Trond Arve Haakonsen

Temporal and Spatial Analyses of Continuous GPS Observations

Doctoral thesis
for the degree of doktor ingeniør

Trondheim, 2005

Norwegian University of Science and Technology
Faculty of Engineering Science and Technology
Department of Civil and Transport Engineering

NTNU
Norwegian University of Science and Technology
Doctoral thesis
for the degree of doktor ingeniør
Faculty of Engineering Science and Technology
Department of Civil and Transport Engineering

ISBN 82-471-7231-3 (printed ver.)
ISBN 82-471-7230-5 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2005:171

Printed by NTNU-trykk

Abstract

The objective of this thesis is to increase the understanding of land motion, using statistical analysis of the available three-dimensional (3-D) time series from Continuously operated Global Positioning System stations (CGPS). A primary goal has been to search for new procedures and determine a spectrum of univariate CGPS time series. A new program package has been developed for adjustment and spectral analysis of the CGPS time series for possible future studies. This will take into account the possibility of weighting and managing, non-stationary and time series with missing data. The least squares spectrum and its relation to covariance and correlation functions are explained in detail. Different estimation methods of such functions that are used to visualise temporal correlations have been tested. A challenge of the work has been the handling data jumps revealed in the numerical investigations, that vitally affect the estimation of correlation functions and spectra.

The work has also included numerical analyses of two independently pre-processed data sets from a Norwegian and a European network of CGPS. The spectral analysis has shown large contributions of annual cycles for most series. In addition semi-annual and three-month cycles have been found. After the removal of these effects, the resulting spectra estimated from the residual series of CGPS time series shows similarities with Markov-processes that are closely related to white noise processes. The detection of a Chandler-like period is more surprising because this period should have been corrected for in the pre-processing stage of CGPS time series. A Principal Component Analysis (PCA) has been implemented to determine the spatial pattern of CGPS data. So far this has not been performed either for a network of Norwegian CGPS stations or for a network covering the whole of Europe. The results from both analyses confirm that the first mode obtained up to 50 percent of the overall variance in all station series. In addition to a high inter-station correlation, only slightly decreasing as a function of distance, this implies at least one common effect for the entire network. The spatial and

partially temporal pattern for the first three modes in CGPS data show similarity with corresponding modes extracted from predictions of atmospheric loading of surface displacements. These relations have been further investigated for the purpose of deriving possible model relations. Numerical investigations using the derived models to correct the CGPS data for the atmospheric loading effect have shown a possible sample variance reduction of up to 10 percent for the vertical component of CGPS data.

Keywords: *GPS permanent stations, Time series analysis, Jump detection, Trend analysis, Weighted regression, Spectral analysis, Correlation analysis, Principal Component Analysis, Empirical Orthogonal Functions*.

Acknowledgements

First and foremost I would like to express my sincere gratitude to my main supervisor Associate Professor Hossein Nahavandchi. Without his inspiring guidance and encouragement, it would be hard to carry out this work. I would also like to thank my co-supervisor Professor Hans-Peter Plag who formulated the objective of this thesis, and introduced me to the exciting world of software engineering. I am indebted to my supervisors for innumerable discussions, valuable comments, professional advice and help through the process of developing the idea for this thesis and making it a reality.

I would like to extend my gratitude to all members of Geomatics Division for their support, especially to head of Geomatics Division, Assistant Professor Terje Skogseth for the interest he showed in my progress and was a person who I always could turn to for assistance. Special thanks also to Professor Terje Midtbø who helped me to formulate the basic approach used in this thesis. Other members of the Division: Professor Ingolf Hådem, Professor Knut Ragnar Holm, Associate Professor Oddgeir Øfsti, Stig Halvorsen, Svein Tore Dahl, and Wenche Waal are all acknowledged. I would like to take this opportunity to thank NTNU for the financial support and Professor Asbjørn Hovd, Head of Department of Civil and Transport Engineering for his help and support.

Many thanks go to my fellow PhD students during the years, especially Mr Erik Nyrnes, Mr Ali Soltanpour and Mr Kourosh Ghazavi who were always willing to help me with practical problems and provide useful advice.

I am grateful to the Geodetic Institute of National Mapping Authority of Norway and its Head Mr Bjørn Engen for the financial support, help and providing

the necessary data. Dr Kierulf, Dr Kristiansen, and Mr Bockmann at this division are cordially acknowledged.

I would also like to thank Tonie van Dam and Pascal Gegout for making their atmospheric loading predictions available. Thanks go to Mike Heflin for providing the CGPS time series and for his positive response to my questions.

Special thanks go to Professor Henning Omre, NTNU, for his enthusiastic lectures, and for pointing out the importance of statistics theory to me. I also thank Assistant Professor Stewart Clark, NTNU, for his lectures on matters of English and for editing the thesis.

Last but not least, I wish to thank my wife Marit, for her patient, positive support and understanding throughout this work.

Contents

Notations

Symbols

A

C

D

E

F

List of Figures

List of Tables

Chapter 1

Background

1.1 Introduction

Since the early 1980s, the International GPS Service (IGS) has stimulated the establishment of increasingly denser regional and global networks of Continuous GPS (CGPS) stations. The main purpose of these activities was to provide very precise satellite orbits and clock parameters required for high-precision analysis of GPS observations. CGPS data processing has been improved over the last decade as well as the approaches to realise a high accuracy and stable reference frame. Present three-dimensional (3-D) CGPS time series allow signals to be detected on the millimetre level [e.g. Smith et al., 2004]. This makes it possible to use time series from CGPS networks to derive the spatial and temporal pattern of the dominant geophysical signals in the displacements of the Earth's surface. The characteristics can be used to identify the major processes causing the displacements. In this way, CGPS-determined time series help to validate the geophysical models used for the predictions of displacements. On the other hand, well-determined effects from improved displacement predictions may be used as future corrections in the pre-processing stage of GPS, and thus improve future GPS models. There is reason to believe that neighbouring CGPS stations are co-varying. In that context, an idea would be to separate out common motion (in time). One possible method is creating a common mode signal that can represent all the stations. This is called Principal Component Analysis (PCA). The method is also known as Empirical Orthogonal Functions (EOF) or Common Modes (CM) analyses. The

theory has been known for a while [Lorenz, 1956]. However, the formulas require de-trended or zero mean expectation univariate data series as input, which CGPS time series are generally not. A great challenge, if really possible, will therefore be to ensure de-trended CGPS series as input, without removing any of the global or regional effects which are searched in the later multivariate PCA. A detailed examination of each single series both in the time and frequency domain will be required. For trend analyses in time domain, classical regression approaches [as in Wetherill, 1986] or [Johnson and Wichern, 1998] seem reasonable. The theory of Least Squares Spectral Analysis (LSSA), described by Wells et al. [1985],Vanicek and Krakiwsky [1986] and Craymer [1998] will be used to investigate non-stationary time series the in frequency domain. The goal of PCA is to identify those modes that contribute most to the variance of the time series. In order to achieve this, the modes can be constructed as a set of linear combinations of the original data. A search is made for the directions of maximum variation, subject to the constraint that they all are orthogonal to each other. Thus the modes can be found from an orthogonal transformation from the original directions. Most often, common modes are considered as disturbing noise and attempts are made to remove them. Scherneck et al. [2000] use EOFs to extract common modes from Scandinavian CGPS time series and attribute the main modes to errors in satellite orbits and clocks. Smith et al. [2004] use the daily average of stations from a local network to remove common effects to all stations, and thus be able to detect minor differential movements with respect to one station. In stable areas not affected by tectonic motion, de-trended daily CGPS time series appear to have high inter station correlations. An EOF analysis of the vertical component for a network of Swedish and Finnish CGPS stations made by [Johansson et al., 2002], revealed a correlated noise across a large region (~ 1000km) only slightly decreasing with distance. They conclude that identification and elimination of this source (or sources) will make it possible to reduce the velocity uncertainties from GPS by approximately 50 percent.

Surface loading due to mass relocations such as in the atmosphere, the ocean, the terrestrial hydrosphere and the cryosphere constantly deforms the solid Earth. Model estimates indicate that radial displacements of the Earth's surface due to loading range from a few centimetres at diurnal time scales to about one cen-

timetre with seasonal to annual time scales. Horizontal displacements are typically one third or less of the radial displacements. For known surface loads, geophysical models are used to predict the surface displacements due to loading with an increasing accuracy [van Dam et al., 2003]. The IERS conventions (2003) [McCarthy and Petit, 2004] specify models for most of these processes. However, uncertainties in the surface loads as well as the Earth model limit the accuracy of the predictions. In tectonically active regions or regions with large man-made subsidence, tectonic and man-made processes may also contribute significantly to such displacements. At present Earth tides and ocean tidal loading can be modelled with high accuracy [McCarthy and Petit, 2004] and are generally taken into account in state-of-the-art analysis of CGPS data. Models for atmospheric surface loading induced displacements that are generally far less accurate mainly due to uncertainties in the surface load itself, but also the geophysical models used. Therefore, these loading signals are normally not taken into account in the processing of CGPS observations. During the last year, several data sets of various atmospheric loading predictions are available for research purposes at http://www.sbl.statkart.no. At the end of this thesis, possible relations between the vertical components of this data and CGPS time series will be investigated. Separate PCAs will be made for a network of common stations of both CGPS data and atmospheric loading predictions. Then an attempt will be made to identify and compare extracted modes for the two data sets. If possible, a model relation will be derived.

1.1.1 The problem of observing land motion with GPS

Considering CGPS sites on the continuously deforming Earth's surface, as the corners or nodes of a polyhedron, this makes the Earth in to a multi-corner body. CGPS networks measure the surface displacements as position changes of the nodes in time (velocities). Some of the contributions to the surface displacement can be modelled while others are still rather uncertain. Therefore, much is included in a model for station movements commonly known as the station motion model. It is a goal to improve this model. For each of the three components defined in a terrestrial reference frame as Cartesian coordinates, or alternatively as north, east and vertical components in a local topocentric system, the position of

a point $x(t)$ of epoch t can be represented in the station motion model by:

$$x(t) = x_0 + v_0(t - t_0) + \varepsilon(t) \tag{1.1}$$

where x_0 and v_0 are the position and velocity at epoch $t = t_0$, and $\varepsilon(t)$ is an error term. Ideally the only unknowns in the station motion model in Eq. (1.1) are x_0 and v_0. Then Eq. (1.1) expresses a general equation for a basic linear regression problem, fitting a straight line through a series of points $x(t)$ at epochs t. The surface of the Earth is constantly displaced due to various internal and external forces. Examples are the tectonic forces from the Earth's interior, post glacial rebound, surface loading due to atmosphere, ocean, cryosphere, terrestrial hydrosphere, man made or not. External forces include Earth tides due to the Moon, the Sun and planets, also a tide-like force induced by polar motion. The various factors have different spatial wave lengths and different temporal scales. Some forces such as solid Earth tides and ocean tides are often considered as reasonably well known and are then already included in the GPS processing stage. Thus the CGPS time series $x(t)$ is already corrected for these effects. From several geophysical or other processes one may in a general set:

$$\Delta(t) = \sum_{j=1}^{k} \delta x_j(t) \tag{1.2}$$

where $\Delta(t)$ is a sum of corrections $\delta x_j(t)$ due to the j-th process including time- and place variable contributions to the coordinates. The corrections can be applied to the station motion model in either of two ways. For known error free effects, the time series $x(t)$ can be corrected directly before the regression. Or, for unknown effects with known functional relations to the observations, $\Delta(t)$ may be added to the right hand side of Eq. (1.1) and the unknowns δx_j be solved in a linear regression. In addition to noise, the time series of residuals after a linear regression still have the effects that are not included in the station motion model. This means that only the known effects that are corrected or solved for are modelled. Possible unknown effects that are not understood are still left in the noise part $\varepsilon(t)$. Perhaps even more dangerous for a parameter interpretation are, that these unknown effects may affect and distort the parameter estimates of the effects that already assumed

to be known and solved for.

1.1.2 Approach to the solution.

Several time series in a regional network are available. A goal is to separate the understandable part, the signal, from the not understandable which is then called noise. Based on a search for realistic uncertainties and measures of accuracy of observations and parameter estimates, the discussion still goes on whether the remaining series of CGPS residuals from regressions represents pure random noise or not. For these purposes stochastic models including the use of a full a priori covariance matrix are commonly investigated. An attempt to improve existing models will require another focus. Searching for a complete separation of all contributing effects to a CGPS time series, the use of other than a diagonal covariance matrix may affect and distort possible interpretations of the solved parameters. From this point of view, a deterministic approach should be preferred. All known effects from land motions will be identified as model predictions in the observed observations. Interacting processes may be difficult to separate from each other. They may appear as highly correlated parameter estimates from the regression. For example, using too short CGPS time series to solve long-term trends may lead to the problem of over-parameterisation. A detailed examination of univariate series is essential to investigate such possible relations.

1.2 Analysis strategy

First, possible outliers, data offsets and highly correlated parameter estimates will be investigated successive to the implementation of spectral analyses of each single series. Second, the PCA procedure is carried out for the residual CGPS series, attempting to characterise any common effects in the region.

An analysis strategy for the deterministic model improvements of univariate time series of CGPS data may be described in a schematic way as in Figure 1.1. It is commonly known as trend analysis, for which a deterministic model will be fitted to the observations. The differences between the observations and a fitted deterministic model, will be denoted residuals. The residuals can be considered

as estimates for the unknown true errors which the derived models are based on. A critical phase using this method is the choice of an optimal deterministic model which is performed in a least squares sense.

In an early phase of model identification, the correct choice of an optimal model is difficult. An incomplete deterministic model will introduce a bias or a systematic deviation or linear dependency between the residuals. The analysing procedure has to contain a test for outliers or suspicious observations. For the introduction of a test variable, some assumptions of density functions representing the errors have to be done. In Subsection 3.2.1 a test for the detection of outliers is used for the numerical investigations in Chapter 6. This test assumes that the time series of normalised errors are independently normally distributed, $N(0,1)$. Some observations may be sorted incorrectly as outliers in an early phase of the analysis. For each improvement of the model, all the observations have to be tested again for possible outliers. A search for an optimal model will thus be an iterative procedure. Diagnostics based on the normalised residuals such as tests of normality or interpretation of their estimated Auto Correlation Function (ACF) will give some information about the success of a chosen model. The residual diagnostic can thus be used in a discussion of the correct choice of model. The results using the likelihood ratio test for simultaneous significance of parameter estimates introduced in Subsection 3.2.3 can also be used. Because a normalised spectrum and the ACF contain identical information,

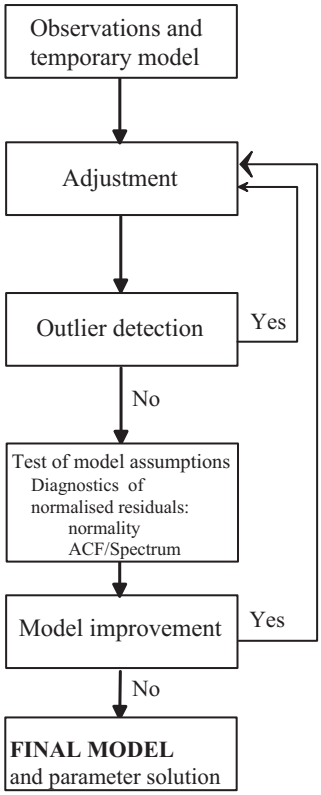

Figure 1.1: Model improvement.

a determination of the spectrum of CGPS data will be an alternative way to find the temporal correlation. A great simplification would be if the adjustment using the final model will generate independent normally distributed normalised residuals. Then the assumptions for the test of simultaneous significance of estimated parameters and the test for outliers will be correct, as well as the use of a diagonal a priori covariance matrix for adjustment.

For the multivariate PCA, it is a goal to use de-trended data. It is a great challenge to ensure that the regressions for univariate series only remove local effects such as jumps and not any of the global effects that should be investigated. As an example, parameters for linear and long periodic trends are highly correlated. Removing a linear trend may influence the long periodic trend in the remaining series. This stresses the investigation of each single series before de-trending is performed.

1.3 Reference systems and frames

The determination of high precision satellite orbits requires a set of differential equations to be solved in a well-defined reference system, or in fact two reference systems are used. A terrestrial reference system which is co-rotating with the Earth "fixed" to its crust, and a celestial reference system "fixed" to the celestial sphere. The relative orientation between the two reference systems is represented by five Earth Orientation Parameters (EOP). Reference systems can be materialised through a reference frame, such as a set of coordinates for a network of stations valid for a certain reference epoch. Two standard reference systems defined for space geodetic purposes are the International Terrestrial Reference System (ITRS) and the International Celestial Reference System (ICRS), defined in the IERS conventions (2003) [McCarthy and Petit, 2004] realised by the International Earth Rotation Service (IERS). Based on available space geodetic observation techniques and the analysis methods for these observations, the reference systems can be realised through appropriate reference frames such as the International Terrestrial Reference Frame (ITRF) or the International Celestial Reference Frame (ICRF)

Assume the points in a global network define the corners of a polyhedron, that is as a multi-corner body. The points may either define the corners of a terrestrial polyhedron, or projecting the directions from very distant radio sources to points on the celestial sphere, they can also define a celestial polyhedron. The Earth's surface is continuously deformed with time due to internal and external forces. In general all points on (and in) the Earth move relative to each other, so the terrestrial polyhedron will deform in time. In extreme cases, the relative motion of the corners can reach up to 20 cm per year [Andersen, 2002]. Also the celestial polyhedron will deform in time, but this deformation rate is several orders of magnitude smaller and can thus be neglected for a time interval of a few decades. If the coordinates of the corners are known, the scale and each of the three axes of a reference frame are given implicitly. If the velocities of the corners are known, the change of direction and scale of the axes with time is given implicitly.

Following the line of thought above, the realisation of ITRS to ITRF are given by a table of coordinates and velocities for the points, valid for a certain reference epoch. The velocities can be considered as estimates for deformations due to geophysical processes. Present computations of coordinates and velocities of the corners in the terrestrial and celestial polyhedrons, and their relating EOP series are based on observations from several space observation techniques (of which GPS is only one), each with its own type specific analysis software. The observations are usually analysed in daily (24h) intervals, thus individual daily realisations of the terrestrial and celestial polyhedrons including consistent estimates of the EOPs are computed. The daily solutions are combined to averages for a time interval of several years. On the basis of analyses carried out by several analysis centres, IERS combines the individual average polyhedrons to one common ITRF solution of the terrestrial polyhedron, as a set of average coordinates and velocities, including full variance matrices, and EOP time series. Currently, ITRF solutions are published nearly annually. The ITRF network has been improved with time in terms of the number of sites and collocations, as well as their distribution over the globe. Mainly because of improved analysis strategies, velocity and coordinate precisions are also improved with time. The current version, the ITRF2000 network contains about 500 sites, intended to be a standard solution for geo-referencing all Earth science applications. Processes to be included in the

analyses and their underlying geophysical models are specified in the IERS conventions (2003) [McCarthy and Petit, 2004]. In May 2001, the latest realisation of ITRS, the ITRF2000 was introduced. It is considered to be the most precise reference frame (RF).

Compared to other space geodetic techniques, monitoring deformations of the Earth's surface, GPS appears to be the more economic solution if highest precision and temporal resolution are required [Prescott, 1996a]; [Prescott, 1996b]; [Savage, 1996]. GPS coordinates and orbits are basically not given in the ITRF frame. Several techniques are used to relate the CGPS time series to a version of ITRF2000, inter- or extrapolated to a certain daily epoch.

1.4 The database

1.4.1 The Norwegian network of 14 CGPS stations

Several data sets are delivered from the Norwegian Mapping Authority (NMA). They have shown to be of various quality. The investigations is mainly based on the third and last data set. These 14 series are of different lengths and contain several intervals of missing data, which makes a consistent multivariate analysis problematic. All Norwegian data have been analysed with the GPS Inferred Position SYstem - Orbit Analysis and SImulation II-software. (GIPSY-OASIS-II), using the fiducial-free Precise Point Positioning technique (PPP) described in [Zumberge et al., 1997]. Fiducial-free precise orbits, time corrections to the satellite clocks, and EOPs from one of the IGS analysis centres, namely NASA's Jet Propulsion Laboratory (JPL), have been used. The satellite orbits, clocks and EOPs are highly constrained, while station clocks and zenith total delays are estimated as stochastic parameters. The station clock bias is estimated as white noise process ($\sigma = 1\mu s$) and the zenith total tropospheric delay as random walk process with $1cm/\sqrt{h}$ as the rate of change of the parameter variance, ($h =$ hour). Furthermore, the azimuth dependency of the ray bending delay of the GPS signals through the troposphere is included by estimating the north-south and the east-west gradient vectors. JPL processes the data from a globally distributed net of about 40 to 50 IGS stations, and the processing is carried out using a free-network

approach [see Heflin et al., 1992]. The resulting orbits, clocks, and EOP are therefore given in a free GPS reference frame, materialised with loose constraints. This free GPS solution then has to be transformed to an already established reference frame such as the latest ITRF. This is done by constraining selected IGS stations given in the ITRF frame. JPL's approach is to convert the solution to the ITRF on a weekly basis. The resulting products are precise orbits, clocks, and EOP given in ITRF. Additionally, JPL provides seven transformation parameters (3 origin offsets (TX, TY, TZ), 3 rotation angles (RX, RY, RZ), and a scale (s)) based on one day's solution. JPL denotes these solutions their quick transformation parameters (qx). The latency on these products are 8-15 days and they are the recommended parameters to be used in precise point positioning using the free-network products. This frame is called the IGS00 reference frame. However, JPL also computes more precise transformation parameters which are based on the solutions from many years which are more rigorously computed. The latency for these products denoted as x-files is up to 6 months. In the transformation from the free GPS solution to ITRF, NMA has only used the precise transformation parameters (x-solutions). The Norwegian data sets of CGPS time series, are closely aligned to ITRF, of which the last one to ITRF2000, is given for the epoch 2002 Jan 01 [Altamimi et al., 2002]. None of the data provided by NMA in this study are corrected for the ocean tidal loading. The CGPS time series used in this study are daily (24h) estimates of station coordinates given for the North, East and Vertical components including standard deviations as a measure of their precision, in a local topocentric system.

1.4.2 The European network of 28 CGPS stations

CGPS data from 28 stations are provided from the JPL website of [Heflin, 2004, (cited: 18 October 2004)]. As for the Norwegian data, these time series are daily estimates found using PPP with no ambiguity resolution. They consist of station coordinates given for the East, North and Vertical components in a local topocentric system. The daily samples given are coordinate variations in ITRF2000 [Altamimi et al., 2002, see] with respect to the station coordinates at the reference epoch 1 January 2000. The samples also include standard deviation of the daily analyses. The satellite offsets and clocks are those provided by JPL [Heflin et al.,

2002] which result in coordinate time series closely aligned to ITRF2000. Only the Vertical component has been investigated, where the variability is at least factor three larger than the horizontal components. The time interval was selected to account for a trade off between maximum spatial coverage and length of time series. Moreover, data quality has improved considerably over time and this gives a higher weight to more recent data. The interval 2000.0 to 2004.0 was found to have a sufficient number of records covering most of the time window. All selected stations have been used for the determination of ITRF2000.

1.4.3 Atmospheric surface loading predictions

Atmospheric loading predictions were taken from the web page of the Special Bureau for Loading (SBL) [http://sbl.gdiv.statkart.no], see van Dam et al. [2003] for a discription of the Global Geophysical Fluid Center (GGFC), which is part of the International Earth Rotation and Reference Systems Service (IERS). Five different sets of time series for most of the ITRF sites are provided, see Table 6.7. These sets (P1-P5) differ mainly with respect to the computation of the air pressure anomaly but also in the computational method, the applied Earth model, and the reference frame. Two pairs are identical in computation and Earth model but different in the reference frame namely, sets P1+P3 and P2+P4, respectively. The first set in each pair uses pressure anomalies computed from the National Center for Environmental Prediction (NCEP) surface pressure data. The second set is created on the basis of the European Centre for Medium Range Weather Forecast (ECMWF) sea level pressure data. In both cases, the anomalies are computed as the difference between each six-hourly pressure field and a reference surface pressure computed from multi-year data sets [van Dam, 2004, in review]. The load-induced displacements are computed using either Green's function for a spatial convolution with the load anomaly [Farrell, 1972] or the corresponding Load Love Numbers in a summation of a spherical harmonic expansion of the load [van Dam et al., 2003]. The Load Love Numbers used are computed for spherical symmetric, non-rotating, elastic and isotropic Earth models, The two reference frames used in the computation of the load-induced displacements have their origin in the Centre of mass of the solid Earth (CE) and in the Centre of Mass of the Earth system (CM), which for the particular loading computations consists of the solid

Earth plus atmosphere, respectively. The difference between these frames is discussed in detail by Blewitt [2002]. The reference frame as realised by the time series of Heflin [2004] is an approximation to ITRF, which with respect to its origin is close to a Centre of Figure of the solid Earth frame (CF). The CF frame is very similar to the CM frame [Blewitt, 2002] and therefore, we expect that the CE series provide predictions better representing the loading signal as captured by the CGPS time series than the CM predictions.

1.5 Author's contribution

CGPS stations have been monitoring surface displacements for up to ten years in increasingly denser regional and global networks. In the early stage of this study (2000), no smoothly working strategy to handle CGPS time series in a statistical defensible way had been implemented. The author hopes that this thesis will make a contribution to the investigation of methods and models applied for CGPS time series, such as outlier detection, jump detection, estimation of correlation functions and spectral analysis for non-stationary CGPS time series. Some statistical tests, such as for outliers, parameters, amplitudes in frequency domain or correlation matrices are found suitable. Hopefully, the introduction and investigation of PCA will make sense for future work with multivariate CGPS time series.

Careful examination of the Norwegian network of CGPS data is made for univariate data series. A detailed description of search procedures for jumps which are unknown in time has been performed, as well as estimation techniques for correlation functions and spectral analysis of non-stationary data. A final goal was to determine and possibly interpret and separate local and regional effects in data. This is done through successive model improvements and interpretations. Finally, PCAs have been made for all three directions to investigate and determine possible common modes in the Norwegian CGPS data.

From a network of European CGPS stations, a PCA for the vertical component has been performed. Common modes have been investigated trying to identify their causes. Separate PCAs are made for five data sets of atmospheric loading surface predictions for 28 identical sites to the European CGPS stations. Extracted modes

from the separate PCAs of atmospheric loading predictions are then compared and correlated with the modes identified in the CGPS data. An interpretation of possible connections has been made in an attempt to find a model that can apply corrections for atmospheric surface loading effects on processed GPS data in the future.

Some parts of this thesis have already been presented at international conferences and also submitted for publication in refereed journals. These are:

T.A. Haakonsen and H. Nahavandchi. Error analysis of GPS/SATREF data. Presentation at Geodesi og Hydrografidagene, Hønefoss, Norway, 2003.

T.A. Haakonsen and H. Nahavandchi. Correlation and spectral analysis of the Norwegian permanent GPS stations using a Least-squares procedure. Presentation in European Geosciences Union, Nice, France, 2004a.

T.A. Haakonsen, H. Nahavandchi, and H-P. Plag. Determination of the Error-Spectrum of Continuous GPS Observations. Presentation in EGS-AGU-EUG Joint Assembly, Nice, France, 2003.

T.A. Haakonsen, H. Nahavandchi, and H-P. Plag. Principal Component Analysis of the Norwegian permanent GPS stations. Presentation in European Geosciences Union, Nice, France, 2004a.

T.A. Haakonsen, H-P. Plag, and H.P. Kierulf. Common modes in GPS-determined vertical land motion in Europe and their relation to atmospheric loading. J. Geophys. Res., 2004b. Submitted.

T.A. Haakonsen, H-P. Plag, H.P. Kierulf, and G. Blewitt. Improving reference frame stability by modelling common modes of surface displacements using Empirical Orthogonal Functions. Presentation AGU Fall Meeting, San Francisco, USA, 2004c.

