenkins 1970). In this model, the difference operators are applied repeatedly to the observed series until the differenced observations resemble a realisation of some stationary processes, from then on, the theory of stationary process is used model and analyse the stationary (and therefore the original) series.

4.6.1. Time series analysis of tropospheric WV

Strategies for quantifying the overall transient variability of WV are not straightforward, and a number of statistical approaches for modelling WV variability have been attempted. Gierens et al. (1997) used the measurements of about 2000 flights within the Ozone and WV by Airbus in-service aircraft (MOZAIC) program and confirmed that fluctuations of humidity and temperature from their local means could be characterized by occasional large fluctuations (i.e. heavy-tailed distributions) in the upper troposphere (at pressure levels 166 to 290 hPa on the general circulation model grid scale). It was found that the fluctuations could then be modelled by the Lorentz distribution rather than the Gauss distribution and this was due to large excursions in the fluctuations of humidity and temperature. Later, data from 3 years of MOZAIC measurements (this is data is described in Gierens et al. 1997) was used to determine the nature of the distribution law of WV; which plays a vital role in testing whether the hydrological cycle in climate models is adequately represented. It was reported that the frequency of occurrence of relative humidity greater than 100% decreased exponentially above ice saturation and that it decreases exponentially for the entire range of values in the lower stratosphere (Gierens *et al*. 1999). A stochastic source-sink model capable of producing such distributions was then formulated.

Data from NASA's Pacific exploratory mission in the tropics phase A, that was conducted between August and September 1996 was used to study the impact of human activity on tropospheric chemistry in the remote regions over the pacific (Cho and Newell, 2000). Based on the empirical multifractal formula for the structure function originally described by Pierrehumbert, (1996), an "anomalous scaling" or multi-fractality between 50 to 100 km horizontal range of the WV distribution was reported (Cho and Newell, 2000). From these findings, it was noted that while WV increase was statistically stationary, the transient WV fields did not exhibit the stationary properties. As a result, the probability distribution (e.g., the variance) of transients could therefore not be characterised accurately from a finite number of observations.

Recently, Jin *et al*., (2009) used the precipitable WV time series determined from co-located space geodetic techniques to quantify the systematic biases between VLBI and GPS in the 5-year co-located measurements. The reported results demonstrated that systematic biases in the geodetic data that describe the atmosphere systems and processes could be accounted for if the co-located observations are utilized. Due to the role played by atmospheric WV in Earth's atmospheric radiation budget, global hydrological cycles and global climate change (Suparta *et al*., 2009), these findings have important applications in weather forecasting, numerical weather prediction and climate change studies as discussed by Gettelman and Fu, (2008).

In this section, the analysis of geodetic WV time series is limited to, a) determining or transforming of the geodetic WV time series to stationarity, b) detecting seasonality using the autocorrelation, partial autocorrelation and automatic spectral plots, and c) deducing the inherent stationary model in the data. In the present analysis, geodetic WV data is assumed to exhibit a systematic pattern and random noise (error). In order to observe the pattern more clearly, some form of noise filtering is done by use of point-averaged smoothing. This methodology involves fitting some function or adjusting/correcting for the trend in the data records. To this end, an adaptive filtering algorithm reported in Wessel and Voss (2000) and described in detail in Section 3.4.2 is applied in smoothing the geodetic WV time series. For instance, Figure 4.1 depicts a sample wet tropospheric linear horizontal gradient data set that has been adaptively filtered in order to eliminate measurement or systematic errors. The filtering procedure employed here is based on the adaptive cumulates (the mean and standard deviation); used as filter coefficients which were adapted spontaneously during the computation process. This is a robust approach because it caters for the sudden changes in the time series. In the current filtering procedure, basic variability of zonal gradients is calculated using a binomial-7-filtering given in equation(106).

$$Y_t^{\Delta} = \frac{Y_{t-3} + 6Y_{t-2} + 15Y_{t-1} + 20Y_t + 15Y_{t-1} + 6Y_{t-2} + Y_{t-3}}{64} \qquad (106)$$

The adaptive mean and standard deviation are calculated and observations that are flagged anomalous based on the procedure described in Section 3.4. The reconstructed series is used to compute the basic variability, the new adaptive mean $\mu_t^{a'}$ and standard deviation $\sigma_t^{a'}$ which are then used to test for anomalous values using the inequality,

$$\left|Y_t^{\Delta} - \mu_t^{a'}\right| > \sigma_t^{a'} k_{f1} + \sigma^k. \qquad (107)$$

Here the filter coefficient vary as $2 \le k_{f1} \le 5$. The diurnal fluctuations of the zonal gradients are represented by a basic variability $\sigma^k \sim 365^{-1}$. Anomalous values are replaced with the respective values determined *a priori*. The filter coefficients take empirical constant values that range between 2.5 and 5.0 (these values are arbitrary determined). The reconstructed series plotted in Figure 4.1 used a filter coefficient that was arbitrary set to 3.0. In order to minimize the filtering biases, an adaptive basic variable of 0.011 was considered. This value was selected because the tropospheric linear horizontal gradients considered in this analysis are often sampled 4 times daily. The weighting factor of 365^{-1} was used to account for daily fluctuations in the observed series. As depicted in Figure 4.1, the unfiltered zonal gradients have anomalous values in the first and last quartiles. These values are not removed from the series. Instead, interval filtering described in Wessel *et al*., (1994) is used to adjust the anomalies. The interval filtering is done through the spontaneous adaptation of the filter coefficients due to the sudden changes in the observed series.

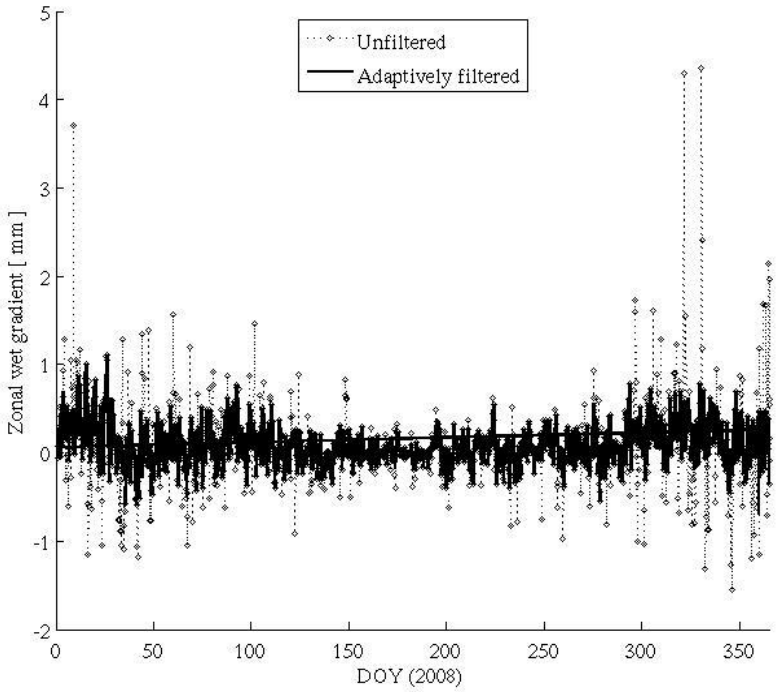

Figure 4.1 Adaptive filtering applied to the zonal linear horizontal wet gradient component observed over HartRAO. The linear zonal gradient has not been detrended.

4.6.2. Investigating stationarity in tropospheric geodetic WV

In this subsection, the stationarity in the geodetic tropospheric WV is examined and described by use of classical Box-Jenkins test. If the test for stationarity fails, then the geodetic WV ought to be transformed into a stationary time series before a model for stationarity is formulated. In the current analysis, stationarity in geodetic WV time series is identified by examining the behaviour of the sample auto-correlation function. The auto-correlation function measures the correlation between Y_t and $Y_{t+\tau_{acf}}$ (where τ_{acf} is time lag). In other words, the auto-correlation function of the geodetic WV time series describes the correlation (which also refers to the degree of dependence) that exists between geodetic WV time series and the same geodetic WV time series but lagged by 1, 2,.., τ_{acf} . The auto-correlation function is then depicted graphically by use of a histogram both quantitatively and qualitatively by, a) determining the period of the oscillation and b) by looking at the shape of the autocorrelation plot which gives some indication of the suitable model parameters of the time series model (e.g., autoregressive (AR), moving average (MA) or autoregressive-moving average (ARMA)). In most data sets, the auto-correlation coefficients are significant for a large number of time lags, τ_{acf}. However, due to the propagation of the auto-correlation at τ_{acf}, the auto-correlation at τ_{acf} >1 might not be explicit. As a result, the partial auto-correlation plot is used to ascertain whether auto-correlation at τ_{acf} *>1* indeed represents the auto-correlation at τ_{acf} =1. The partial auto-correlation plot is derived from the partial auto-correlation function. The partial auto-correlation plot often has a spike at τ_{acf} =1 which could imply that all the higher order auto-correlations are effectively explained by the auto-correlation at τ_{acf} =1.

In most time series analysis, there is no theoretical reference result that could be used to select and characterize a stochastic process. In this thesis, we consider a general model ARMA *{p, q}* and then automatically identify the appropriate model parameters hereafter, the model orders, *{p, q}* based on the underlying temporal pattern embedded in the time series data through:- a) identifying the appropriate time dependence by ensuring that the auto-covariance, auto-correlation and the partial auto-correlation described in the time series model represents a stationary stochastic process, b) ensuring that the filtered biases from the appropriate models are normally distributed, and c) ensuring that the model choice is parsimonious. After obtaining stationary geodetic WV time series, the sample auto-correlation function and partial correlation function could be used to identify a Box-Jenkins model that appropriately describes the geodetic WV time series. In order to assess the

specific statistical structure of geodetic WV sequence, a stochastic approach that models the dependent structure embedded in geodetic WV time series is considered. As a result, the Box-Jenkins statistical methodology briefly described in the following is used.

If the values of geodetic WV series are denoted by Y_t, Y_{t-1}, Y_{t-2}, then Box-Jenkins methodology is based on an ARMA *(p, q)* model given by Equation (108);

$$Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} + U. \quad (108)$$

Here the a_t's are independent and identically distributed random shocks with a zero mean and a finite variance σ_a^2. Further, $\phi_i \ \forall i = 1,2,\ldots,p$ and $\theta_j, \ \forall j = 1,2,\ldots,q$ respectively denote the AR and MA coefficients while *U* is the model constant which is related to the mean of geodetic WV series. In the model characterized by Equation(108), the current geodetic WV series observation, Y_t is explained by a linear combination of the *p* previous observations, Y_{t-1}... Y_{t-p}, a linear combination of the *q* previous random shocks a_{t-1}..., a_{t-q} and a constant term *U*. The error term is given by a_t. If *q=0*, a pure AR (*p*) is derived and if *p=0*, the class of pure MA (*q*) is retrieved. The backward shift operator, *B* can be used such that $BY_t=Y_{t-1}$, a compact ARMA (*p, q*) model then becomes;

$$\phi(B)Y_t = U + \theta(B)a_t. \quad (109)$$

Here, $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q$ represent the AR and MA operators respectively. For stationarity, $\phi(B)$ roots ought to lie outside the unit circle. The model is described by Equation (109) could be visualized in the sample and partial autocorrelation plot with characteristics given in Table 4.1.

Table 4.1 Sample AC and PAC model behaviour

Model	**SAC**	**SPAC**
MA	Cuts off at lag p	Tails off
AR	Tails off	Cuts off at lag p
ARMA	Tails off	Tails off

From Table 4.1, the plots of the SAC and SPAC could be visualised and a pure AR (*p*) and/or MA (*q*) processes established. However, estimates of (*p, q*) are not trivial from ARMA (*p, q*) processes. In the current analysis, a large number of candidate geodetic WV time series

models are computed and some statistical criteria (e.g., AIC (Akaike, 1969) or SIC in equations(110)), where μ_t and $\sigma^2=\sum_{t=1}^{T}\left(\mu_t\ T^{-1}\right)$ are the estimated residuals and variances respectively) is used to select a suitable model that is representative of the data. To this end, an ARMA (*p, q*) model is estimated based on the method of maximum likelihood. The likelihood function of the ARMA (*p, q*) model is non-linear in the unknown parameters and therefore non-linear optimization techniques are often used to solve for the unknown parameters.

$$\begin{aligned} \mathrm{AIC} &= \log\tilde{\sigma}^2+\frac{2}{T}(p+q+1); \\ \mathrm{SIC} &= \log\tilde{\sigma}^2+\frac{(p+q+1)\log T}{T}. \end{aligned} \tag{110}$$

In the current case study, WV time series data (for the period 1998 to 2008) derived from surface temperature measurements at HartRAO and numerical simulations of the ECMWF was used to investigate the stationarity properties. From the results, some indication of the broad correlation characteristics were averred from the sample auto-correlation and partial auto-correlation plots of the tropospheric WV time series as depicted in Figure 4.2. From the sample autocorrelation plot, it is evident that the line graphs exhibit damped oscillations which are the absolute sinusoidal components. These oscillations tail off slowly to zero, therefore indicating that the fluctuations in geodetic tropospheric WV time series are driven by non-stationary stochastic processes. In the present analysis, the auto-correlation function plot of the WV series exhibits spikes at several lags (e.g. at lags 1, 2, 4, 5, 6, 7, 8, 10, 11 and 12). This is expected since WV time series possess inherent diurnal, seasonal and trend components. One way to formulate a model for data with such auto-correlation pattern is to use the Space Time Auto-Regressive Moving Average (STARMA) models proposed by Pfeifer and Deutsch, (1980). However, the STARMA models do not take into account the embedded nonlinear behaviour that is representative of the underlying process. In particular, dynamical processes with unusual jumps cannot be effectively studied using STARMA models.

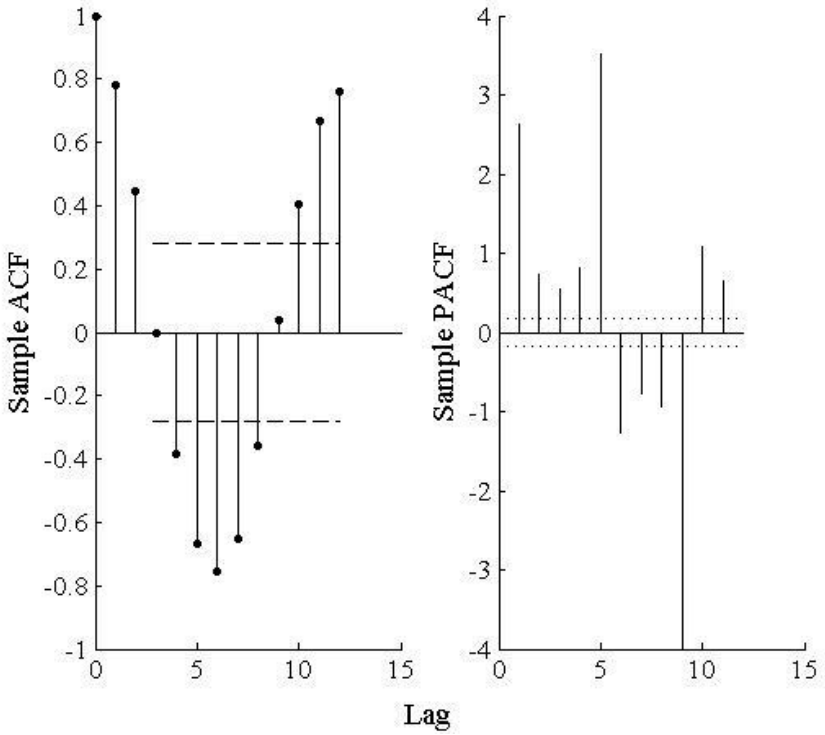

Figure 4.2 Sample autocorrelation and partial autocorrelation function for WV

To use ARMA models adequately, the WV time series must be stationary with respect to the mean and variance. Differencing between consecutive observations is one way to efficiently achieve stationarity. If the difference operator is defined as $\nabla = 1 - B$, such that $\nabla Y_t = Y_t - Y_{t-1}$ and the differencing goal is to stabilize the mean, then the model corresponding to the original WV series could be called Auto-Regressive Integrated Moving Average (ARIMA) model. An ARIMA (*p, d, q*) model is defined by the equation;

$$\phi(B)\nabla^d Y_t = U + \theta(B)a_t. \qquad (111)$$

Here, the d^{th} is the difference of the original series $\nabla^d Y_t$ is stationary and could be represented by a stationary ARMA (*p, q*) model.

The WV time series is often computed daily, monthly, quarterly or annually. Therefore WV series exhibit strong diurnal, seasonal or annual periodic fluctuations that often recur at pre-determined phases. In addition, it is expected that seasonal nonstationary features could be embedded in the series. For a WV time series of period k, (k =12 for

monthly data, and k = 36 for quarterly data), stationarity is often achieved by calculating the seasonally differenced series $Y_t - Y_{t-k}$. The seasonal difference operator of period k is denoted by $\nabla_k = 1 - B^k$ and therefore a seasonal ARIMA model could be defined by an equation of the form;

$$\phi(B)\nabla^d \nabla_k^D Y_t = U + \theta(B)a_t \qquad (112)$$

The model in Equation (112) takes into account seasonal and regular differencing due to the presence of trend components in the data.

In ARMA modelling applied in this research, the data series ought to be stabilised by using various transformations such as square root or the logarithmic transformations. A practical tool for the choice of the appropriate transformation which is based on the power transformation is the mean-range plot where the range of data is plotted against the mean of each seasonal period. For a detailed discussion on the mean-range plot, refer to Helfenstein, (1986). In the current analysis, the Box-Cox transformation was used to transform the data to a stationary time series. The Box-Cox transformation can be taken as a general time deformation process applied to WV series. This type of transformation is only in the time domain and therefore the notion of stationarity is restricted to a linear transformation. The resultant data series is then subjected to the two-sided Lilliefors and the Jarque-Bera goodness-of-fit test of composite normality which performs the normality test based on the hypothesis that the data in the WV comes from an unspecified normal distribution.

The Lilliefors test evaluates the hypothesis that the WV observations have a normal distribution with unspecified mean and variance, against the alternative that the WV observations do not have a normal distribution. This test compares the empirical distribution of WV with a normal distribution having the same mean and variance as WV. The test is similar to the Kolmogorov-Smirnov test, but since the parameters of the normal distribution are estimated from WV rather than specified *a priori*, the test becomes more data adaptive. From the test, if the result *H=1*, then the hypothesis that WV observations have a normal distribution is rejected. However, if *H =0*, then it implies that null hypothesis cannot be rejected. Additionally, the P-value is computed by interpolation into the Lilliefors simulation table. Table 4.2 illustrates that at 5% significance level, the result of the test is *H=0*. This indicates that the null hypothesis (i.e. the data are normally distributed) cannot be rejected at 7.2 % significance level. In addition, the Lilliefors test statistic of 0.0713 is smaller than the

cut-off value of 0.0745 at 5% significance level, therefore the hypothesis of normality cannot also be rejected.

Table 4.2. Test statistics of the Box-Cox transformed WV normality tests

Test type	**H**	**P-value**	**Statistic**	**Critical value**
Lilliefors	0.0	0.072	0.0713	0.0745
Jarque-Bera	0.0	0.055	5.3233	5.5782

Additionally, the Jarque-Bera test evaluates the hypothesis that WV observations have a normal distribution with unspecified mean and variance, against the alternative that WV does not have a normal distribution. The test is based on the sample skewness and kurtosis of WV. For a true normal distribution, the sample skewness should be near 0 and the sample kurtosis should be near 3. The Jarque-Bera test determines whether the sample skewness and kurtosis are unusually different than their expected values, as measured by a chi-square statistic. The Jarque-Bera test for normal distribution in WV results yielded similar result of $H=0$ obtained from the Lilliefors test and therefore the normality hypothesis at the 5% significant level and 5.5% P-value could not be rejected. The derived time series models could be used to analyze the power spectral density and covariance function of the stochastic WV observations. These models could be suitable for characterizing the spectral density of the random WV observations with known model type and order. However, it should be noted that the spectral characteristics of the WV observations are often unknown *a priori* and therefore a large number of candidate models ought to be computed.

An automatic algorithm that estimates a suitable ARMA (p, q) model to the monthly averaged stationary tropospheric WV revealed that the 1999 to 2008 monthly tropospheric WV data could be modelled by AR (100), MA (20) and ARMA (10, 9). It can be seen that MA order of 20 is much lower than the selected AR order of 100 partly due to the fact that a high AR order model is often used as an intermediate parameter for the estimation of MA models. The optimal coefficient vector of the AR model consists of the set, {ϕ: 1.00, -0.72, -0.11, 0.32, -0.05, 0.11, 0.03, 0.13, 0.02, and -0.25} and the optimal coefficients of the MA model were found to be the vector set, {θ: 1.00, -0.25, -0.15, 0.12, -0.28, 0.27, 0.33}. The prediction error of the three best-selected models is estimated based on the measured and

given values of the residual in variance described by Broersen (2002). For MA and ARMA models, the prediction error is given by Equation (113);

$$e_p(m) = \{\delta(m)\}\frac{1+\frac{m}{N}}{1-\frac{m}{N}}. \qquad (113)$$

Here, m is the number of estimated parameters in the model and δ is the residual of the variance. For AR (*p*) models, the prediction error is given by Equation (114);

$$e_p(p) = \{\delta(p)\}\prod_{m=1}^{p}\frac{1+\frac{1}{N+1-m}}{1-\frac{1}{N+1-m}} \qquad (114)$$

Equation (114) is significantly different from Equation (113) for m > *0*. A single time series model, with selected model order and type, with the smallest estimate of e_p could therefore be easily selected. Using the estimated parameters of the selected model, the spectral density and other statistically significant details such as the second-order characteristics of WV observations could be inferred. The model error denoted as ME is defined as the measure of the accuracy of the estimated model which is the difference between the estimated model and a true stationary process. This measure is simply a scaled transformation of the one step ahead squared error of e_p. Based on the vector set $\{\phi, \theta\}$ of the tropospheric WV, a true stochastic stationary process was modelled and the difference between the stationary process and the estimated model from the WV data. An ME value of ~ 262 was obtained. However, the difference between the selected ARMA and AR models from the same WV data set was found to be ~ 15.