1.6 Outline of the thesis

The first chapter of this thesis is meant to explain the background for the geodetic tasks to be investigated. Observations or measurements taken as functions of time are rarely independent. A vital question is how to get a measure of this temporal dependency when it is possible to get only one sample at one specific place and interval of time. The time series analysis makes use of the assumption weak stationarity, discussed in Chapter 2. Estimates of Covariance- and Correlation functions may act as measures of linear dependency. Most common estimators for these functions assume complete data series, free of gaps, recorded at equidistant intervals. The CGPS data do not fulfil these requirements, neither are they weakly stationary. To overcome these problems, the theory of time series analysis is combined with the use of regression models, described in Chapter 3. A linear regression model is commonly known by geodesists as a linear version of the Least Squares- or the Gauss-Markov model. These models can be partitioned in one mathematical functional or a deterministic part, and one random or stochastic part. A correct decomposition of a time series in these two parts may be hard or impossible to carry out. A main challenge will be to find the best fitted model to the observations. To represent the series in the frequency domain, the combination of non-stationary time series and regression models leads to the theory of LSSA described in Chapter 4. At the end of the chapter, some simulations with synthetic data are presented to illustrate the separation problem. In Chapter 5, the PCA theory that assumes de-trended data is presented. Finally Chapter 6 is devoted to the numerical investigations, followed by the conclusions and recommendations in Chapter 7

Chapter 2

Stochastic processes and spectral analysis

This chapter will give a brief introduction to the theory of stochastic processes and spectral analysis. For further details about these subjects see Shumway and Stoffer [2000], Wei [1990], Newland [1993], Box et al. [1994], Kedem and Fokianos [2000],Priestley [1981] and Howell [2001].

CGPS time series is expected to be influnced by periodicity, such as points on the Earth's surface responding with differing vertical displacements in the summer than in the winter. A single time series is commonly plotted in a space or domain, using time as argument along the horizontal axis. If periodicity in a time series should be investigated, it is possible to reproduce the series from the time domain in another space or domain with the period as an argument along the horizontal axis. This also applies its inverse, the frequency as argument. The vertical axis in this new domain, the frequency or spectral domain, represents the influence from each frequency component to the original time series. The relations between the two domains can be mathematically expressed through a mapping function or a transformation which introduces an evident tool for time series analysis. Depending on the purposes, one can swich between the two domains for any view of the series that is required. Under certain conditions restricted to the time series, one possible transformation can be the well-known Fourier transform. This is a linear transformation designed using orthogonal base functions, thus this gives an

ortogonal transformation, with several advantages. A complete reconstruction of the series from the frequency to time domain is possible with the inverse Fourier transform. Considering that a basis for the searched transformation is a representation of the original time series in frequency domain, it may not be a surprise that the coefficients representing the linear relation between the two spaces are found from functions represented by sine- and cosine terms. Thus the transformation from the time to the frequency domain may also be interpreted as a correlation procedure. Frequencies can be represented by combinations of sine- and cosine terms. Composed frequency components in a certain range are all correlated with the time series, one by one. The response becomes large for similarities between the series and the composed frequencies.

2.1 Stochastic processes

2.1.1 Time series analysis

The main intention in time series analysis is often to find an optimal model describing the stochastic process that may have generated the time series. It is usually done by attempting to separate the time series information into one deterministic or "signal" part and one stochastic or "noise" part. Careful modelling is necessary to obtain meaningful results.

A time series may be defined either as:

- A collection of stochastic variables representing a stochastic process, indexed according to the order they are obtained in time, or as
- The sequence of values these stochastic variables may take on during a sample, a data series or a realisation of the stochastic process.

A stochastic process is also explained as an ensemble or (theoretically regarded) an infinite collection of sample functions: $x_j(t)$, $j = 1, 2, ...$, where the time argument t may be continuous: $t \in [-\infty, \infty]$ or discrete: $t = ... - 2, -1, 0, 1, 2, ...$ (with the time unit selected so that the sampling interval corresponds to one time

unit). Time series taken only at specific time intervals, are said to be discrete. Continuous time series also provide only digitised values at discrete intervals. For a discussion of the distinction between continuous and discrete time processes, the reader is referred to Priestley [1981]. In most cases, only one realisation of a stochastic process is available. Dealing with univariate time series only, the index j will be omitted. A usual notation for univariate time series is: $x(t)$ or x_t, where the index $t = 1, 2, ..., n$ will refer to the order the value is obtained in time. A realisation of n stochastic variables $x_1, x_2, ..., x_n$ may also be seen as a sample from a n-dimensional joint distribution function. This joint distribution function should be used for a complete description of the data. However, apart from that, this joint distribution function is hardly ever known. Consequently, this is an unwieldy tool for an analysis of time series.

2.1.2 Stationary and ergodic processes

A process is said to be completely or strictly stationary if its statistical properties do not change with absolute time. A process is $n-$th order weak stationary if all its joint statistical moments up to n exists and are time invariant. A second order weak stationary or a covariance stationary process has constant first and second order statistical moments which means that the mean and variance are constant through time. The covariance and correlation will be functions of the time difference only, and thus they are independent of the absolute time. A Gaussian process is completely defined with its first two statistical moments. In that case second order stationarity is equivalent to completely stationarity. A process will obviously not be stationary in the mean (the first order statistical moment) if a trend appears in the data. If sufficient realisations of a stochastic process are available it would be possible to approximate suitable probability density functions for all stochastic variables at requested times. Parameters, such as for the mean and variance for these density functions could then be estimated using averaging across the whole ensemble. In most cases only one realisation is available. Some assumptions have to be done to make the theory applicable. A stationary stochastic process is said to be ergodic if every particular member of the ensemble is representative for all members Priestley [1981]. Then ensemble averaging is equivalent to averaging along the time axis and implies that all statistical information about the process

can be found in one single realisation. An ergodic process will always be stationary, but a stationary process will not necessarily be ergodic. However, it is common to assume that a stochastic process is ergodic, but this cannot be verified.

2.1.3 General covariance and correlation functions

Different definitions of the covariance and correlation functions exist. The formulas presented in this subsection can be used for equidistant and equally weighted data and are identical with the definitions of Shumway and Stoffer [2000]. The prefix auto is often used when covariances and correlations within a time series is computed. Handling two different series, the prefix cross will be used. Let $\triangleq$ denote a definition and E represents the usual expected value operator from probability and statistics theory.

The general Auto Covariance Function (ACvF) is defined as the second moment product:

$$\gamma_x(t,s) \triangleq E[(x_t - \mu_{x_t})(x_s - \mu_{x_s})] \tag{2.1}$$

The auto covariance is thus a measure of linear dependence between two points in a series observed at different times t and $s = t + \tau$. τ is the time difference, delay or lag between the times s and t. Recall from classical statistics that if $\gamma_x(t,s) = 0$, x_t and x_s are not linear related, but there still may be some dependence structure between them. However, if x_t and x_s are bivariate normal, $\gamma_x(t,s) = 0$ ensures their independence. For $s = t$ the auto covariance reduces to the variance:

$$\gamma_x(t,t) \triangleq E[(x_t - \mu_{x_t})^2] \tag{2.2}$$

Analogous to the correlation coefficient in classical statistics, it is often more convenient to deal with a measure between -1 and 1. This leads to a definition of the general Auto Correlation Function (ACF):

$$\rho_x(t,s) \triangleq \frac{\gamma_x(t,s)}{\sqrt{\gamma_x(s,s)\gamma_x(t,t)}} \tag{2.3}$$

Equivalent functions are also defined for covariances and correlations between two different time series x_t and y_s, the general Cross Covariance Function (CCvF):

$$\gamma_{xy}(t,s) \triangleq E[(x_t - \mu_{x_t})(y_s - \mu_{y_s})] \tag{2.4}$$

and a scaled version of the CCvF with possible values between -1 and 1 is the general Cross Correlation Function (CCF) defined as:

$$\rho_{xy}(t,s) \triangleq \frac{\gamma_{xy}(t,s)}{\sqrt{\gamma_x(t,t)\gamma_y(s,s)}} \tag{2.5}$$

2.1.4 Covariance and correlation functions of ergodic processes

Assuming ergodic processes, it is possible to simplify the more general formulas for covariance- and correlation functions. From second order stationarity assumptions it follows that mean μ, and variance σ^2, will be constant and independent of absolute time, thus:

$$E[x(t)] = E[x(t+\tau)] = \mu_x \qquad \text{and} \qquad \sigma^2_{x(t)} = \sigma^2_{x(t+\tau)} = \sigma^2_x \tag{2.6}$$

and equivalent for $y(t)$. Auto covariance and auto correlations are computed from one time series only. The index x is unnecessary and often removed. so the notation for the ACvF in Eq. (2.1) is simplified to:

$$\gamma(\tau) \triangleq E[(x_t - \mu_x)(x_{t+\tau} - \mu_x)] \tag{2.7}$$

and the ACF:

$$\rho(\tau) \triangleq \frac{\gamma(\tau)}{\gamma(0)} \tag{2.8}$$

where the auto covariance for $\tau = 0$, $\gamma(0)$ is equal to the variance σ^2_x in Eq.(2.6). Now, let $x(t)$ and $y(t)$ be two different ergodic processes of time. The CCvF is

the covariance between $x(t)$ and $y(t)$ as a function of the time delay τ:

$$\text{CCvF} = \quad \gamma_{xy}(\tau) \triangleq \quad E[(x_t - \mu_x)(y_{t+\tau} - \mu_y)] \tag{2.9}$$

Finally the cross correlation function CCF which describes linear dependency between two ergodic processes $x(t)$ and $y(t)$ as a function of time delay τ is:

$$\text{CCF} = \quad \rho_{xy}(\tau) \triangleq \quad \frac{\gamma_{xy}(\tau)}{\sqrt{\gamma_x(0)\gamma_y(0)}} \tag{2.10}$$

2.1.5 Estimation of covariance and correlation functions

Direct estimation for equally spaced and equally weighted data

For discrete time series sampled with constant interval, auto and cross covariances and correlations are computed at the integer steps of the sampling interval. In the discrete case, an integer k will be used as the lag-argument. Let the symbol $\forall$ denote for all.

Estimated/Sample means :

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t \quad (= \hat{\mu}_x) \qquad , \qquad \bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t \quad (= \hat{\mu}_y) \tag{2.11}$$

Estimated/Sample ACvF

$$\hat{\gamma}(k) = \frac{1}{n}\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x}) \qquad \forall \quad k \tag{2.12}$$

Estimated/Sample ACF :

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)} \qquad \forall \quad k \tag{2.13}$$

Estimated/Sample CCvF:

$$\hat{\gamma}_{xy}(k) = \begin{cases} \frac{1}{n}\sum_{t=1}^{n-k}(x_t - \bar{x})(y_{t+k} - \bar{y}) & k \geq 0 \\ \frac{1}{n}\sum_{t=1}^{n+k}(y_t - \bar{y})(x_{t-k} - \bar{x}) & k \leq 0 \end{cases} \tag{2.14}$$

Estimated/Sample CCF :

$$\hat{\rho}_{xy}(k) = \frac{\hat{\gamma}_{xy}(k)}{\sqrt{\hat{\gamma}_x(0)\hat{\gamma}_y(0)}} \qquad \forall \quad k \tag{2.15}$$

Some authors prefer $(n-k)$ instead of n in the denominator of Eqs. (2.12) and (2.14). None of the four estimators are unbiased. This is proved by reseachers, such as Wei [1990]. However, if $\rho(k) \to 0$ as $k \to \infty$ then they are all asymptotically unbiased. As shown in Appendix C these equations can be used to compute the elements in covariance and correlation matrices. Eqs. (2.12) and (2.14) ensure such matrices are positive semidefinite, thus they will be preferred. For a number of n observations, Wei [1990] recommends at most $k = n/4$ estimates are to be computed.

2.1.6 Auto and cross correlation estimation with gaps in data

Most derived estimators require equally spaced data in complete series without gaps. Experiences during the recording phase have often shown that this is hard to make. If one succeeds to sample data with a constant sampling frequency, just a small problem with measuring equipment will cause gaps in sampled observations. A discussion how to handle such gaps still goes on. Some recommend to fill these holes with artificial observations. Others are warning just to do that.

(i) Direct estimation using interpolation

Artificial observations may be estimated as the mean of the total series or values computed from a probability density function approximated with a part of the time series around the "gap"area. Then the estimators in Eqs. (2.12) - (2.15) can be used straightforwardly.

(ii) Direct estimation using valid combinations

Another direct method would be to use valid combinations, representing only the common epochs of two data series in the estimation. For each k the number of valid combinations in two data series $y(t)$ and $x(t)$ is counted, see Figure 2.1. Actually data is stored in two new arrays. Then for each k, Eqs. (2.14) or (2.15) can be used. To find the ACF or ACvF for one data series $x(t)$, the data series $y(t)$ can be set to $x(t)$ in Figure 2.1.

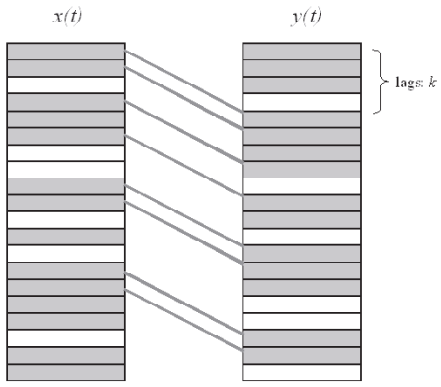

Figure 2.1: Cross correlating two arrays with gaps. White cells represent data gaps.

(iii) Indirect estimation using the inverse Fourier transform

As will be discussed later in Subsection 2.2.3, there is a relation between the ACvF and the Fourier spectrum, and between the ACF and the normalised Fourier spectrum. They are connected through the Fourier Transform and the Inverse Fourier Transform and hence they are said to be Fourier Transform pairs. Provided that the Fourier spectrum or normalised Fourier spectrum is given, it is possible to compute the ACvF or ACF through an inverse Fourier transform.

2.1.7 Weighted and unequally spaced data

A transformation comparable to the Fourier transform is derived for unequally spaced data, [see Vanicek and Krakiwsky, 1986]. This transform denoted the Least Squares Transformation (LST) is developed to also manage observations of different weights, [see Craymer, 1998].

2.2 Spectral analysis, stationary data

The most common method to detect the contents of periodic components of a time series is Fourier Analysis, named after the French mathematican Jean-Baptiste

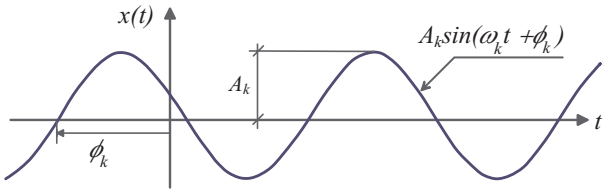

Figure 2.2: A sequence of sine waves

Joseph Fourier (1768-1830). The method is well suitable to investigate time invariant signals of sine wave form, but there are some limitations which should be mentioned. The data have to be second order (weak) stationary, and no trends are allowed. The theory will not be directly applicable to the CGPS data, but an introduction is necessary to point out similarities and differences to the Least Squares Spectral Analysis, introduced in Section 4.2. When data are equally spaced, the use of a computing algoritm, the Fast Fourier Transform (FFT) is customary to find the spectrum. The FFT alogoritm is an efficient tool to improve the computational speed reducing the number of operations from a function of n^2 to a function of $nlog_2 n$. The algorithm is well described in the literature, [see Press et al., 1992, among others] and will be disregarded here.

2.2.1 Wave theory

A sequence of sine waves as in Figure 2.2 may be described using three variables:

1. The amplitude A_k decides the "height" of the wave, e.g. $A_k = 1$ keeps the ordinate within the interval $[-1.1]$.

2. The phase or phase angle $\phi_k \in [-\pi, \pi]$ defines a shift relative to the time origo. A change of phase implies a translation along the abcissa-axis. The positive sign in the paranthesis defines the positive ϕ_k-direction towards the left.

3. The angular frequency $\omega_k = 2\pi k/p$. The index k decides the number of oscillations per period p and can thus distinguish different wavelengths or angular frequencies.

This sequence can be written as:

$$x_k(t) = A_k sin(\omega_k t + \phi_k) \tag{2.16}$$

where the index k is used to distinguish different waves or frequency components. It is not related to the integer k denoting the delay or lags for correlation functions.

2.2.2 Fourier analysis

A periodic function represented by trigonometric series

Fourier's basic premise was that any finite length, infinitely repeated time series $x(t)$, defined over the principal interval $[0, L]$ with special choices of ω_k, can be exactly reproduced using a sufficient sum of sequences or sine waves as in Eq. (2.16) at the form:

$$x(t) \triangleq A_0 + \sum_k A_k sin(\omega_k t + \phi_k) \tag{2.17}$$

By summing up just a few simple sine waves, it is possible to create rather complicated looking functions. With the indentity:

$$sin(\alpha + \beta) = sin\alpha \cdot cos\beta + cos\alpha \cdot sin\beta \tag{2.18}$$

$x(t)$ can also be written as:

$$x(t) = A_0 + \sum_k A_k \left[sin(\omega_k t) cos\phi_k + cos(\omega_k t) sin\phi_k\right] \tag{2.19}$$

and by setting:

$$a_0 = 2A_0 \qquad a_k = A_k sin\phi_k \qquad b_k = A_k cos\phi_k \tag{2.20}$$

one finally arrives the more common expression:

$$x(t) = \frac{a_0}{2} + \sum_k \left[a_k cos(\omega_k t) + b_k sin(\omega_k t)\right] \tag{2.21}$$

known as the Fourier series of $x(t)$ in which $a_0/2$ is the mean value of the record

of length L, and a_k and b_k are constants. In signal analysis, the record average is also denoted the DC-offset or DC-level (Direct Current). Limiting k to integers makes some favourable properties. The coefficients a_k and b_k are then called Fourier coefficients and the periodic trigonometric functions in Eq. (2.21) become orthogonal base functions, which means that the coefficients for a given angular frequency can be computed independently from the others. Provided enough Fourier components are used, each value of the original series can be totally reconstructed over the principle interval.

Amplitudes A_k and phase angles ϕ_k are now given by:

$$A_k = \sqrt{a_k^2 + b_k^2} \qquad \phi_k = \text{arctan}(a_k/b_k) \qquad (k \geq 1) \tag{2.22}$$

Note that other definitions of the phase angle exists. Some refer the phase angle to a cosine-wave, which is in fact nothing but a phase shifted sine-wave.

The Fourier coefficients, can be written as [see e.g Newland, 1993] :

$$\underset{k=0}{A_0} = \frac{1}{L}\int_0^L x(t)dt = \frac{1}{L}\int_{-L/2}^{L/2} x(t)dt \tag{2.23}$$

$$\underset{k\geq 1}{a_k} = \frac{2}{L}\int_0^L x(t)cos(\omega_k t)dt = \frac{2}{L}\int_{-L/2}^{L/2} x(t)cos(\omega_k t)dt \tag{2.24}$$

$$\underset{k\geq 1}{b_k} = \frac{2}{L}\int_0^L x(t)sin(\omega_k t)dt = \frac{2}{L}\int_{-L/2}^{L/2} x(t)sin(\omega_k t)dt \tag{2.25}$$

The kernals in Eqs. (2.23) - (2.25) are all periodic functions with period L. Thus the values of these integrals remain unchanged if the interval of integration $(0, L)$ is replaced with any other interval of length L, for instance the symmetric interval $(-L/2, L/2)$.

The angular frequency ω_k can be expressed as an integer multiple of the fundamental angular frequency $\omega_1 = 2\pi/L$:

$$\omega_k = k\omega_1 = \frac{2\pi k}{L} \tag{2.26}$$

The space between the Fourier- or integer angular frequencies is :

$$\Delta\omega = \frac{2\pi(k+1)}{L} - \frac{2\pi k}{L} = \frac{2\pi}{L} \qquad (= \omega_1) \tag{2.27}$$

The angular frequency is related to the frequency f as:

$$\omega = 2\pi f \tag{2.28}$$

which again is the inverse of a period or a wave length. The Fourier coefficients represent the strength of the different frequencies in the analysed time series. A possible way to visualise the frequency content of a data series would be to use the obtained Fourier coefficients to compute the amplitudes A_k and phases from Eq. (2.22) , and plot them in the frequency domain (as a function of frequency). They are commonly known as the amplitude- and phase spectrum. However, different kinds of spectra are used. Later in this thesis, focus will be put on finding the frequency content of stochastic time series, which is related to the corresponding Spectral Density Function (SDF).

Non periodic functions

Non periodic functions cannot be represented by Fourier series, as in Eq.(2.21), but subject to certain conditions, one can still follow the line of thought as above. It is possible to define a new periodic function with a period equal to the length of the non periodic function in a finite interval, and then assume this function will repeat itself at the infinity t-axis, outside this interval. The reader is refered to Priestley [1981].

Discrete Fourier Transformation (DFT)

It is not possible to record a continuous process of time for all time points or for infinitely long time. Observations or measurements of a stochastic process are often attempted to be digitised at equidistant intervals, as a discrete time series of finite duration. A goal is to find the frequency content of the infinite long continuous series, analysing the available recorded discrete time series. Assume x_i is sampled with constant interval Δt with finite duration L. The total of n sample points exists. Their locations in time are: $x(i\Delta t) \triangleq x_i (i = 1, 2, ..., n)$. One

wants to reproduce the original time series as a sum of trigonometric terms of different amplitudes and phases, equivalently a sum of Fourier coefficient pairs. Assume an even number n, of available observations. The integrals in Eqs. (2.24) and (2.25) may be replaced approximately by the sums [see Emery and Thomson, 2001, among others];

$$a_k = \frac{2}{n}\sum_{i=1}^{n} x_i cos(2\pi ki/n) \qquad k = 0, 1, 2, ...(n/2) \tag{2.29}$$

$$b_k = \frac{2}{n}\sum_{i=1}^{n} x_i sin(2\pi ki/n) \qquad k = 1, 2...(n/2) \tag{2.30}$$

Eqs. (2.29) and (2.30) are known as Discrete Fourier Transforms (DFT). Because of symmetry and the difficulties in an interpretation of negative frequencies, these equations are simplified by a factor 2, and derived to be computed only for positive integers of k. The finite discrete trigonometric Fourier series will be denoted the Inverse Discrete Fourier Transform (IDFT) and become:

$$x_i = \frac{a_0}{2} + \sum_{k=1}^{n/2} [a_k cos(2\pi ki/n) + b_k sin(2\pi ki/n)] \qquad i = 1, ..., n \tag{2.31}$$

Introducing j as the imaginary unit and set:

$$cos(2\pi ki/n) = \frac{e^{j2\pi ki/n} + e^{-j2\pi ki/n}}{2} \tag{2.32}$$

and

$$sin(2\pi ki/n) = \frac{e^{j2\pi ki/n} - e^{-j2\pi ki/n}}{2j} \tag{2.33}$$

the finite trigonometric Fourier series in Eq. (2.31) may also be rewritten in terms of complex exponentials as the Inverse Complex Discrete Fourier Transform (ICDFT):

$$x_i = \sum_{k=0}^{n-1} X_k e^{j2\pi ki/n} \qquad i = 1, 2, ..., n \tag{2.34}$$

and the Complex Discrete Fourier Transform (CDFT) as:

$$X_k = \frac{1}{n}\sum_{i=1}^{n} x_i e^{-j(2\pi ki)/n} \qquad k = 0, 1, 2, ..., n \tag{2.35}$$

The Nyquist frequency

As can be seen from the Eqs. (2.29) and (2.30), an upper limit for the number of integer frequencies to be computed is decided by the Nyquist frequency or the folding frequency when $k = n/2$. Outside this range of frequencies, the Fourier coefficients will repeat themselves, and even more dangerous, the constituents of frequency components in data outside this range will be reflected or mirrored inside the range to be investigated and distort the spectrum. The phenomena is called aliasing, and care must be taken also for estimated values inside the limits, particularly if they are close to the Nyquist frequency. In fact, it is not re-commended to estimate Fourier coefficients from Eqs. (2.29) and (2.30) for both $k = 0$ (the DC-offset) and $k = n/2$ (the Nyquist frequency) because they are aliased and thus contain some common frequency information.

The first or lower angular frequency to be computed is the fundamental angular frequency $\omega_1 = \omega_{lowlim}$, limited by the record length $L = n\Delta t$. Setting maximum $k = (n/2)$ we find the upper limit ω_{uplim} to be detected from a sample as:

$$(\omega_{n/2} =) \qquad \omega_{uplim} = \frac{n}{2} \cdot \omega_{lowlim} = \frac{n}{2} \cdot \frac{2\pi}{L} = \frac{\pi}{\Delta t} \tag{2.36}$$

which is decided by the sampling interval, Δt. Deciding a correct sampling interval, Δt is therefore of great importance. As an example, the satellites in the GPS system have approximately a semi-diurnal cycle. CGPS data used to compute the Fourier coefficients, have to be averaged over one or several complete GPS periods, e.g. a diurnal cycle. If not, estimates of Fourier coefficients for angular frequencies greater than the Nyquist frequency, will reflect and distort coefficients found for smaller frequencies.

Table 2.1: Limits for different spectrum arguments

Argument	**Symb.**	**Unit**	**Lower limit**	**Upper limit**
angular freq.	ω	[rad/time unit]	$2\pi/L$	$\pi/\Delta t$
frequency	f	[cycles/time unit]	$1/L$	$1/2\Delta t$
period	f^{-1}	[time units/cycle]	$2\Delta t$	L

There are different kinds and definitions of a spectrum. Because of the random nature of measured data (recorded time series), spectral density functions have been defined. A better interpretation may be made using the frequency $f = 2\pi/\omega$, or the inverse frequency, the period f^{-1} as arguments. Limits for different spectrum arguments are listed in Table 2.1. Their relations can be seen from the example of an amplitude spectrum shown in Figure 2.3. For stochastic time series, the spectrum is commonly written as continuous over the interval of permitted frequencies.

Phenomena in the nature is often expected to be of periodic character. The main task of spectral analysis is to find the unknown angular frequencies ω_k. and the coefficients a_k and b_k based on the measurements of the phenomena e.g. a recorded time series. Periodic functions as in Eq. (2.21) are often unknown. The angular frequency ω_k makes this to a non linear problem. However, if the angular frequencies are known (for instance chosen to be Fourier- or integer frequencies) the remaining task, to evaluate a_k and b_k may be solved using linear models. Significant frequency components found from a recorded time series hardly ever coincide with the integer Fourier frequencies. As an example, assume we want to investigate the importance of periods like $f^{-1} = 2L/3$ in the lower part of Figure 2.3. The Fourier analysis will not give a proper answer, because it only computes values for the periods $f^{-1} = L/2$ and $f^{-1} = L$. Later, the Least Squares Spectral Analysis (LSSA) will be presented. A main advantage of using LSSA is the possibility to make a zoomable argument. However, it is very demanding in terms of orthogonality to make this possible. The advantages of orthogonality will be left and coefficients a_k and b_k cannot be computed independent of each other any longer.

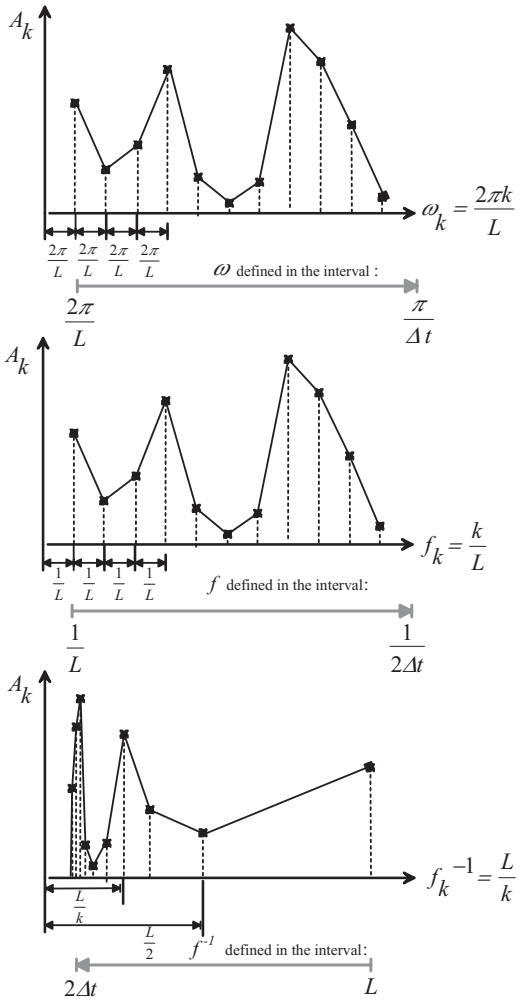

Figure 2.3: Example of Amplitude spectrum with different arguments (1) angular frequency ω, (2) frequency f, (3) period (inverse frequency) f^{-1}

2.2.3 Spectral density

Consider $x(t)$ as a zero mean, weak stationary process with $E(x_t) = \mu_x = 0$.
The time history of a weak stationary sample function $x(t)$ is not periodic. The

condition:

$$\int_{-\infty}^{\infty} |x(t)|dt < \infty \tag{2.37}$$

is not satisfied so $x(t)$ cannot be represented by a discrete Fourier series, but nevertheless the ACvF contains information about the frequency contents of the process $x(t)$. Instead of $x(t)$, the ACvF, $\gamma_x(\tau)$ of $x(t)$ is analysed. Provided that $\gamma_x(\tau)$ in Eq. (2.7), fulfils the conditions:

$$\gamma_x(\tau) \to 0 \quad \text{when} \quad \tau \to \infty \qquad \text{and} \qquad \int_{-\infty}^{\infty} |\gamma_x(\tau)|d\tau < \infty \tag{2.38}$$

the Fourier Transform of $\gamma_x(\tau)$ and its inverse may be defined [Newland, 1993, see] as:

$$S_x(\omega) \triangleq \frac{1}{2\pi} \int_{-\infty}^{\infty} \gamma_x(\tau) e^{-j\omega\tau} d\tau \tag{2.39}$$

and

$$\gamma_x(\tau) \triangleq \int_{-\infty}^{\infty} S_x(\omega) e^{j\omega\tau} d\omega \tag{2.40}$$

$S_x(\omega)$ is a function of angular frequency ω, denoted the spectral density function of the process $x(t)$.

$S_x(\omega)$ has the properties:

1. $S_x(\omega)$ is a real even function of ω.
2. $S_x(\omega)$ is never negative.
3. $S_x(\omega)$ is symmetric around $(\omega = 0)$ (if $\gamma_x(\tau)$ is real).

In the special case when $(\tau = 0)$ in Eq. (2.40) one gets:

$$\gamma_x(\tau = 0) = \int_{-\infty}^{\infty} S_x(\omega) d\omega \tag{2.41}$$

which is the total variance, $\gamma_x(0)$. Graphically it represents the area under the $S_x(\omega)$-curve. If $x(t)$ is a zero mean process, the total variance equals the mean square value so it is possible to write

$$\gamma_x(0) = E[x(t)^2] = \frac{1}{n}\int_{-\infty}^{\infty}[x(t)]^2 dt = \int_{-\infty}^{\infty} S_x(\omega)d\omega \tag{2.42}$$

which is a form of Bessel's equality scaled with $1/n$, see Howell [2001]. Note that the area of $S_x(\omega)$ between the limits $-\omega_k$ and ω_k is

$$\int_{-\omega_k}^{\omega_k} S_x(\omega)d\omega \tag{2.43}$$

for any ω_k, and whenever ω_k is an integer frequency or not, this part of the total area is a positive value representing the portion of the total variance of x_t, attributed by components of angular frequencies with absolute value less than ω_k.

Normalised spectral density function versus ACF

The spectral density function in Eq.(2.39) and the ACvF (2.40) is related through the FT and its inverse. Dividing Eqs. (2.39) and (2.40) by $\gamma_x(0)$, it can be shown that the ACF and a normalised version of the spectral density function also are FT-pairs. $S_x(\omega)_{norm}$

$$S_x(\omega)_{norm} \quad \triangleq \frac{1}{2\pi}\int_{-\infty}^{\infty}\rho_x(\tau)e^{-j\omega\tau}d\tau \tag{2.44}$$

and

$$\rho_x(\tau) \triangleq \int_{-\infty}^{\infty} S_x(\omega)_{norm} \quad e^{j\omega\tau}d\omega \tag{2.45}$$

$S_x(\omega)_{norm}$ satisfies:

$$\begin{gathered} S_x(\omega)_{norm} \geq 0 \quad \forall \quad \omega \\ \int_{-\infty}^{\infty} S_x(\omega)_{norm}d\omega = 1 \end{gathered} \tag{2.46}$$

which means $S_x(\omega)_{norm}$ is a probability density function

The contribution to the total variance of x_t caused by angular frequency components in the band ω_k to $\omega_k + \Delta\omega$ is:

$$\int_{\omega_k}^{\omega_k+\Delta\omega} S_x(\omega)_{norm} d\omega \tag{2.47}$$

Three different representations of $x(t)$

A time series may now be given as three different representations, as the time series $x(t)$ itself, as the ACvF, or as the spectral density function $S_x(\omega)$. Investigations of each of these representations give valuable information about which kind of process that may have generated the time series. The ACvF, $\gamma_x(\tau)$ and the $S_x(\omega)$ contain identical information about the frequencies contributing in $x(t)$. They are related through the FT and IFT and are thus denoted as Fourier pairs. Depending of what their spectrum looks like, which is the picture of their frequency content, they can be classified in narrow- or broad band processes.

The two-sided spectral density function is defined for both negative and positive angular frequencies. Recalling that $S_x(\omega)$ is symmetric, and $\gamma_x(\tau)$ is real valued, one can avoid the difficulties handling the negative frequencies using the relation:

$$\int_{-\omega_k}^{\omega_k} S_x(\omega) d\omega = 2 \int_0^{\omega_k} S_x(\omega) d\omega \tag{2.48}$$

to find the one-sided spectrum.

Estimation of a spectrum (Spectral values)

In the past, a common way to compute values in a spectrum, estimated spectral values $\hat{S}(\omega_k)$ or spectral coefficients S_k, was first to estimate the ACvF from Eq. (2.40), then during a Fourier transform computing the estimates for the spectral density Eq. (2.39). After the invention of Fast Fourier Transform (FFT), a newer so-called direct method was produced to obtain the estimates S_k of $S_x(\omega)$. The computation of the ACvF is not really necessary. Instead, the CDFT of the time

series itself will be computed to get estimates for X_k in Eq. (2.35). The relation between S_k and X_k is

$$S_k = X_k^* X_k = X_{-k} X_k = |X_k|^2 \tag{2.49}$$

where $X_{-k} = X_k^*$ is the complex conjugate of X_k, that can be computed directly from the coefficients in Eqs. (2.29) and (2.30) with:

$$X_k = \begin{cases} \frac{1}{2}(a_k - jb_k) & k \geq 1 \\ \frac{1}{2}a_0 & k = 0 \\ \frac{1}{2}(a_{|k|} + jb_{|k|}) & k \leq -1 \end{cases} \tag{2.50}$$

Estimates for the normalised spectral values (NSV) may be found as:

$$S_{k-norm} = \frac{S_k}{\sum_{k=1}^{n} S_k} \tag{2.51}$$

Physical interpretation

The total energy of a process $x(t)$ may be defined as

$$\text{Total energy} \triangleq \int_{-\infty}^{\infty} [x(t)]^2 dt \qquad = n\gamma_x(0) \tag{2.52}$$

which is not a finite measure. It is usually more convenient to define the total power as total energy per unit of time over the finite interval equal to the length L of the recorded time series.

$$\text{Total power} \triangleq \frac{1}{L} \int_0^L [x(t)]^2 dt \tag{2.53}$$

Using the available discrete data of finite length $L = n\Delta t$, the integral of Eq. (2.53) can be approximated by a Riemann sum, and a discrete expression for the total power would be

$$\frac{1}{n\Delta t} \sum_{i=1}^{n} x_i^2 \Delta t = \frac{1}{n} \sum_{i=1}^{n} x_i^2 = \hat{\gamma}_x(0) \tag{2.54}$$

for a zero mean process, which again can be expressed as the finite sum of spectral values in the range $k = 0$ to $k = n - 1$, $\sum_{k=0}^{n-1} |X_k|^2$ and explains why the phrase power spectrum is commonly used for $S_x(\omega)$.

2.2.4 Matrix notations

Let the time series of a zero-mean process (with $a_0/2$-term equal zero), be represented in a time shifted even numbered observation vector $\boldsymbol{y} = [x_1, x_2, ..., x_i, ..., x_n]^T$. The Fourier coefficients up to the Nyquist frequency may be collected in a vector:

$$\boldsymbol{\beta} = \underbrace{[a_1, b_1, a_2, b_2, ..., a_k, b_k, ..., a_{(n/2)-1}, b_{(n/2)-1}]^T}_{(1 \cdot (n-2))} \tag{2.55}$$

For $k = 1, 2, ..., (n/2) - 1$ defining column vectors for cosine- and sine terms of length n:

$$\boldsymbol{c}_{\omega_k} = \begin{bmatrix} cos(2\pi k1/n) \\ cos(2\pi k2/n) \\ \vdots \\ cos(2\pi ki/n) \\ \vdots \\ cos(2\pi kn)/n) \end{bmatrix} \qquad \boldsymbol{s}_{\omega_k} = \begin{bmatrix} sin(2\pi k1/n) \\ sin(2\pi k2/n) \\ \vdots \\ sin(2\pi ki/n) \\ \vdots \\ sin(2\pi kn/n) \end{bmatrix} \tag{2.56}$$

The column vectors may be collected again in a design matrix:

$$\boldsymbol{X} = \underbrace{[\boldsymbol{c}_{\omega_1} \vdots \boldsymbol{s}_{\omega_1} \vdots \boldsymbol{c}_{\omega_2} \vdots \boldsymbol{s}_{\omega_2} \vdots ... \vdots \boldsymbol{c}_{\omega_k} \vdots \boldsymbol{s}_{\omega_k} \vdots ... \vdots \boldsymbol{c}_{\omega_{(n/2)-1}} \vdots \boldsymbol{s}_{\omega_{(n/2)-1}}]}_{(n \cdot (n-2))} \tag{2.57}$$

So the system of linear equations in Eq. (2.31) can be expressed as:

$$\underset{(n \cdot 1)}{\boldsymbol{y}} = \underset{(n \cdot (n-2)) \cdot ((n-2) \cdot 1)}{\boldsymbol{X}\boldsymbol{\beta}} \tag{2.58}$$

Because of length($\boldsymbol{y}$) $>$ length($\boldsymbol{\beta}$), this gives an overdetermined system of linear

equations for a solution of $\boldsymbol{\beta}$, that can be found using the least squares estimator:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y} \tag{2.59}$$

with:

$$\boldsymbol{X}^{\boldsymbol{T}} = [\boldsymbol{c}_{\omega_1}^{\boldsymbol{T}} \vdots \boldsymbol{s}_{\omega_1}^{\boldsymbol{T}} \vdots \boldsymbol{c}_{\omega_2}^{\boldsymbol{T}} \vdots \boldsymbol{s}_{\omega_2}^{\boldsymbol{T}} \vdots \ldots \vdots \boldsymbol{c}_{\omega_k}^{\boldsymbol{T}} \vdots \boldsymbol{s}_{\omega_k}^{\boldsymbol{T}} \vdots \ldots \vdots \boldsymbol{c}_{\omega_{(n/2)-1}}^{\boldsymbol{T}} \vdots \boldsymbol{s}_{\omega_{(n/2)-1}}^{\boldsymbol{T}}]^{\boldsymbol{T}} \tag{2.60}$$

and taking the advantage of special properties because the columns of $\boldsymbol{X}$ are constructed from orthogonal base functions, one get

$$\underset{((n-2)\cdot(n-2))}{\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}} = \begin{bmatrix} \boldsymbol{c}_{\omega_1}^{\boldsymbol{T}}\boldsymbol{c}_{\omega_1} & \boldsymbol{c}_{\omega_1}^{\boldsymbol{T}}\boldsymbol{s}_{\omega_1} & \boldsymbol{c}_{\omega_1}^{\boldsymbol{T}}\boldsymbol{c}_{\omega_2} & \cdots & \cdots \\ \boldsymbol{s}_{\omega_1}^{\boldsymbol{T}}\boldsymbol{c}_{\omega_1} & \boldsymbol{s}_{\omega_1}^{\boldsymbol{T}}\boldsymbol{s}_{\omega_1} & \boldsymbol{s}_{\omega_1}^{\boldsymbol{T}}\boldsymbol{c}_{\omega_2} & \cdots & \cdots \\ \boldsymbol{c}_{\omega_2}^{\boldsymbol{T}}\boldsymbol{c}_{\omega_1} & \boldsymbol{c}_{\omega_2}^{\boldsymbol{T}}\boldsymbol{s}_{\omega_1} & \boldsymbol{c}_{\omega_2}^{\boldsymbol{T}}\boldsymbol{c}_{\omega_2} & \cdots & \cdots \\ \vdots & \vdots & \vdots & \ddots & \\ \vdots & \vdots & \vdots & & \ddots \end{bmatrix} = \frac{n}{2}\cdot\boldsymbol{I} \tag{2.61}$$

because of the sums:

$$\begin{aligned} \boldsymbol{c}_{\boldsymbol{\omega_k}}^{\boldsymbol{T}}\boldsymbol{c}_{\boldsymbol{\omega_l}} &= \textstyle\sum_{i=1}^{n} cos(2\pi ki/n)cos(2\pi li/n) \\ &= \begin{cases} n & \text{if} \quad k = l = n/2 \quad \text{or} \quad n \\ n/2 & \text{if} \quad k = l, \quad l \neq n/2 \quad \text{or} \quad n \\ 0 & \text{else} \quad \forall \quad k \neq l \end{cases} \\ \boldsymbol{s}_{\boldsymbol{\omega_k}}^{\boldsymbol{T}}\boldsymbol{s}_{\boldsymbol{\omega_l}} &= \textstyle\sum_{i=1}^{n} sin(2\pi ki/n)sin(2\pi li/n) \\ &= \begin{cases} 0 & \text{if} \quad k = l = n/2 \quad \text{or} \quad n \\ n/2 & \text{if} \quad k = l, \quad l \neq n/2 \quad \text{or} \quad n \\ 0 & \text{else} \quad \forall \quad k \neq l \end{cases} \\ \boldsymbol{c}_{\boldsymbol{\omega_k}}^{\boldsymbol{T}}\boldsymbol{s}_{\boldsymbol{\omega_l}} &= \textstyle\sum_{i=1}^{n} cos(2\pi ki/n)sin(2\pi li/n) = 0, \forall \quad k \quad \text{and} \quad l \end{aligned} \tag{2.62}$$

and finally a solution for the vector of Fourier coefficients can be found, avoiding a matrix inversion as:

$$\hat{\boldsymbol{\beta}} = \frac{2}{n}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y} \tag{2.63}$$

Chapter 3

Regression models and time series

The formulas in Chapter 2, assume second order weak stationarity, but this is hardly ever the case for recorded data. A question is how to handle the non-stationary time series data. Most of the statistical literature suggest, a differentiation technique to remove all kinds of polynomial trends. This may work for equally spaced data, but for series of unevenly spaced or missing data, this is not a straightforward method.

A more common way to handle non-stationary data may be to use a functional or a deterministic model taking care of all kinds of trends, of which a classical regression model can be one choice. The non-stationary time series will now be used as observations, and the task of the deterministic model is to absorb all kind of trends in the non-stationary time series. In accordance with the principles of least squares adjustment, the outcome of using such a model, will be a new time series of residuals with constant zero mean. For equally weighted input data, the outcome will also satisfy constant variance, and thus the residuals satisfy the second order weak stationarity assumptions. The residuals can be interpreted as estimates for the unknown true errors remaining after the effect of a fitted model has been removed. They are used, both for validation of the chosen deterministic model, and in further analyses, such as a spectral analysis. However, this is a subject that is full of pitfalls. A main challenge is trying to separate the observations in one deterministic and one stochastic part. Only in a very few cases is this possible. If one does not succeed, an erroneous effect will affect the parameters

to be solved in some way. Interpretations and attempts to improve the first chosen models for detected effects during the analysis are possible but difficult. This is mainly because of correlated parameter estimates in the deterministic and the stochastic parts of the model, and also between the two parts.

CGPS time series should actually be seen as 3-D observations as a basis for adjustment, of which the multivariate regression model would be a better choice of model. For very long time series this implies a serious computational burden. In the case of in-dependency among the components north, east and height, a great simplification would be to split them into isolated time series, and then separately use univariate regression models for each components. A likelihood ratio test for the multivariate case will be introduced to test the assumptions of in-dependency between components for each station. Numerical investigations of CGPS data will show that the assumption of independent components can be justified. However, some parameters to be solved for have to be chosen as common for each components in one station. As an example, if a jump is detected in the height component, the jump time is assumed common for all the three directions.

3.1 Classical linear regression

Classical regression models are derived under the assumptions of independent errors, but time series are often time correlated. The temporal correlation cannot be ignored, so an investigation of the residuals remaining after the model fit is essential.

3.1.1 Classical linear regression model

A classical linear regression model [see Johnson and Wichern, 1998, among others] can be described as:

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix}}_{n\times 1} = \underbrace{\begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1r} \\ 1 & x_{21} & x_{22} & \dots & x_{2r} \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{i1} & x_{i2} & \dots & x_{ir} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nr} \end{bmatrix}}_{n\times(r+1)} \cdot \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_j \\ \vdots \\ \beta_r \end{bmatrix}}_{(r+1)\times 1} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{bmatrix}}_{n\times 1} \tag{3.1}$$

or

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{3.2}$$

where $\boldsymbol{y}$ is an observed response vector (observations) and $\boldsymbol{X}$ is the matrix of values from corresponding predictor variables. $\boldsymbol{X}$ may also be denoted as the design matrix. It describes the geometrical and physical relation between the observations in $\boldsymbol{y}$ and the unknown parameters in $\boldsymbol{\beta}$ and thus the design of the experiment. n is an integer number of observations and $(r+1)$ is the integer number of unknown parameters to be estimated. For an overdetermined solution of the system of linear equations one need $n > (r+1)$. $\boldsymbol{\varepsilon}$ is the vector of errors which is assumed to have the properties:

$$E(\boldsymbol{\varepsilon}) = \mathbf{0} \qquad , \qquad \boldsymbol{cov}(\boldsymbol{y}) = \boldsymbol{cov}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I} \tag{3.3}$$

where $\mathbf{0}$ is a $(n \times 1)$ zero vector, $\boldsymbol{I}$ the $(n \times n)$ identity matrix and σ^2 the a priori reference variance or the variance of unit weight. It is common to set $(\sigma^2 = 1)$. The assumptions in Eq.(3.3) mean:

- Single errors ε_i in $\boldsymbol{\varepsilon}$ do have the same constant variance, and
- They are expected to be (linear) independent.

The values of the (theoretical expression) true errors in $\boldsymbol{\varepsilon}$ are impossible to find in most cases. Estimation of this error vector is common. The estimate of $\boldsymbol{\varepsilon}$ is $\hat{\boldsymbol{\varepsilon}}$ and will be denoted the residual vector.

3.1.2 Least squares estimation

Setting $\sigma^2 = 1$, the least squares estimator for the unknown parameters β_j , *j=0,...,r* is the specific vector $\boldsymbol{\beta}$ which minimises the main function m of sum of error squares: $m(\varepsilon^T\varepsilon) = m((\boldsymbol{y} - \boldsymbol{X\beta})^T(\boldsymbol{y} - \boldsymbol{X\beta}))$.

$$\frac{\partial m(\boldsymbol{\varepsilon^T\varepsilon})}{\partial(\boldsymbol{\beta})} = \mathbf{0} \qquad \Leftrightarrow \qquad \hat{\boldsymbol{\beta}} = (\boldsymbol{X^TX})^{-1}\boldsymbol{X^Ty} \tag{3.4}$$

This estimator can be shown to give a minimum value for the sum of error squares, and is identical to the maximum likelihood estimator, assuming errors (or observations) to be normally distributed. The fitted values of $\boldsymbol{y}$ can be found as: $\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{Hy}$, where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X^TX})^{-1}\boldsymbol{X^T}$ is denoted the "hat" matrix. The vector of residuals becomes:

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \hat{\boldsymbol{y}} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y} \tag{3.5}$$

The residual vector $\hat{\boldsymbol{\varepsilon}}$, contains information about the remaining unknown parameter σ^2. It can be shown that the maximum likelihood estimator for the constant reference variance, also known as the variance of unit weight is

$$\hat{\sigma}^2 = \frac{(\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}})}{n} \tag{3.6}$$

The unbiased estimator is more commonly used

$$\hat{\sigma}^2 = \frac{(\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}})}{(n - (r + 1))} \tag{3.7}$$

which is divided by the redundancy or the degrees of freedom, that is corrected by the number of parameters to be solved for. If n is large, the choice between the two estimators has no consequence.

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X^TX})^{-1}\boldsymbol{X^Ty} = (\boldsymbol{X^TX})^{-1}\boldsymbol{X^T}(\boldsymbol{X\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{\beta} + (\boldsymbol{X^TX})^{-1}\boldsymbol{X^T\varepsilon}$$

Taking the expectation on both sides of the above expression gives

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + (\boldsymbol{X^TX})^{-1}\boldsymbol{X^T}\underbrace{\boldsymbol{E}(\boldsymbol{\varepsilon})}_{\mathbf{0}} = \boldsymbol{\beta} \qquad \Rightarrow \hat{\boldsymbol{\beta}} \text{ an unbiased estimator for } \boldsymbol{\beta}$$

Using the law of error propagation (LEP), see Appendix A, on Eq. (3.4), the covariance matrix for $\hat{\boldsymbol{\beta}}$ can be derived as:

$$\boldsymbol{cov}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T cov}(\varepsilon)((\boldsymbol{X^T X})^{-1}\boldsymbol{X^T})^T = \sigma^2(\boldsymbol{X^T X})^{-1} \quad (3.8)$$

Note that:

$$\begin{aligned} \boldsymbol{X^T}\hat{\varepsilon} = \quad & \boldsymbol{X^T}(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{y} = (\boldsymbol{X^T} - \underbrace{\boldsymbol{X^T X}(\boldsymbol{X^T X})^{-1}}_{\boldsymbol{I}}\boldsymbol{X^T})\boldsymbol{y} = \\ & (\boldsymbol{X^T} - \boldsymbol{X^T})\boldsymbol{y} = \boldsymbol{0} \\ \hat{\boldsymbol{y}}^{\boldsymbol{T}}\hat{\varepsilon} = \quad & (\boldsymbol{X}\hat{\boldsymbol{\beta}})^{\boldsymbol{T}}\hat{\varepsilon} = \hat{\boldsymbol{\beta}}^{\boldsymbol{T}}\underbrace{\boldsymbol{X^T}\hat{\varepsilon}}_{\boldsymbol{0}} = 0 \end{aligned} \quad (3.9)$$

and the matrix $[\boldsymbol{I}-\boldsymbol{H}] = [\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X^T X})^{-1}\boldsymbol{X^T}]$ satisfies the properties

- $[\boldsymbol{I}-\boldsymbol{H}]^{\boldsymbol{T}} = [\boldsymbol{I}-\boldsymbol{H}]$ $\Rightarrow$ Symmetric
- $[\boldsymbol{I}-\boldsymbol{H}][\boldsymbol{I}-\boldsymbol{H}] = [\boldsymbol{I}-\boldsymbol{H}]$ $\Rightarrow$ Idempotent

Using the LEP, see the Appendix A, again on the expression $\hat{\varepsilon} = (\boldsymbol{I}-\boldsymbol{H})\boldsymbol{y}$ gives:

$$\begin{aligned} \boldsymbol{cov}(\hat{\varepsilon}) = \quad & (\boldsymbol{I}-\boldsymbol{H})\boldsymbol{cov}(\varepsilon)(\boldsymbol{I}-\boldsymbol{H})^{\boldsymbol{T}} = (\boldsymbol{I}-\boldsymbol{H})\boldsymbol{cov}(\varepsilon) = \\ & \sigma^2[\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X^T X})^{-1}\boldsymbol{X^T}] \end{aligned} \quad (3.10)$$

because of the properties of $(\boldsymbol{I}-\boldsymbol{H})$.

Because $\hat{\boldsymbol{\beta}}$ and $\hat{\varepsilon}$ are uncorrelated, one gets: $\boldsymbol{cov}(\hat{\boldsymbol{\beta}}, \hat{\varepsilon}) = \boldsymbol{0}$

3.1.3 The weighted linear regression model

The observed responses y_i are sometimes assumed to be of variable accuracy. In such cases it is usual to give an observed response a specific weight. Finding suitable approximations for these weights might often be a problem. Variance- and covariance estimates from earlier experience may often be the best a priori values of accuracy that can be achieved.

Independent observations

If the observed responses y_i are independent, it is sufficient to use a simple diagonal weight matrix in the adjustment. It is reasonable that a precise observation, in the meaning of low a priori variance, should influence the estimated parameters more than observations with greater variance. The inverse of an a priori covariance matrix, may be used as an input weight matrix $\boldsymbol{W}$ for adjustment. A diagonal matrix is straightforward to invert.

$$\boldsymbol{W} = \boldsymbol{cov}(\boldsymbol{y})^{-1} = \underbrace{\begin{bmatrix} \frac{1}{\sigma_{11}^2} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_{22}^2} & 0 & \dots & 0 \\ 0 & 0 & \ddots & & \vdots \\ 0 & \vdots & & \ddots & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sigma_{nn}^2} \end{bmatrix}}_{n \times n} \tag{3.11}$$

Weights are relative quantities. Thus the weight matrix can also be defined as the inverse of a co-factor matrix $\boldsymbol{\Sigma_{yy}}$, which is related to covariance matrices through:

$$\boldsymbol{cov}(\boldsymbol{y}) = \sigma^2 \boldsymbol{\Sigma_{yy}} \quad \Leftrightarrow \quad \boldsymbol{cov}(\boldsymbol{y})^{-1} = \frac{1}{\sigma^2} \boldsymbol{\Sigma_{yy}^{-1}} \tag{3.12}$$

Using a trick, setting the a priori reference variance equal to one ($\sigma^2 = 1$), the co-factor and the a priori covariance matrix become identical. For a special case of all weights equal to one, we got $\boldsymbol{\Sigma_{yy}} = \boldsymbol{I}$ which leads back to the unweighted case of Eq. (3.3).

Dependency among observations

If the observed y_i is not independent, it is possible to use a complete a priori-covariance matrix. Outside the diagonal consisting of variances, the covariances can describe the grade of linear dependency between the observations. An inversion of a full symmetric quadratic matrix requires a number of multiplications as a function of n^3, [see Press et al., 1992]. For large samples n, this might cause numerical and processing-time problems. Obviously there is a relationship between

the use of a complete a priori covariance matrix and possible linear dependency between the corresponding errors. For the time being $\boldsymbol{cov}(\boldsymbol{y})$ or $\boldsymbol{\Sigma}_{\boldsymbol{yy}}$ are assumed to be diagonal. However, the estimators to be derived in Subsection 3.1.4 are valid also for complete co-factor and covariance matrices.

3.1.4 Weighted least squares estimation

Using the same reasoning as in the unweighted case, a least squares estimator for $\boldsymbol{\beta}$ can be derived, minimising the main function of weighted error sum of squares:

$$\frac{\partial m(\boldsymbol{\varepsilon}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{-\boldsymbol{1}}\boldsymbol{\varepsilon})}{\partial(\boldsymbol{\beta})} = 0 \qquad \Leftrightarrow \qquad \hat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{-\boldsymbol{1}}\boldsymbol{X})^{-\boldsymbol{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{-\boldsymbol{1}}\boldsymbol{y} \tag{3.13}$$

which also can be shown to give an unbiased estimator for $\boldsymbol{\beta}$.
The fitted values of $\boldsymbol{y}$; $\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$, and the residuals:

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{-\boldsymbol{1}}\boldsymbol{X})^{-\boldsymbol{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{-\boldsymbol{1}})\boldsymbol{y} \tag{3.14}$$

The covariance matrix of the estimated parameters:

$$\boldsymbol{cov}(\hat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}} = \sigma^2(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{-\boldsymbol{1}}\boldsymbol{X})^{-\boldsymbol{1}} \tag{3.15}$$

An unbiased estimator for the variance of unit weight:

$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{\varepsilon}}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{-\boldsymbol{1}}\hat{\boldsymbol{\varepsilon}}}{n-(r+1)} \tag{3.16}$$

The covariance matrix of the residuals

$$\boldsymbol{cov}(\hat{\boldsymbol{\varepsilon}}) = \sigma^2[\boldsymbol{\Sigma}_{\boldsymbol{yy}} - \boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{-\boldsymbol{1}}\boldsymbol{X})^{-\boldsymbol{1}}\boldsymbol{X}^{\boldsymbol{T}}] \tag{3.17}$$

3.1.5 Non-linearity, linearising

This thesis only discusses linear problems. Sometimes the relations between observed responses and unknown parameters are non-linear. However, linear models may even be used to solve non-linear problems, making use of linearising together with an iteration procedure. Assuming that preliminary values of proper

accuracy exist for unknown parameters, it is possible to expand the original non-linear functional model in a Taylor series. Here it is common to neglect terms of the second and higher powers. The final solution is found by several iterations, and subsequent improvements of preliminary values, until the corrections are too small to make any sense. In mathematics, a similar procedure is known as the Newton-Raphson method [see Kreyszig, 1999, p.841 among others]. Introducing stochastic variables, stochastic errors will also affect the corrections. The extended method is then called Gauss-Newton's method [see Koch, 1999]. The Gauss-Newton iterative method has often been shown to give a fast converging solution. If one uses "good" initial values, only one computation is enough. However, it is usual to make several iterations where initial values are updated, until their improvements converge to a chosen limit that is close enough to zero. The linearising procedure reduces an original non-linear model problem to a linear regression problem. In geodesy the combination, non-linear model, reduced to a linear one, and then used an iterative way, is usually known as the least squares method.