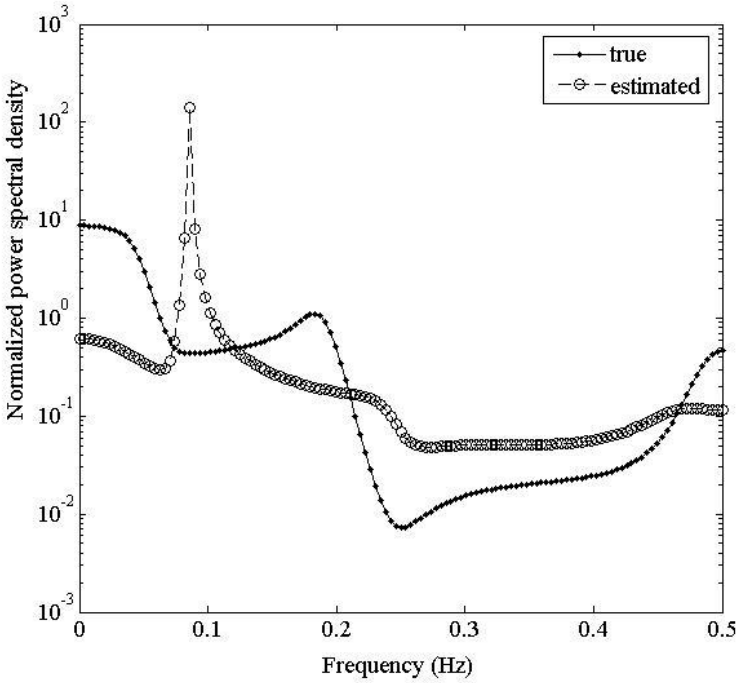

Figure 4.3. True and estimated power spectral density (log scale)

Figure 4.3 compares the power spectral densities of a true stationary process and the spectrum estimated from WV model on the log-scale. It is evident from the figure that the estimated model approximates to the true stationary process with subtle differences which are quantified by the ME. Furthermore, the estimated model accuracy as a function of model order and type is depicted by Figure 4.4. It is clear from Figure 4.4 that AR (50) and ARMA (10, 9) models have higher accuracy than the MA (20) based on the underlying processes in WV data. In all the estimated models, the accuracy increases with increase in the model order up to the $\{p, q\}$ of the ARMA model.

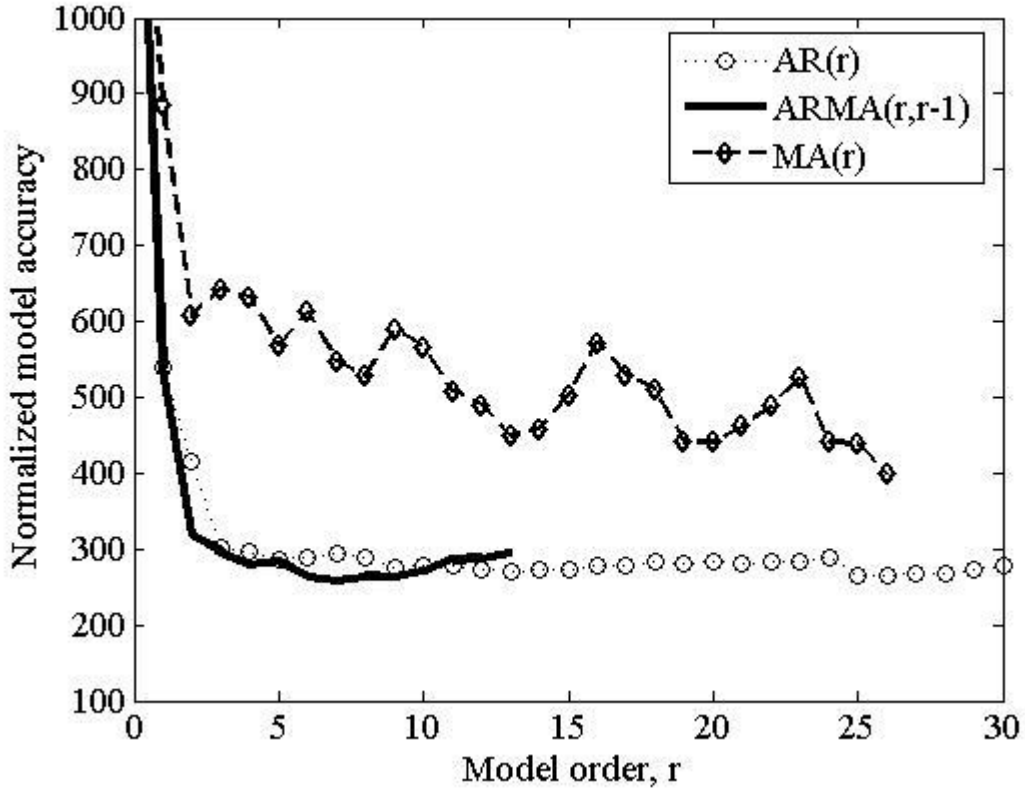

Figure 4.4. Estimated model accuracy as a function of the model type and order.

Assuming that the estimated model is representative of the data, the resultant spectrum could be robust and more accurate compared to the parametric spectrum which could be computed from, for instance the classical periodogram. This is evident from the left panel of Figure 4.5. As can be seen from the left panel of Figure 4.5, the spectral density of the periodogram do not have a smooth curve; this could be attributed to the distortion of the spectrum by convolution of the window function whose width equal the length of the WV. In addition, the periodic oscillations of WV are often treated differently in the discrete Fourier transform. For instance, if there are narrow spectral features (e.g. high frequency components in WV fluctuations), those narrow components will be broadened by the convolution. On the other side, there will be less broadening of the spectral peaks for a broad window function in the time domain but whose spectral main lobe is narrow in the frequency domain.

As depicted in Figure 4.5, the power spectral density based on the periodogram vary significantly from the true power spectral density of the WV series. A common approach (also called the Bartlett method) used to correct this inconsistency is to divide the data record into small subsets and compute the periodogram separately. Thereafter, the results are averaged over all the small records. Averaging periodograms reduces the variance in the

estimated power spectral density and therefore provides a better estimate of the spectral properties of the WV observations. Furthermore, the spectrum of the true stationary process (in the left panel) compares well with the spectrum of the true and the estimated time series (this is plotted in the right panel).

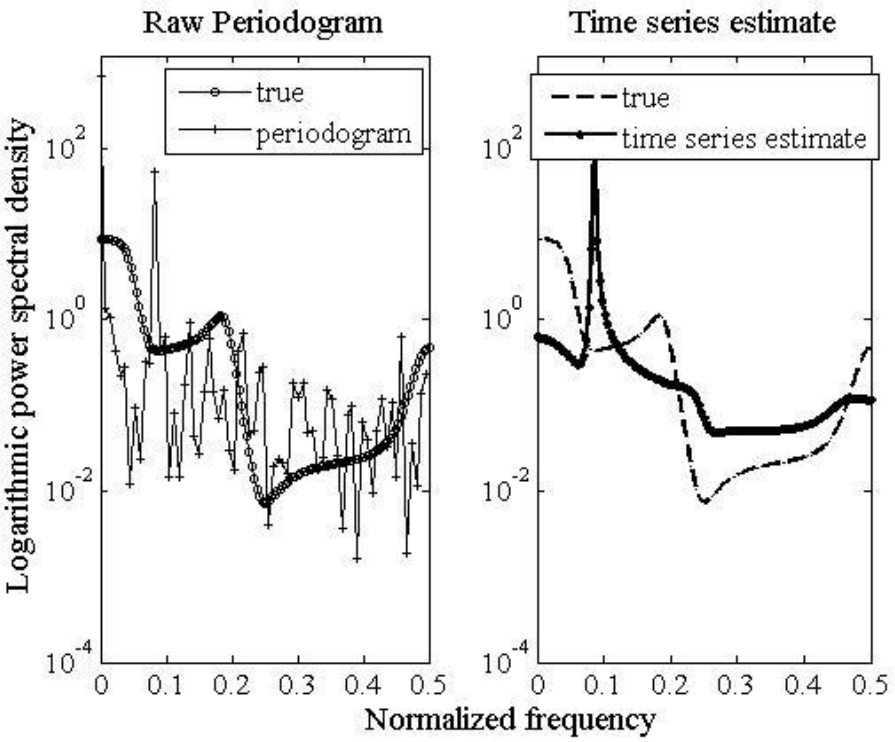

Figure 4.5. Periodogram and the spectra of the true and estimated time series.

4.7. Concluding remarks

Recent advances in atmospheric remote sensing have availed WV data from a variety of sources and sensors with improved spatial-temporal resolution. As a result, data sets that could be used to compute WV for investigating the structure and dynamics in the troposphere have increased. Further, the long time series of WV allows detailed studies on WV acting as a major component of the global hydrological cycle, as a greenhouse gas as well as the variability of WV at different spatio-temporal scales in the climate system. In this chapter, the stochastic behaviour of stationary WV fluctuations has been characterised using a general auto-regressive moving-average (ARMA) time series model. In the analysis, an automatic algorithm which estimates the appropriate model parameters has been used to formulate a model that is used to investigate the nature of the underlying processes that drives the

variability of tropospheric WV. In the present analysis, monthly averaged stationary tropospheric WV derived from geodetic VLBI measurements for the period from 1999 to 2008 is modelled by AR (100), MA (20) and ARMA (10, 9). The power spectral densities of a true stationary process and the spectrum estimated from WV model were compared on the log-scale. Results showed that the estimated model approximates to the true stationary process. Furthermore, the estimated model accuracy as a function of model order and type showed that AR (50) and ARMA (10, 9) models have higher accuracy than the MA (20) based on the underlying processes in WV data.

5. Multi-scale WV fluctuation characteristics over southern Africa

A small forcing can cause a small change or a huge one
- National Academy of Sciences (2002)

This Chapter presents results and discussion on the vertical WV model derived from SHADOZ network and HALOE satellites. Results depict spatial-dependent vertical differences in the WV model. Furthermore, the multi-scale organization is clearly evident in the spatial-temporal WV variability over the low- and mid-tropical Africa by use of NCAR/NCEP reanalysis and the in situ SHADOZ network data. In particular, the present chapter considers the use of WV derived from model simulations and in suit radiosonde data to assess the power law scaling behaviour of WV. Understanding of this complex behaviour contributes towards understanding the contribution of meteorological factors that influence geodetic tropospheric delay modelling.

5.1. Introduction

Analysis of WV variability in the low- and mid-tropical Africa is based on the *in situ* radiosonde observations of the SHADOZ station network comprising of Ascension, Irene (South Africa), Reunion (Reunion) and Nairobi (Kenya) and the numerical model simulations for the period from 1998 to 2006. The motivation for analysing WV fluctuations in the tropical Africa is driven by the desire to obtain an in depth understanding of the spatial-temporal WV fluctuations as well as study the mechanisms driving WV variability and its link to the climatic variables. Analysis of the climatic variables influenced by WV is essential for the accurate modelling of the influence of WV on the estimation of the geodetic tropospheric delay observable and of the regional hydrological cycle in Africa-South of the Equator. The analysis of WV fluctuations is based on the multiscale organisation paradigm as detailed in section 5.2, where a mean vertical profile of the troposphere WV is developed using the radiosonde measurements at Irene-South Africa and Malindi- Kenya. This research

work has been published as journal articles; see Botai *et al*., (2010) and Sivakumar *et al*., (2010).

The principal effects of variations of WV in the troposphere are; (i) the effect on the radiative balance, and (ii) the effect on cloudiness (which indirectly influences the radiative field). Throughout the troposphere, WV exerts a strong influence over how Earth loses radiative energy to space and this sets a balance between the energy received and absorbed from the sun. The WV feedback (i.e., the feedback on global temperature caused by changes to WV resulting from increases in CO_2 and other gases) is now almost universally accepted to be positive and strong. Based on data from NASA's satellite borne atmosphere infrared sounder (AIRS) over the period 2003-2008, Dessler *et al*., (2008) assessed the tropospheric WV response to global-average surface temperature of the Earth and reported strong positive WV feedback with a magnitude of $\lambda_q \sim 2.04\ Wm^{-2}K^{-1}$. This finding corroborates results from climate model simulations.

Gettelman and Fu, (2008) used humidity and temperature data from AIRS to analyse how the upper troposphere responded to changes in the underlying surface temperature. These observations were compared with simulations of the NCAR community atmosphere model version 3 (CAM) described by Collins *et al*., (2006). The results from AIRS and CAM simulations found a positive WV feedback i.e., as the temperatures increase, the WV in the upper troposphere also increases to keep the relative humidity at nearly equilibrium.

The abundance of WV in the atmosphere has a significant consequence on the earth's climate. This is due to its large energy transfer associated with phase transition where short-term dynamics of the atmosphere is also affected. WV therefore plays a key role in both the radiative and dynamic processes of the climate system (Zveryaev *et al*., 2007). The sensitivity of precipitation, WV and temperature changes in large-scale atmospheric circulation makes identification of the regional trends in precipitation and their contribution to temperature and WV variability critical. This involves the formation of the Polar Stratospheric Clouds (PSC) which are the reservoirs of halogenated molecules involved in the spring ozone depletion. Acid rain in the form of H_2CO_3, HNO_3, H_2SO_4, etc., is formed by the reaction of CO_2, nitrogen dioxide (NO_2) and Sulphur dioxide (SO_2) in their aqueous states.

Global distribution and variability of the atmospheric WV has been well documented (e.g., Dai, 2006). In addition, evidence of WV variability over regional scales has also been documented (see e.g. Trenberth *et al*., 2005). Further, WV is unique among atmospheric trace constituents due to the role it plays in creating saturation conditions prevalent in the

atmosphere. This property is the most important factor governing the distribution of WV in the atmosphere, both in the troposphere and in the stratosphere.

In the troposphere WV varies by as much as four orders of magnitude in a vertical profile while in the stratosphere, variations are much smaller (~10% lower, as reported by Stenke and Grewe, 2005) but still significant. Furthermore, WV in the upper troposphere and low stratosphere plays a key role in atmospheric chemistry (Stenke and Grewe, 2005). For instance, the hydrogen oxides involved in catalytic reaction cycles which are responsible for the control of the production and destruction of ozone layer in the lower stratosphere are produced from the oxidation of WV and methane by excited oxygen atoms. In addition, the partitioning of the nitrogen and halogen family is influenced by the hydrogen oxides. Nitrogen and halogen elements are crucial for ozone removal in the stratosphere.

One way to investigate the variability of WV is to assess the variability of precipitable WV (*PWV*) which is derived from the Integrated WV (*IWV*) along the path of the balloon sounding, i.e.,

$$IWV = \int_{h_s}^{h_a} \rho_v \mathrm{d}h, \qquad (115)$$

Where ρ_v, and h_a are the density of WV and the top of the troposphere respectively. The IWV is then mapped into PWV using Equation (116),

$$PWV = \frac{IWV}{\rho_w}, \qquad (116)$$

where ρ_w is the density of liquid water. Using the gas state Equation, ρ_v can be obtained from

$$\rho_v = \frac{P_v}{R_v \times T}, \qquad (117)$$

where R_v= 461.495 $JK^{-1}Kg^{-1}$ is the specific gas constant for WV. The partial pressure, P_v of WV which is obtained from Relative Humidity (RH), as expressed in Equation (118)

$$P_v = RH \times e^{-\left[37.2465+0.213166\times T-2.56908\times 10^{-4}\times T^2\right]}. \qquad (118)$$

Here T is the absolute temperature in Kelvin. Since radiosonde observations are discrete data series of temperature and RH at different heights, the atmosphere layer could then be subdivided into discrete layers. This implies that, if the parameter field $\{\rho_v, T\}$ at each layer is assumed to be linear, then Equation (119) could be approximated as

$$PWV = \frac{1}{\rho_w} \sum_j \left(h_{j+1} - h_j\right) \times \frac{1}{2}\left(\rho_v^{j+1} + \rho_v^j\right). \qquad (119)$$

Radiosonde measurements have important applications in verifying WV computed from numerical weather prediction models: they are currently one of the main observation techniques to provide atmospheric water vapour profiles in an operational NWP system. Yang *et al*., (1999) stated that the horizontal scale of WV is, on average, larger than the existing model grid resolution of most the NWP systems. Therefore, even if WV fluctuations with short wavelengths exist, the fluctuations are not measured in most cases and ought to be ignored. This assumption would be acceptable in NWP system. However in space geodetic applications, such assumption could not be favourable at all. In such cases where geodetic stations are close to radiosonde stations, the data from radiosonde measurements is invaluable with regard to verifying WV derived from space geodetic techniques such as GPS, VLBI and WVR for tropospheric modelling of the geodetic delay observable.

Despite the emerging long records of satellite-based observations of atmospheric parameters that describe the structure and dynamics of the atmosphere, radiosonde measurements continue to prove useful in diagnosing the variations in the vertical temperature, humidity and wind speed and direction. As a result, the derived scale height of WV (this is based on the humidity information) distribution can be studied in order to determine the relations between the scale height vertical WV distribution and the rate of de-correlation of the integrated WV over horizontal separation. As reported by Ruf and Beus (1997), this relationship is as a result of the departure from the simple Kolmogorov behaviour of WV turbulence structure, since the horizontal separation approaches the scale height dimension.

Assessment of WV variability could be done using the WV mixing ratio (π) where the relation in Equation (120) and (121) is used to compute the precipitable WV.

$$PWV = \int_{h_s}^{h_a} \frac{\pi \mathrm{d}p}{g}, \qquad (120)$$

Here, dp and g are the incremental pressure change with height in Pascal units and gravititational constant respectively, π is the mixing ratio of WV per gram of air.

In first section of this chapter, a mean vertical profile of WV in the South of the tropical Africa is modelled based on the data from selected radiosonde stations which are the part of SHADOZ network. The general approach in this chapter is to use SHADOZ *in situ* measurements and numerical weather model simulations to investigate the regional

variability of WV. A description of the data sets and the methodology used in obtaining the reference profiles for different location is reported by Sivakumar *et al*., (2010). In the second part of the chapter, the multi-scale organisation of the regional WV variability noted by Botai *et al*., (2010) is reported.

5.2. Vertical profile of WV from SHADOZ data

The global time-mean distribution and large-scale variations of WV are fairly well characterised by satellite data sets, especially by HALOE satellites. In addition, *in situ* and ground-based data sets augment satellite information and present a picture consistent with satellite observations. *In situ* and ground-based data sets are also essential for revealing the behaviour of WV at smaller spatial scales, for long-term monitoring and for validation of satellite data sets. As one of the aims of the research work in this thesis a model (mean) profile for WV in Southern hemisphere latitude using about 10 years (1998-2007) of the SHADOZ balloon borne measurement from Nairobi-Kenya (1.29^0 S; 36.80^0 E; 1795 m), Malindi-Kenya (2.99^0 S; 40.19^0 E; -6 m), and Irene-South Africa (25.90^0 S; 28.22^0 E; 1524 m) was constructed and is reported in the ensuing sections.

The vertical profile of mean WV computed in this section is based on the data obtained from the SHADOZ measurements to obtain a height profile of WV in Southern region of Africa. Such profiles can be used as a reference for comparisons with other measurements such as satellite observations. The details about the data and quality of ozonesonde measurements can be found in several publications (e.g., Borchi *et al*., 2005; Sivakumar *et al*., 2007). At each radiosonde station, about 10 years of ozonesonde data gathered from 1998 to 2007 of Irene, from 1999 to 2006 of Malindi and from 1998 to 2007 of Nairobi stations was used. The measurement data for height region up to 30 km altitude are collected from SHADOZ data which are archived at http://croc.gstc.nasa.gov/shadoz/site2.html/. The SHADOZ measurement contains pressure, temperature, relative humidity and ozone. The mean value of WV mixing ratio in ppmv is found from the relative humidity data using the relation given by Equation (121);

$$\pi = 61121 \times RH \times \frac{e^{\frac{17502T(z)}{24097+T(z)}}}{P(z)}. \qquad (121)$$

Here, π is the mixing ratio, RH is relative humidity, T(z) is temperature in degree centigrade.

The height profile of WV is obtained for the regions of southern latitude hemisphere. The mean values of 10 years of SHADOZ *in-situ* measurement data are further used for making comparison. The mean WV profile calculated from the SHADOZ measurement for stations at Malindi and Irene is displayed in Figure 4.1 which indicates that the variability of WV increases with altitude and that there are notable high variations (this can been seen in terms if the spread or variances in the mixing rations) between 2 and 7 km. It can be observed that above 2 km, the relative variability of WV is greater than 20% in Malindi and 43% in Irene. Such difference indicates that the variation of WV concentration with latitude region. The variation of relative humidity with temperature also contributes to the WV variability in the stratospheric region. The questionable accuracy with altitude can affect the amplitudes of WV variability.

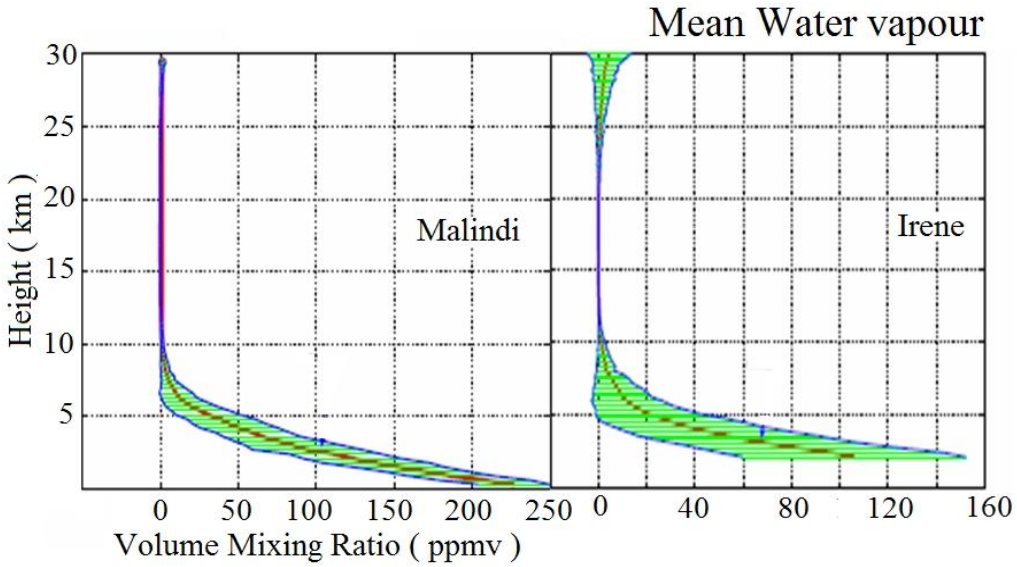

Figure 5.1. Height profile of mean water vapour obtained from SHADOZ datasets.

The mean WV profile of about 10 years of SHADOZ network stations (in this case Malindi and Irene) was derived and has been used to verify WV profiles retrieved from other satellite measurements, such as; GPS and HALOE (see for instance, Sivakumar *et al.*, (2009a)). Additionally, these results would be useful for initializing numerical models and improving parameterizations of radiative and cloud processes. In general, the vertical structure of WV can be used to investigate the occurrence of turbulent events in the atmosphere which could be associated with local weather conditions and passage of fronts. Refractivity profiles computed from radiosonde data could also be used to describe the turbulent fluctuations through the analysis of the structure constant. These measurements have vital applications in

assessing the occurrence of ducting conditions which could have adverse effects in geodetic observations (especially microwave measurements) as well as radio communication.

5.3. Multiscale organisation of WV in mid- and low-tropical Africa

A number of ground based and space borne remote sensors are available to provide the vertical and horizontal profiles of WV, for e.g. radiosondes, Light Detection And Ranging (LiDAR-Raman), GPS, VLBI, WVR(Raschke, 2002). Furthermore, considering the high temporal and spatial variability of WV, depicting how the WV fluctuations are organised into diurnal, synoptic, seasonal and climatic categories, could provide useful information in meteorology (numerical weather prediction and climatology) and space geodetic studies. The variability of WV is associated with spatial structure and unique modes of variability inherent in the rotated Principal Component Analysis (PCA) of the WV energy spectra. The PCA rotating component of WV captures the dominant modes of the WV in temporal scales with similar spatial organisation (Petr, 2005). This linear transformation of PCA allows easy interpretation of the strongest spatial relationships of WV features that drive atmospheric weather systems over a particular region. Saco and Kumar (2000) used similar methodology to capture the spatial patterns of the coherence in the temporal scales of variability of stream flow response. On the other hand, Schubert *et al*., (1998) separated different temporal scales (by choosing the scales independently) of precipitation using a pre-designed filter.

Multiscale spatial-temporal structures of WV describe the movement of water within and between the Earth's atmosphere, oceans and continents (Trenberth *et al*., 2005; Zveryaev and Allan, 2005). Soden and Fu (1995) assessed the temporal structure of WV using satellite-derived upper-tropospheric relative humidity over the tropical region (30^0 S - 30^0 N) and concluded that a positive relationship between relative humidity and deep convective processes exists. Though, the global spatial distribution and trends in WV are dominated by large-scale dynamics, such as., El Nino- Southern Oscillation (ENSO) rather than the thermodynamics, (See., Zveryaev and Allan, 2005), the linkage between WV anomalies and atmospheric circulation processes is difficult to establish due to the complexity of the spatial-temporal structures of WV. The spatial and temporal variability of WV in the mid- and low-tropical Africa ranges from a few kilometres to thousands of kilometres and from a few minutes to several days, similar to the meso- /synoptic scale processes, respectively (Husak, 2005). Therefore, the analysis of the correlations of WV between the spatial grids is of great

practical importance for studying the conditions that lead to the development of hazardous weather systems.

In order to understand the feedback processes operating within the mid- and low-tropical Africa, a robust methodology of examining the spatial-temporal structure is required. In this study, the spatial and temporal organisation of WV is analysed simultaneously using orthogonal wavelet transformation which allows for calculating the total energy of WV by summing up their individual scales either in spatial or temporal regimes. The spatial-temporal fluctuations of WV are investigated which further helps to understand the regional weather patterns in the East, Central and Southern Africa. Results obtained from this study would also form the basis for future comprehensive analysis of the relation between WV variability and the associated atmospheric weather systems as well as any other forcing mechanisms observed in the low- and mid-tropical Africa. In addition, the goal of this study is to understand the mechanisms driving WV variability and its link to the climatic variables which are essential for accurate modelling of the regional hydrological cycle.

The main data source used in this particular study is about 8 years of upper air radiosonde/ozonesonde data archived at the SHADOZ station network consisting of Ascension, Irene, Reunion and Nairobi). The publication by Thompson *et al.*, (2003) provides further details about the SHADOZ network. The geographical positions of the SHADOZ stations and the details about data considered for the present study are tabulated in Table 5.1.

Table 5.1 SHADOZ network stations used in the current study.

Station	Longitude	Latitude	Elevation	# of Launch	Time Period
Nairobi	36.80 E	1.27 S	1795.00	370	Jan 1998 to Aug 2007
Reunion	55.48 E	21.06 S	24.00	293	Jan 1998 to Oct 2006
Irene	28.22 E	25.90 S	15.24	232	Nov 1998 to Dec 2006
Ascension	14.42 W	7.98 S	91.00	397	Jan 1998 to Dec 2006

The SHADOZ stations were configured to obtain mean information of WV over four grids, and designated WV_g time series. In addition, a time series of integrated WV (hereafter, WV_{ncep}) was constructed from the average over four grid points of each SHADOZ station

(See, Figure 5.2), using the NCEP/NCAR reanalysis (Kalnay *et al*., 1996) data whose zonal and meridional spatial resolution is 2.5^0.

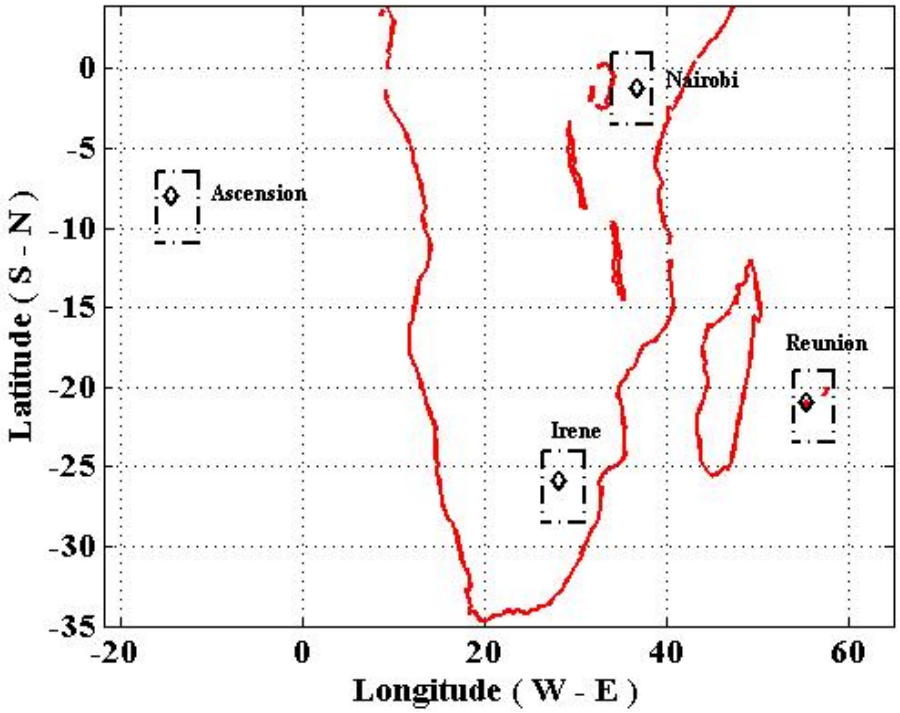

Figure 5.2. The SHADOZ stations with the corresponding grid boxes formed by the closest four grid points of reanalysis data from the National Centres for Environmental Prediction and Atmospheric Research.

For each NCEP/NCAR grid point, the temporal series of WV_g is tested manually for inherent normal distribution and then transformed by Box-Cox transformation (Box and Cox, 1964) which ensures a normal distribution. Prior to the Box-Cox transformation, the WV data sets are detrended. Further, in order to account for latitudinal distortions, each point of WV_g anomalies are weighted by the square root of the cosine of latitude (North *et al*., 1982). The resulting time series has been linearly detrended and subjected to non-decimal Haar wavelet transformation (Lindsay *et al*., 1996) to capture localised temporal fluctuations.

In contrast to the Fast Fourier Transform (FFT), the wavelet power spectrum (absolute value squared of the wavelet transform) provides the total energy of the WV_g time series at a given scale while FFT gives information about what frequencies are present in the signal, but lacks the ability to correlate the frequencies with the time of their presence. In general, the difference between Fourier and wavelet coefficients is that the former is influenced by a function on its entire domain (global measure), while the latter is influenced

by local features. The wavelet power spectrum is therefore chosen in this study as a better measure of variance attributed to localised events. The wavelet coefficients at each time scale were used to compute the energy spectrum per spatial scale to form a temporal scale series (S) over the grid points (G) to form a matrix D with dimensions S × G.

The calculated WV_g values from radiosonde measurements at the SHADOZ stations and from the NCEP/NCAR reanalysis data are plotted in Figure 5.3 (gridded NCEP/NCAR reanalysis data is plotted in the left panel and while the four SHADOZ stations; (a) Nairobi (b) Ascension (c) Irene and (d) Reunion are plotted in the right panel. It is clear from the figure that the NCEP/NCAR reanalysis data exhibit a cyclic trend over the period of observations, whereas such cycles are not evident in the SHADOZ observations. The difference might be due to the coarse latitude and longitude resolution of NCEP/NCAR data that were averaged over the station grid box, while each SHADOZ station corresponds to a particular location. In addition, sensitivity of the balloon measurements may have contributed to the differences in WV_g from the two measurements. Further to this, NCEP/NCAR reanalysis data are based upon simulation with possible inherent biases. The differences between the NCEP/NCAR reanalysis data and SHADOZ station data were calculated for each station. Results concluded that the Irene and Reunion stations have higher mean deviations (~ 40 mm) while the Nairobi and Ascension stations have a mean WV_g deviation of ~ 30 mm (this is depicted in Figure 5.4).

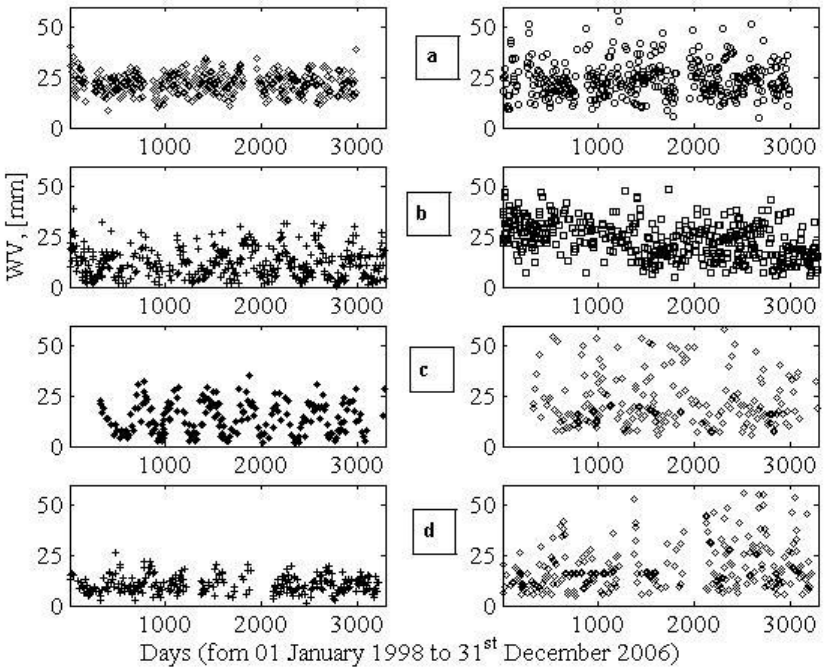

Figure 5.3. Daily integrated spatially averaged Water Vapour, WV_g [mm] for (a) Nairobi (b) Ascension (c) Irene and (d) Reunion.

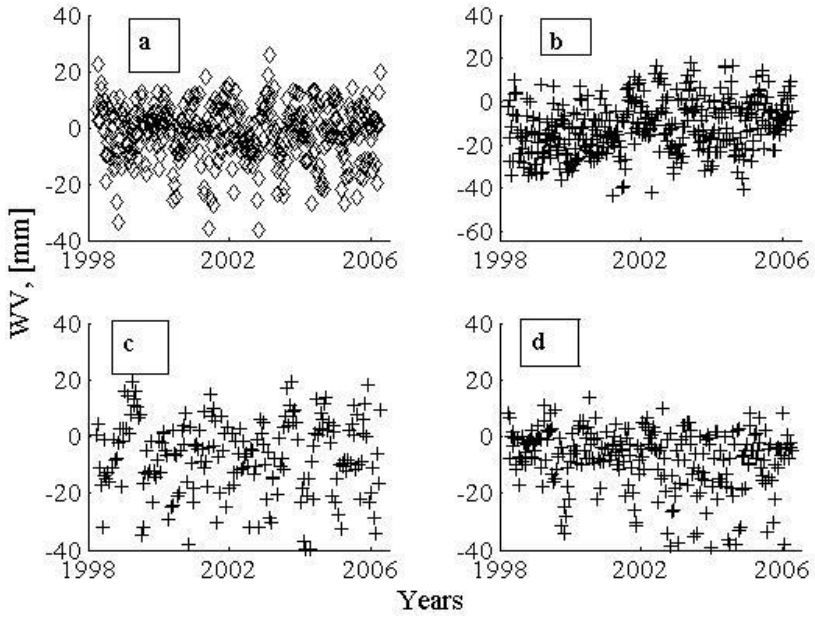

Figure 5.4. Differences of Water Vapour in mm, calculated from four SHADOZ stations (a: Nairobi, b: Ascension, c: Irene and d: Reunion) and the gridded NCEP/NCAR reanalysis.

It is understandable from Figure 5.4 that the variability pattern in WV is difficult to discern from the time series. The excursions from the mean signify the presence of exogenous processes that play a significant role in WV_g fluctuations. These stochastic processes are manifestations of local weather system processes (eg., convection, precipitation). To better understand these fluctuations, the nature of distribution of WV needs to be known. The standard probability distributions of WV_g are used and are compared to the normal Guassian distribution. The normal Guassian distribution has been generated by selecting random data sets.

To assess the normal (Gaussian) distributions of WV_g, the QQ-plots were drawn between the Guassian generated and WV_g probability distribution. A linear variation in the QQ plot could signify a normally distributed time series. This distribution has been tested, individually for each station as shown in Figure 5.5. The obtained regression coefficients illustrate that the SHADOZ station; Ascension has high linearity in comparison to that of Nairobi, Irene and Reunion. A maximum non-linear fluctuation component of ~10% was obtained for Reunion. On the other hand, Irene, Nairobi and Ascension have values of ~8%, 5% and 1%, respectively. The results reported here imply that WV over Ascension follow a normal distribution and appear not to be affected by non-linear local weather conditions.

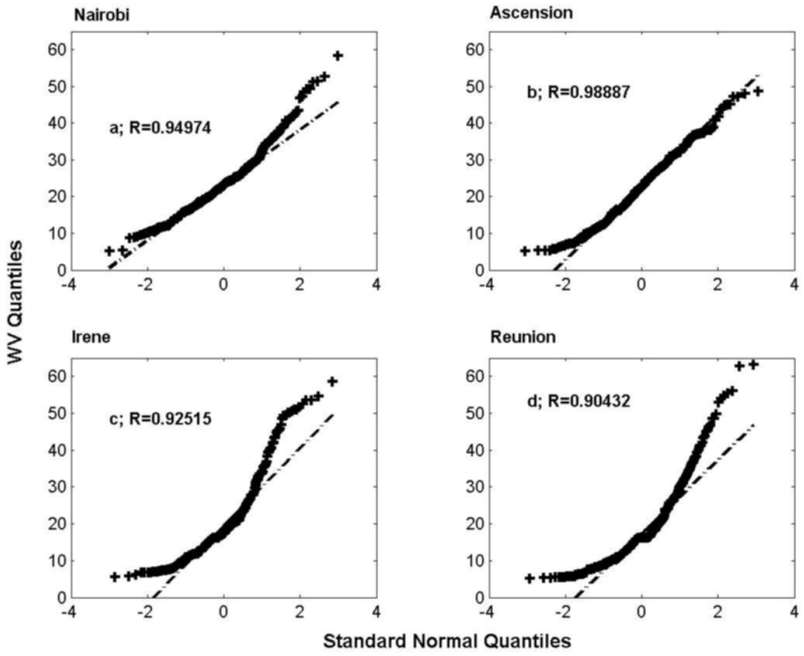

Figure 5.5. Quartile-quartile (QQ) plot of a Gaussian distribution, and the probability distribution of WV_g at the four SHADOZ stations under consideration.

To study the local temporal fluctuations of WV, the Haar wavelet transform of maximum overlap discrete technique has been applied. The wavelet coefficients derived from the WT of WV are used to assess, capture and discriminate between the different modes of local fluctuations in time series in the frequency-time space. Each SHADOZ stations' mean WV data were grouped into years and months. The corresponding monthly mean over the 8-year period of data is subjected to WT after performing de-trending. Figure 5.6 depicts the obtained wavelet coefficient (amplitude) at different temporal scales of 3, 8, 12 and 36 months (from bottom to top) or the time period of oscillation of WV_g at a given location.

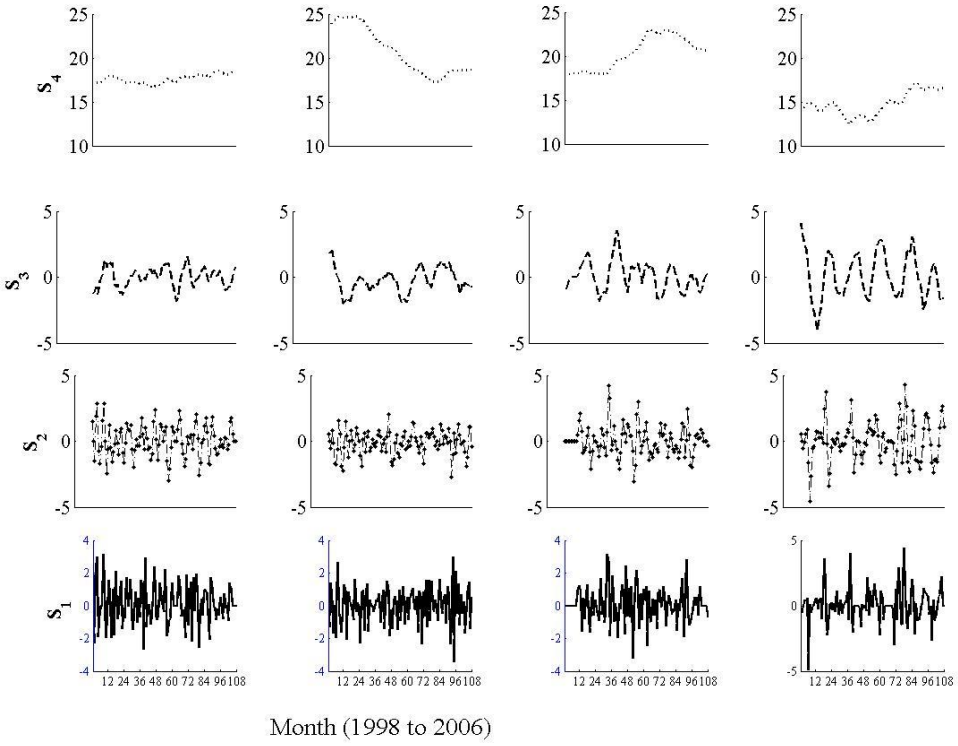

Figure 5.6. Haar wavelet spectra at different scales and at different station locations (Ascension, Reunion, Irene and Nairobi) - from left to right, respectively.

The relation between the period of oscillation of WV_g fluctuations and the wavelet scale index is obtained from the equation $s=2^{j-1}$, where the j^{th} index denotes the period. The method of deducing the wavelet co-efficient is reported by Percival and Walden (2000). Although, scale-1 (~3 month) does not offer any clear information on the fluctuations, other higher order scales show a significant oscillation at all the stations. Notably, the annual oscillation

(scale-3) is clearly distinguishable at all the stations. In comparison to all the stations, Nairobi exhibits a clear cyclic variation. For almost all the stations, the scale-4 (3-year) component does not complete one period of a cycle, inferring that the periodicity is more than 12-years. It is noted here that the maximum possible number of scales obtained depends on the length of data period used. The log-log plot of the wavelet energy which is depicted in Figure 5.7 reveals an approximate power law scaling at lower time scales, which break down at high time scales. These results are consistent with the results reported by Lay (1997) and Cho *et al.*, (2000). At high time scales, the break down in the linear relationship is associated with response of WV_g fluctuations to tele-connection patterns such as the influence of ENSO in the low and mid- tropical Africa; see for example Trenberth *et al.*, (2005).

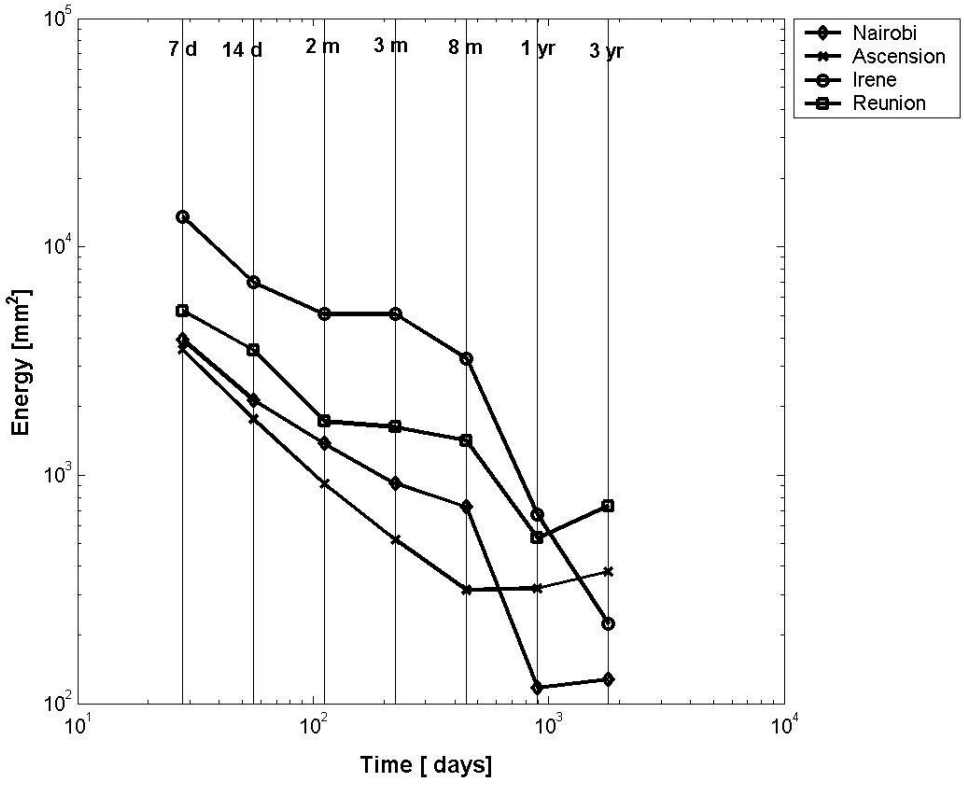

Figure 5.7. Approximate power law scaling of the WV derived wavelet energy.