3.2 Statistical tests

In a weighted regression model, the use of a diagonal $\boldsymbol{\Sigma_{yy}}$ is equivalent to expecting that the errors in the $\boldsymbol{\varepsilon}$-vector are independent. Some assumptions have to be made to derive test statistics. Each element in the $\boldsymbol{\varepsilon}$-vector is assumed to be normally distributed, with zero expectation, but different variances are allowed. Each error may then be looked upon as a single realisation from n different normal distributions.

3.2.1 Outlier detection

A normalisation has to be made in order to make the residuals remaining after a weighted regression comparable to each other. According to Johnson and Wichern [1998], most statisticians prefer diagnostics based on studentised residuals, computed by dividing each residual with the square root of the corresponding element in $\widehat{\boldsymbol{cov}}(\hat{\boldsymbol{\varepsilon}})$. Using $\hat{\sigma}^2$ as estimator for σ^2, an estimate for $\boldsymbol{cov}(\hat{\boldsymbol{\varepsilon}})$ can be

computed as:

$$\widehat{\boldsymbol{cov}}(\hat{\boldsymbol{\varepsilon}}) = \hat{\sigma}^2[\boldsymbol{\Sigma_{yy}} - \boldsymbol{X}(\boldsymbol{X^T\Sigma_{yy}^{-1}X})^{-1}\boldsymbol{X^T}] \tag{3.18}$$

Out-of-context test for outliers

Vanicek and Krakiwsky [1986] use the term standardised residuals. Their definition is:

$$\text{Standardised residuals} \triangleq \hat{\varepsilon}_i^* = \frac{\hat{\varepsilon}_i}{\sqrt{(\widehat{\boldsymbol{cov}}(\hat{\boldsymbol{\varepsilon}}))_{ii}}} \tag{3.19}$$

According to Vanicek and Krakiwsky [1986] the standardised or normalised residuals computed using estimates for both unknown mean and variance, these are τ-distributed ($\tau = tau$) with $n-(r+1)$ degrees of freedom. The τ-distribution has been investigated by Pope [1976]. An approximation formula using the relation between τ-distribution and the student T distribution is presented in Leick [2004] . However, for large samples the difference is negligible. For a large number of observations n, the standardised residuals will be approximately student-T distributed with $n-(r+1)$ degrees of freedom. In its turn, this is then approximately normally distributed with zero mean and variance one, so:

$$\text{Standardised residuals} \approx N(0,1) \qquad \text{if } n \text{ is large} \tag{3.20}$$

Normality assumption is common for most outlier tests. The magnitudes of the standardised residuals are often used to identify a possible outlier in corresponding observations. Here it is common to exclude observations where the absolute value of corresponding standardised residuals exceeds a limit or a k-factor. However, the k-factor is related to a significance level α. A problem might be that a possible outlier will influence the size of $\hat{\sigma}^2$ and thus affects the values of all computed standardised residuals, especially in cases of few observations. As it is best to avoid this, it is customary to standardise $\hat{\varepsilon}_i$ using the delete-one estimated variance , which is the residual mean square when the $i-$th observation is dropped from the estimation. The standardising trick ignores the covariances between the residuals, so $\hat{\varepsilon}_i$ has actually been taken out of context of the other residuals. This means that, the existence of the other members of the series has been disregarded. Each observation is then tested independently of the others. Vanicek and Krakiwsky [1986] use the expression out-of-context. The test is derived, assuming only

one outlier is hidden in the observations. If an outlier is found, the test has to be redone. If the test recommends a total significance level α_{tot}, each standardised residual $\hat{\varepsilon}_i^*$, has to be tested with a smaller, individual significance level α_{ind}. Assuming that the probability to detect an individual outlier is $p_{ind} = (1 - \alpha_{ind})$ and the errors are independent, the total probability to detect one outlier out of n observations is $p_{tot} = (1 - \alpha_{tot})$. The multiple of the n single probabilities will be

$$(1 - \alpha_{tot}) = (1 - \alpha_{ind})^n \quad \Leftrightarrow \quad \alpha_{ind} = 1 - (1 - \alpha_{tot})^{\frac{1}{n}} \approx \alpha_{tot}/n \qquad (3.21)$$

because $(1 - \alpha)^{\frac{1}{n}} \approx 1 - (\alpha/n)$ is a good approximation when n is large.

Outlier test based on conditional probabilities (*in-the-context*)

The individual observations y_i tested through an investigation of $\hat{\varepsilon}_i$ can also be examined in-the-context of being member of a series. According to Vanicek and Krakiwsky [1986], this is not a straightforward method, but can be made by setting a modified $\alpha_{sim} \neq \alpha_{ind}$ in Eq. (3.21) which accounts for the simultaneity of all the tested elements in the vector $\boldsymbol{y}$. A test using conditional probabilities to treat all the observations simultaneously, or in-the-context with each other, will request a value $\alpha_{sim} > \alpha_{ind}$ in Eq.(3.21).

A choice of significance level, and thus a proper value for the k-factor is difficult. Using this approach for time series, the correlations between errors are requested to be small for a final chosen model. An indication of correlations may be found looking at the standardised residuals as a time series and estimate their ACF. However, it is clear that the estimated auto correlations is a very uncertain measure of the true dependency. A wrong chosen deterministic model may introduce large estimates for the ACF. Each statistical test should therefore be handled with care.

Each time a possible outlier has been detected and excluded, a new adjustment and a new outlier-test has to be done for all remaining observations. When several outliers are detected, an optimal procedure should, in an iterative way, replace the first out-sorted outliers one by one and test if they could be included again. In most software a variable k-factor is used to evaluate if the total number of out-

sorted observations is reasonable for the whole sample. In this way the k-factor can be used as an adjusting screw, to filter the number of input observations.

3.2.2 Goodness-of-fit test

The residuals lack of fit for the final chosen model may be examined using a Goodness-of-fit test, [see Emery and Thomson, 2001], based on the test statistics:

$$X^2 = \frac{\sum_{i=1}^{k}(f_i - F_i)^2}{F_i} \tag{3.22}$$

X^2 is then χ^2_ν.-distributed (chi-square) with ν degrees of freedom. The test assumes the observations from a variable X, to be independent. The idea is to group the total of n observations into k class intervals, whose plots form a frequency histogram or observed probability density function $p_o(X)$. f_i is the observed frequency in the ith class interval. The test compares f_i with the expected frequency F_i, in the ith class for the theoretically probability density function $p(X)$, that the observations are assumed to be sampled from. Generally the null hypothesis will be:

$$\boldsymbol{H_0} : p_o(X) = p(X) \quad \text{versus} \quad \boldsymbol{H_A} : p_o(X) \neq p(X) \tag{3.23}$$

where $\boldsymbol{H_A}$ is the alternative hypothesis. A normality test considering if the standardised residuals are $N(0,1)$-distributed can then be made forming the null hypothesis:

$$\boldsymbol{H_0} : p_o = \frac{e^{\frac{x^2}{2}}}{\sqrt{2\pi}} \quad \text{versus} \quad \boldsymbol{H_A} : p_o(X) \neq \frac{e^{\frac{x^2}{2}}}{\sqrt{2\pi}} \tag{3.24}$$

The number of degrees of freedom to be used for the χ^2-distribution is $\nu = k - 3$. One degree is lost through the restriction that if $(k-1)$ class intervals are determined, the k-th class interval follows automatically. Two degrees of freedom are lost when mean and variance has to be estimated. A good fit means generally f_i close to F_i for all k groups. X^2 then becomes small and vice versa, so the upper part of the χ^2-distribution has to be used. The null hypothesis will be rejected at a

chosen significance level α, if the observed value

$$X^2 > \chi^2_\nu(\alpha) \tag{3.25}$$

3.2.3 Likelihood ratio test for regression parameters

A test of simultaneous significance for a group of parameters in $\boldsymbol{\beta}$ will be introduced, for the case of different weighted, but independent observations. One of many applications is to test the simultaneous significance of several frequency components (constructed from pairs of spectral parameters) in the weighted least squares procedure introduced in Section 4.2. The deterministic part of the regression model will be divided in two parts. The first part will be denoted with subscript 1, as $\boldsymbol{X_1\beta_1}$. The second part with subscript 2, as $\boldsymbol{X_2\beta_2}$:

$$\boldsymbol{y - \epsilon = X\beta} = \underset{n\cdot(q+1)}{\left[\begin{array}{c|c} \boldsymbol{X_1} & \boldsymbol{X_2} \\ & \end{array}\right]} \underset{(r-q)\cdot 1}{\left[\begin{array}{c} \boldsymbol{\beta_1} \\ \hline \boldsymbol{\beta_2} \end{array}\right]} \underset{(q+1)\cdot 1}{} = \boldsymbol{X_1\beta_1 + X_2\beta_2} \tag{3.26}$$

A test is required which reveals the simultaneous influence from the predictors $x_{(q+1)}, x_{(q+2)}, ..., x_r$ on the responses in $\boldsymbol{y}$. This is analogous to a simultaneous test if the parameters in $\boldsymbol{\beta_2}$ are significantly different from the zero vector. The null hypothesis will be:

$$\boldsymbol{H_0} : \boldsymbol{\beta_2} = \begin{bmatrix} \beta_{q+1} \\ \beta_{q+2} \\ \vdots \\ \beta_r \end{bmatrix} = \boldsymbol{0} \quad \text{versus} \quad \boldsymbol{H_A : \beta_2 \neq 0} \tag{3.27}$$

where $\boldsymbol{H_A}$ is any alternative hypothesis. Rejecting $\boldsymbol{H_0}$ at the chosen significance level (α) indicates that the parameters in $\boldsymbol{\beta_2}$ are significant and cannot be omitted.

Likelihood function

Assume the design matrix $\boldsymbol{X}$ has full rank, and $\boldsymbol{\Sigma_{yy}^{-1}}$ is diagonal. Given the data and the normal assumptions, the likelihood associated with the parameters $\boldsymbol{\beta}$ and

σ^2 is

$$L(\boldsymbol{\beta}, \sigma^2) \triangleq \frac{1}{(2\pi\sigma^2)^{n/2}|\boldsymbol{\Sigma_{yy}^{-1}}|^{1/2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})^{\boldsymbol{T}}\boldsymbol{\Sigma_{yy}^{-1}}(\boldsymbol{y}-\boldsymbol{X\beta})} \tag{3.28}$$

σ^2 is often set equal to one and omitted, but in the context of deriving a likelihood ratio test, it will later be replaced with other expressions. The maximum functional value is achieved when $\boldsymbol{\beta}$ and σ^2 are estimated using the Maximum Likelihood Estimators (MLE). The MLE for σ^2 in the case of weighted observations analogous to the one in Eq. (3.6) is:

$$\hat{\sigma}^2_{MLE} = \frac{(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T \boldsymbol{\Sigma_{yy}^{-1}} (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})}{n} \tag{3.29}$$

It can be shown that the MLE for $\boldsymbol{\beta}$ in the weighted case is equal to the weighted Least Squares Estimator (LSE) [see Koch, 1999].

$$\hat{\hat{\boldsymbol{\beta}}}_{MLE} = (\boldsymbol{X^T\Sigma_{yy}^{-1}X})^{-1}\boldsymbol{X^T\Sigma_{yy}^{-1}y} \tag{3.30}$$

All other estimators than the MLE will give a smaller functional value in Eq. (3.28). Under $\boldsymbol{H_0}$ (If $\boldsymbol{H_0}$ is correct), one gets $\boldsymbol{y} = \boldsymbol{X_1\beta_1} + \boldsymbol{\varepsilon}$ and:

$$\hat{\hat{\boldsymbol{\beta}}}_1 = (\boldsymbol{X_1^T\Sigma_{yy}^{-1}X_1})^{-1}\boldsymbol{X_1^T\Sigma_{yy}^{-1}y} \tag{3.31}$$

$$\hat{\sigma}^2_1 = \frac{(\boldsymbol{y} - \boldsymbol{X_1}\hat{\boldsymbol{\beta}}_1)^{\boldsymbol{T}}\boldsymbol{\Sigma_{yy}^{-1}}(\boldsymbol{y} - \boldsymbol{X_1}\hat{\boldsymbol{\beta}}_1)}{n} \tag{3.32}$$

Test statistics

A test statistic W will be presented in order to test the hypothesis in Eq. (3.27). W should be understood as the ratio of the maximum of estimated values of the likelihood function with and without the constraint of the null hypothesis.

$$\text{Likelihood ratio} = W = \frac{\max\limits_{\boldsymbol{\beta_1}, \sigma_1^2} L(\boldsymbol{\beta}, \sigma^2)}{\max\limits_{\boldsymbol{\beta}, \sigma^2} L(\boldsymbol{\beta}, \sigma^2)} \tag{3.33}$$

The denominator is always equal or greater than the numerator. For large differences, W is much smaller than one and $\boldsymbol{H_0 : \beta_2 = 0}$ will be rejected. Inserting for the MLE in Eq. (3.28) one gets:

$$\underset{\boldsymbol{\beta_1},\sigma_1^2}{max} L(\boldsymbol{\beta},\sigma^2) = \frac{1}{(2\pi\hat{\hat{\sigma}}_1^2)^{n/2}|\boldsymbol{\Sigma_{yy}}|^{1/2}} e^{\frac{-(\boldsymbol{y}-\boldsymbol{X_1}\hat{\boldsymbol{\beta}}_1)^T\boldsymbol{\Sigma}_{yy}^{-1}(\boldsymbol{y}-\boldsymbol{X_1}\hat{\boldsymbol{\beta}}_1)}{2(\boldsymbol{y}-\boldsymbol{X_1}\hat{\boldsymbol{\beta}}_1)^T\boldsymbol{\Sigma}_{yy}^{-1}(\boldsymbol{y}-\boldsymbol{X_1}\hat{\boldsymbol{\beta}}_1)/n}} \tag{3.34}$$

and:

$$\underset{\boldsymbol{\beta},\sigma^2}{max} L(\boldsymbol{\beta},\sigma^2) = \frac{1}{(2\pi\hat{\hat{\sigma}}^2)^{n/2}|\boldsymbol{\Sigma_{yy}}|^{1/2}} e^{\frac{-(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})^T\boldsymbol{\Sigma}_{yy}^{-1}(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})}{2(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})^T\boldsymbol{\Sigma}_{yy}^{-1}(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})/n}} \tag{3.35}$$

and the likelihood ratio reduces to:

$$W = \left(\frac{\hat{\hat{\sigma}}_1^2}{\hat{\hat{\sigma}}^2}\right)^{-n/2} = \left(\frac{\hat{\hat{\sigma}}^2 + \hat{\hat{\sigma}}_1^2 - \hat{\hat{\sigma}}^2}{\hat{\hat{\sigma}}^2}\right)^{-n/2} = \left(1 + \frac{\hat{\hat{\sigma}}_1^2 - \hat{\hat{\sigma}}^2}{\hat{\hat{\sigma}}^2}\right)^{-n/2} \tag{3.36}$$

Rejecting $\boldsymbol{H_0}$ for small values of W is equivalent to rejecting $\boldsymbol{H_0}$ for large values of $\frac{(\hat{\hat{\sigma}}_1^2 - \hat{\hat{\sigma}}^2)}{\hat{\hat{\sigma}}^2}$, or with the relations to the unbiased LSE's we find the scaled version, [see Johnson and Wichern, 1998]:

$$\frac{n(\hat{\hat{\sigma}}_1^2 - \hat{\hat{\sigma}}^2)/(r-q)}{n\hat{\hat{\sigma}}^2/(n-r-1)} = \frac{(RWSS_1 - RWSS)/(r-q)}{RWSS/(n-r-1)} = F \tag{3.37}$$

where $RWSS$ and $RWSS_1$ denotes the Residual Weighted Sum of Squares (RWSS) from adjustments, respectively, with parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\beta_1}$. This F-ratio of two χ^2-distributed quantities has a F-distribution. A test of simultaneous significance of several parameters in a vector $\boldsymbol{\beta_2}$ Eq. (3.27), will be rejected if the ratio:

$$\frac{(RWSS_1 - RWSS)}{RWSS} > \frac{(r-q)}{(n-r-1)} F_{(r-q,n-r-1)}(\alpha) \tag{3.38}$$

where $F_{(r-q,n-r-1)}(\alpha)$ is the upper 100α percentile of the F-distribution with $(r-q)$ and $(n-r-1)$ degrees of freedom.

3.3 Heaviside function and detection of data offsets

Recorded time series, often include jumps or offsets. For various reasons, they may be known or unknown in time.

Known jumps

If the time t_{β_j}, for a jump j is known, the modelling of jump size β_j can be done using a unit step function, also called the Heaviside function. Using time as argument, the definition of the Heaviside function is [See Kreyszig, 1999, among others]

$$h(t - t_{\beta_j}) = \begin{cases} 0 & \text{if} \quad t < t_{\beta_j} \\ 1 & \text{else} \end{cases} \tag{3.39}$$

An observation equation for the model in Eq. (3.1) including parameters for constant term, linear trend (slope) and several jumps, let us say the total number of jumps are: n_{jump}, will be of the form:

$$y(t) = \beta_0 + \beta_1(t) + \sum_{j=2}^{n_{jump}+1} \beta_j h(t - t_{\beta_j}) + \varepsilon(t) \tag{3.40}$$

or for a general observation i:

$$y_i = \beta_0 + \beta_1 t_i + \sum_{j=2}^{n_{jump}+1} \beta_j h(t_i - t_{\beta_j}) + \varepsilon_i \tag{3.41}$$

Using matrix notations for a set of linear equations, for the estimation of one jump at the time ($t_{\beta_j} = t_i$) may look like:

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix}}_{n\times 1} = \underbrace{\begin{bmatrix} 1 & t_1 & 0 \\ 1 & t_2 & 0 \\ \vdots & \vdots & \vdots \\ 1 & t_{i-1} & 0 \\ 1 & t_i & 1 \\ \vdots & \vdots & \vdots \\ 1 & t_n & 1 \end{bmatrix}}_{n\times 3} \cdot \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}_{3\times 1} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{bmatrix}}_{n\times 1} \tag{3.42}$$

or

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{3.43}$$

The model can easily be expanded for several jumps, by adding more columns to the $\boldsymbol{X}$ matrix and jump-parameters to the $\boldsymbol{\beta}$-vector.

Unknown jumps

However it would be interesting to clarify even if a time series contains unknown jumps. This is of special interest because such jumps in a time series will violate the assumption of constant expectation for the residuals which is essential for all further analyses.

Assume there are one or several unknown jumps hidden in the data. For which time do these jumps occur? One possible way to investigate the problem will be to estimate a jump for all possible points in time. This can be seen as a "sliding" of the Heaviside function through the last column of the $\boldsymbol{X}$-matrix in Eq. (3.42). Making a new adjustment for each step it is possible to estimate both the jump size and the a posteriori reference variance. The selection of a possible jump may be done using either the criterion:

- L1-norm: A jump exists corresponding to the time which gives the maximum estimated absolute value for the parameter itself, or
- L2-norm: A jump exists at the time corresponding to minimum estimated reference variance

Mainly the second (least squares) criterion will be used to detect possible unknown jumps in the time series of CGPS data in this thesis. As an alternative to plot the estimated reference variance $\hat{\sigma}_0^2$, as a function of the jump time $\hat{\sigma}_0^2(t)$ to find a minimum value, a new variable $VR(t)$ is constructed. $VR(t)$ shows the reference Variance Reduction (VR) as a function of time and will give a maximum value at the time for which we expect a jump to appear. The reduced reference variance is compared to the reference variance estimated without any jumps, denoted $\hat{\sigma}_0^2(0)$:

$$VR(t_i) = \frac{\hat{\sigma}_0^2(0) - \hat{\sigma}_0^2(t_i)}{\hat{\sigma}_0^2(0)} = 1 - \frac{\hat{\sigma}_0^2(t_i)}{\hat{\sigma}_0^2(0)} \tag{3.44}$$

The biased estimator that does not correct for degrees of freedom in Eq. (3.6) will be preferred because it ensures $\hat{\sigma}_0^2(0) \geq \hat{\sigma}_0^2(t_i)$ and $(0 \leq VR(t_i) \leq 1)$.

In the analysis of CGPS data, the method has shown some weaknesses, as will be discussed in Subsection 6.1.1. For a network represented by several co-varying series, cross correlation functions can be an alternative tool that can be used for the detection of data offsets.

3.4 Multivariate linear regression

An optimal adjustment of CGPS data should be made using observations from all three components in one simultaneous estimation. This is possible by an expansion of the observation vector $\boldsymbol{y}$ to a $\boldsymbol{Y}$-matrix with the number of columns equal to the number of components to be estimated simultaneously. A multivariate regression model is described in Johnson and Wichern [1998]. Very long time series may cause serious computing problems. If an observed time series from different components behaves independently of each other, it becomes a great advantage to treat them as three univariate time series. A significance test of correlations between the three components will be introduced, to test whether it is possible to avoid the time consuming simultaneously adjustment,

3.4.1 Likelihood ratio test of the correlation matrix

The purpose of this subsection is to derive a test of cross correlations between data from an expected p-dimensional multi-normal distribution. If correlations between dimensions can be ignored, it may be possible to reduce an original p-dimensional problem to p univariate ones.

Assume data are available in a data matrix consisting of:
p = number of variables (not necessary independent) in columns.
n = number of independent samples (observations) in rows.

Data matrix:

$$\underset{(n\times p)}{\boldsymbol{x}} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ x_{j1} & x_{j2} & \dots & x_{jp} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x_1}^T \\ \boldsymbol{x_2}^T \\ \vdots \\ \vdots \\ \boldsymbol{x_j}^T \\ \vdots \\ \boldsymbol{x_n}^T \end{bmatrix} \tag{3.45}$$

The row vectors: $\boldsymbol{x_j}^T = [x_{j1}, x_{j2} \dots\dots x_{jp}], \quad (j = 1, 2, ..., n)$ represent independent observations and are said to form a random sample from a common joint distribution with density function $f(\boldsymbol{x}) = f(x_1, x_2, \dots, x_p)$. $\boldsymbol{x}$ is then said to be (p-dimensional) multi-normal distributed, with expectation vector $\boldsymbol{\mu}$ and a priori covariance matrix $\boldsymbol{\Sigma}$:

$$\boldsymbol{x} \sim N_p(\underset{(p\times 1)}{\boldsymbol{\mu}}, \underset{(p\times p)}{\boldsymbol{\Sigma}}) \tag{3.46}$$

Likelihood ratio test

As in the test in Subsection 3.2.3, a test statistic will be presented based on the ratio between maximum values of two maximum likelihood functions, with and without constraints. In Subsection 3.2.3, the significance of parameters was considered and in a similar way it will be possible to derive a test statistic for co-

variance matrices. In accordance with Koch [1999], the test may be applied if estimated covariance matrices are to be replaced by matrices of simpler structure, possibly by diagonal matrices. The hypothesis will be:

$$\boldsymbol{H_0 : \Sigma = \Sigma_{Diag}} \quad \text{versus} \quad \boldsymbol{H_A : \Sigma \neq \Sigma_{Diag}} \tag{3.47}$$

An alternative to a test of covariance matrices would be a test of the correlation matrix $\boldsymbol{\rho}$ versus the identity matrix $\boldsymbol{I}$. In a special case, when observations are all $N(0,1)$-distributed, the correlation and covariance matrix are identical. Then the hypothesis will be:

$$\boldsymbol{H_0 : \rho = I} \quad versus \quad \boldsymbol{H_A : \rho \neq I} \tag{3.48}$$

A likelihood function (joint density function) valid for multi-normal distributed observations is:

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \triangleq \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} e^{-\frac{1}{2}\sum_{j=1}^{n}(\boldsymbol{x_j}-\boldsymbol{\mu})^T \boldsymbol{\Sigma^{-1}} (\boldsymbol{x_j}-\boldsymbol{\mu})} \tag{3.49}$$

Maximum functional value is achieved when $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are estimated using the Maximum Likelihood Estimators (MLE):

$$\tilde{\boldsymbol{\mu}} = \bar{\boldsymbol{x}} = \begin{bmatrix} \bar{\boldsymbol{x}}_{\boldsymbol{1}} \\ \bar{\boldsymbol{x}}_{\boldsymbol{2}} \\ \vdots \\ \vdots \\ \bar{\boldsymbol{x}}_{\boldsymbol{p}} \end{bmatrix} \qquad \text{and} \qquad \tilde{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{j=1}^{n} (\boldsymbol{x_j} - \bar{\boldsymbol{x}})(\boldsymbol{x_j} - \bar{\boldsymbol{x}})^{\boldsymbol{T}} \tag{3.50}$$

All other estimators will give a smaller function value in Eq. (3.49). A test statistic Λ will be derived, to test the hypothesis in Eq. (3.47). Λ should be understood as the ratio of the maximum of estimated values of the likelihood function with and without the constraint of the null hypothesis.

$$Likelihood \quad ratio = \Lambda = \frac{\underset{\boldsymbol{\mu},\boldsymbol{\Sigma}\in\boldsymbol{\Sigma}_{Diag}}{maximum} \quad L(\boldsymbol{\mu}, \boldsymbol{\Sigma_{Diag}})}{\underset{\boldsymbol{\mu},\boldsymbol{\Sigma}}{maximum} \quad L(\boldsymbol{\mu}, \boldsymbol{\Sigma})} < C \tag{3.51}$$

Maximum for the likelihood function in the numerator is obtained by allowing $\boldsymbol{\mu}$ to vary over all values, but keeping $\boldsymbol{\Sigma}$ restricted to the subset $\boldsymbol{\Sigma}_{Diag}$. Without restrictions, the maximum for the denominator is achieved when the MLEs $\bar{\boldsymbol{x}}$ and $\tilde{\boldsymbol{\Sigma}}$ are inserted. The nominator will always be smaller than the denominator, except when $\boldsymbol{\Sigma}_{Diag} = \tilde{\boldsymbol{\Sigma}}$. Hence $0 \leq \Lambda \leq 1$ and the null hypothesis in Eq. (3.47) has to be rejected for values of Λ very different from one. If the sampling distribution of the test statistic Λ is known, C can be selected to choose a test with a specified significance level α. Johnson and Wichern [1998] derived the exponent in Eq. (3.49) as:

$$\begin{aligned} &-\frac{1}{2}\sum_{j=1}^{n}(\boldsymbol{x_j}-\boldsymbol{\mu})^{\boldsymbol{T}}\boldsymbol{\Sigma^{-1}}(\boldsymbol{x_j}-\boldsymbol{\mu}) = \\ &-\frac{1}{2}tr[\boldsymbol{\Sigma^{-1}}(\sum_{j=1}^{n}(\boldsymbol{x_j}-\bar{\boldsymbol{x}})(\boldsymbol{x_j}-\bar{\boldsymbol{x}})^{\boldsymbol{T}}] \\ &+\sum_{j=1}^{n}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})^{\boldsymbol{T}}\boldsymbol{\Sigma^{-1}}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}) \end{aligned} \tag{3.52}$$

Replacing $\boldsymbol{\mu}$ with the estimator $\tilde{\boldsymbol{\mu}} = \bar{\boldsymbol{x}}$, the last term will disappear. The expression $\sum_{j=1}^{n}(\boldsymbol{x_j}-\bar{\boldsymbol{x}})(\boldsymbol{x_j}-\bar{\boldsymbol{x}})^{\boldsymbol{T}}$ can be written as $n\tilde{\boldsymbol{\Sigma}}$, see Eq.(3.50), so one gets:

$$\begin{aligned} \Lambda &= \frac{e^{(-\frac{1}{2}tr[\boldsymbol{\Sigma_{Diag}^{-1}}(\sum_{j=1}^{n}(\boldsymbol{x_j}-\bar{\boldsymbol{x}})(\boldsymbol{x_j}-\bar{\boldsymbol{x}})^{\boldsymbol{T}}])}}{(2\pi)^{np/2}\left|\boldsymbol{\Sigma}_{Diag}\right|^{n/2}} \cdot \frac{(2\pi)^{np/2}\left|\tilde{\boldsymbol{\Sigma}}\right|^{n/2}}{e^{(-\frac{1}{2}tr[\boldsymbol{\tilde{\Sigma}^{-1}}(\sum_{j=1}^{n}(\boldsymbol{x_j}-\bar{\boldsymbol{x}})(\boldsymbol{x_j}-\bar{\boldsymbol{x}})^{\boldsymbol{T}}])}} \\ &= \left(\frac{|\tilde{\boldsymbol{\Sigma}}|}{|\boldsymbol{\Sigma}_{Diag}|}\right)^{(n/2)} \cdot \frac{e^{-\frac{1}{2}tr[\boldsymbol{\Sigma}_{Diag}^{-1}n\tilde{\boldsymbol{\Sigma}}]}}{e^{-\frac{1}{2}tr[\tilde{\boldsymbol{\Sigma}}^{-1}n\tilde{\boldsymbol{\Sigma}}]}} = \left(\frac{|\tilde{\boldsymbol{\Sigma}}|}{|\boldsymbol{\Sigma}_{Diag}|}\right)^{(n/2)} \cdot \frac{e^{\frac{n}{2}tr[\boldsymbol{I}_{(p\cdot p)}]}}{e^{\frac{1}{2}tr[\boldsymbol{\Sigma}_{Diag}^{-1}n\tilde{\boldsymbol{\Sigma}}]}} \\ &= \left(\frac{|\tilde{\boldsymbol{\Sigma}}|}{|\boldsymbol{\Sigma}_{Diag}|}\right)^{(n/2)} \cdot \frac{e^{\frac{np}{2}}}{e^{\frac{n}{2}tr[\boldsymbol{\Sigma}_{Diag}^{-1}\tilde{\boldsymbol{\Sigma}}]}} = \left(\frac{|\tilde{\boldsymbol{\Sigma}}|}{|\boldsymbol{\Sigma}_{Diag}|}\right)^{(n/2)} \cdot e^{(p-tr[\boldsymbol{\Sigma}_{Diag}^{-1}\tilde{\boldsymbol{\Sigma}}])} \end{aligned} \tag{3.53}$$

An unbiased estimator may be used instead of the biased MLE for the covariance matrix $\tilde{\boldsymbol{\Sigma}}$. However, if n is large, this will not make a difference thus this correction is omitted here. While testing a special variant of covariance matrices,

the estimated correlation matrix $\boldsymbol{R}$ with the hypothesis in Eq. (3.48), the last expression for Λ in Eq.(3.53) will be simplified by: $\boldsymbol{\Sigma}_{Diag}^{-1} = \boldsymbol{I}$ and $|\boldsymbol{\Sigma}_{Diag}| = 1$. $\tilde{\boldsymbol{\Sigma}} = \boldsymbol{R}$ which has trace p, so the final test statistic is reduced to:

$$\Lambda = |\boldsymbol{R}|^{(n/2)} < C \tag{3.54}$$

Note: The determinant of $\boldsymbol{R}$ is one, if all off diagonal elements in $\boldsymbol{R}$ are zero. In any other case, the determinant of $\boldsymbol{R}$ is smaller than one. For a large sample size $-2\ln\Lambda$ is approximately $\chi^2_{p(p-1)/2}$-distributed [see Johnson and Wichern, 1998, p.505]. Then $\boldsymbol{H_0}$ in Eq. (3.48) can be expressed:

$$\text{reject} \quad \boldsymbol{H_0} \quad \text{if:} \qquad -2ln\Lambda > \chi^2_{p(p-1)/2}(\alpha) \tag{3.55}$$

The significance of correlations between directions for each of the Norwegian CGPS-stations is tested in Subsection 6.1.5 using the CGPS time series of normalised residuals as input in the data matrix in Eq. (3.45).