PCA has been determined for the wavelet coefficients of all the four stations, and the calculated variance is presented in Figure 5.8. The first three variance components account for 98% of the WV_g variations. The first component represents high frequency temporal fluctuations (monthly time scales) and accounts for 67% of the variability. Component two

represent the variance associated with annual fluctuations, and accounts for about 27% of the WV_g fluctuations. About 4% of WV_g variability is associated with low frequency fluctuations (1< timescales < 9-year).

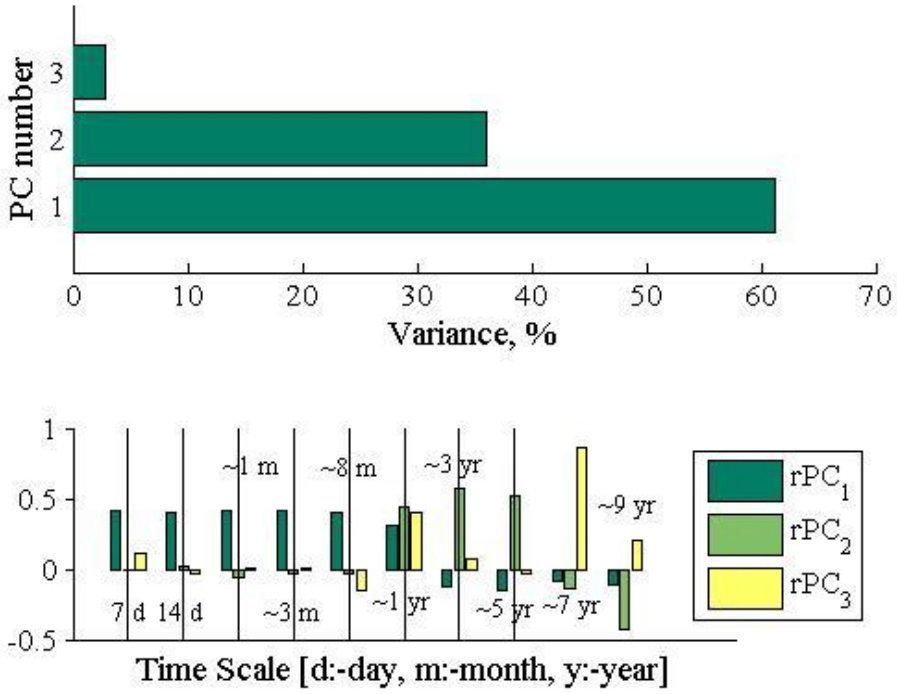

Figure 5.8. Co-variance of the Principle Component Analysis (PCA) components obtained from the four stations under consideration.

Decadal fluctuations cannot be inferred convincingly due to the short time-span of the data (8-years from 1998 to 2006). These results indicate that there is a distinct spatial structure for each short term temporal WV_g variation in the low and mid-tropical Africa region that could be attributed to synoptic/seasonal-scale weather systems, which is consistent with findings from Husak (2005) who reported that seasonal weather systems, topography, the Inter-Tropical Convergence Zone (ITCZ) and monsoon winds affect WV distribution and fluctuation. Jin *et al*. (2008) also reported that the variability of WV in China is dominated by seasonal variations. In addition, the spatial distribution of WV dependence on the thermodynamic relationship between WV and temperature has been reported in Zveryaev and Allan, (2005).

The marked differences between WV fluctuations at longer timescales could be attributed to the WV response to tele-connection patterns such as ENSO in the low and mid-

tropical Africa; this is in line with the findings of Trenberth *et al*., (2005) who had indicated that the variability of WV is dominated by the evolution of ENSO. This link shows a strong relationship over the oceans between WV and Sea Surface Temperatures (SSTs). Further, the African low and mid- latitude WV has a strong link to rainfall due to its close association with the mean wind flow, and convergence of moisture by trade winds as well as their links to SSTs. In addition, the correlation analyses performed between surface temperature and WV show that a link exists between WV anomalies and regional air temperature variations with marked seasonal dependence (the results are not presented here) over all four SHADOZ stations.

5.4. Concluding remarks

In an effort to analyse regional spatial and temporal features of WV variability over low and mid-tropical Africa, NCEP/NCAR reanalysis data around the SHADOZ network of four stations were used to calculate spatially averaged WV (VW_g) over the period from 1998 to 2006. The WV_g was calculated as the spatial average of the four closest NCEP/NCAR grid points around the SHADOZ stations to form grid cells. Based on these grid cells, data from NCEP/NCAR reanalysis data were also used to calculate the vertically integrated column of WV over the same time epoch for comparison. For the first time, the WV_g variability in the low and mid-tropical Africa was analysed using *in-situ* data from the SHADOZ network. The results show that WV exhibits high frequency fluctuations in the wavelet space. Common to the entire SHADOZ network considered in this study is the pattern of temporal WV_g fluctuations with monthly time scales dominating. This dominant variance appears to be associated with locally driven WV variations such as the local weather systems. Our results show that WV also exhibits the power law scaling in the wavelet energy. The approximate log-log linear relationship at smaller temporal scales that breaks down at synoptic scales suggests that the energy-times spectra of WV on different temporal scales are correlated. Furthermore, based on PCA, three dominant modes emerge. These modes explain ~ 98% of the total spatial variance of the normalized energy in WV fluctuations. To validate the current findings, future studies will involve the use of observations such as HALOE (Russell *et al*., 1993) and regional numerical simulation model data sets to determine the temporal and spatial organisation of PWV data at finer spatial and temporal scales.

In general, from the current analysis, results indicate that WV exhibits high frequency fluctuations in the time-frequency space. For the entire SHADOZ network considered in this

study, monthly time-scales dominate the pattern of temporal WV fluctuations. This dominant variance appears to be associated with locally driven WV variations such as the local weather systems. Our results show that WV also exhibits the power law scaling in the wavelet energy. The approximate log-log linear relationship at smaller temporal scales breaks down at synoptic scales. This behaviour suggests that the energy-times-frequency spectra of WV on different temporal scales could be correlated.

Mechanisms that influence the global WV distribution can also be assessed in the context of their role in regional tropospheric WV fluctuations. Tropospheric regional WV is influenced by the dynamics and the seasonal changes in temperature. The ability of the atmospheric dynamics influencing water follows from the steep slope of the Clausius-Clapeyron equation: this relates to the rapid increase in water holding capacity of the atmosphere. The environmental lapse-rate then rapidly decreases the WV with altitude (at a scale height of ~2 km).

Our understanding of factors controlling long-term changes in tropospheric WV is inadequate to explain the observed variations or to provide good projections in the near future. In the lower stratosphere, the observed changes in WV could be linked to changes in other greenhouse gases, such as, ozone and methane. It is fairly clear that atmospheric processes (e.g., transport, convection, and clouds) are involved in determining the distribution of tropospheric WV, but their influences are very difficult to quantify. Since it is also difficult to predict how these might change in response to natural and human-induced climate change, future changes in the distribution and variability of tropospheric WV still remain unknown.

It is generally accepted that there is a cancellation effect between increasing humidity and decreasing temperature with height. Furthermore, the uncertainty over the precise WV concentration in the troposphere, and even greater uncertainty over trends in upper tropospheric humidity and temperatures, makes it impossible to carry out a firm quantitative estimate of the tropospheric radiative consequences of long-term changes in tropospheric WV. However, present knowledge indicates that the radiative response to long-term changes in upper tropospheric WV will almost certainly be significant. Due to this, the following recommendations are proposed. Firstly, upper tropospheric WV should be monitored with a view to determine long-term variations. More observations of the tropical tropopause region (15-20 km), by both *in situ* and remote sensing methods, are needed to improve our understanding of stratosphere-troposphere exchange. Furthermore, in order to assess robust estimates of WV variability, complementary observations and techniques are recommended.

Second, in order to have a balanced view of all the mechanisms that drive WV variability in the troposphere, combined measurements of WV, cloud microphysical properties, and chemical species are recommended.

6. Non-linear and nonstationary processes in geodetic TD/WV

So, Nat'ralists observe, a Flea
Hath smaller Fleas that on him Prey,
And these have smaller Fleas to bit' em,
And so proceed, proceed ad infinitum
- J'onathan Swift (1667-1745)

This Chapter presents results and discussion on the non-linear properties in the fluctuations of TD/WV based on the statistical information derived from DFA. In order to account for the non-linear and nonstationary properties in TD/WV fluctuations, the EMD/EEMD is used. Self-similar behaviour in TD/WV is assessed and found to be present. The lasting periods of synchronization with fluctuating phase coincidence (and correlations by pairs thereof) between different oscillation modes of WV and atmospheric mean temperature demonstrate that fluctuations (which are non-linear and depicted in the phase shifts) in WV and temperature are driven by common underlying processes (exhibiting possible stochastic resonance). Assessing the significance of non-linearity and nonstationarity in geodetic data analysis is demonstrated by the introduction of non-linear function to account for azimuth dependent atmospheric range correction. Using a one-month SLR data, more than 15% improvement of the O-C residuals is achieved for both LAGEOS 1 and 2.

6.1. Introduction

In the research reported in this Chapter, the emphasis is on the analysis of the variability of geodetic tropospheric parameters by extracting the local time/frequency scales of variability embedded in the geodetic time series. A data adaptive decomposition (by using the EMD and/or EEMD) is considered, whereby complete, orthogonal and local components are obtained. Often EMD/EEMD decomposes the time series into IMF according to the levels of their local oscillation or frequency. The EMD effectively captures the non-linear characteristic in respect of the amplitude and frequency modulation with local time scales. In

this regard, the local time/frequency scales are defined by the instantaneous frequency. Once IMFs are obtained, Hilbert Transform spectral analysis is performed on the time series to provide the temporal-frequency variation of information which is a key component of time-frequency analysis for nonstationary geodetic-tropospheric time series.

In particular, a non-decimated Haar wavelet transform and EMD/EEMD are used to analyse nonstationary processes in WV and other derived tropospheric parameters such as tropospheric delay due to WV. In the first part, geodetic WV measured at HartRAO is tested for nonstationarity and non-linearity using DFA in the wavelet space. Here the geodetic WV was decomposed into the wavelet space and thereafter, the characteristic fluctuation properties of the power spectra are investigated using DFA. Furthermore, the EEMD was used for correlation analysis of some of the important tropospheric parameters such as wet zenith delay and mean tropospheric temperature that affect the accuracy of the geodetic delay observable over geodetic stations; here, we have used the geodetic station at HartRAO, South Africa. The correlation analysis considered here illustrate the synchronisation or the degree of closeness of the modes in WV and mean atmospheric temperature time series derived from the phase variance between the time series (Botai *et al*., 2009b).

6.2. Nonstationary processes in tropospheric WV using wavelet analysis

Non-linear processes exist in nature; examples of the dynamical systems that exhibit non-linear and nonstationary characteristics include the Earth's climate system (Rial *et al*., 2004), river flow and discharge (Montanari *et al*., 2000), atmosphere (Ivanova, *et al*. 1999; Ausloos *et al*., 2001). If a geophysical field such as PWV is decomposed additively into structural components such as the trend, the cyclical and seasonal and irregular components, then the coherence between the properties of the observed PWV series and those of the structural components could be assessed. The characteristics structure in the components of the time series could be used to evaluate nonstationarity manifested in the second order properties such as the variance and the mean. This may be an indication that the dynamical processes in the geophysical field are driven by complex processes as reported by Taqqu *et al*., (1995), Verdes *et al*., (2001) and Hu *et al*., (2001).

In the current analysis, the presence of memory in the tropospheric PWV is assessed using statistical theory. Using a global PWV parameter (gPWV) estimated at HartRAO, the PWV time series during 2000-2006 was reconstructed with sliding window using Singular

Spectrum Analysis (SSA). See Ghil *et al*., (2001) for a detailed account on SSA. The gPWV is defined (here) as the daily mean PWV value computed from two geodetic techniques (GPS and VLBI) and NWP models. The wavelet transform of the reconstructed gPWV time series was computed using the non-decimated Haar wavelet technique (Percival *et al*., 2000). As described in Chapter 3, the presence of discontinuities due to boundaries in the data is accounted for by reflecting the time series onto the last observation. This method of handling discontinuities in the geodetic PWV is tractable because the sample mean and variance are not affected and the seasonal patterns in PWV time series exhibits are preserved.

The DFA and wavelet joint estimator were used to calculate the second-order parameters (e.g., the mean, variance and scaling exponent measure the scaling behaviour) that are used to infer the (non-)stationary properties in the reconstructed time series. The purpose of wavelet joint estimator is to eliminate the deterministic trends and non-discontinuous changes in the second-order properties in the WV fluctuations.

Figure 6.1 depicts the reconstructed gPWV computed from leading principal components projected onto a reduced subspace with minimum redundancy in SSA. In this way, the gPWV was smoothed and any noise in the time series eliminated through filtering. This method of filtering preserves the second order statistical parameters which characterise nonstationarity and/or local stationarity.

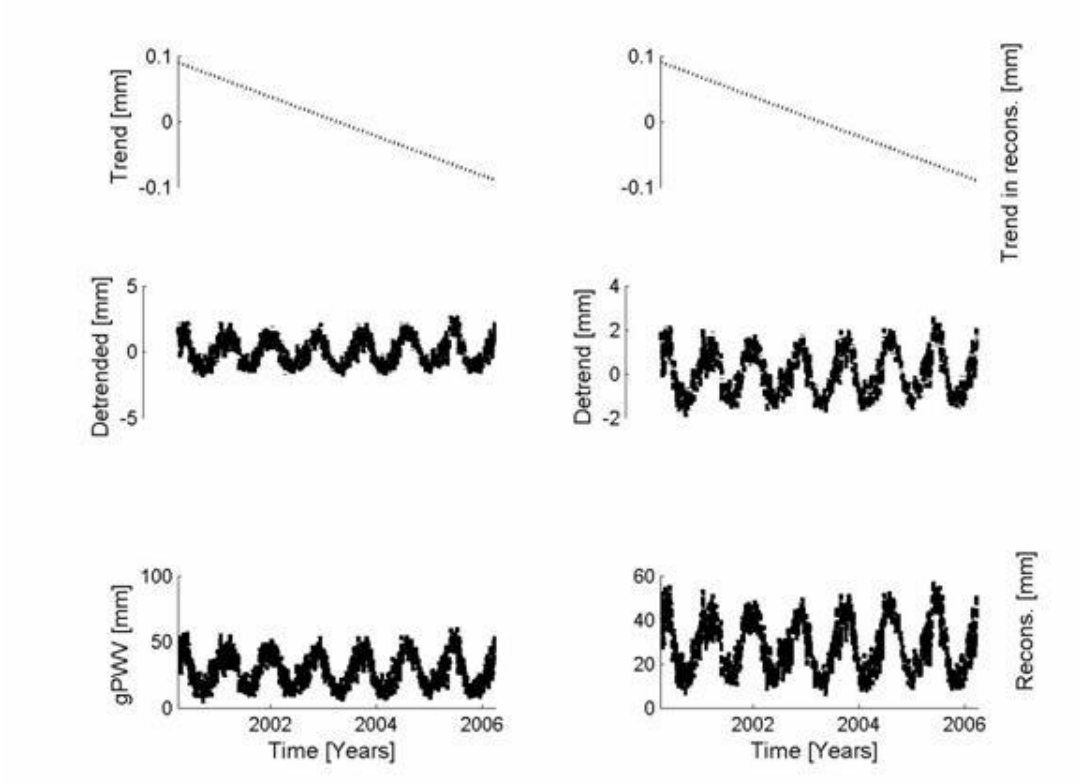

Figure 6.1. PWV time series. Left: Non-detrended PWV time series. Right: PWV derived from the singular spectrum analysis.

Figure 6.2 depicts the 'detail' (even scales) of the gPWV in the wavelet space derived from the Maximum Overlap Discrete Wavelet Transform (MODWT) analysis (see Whitcher 1998; Percival *et al.*, 2000). The MODWT technique allows the non-linear and nonstationary structure (e.g., trends, jumps and spikes) in the data to be examined in the wavelet space. The illustrated temporal scales correspond to the physical time scales 1 week, 2 month, 4 months, and 8 months respectively. In order to get a better understanding of the inherent PWV fluctuation patterns, the individual frequency levels are separated out and plotted as functions of time as shown in Figure 6.2. Each frequency band spans over the time span with gradual decreasing time resolution. From Figure 6.2, we suppose that some contaminating observations other than the gPWV observations (the outliers) are embedded in the high modes of oscilations (Greenblatt, 1994). Therefore caution should be taken when interpreting the oscillatory components, especially the high frequency wavelength coefficients. It is evident from Figure 6.2 that gPWV exhibits temporal nonstationary characteristics.

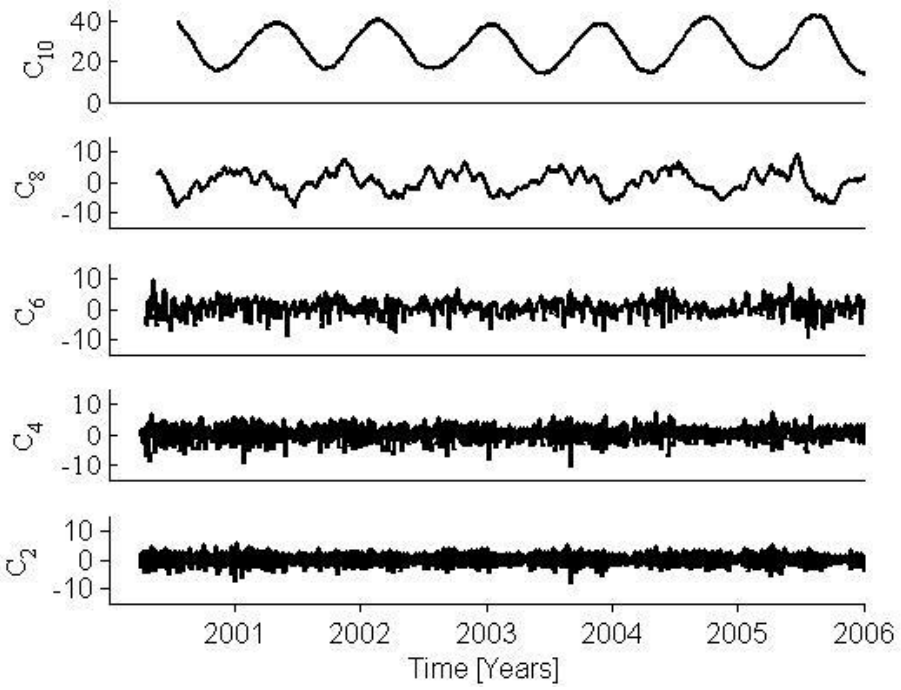

Figure 6.2. Haar wavelet spectra of the reconstructed gPWV at different even-scales.

In order to study some aspects of self-similar behaviour present in the reconstructed gPWV times-series, local linear least square function was fit to a standardised integrated time series constituted from segments of gPWV. The average fluctuation for each segment was used to compute the scaling exponent. Figure 6.3 indicates that a single linear fit on the gPWV power spectral density has different values of β at different scales (the wavelet scale and β are linearly dependent). As depicted in Figure 6.3, the SS and LRD measures are dependent on the 'detail' of WT.

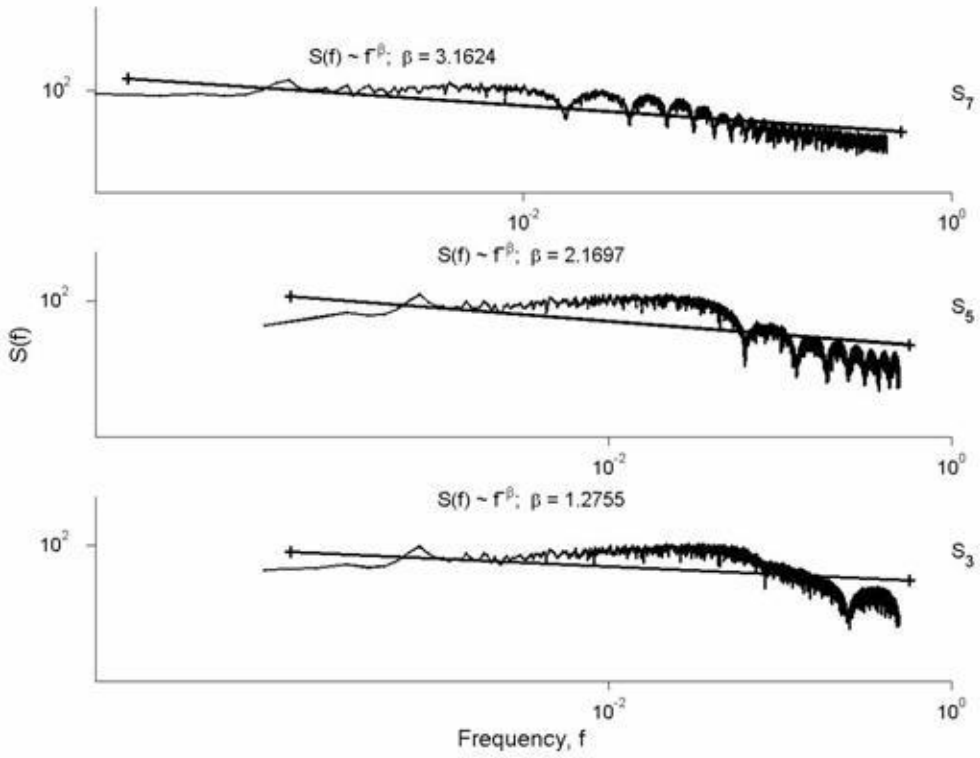

Figure 6.3. The power spectral density of gPWV at different scales.

Figure 6.4 illustrates that the root mean square fluctuations and the sliding window width have different scaling exponents (α), suggesting that the long-range power-law correlation in gPWV (i.e. $\alpha > 0.5$) is dependent on the physical time scales. These results indicate that atmospheric processes manifested in the PWV fluctuations are long-range correlated.

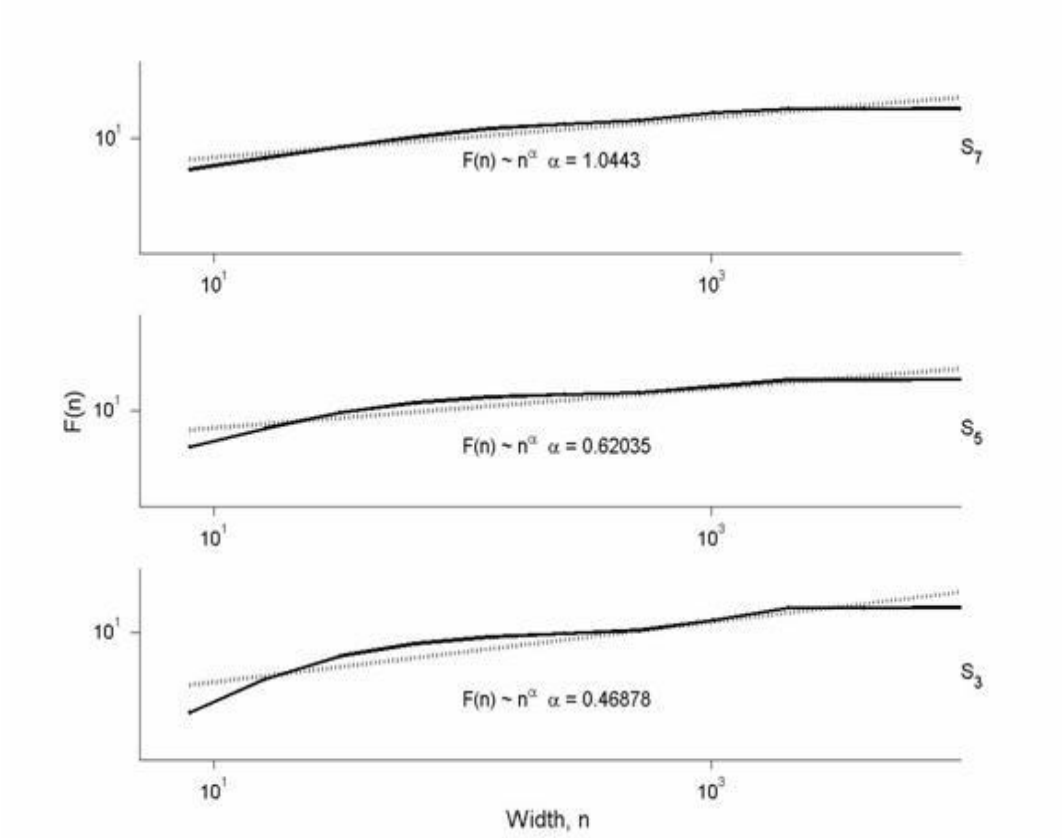

Figure 6.4. Scaling behaviour of the reconstructed gPWV time Series.

The temporal variation of the mean, variance and α of gPWV is depicted in Figure 6.5. The temporal variation in the mean and variance (and also correlations) can be used to measure the presence of SS and/or LRD. In our analysis using the wavelet estimator of Abry and Veitch, (1998) of second-order parameters, the test for SS and/or LRD properties in gPWV time-series showed that the mean and variance are not time invariant as illustrated in Figure 6.5. The experimental analysis for stationarity reported here considered four vanishing moments, the sub-series scale $j_{1,2} = \{2:4\}$ while the scales for the whole series was set to $j_{1,2} = \{2:8\}$. This shows that tropospheric PWV has SS and LRD properties.

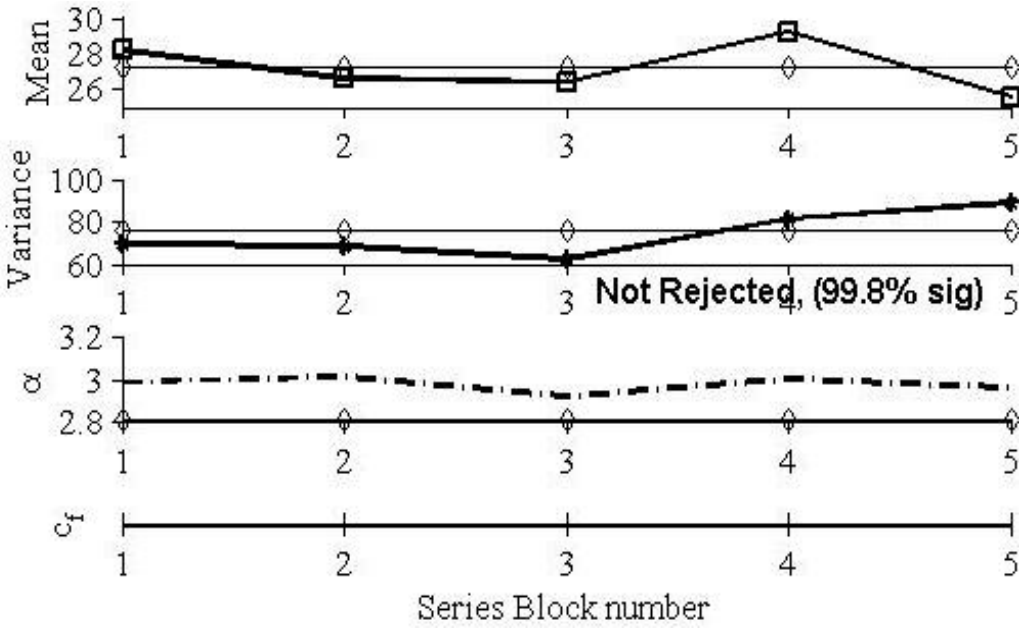

Figure 6.5. Wavelet estimator for SS and LRD at the 10th scale.

6.3. On the noise-Assisted geodetic data analysis

Analysing non-linear and nonstationary data series using the HHT is a new cutting-edge methodology that is widely used in many scientific fields such as finance (e.g., Huang *et al*., 2003b), biomedical applications (e.g., Huang *et al*., 1999b), natural sciences (e.g., Salisbury and Wimbush 2002; Pan *et al*., 2002) and engineering sciences (e.g., Huang *et al*., 2005). For example, Salisbury and Wimbush,(2002) used Southern Oscillation Index (SOI) data and applied the HHT technique to determine whether the SOI data are sufficiently noise free; if useful predictions can be made and whether future ENSO events can be predicted from SOI data. Pan *et al*., (2002) used HHT to analyse satellite scatterometer wind data over the North-Western Pacific and compared the results to Vector Empirical Orthogonal Function (VEOF) results. More recently, Pegram *et al*., (2008) suggested an improvement to the original EMD algorithm using rational splines and the flexible treatment of the end conditions and applied it to rainfall time series analysis.

Figure 6.6 depicts the temporal variations of the daily-averaged WV, zonal and meridional gradients derived from ECMWF data at HartRAO over the period from 2006-2009. From Figure 6.6, it is evident that the WV and linear horizontal gradients time series show that fluctuations occur on time scales much shorter than the length of the entire time series. The short-time fluctuations are then decomposed into IMFs. In this report, IMFs derived from WV daily time series data are given in Figure 6.7. The current WV data resulted

in 11 IMFs (the trend inclusive). These oscillations are not visible in the WV raw data due to their superposition. Looking at individual IMFs, it is possible to decipher information regarding WV variability at different time-scales. In particular, δ_{3-5} clearly shows intra-seasonal modes whose period of oscillation varies from days to months. However, δ_6 shows an almost periodic envelope with amplitudes having seasonal dependence. When examining the low frequency modes of WV fluctuations depicted in Figure 6.7, we notice that δ_8 exhibits regular intra-annual fluctuations that corroborate with the sub-tropical seasons. This component is also visible in the raw WV data.

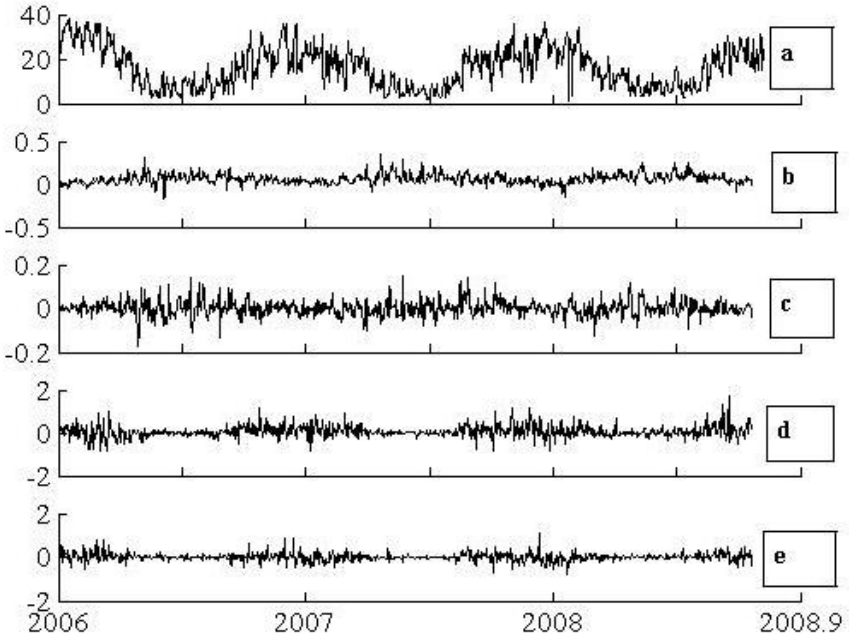

Figure 6.6. Time series (time is plotted in the x-axis) of a) Water vapour, mm; b) Meridional hydrostatic; c) Zonal hydrostatic; d) Meridional wet and e) Zonal wet; linear horizontal gradients , mm/degree.

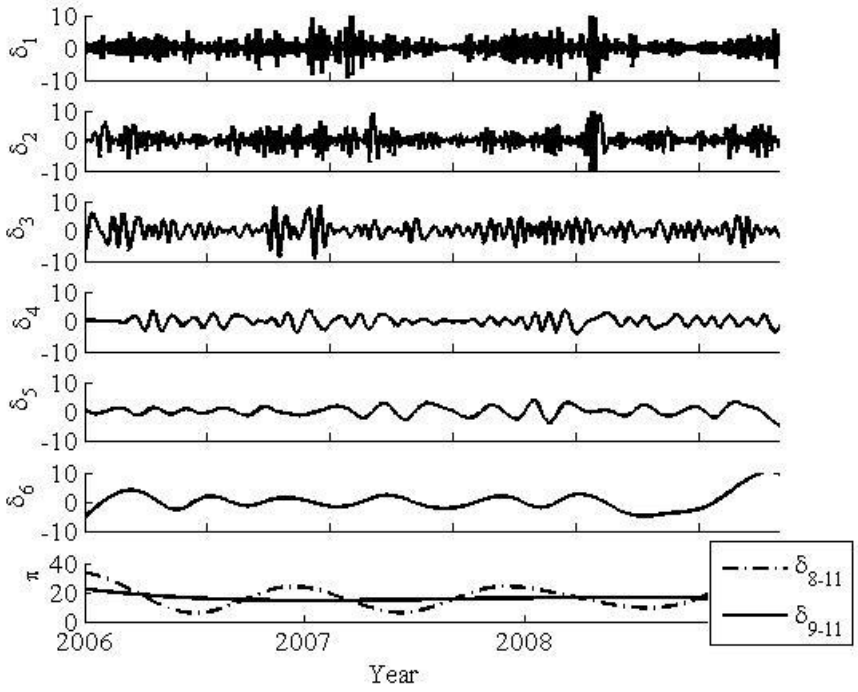

Figure 6.7 WV derived IMFs ($\delta_{1\to6}$). The bottom panel illustrates the adaptive trends.

Given that the trend is a local non-oscillatory function defined for a local time scale, the trend is also one of the many local properties of the data, (see Zhaohua *et al*., 2007). From the tropospheric WV, we see oscillatory modes are superimposed onto a monotone (oscillatory and decreasing) base function. In particular, various trends such as diurnal, intra-annual and annual trend are plotted in Figure 6.8. From Figure 6.8, the overall adaptive trend (π_{11}) is approximately identical to the linear trend (see bottom panel). The overall adaptive trend is derived from the entire data span. Annual and intra-annual trends are formed from the combination of trends derived from δ_{8-11} IMFs.

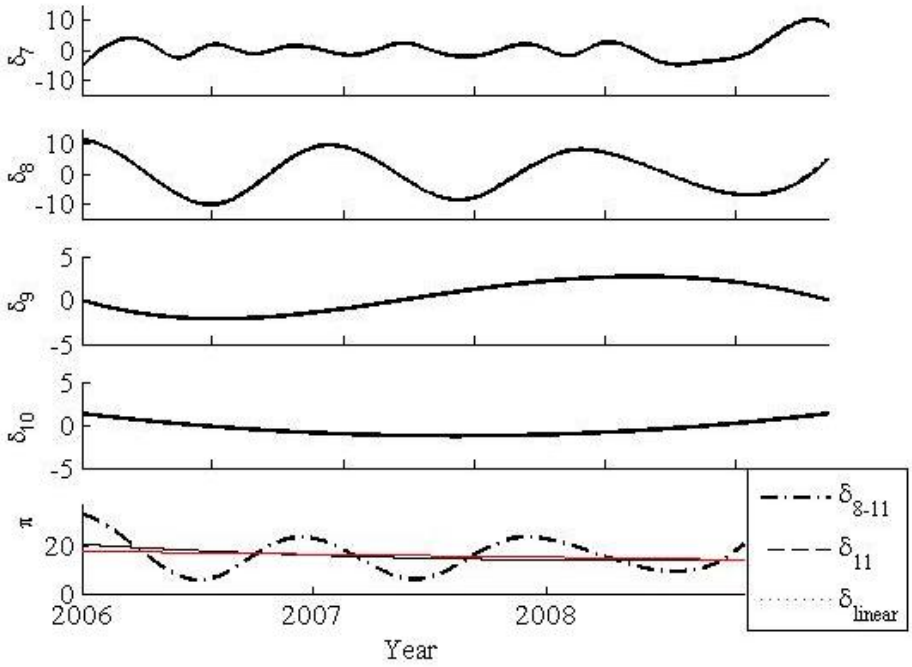

Figure 6.8. Low frequency IMFs and linear and adaptive trends at different time scales.

To completely characterise the tropospheric WV and linear horizontal gradients fluctuations, local time/frequency scales (defined by the instantaneous frequencies of IMFs) of the decomposed time series were extracted. The instantaneous frequencies of the 6^{th} IMF for WV and the zonal linear horizontal (hydrostatic and wet) gradients are plotted in Figure 6.9.

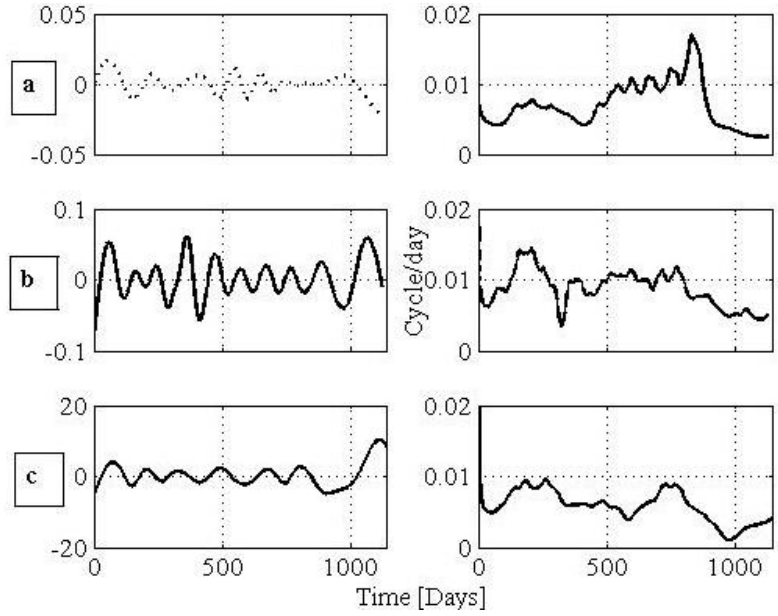

Figure 6.9. The Instantaneous frequency of the 6^{th} IMF. From top left to right: a) Zonal hydrostatic gradient and instantaneous frequency, b) Zonal wet gradient and instantaneous frequency and c) WV and instantaneous frequency.

The probability distribution of individual IMFs examined using WV and the linear horizontal components. The results show that the probability distribution is approximately normally distributed (see Figure 6.10 to Figure 6.12; the fitted normal distribution curve is also depicted). From the current variables, it can be noticed that, the first three IMFs derived from WV, hydrostatic and zonal linear horizontal gradients exhibit normal probability distribution. These results are consistent with those reported in Wu and Huang, (2004) and they are also in agreement with the central limit theorem.

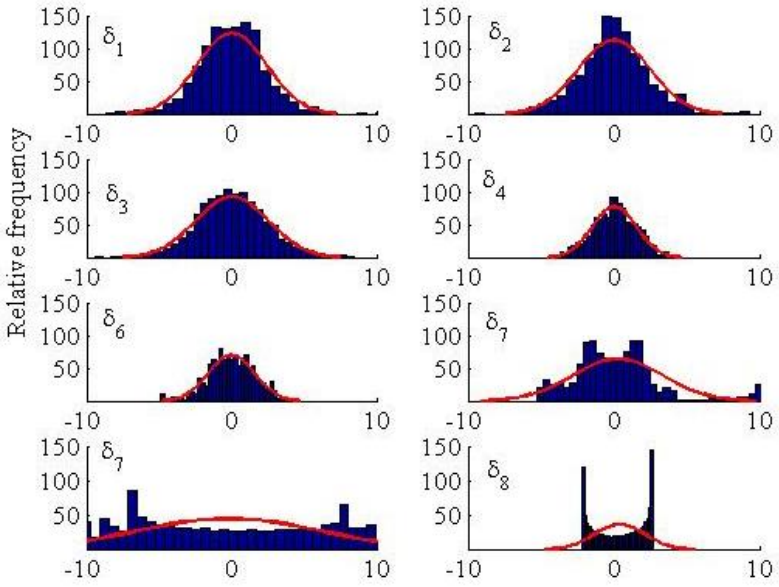

Figure 6.10. Probability distribution of WV derived IMFs.

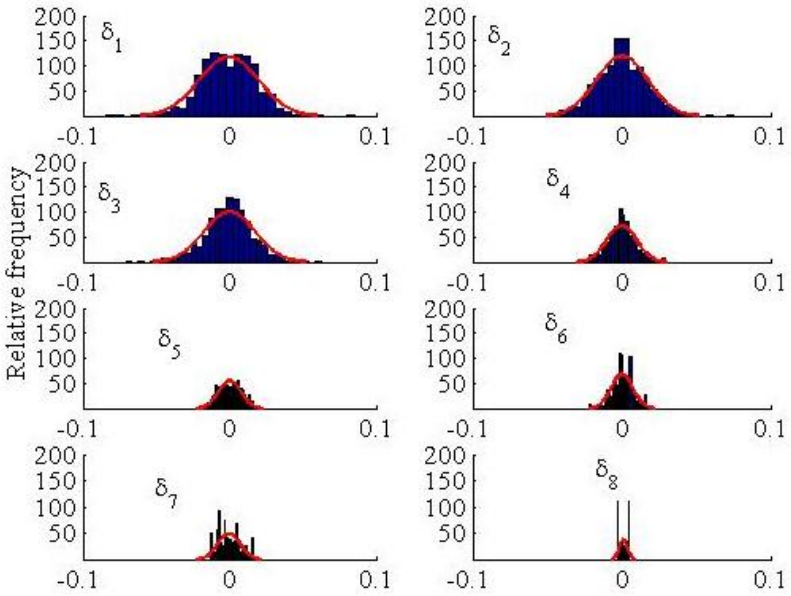

Figure 6.11. Probability distribution of the zonal linear horizontal hydrostatic gradients.

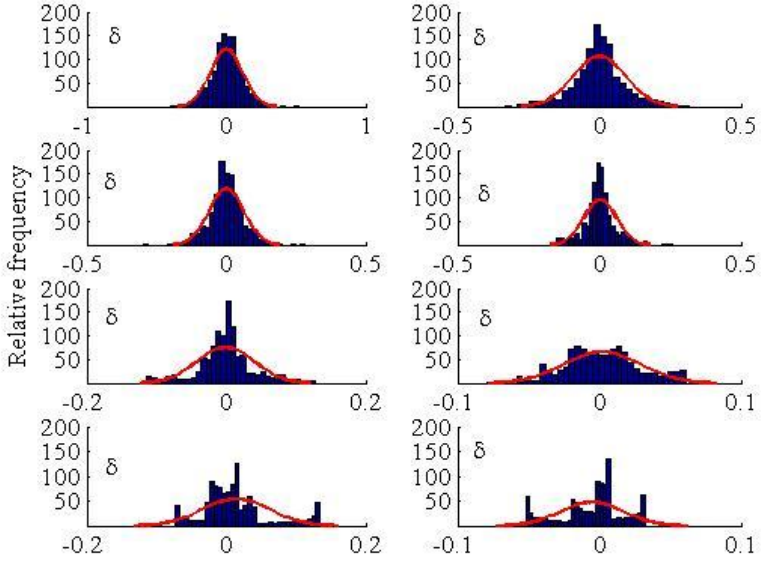

Figure 6.12. Probability distribution of the zonal linear horizontal wet gradients.

From the probability distribution depicted above, it can be observed that the deviation from the normal distribution function increases with an increase in the frequency modes of WV and linear horizontal gradients. This behaviour could be attributed to the decrease in the number of oscillations which reduces the ensemble sample size and therefore results in the less smooth distribution. Higher number of oscillations (number of IMFs with high frequency modes) follows the normal distribution according to the central limit theory. Following

Papoulies, (1996), it could be expected that WV and linear horizontal gradients therefore exhibit a Chi-square distribution.

The combination of EMD and Hilbert spectrum analyses provides an alternative adaptive method to analyse nonstationary and non-linear time series. It can perform and enhance most of the traditional data analysis tasks, such as filtering, regressions, and spectral analysis adaptively. To accommodate data from nonstationary processes, a number of methods such as spectrogram, Wigner- Ville distribution, and Wavelet analysis have all been used extensively with some degree of success. Recently, the EMD method has attracted considerable attention and been used widely in many fields. While the EMD methodology has proved to be versatile and robust, it cannot reveal the signal characteristic information accurately. This is because it has a shortcoming of mode mixing; a condition where an IMF exhibits local oscillations with clear different time/frequency scales, as reported in Huang *et al*., (1999). A solution to this problem is by subjectively introducing constraints for mode mixing, which eliminates the adaptability property in the EMD data.

Geodetic data is often measured at different spatial-temporal scales. On decomposing the intermittent sampled geodetic data, the resultant IMFs exhibits scales with a broad spectrum or redundant signals. This condition causes aliases in the time-frequency distribution. Redundancy in IMFs oscillations often conceals important characteristics that depict physically meaningful information in the data. To alleviate the mode mixing problem occurring in EMD, the EEMD method is applied. With EEMD, the components with a truly physical meaning can be extracted from the signal.

A finite, non-infinitesimal and amplitude white noise is used to force the ensemble to exhaust all possible solutions in the sifting process, thus require different scale signals to collate in the proper IMF dictated by the dyadic filter banks (Zhaohua and Huang, 2009). The effect of the added white noise is to present a uniform reference frame in the time-frequency and time-space. This implies that the added noise provides a natural reference for the signals of comparable scale to collate in one IMF. With this ensemble mean, the scale can be clearly and naturally separated without any priori subjective criterion selection, such as in the intermittence test for the original EMD algorithm. This new approach fully utilises the statistical characteristics of white noise to perturb the data in its true solution neighbourhood and then cancel out through the ensemble averaging.