Chapter 4

Spectral analysis of non-stationary data

The most common method for spectral analysis, i.e. original Fourier analysis, is not recommended for the non-stationary CGPS time series including several gaps. In this chapter, methods for spectral analysis of non-stationary non-equidistant data will be introduced. Some methods using least squares curve fit in the time domain will be discussed. They are all different variants of weighted regression analysis and differ basically in the model function, and consequently, the design of normal equations. The theory of Least Squares Spectral Analysis (LSSA) is described by Wells et al. [1985],Vanicek and Krakiwsky [1986] and Craymer [1998]. A similar method is used by Plag [1988]. He denotes the resulting spectrum as a Variance Spectrum (VS). An alternatively relation to FT exists between the Least Squares Spectrum (LSS) and the ACvF. Craymer [1998] denotes this relation the Least Squares Transform (LST)

In the presentation of Fourier analysis in Chapter 2, integer frequencies were chosen to estimate the Fourier coefficients a_k and b_k. Values for the coefficients will not be bounded to integers of k any longer, but they will still belong to an interval limited by the Nyquist frequency and the length of the time series as described in Table 2.1 . To separate the notations, the parameters will from now on be written as $a(\omega_k)$ and $b(\omega_k)$. Because of their relations to later introduced spectral values, they will be referred to as spectral parameters. The main task of spectral analysis

is to find the frequency content of a time series. A model for non-stationary data has to contain both trend- and spectral parameters. The model will be separated in two parts, a deterministic part for trend parameters and spectral part for the spectral parameters. In the context of spectral analysis, the solution of trend parameters is of minor interest, but the correlations between estimates for spectral components and other parameters have to be considered. Some simulations will be performed to investigate the LSSA-method.

4.1 Weighted Least Squares (WLS) procedure

4.1.1 Weighted Sum of Squares (WSS) decomposition

The WLS estimator from Eq. (3.13) can be written:

$$\hat{\boldsymbol{\beta}} = \underbrace{(\boldsymbol{X}^T\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{X})^{-1}}_{N^{-1}=\boldsymbol{\Sigma}_{\hat{\beta}\hat{\beta}}} \underbrace{\boldsymbol{X}^T\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{y}}_{\boldsymbol{u}} = \boldsymbol{N}^{-1}\boldsymbol{u} \tag{4.1}$$

The weighted sum of squared observations is a scalar and can be decomposed as:

$$\begin{aligned} \boldsymbol{y}^T\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{y} = \; & (\hat{\boldsymbol{y}}+\boldsymbol{y}-\hat{\boldsymbol{y}})^T\boldsymbol{\Sigma}_{yy}^{-1}(\hat{\boldsymbol{y}}+\boldsymbol{y}-\hat{\boldsymbol{y}}) = \\ & (\hat{\boldsymbol{y}}+\hat{\boldsymbol{\varepsilon}})^T\boldsymbol{\Sigma}_{yy}^{-1}(\hat{\boldsymbol{y}}+\hat{\boldsymbol{\varepsilon}}) = \\ & \hat{\boldsymbol{y}}^T\boldsymbol{\Sigma}_{yy}^{-1}\hat{\boldsymbol{y}} + \hat{\boldsymbol{\varepsilon}}^T\boldsymbol{\Sigma}_{yy}^{-1}\hat{\boldsymbol{y}} + \hat{\boldsymbol{y}}^T\boldsymbol{\Sigma}_{yy}^{-1}\hat{\boldsymbol{\varepsilon}} + \hat{\boldsymbol{\varepsilon}}^T\boldsymbol{\Sigma}_{yy}^{-1}\hat{\boldsymbol{\varepsilon}} = \\ & \hat{\boldsymbol{y}}^T\boldsymbol{\Sigma}_{yy}^{-1}\hat{\boldsymbol{y}} + \hat{\boldsymbol{\varepsilon}}^T\boldsymbol{\Sigma}_{yy}^{-1}\hat{\boldsymbol{\varepsilon}} \end{aligned} \tag{4.2}$$

The WLS estimator can be interpreted as an orthogonal projection. [see Teunissen, 2000], Analogous to the unweighted case of Eq. (3.9), the term $\hat{\boldsymbol{\varepsilon}}^T\boldsymbol{\Sigma}_{yy}^{-1}\hat{\boldsymbol{y}} + \hat{\boldsymbol{y}}^T\boldsymbol{\Sigma}_{yy}^{-1}\hat{\boldsymbol{\varepsilon}}$ vanishes.

Defining the different parts of WSS as:

$$\begin{aligned} &\text{Observation Weighted Sum of Squares:} && \text{OWSS} \triangleq \boldsymbol{y}^T\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{y} \\ &\text{Fitted Model Weighted Sum of Squares:} && \text{MWSS} \triangleq \hat{\boldsymbol{y}}^T\boldsymbol{\Sigma}_{yy}^{-1}\hat{\boldsymbol{y}} \\ &\text{Residual Weighted Sum of Squares:} && \text{RWSS} \triangleq \hat{\boldsymbol{\varepsilon}}^T\boldsymbol{\Sigma}_{yy}^{-1}\hat{\boldsymbol{\varepsilon}} \end{aligned} \tag{4.3}$$

and using the shorter notations $\boldsymbol{N} = \boldsymbol{X}^T\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{X}$ and $\boldsymbol{u} = \boldsymbol{X}^T\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{y}$, it is possible to derive:

$$\hat{\boldsymbol{y}}^T\boldsymbol{\Sigma}_{yy}^{-1}\hat{\boldsymbol{y}} = (\boldsymbol{X}\hat{\boldsymbol{\beta}})^T\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{X}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^T(\boldsymbol{X}^T\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{X})\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^T\boldsymbol{N}\hat{\boldsymbol{\beta}} \tag{4.4}$$

$$\hat{\boldsymbol{\beta}}^{\boldsymbol{T}}\boldsymbol{N}\hat{\boldsymbol{\beta}} = (\boldsymbol{N}^{-\boldsymbol{1}}\boldsymbol{u})^{\boldsymbol{T}}\underbrace{\boldsymbol{N}\boldsymbol{N}^{-\boldsymbol{1}}}_{\boldsymbol{I}}\boldsymbol{u} = \boldsymbol{u}^{\boldsymbol{T}}\boldsymbol{N}^{-\boldsymbol{1}}\boldsymbol{u} \tag{4.5}$$

as alternatively expressions for MWSS, so:

$$MWSS = \hat{\boldsymbol{y}}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{-\boldsymbol{1}}\hat{\boldsymbol{y}} = \hat{\boldsymbol{\beta}}^{\boldsymbol{T}}\boldsymbol{N}\hat{\boldsymbol{\beta}} = \boldsymbol{u}^{\boldsymbol{T}}\boldsymbol{N}^{-\boldsymbol{1}}\boldsymbol{u} \tag{4.6}$$

4.1.2 Stationary data

First assume the vector $\boldsymbol{y}$ consisting of a time series from a zero mean process, with known a priori covariance matrix, $\boldsymbol{cov}(\boldsymbol{y}) = \sigma^2\boldsymbol{\Sigma}_{\boldsymbol{yy}}$. The vector $\boldsymbol{\beta}$ consists of chosen pairs of parameters $a(\omega_k), b(\omega_k)$, corresponding to frequencies which are not necessarily limited to the Fourier frequencies:

$$\boldsymbol{\beta} = \underbrace{[a(\omega_1), b(\omega_1), a(\omega_2), b(\omega_2), ..., a(\omega_k), b(\omega_k), ..., a(\omega_m), b(\omega_m)]}_{(1\cdot 2m)}^{\,T} \tag{4.7}$$

For $k = 1, 2, ..., m$, corresponding column vectors for cosine- and sine terms of length n could be written as:

$$\boldsymbol{c}_{\omega_k} = \begin{bmatrix} cos(\omega_k t_1) \\ cos(\omega_k t_2) \\ \vdots \\ cos(\omega_k t_i) \\ \vdots \\ cos(\omega_k t_n) \end{bmatrix} \qquad \boldsymbol{s}_{\omega_k} = \begin{bmatrix} sin(\omega_k t_1) \\ sin(\omega_k t_2) \\ \vdots \\ sin(\omega_k t_i) \\ \vdots \\ sin(\omega_k t_n) \end{bmatrix} \tag{4.8}$$

The column vectors may be collected again in a design matrix:

$$\boldsymbol{X} = \underbrace{[\boldsymbol{c}_{\omega_1}\vdots\boldsymbol{s}_{\omega_1}\vdots\boldsymbol{c}_{\omega_2}\vdots\boldsymbol{s}_{\omega_2}\vdots....\vdots\boldsymbol{c}_{\omega_k}\vdots\boldsymbol{s}_{\omega_k}\vdots...\vdots\boldsymbol{c}_{\omega_m}\vdots\boldsymbol{s}_{\omega_m}]}_{n\cdot 2m} \tag{4.9}$$

With the choice $m < n/2$, the system of linear equations is overdetermined, and a WLS estimator for $\boldsymbol{\beta}$ is:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{-\boldsymbol{1}}\boldsymbol{X})^{-\boldsymbol{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{-\boldsymbol{1}}\boldsymbol{y} \tag{4.10}$$

The relation to Fourier

Assume a complete, equally spaced and equally weighted ($\boldsymbol{\Sigma_{yy}} = \boldsymbol{I}$) data series. For the choice of Fourier frequencies in the $\boldsymbol{X}$-matrix, the estimated parameters in $\hat{\boldsymbol{\beta}}$ will be identical to the Fourier coefficients. In other words, Fourier analysis is a special case of the more general WLS procedure. This is shown by Craymer [1998]

Using Fourier frequencies ω_k, the column vectors in the $\boldsymbol{X}$-matrix will be orthogonal, which means the parameter estimates in $\hat{\boldsymbol{\beta}}$ become independent of each other. Sub-matrices and sub-vectors representing the frequency ω_k will be:

$$\underset{(n\cdot 2)}{\boldsymbol{X_k}} = [\boldsymbol{c}_{\omega_k} \vdots \boldsymbol{s}_{\omega_k}] \qquad \underset{(2\cdot 1)}{\boldsymbol{\beta_k}} = [a(\omega_k) \vdots b(\omega_k)]^{\boldsymbol{T}} \tag{4.11}$$

It will be possible to solve the pairs of parameters $\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}$, independently from each other, and separate their contributions to the total observation sum of squares in partial sums from each pair of estimated Fourier coefficients:

$$\boldsymbol{y^T y} = \hat{\boldsymbol{y}}^{\boldsymbol{T}} \hat{\boldsymbol{y}} + \hat{\boldsymbol{\varepsilon}}^{\boldsymbol{T}} \hat{\boldsymbol{\varepsilon}} = \hat{\boldsymbol{\beta}}^{\boldsymbol{T}} \boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{X} \hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\varepsilon}}^{\boldsymbol{T}} \hat{\boldsymbol{\varepsilon}} \tag{4.12}$$

with

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{\boldsymbol{T}} \boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{X} \hat{\boldsymbol{\beta}} = \; & \hat{\boldsymbol{\beta}}_{\mathbf{1}}^{\boldsymbol{T}} \boldsymbol{X}_{\mathbf{1}}^{\boldsymbol{T}} \boldsymbol{X}_{\mathbf{1}} \hat{\boldsymbol{\beta}}_{\mathbf{1}} + \hat{\boldsymbol{\beta}}_{\mathbf{2}}^{\boldsymbol{T}} \boldsymbol{X}_{\mathbf{2}}^{\boldsymbol{T}} \boldsymbol{X}_{\mathbf{2}} \hat{\boldsymbol{\beta}}_{\mathbf{2}} + \ldots + \\ & \hat{\boldsymbol{\beta}}_{\boldsymbol{k}}^{\boldsymbol{T}} \boldsymbol{X}_{\boldsymbol{k}}^{\boldsymbol{T}} \boldsymbol{X}_{\boldsymbol{k}} \hat{\boldsymbol{\beta}}_{\boldsymbol{k}} + \ldots + \hat{\boldsymbol{\beta}}_{\boldsymbol{m}}^{\boldsymbol{T}} \boldsymbol{X}_{\boldsymbol{m}}^{\boldsymbol{T}} \boldsymbol{X}_{\boldsymbol{m}} \hat{\boldsymbol{\beta}}_{\boldsymbol{m}} \end{aligned} \tag{4.13}$$

where the contribution to the total sum of squares explained by the wave, or frequency k is:

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}^{\boldsymbol{T}} \boldsymbol{X}_{\boldsymbol{k}}^{\boldsymbol{T}} \boldsymbol{X}_{\boldsymbol{k}} \hat{\boldsymbol{\beta}}_{\boldsymbol{k}} \tag{4.14}$$

A graphical interpretation is possible. Assume that a wave k as in Figure 2.2 is fitted to a zero mean stationary data series using a least squares procedure. In a special case when the wave does not fit the data at all, the estimated amplitude:

$$\hat{A}_k = \sqrt{\hat{a}(\omega_k)^2 + \hat{b}(\omega_k)^2} \tag{4.15}$$

becomes zero (and the fitted wave a straight line). The residual sum of squares, or the reference variance will remain unchanged. For all other fitted waves, the remaining residual sum of squares, or the reference variance will always be smaller. The contribution to the total variance (for the zero mean process) explained by the wave or frequency k is:

$$S_{k-norm} = \frac{\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}^{\boldsymbol{T}} \boldsymbol{X}_{\boldsymbol{k}}^{\boldsymbol{T}} \boldsymbol{X}_{\boldsymbol{k}} \hat{\boldsymbol{\beta}}_{\boldsymbol{k}}}{\boldsymbol{y}^{\boldsymbol{T}} \boldsymbol{\Sigma}_{\boldsymbol{yy}}^{-\boldsymbol{1}} \boldsymbol{y}} \tag{4.16}$$

where

$$\begin{aligned} &\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}^{\boldsymbol{T}} \boldsymbol{X}_{\boldsymbol{k}}^{\boldsymbol{T}} \boldsymbol{X}_{\boldsymbol{k}} \hat{\boldsymbol{\beta}}_{\boldsymbol{k}} = \\ &\left[\begin{array}{c:c} \hat{a}(\omega_k) & \hat{b}(\omega_k) \end{array} \right] \left[\begin{array}{cc} \sum_{i=1}^{n} cos^2(\omega_k t_i) & 0 \\ 0 & \sum_{i=1}^{n} sin^2(\omega_k t_i) \end{array} \right] \left[\begin{array}{c} \hat{a}(\omega_k) \\ \cdots \\ \hat{b}(\omega_k) \end{array} \right] = \\ &n[\hat{a}(\omega_k)^2 + \hat{b}(\omega_k)^2] = n\hat{A}_k^2 \end{aligned} \tag{4.17}$$

using analogous expressions for sums as in Eq. (2.62). The expression in Eq. (4.16) is always between zero and one and may thus be seen as an estimate for the normalised spectral density function for the one-sided Fourier or amplitude spectrum.

4.1.3 Non-stationary weighted data

For the challenge to find the spectral contents of a non-stationary data series with different kinds of trends, several difficult problems have to be considered.

1. Using other than the Fourier frequencies, it is not possible to separate the partial sums as for $\hat{\boldsymbol{\beta}}^{\boldsymbol{T}} \boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{X} \hat{\boldsymbol{\beta}}$ in Eq. (4.13), because the spectral parameter estimates in $\hat{\boldsymbol{\beta}}$ will be correlated.

2. For non-stationary data, the effect of trend parameters in the deterministic part of the model have to be accounted for when solving the parameter estimates in the spectral part of the model. This problem is further complicated

because very strong correlations between parameter estimates in the deterministic and the spectral part of the model may cause numerical problems. The parameter solutions can be incorrect and the spectrum distorted.

To avoid the first problem, assume the spectral contents for only one frequency k should be computed. The WSS explained by the trend or the deterministic part of the model will be denoted by subscript d. To avoid the second problem, assume there is no linear dependency between parameters in the deterministic and the spectral part of the model. Then the WSS is decomposed to:

$$\boldsymbol{y^T\Sigma_{yy}^{-1}y} = \boldsymbol{\hat\beta_d^T X_d^T\Sigma_{yy}^{-1}X_d\hat\beta_d} + \boldsymbol{\hat\beta_k^T X_k^T\Sigma_{yy}^{-1}X_k\hat\beta_k} + \boldsymbol{\hat\varepsilon^T\Sigma_{yy}^{-1}\hat\varepsilon} \qquad (4.18)$$

The contribution from frequency k to the total variance remaining after the trend has been removed will be:

$$S_{k-norm} = \frac{\boldsymbol{\hat\beta_k^T X_k^T\Sigma_{yy}^{-1}X_k\hat\beta_k}}{\boldsymbol{y^T\Sigma_{yy}^{-1}y - \hat\beta_d^T X_d^T\Sigma_{yy}^{-1}X_d\hat\beta_d}} \qquad (4.19)$$

One weakness of the method is the required time of computations. For different frequencies k, not only the sum $\boldsymbol{\hat\beta_k^T X_k^T\Sigma_{yy}^{-1}X_k\hat\beta_k}$, but also the sum $\boldsymbol{\hat\beta_d^T X_d^T\Sigma_{yy}^{-1}X_d\hat\beta_d}$ will change. For each frequency k, a new adjustment will be necessary.

The second problem has to be considered. When linear dependency between parameters in the deterministic and the spectral part of the model exists, then the terms:

$$\boldsymbol{\hat\beta_k^T X_k^T\Sigma_{yy}^{-1}X_k\hat\beta_k + \hat\beta_d^T X_d^T\Sigma_{yy}^{-1}X_d\hat\beta_d} \qquad (4.20)$$

in Eq. (4.18) cannot possibly be separated. A separate adjustment in a frequency free model with trend parameters only, has to be done to find the RWSS in the denominator of Eq. (4.19) as:

$$\boldsymbol{\hat\varepsilon_d^T\Sigma_{yy}^{-1}\hat\varepsilon_d = y^T\Sigma_{yy}^{-1}y - \hat\beta_d^T X_d^T\Sigma_{yy}^{-1}X_d\hat\beta_d} \qquad (4.21)$$

For the problem of very strong correlations between parameter estimates in the deterministic and the spectral part of the model, one could try first to remove the trend $\boldsymbol{X}\hat{\boldsymbol{\beta}}$ from the observations $\boldsymbol{y}$ and then compute a spectrum of the remaining residuals $\hat{\boldsymbol{\varepsilon}}$. To handle jumps in data, some simulations using this method have been made in Section 4.4.

The relation to Variance Spectrum (VS)

The numerator in Eq. (4.19) referred to Eq. (4.18) can be written:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}^{\boldsymbol{T}}\boldsymbol{X}_{\boldsymbol{k}}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{\boldsymbol{-1}}\boldsymbol{X}_{\boldsymbol{k}}\hat{\boldsymbol{\beta}}_{\boldsymbol{k}} = \;& \boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{\boldsymbol{-1}}\boldsymbol{y} - \hat{\boldsymbol{\beta}}_{\boldsymbol{d}}^{\boldsymbol{T}}\boldsymbol{X}_{\boldsymbol{d}}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{\boldsymbol{-1}}\boldsymbol{X}_{\boldsymbol{d}}\hat{\boldsymbol{\beta}}_{\boldsymbol{d}} - \hat{\boldsymbol{\varepsilon}}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{\boldsymbol{-1}}\hat{\boldsymbol{\varepsilon}} = \\ & \hat{\boldsymbol{\varepsilon}}_{\boldsymbol{d}}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{\boldsymbol{-1}}\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{d}} - \hat{\boldsymbol{\varepsilon}}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{\boldsymbol{-1}}\hat{\boldsymbol{\varepsilon}}\end{aligned} \tag{4.22}$$

so Eq. (4.19) can be written:

$$1 - \frac{\hat{\boldsymbol{\varepsilon}}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{\boldsymbol{-1}}\hat{\boldsymbol{\varepsilon}}}{\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{d}}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{\boldsymbol{-1}}\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{d}}} = VS(k) \tag{4.23}$$

Accounting for the effect of $\boldsymbol{cov}(\hat{\boldsymbol{\beta}}_{\boldsymbol{d}}, \hat{\boldsymbol{\beta}}_{\boldsymbol{k}}) \neq \boldsymbol{0}$, a separate adjustment is made to estimate $\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{d}}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{\boldsymbol{-1}}\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{d}}$. This is equivalent to the weighted form of variance spectrum (VS), used by Plag [1988]. The values of $VS(k)$ will always be in the interval $0 \leq VS(k) \leq 1$, in fact this are the normalised spectral values from a spectral density function.

From regression analysis the expression for $VS(k)$ in Eq. (4.23) is known as the coefficient of determination R^2 [Walpole and Myers, 1993] that measures the quality of model fit. If the fitted wave k passes through all data points, so $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{0}$, then $R^2 = 1$. At the other extreme, $R^2 = 0$.

Test statistics for spectral parameters

Rewriting $VS(k)$ in Eq. (4.23) as:

$$VS(k) = \frac{(RWSS_d - RWSS_{(s+d)})}{RWSS_d} \tag{4.24}$$

Comparing this expression with Eq. (3.38), it is possible to derive a similar test for several pairs of spectral parameters. Let d be the number of parameters in the deterministic model, and $(s + d)$ the sum of parameters in both the deterministic and the spectral part of the model. For a test of one frequency only, the number of spectral parameters are $(s = 2)$, but it is possible to derive an extended test, for more than one frequency, using several pairs of spectral parameters when computing $RWSS_{(s+d)}$. Several pairs of spectral parameters are simultaneously significant if:

$$\frac{(RWSS_d - RWSS_{(s+d)})}{RWSS_d} > \frac{(s)}{(n-d)} F_{(s,n-d)}(\alpha) \tag{4.25}$$

4.2 Weighted Least Squares Spectral Analysis

In LSSA, the parameters in the deterministic part of the model are in fact not estimated at all, but within the estimation of spectral components there have to be taken account for the effect of the deterministic model. A trick is to separate the parameter vector $\boldsymbol{\beta}$ in two parts, one for parameters as a solution of the deterministic part of the model $\boldsymbol{\beta_d}$, with subscript d. Another one $\boldsymbol{\beta_s}$, for the parameters $a(\omega_k)$ and $b(\omega_k)$, we want to solve in the spectral part of the model, with subscript s. The parameter vector $\boldsymbol{\beta_s}$ will include pairs of spectral parameters for all the frequencies k wanted to be solved for, and is not bounded to a single frequency $\boldsymbol{\beta_k}$ as in Subsection 4.1.3. A second design matrix $\boldsymbol{X_s}$ describes the linear combinations of the residuals in sin- and cos terms, as in Eq. (4.9).

An expression for the spectral parameters in $\boldsymbol{\beta_s}$ which are related to the later computed spectral values will be derived without computing the solution of $\boldsymbol{\beta_d}$. This is made splitting the full normal matrix $\boldsymbol{N}$ into four blocks, which results in a less computable demanding form for an inversion to the cofactor matrix $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}}$. The final matrix expressions may look rather complex, but can be collecting into reduced matrices an vectors (denoted with stars) that can be used in classical least squares algorithms.

The formula for the common solution vector $\boldsymbol{\beta} = [\boldsymbol{\beta_d} \vdots \boldsymbol{\beta_s}]^T$ is an estimator de-

rived from the principle of least squares:

$$E[\boldsymbol{y}] = \boldsymbol{y} - \boldsymbol{\epsilon} = \boldsymbol{X}\boldsymbol{\beta} = \left[\begin{array}{c:c} \boldsymbol{X_d} & \boldsymbol{X_s} \end{array} \right] \left[\begin{array}{c} \boldsymbol{\beta_d} \\ \dots \\ \boldsymbol{\beta_s} \end{array} \right] = \boldsymbol{X_d}\boldsymbol{\beta_d} + \boldsymbol{X_s}\boldsymbol{\beta_s} \quad (4.26)$$

Least squares estimators:

$$\hat{\boldsymbol{\beta}} = \underbrace{(\boldsymbol{X}^T\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{X})^{-1}}_{\boldsymbol{N^{-1}}}\underbrace{\boldsymbol{X}^T\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{y}}_{\boldsymbol{u}} = \boldsymbol{N^{-1}}\boldsymbol{u} \quad (4.27)$$

or divided:

$$\begin{gathered} \left[\begin{array}{c} \hat{\boldsymbol{\beta}}_{\boldsymbol{d}} \\ \dots \\ \hat{\boldsymbol{\beta}}_{\boldsymbol{s}} \end{array} \right] = \left[\begin{array}{c:c} \boldsymbol{X_d^T\Sigma_{yy}^{-1}X_d} & \boldsymbol{X_d^T\Sigma_{yy}^{-1}X_s} \\ \dots & \dots \\ \boldsymbol{X_s^T\Sigma_{yy}^{-1}X_d} & \boldsymbol{X_s^T\Sigma_{yy}^{-1}X_s} \end{array} \right]^{-1} \cdot \left[\begin{array}{c} \boldsymbol{X_d^T\Sigma_{yy}^{-1}y} \\ \dots \\ \boldsymbol{X_s^T\Sigma_{yy}^{-1}y} \end{array} \right] = \\ \left[\begin{array}{c:c} \boldsymbol{N_{dd}} & \boldsymbol{N_{ds}} \\ \dots & \dots \\ \boldsymbol{N_{sd}} & \boldsymbol{N_{ss}} \end{array} \right]^{-1} \cdot \left[\begin{array}{c} \boldsymbol{u_d} \\ \dots \\ \boldsymbol{u_s} \end{array} \right] = \left[\begin{array}{c:c} \boldsymbol{M_{dd}} & \boldsymbol{M_{ds}} \\ \dots & \dots \\ \boldsymbol{M_{sd}} & \boldsymbol{M_{ss}} \end{array} \right] \cdot \left[\begin{array}{c} \boldsymbol{u_d} \\ \dots \\ \boldsymbol{u_s} \end{array} \right] \end{gathered} \quad (4.28)$$

The inverse of the normal equation hyper matrix $\boldsymbol{N}$ can be written:

$$\left[\begin{array}{c:c} \boldsymbol{N_{dd}} & \boldsymbol{N_{ds}} \\ \dots & \dots \\ \boldsymbol{N_{sd}} & \boldsymbol{N_{ss}} \end{array} \right]^{-1} = \left[\begin{array}{c:c} \boldsymbol{M_{dd}} & \boldsymbol{M_{ds}} \\ \dots & \dots \\ \boldsymbol{M_{sd}} & \boldsymbol{M_{ss}} \end{array} \right] \quad (4.29)$$

where:

$$\boldsymbol{M_{dd}} = \boldsymbol{N_{dd}^{-1}} + \boldsymbol{N_{dd}^{-1}N_{ds}}\underbrace{\left(\boldsymbol{N_{ss}} - \boldsymbol{N_{sd}N_{dd}^{-1}N_{ds}}\right)^{-1}}_{\boldsymbol{N^{*-1}}}\boldsymbol{N_{sd}N_{dd}^{-1}} \quad (4.30)$$

$$\boldsymbol{M_{ds}} = \boldsymbol{N_{dd}^{-1}N_{ds}}\underbrace{\left(\boldsymbol{N_{ss}} - \boldsymbol{N_{sd}N_{dd}^{-1}N_{ds}}\right)^{-1}}_{\boldsymbol{N^{*-1}}} \quad (4.31)$$

$$\boldsymbol{M_{sd}} = -\underbrace{\left(\boldsymbol{N_{ss}} - \boldsymbol{N_{sd}N_{dd}^{-1}N_{ds}}\right)^{-1}}_{\boldsymbol{N^{*-1}}}\boldsymbol{N_{sd}N_{dd}^{-1}} \quad (4.32)$$

$$\boldsymbol{M_{ss}} = \underbrace{\left(\boldsymbol{N_{ss}} - \boldsymbol{N_{sd}N_{dd}^{-1}N_{ds}}\right)^{-1}}_{\boldsymbol{N^{*-1}}} \quad (4.33)$$

A proof of the matrix inversion method is given in Appendix B. The solution for $\boldsymbol{\beta_s}$ is now found inserting the expressions for $\boldsymbol{M_{sd}}$ from Eq. (4.32) and $\boldsymbol{M_{ss}}$ from Eq. (4.33) into Eq. (4.28).

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\boldsymbol{s}} &= \boldsymbol{M_{sd}}\boldsymbol{u_d} + \boldsymbol{M_{ss}}\boldsymbol{u_s} \\ &= -\boldsymbol{N}^{*-1}\boldsymbol{N_{sd}}\boldsymbol{N_{dd}^{-1}}\boldsymbol{u_d} + \boldsymbol{N}^{*-1}\boldsymbol{u_s} \\ &= \boldsymbol{N}^{*-1}\underbrace{(\boldsymbol{u_s} - \boldsymbol{N_{sd}}\boldsymbol{N_{dd}^{-1}}\boldsymbol{u_d})}_{\boldsymbol{u}^*} \\ &= \boldsymbol{N}^{*-1}\boldsymbol{u}^*\end{aligned} \tag{4.34}$$

which is a solution for $\hat{\boldsymbol{\beta}}_{\boldsymbol{s}}$ without solving $\hat{\boldsymbol{\beta}}_{\boldsymbol{d}}$. As an alternative to the derived formula (4.34) it is possible to compute $\hat{\boldsymbol{\beta}}_{\boldsymbol{s}}$ making use of a corrected or reduced inverse covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{*-1}$. Looking back at Eq. (4.28) one get:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\boldsymbol{s}} = \boldsymbol{N}^{*-1}\boldsymbol{u}^* = \; & (\boldsymbol{N_{ss}} - \boldsymbol{N_{sd}}\boldsymbol{N_{dd}^{-1}}\boldsymbol{N_{ds}})^{-1}(\boldsymbol{u_s} - \boldsymbol{N_{sd}}\boldsymbol{N_{dd}^{-1}}\boldsymbol{u_d}) = \\ & (\boldsymbol{X_s^T}\boldsymbol{\Sigma_{yy}^{-1}}\boldsymbol{X_s} - \boldsymbol{X_s^T}\boldsymbol{\Sigma_{yy}^{-1}}\boldsymbol{X_d}\boldsymbol{N_{dd}^{-1}}\boldsymbol{X_d^T}\boldsymbol{\Sigma_{yy}^{-1}}\boldsymbol{X_s})^{-1}\cdot \\ & \quad (\boldsymbol{X_s^T}\boldsymbol{\Sigma_{yy}^{-1}}\boldsymbol{y} - \boldsymbol{X_s^T}\boldsymbol{\Sigma_{yy}^{-1}}\boldsymbol{X_d}\boldsymbol{N_{dd}^{-1}}\boldsymbol{X_d^T}\boldsymbol{\Sigma_{yy}^{-1}}\boldsymbol{y}) = \\ & [\boldsymbol{X_s^T}\underbrace{(\boldsymbol{\Sigma_{yy}^{-1}} - \boldsymbol{\Sigma_{yy}^{-1}}\boldsymbol{X_d}\boldsymbol{N_{dd}^{-1}}\boldsymbol{X_d^T}\boldsymbol{\Sigma_{yy}^{-1}})}_{\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{*-1}}\boldsymbol{X_s}]^{-1}\cdot \\ & \boldsymbol{X_s^T}\underbrace{(\boldsymbol{\Sigma_{yy}^{-1}} - \boldsymbol{\Sigma_{yy}^{-1}}\boldsymbol{X_d}\boldsymbol{N_{dd}^{-1}}\boldsymbol{X_d^T}\boldsymbol{\Sigma_{yy}^{-1}})}_{\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{*-1}}\boldsymbol{y} = \\ & \hat{\boldsymbol{\beta}}_{\boldsymbol{s}} = [\boldsymbol{X_s^T}\boldsymbol{\Sigma_{yy}^{*-1}}\boldsymbol{X_s}]^{-1}\cdot\boldsymbol{X_s^T}\boldsymbol{\Sigma_{yy}^{*-1}}\boldsymbol{y}\end{aligned} \tag{4.35}$$

So changing the a priori covariance matrix to:

$$\boldsymbol{\Sigma_{yy}^{*-1}} = (\boldsymbol{\Sigma_{yy}^{-1}} - \boldsymbol{\Sigma_{yy}^{-1}}\boldsymbol{X_d}\boldsymbol{N_{dd}^{-1}}\boldsymbol{X_d^T}\boldsymbol{\Sigma_{yy}^{-1}}) \tag{4.36}$$

makes it possible to find the solution for the parameters in $\hat{\boldsymbol{\beta}}_{\boldsymbol{s}}$ in Eq. (4.34) or Eq. (4.35), and nevertheless account for the effect of the deterministic part of the model, using a classical least-squares algorithm.

4.2.1 Normalised Spectral Values

Accounting for the effect of the deterministic model, spectral estimates S_k^*, may be found handling the reduced normal matrix $\boldsymbol{N}^*$ in two different ways. Craymer [1998] denotes the methods as independent and simultaneous spectral estimations.

Simultaneous spectral estimation

First compute $\boldsymbol{N}^*$ which is diagonal only in the case when Fourier frequencies are chosen. Then invert the total $\boldsymbol{N}^*$ to find $\boldsymbol{N}^{*-1} = \boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}_s \hat{\boldsymbol{\beta}}_s}$. To avoid singularity problems, care must be taken for the choice of frequencies. Spectral estimates S_k^* for frequency k are found, picking two corresponding elements from the $\hat{\boldsymbol{\beta}}_s$-vector, alternatively the easier computable $\boldsymbol{u}^*$-vector and the $(2{\cdot}2)$ quadratic block element from the $\boldsymbol{N}^{*-1}$-matrix

$$\underset{(1\cdot1)}{S_k^*} = \underset{(1\cdot2)}{(\hat{\boldsymbol{\beta}}_s)_k^{\boldsymbol{T}}} \underset{(2\cdot2)}{(\boldsymbol{N}^*)_k} \underset{(2\cdot1)}{(\hat{\boldsymbol{\beta}}_s)_k} = \underset{(1\cdot2)}{\boldsymbol{u}_{\boldsymbol{k}}^{*\boldsymbol{T}}} \underset{(2\cdot2)}{(\boldsymbol{N}^{*-1})_k} \underset{(2\cdot1)}{\boldsymbol{u}_{\boldsymbol{k}}^*} \tag{4.37}$$

The relation with other frequency components is accounted for during the inversion of $\boldsymbol{N}^*$, thus this method is denoted simultaneous or in-the-context spectral estimation. Normalised spectral values, always between zero and one are found dividing this value at the reduced OWSS (that will be introduced in Eq. (4.41)):

$$S_{k-norm}^* = \frac{S_k^*}{\boldsymbol{y}^{\boldsymbol{T}} \boldsymbol{\Sigma}_{\boldsymbol{yy}}^{*-1} \boldsymbol{y}} \tag{4.38}$$

Independent spectral estimation

An alternative more robust computation of spectral values can be made inverting only separately picked $(2 \cdot 2)$ quadratic block elements from $\boldsymbol{N}$. Each spectral estimate is computed without concerning correlations with other spectral estimates. The method is denoted independently or out-of-context spectral estimation.

$$\underset{(1\cdot1)}{S_k^*} = \underset{(1\cdot2)}{\boldsymbol{u}_{\boldsymbol{k}}^{*\boldsymbol{T}}} \underset{(2\cdot2)}{(\boldsymbol{N}_{\boldsymbol{k}}^*)^{-1}} \underset{(2\cdot1)}{\boldsymbol{u}_{\boldsymbol{k}}^*} \tag{4.39}$$

Normalised spectral values:

$$S_{k-norm}^* = \frac{S_k^*}{\boldsymbol{y}^{\boldsymbol{T}} \boldsymbol{\Sigma}_{\boldsymbol{yy}}^{*-1} \boldsymbol{y}} \tag{4.40}$$

Going backwards, it may be easier to show that the independent estimated spectral value above is equivalent with the expression in Eq.(4.19). The reduced OWSS in the denominator of Eqs. (4.40) and (4.38) is:

$$\begin{aligned} \boldsymbol{y^T\Sigma_{yy}^{*-1}y} = \quad & \boldsymbol{y^T[\Sigma_{yy}^{-1} - \Sigma_{yy}^{-1}X_dN_{dd}^{-1}X_d^T\Sigma_{yy}^{-1}]y} = \\ & \boldsymbol{y^T\Sigma_{yy}^{-1}y - y^T\Sigma_{yy}^{-1}X_dN_{dd}^{-1}X_d^T\Sigma_{yy}^{-1}y} = \\ & \boldsymbol{y^T\Sigma_{yy}^{-1}y - (X_d^T\Sigma_{yy}^{-1}y)^TN_{dd}^{-1}(X_d^T\Sigma_{yy}^{-1}y)} = \\ & \boldsymbol{y^T\Sigma_{yy}^{-1}y - u_d^TN_{dd}^{-1}N_{dd}N_{dd}^{-1}u_d} = \\ & \boldsymbol{y^T\Sigma_{yy}^{-1}y - \hat{\beta}_d^TX_d^T\Sigma_{yy}^{-1}X_d^T\hat{\beta}_d} = \\ & \boldsymbol{\hat{\varepsilon}_d^T\Sigma_{yy}^{-1}\hat{\varepsilon}_d} \end{aligned} \tag{4.41}$$

For only one frequency k, the numerator in Eq. (4.40) can also be shown to be equal to the numerator in Eq. (4.23) and hence the normalised spectral value found for the Independent Least Squares Spectrum (ILSS) is also equivalent with a weighted form of variance spectrum from Plag [1988]. The derived spectra in the simulations in Section 4.4 and in the numerical investigations in Chapter 6 are all Independent Least Squares Spectra.

4.2.2 Merging effects

In addition to the problem of aliasing, already discussed in Section 2.2.2, there are some other effects affecting the estimated spectrum which has to be mentioned.

Spectral leakage

Simplifying from a time series of length ($L \rightarrow \infty$), to a time series of finite duration L, will cause some distortions in the Fourier Spectrum called spectral leakage. Peaks are spread over the frequency domain. Close peaks may be overlapped or merged together, and side lobes of the original peaks may appear as ghost peaks.

Spectral resolution

Spectral resolution depends on the length L of the time series [see Abbasi, 1999]. The difference between two frequencies in a Fourier spectrum has to be:

$$|f_1 - f_2| > \frac{1}{L} \tag{4.42}$$

to be separated.

Primary, secondary and higher order spectra

The problems mentioned above may appear even more noticeably with trends in data and the use of LSSA. Investigations of the first or primary estimated spectrum may not be enough to give a sufficient picture of a time series frequency contents. Moving selected parameters for significant frequencies from the primary spectrum over to the deterministic part of the model, a secondary spectrum can be estimated. This may reveal some frequencies that are hidden in the primary estimated spectrum. One should beware of estimation of frequencies in the secondary spectrum close to frequencies moved to the deterministic part. It will often cause ill-conditioning and numerical problems.

Finally, a number of significant frequencies have to be chosen. Iterative numerical routines may compensate for some of the merging effect, caused by correlations or spectral leakage between spectral values. A simultaneous test of significance, may be done for the selected choice of frequencies. The result may cause a decision of adding or removing a frequency. In that case the demanding iterative merging routine computation has to be redone.

4.3 Iteratively re-weighted least squares estimation

Craymer [1998] describes a procedure for stochastically modelling residuals from different weighted and unequally spaced time series of observations, based on a weighted regression model named iteratively re-weighted least squares estimation. A short description of this procedure follows:

Assume that an optimal functional model is found. Start with some known a priori (often diagonal) covariance matrix $\boldsymbol{cov}(\boldsymbol{y}) = \sigma^2\boldsymbol{\Sigma}_{\boldsymbol{yy}}$. Make an adjustment and estimate a normalised spectrum of the residuals. This spectrum, named the weighted Least Squares Spectrum (LSS) is derived in Section 4.2. He denotes the relation $\boldsymbol{u} = \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{\boldsymbol{-1}}\boldsymbol{y}$ (analogous to the Fourier Transform) a Least Squares Transform (LST), which is valid also for a weighted data-series with gaps. An estimated ACF for the residuals can be found during an inverse LST of the normalised weighted LSS. Using a priori variances combined with the indirect estimated ACF, a new full covariance matrix $\boldsymbol{cov}(\boldsymbol{y})$ can be generated. This one is included in a new least squares adjustment, to estimate a new set of residuals, spectrum and ACF. The procedure will then be repeated until the solution for the deterministic model and the covariance matrix converge to a stable form. The procedure is illustrated in Figure 4.1.

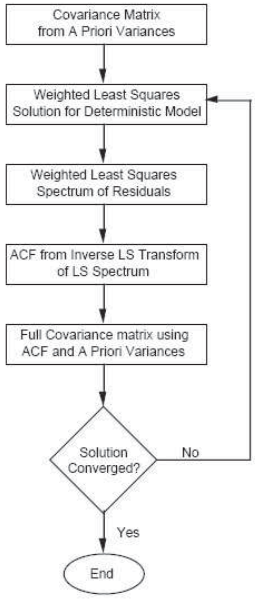

Figure 4.1: Iteratively re-weighted least squares estimation. Figure from Craymer [1998]

Investigations by Craymer [1998] on simulated data showed that even for highly autocorrelated observations, the estimated parameters were not significantly affected. From the conclusion on p.157-158 in Craymer [1998]: *Ignoring the correlations was found to have little effect on the results (estimated spectra and model*

parameters) but it significantly increases the computational efficiency. On the other hand, the use of a priori correlations among the observations generally gives more realistic estimates of the uncertainties associated with the estimated model parameters. Ignoring correlations generally results in overly optimistic estimates of the standard deviations.

4.4 Simulations with non-stationary data

Some simulations have been performed to validate the new written software and for a detailed study of the theory of spectral analysis used for the non-stationary data. The simulations have shown that detection of all data jumps and their includsion in the deterministic part of the model are of vital importance to derive correct spectra.

4.4.1 Distortions of the Independent Least Squares Spectrum

Several equally weighted time series, more than three years long have been created. The values chosen for the synthetic series are typical for the later investigation of CGPS data. The time array that is used with gaps included is identical to the one from the Norwegian series **trh3** of station Trondheim. Observations have been derived from the function:

$$f(t) = \beta_0 + \beta_1 t + \sum_{j=2}^{3} \beta_j h(t - t_{\beta_j}) + \sum_{k=1}^{2} A_k sin(2\pi f_k t + \phi_k) + \epsilon(t) \quad (4.43)$$

The first three terms are identical to the definitions from the model in Eq. (3.40). The fourth term represents two wave components as described in Eq. (2.16) with $\omega_k = 2\pi f_k$. The last term $\epsilon(t)$, is independent Gaussian white noise from a $N(0, \sigma_\epsilon^2)$-distribution. Values used to form different synthetic time series are listed in Table 4.1. As an example, the synthetic time series used for the two first simulations **sim01** and **sim02** includes all parts of Eq.(4.43) except for the noise term. The series is plotted in Figure 4.2. Except for the missing jump effect, the synthetic series used for the simulation **sim03** is identical to the one used for simulation **sim01** and so on. An identical series to the one in Figure 4.2, except for

Table 4.1: Simulated time series

Synthetic time series (2000JAN01-2003APR26)													
Par:	β_0	β_1	β_2	t_{β_2}	β_3	t_{β_3}	f_1	A_1	ϕ_1	f_2	A_2	ϕ_2	σ_ϵ
$\frac{\text{unit}\Rightarrow}{\text{name}\Downarrow}$	mm	$\frac{mm}{y}$	mm	y	mm	y	$\frac{c}{y}$	mm	rad	$\frac{c}{y}$	mm	rad	mm
sim01	-10000	5	40	2001	-30	2002	1	3	0	2	5	$-\frac{\pi}{2}$	
sim02	-10000	5	40	2001	-30	2002	1	3	0	2	5	$-\frac{\pi}{2}$	
sim03	-10000	5					1	3	0	2	5	$-\frac{\pi}{2}$	
sim04	0	0					1	3	0	2	5	$-\frac{\pi}{2}$	
sim06	-10000	5	200	2001	-100	2002	1	3	0	2	5	$-\frac{\pi}{2}$	
sim07	-10000	5	40	2001			1	3	0	2	5	$-\frac{\pi}{2}$	
sim08	-10000	5	40	2001	-30	2002	1	3	0	2	5	$-\frac{\pi}{2}$	2

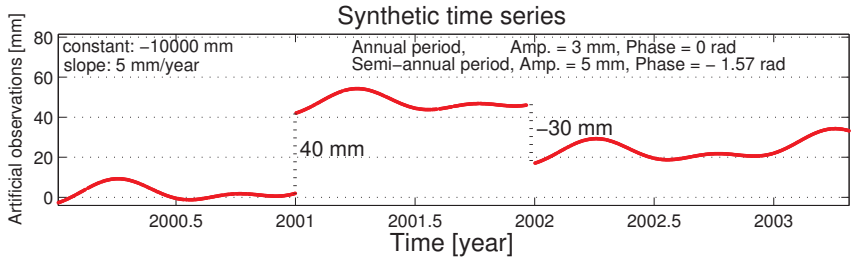

Figure 4.2: Synthetic noise free time series used for simulations: **sim01** and **sim02**.

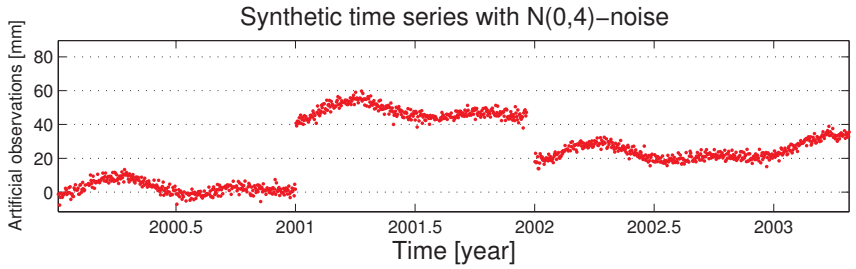

Figure 4.3: The synthetic time series with noise used for simulation: **sim08**. The series **sim08** are identical to **sim01** in Figure 4.2, except for the noise added from a $N(0, 4)$-distribution.

Table 4.2: Parameters solved for in the deterministic model under the LSSA

Parameters in det. model				
Par./name	β_0	β_1	β_2	β_3
sim01	*	*		
sim02	*	*	*	*
sim03	*	*		
sim04	*	*		
sim06	*	*	*	*
sim07	*	*		
sim08	*	*		

noise added from a $N(0, 4)$-distribution is plotted in Figure 4.3 and used in simulation **sim08**. Several deterministic models are used in the investigations. The parameters included and solved for in the various deterministic models are listed in Table 4.2. The resulting spectra are shown in Figures 4.4 to 4.7.

Interpretation of LSSA-simulations

Simulation **sim01** illustrates the situation of two jumps in data which is not included and solved for in the deterministic model. The depicted ILSS is shown on the left-hand side of Figure 4.4. The resulting spectrum is completely useless. The same happens with the ILSS from simulation **sim07** shown in the right-hand side of Figure 4.6. One jump , β_2 at time t_{β_2} was used when the synthetic time series was created, but not solved for in the deterministic under the ILSS estimation. The resulting spectrum indicates somehow the two periods, but they are more or less distorted.

For simulations **sim03** and **sim04**, Figure 4.5 shows that the LSSA method has detected the annual and semi-annual periods very well, independent of values chosen for β_0 and β_1 in the synthetic series, that distinguishes the two simulations. A small distortion can be seen for long periods. The problem increases if the data series includes jumps as shown for **sim02** in the right-hand part of Figure 4.4 and for **sim06** in the left-hand part of Figure 4.6. A deterministic model containing two jump parameters β_2 and β_3 at times t_{β_2} and t_{β_3} was used for these estimations. The annual and semi-annual periods can still be precisely detected,

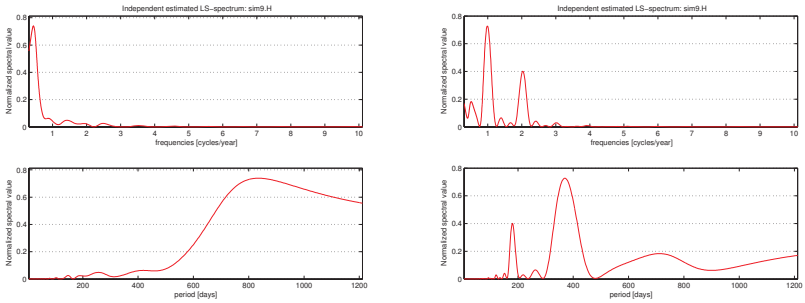

Figure 4.4: ILSS, simulations **sim01** and**sim02**

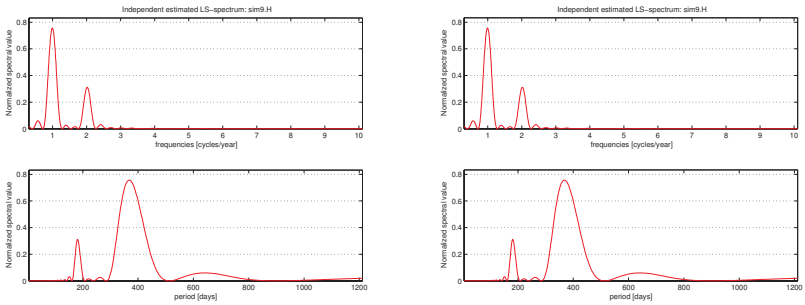

Figure 4.5: ILSS, simulations **sim03** and **sim04**

but the spectra for long periods show some larger distortions. However, the distortion is independent of the jump values which differs from the synthetic series **sim02** and **sim06**. The series **sim08** is identical to **sim02**, except that noise from a $N(0, 4)$-distribution is added. The spectrum in Figure 4.7 shows that annual and semi-annual periods still seem to be precisely detected almost unaffected by the noise, but this ILSS also has distortions for low frequencies.

The simulations have shown that the choice of a correct deterministic model when computing the ILSS is of vital importance. An insufficient selected number of jumps or steps in the deterministic model will completely destroy the ILSS as in **sim01** and **sim07**. Even if jump parameters are included precisely in time, some distortion are found in the lower frequency band (long periods) as shown in the spectra for **sim02** in Figure 4.4 and **sim06** in Figure 4.6. The distortion may be

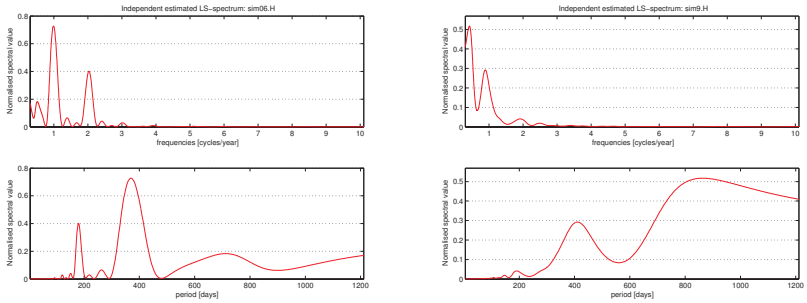

Figure 4.6: ILSS, simulations **sim06** and **sim07**

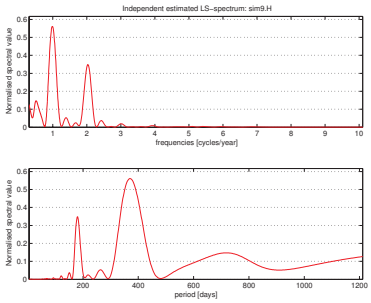

Figure 4.7: ILSS, simulation **sim08**

explained by correlations between estimates for the spectral values and jump parameters and will be further investigated in Subsection 4.4.3. A small distortion or spectral leakage can also be seen in the spectra for the jump free series **sim03** and **sim04** in Figure 4.5. This probably belongs to the correlations between the linear trend and long period wavelengths. However, the investigations have shown that care must be taken when selecting the spectral values to be estimated, especially in the low frequency band.

4.4.2 ILSS-distortions caused by jumps

A strict proof is not shown here, but an explanation of why undetected jumps in data completely destroy the spectrum will be given. The main advantages of the

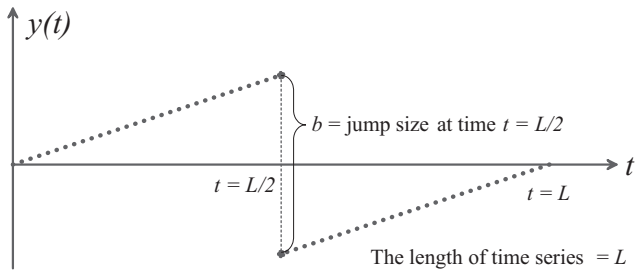

Figure 4.8: Synthetic time series of observations

VS and LSSA methods are their handling of correlations between parameter estimates in the deterministic and the stochastic parts of the model. To simplify this explanation we will disregard these correlations for a while. This can be done because simulations in Subsection 4.4.1 have shown that this correlation influence is a lot smaller than the effect of a missing jump in the model which is pointed out here. Consequently, in this explanation, a technique will be used which removes the trend before the spectrum computation.