6.3.1. Correlation of tropospheric WV and temperature using phase differences

Variability of troposphere WV and mean temperature plays an important role in driving the global energy and water cycles. In space geodesy, WV and temperature affect the accuracy of the delay observable. It is therefore important to study the temporal correlations of WV and temperature to understand the bias contribution from each of these parameters to the geodetic delay observable. A flexible methodology for analysing the WV fluctuations and mean temperature that adapts to their noise and uneven spectral measurements is reported in this section. The correlation strategy employed here involves the linkage of instantaneous phase differences among the associated WV and temperature IMF modes derived from the EEMD.

In this particular analysis, the amplitude level of the first IMF is taken as reference amplitude for the noisy IMFs in the data. The assumption taken here is that, the first IMF (with highest frequencies) is corrupted with noise. Therefore the criterion for selecting meaningful empirical IMFs is based on the proportion amplitude of target IMF to the reference amplitude where the amplitude of the selected amplitude ought to be about 25% of the maximum amplitude of the reference IMF. The rationale behind this criterion is derived from the relationship between the IMF amplitude and the total energy of the IMF which is the square of the amplitude is the equivalent to the energy of the IMF. As a result, the analysis uses only significant oscillating IMFs.

The phase differences between the selected IMF derived from WV and calculated temperature, provides the degree of linkage between different modes of the respective IMFs. Using the Hilbert transform, local frequencies of each IMF mode can be computed by using Equation (122). In this research, the focus is projection of the phase shift defined by Equation (122)

$$\varpi_{i,j}^{t} = \mathbb{R}e\left(e^{i\Delta\theta_{i,j}^{t}}\right) \tag{122}$$

where $\Delta\theta_{i,j}^{t} = \left|\theta_{i}^{t} - \theta_{j}^{t}\right|$. To determine the closeness of the modes, a global indicator of the constant phase shift between WV and temperature: the variance of $\varpi(:)$ is computed. The constraint used to detect pairs of modes with constant phase shift in this study is that of $\mathrm{var}\left[\varpi(:)\right] \leq 0.33$. Thereby, the IMFs with $\mathrm{var}\left[\varpi(:)\right] \geq 0.33$ have different local frequencies and hence have weak or no correlation.

Raw daily WV and mean atmospheric temperature data corresponding to the atmosphere over HartRAO exhibits certain unusual magnitudes (see Figure 6.13) due to systematic or instrumental errors. These errors were filtered out by ensuring that the data sets have been pre-proposed using the adaptive filtering procedure as described in Chapter 3 prior to the IMFs extraction. The adaptive filtering stage also gives more stability to the produce IMFs extraction. The adaptively filtered time series data are plotted in Figure 6.13. In the figure, time is expressed in days since January 2005. The adaptive filtering procedure was done using a basic variability value of 0.08, the filter coefficient of 3.0 and 0.05 controlling coefficient. As observed from the filtered time series, there exists nontrivial structure in the WV fluctuations and mean atmosphere temperature. In Chapter 3, it is reported that data could fail a test for Gaussian, thus ruling out a Gaussian linear stochastic process as the source of fluctuations. Due to this inconsistency, application of the *model-based*[11] methods to assess the structure components in WV fluctuations might be not yield representative properties.

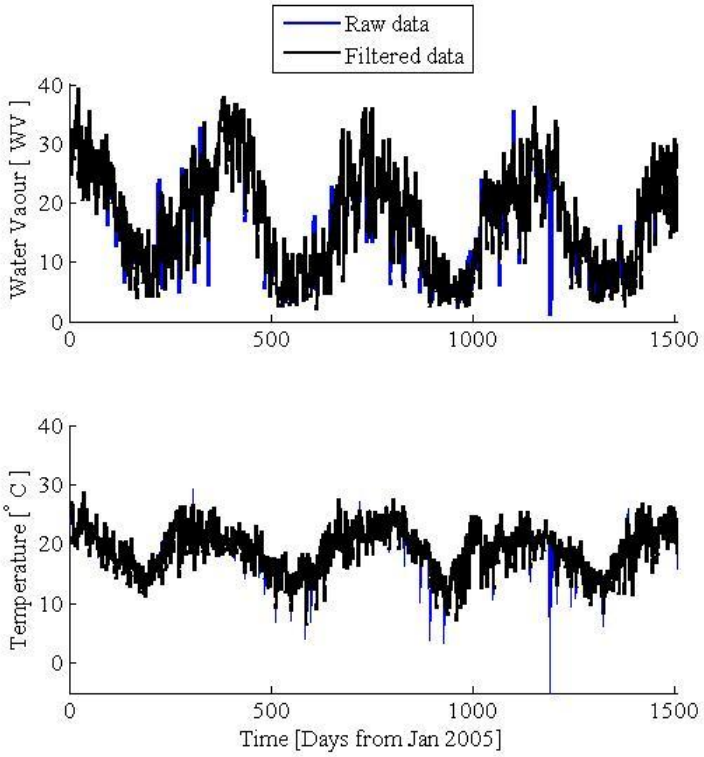

Figure 6.13. Time series of water vapour (top) and mean atmosphere temperature (bottom).

[11] Ad-hoc and model-based (e.g., ARMA & ARIMA) approaches are used to decompose a time series into structural components such as trend, seasonal/cyclic and irregular components.

The IMFs of WV and temperature data are shown in Figure 6.14 and Figure 6.15 respectively. The original WV and temperature data and the eight IMFs (the components from short to longer time periods) are plotted from top to bottom. Here IMF_8 corresponds to the trend in the data set.

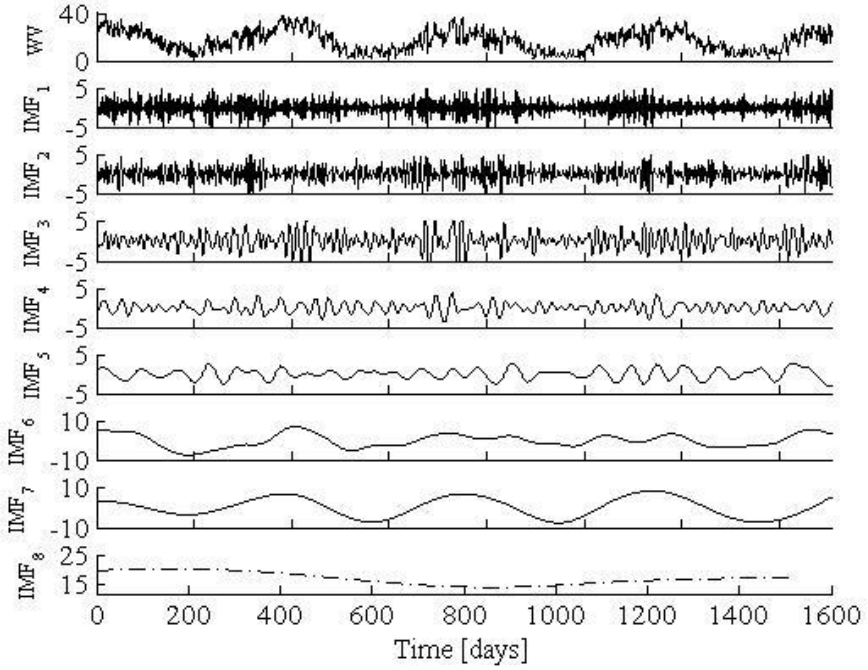

Figure 6.14. Intrinsic mode function components of water vapour over HartRAO.

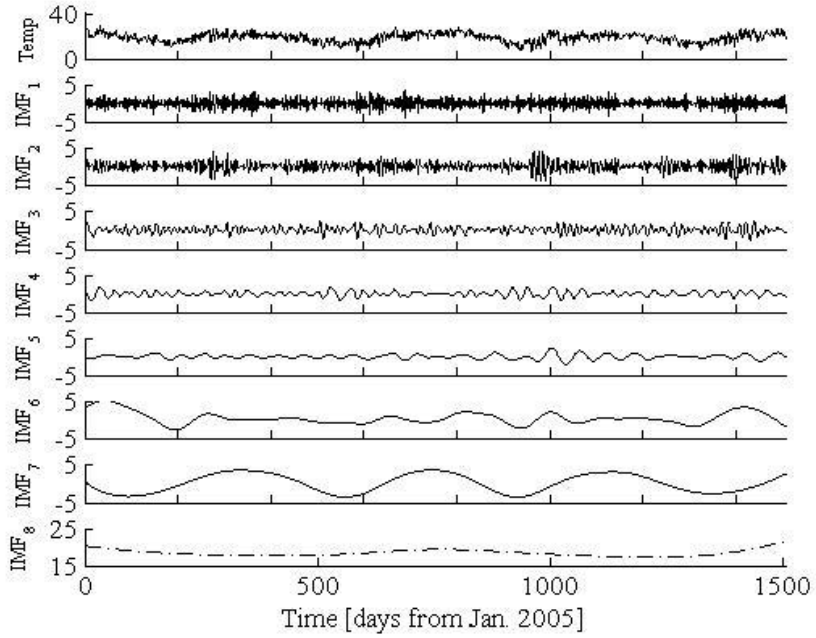

Figure 6.15. Intrinsic mode function component of mean atmosphere temperature over HartRAO.

The observed low frequency IMFs could be associated to the fluctuations that are driven by physical processes, while the high-frequency IMFs dependent on noise and independent external forces. As observed from the Figures 6.14 and Figure 6.15, the local periods are not constant but for each fixed IMF, they are constrained within different ranges. The IMFs of WV depicted in Figure 6.14 exhibits oscillation patterns with characteristic periods; annual ($IMF_{6,\,7}$ mode), seasonal ($IMF_{4,\,5}$ mode), monthly (IMF_3 mode) and diurnal ($IMF_{<2}$ mode) components. IMF_8 corresponds to the non-linear trend which exhibits both positive and negative trend. The IMF_s of the mean atmosphere temperature over HartRAO (see Figure 6.15) exhibit similar local periods ranging from 400 to 500 days ($IMF_{6,\,7}$ mode), 60 to 100 days (IMF_s mode), 30 days (IMF_3 mode) and diurnal ($IMF_{<2}$ mode) fluctuating components.

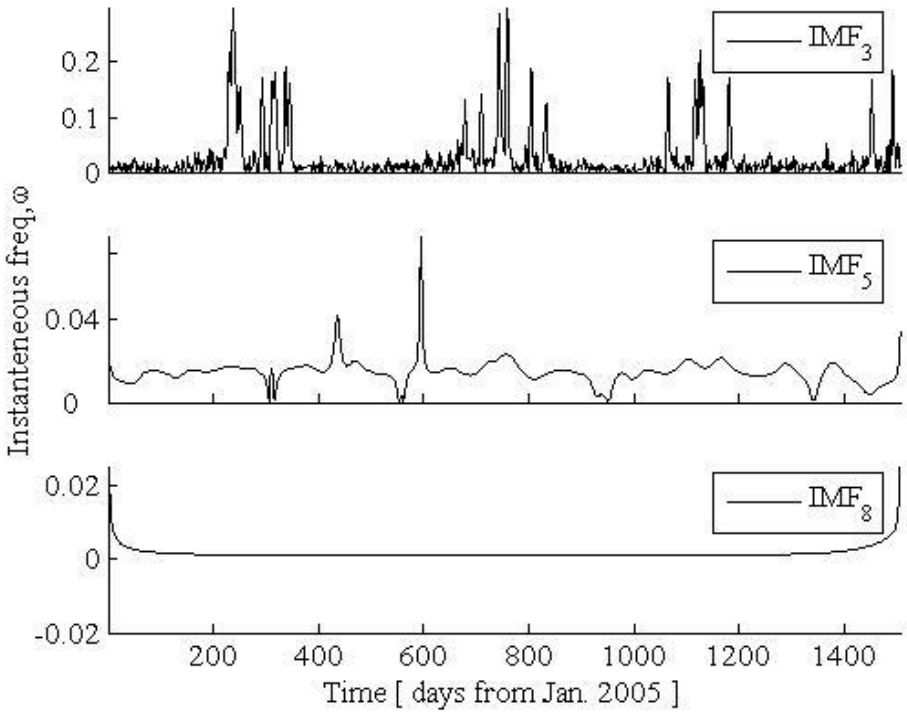

Figure 6.16. Instantaneous frequency of selected WV Intrinsic Mode Functions.

As mentioned earlier, not all the extracted IMFs from WV and mean atmosphere temperature are physically significant. Therefore on application of the criteria outlined earlier, the modes $IMF_{3,\,5,\,8}$ and $IMF_{2,\,5,\,6}$ from WV and mean atmospheric temperature respectively were considered for calculating the phase shift. As a result, Figure 6.16 and Figure 6.17 illustrate

the local frequencies of the IMFs used to derive the variance matrix for evaluating the degree of synchronisation between WV and mean troposphere temperature fluctuations. It is observed that the instantaneous frequencies (unit: cycles per day) IMF_3 and IMF_2 of WV and mean atmosphere temperature have higher values during summer months.

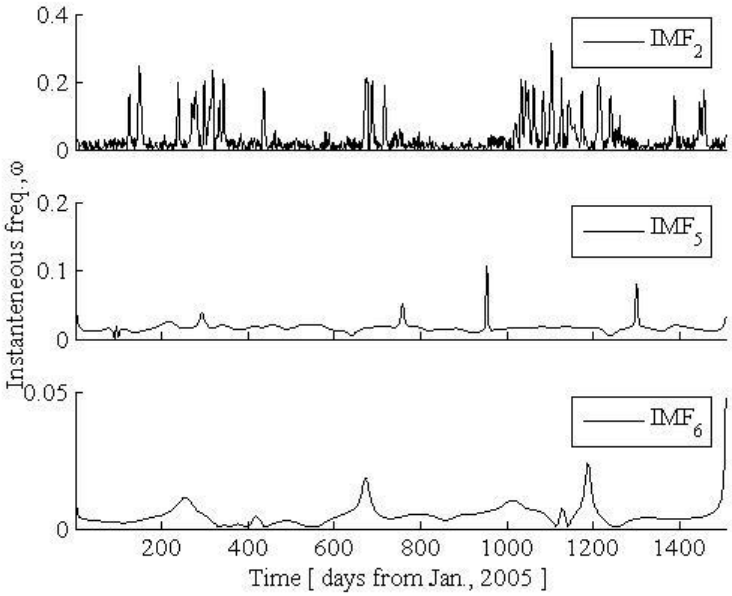

Figure 6.17. Instantaneous frequency of selected Intrinsic Mode Functions of the mean temperature.

From the variance matrix, only five pairs of modes of IMFs were considered as exhibiting lasting periods of synchronisation with fluctuating phase coincidence as illustrated in Figure 6.18. It is noted that in most of the IMFs couples selected with synchronisation, the phase coincidence fluctuates between 1 and -1 over the entire time period of this data. The correlations by pairs indicate that common fluctuations in WV and mean atmospheric temperature could be associated to both local and non-local processes. In particular, the local processes are partly responsible for driving these fluctuations i.e. heat waves and cold fronts which are common in the Highveld climatic region. The temporal dependence of phase shift seems to suggest that the WV and temperature fluctuations are strongly non-linear.. The non-linear oscillating structures visible in the phase shifts could have been triggered possibly by a stochastic resonance phenomenon such as the Inter-Tropical Convergence Zone (ITCZ) and other trade winds.

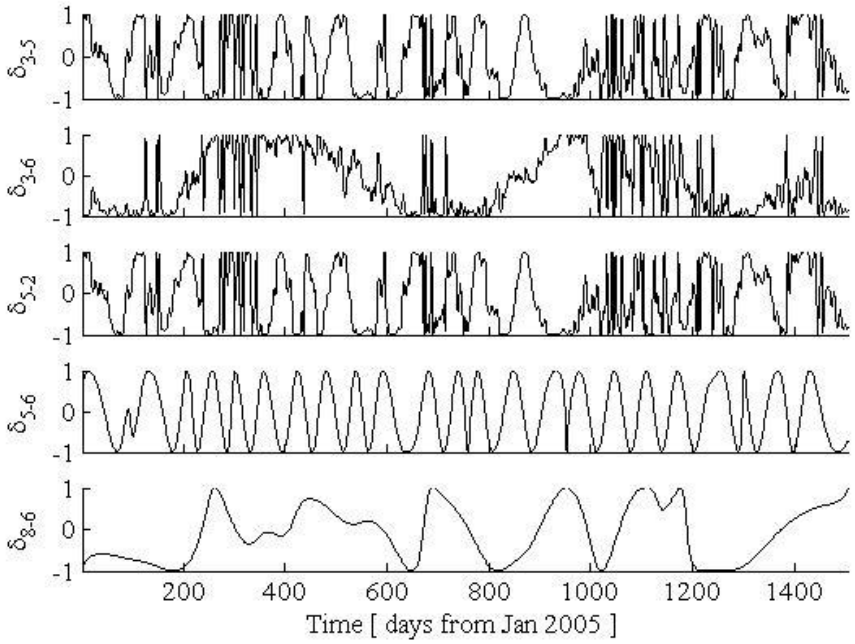

Figure 6.18. Phase shift of WV and mean atmosphere temperature IMFs modes.

6.4. Assessing the effect of non-linearity/stationarity in atmospheric range correction

Until now, the detection, analysis and characterisation of non-linearity and nonstationarity in geodetic data has been described and presented in the preceding sections of this thesis. In the present section, the benefit of introducing non-linearity and nonstationarity in atmospheric correction to the SLR range is investigated. In particular, a non-linear function is introduced to model the azimuth dependent atmospheric range correction. From the analysis of one-month of SLR data, results of azimuth dependent atmospheric range correction indicate that the introduced second-order non-linear function improves the Observed-Computed (O-C) residual by over 15%. The computed azimuth atmospheric range correction is in general negative suggesting that the current atmospheric range correction models generally overestimate the atmosphere range bias.

6.4.1. SLR atmospheric range correction

The atmosphere-Earth and Ocean is an intrinsically non-linear and nonstationary physical system that is the subject of space geodetic research. This non-linearity/nonstationarity is manifested in many series of geodetic variables depicted in Table 2.1, in Chapter 2. Detecting

and quantifying this non-linearity/nonstationarity depends on a) the type of variable, b) the geographic area and/or c) the data span. As described in the foregoing section (see Chapter 2), the range Equation used in SLR processing includes an atmospheric term which is currently modelled as an elevation only model (this is the Pavlis and Mendes (2004) model adopted by the ILRS working group). This model does not take into account the azimuth dependence and this is the point of departure. The current elevation-only model could be viewed to be linear and therefore may not be realistic since other factors such as measurement errors and unknown noise sources, may introduce non-linearities during SLR processing (and this may impact on the accuracy parameters such as station position and EOPs). In general, the measurements of e.g., the barometric pressure, temperature and relative humidity exhibit second-order terms (gradients) that are sensitive to the asymmetry of the atmosphere and could introduce some non-linearity in the atmospheric range correction term.

Detecting, characterising/analysing and incorporating non-linearity and non-stationarity into models of the SLR observation equation terms is of extraordinary importance for accurate computation of the O-C residuals. Detection and analysis of non-linearity and nonstationarity in moisture fluctuations have been achieved, in part, by calculating persistence/scaling behaviour in tropospheric delay due to WV. In order to assess the contribution of incorporating non-linear/nonstationary aspects in geodetic analysis, the current atmospheric correction used in SLR analysis has been modified by adding a non-linear term to account for the azimuth dependent atmospheric range bias. The proposed modification (see Equation (123)) has been tested on one-month of SLR data analysed by the SLR analysis software developed at HartRAO (see Combrinck, 2010).

$$O\text{-}C = R_o - \left[R_c + \Delta R_{a,e} + \Delta R_{a,az} \right] \tag{123}$$

Here, R_0 and R_c are the observed and computed ranges. Similarly, $\Delta R_{a,e}$ and $\Delta R_{a,az}$ are the elevation-only and azimuth dependent (this is modelled as a second-order function) atmospheric range correction terms. There could still be remnants of other factors such as time bias and range bias; however these were estimated before the estimate of the azimuthal non-linearity so should be eliminated to a large extent.

Figure 6.19 depicts differences between O-C (for both LAGEOS 1 and 2) before and after incorporating the non-linear function. From the results, a 15 - 20% improvement in the estimated O-C derived from the analysis of one-month SLR data for the orbits of LAGEOS 1 and 2 due is attained due to incorporation of the non-linear term in Equation (123).

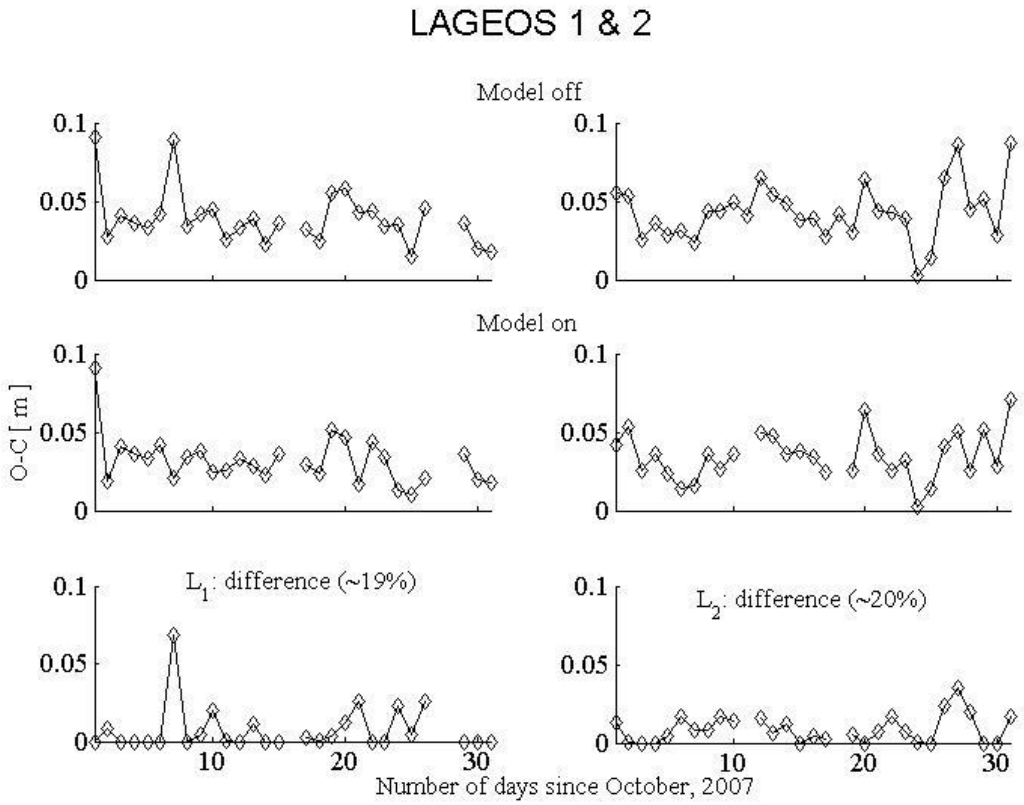

Figure 6.19. LAGEOS 1 (left column) and LAGEOS 2 (right column) observed-computed values, before (top panel), after (middle panel), percentage difference (bottom panel).

The estimated average value, minimum, maximum and standard deviation before and after incorporating the non-linear term, for O-C for LAGEOS 1 & 2 are depicted in Table 6.1.

Table 6.1. Statistical description of the observed-computed residual before and after incorporating the azimuth dependent atmospheric correction term (non-linear) in SLR processing.

Statistical parameters (O-C)	LAGEOS 1		LAGEOS 2	
	Model off	Model on	Model off	Model on
Maximum [m]	0.0908	0.0908	0.0874	0.0708
Minimum [m]	0.0145	0.0098	0.0027	0.0020
Mean [m]	0.0392	0.0317	0.0434	0.0347
Variance $\times 10^{-4}$ [m^2]	3.1121	2.4538	3.3242	2.3163

The huge improvement in O-C could be attributed to the azimuth dependent atmospheric range correction term which partly accounts for the asymmetry of the atmosphere (and the local physical conditions, such as pressure, temperature and relative humidity, along the path from the SLR station to the target satellite). Furthermore, the physical conditions (e.g., thermal radiation) of the surface in the vicinity of the SLR station will also contribute to the non-linearity and nonstationarity inherent in the SLR range. Figure 6.20 and Figure 6.21 demonstrates the azimuth dependent atmospheric range correction values based on the one-month SLR data processed by SLR analysis software developed at HartRAO.

As depicted in Figure 6.20 and Figure 6.21, there is generally strong azimuth dependence in the values. While the current one-month data might not be sufficient to demonstrate the actual quadrant dependence of the atmospheric range bias correction, the preliminary results point to specific arcs depicting non-linear relation between the angular direction and the atmospheric range correction. Furthermore, the average correction value of the present analysis is negative (see Figure 6.21), therefore suggesting that the current model used for atmospheric range correction could probably be overestimating the atmospheric range bias. This result is of extraordinary importance to the ILRS working group as well as to the research community (though long term analysis is required) with regard to the benefit of incorporating non-linear and nonstationary models in the analysis geodetic data.

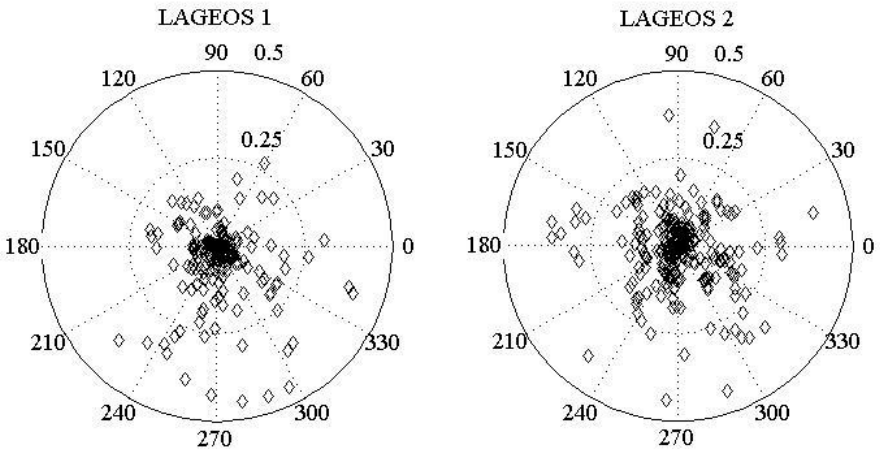

Figure 6.20. Polar description of azimuth dependent atmospheric range correction for LAGEOS 1 and 2. Only atmospheric range correction values below 0.5 (accounts for 98% of the total correction) are plotted for clarity.

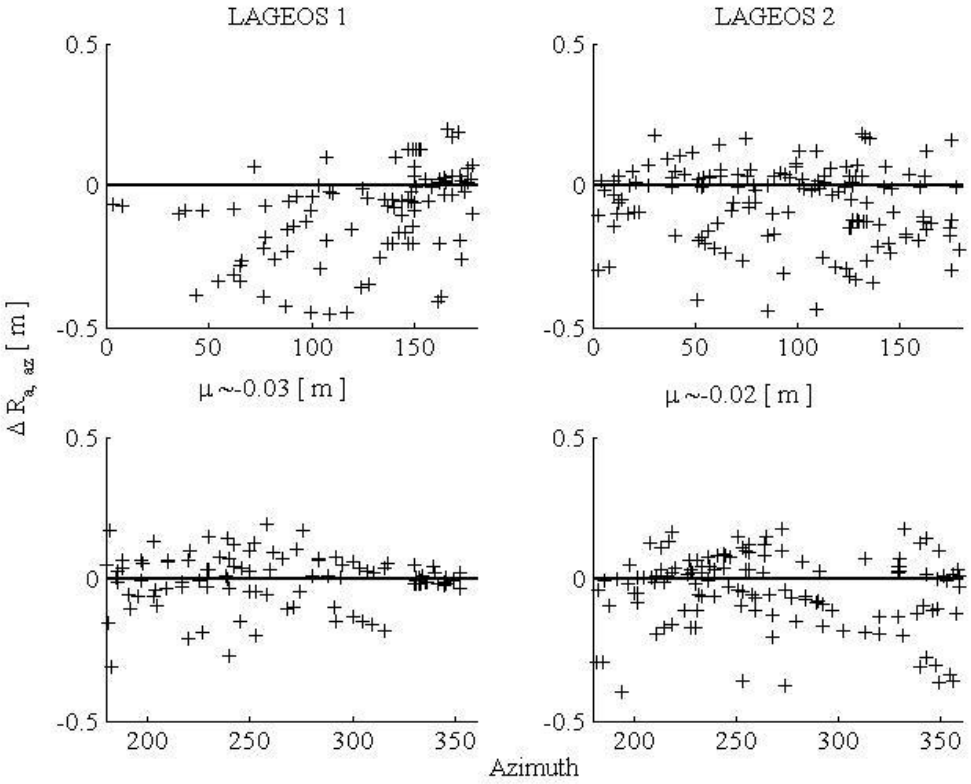

Figure 6.21. Southern (top row) and Northern (bottom row) hemisphere azimuth dependent atmosphere range correction. A negative mean value suggest that the current atmospheric range correction models generally overestimates the atmospheric range bias.

6.5. Concluding remarks

The two critical strengths for non-linear approach to the analysis of WV fluctuations are as follows:-. First, they are intrinsic to the signal where in this case, linear methods have been exploited, yet certain structures in WV fluctuations have not been accounted for. Secondly, the atmospheric system is intuitively known to include non-linear components and therefore a linear description could be unsatisfactory. It is rather subjective to presuppose that the non-linear components in the atmosphere prove enough that the non-linearity is also reflected in troposphere WV fluctuations. As a result and without any prejudice, the application of non-linear analysis methods of WV fluctuations reported here was first justified by the establishment of non-linearity in WV fluctuations in the first part of this chapter.

Different non-linear approaches to the analysis have been introduced and applied to the study WV fluctuations. The DFA and wavelet based analysis tools are widely used in several fields of complexity analysis in science. For the first time however, the geodetic

troposphere WV time series has been decomposed into segments and nonstationary processes investigated using statistical information. It is shown from the wavelet space analysis that the tropospheric WV has trends and unusual magnitudes indicating nonstationarity. Further, the second-order statistical properties (means, variances and correlations) show that they are time variants. This is the additional evidence that WV variability is driven by forces that are nonstationary. The PWV time series also has self-similar and long-range dependence properties which are dependent on the WT 'detail'.

While wavelet based approach is a good tool for detection, identification and measurement of features such as scaling in the data, it is a poor method for analysing the time-energy-frequency distributions. The extracted statistical properties are vital for establishing the presence and nature of non-linear and nonstationary properties in the data. In order to accommodate the non-linear and nonstationary structures in WV fluctuations, an alternative strategy: the EMD (EEMD) is used to adaptively decompose the WV fluctuations into different scales based on the local temporal characteristic of the data. The use of EEMD enables comparison of WV and the mean troposphere temperature, thereby giving a powerful method to extract their features that could be used to describe their dependence. Such assertion has meteorological and geodetic applications.

In order to assess the contribution of incorporating non-linear/nonstationary aspects in geodetic analysis, a non-linear term which accounts for the azimuth dependent atmospheric range bias was incorporated onto the SLR analysis software developed at HartRAO. Results show that by incorporating non-linearity and nonstationarity to the atmospheric range correction model, a 15 - 20% improvement in the estimated O-C residuals derived from the analysis of one-month SLR data was achieved for the orbits of LAGEOS 1 and 2.

7. Concluding remarks

Now a whole is that which has a beginning, middle, and end.
- Aristotle, c330 BC.

Data analysis is an important tie in the scientific research cycle of observation, analysis, interpretation and theorising. In particular, time series analysis: this often entails modelling and theoretical studies, and are often required in order to understand fluctuations of tropospheric parameters such as tropospheric delay due to WV and delay gradients and WV. These parameters are used in meteorology and climate studies as well as in space geodetic applications. The research work reported in this thesis is aimed at contributing to both these fields. We have used both parametric and non-parametric tools and studied particular process properties as inferred from the variability of tropospheric parameters. This is the first attempt to systematically apply model-based and non-linear analyses strategies to investigate the dynamic structure embedded tropospheric parameters.

To summarise, the current research reports that the variability of WV and tropospheric delays due to WV over southern Africa a) exhibits multiscale properties b) are driven by non-linear and non-stationary processes, and c) over a long period of time, the fluctuations exhibit complex scaling properties, which suggest that the fluctuations have temporal memory. As a result, tropospheric geodetic parameters do not satisfy the stationary conditions. Furthermore, incorporating non-linearity and nonlinearity to the atmospheric range bias correction in SLR analysis improves the O-C residuals by more than 15%. To conclude, at various points in this thesis, we have assessed and characterised the:

a) *Stochastic behaviour of tropospheric WV or tropospheric delay due to WV*

   In the course of the thesis, the previous literature on geodetic modelling of tropospheric parameters was examined. Time series analysis of tropospheric parameters by use of the automatic model-based approach was assessed in detail. The Box-Jenkins approach of automatic model identification and selection of the ARMA model, which models the dependence structure embedded in WV time series, was proposed, illustrated and analysed. It was found that if a parameter series is transformed to stationarity, then the underlying model structure could be represented by the ARMA model with model degree and order determined via the maximum likelihood criteria.

b) Multiscale variability of WV in the low and mid-tropical Africa

Empirical studies and simulations of WV variability over the last decade presume that the inherent fluctuations are stationary and Gaussian. The multiscale structure of WV in the low- and mid-tropical Africa region was investigated, based on *in situ* radiosonde observations of the SHADOZ station network comprising of Ascension (Ascension Island), Irene (South Africa), Reunion (Reunion) and Nairobi (Kenya) and the numerical model simulations for the period 1998 to 2006. The purpose of analysing WV fluctuations in tropical Africa was to obtain an in-depth understanding of the spatial-temporal WV fluctuations as well as to study the mechanisms driving WV variability and its link to the climatic variables. The climatic variables influenced by WV are essential for accurate geodetic tropospheric modelling. The results from the analysis show that WV exhibits high frequency fluctuations in the wavelet space. Furthermore, the embedded pattern of temporal WV fluctuations over the SHADOZ network has a dominating monthly signature. This dominant variance appears to be associated with locally-driven WV variations such as the local weather systems. The power law scaling of WV wavelet energy is a critical finding in this research. In the study, the approximate log-log linear relationship at smaller temporal scales, which breaks down at synoptic scales, suggests that the energy-times spectra of WV on different temporal scales are correlated. Furthermore, based on PCA, three dominant modes emerge. These modes explain ~ 98% of the total spatial variance of the normalised energy in WV fluctuations.

c) *Self-similar behaviour in tropospheric WV*

Studying the dynamic structure in WV or tropospheric delay caused by WV based on the wavelet approach is sufficient with regard to detection, identification and measurement of such as second order statistical parameters (which are the mean and variance) and scaling in the data. In the thesis, self-similar behaviour is assessed by use of DFA in the time-energy-frequency distributions. The extracted statistical properties are vital for establishing the presence and nature of non-linear and non-stationary properties in the data. In order to accommodate the non-linear and non-stationary structures in WV fluctuations, an EEMD is used to adaptively decompose the WV fluctuations into different scales, based on the local temporal characteristic of the data. The use of EEMD enables comparison of WV and the

mean troposphere temperature, thereby giving a powerful method to extract their features that could be used to describe their dependence.

d) *Detection, characterization and incorporation of non-linearity and nonstationarity for atmospheric range bias correction in SLR analysis*

The atmosphere-Earth and Ocean is an intrinsically non-linear and nonstationary physical system that is the subject of space geodetic research. This non-linearity/nonstationarity is manifested in many series of geodetic variables. The benefit of introducing non-linearity and nonstationarity in atmospheric correction, for example, to the SLR range has been assessed. In particular, a non-linear function is introduced to model the azimuth dependent atmospheric range correction. Based on the present analysis of one-month of SLR data, the azimuth dependent atmospheric range correction results suggest that introducing a second-order non-linear function could improve the O-C residuals by over 15% and the computed azimuth atmospheric range correction is generally negative suggesting that the current atmospheric range correction models could be overestimating the atmosphere range bias.

The studies in this thesis raised some outstanding questions and leads to many possible future extensions. These questions emerge from the ideas and concepts proposed in this work and therefore initiate new fields of research. An interesting perspective in this regard is the refinement of tropospheric modelling strategies so as to include the non-stationary behaviour of tropospheric WV or tropospheric delay due to WV. Geodetic analysis strategies often model tropospheric delays and delay gradients based on the concept of stationarity. However, and as can be inferred from the present work, tropospheric parameters are non-linear and non-stationary. Therefore an adjustment of tropospheric modelling strategies for geodetic applications is suggested.

The analyses of self-similar behaviour in tropospheric parameters by use of a scaling parameter have important climatic applications. In the current study, we have investigated and found that tropospheric parameters have memory. This memory has spatial dependence and therefore a global climatology of scaling parameters derived from various meteorological parameters could be an important product to climate change studies. An important research topic could therefore be the development of methodology to generate scaling parameters from various non-parametric approaches for the purpose of characterising the spatial self-similar patterns at global scales.

Bibliography

1. Abry A., P. Flandrin, M. Taqqu and D. Veitch, (1999). Wavelets for the analysis, estimation and synthesis of scaling data, in Self-similar Network Traffic and Performance Evaluation, Wiley.
2. Abry P. and S. Veitch, (1998). Wavelet analysis of long range dependent traffic, IEEE Trans. Info. Theor., **44**: 2-15.
3. Abry P., P. Flandrin, M. S. Taqqu and S. Veitch, (2000). Self-similar network traffic and performance evaluation, Park and Willinger (eds), John Wiley and Sons, ISBN: 9780471319740.
4. Alber C., R. Ware, C. Rocken and J. Braun, (2000). Obtaining single path phase delays from GPS double differences, Geophys. Res. Lett., **27**(17): 2661-2664.
5. Ao C. O., (2007). Effect of ducting on radio occultation measurements: An assessment based on high-resolution radiosonde soundings, Radio Sci., **42**(RS2008): doi.:10.1029/2006RS003485.
6. Ausloos M. and K. Ivanova, (2001). Power-law correlations in the southern-oscillation-index fluctuations characterizing El Nino. Phys. Rev., E**63**: 047201.
7. Arinc, (1993). Interface Control document. Navstar GPS Space Segment/Navigation User Interfaces. Technical Report ICD-GPS-200C, ARINC Research Corp., El Segundo, CA.
8. Askne J. and H. Nordius, (1987). Estimation of tropospheric delay for microwaves from surface weather data. Radio Science, **22**(3): 379-386.
9. Bates J.J., (1994). Variability of global upper-tropospheric water vapour derived from satellite infrared observations. IGARSS '94, **4**: 2060-2062, ISBN: 0-7803-1497-2.
10. Banks P. M., W. T. Huntress, R. D. Hudson, C. A. Reber, C. A. Barth, C. B. Farmer, M. A. Geller, J. C. Gille, J. London, L. R. Megill, J. M. Russell III, R. G. Roble, R. S. Stolarski and J. W. Waters, (1978). Upper Atmosphere Research Satellite Program, JPL Publication: 78-54.
11. Beran J., (1994). Statistics of long-memory processes, Chapman and Hall, New York.
12. Behrend D., J. Boehm, P. Charlot, T. Clark, B. Corey, J. Gipson, R. Haas, Y. Koyama, D. MacMillan, Z. Malkin, A. Niell, T. Nilsson, B. Petrachenko, A. E. E. Rogers, G. Tuccari

and J. Wresnik, (2008). Recent progress in the VLBI2010 development. International Association of Geodesy Symposia, **133** (5): 833-840.

13. Beutler G., M. Rothecher, S. Scher, T. A. Springer, J. Kouba and R. E. Neilan, (1999). The International GPS Service (IGS): an interdisciplinary service in support of earth sciences. Adv. Space Res., **23**(4): 632-635.
14. Bevis M., S. Businger, T. A. Herring, C. Rocken, R. A. Anthes and R. H. Ware, (1992). GPS meteorology: Remote sensing of atmospheric water vapour using the global positioning system, J. Geophys. Res., **97(**D14-15): 15,787-15,801.
15. Bevis M., S. Businger, S. Chiswell, T. A. Herring, R. A. Anthes, C. Rocken and R. H. Ware, (1994). GPS meteorology: mapping zenith wet delays onto precipitable water. J. Appl. Meteor., **33**: 379-386.
16. Brockwell P. J. and R. A Davis, (1996). Introduction to Time Series and Forecasting, New York: Springer-Verlag.
17. Broersen P. M. T., (2002). Finite sample effects in vector autoregressive modelling. IEEE Trans. Instrument. and Measur., **51**(5): 917-922.
18. Botai O. J., W. L. Combrinck and C.J. de W. Rautenbach, (2008). Nonstationary tropospheric processes in geodetic precipitable water vapor time series, Michael G. Sideris (ed). Observing our Changing Earth, International Association of Geodesy Symposia, Springer Berlin Heidelberg, **133**: 625-630, doi.: 10.1007/978-3-540-85426-5.
19. Botai O. J., S. Venkatarama, C. J. de W. Rautenbach and W. L. Combrinck, (2010), Multi-scale organization of water vapour in the low- and mid- tropical Africa, Advances in Geosciences, **16**: 241-251.
20. Botai O. J., W. L. Combrinck and V. Sivakumar, (2009). Assessing the degree of synchronisation between geophysical records using the method of instantaneous phase differences. In 11th SAGA biennial technical meeting and exhibition. D. Vogt and S. Fourie (eds). ISBN. 978-0-620-44602-0: 588-593.
21. Boehm J., B. Werl and H. Schuh, (2006). Troposphere mapping functions for GPS and very long baseline interferometry from European Centre for Medium-Range Weather Forecasts operational analysis data, J. Geophys. Res., **111** (B02406), doi.:10.1029/.
22. Boehm J. and H. Schuh, (2004). Vienna mapping functions in VLBI analyses, Geophys. Res. Lett., **31**(L01603), doi.:10.1029/2003GL018984.
23. Boehm J. and H. Schuh, (2007). Troposphere gradients from the ECMWF in VLBI analysis. Journal of Geodesy, doi.:10.1007/s00190-007-0144-2.

24. Boehm J., A. Niell, P. Tregoning and H. Schuh, (2006). Global mapping function (GMF): A new empirical mapping function based on numerical weather model data. Geophys. Res. Lett., **33**(L07304), doi.: 10.1029/2005GL025546
25. Bock O. and E. Doerflinger, (2001). Atmospheric modelling in GPS data analysis for high accuracy positioning. Physics and Chemistry of the Earth, Part A: Solid Earth and Geodesy, **26**(6-8): 373-383.
26. Box G. E. P. G. M. and G.M. Jenkins, (1970). Time series analysis: Forecasting and control, San Francisco: Holden-Day.
27. Cannon W. H., (1978). The classical analysis of the response of a long baseline radio interferometry, Geophys., J. R. Astronomy Soc., **53**: 503
28. Chen Z., P. Ch. Ivanov, K. Hu and H. E. Stanley, (2002). Effects of nonstationarities on detrended fluctuation analysis. Phy. Rev., E**65**: 041107.
29. Collins W. D., P. J. Rasch, B. A. Boville, J. J. Hack, J. R. McCaa, D. L. Williamson, B. P. Briegleb, C. M. Bitz, S-J Lin and M. Zhang, (2006). The formulation and atmospheric simulation of the community atmosphere model version 3 (CAM3). J. Climate, **19**: 2144–2161.
30. Combrinck L., (2010). Satellite Laser Ranging; In Sciences of Geodesy I, Chap 9, G. Xu, (ed). Springer Verlag, Berlin.
31. Cucurull L., B. Navascues, G. Ruffini, P. Elòsegui, A. Rius and J. Vilà, (2000). The use of GPS to validate NWP systems: The HIRLAM Model. J. Atmos. Oceanic Technol., **17**: 773-787.
32. Davis J. L., (2001). Atmospheric water vapour signals in GPS data: Synergies, correlations, signals, and errors. Phys. Chem. Earth, **26**: 513–522.
33. Davis J. L., P. Elòsegui, J. X. Mitrovica and M. E. Tamisiea, (2004). Climate-driven deformation of the solid Earth from GRACE and GPS, Geophys. Res. Lett., **31**(L24605): doi.: 10.1029/2004GL021435.
34. Davis J. L., (1992). The effect of turbulence on atmospheric gradient parameters determined from ground-based radiometric and space geodetic measurements, Geophys. Res. Lett., **19**: 2183-2186.
35. Davis J. L., G. Elgered, A. E. Niell and C.E. Kuehn, (1993). Ground-based measurements of gradients in the "wet" radio refractivity of air, Radio Science, **28**: 1003-1018.