Consider a noise free equally spaced synthetic time series with sampling interval dt and length L has been created using the parameters, const = 0, slope = a and one jump = b at time $t = L/2$. A plot of such a series is shown in Figure 4.8. Assume that a correct 3-parameter model with parameters const, slope and jump at the time $t = L/2$ was used. After an adjustment, basically no frequency components should fit the remaining residuals which would all be zero. Missing the fact that the time series contains a jump, an incomplete 2-parameters model will be used. As can be seen in Figure 4.8, the two estimated parameters const and slope in found from a least squares fit using such a model would both be zero. Remaining residuals after the adjustment, using this model, would then be just like the original data. A best fit of periodic components, in the least-squares sense, to the residuals using this incomplete deterministic model will be explained. In the fit of a sequence of sine waves $Asin(\omega t)$, simplified by setting the phase to zero, this includes only two variables: the angular frequency ω and the amplitude A. In this process one is looking for the values of A and ω which give the lowest estimated sum of squares or alternatively the minimum reference variance. New residuals will be the deviations between the observations in the $\boldsymbol{y}$-vector, simu-

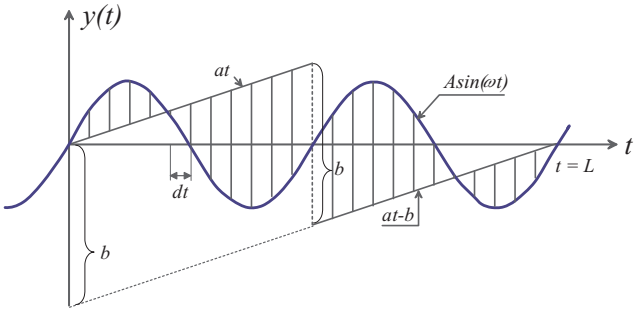

Figure 4.9: Illustration of the Sum of Squares = SS in Eq.(4.44) (= The residuals, squared and then summed).

lated by the functions $y(t) = at$ or $y(t) = at - b$, and the variable sine wave sequence $Asin(\omega t)$, see Figure 4.9. In the special case, when a jump occurs at time $t = L/2$, the sum of squares is:

$$SS = \int_{t=0}^{t=\frac{L}{2}} (at - Asin(\omega t))^2 dt + \int_{t=\frac{L}{2}}^{t=L} (at - b - Asin(\omega t))^2 dt \tag{4.44}$$

Looking for a minimum, extreme values may be found, taking the partial derivatives of the integral solution in Formula (4.44) and setting them equal to zero:

$$\partial SS(A, \omega)/\partial A = 0 \tag{4.45}$$

$$\partial SS(A, \omega)/\partial \omega = 0 \tag{4.46}$$

Omitting the mathematical problems to solve the integration, it may be easier to see graphically that an optimal fit of a wave to this data in a least-squares sense is found when the wavelength is approximately equal to L, see Figure 4.10. Note that the estimated amplitude for such a fitted (false) wave would be less than half of the jump size, provided that the jump missed in the deterministic model appears at time $t = L/2$. A wavelength of L corresponds to $\omega = 2\pi/L$ which is known as the fundamental frequency or the lowest frequency that can be found in a time series. The conclusion is not unexpected: Using an incomplete deterministic model caused by unidentified jumps will give large distortions in the lower frequency band or for long periods in the ILSS.

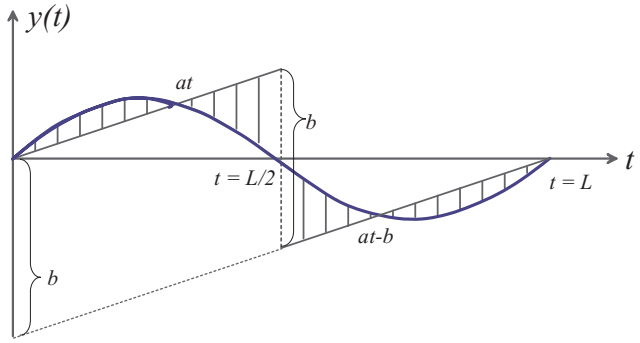

Figure 4.10: Least Sum of Squares .

4.4.3 Jump detection and correction before LSSA

The simulations in Subsection 4.4.1 have shown:

- The importance of detecting all possible jumps before the estimation of a spectrum.
- Distortions for low frequencies, probably caused by correlations between parameter estimates of jumps, linear trend and spectral values for low frequencies.

In this Subsection some further simulations will be done to demonstrate the jump detection procedure described in Section 3.3 and also show the changes of the estimated ACF through different phases of model improvement. The possibilities of correcting the time series for the effect of jumps before the LSSA will be investigated. The purpose is to improve the estimation of spectral values for low frequencies. Using synthetic time series with periodic components , it is clear that jump times and sizes found during the procedure will be different from the true values which generated the synthetic series. A procedure for this investigation will be:

1. Make a synthetic series with the function in Eq.(4.43).
2. Detect jump times t_{β_j} and jump size estimates $\hat{\beta}_j$ as described in the procedure in Section 3.3.

3. Remove the effect of jumps from the real series using:

$$y(t)_{new} = y(t)_{old} - \sum_{j=2}^{n_{jump}+1} \hat{\beta}_j h(t - t_{\beta_j}) \tag{4.47}$$

4. Do a new LSSA with the jump free data series and evaluate the differences with ILSS from Subsection 4.4.1 using models including jump parameters. The differences in the low frequency band may give particularly interesting results.

The synthetic series **sim02** containing parameters for constant term, linear trend, jumps at times $t_{\beta_2} = 2001$ and $t_{\beta_3} = 2002$ and annual and semi-annual frequency components will be investigated. For parameter values, see Table 4.1 and Figure 4.2. These computations for five different models are denoted **sim12-sim16**.

sim12

A deterministic model containing only two parameters, a constant term and linear trend was used. The left-hand side of Figure 4.11 shows the estimated ACF of normalised residuals. It indicates very large ACF-values using this temporary or incomplete model. The method explained in Section 3.3 have been used to detect jumps. The upper part of the right-hand side of Figure 4.11 shows the possible reference variance reduction for a single new jump inserted to each epoch for the whole time series. The maximum reduction was correctly found on 1 January 2001. Introducing a jump parameter at this point in time would reduce the sum of squares or the reference variance with more than 80 percent. The red coloured line in the lower part of the right-hand side of Figure 4.11 shows the estimated jump value for all epochs. However, the size of jumps is computed incorrectly i.e. a value of more than 50 mm on 1 January 2001, and a negative value of about 45 mm for a jump around turn of the year 2001 and 2002. The correct jump sizes as used in the synthetic series are: $\beta_2 = 40$ mm and $\beta_3 = -30$ mm.

sim13

One jump parameter at time 2001.0000 was added to the earlier deterministic model from **sim12**. Estimated ACF and results from a new jump detection are

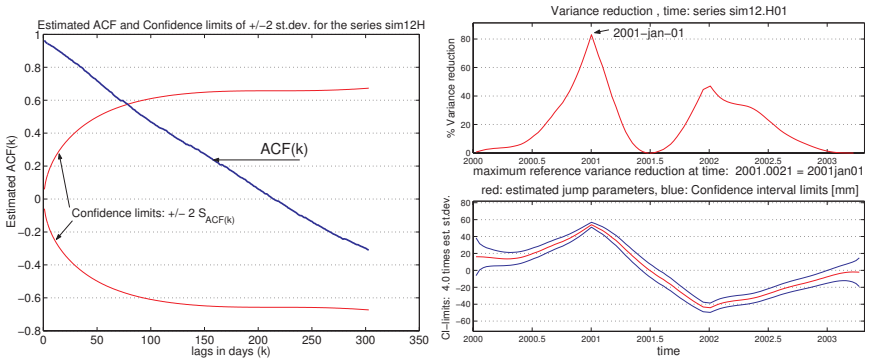

Figure 4.11: Estimated ACF of normalised residuals and results of jump detection for simulation **sim12**.

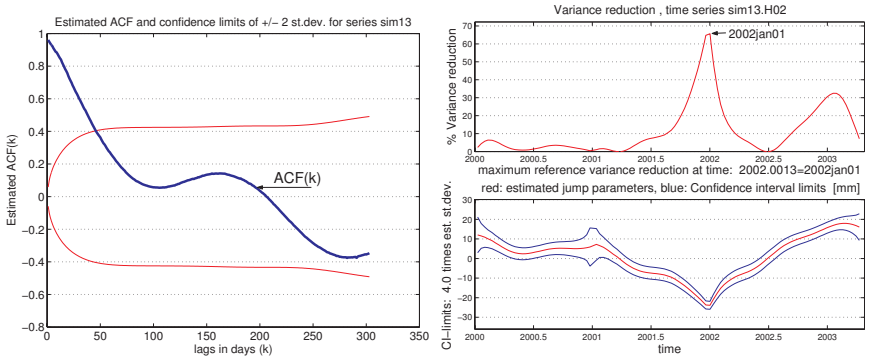

Figure 4.12: Estimated ACF of normalised residuals and results of jump detection for simulation **sim13**.

shown in Figure 4.12. The estimated ACF function decreases faster and values are lower than for **Sim12**. The second jump has been correctly detected 1 January 2002. Adding such a new jump parameter at time 2002.0000 will reduce the new reference variance by approximately 65 percent.

sim14

The model was extended with a second new jump parameter inserted at time 2002.0. Data, fitted model, estimated jumps and remaining residuals are shown in Figure 4.13. The estimated jump sizes $\hat{\beta}_2 = 44.6$ mm and $\hat{\beta}_3 = -23.9$ mm should

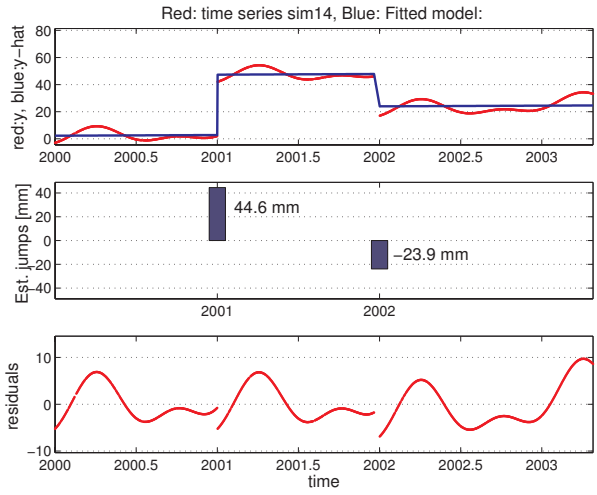

Figure 4.13: In the upper plot: Observations and Fitted model, in the middle: Estimated jumps and the lower plot: remaining residuals for series **sim14**.

be compared to the true input values in the simulation $\beta_2 = 40$ mm and $\beta_3 = -30$ mm. Figure 4.14 shows the estimated ACF of normalised residuals and the result of a new possible jump detection. Again estimated ACF has decreased. The jump detection plot gives rise for concern. The plot shows a reference variance reduction of almost 40 percent inserting a new jump parameter at time 2003.0856. In an analysis of real data, this may be one of the pitfalls. Making the mistake to model such erroneous jump will probably introduce errors in ILSS.

In order to investigate the influence on LSSA, the data were (incorrectly) adjusted with the effect of the estimated jump parameters $\hat{\beta}_2 = 44.6$ mm and $\hat{\beta}_3 = -23.9$ mm found using Eq. (4.47). The resulting estimated ILSS is shown on the left-hand side of Figure 4.15 (**sim24**). The distortion for long periods was somehow reduced compared to the estimated spectrum **sim02** shown in Figure 4.4. The differences between the incorrect estimated jump parameters in $\hat{\boldsymbol{\beta}}$ and their real values cannot be found in a real situation. To compare this effect in the spectrum an ILSS was computed using true values $\beta_2 = 40$ mm and $\beta_3 = -30$ mm (**sim25**). The results are shown on the right-hand side of Figure 4.15 and as it

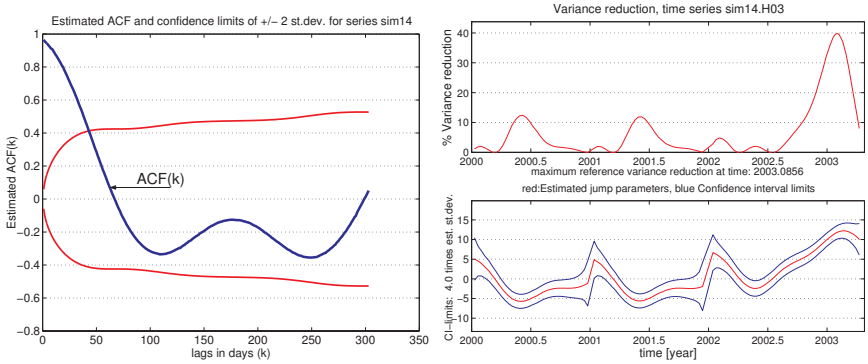

Figure 4.14: Estimated ACF of normalised residuals and results of jump detection for simulation **Sim14**.

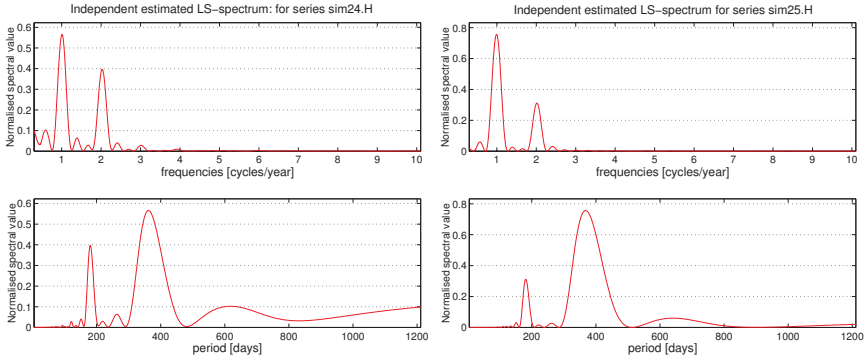

Figure 4.15: ILSS for series corrected for jump estimates: **sim24** (left) and real jump values **sim25** (right).

was expected, these results are then comparable to the situations from **sim03** and **sim04** with spectra shown in Figure 4.5.

sim15

To show the further behaviour of the estimated ACF, a model containing parameters as in **sim14** plus parameters for only an annual period were fitted to the data. The results are shown in Figure 4.16.

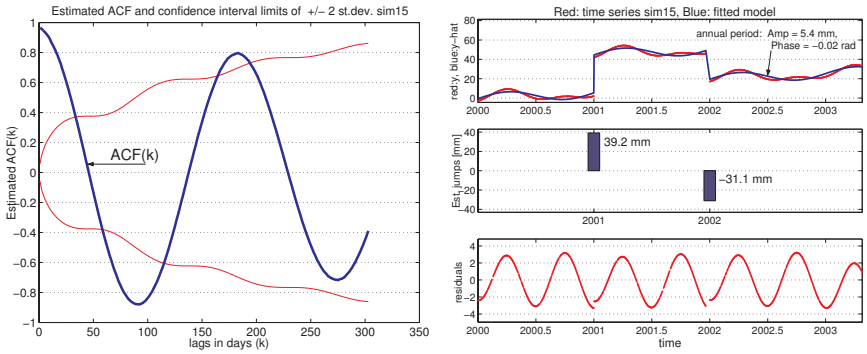

Figure 4.16: Estimated ACF of normalised residuals and fitted model for simulation **sim15**.

sim16

The result of LSSA of the series **sim14** showed two significant periods, so the final chosen deterministic model includes parameters for:

- constant term
- linear trend
- jumps at times $t_{\beta_2} = 2001.0000$ and $t_{\beta_3} = 2002.0000$
- annual and semi-annual frequency components

The results are shown in Figure 4.17. All parameters were estimated exactly equal to true values. The remaining residuals are all very close to zero. The estimated ACF are almost, but not exactly, zero for all lags $k \neq 0$. Small estimated ACF are probably caused by some numerical instabilities when all residuals are close to zero.

sim17

A last simulation has been performed as in the case **sim16**, but with noise from a $N(0, 4)$-distribution added to the synthetic series as shown in Figure 4.3. The results are shown in Figure 4.18. Estimated parameters largely agree with the true values.

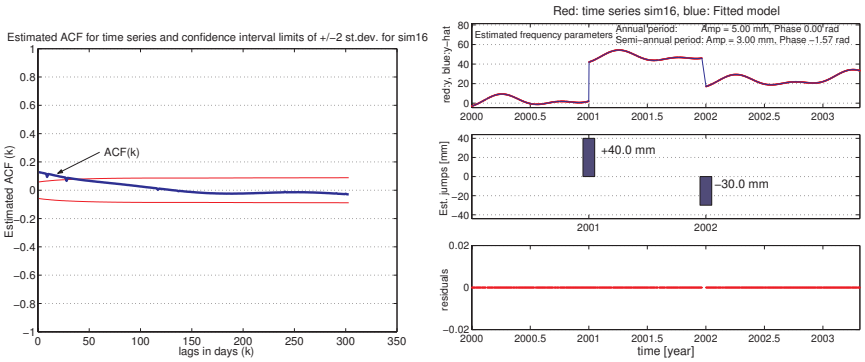

Figure 4.17: Estimated ACF of normalised residuals and fitted model for simulation **sim16**. (Noise free time series)

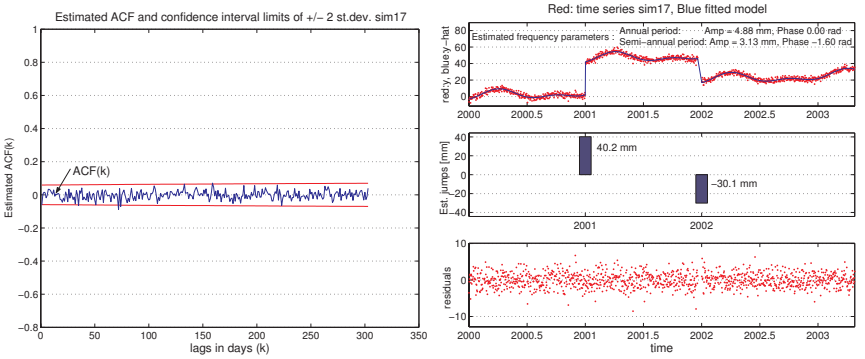

Figure 4.18: Estimated ACF of normalised residuals and fitted model for simulation **sim17**. (Time series with noise)

Summary

- Using an incomplete model generates rather large ACF estimates.
- The estimated ACF decreases successively through stepwise model improvement.
- Residual diagnostics such as tests for goodness-of-fit or (linear) dependencies cannot be done before the best possible model is found.

Chapter 5

Principal component analysis

Principal Components Analysis (PCA) is often mentioned as one specific method in the more general theory of Factor Analysis. In PCA, the data is not restricted to be normally distributed, which is a great advantage compared to other methods. PCA is commonly known as Empirical Orthogonal Function (EOF) analysis in oceanography [Emery and Thomson, 2001]. The main purpose of PCA is often to consider a data reduction. In an expanded sense, the idea can be used to determine and investigate influences from a few extracted common modes or factors that contribute most to the variance of the time series. Understanding the relations between the original variables and these factors may give valuable contributions to model improvements. A transformation from the time series to their frequency representations was focused on in spectral analysis. For the multivariate PCA analysis of CGPS data another linear transformation is sought. That is a transformation in the multi-dimensional orthogonal space, to determine the directions of maximum variance. The coefficients expressing these linear relations or rotations, are given by another set of orthogonal base functions than those in the Fourier transform. These are the eigenvector elements found from a spectral decomposition of the sample covariance matrix of the original series. Because the covariance matrix is computed from the sample, the eigenvector elements are said to be found empirically. They are commonly denoted as Empirical Orthogonal Functions, and the method is termed EOF analysis. The time series representation related to the new set of (principal) axes is denoted the time series of amplitudes. From a statistical point of view these series represent the principal components or

modes. A complete reconstruction from the amplitudes to the original time series is possible with the inverse transformation.

Suppose that we have a set of p-dimensional stochastic variables in $\boldsymbol{X}$. Typically the stochastic variables will be represented by a p-dimensional sample of n observations, e.g. as a p-dimensional time series of length n. Thus parameters such as means, variances and covariances can be estimated from the values of the observations. Assume now that this sample contains some common variation representing one or more common modes. Information about such common modes will be hidden in the covariance or correlation structure of the p variables. The idea behind PCA is to express a new set of p stochastic variables $\boldsymbol{Y}$ or Principal Components, that expresses possible modes as uncorrelated linear combinations of the original variables $\boldsymbol{X}$. As in Fourier Analysis, the linear combinations expressing the relation between $\boldsymbol{X}$ and $\boldsymbol{Y}$, will be represented by a set of orthogonal basis functions. However, the j-th orthogonal function searched in PCA will be the one that maximises the variance of Y_j. These orthogonal functions will be collected in a square orthogonal matrix and sorted after the magnitudes of their variances in descending order. This matrix can be interpreted as an operator or a orthogonal transformation matrix that represents the transformation from a p-dimensional X-space to a p-dimensional Y-space. It will be shown that the orthogonal matrix contains the eigenvectors $\boldsymbol{e_j}$ of $Cov(\boldsymbol{X})$ and their corresponding eigenvalues λ_j are the diagonal elements in $Cov(\boldsymbol{Y})$. Considering time series representing the stochastic variables in $\boldsymbol{X}$ and $\boldsymbol{Y}$, the PCA will be characterised by three quantities. First, amplitude time series $\boldsymbol{y_j}$ describing the temporal variation of the j-th mode. Second, eigenvector vectors $\boldsymbol{e_j}$ as orthogonal base functions describing the contribution of this mode to each individual time series $\boldsymbol{x_j}$ (and for the data used in this study, describing the geographical pattern of the mode). And third, eigenvalues λ_j quantifying the contribution of this mode to the overall variance of the data set.

For the moment the distribution of the underlying stochastic variables will be considered, rather than the particular observations.

5.1 Theory

Let $\boldsymbol{X} = [X_1, X_2, ..., X_j, ...X_p]^T$ be a p-dimensional vector of stochastic variables with expectation vector $E(\boldsymbol{X}) = \boldsymbol{\mu} = \mathbf{0}$

Let the reference variance be $\sigma^2 = 1$ and $Cov(\boldsymbol{X}) = \boldsymbol{\Sigma}$ be the covariance matrix of $\boldsymbol{X}$:

$$\underset{(p\times p)}{\boldsymbol{\Sigma}} = \begin{bmatrix} Var(X_1) & Cov(X_1X_2) & \dots & Cov(X_1X_j) & \dots & Cov(X_1X_p) \\ Cov(X_2X_1) & Var(X_2) & \dots & Cov(X_2X_j) & \dots & Cov(X_2X_p) \\ \vdots & \vdots & \ddots & & & \vdots \\ Cov(X_jX_1) & Cov(X_jX_2) & & Var(X_j) & & Cov(X_jX_p) \\ \vdots & \vdots & & & \ddots & \vdots \\ Cov(X_pX_1) & Cov(X_pX_2) & \dots & Cov(X_pX_j) & \dots & Var(X_p) \end{bmatrix} \tag{5.1}$$

Consider linear relations, given by the coefficients in a square matrix

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2p} \\ \vdots & \vdots & & \vdots & \dots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pj} & \dots & a_{pp} \end{bmatrix} = [\boldsymbol{a_1} \vdots \boldsymbol{a_2} \vdots \dots \vdots \boldsymbol{a_j} \vdots \dots \vdots \boldsymbol{a_p}] \tag{5.2}$$

between the $\boldsymbol{X}$-vector and another stochastic vector $\boldsymbol{Y} = [Y_1, Y_2, ..., Y_j, ...Y_p]^T$:

$$\underset{(p\times 1)}{\boldsymbol{Y}} = \underset{(p\times p)(p\times 1)}{\boldsymbol{A^T}\ \boldsymbol{X}} \begin{bmatrix} \boldsymbol{a_1^T} \\ \hline \boldsymbol{a_2^T} \\ \hline \vdots \\ \hline \boldsymbol{a_j^T} \\ \hline \vdots \\ \hline \boldsymbol{a_p^T} \end{bmatrix} \cdot \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_j \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{a_1^T X} \\ \hline \boldsymbol{a_2^T X} \\ \hline \vdots \\ \hline \boldsymbol{a_j^T X} \\ \hline \vdots \\ \hline \boldsymbol{a_p^T X} \end{bmatrix} \tag{5.3}$$

with e.g. : $\boldsymbol{a_1^T X} = [a_{11}, a_{21}, ..., a_{p1}]\boldsymbol{X} = \sum_{j=1}^{p} a_{j1}X_j$.

Applying the LEP, from Appendix A tp Eq. (5.3), the elements in the covariance matrix of $\boldsymbol{Y}$ is found to be:

$$Var(Y_j) = \boldsymbol{a_j^T \Sigma a_j} \qquad j = 1, 2, ..., p \tag{5.4}$$

$$Cov(Y_j, Y_k) = \boldsymbol{a_j^T \Sigma a_k} \quad j, k = 1, 2, ..., p \qquad j \neq k \tag{5.5}$$

or

$$\underset{(p\times p)}{Cov(\boldsymbol{Y})} = \boldsymbol{A^T \Sigma A} = \begin{bmatrix} \boldsymbol{a_1^T \Sigma a_1} & \boldsymbol{a_1^T \Sigma a_2} & \dots & \boldsymbol{a_1^T \Sigma a_j} & \dots & \boldsymbol{a_1^T \Sigma a_p} \\ \boldsymbol{a_2^T \Sigma a_1} & \boldsymbol{a_2^T \Sigma a_2} & \dots & \boldsymbol{a_2^T \Sigma a_j} & \dots & \boldsymbol{a_2^T \Sigma a_p} \\ \vdots & \vdots & \ddots & & & \vdots \\ \boldsymbol{a_j^T \Sigma a_1} & \boldsymbol{a_j^T \Sigma a_2} & & \boldsymbol{a_j^T \Sigma a_j} & & \boldsymbol{a_j^T \Sigma a_p} \\ \vdots & \vdots & & & \ddots & \vdots \\ \boldsymbol{a_p^T \Sigma a_1} & \boldsymbol{a_p^T \Sigma a_2} & \dots & \boldsymbol{a_p^T \Sigma a_j} & \dots & \boldsymbol{a_p^T \Sigma a_p} \end{bmatrix} \tag{5.6}$$

The uncorrelated linear combinations of $Y_j \quad ; j = 1, 2, .., p$ are searched, whose variances are as large as possible. The variance in Eq. (5.4) can be increased by multiplying any $\boldsymbol{a_j}$ by some constant, so the maximum has to be reached under some constraints to the $\boldsymbol{a_j}$-vectors. Bounding the scalar $\boldsymbol{a_j^T a_j}$ to be one is equivalent to claim the vectors $\boldsymbol{a_j}(j = 1, 2, ..., p)$ to be of unit length.