36. Davis J. L., T. A. Herring, I. I. Shapiro, A. E. E. Rogers and G. Elgered, (1985). Geodesy by radio interferometry: Effects of atmospheric modelling errors on estimates of baseline length, Radio Sci., **20**(6): 1593-1607.
37. Deuber B., J. Morland, L. Martin, and N. Kàmpfer, (2005). Deriving the tropospheric integrated water vapor from tipping curve-derived opacity near 22 GHz, Radio Sci., **40** (RS5011): doi.:10.1029/2004RS003233.
38. Dinoev T., P. Ristori, Y. Arshinov, S. Bobrovnikov, I. Serikov, B. Calpini, H. van den Bergh, V. Simeonov, (2006). Meteorological water vapor Raman lidar: advances, Lidar Technologies, Techniques, and Measurements for Atmospheric Remote Sensing II. Edited by Singh, Upendra N. Proceedings of the SPIE, **6367**: 63670C.
39. Emardson T. R. and H. J. P. Derks, (1998). On the relation between the wet delay and the integrated precipitable water vapour in the European atmosphere, Meteorological Applications, **7**: 61-68, Doi.:10.1017/S1350482700001377.
40. Gradinarsky L. P., R. Haas, G. Elgered and J. M. Johansson, (2000). Wet path delay and delay gradients inferred from microwave radiometer, GPS and VLBI observations, Earth, Planets and Space, **52**: 695-698.
41. Greenblatt S., (1994). Wavelets in Econometrics: An application to outlier testing, manuscript, Ewp-em/9410001.
42. Ghil M., M. R. Allen, M. G. Dettinge, K. Ide, D. Kondrasshov, M. E. Mann, A. Robertson, A. Sounders, Y. Tian, F. Varadi, and P. Yiou, (2001). Advanced Spectral Methods for Climatic Time Series. Rev. of Geophysics, **40**(1): 1.1-1.41, doi.: 10.1029/.
43. Grimmet G. and D. R. Stirzaker, (2001). Probability and random processes. Oxford University Press, USA; 3rd ed. ISBN-10: 0198572220.
44. Griffiths D. J., (1999). Introduction to Electrodynamics 3rd edition. Upper Saddle River: Prentice Hall.
45. Haase J., M. Ge, H. Vedel and E. Calais, (2003). Accuracy and variability of GPS tropospheric delay measurements **o**f water vapour in the Western Mediterranean., **42**: 1547-1568.
46. Haas R., E. Gueguen, H. G. Scherneck, A. Nothnagel and J. Campbell, (2000). Crustal motion results derived from observations in the European geodetic VLBI network. Earth Planets Space, **52**: 759-764.

47. Haas R., A. Nothnagel J. Campbell and E. Gueguen, (2003). Recent crustal movements observed with the European VLBI network: geodetic analysis and results, Journal of Geodynamics, **35**(4-5): 391-414, doi.: 10.1016/S0264-3707(03)00003-6.
48. Heinkelmann R., J. Boehm, H. Schuh, S. Bolotin, G. Engelhardt, D. S. MacMillan, M. Negusini, E. Skurikina, V. Tesmer and O. Titov, (2007). Combination of long time-series of troposphere zenith delays observed by VLBI. J. Geod., **81**: 483-501, doi.: 10.1007/.
49. Herring T. A., J. L. Davis and I. I. Shapiro, (1990). Geodesy by radio interferometry: the application of Kalman filtering to the analysis of Very Long Baseline Interferometry data. J. Geophys. Res., **95**(B8): 12561-12581.
50. Hipel K. W. and A. A. McLeod (1994). Time series modelling of water resources and environmental systems, Elsevier, Amsterdam. ISBN: 0-444-89270-2.
51. Hobiger T., T. Kondo and H. Schuh (2006). Very long baseline interferometry as a tool to probe the ionosphere, Rad., Sci., **41**(RS1006): doi.:10.1029/2005RS003297.
52. Hofmann-Wellenhoff B., H. Lichtenegger and J. Collins, (1997). Global Positioning System, Theory and Practice. Springer, N. Y., 4th revised edition.
53. Hofmann-Wellenhoff B., H. Lichtenegger and J. Collins, (2001). GPS, Theory and Practice, 5th ed., Springer-verlag.
54. Houghton J. T., (1986). The Physics of the atmosphere, 2nd ed., Cambridge University Press, Cambridge, UK.
55. Ho C. M., B. D. Wilson, A. J. Mannucci, U. J. Lindquister and D. N. Yuan, (1997). A comparative study of ionospheric total electron content measurements using global ionospheric maps of GPS, TOPEX radar, and the Bent model, Radio Sci., **32**(4): 1499-1512.
56. Hogg D. C., F. O. Guiraud and W. B. Sweezy, (1981). The short-term temporal spectrum of precipitation water vapour. Science, **213**(4512): 1112.
57. Huang N. E., C. C. Chen, K. Huang, L. W. Salvino, S. R. Long and K. L. Fan, (2001). A new spectral analysis of station TCU129, Chi-Chi, Taiwan, 21 September 1999. Bulletin of the seismological of America, **91**(5): 1310-1338.
58. Huang N. E. and N. O. Attoh-Okine, (2005). The Hilbert-Huang transform in engineering, Taylor & Francis.
59. Hu K., P. Ch. Ivanov, Z. Chen, P. Carpena and H. E. Stanely, (2001). Effect of trends on detrended fluctuation analysis. Phys. Rev. E**64**, 011114.

60. Huang N. E., M. Wu, W. Qu, S. R. Long and S. P. Shen, (2003b). Applications of the Hilbert-Huang transform to nonstationary financial time series analysis. Appl. Stoch. Models in Business and Industry, **19**: 245-268.
61. Huang W., Z. Shen, N. E. Huang and Y. C. Fang, (1999b). Nonlinear indicial response of complex nonstationary oscillations as pulmonary hypertension responding to step hypoxia. Proc. Natl. Acad. Sci. USA, 96.
62. Huang N. E., Shen Z., Long S. R., Wu M. C., Shih H. H., Zheng Q., Yen N. –C., Tung C. C. and Liu H. H., (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. Roy. Soc. Lond., **454**: 903-993.
63. ITU-R., (2003). The refractive index: its formula and refractivity data: 453-459.
64. Ivanova K. and M. Ausloos, (1999). Application of the detrended fluctuation analysis (DFA) method for describing cloud breaking. Physica A**274**: 349.
65. Janes H. W., R. B. Langley and S. P. Newby, (1991). Analysis of tropospheric delay prediction models: comparisons with ray-tracing and implications for GPS relative positioning, J. Geodesy, **65**(3): 151-161: doi.: 10.1007/BF00806344.
66. James C. O., (1967). Optical refractive index of air: dependence on pressure, temperature and composition, Appl. Opt. **6**: 51-59.
67. Jarlemark P. O. J. and G. Elgered, (1998). Characterization of temporal variations in atmospheric water vapour. IEEE Trans., Geosci., Remote Sens., **32**(1): 319 – 321.
68. Jarlemark P. O. J., (1997). Analysis of temporal and spatial variations in atmospheric water vapor using microwave radiometry, PhD thesis, Techn. Rep. 308, School of Electrical Computer Engineering, Chalmers Univ. Techn., Gotebörg, Sweden.
69. Jin S. G. and P. H. Park, (2005). A new precision improvement of zenith tropospheric delay estimates by GPS. Current Science, **89**(6): 997-1000.
70. Jin S., O. F. Luo and J. Cho, (2009). Systematic errors between VLBI and GPS precipitable water vapour estimations from 5-year co-located measurements. J. atmos., solar-terr., phys., **71**(2): 264-272.
71. Kalkany E., M. Kanmistu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, B. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. Mo, C. Ropelewski, J. Wang, R. Jenne and D. Joseph, (1996). The NCEP/NCAR 40- year reanalysis project, Bull. Amer. Meteor. Soc., **77**: 437-471.

72. Kruger A. C., (2006). Observed trends in daily precipitation indices in South Africa: 1910-2004. Int. J. Climatology, **26**(15): 2275-2285.

73. Li J., S. Miyashita, T. Kato and S. Miyazaki, (2001). GPS time series modelling by autoregressive moving average method: Application to crustal deformation in central Japan. Earth Planets Space, **157**: 155-162.

74. Lin S-L., P-C. Tung and N. E. Huang, (2009). Data analysis using a combination of independent component analysis and empirical mode decomposition. Phys. Review, E **79**: 066705-1-6.

75. Mallat S. (1999). A Wavelet Tour of Signal Processing, 2nd Edition Academic Press.

76. Montanari A., R. Rosso and M. S. Taqqu, (2000). A seasonal fractional ARIMA model applied to the Nile River monthly flows at Aswan. Water Res. Res., **36**: 1249,

77. Niell A. E., (2000). Improved atmospheric mapping functions for VLBI and GPS, Earth Planets Space, **52**: 699-702.

78. Niell A. E., (2001). Preliminary evaluation of atmospheric mapping functions based on numerical weather models, Phys. Chem. Earth, **26**: 475-480.

79. Niell A. E., A. J. Coster, F. S. Solheim, V.B. Mendes, P. C. Toor, R. B. Langley and C. A. Upham, (2001). Comparison of measurements of atmospheric wet delay by radiosonde, water vapor radiometer, GPS, and VLBI. J. Atmos. Oceanic Technol., **18**: 830-850.

80. Niell A. E., (2006). Interaction of Atmosphere Modelling and GPS Analysis Strategy. In: Behrend D, Rius A. (eds) Proceedings of IVS 2006 general meeting, NASA/CP-2006.

81. Niell A., (2005). VLBI2010: Current and Future Requirements for Geodetic VLBI systems. In: International VLBI Service for Geodesy and Astrometry 2005 Annual Report, D. Behrend and K. Baver (eds), NASA/TP-2006-214136: 13-40.

82. Nilsson T. and R. Haas, (2008). Modelling tropospheric delays with atmospheric turbulence models. Proceedings of the Fifth IVS General meeting: Measuring the future. ISBN/ISSN: 978-5-02-025332-2: 361-370.

83. Nilsson T., G. Elgered and L. Gradinarsky, (2006). Characterising atmospheric turbulence and instrumental noise using two simultaneously operating microwave radiometers. IEEE MicroRad: 270-275.

84. Nilsson T. and G. Elgered, (2008). Long-term trends in the atmospheric water vapour content estimated from ground-based GPS data. J. Geophys. Res., **113**(D19101): doi.: 10.1029/2008JD010110.

85. Nilsson T., R. Haas and G. Elgered, (2007). Simulations of atmospheric path delays using turbulence models. In: Proc. of the 18th European VLBI for Geodesy and Astrometry Working Meeting, J. Boehm, A. Pany, H. Schuh (eds.), Goewissenschaftliche Mitteilungen, Schriftenreihe der Studienrichtung Vermessung und Geoinformation, Technische Universität Wien, **79**: 175-180.

86. Nyst M., T. Nishimura, F. F. Pollitz and W. Thatcher, (2006). The 1923 Kanto earthquake re-evaluated using a newly augmented geodetic data set, J. Geophys. Res., **111**(B11306): doi:.10.1029/2005JB003628.

87. Pacione R. and F. Vespe, (2003). GPS zenith total delay estimation in the Mediterranean area for climatological and meteorological applications. J. Atmos. Oceanic Technol., **20**: 1034-1042.

88. Pan J., X. Yan, Q. Zheng, W. T. Liu and V. V. Klemas, (2002). Interpretation of scatterometer ocean surface wind vector EOFs over the Northwestern Pacific. Remote sensing of the environment, **84**: 53-68.

89. Papoulies A., (1986). Probability, random variable and stochastic processes, second edition, McGraw Hill.

90. Petrov L. and J.-P. Boy, (2004). Study of the atmospheric pressure loading signal in VLBI observations, J. Geophys. Res., **109**(NB03405): 10.1029/2003JB002500.

91. Petrov L., D. Gordon, J. Gipson, D. MacMillan, C. Ma, E. Fomalont, R-C. Walker and C. Carabajal, (2009). Precise geodesy with the Very Long Baseline Array, J. Geodesy, doi.: 10.1007/s00190-009-0304-7.

92. Pegram G. G. S., M. C. Peel and T. A. McMahon, (2008). Empirical mode decomposition using rational splines: an application to rainfall time series, Proc. R. Soc. A., **464**: 1483-1501, doi.: 10.1098/rspa.2007.0311.

93. Percival D. B. and A. Walden, (2000). Wavelet Methods for Time Series Analysis. Cambridge University.

94. Peña M. A., J. C. Echeverría, M. T. García and R. González-Camarena, (2009). Applying fractal analysis to short sets of heart rate variability data, Med. Biol. Compu., **47**: 709-717, doi.: 10.1007/s11517-009-0436-1.

95. Pfeifer P. E. and S. J. Deutsch, (1980). A three-stage iterative procedure for space-time modelling. Technometrics, **22**(1): 35- 47.

96. Qian S. (2002). Introduction to Time-Frequency and Wavelet Transforms Prentice Hall PTR, 2002.

97. Qian X-Y., W-Z. Zhou and G-F Gu, (2008). Modified detrended fluctuation analysis based on empirical mode decomposition, arXiv: 0907.3284v1.

98. Ray R., (1999). A Global Ocean Tide Model From TOPEX/Poseidon Altimetry/GOT99.2-NASA/TM-1999-209478: 58, Goddard Space Flight Center/.

99. Rial A. R., R. A. Pielke, M. Beniston, M. Claussen, J. Canadell, P. Cox, H. Held, N. De Noblet-Ducoudre, R. Prinn, J. F. Reynolds and J. D. Salas, (2004). Nonlinearities, feedbacks and critical thresholds within the Earth's Climate. Climate Change, **65:** 11-38.

100. Riepl S. and W. Schlűter, (2000). Normal point algorithm for reduction of two color SLR operations. [Internet] \http://www.wettzell.ifag.de/publ/2cnnpt/nizza.html. [Date of access: 10th Dec. 2008].

101. Rothacher M. (2002). A combination of space geodetic techniques. In: International VLBI service for Geodesy and Astrometry General Meeting Proceedings, Tsukuba, Japan, February 4-7, edited by N. R. Vandenberg and K. Baver: 33-43, NASA/CP-2002-210002.

102. Rocken C., R. Ware, T. Van Hove, F. Solheim, C. Alber, J. Johnson, M. Bevis and S. Businger, (1993). Sensing atmospheric water vapour with the global positioning system, Geophys. Res. Lett., **20**(23): 2631-2634.

103. Ruf C. S. and S. E. Beus, (1997). Retrieval of tropospheric water vapour scale height from horizontal turbulence structure. IEEE Trans. Geosci. Remote Sens., **35**(2): 203-211.

104. Rybski D. and A. Bunde, (2009). On the detection of trends in long-term correlated records, Physica A: Statistical Mechanics and its applications, 388(8), 1687-1695.

105. Salisbury J. I. and M. Wimbush, (2002). Using modern time series analysis techniques to predict ENSO events from the SOI time series. Nonlinear Processes in Geophysics, **9**: 341–345.

106. Saastamoinen J., (1972). Atmospheric correction for the troposphere and the stratosphere in radio ranging of satellites, In: The Use of Artificial Satellites for Geodesy. Geophys. Monogr. Ser. **15**: 247-251.

107. Saco P. and P. Kumar, (2000). Coherent modes in multiscale variability of streamflow over the United States. Water Resour. Res., **36**: 1049–1067.
108. Sansò and A. J. Gil, (2006). Estimating Crustal Deformation Parameters from Geodetic Data: Review of Existing Methodologies, Open Problems and New Challenges, International Association of Geodesy Symposia, **131**: 7-18, doi.: 10.1007/978-3-540-38596-7.
109. Scherneck H.-G., (1991). A parameterized solid earth tide model and ocean tide loading effects for global geodetic baseline measurements, Geophys. J. Int., **106**: 677-694.
110. Stenke A. and V. Grewe, (2005) Simulation of stratospheric water vapour trends: Impact on stratospheric Ozone chemistry. Atmos. Chem. Phys., **5**: 1257-1272.
111. Shumway R. H. and D.S. Stoffer, (2006). Time series analysis and its applications: With R Examples, 2nd ed., Springer.
112. Schubert S. D., H. M. Helfand, C.-Y. Wu and W. Min, (1998). Sub-seasonal variations in warm-season moisture transport and precipitation over the central and eastern United States. J. Climate, **11**: 2530–2555.
113. Soden B. J. and R. Fu, (1995). A satellite analysis of deep convection, upper-tropospheric humidity, and the greenhouse effect. J. Climate, **8**: 2333–2351.
114. Schuh H., and J. Boehm, (2003). Determination of Tropospheric Parameters Within the new IVS Pilot Project, In: Proceedings of the 16th Working Meeting on European VLBI for Geodesy and Astrometry, W. Schwegmann and V. Thorandt (eds), Bundesamt für Kartographie und Geodäsie, Frankfurt/Leipzig: 257 - 264.
115. Shapiro I. I., (1976). Estimation of astrometric and geodetic parameters, In methods of experimental physics, 12C, M. L. Meeks, Ed., Academic Press, N. Y. :261-276.
116. Sivakumar V., D. Tefera, G. Mengistu and O. J. Botai, (2010). Mean ozone and water vapour height profiles for southern hemisphere region using Radiosonde/Ozonesonde and HALOE satellite data, Advances in Geonsciences, **16**: 263-270.
117. Smith E. K. and S. Weintraub, (1953). The constants in the equation for atmospheric refractive index at radio frequencies, Proceedings of the IRE, **41**(8): 1035-1037.
118. Standish E. M., (1990). The observational basis for JPL's DE200, the planetary ephemerides of the astronomical almanc. Astron. Astrophys., **233**: 252-271.
119. Stoew B. and G. Elgered, (2005). Spatial and temporal correlations of the GPS estimation errors, TOUCH report, 22. [Internet] http://web.dmi.dk/pub/touch/. [Date of access: 12th Jan. 2008].

120. Stull R. B., (1994). An introduction to boundary layer meteorology (Atmospheric Sciences Library), Kluwer Academic, Dordrecht, Netherlands.
121. Talkner P. and R. O. Weber, (2000). Power spectrum and detrended fluctuation analysis: application to daily temperatures, Phys. Rev. E., **62**: 150-160.
122. Taqqu M. S., V. Teverovsky and W. Willinger, (1995). Estimators for Long-Range Dependence: An Empirical Study. Fractals, **3**(4): 785-798.
123. Tarniewicz J., O. Bock, J. Pelon and C. Thom, (2002). Raman lidar for external GPS path delay calibration devoted to high accuracy height determination. Physics and Chemistry of the Earth, Parts A/B/C, **27**(4): 329-333(5).
124. Thompson A. R., J. M. Moran and G.W. Swenson, (2001). Interferometry and synthesis in radio astronomy. Wiley, New York.
125. Tralli D. M., T. H. Dixon and S. A. Stephens, (1988). Effect of wet tropospheric path delays on estimation of geodetic baselines in the Gulf of California using the global positioning system. J. Geophys. Res., **93**(B6): 6545-6557.
126. Trauhaft R. N. and G. Lanyi, (1987). The effect of the dynamic wet troposphere on radio interferometry measurements. Radio Sci., **22**: 251-265.
127. Trömel S. and C. D. Schönwiese, (2008). Robust trend estimation of observed German precipitation, Theor. Appl. Climatol. **93**:107–115, doi.: 10.1007/s00704-007-0341-1.
128. Thayer G. D., (1974). An improved equation for the radio refractive index of air, Radio Sci., **9**(10): 803-807.
129. Torrence C. and G. P. Compo, (1998). A practical guide to wavelet analysis, Bull. Amer. Meteorol. Soc., **79**: 61–78.
130. Varotsos C., M. Efstathiou and C. Tzanis, (2009). Scaling behaviour of global tropopause, Atmos. Chem. Phys., **9**: 677-683.
131. Verdes P. F., P. M. Granitto, H. D. Novane, and H. A. Ceccatto, (2001). Nonstationary time-series analysis: Accurate reconstruction of driving forces. Phy. Rev. Lett., **87**(12).
132. Wang C. -S., Y. -A. Liou and T. -K. Yeh, (2008). Impact of surface meteorological measurements on GPS height determination, Geophys. Res. Lett., **35**(L23809), doi.:10.1029/2008GL035929.
133. Wessel N., A. Voss, H. Malberg, C. H. Ziehmann, H. U. Voss, A. Schirdewan, U. Meyerfeldt and J. Kurths, (2000). Nonlinear analysis of complex phenomena in cardiological data, Herzschr. Elektrophys., **11**(3): 159-173

134. Wicki F., (2001). Robust estimator for the adjustment of geodetic networks. In: Carosio A., Kutterer H. (eds), Proceedings of the first international symposium on robust statistics and fuzzy techniques in geodesy and GIS, IAG-SSG 4.190 Non-probabilistic assessment in geodetic data analysis: 53-60.

135. Wolfgang R. D. and B. Richter (eds), (2008). IERS Annual Report 2006. International Earth Rotation and Reference Systems Service, Central Bureau. Frankfurt am Main: Verlag des Bundesamts fűr Kartographie und Geodäsie, **187**: (in print).

136. Wresnik J., J. Boehm, A. Pany and H. Schuh, (2008). VLBI2010 simulations at IGG ViennaIn: Measuring the Future, Proceedings of the 5th IVS General Meeting, Andrey Finkelstein and Dirk Behrend (eds.), Saint Petersburg, "Nauka", ISBN: 978-5-02-025332-2:421-426, 3-6 March, St. Petersburg, Russia.

137. Wresnik J., R. Haas, J. Boehm and H. Schuh, (2005). Thermal deformation of VLBI antennas, In: Proceedings of the 17th Working Meeting on European VLBI for Geodesy and Astrometry, edited by M. Vennebusch and A. Nothnagel, April 22-23.

138. Whitcher B. J., (1998). Assessing Nonstationary time Series Using Wavelets. Ph. D. Thesis, University of Washington.

139. Wu Z. and N. E. Huang, (2004). A study of the characteristics of white noise using the empirical mode decomposition method, Proceedings Royal Society of London, A**460**: 1597-1611.

140. Zhang C., B. E. Mapes and B. J. Soden, (2003). Bimodality in tropical water vapour. Q.J. R. Meteorol. Soc., **129**: 2847-2866.

141. Zhaohua W., Huang N. E., Long S. R. and C. –K. Peng, (2007). On the trend, detrending and variability of nonlinear and nonstationary time series, PNAS, **104**(38): 14889-14894.

142. Zhaohua W. and N. E. Huang, (2009). Ensemble empirical mode decomposition: a noise-assisted data analysis method, Advances in Data Adaptive Analysis, **1**(1): 1-41.

143. Zhengdong B. and F. Yanming, (2003). GPS Water Vapour Estimation Using Interpolated Surface Meteorological from Australian Automatic Weather Stations. In: 10th Australian International Aerospace Congress, 29 July - 1 August, Brisbane, Queensland.