5.1.1 Definition of principal components

The first principal component is defined as the linear combination $PC(1) = Y_1 = \boldsymbol{a_1^T X}$ that maximises $Var(Y_1)$ subject to $\boldsymbol{a_1^T a_1} = 1$. This constrained maximisation problem could be solved using Lagrange multipliers, considering the main function:

$$m(\boldsymbol{a_1}, \lambda) = \boldsymbol{a_1^T \Sigma a_1} - \lambda(\boldsymbol{a_1^T a_1} - 1) \tag{5.7}$$

and use matrix derivation:

$$\frac{\partial m}{\partial \boldsymbol{a_1}} = 2(\boldsymbol{\Sigma} - \lambda \boldsymbol{I})\boldsymbol{a_1} \tag{5.8}$$

and set this expression equal to the zero vector one gets:

$$(\boldsymbol{\Sigma} - \lambda \boldsymbol{I})\boldsymbol{a_1} = \boldsymbol{\Sigma a_1} - \lambda \boldsymbol{a_1} = \boldsymbol{0} \tag{5.9}$$

This expression frequently turns up in the matrix theory and states that:
$\boldsymbol{a_1}$ is an eigenvector of $\boldsymbol{\Sigma}$ corresponding to to the eigenvalue λ

The second principal component is the linear combination $PC(2) = Y_2 = \boldsymbol{a_2^T X}$ that maximises $Var(Y_2)$ subject to the two restrictions, $\boldsymbol{a_2^T a_2} = 1$ and $Cov(Y_1, Y_2) = \boldsymbol{a_1^T \Sigma a_2} = 0$, and so on. Thus all inner products are restricted to:

$$\boldsymbol{a_j^T a_j} = 1 \qquad j = 1, 2, ..., p \tag{5.10}$$

$$\boldsymbol{a_j^T a_k} = 0 \quad j, k = 1, 2, ..., p \qquad j \neq k \tag{5.11}$$

which ensure all column vectors of $\boldsymbol{A}$ to be orthogonal and even also orthonormal, and thus $\boldsymbol{A}$ becomes an orthogonal matrix. Orthogonal matrices have some special useful properties [see Kreyszig, 1999, among others]. If $\boldsymbol{A}$ is orthogonal, then $\boldsymbol{A^T} = \boldsymbol{A^{-1}}$ so

$$\boldsymbol{A A^T} = \boldsymbol{I} = \boldsymbol{A^T A} \tag{5.12}$$

which means that both the row vectors and column vectors in $\boldsymbol{A}$ are orthogonal and thus $\boldsymbol{A^T}$ is also an orthogonal matrix.

The $\boldsymbol{A^T}$-matrix may be interpreted as an operator that transforms points given in an p-dimensional orthogonal coordinate system X to another p-dimensional orthogonal coordinate system Y. The relations between directions in the two orthogonal systems X and Y are given by the elements in the transformation matrix $\boldsymbol{A^T}$. An orthogonal transformation does not include shifts or scale changes, these relations include rotations only. The characteristics of PCA is the choice of Y-axis in directions with maximum variability to provide a simpler description of the covariance structure.

5.1.2 Spectral decomposition

The covariance matrix

Let $\lambda_1 > \lambda_2 > \ldots > \lambda_j > \ldots > \lambda_p > 0$ be the eigenvalues of descending order of the positive definite symmetric quadratic covariance matrix $\boldsymbol{\Sigma}$, and $\boldsymbol{e_1, e_2, \ldots, e_j, \ldots, e_p}$ their corresponding normalised eigenvectors. Then it can be shown that [see Johnson and Wichern, 1998, p.81-84]:

$$\max_{\boldsymbol{a_j} \neq \mathbf{0}} \left(\frac{\boldsymbol{a_j^T \Sigma a_j}}{\boldsymbol{a_j^T a_j}} \right) = \lambda_j \qquad \text{when} \quad \boldsymbol{a_j = e_j} \tag{5.13}$$

so then we find for $j = 1, 2, ..., p$:

$$\begin{aligned} Var(Y_j) \quad &= \max_{\boldsymbol{a_j} \neq \mathbf{0}} \left(\frac{\boldsymbol{a_j^T \Sigma a_j}}{\boldsymbol{a_j^T a_j}} \right) = \lambda_j \qquad \text{and} \\ Y_j = PC(j) \quad &= \boldsymbol{e_j^T X} \end{aligned} \tag{5.14}$$

Collecting the eigenvalues of descending order in a square diagonal matrix:

$$Cov(\boldsymbol{Y}) = \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{bmatrix} \tag{5.15}$$

and their corresponding normalised eigenvectors in a matrix:
$\boldsymbol{E} = [\boldsymbol{e_1} \vdots \boldsymbol{e_2} \vdots \dots \vdots \boldsymbol{e_j} \vdots \dots \vdots \boldsymbol{e_p}]$, and then using the LEP from Appendix A on $\boldsymbol{Y = E^T X}$ one gets:

$$\boldsymbol{\Lambda = E^T \Sigma E} \tag{5.16}$$

Because both the column and the row vectors in $\boldsymbol{E}$ are orthogonal, and of unit length we have: $\boldsymbol{EE^T = E^T E = I}$ and thus the $\boldsymbol{E}$-matrix is orthogonal and

$$\boldsymbol{\Sigma = E \Lambda E^T} = \sum_{j=1}^{p} \lambda_j \boldsymbol{e_j e_j^T} \tag{5.17}$$

The last terms are also known as the spectral decomposition of $\boldsymbol{\Sigma}$ [see Johnson and Wichern, 1998, p. 104].

The total population variance is:

$$\sum_{j=1}^{p} Var(X_j) = tr(\boldsymbol{\Sigma}) = tr(\boldsymbol{E\Lambda E^T}) = \\ tr(\boldsymbol{\Lambda E^T E}) = tr(\boldsymbol{\Lambda}) = \sum_{j=1}^{p} Var(Y_j) \tag{5.18}$$

The part of the total variance explained by the j-th principal component is:

$$\frac{\lambda_j}{\lambda_1 + \lambda_2 + ... + \lambda_j + ... + \lambda_p} = \frac{\lambda_j}{\sum_{j=1}^{p} \lambda_j} \qquad j = 1, 2, ..., p \tag{5.19}$$

The accumulated value of the first few λ's, is of great interest, considering a variable reduction. If most of the total population variance can be attributed to a few, let us say q components, then these components can replace the original p variables without much loss of information. The elements in the eigenvectors $\boldsymbol{e_j} = [e_{1j}, e_{2j}, ..., e_{kj}, ..., e_{pj}]^{\boldsymbol{T}}$ also have an interpretation. The magnitude of e_{kj} measures the importance of the k-th variable to the j-th principal component irrespective of the other variables.

The inverse covariance matrix

If the inverse covariance matrix $\boldsymbol{\Sigma^{-1}}$ is available, it is not necessary to make an inversion to find the eigenvalues and eigenvectors. For a positive definite $\boldsymbol{\Sigma}$ and an eigenvector $\boldsymbol{e} \neq \boldsymbol{0}$ one obtains:

$$\boldsymbol{\Sigma e} = \lambda \boldsymbol{e} \quad \Longrightarrow \quad \boldsymbol{\Sigma^{-1} e} = \frac{1}{\lambda} \boldsymbol{e} \tag{5.20}$$

So the eigenvalue-eigenvector pair $(\lambda, \boldsymbol{e})$ for $\boldsymbol{\Sigma}$ corresponds to the eigenvalue-eigenvector pair $(\frac{1}{\lambda}, \boldsymbol{e})$ for $\boldsymbol{\Sigma^{-1}}$ and thus the spectral decomposition of $\boldsymbol{\Sigma^{-1}}$ is:

$$\boldsymbol{\Sigma^{-1}} = \boldsymbol{E\Lambda^{-1}E^T} = \sum_{j=1}^{p} \frac{1}{\lambda_j} \boldsymbol{e_j e_j^T} \tag{5.21}$$

A proof is given in Johnson and Wichern [1998].

5.1.3 Loadings

To visualise the contribution from common factors, the elements in the normalised eigenvectors corresponding to the largest eigenvalues can be investigated. Alternatively another version, scaled with the root of the corresponding eigenvalues is often used. They are denoted factor loadings, component loadings or only loadings and is defined as:

$$\underset{(p\cdot p)}{\boldsymbol{L}} = \left[\sqrt{\lambda_1}\boldsymbol{e_1}\vdots\sqrt{\lambda_2}\boldsymbol{e_2}\vdots\ldots\vdots\sqrt{\lambda_q}\boldsymbol{e_q}\vdots\ldots\vdots\sqrt{\lambda_p}\boldsymbol{e_p}\right] \quad \Leftrightarrow \quad \boldsymbol{LL^T} = \boldsymbol{\Sigma} \tag{5.22}$$

The loadings reflect the contribution from the original variables on the PC's. Accepting a small loss of information, it is then possible to explain the covariance structure of the original p variables in terms of just a few $(q < p)$ new variables or common factors. If the last $(p - q)$ eigenvalues are small, the contributions from the $(p - q)$ last columns of $\boldsymbol{L}$, can be neglected and the covariance matrix can be approximated with $\underset{(p\cdot q)(q\cdot p)}{\boldsymbol{L}\ \boldsymbol{L^T}}$.

5.1.4 Statistical modes

For a choice of $q = p$ statistical modes in $\boldsymbol{Y}$ one gets:

$$\underset{(p\cdot 1)}{\boldsymbol{Y}} = \underset{(p\cdot p)(p\cdot 1)}{\boldsymbol{E^T}\ \boldsymbol{X}} \tag{5.23}$$

and because $\boldsymbol{E}$ is an orthogonal matrix, a complete reconstruction of the original variables $\boldsymbol{X}$ from the statistical modes is possible with:

$$\underset{(p\cdot 1)}{\boldsymbol{X}} = \underset{(p\cdot p)(p\cdot 1)}{\boldsymbol{E}\ \boldsymbol{Y}} \tag{5.24}$$

Extracting only a few $q < p$ of the modes can easily be done, using only the eigenvectors corresponding to the q largest eigenvalues:

$$\underset{(q\cdot 1)}{\boldsymbol{Y}^{(q)}} = \underset{(q\cdot p)(p\cdot 1)}{\boldsymbol{E^T}\ \boldsymbol{X}} \tag{5.25}$$

The effect on the original variables, only from the first q selected modes, interpreted as a common mode signal can be found as:

$$\underset{(p\cdot 1)}{\boldsymbol{X}^{(q)}} = \underset{(p\cdot q)}{\boldsymbol{E}}\ \underset{(q\cdot 1)}{\boldsymbol{Y}^{(q)}} \tag{5.26}$$

5.2 Data application and interpretation

The principle of PCA can be explained analogously to Fourier Analysis. In Fourier Analysis, a finite length time series can be completely reproduced using a linear summation of sines and cosines as in Eq. (2.21). Because the principle of Fourier Analysis is to determine the contribution of periodic components in the data, different frequency components expressed in $\sin$ and $\cos$ terms are fitted to the data. To ensure the basis functions $sin(\omega_k t)$ and $cos(\omega_k t)$ becoming orthogonal, only equidistant angular frequencies (Fourier frequencies) are chosen. This makes it possible to solve each Fourier coefficient independent of the others. The amplitudes computed from the Fourier coefficients measures the frequency contribution in the data. Using matrix notation, the basis functions are commonly used in the columns of a coefficient matrix as in Eq. (4.11). In PCA, the frequency content is not focused. Another set of orthogonal basis functions, the set that can explain the largest variance contribution is searched. The eigenvectors in $\boldsymbol{E}$ represents such orthogonal functions.

Now data series will be introduced as samples for the stochastic variables in Eq. (5.23). Assume a data matrix $\underset{(p\cdot n)}{\tilde{\boldsymbol{X}}}$ with p rows of observation series, each of length n. To satisfy the model assumptions $E(\boldsymbol{X}) = \boldsymbol{0}$, the $\tilde{\boldsymbol{X}}$-matrix now represents, the data after their row-means have been subtracted, so that the row totals of $\tilde{\boldsymbol{X}}$ are all zero. In practise $Cov(\boldsymbol{X})$ is often unknown. It is common to estimate $Cov(\boldsymbol{X})$ by the sample covariance matrix $\boldsymbol{S}$ as:

$$\boldsymbol{S} = \frac{\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^{\boldsymbol{T}}}{n} \qquad \text{or} \qquad \boldsymbol{S}_{unbiased} = \frac{\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^{\boldsymbol{T}}}{n-1} \tag{5.27}$$

in which case one finds the estimates:

$$\hat{\lambda}_j \quad \text{of} \quad \lambda_j \qquad \text{and} \qquad \hat{\boldsymbol{e}}_{\boldsymbol{j}} \quad \text{of} \quad \boldsymbol{e}_{\boldsymbol{j}} \tag{5.28}$$

to be placed in $\hat{\boldsymbol{\Lambda}}$ and $\hat{\boldsymbol{E}}$. The orthogonal functions in the columns of $\hat{\boldsymbol{E}}$ are found an empirical way, thus they are denoted Empirical Orthogonal Functions, which explains why PCA also is also known as EOF analysis. With the relation:

$$\underset{(p\cdot n)}{\tilde{\boldsymbol{Y}}} = \underset{(p\cdot p)(p\cdot n)}{\hat{\boldsymbol{E}}^{\boldsymbol{T}}\,\tilde{\boldsymbol{X}}} \tag{5.29}$$

the p series of length n in $\underset{(p\cdot n)}{\tilde{\boldsymbol{Y}}}$ can be found, which also has an interpretation. Analogous to the theory of Fourier Analysis, they are often denoted the series of amplitudes representing each statistical mode. A complete reconstruction of the p observation series is possible when using all p linear combinations of the amplitude series.

$$\underset{(p\cdot n)}{\tilde{\boldsymbol{X}}} = \underset{(p\cdot p)(p\cdot n)}{\hat{\boldsymbol{E}}\,\tilde{\boldsymbol{Y}}} \tag{5.30}$$

5.3 Model validation

Assume the common mode signals from two separately performed PCAs, recomputed to effects influencing each of the original variables. Consider these effects represented in two data matrices, $\tilde{\boldsymbol{X}}^{(q)}$ and $\tilde{\boldsymbol{Z}}^{(q)}$ of the form:

$$\underset{(p\cdot n)}{\tilde{\boldsymbol{X}}}^{(q)} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2k} & \dots & x_{2n} \\ \vdots & \vdots & & \vdots & \dots & \vdots \\ x_{j1} & x_{j2} & \dots & x_{jk} & \dots & x_{jn} \\ \vdots & \vdots & & \vdots & \dots & \vdots \\ x_{p1} & x_{p2} & \dots & x_{pk} & \dots & x_{pn} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x_1} \\ \boldsymbol{x_2} \\ \vdots \\ \boldsymbol{x_j} \\ \vdots \\ \boldsymbol{x_p} \end{bmatrix} \tag{5.31}$$

In order to compare the effects from different PCAs, regression coefficients γ_j can be estimated for the corresponding p series of n observations in the rows of each data matrix. The matrix form of least squares models, handling the row vectors

$\boldsymbol{x_j}$ in $\boldsymbol{X}^{(q)}$ as observations and the row vectors $\boldsymbol{z_j}$ in $\boldsymbol{Z}^{(q)}$ as predictors is:

$$\boldsymbol{x_j} = \boldsymbol{z_j}\gamma_j + \boldsymbol{\varepsilon_j} \qquad \text{for} \quad j = 1, 2, \ldots, p \tag{5.32}$$

$\boldsymbol{\varepsilon_j}$ are vectors of independent identical distributed errors. Minimising the sum of squared errors leads to the least squares estimators for the scalar γ_j:

$$\hat{\gamma}_j = (\boldsymbol{z_j^T z_j})^{-1} \boldsymbol{z_j^T x_j} \tag{5.33}$$

Another factor for a comparison is the correlation coefficients $\hat{\rho}_j$ computed from the series $\boldsymbol{x_j}$ and $\boldsymbol{z_j}$. However, for zero mean time series in $\boldsymbol{z_j}$ and $\boldsymbol{x_j}$, $\hat{\rho}_j$ and $\hat{\gamma}_j$ are related through:

$$\hat{\rho}_j = \hat{\gamma}_j \sqrt{\frac{s^2_{\boldsymbol{z_j}}}{s^2_{\boldsymbol{x_j}}}} \tag{5.34}$$

where $s^2_{\boldsymbol{z_j}}$ and $s^2_{\boldsymbol{x_j}}$ are the sample variances from the series $\boldsymbol{z_j}$ and $\boldsymbol{x_j}$ respectively.

5.4 Normalised data

Sometimes data series appear to be of different length, variability, or even of different units. Variables of great variability may dominate a PCA. Normalising the data to have unit variances before the analysis implies that the variables are to be of equally importance or gives them equal weights. For normalised data, the covariance matrix $\boldsymbol{S}$ then becomes identical with the correlation matrix $\boldsymbol{R}$. However, it is important to realise that the eigenvalues and eigenvectors of $\boldsymbol{R}$ is generally not the same as those of $\boldsymbol{S}$.

Chapter 6

Numerical investigations

6.1 The Norwegian CGPS network

Several CGPS data sets for the numerical investigations analysed in this study are obtained from NMA. A preliminary data description is given in Subsection 1.4.1. The data anaysis, in the first step, was a very time-consuming work [See also Haakonsen and Nahavandchi, 2003]. Based on the different analysis and tests it was decided to choose the third Norwegian data set, which is assumed to be the best one, for further analysis. These are series from 14 stations from the Norwegian network of permanent GPS stations, "SATelittbasert REFeransesystem" (SATREF) of fairly various lengths, and include several intervals of missing data. To avoid confusion, the new series were renamed. The new and the official station abbreviation (four letters) used on the Norwegian CGPS data are listed in Table 6.1 together with the complete station names. A map with locations of the sites in this data set is shown in Figure 6.1. In the following subsections the different steps of model improvement will be performed for univariate series as illustrated in Figure 1.1 see also [Haakonsen and Nahavandchi, 2004]. The time series remaining after outlier detections to be used for final model fits are shown in the upper plots of Figures 6.15 to 6.35

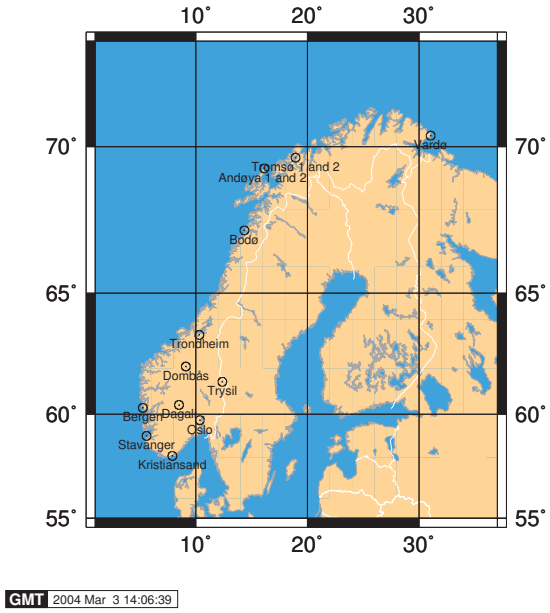

Figure 6.1: Map with locations of the Norwegian sites

6.1.1 Jump detection

The data series to be analysed revealed some rather large unexpected jumps (offsets or shifts). Such jumps can be explained by incidents, known or unknown in time and quantity. Coordinate offsets may be introduced by all changes in the antenna configuration or the up-grading of the receiver software [see Scherneck et al., 1998] and for some examples due to changes in the radome. The selection of the correct number of, and time for possible offsets in the data series was shown to be of vital importance, especially for the later spectral analysis. Some of them are introduced because of known incidents at the specific sites. Others are inexplicable, but were detected by the use of the jump detection procedure described in Section 3.3. The jump detection procedure was executed for each series isolated. There is reason to believe that if a jump appears in one direction, jumps are also expected at the same time for the other two directions for a

Table 6.1: Table of station names, official and renamed abbreviations of the Norwegian CGPS data

Site	Official	New abbr.
Andøya 1	ande	an3e
Andøya 2	ando	an3o
Bergen	berg	ber3
Bodø	bodo	bod3
Dagali	dags	dag3
Dombås	doms	dom3
Kristiansand	kris	kri3
Oslo	oslo	osl3
Stavanger	stav	sta3
Trondheim	tron	trh3
Tromsø 2	trom	trm3
Tromsø 1	tro1	tro3
Trysil	trys	try3
Vardø	vard	var3

station. For all computations in this study, identical station-offset files of jump times are used for all three directions of a station. The method has shown some weaknesses. Sometimes a second jump with the opposite sign is detected close to earlier found jumps. A visual control of the series has turned out to be necessary. In the case of small correlations between estimates of jumps and other parameters, the estimated variance for each jump value gives a preliminary indication of their importance. In a later phase, after the final deterministic models have been chosen, time series of normalised residuals can be correlated with the neighbouring stations, trying to reveal any missed offsets. For several series, an offset around time: $t_\beta = 2000.0000$ was found. A satisfying explanation of this jump time has not been given. Another small offset around time: $t_\beta = 2000.7338$ was located for some series, but because of the risk of mistaken it for an annual change from summer to winter, it was considered to be omitted in the further analysis, . The final selected jump times for the Norwegian CGPS data set are listed in Table 6.2.

Table 6.2: Table of offset times t_β for the Norwegian CGPS data set

Data series	t_{β_2}	t_{β_3}	t_{β_4}
an3o	2000.3285		
ber3	1998.7516	2000.0000	2000.2985
bod3	2000.0000		
kri3	1998.6300	1999.0938	2000.0000
osl3	2001.0184	2001.0706	
sta3	1998.3326	2000.0000	
trh3	1998.4723	1998.7214	2000.0000
trm3	1998.8884	2000.0000	
var3	1998.4586	2000.0000	

A jump detection example

A graphical example of the jump detection procedure described in Section 3.3 is demonstrated in Figure 6.2. Possible jumps in the time series **ales.h** (Series representing the vertical component from station Ålesund) from an earlier data set are investigated. The example shows that estimating a possible new jump in the data series at time $t_i = 1997.7605$ will reduce the reference variance more than 60 percent (upper plot) and its parameter value will be almost 50 mm (lower plot). The blue confidence limits around the estimated jump value in the lower plot is computed as four times the estimated standard deviation for the jump parameter.

6.1.2 Independent least squares spectrum estimation

Determination of the ILSS, may be seen as an early phase of model improvement. In that context it is desirable to filter out a minimum of observations. The k-factor used in the outlier detection routine was set to a large value $(k = 7)$, for the estimation of primary ILSS for the time series of the Norwegian CGPS data, Too strong filtering may impact on the frequency components with large amplitudes, and some correct observations can erroneously be sorted out. For data series including jumps, two different estimations have been made. Strong correlations between parameter estimates indicate a dangerous situation of over-parameterisation that also might cause numerical problems. Simulations in Section 4.4 have shown that jump- and spectral parameters for long periods are strongly correlated. This makes

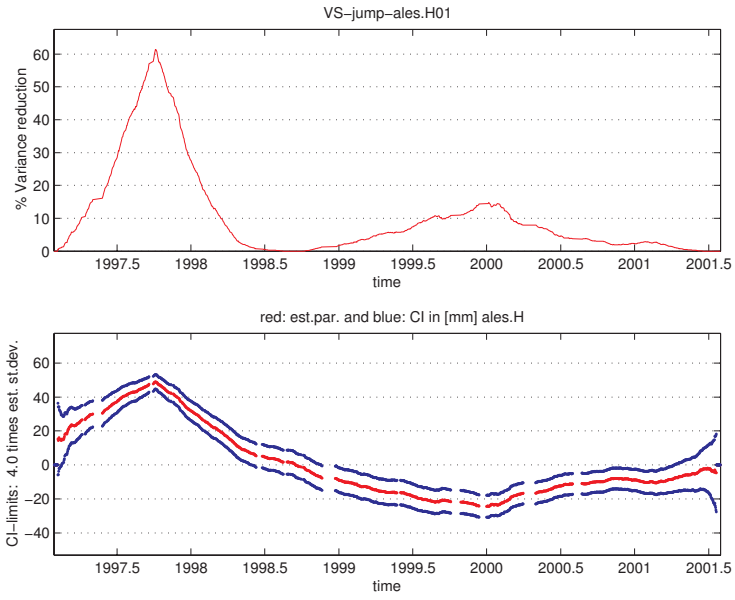

Figure 6.2: Jump detection example, time series **ales.h**.

the spectral estimates sensitive to jumps in the deterministic model as shown in the simulation part in Section 4.4. The effect of a jump may reflect the corresponding correlated spectral estimates and destroy the spectrum for long periods and vice versa. Long periodic components in the data may also be erroneously reflected to the modelled jump parameters.

For the nine stations with series including one or more jumps, two different deterministic models have been used for the estimation of the ILSS. The spectra in the left-hand side plots of Figures 6.3 to 6.11 are the results from estimations with deterministic models containing parameters for constant term, linear trend and jumps as in Eq. (3.3). The spectra in the right-hand side plots of Figures 6.3 to 6.11 are results for which jump sizes for the known jump times from Table 6.2 first were computed, and then their effect was subtracted from the original time series as illustrated in Eq. (4.47). The spectra are thus estimated from corrected series using deterministic models with parameters for constant term and

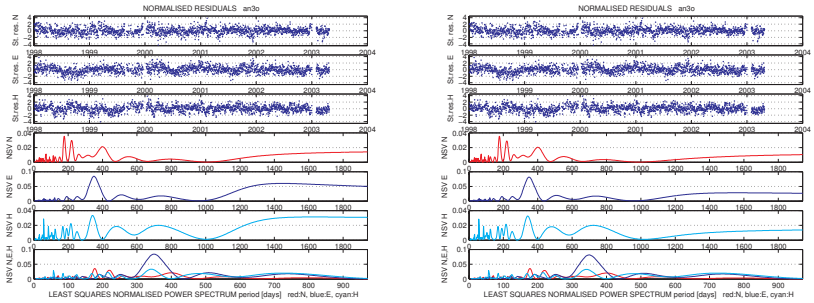

Figure 6.3: Independent Least Squares Spectrum for series: **an3o**

(left) and **an3o** corrected for the effect of jumps (right).

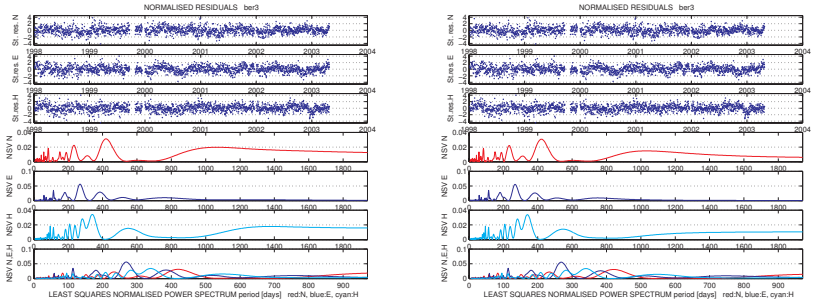

Figure 6.4: Independent Least Squares Spectrum for series: **ber3**

(left) and **ber3** corrected for the effect of jumps (right).

linear trend only. For the rest of the spectra, representing the stations assuming no jumps in data, deterministic models including parameters for constant term, linear trend and jumps are used. These spectra are shown in Figures 6.12 to 6.14

Upper three subplots in Figures 6.3 to 6.14 are the normalised residuals after adjustments. The lower four subplots are respectively LSS for the components North, East, Height separately and North, East, Height together. The first impression is that the spectra are rather different. However most of them show domi-

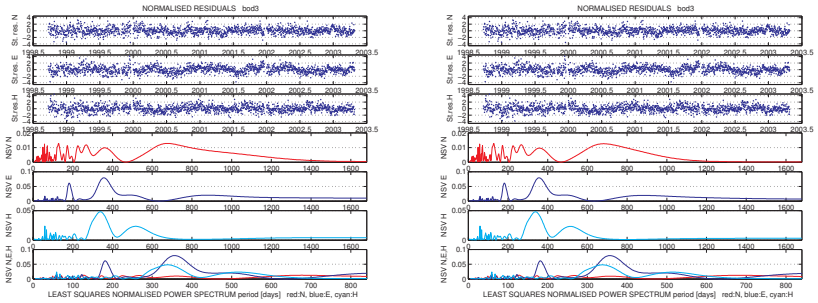

Figure 6.5: Independent Least Squares Spectrum for series: **bod3**

(left) and **bod3** corrected for the effect of jumps (right).

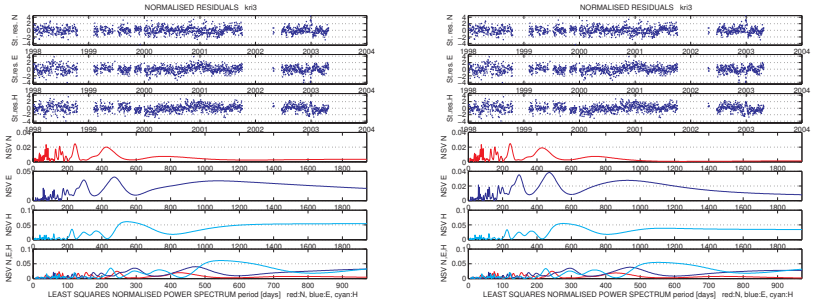

Figure 6.6: Independent Least Squares Spectrum for series: **kri3**

(left) and **kri3** corrected for the effect of jumps (right).

nating peaks more or less close to annual periods. Some spectra for the height components show periods close to the annual and semi-annual, but seem to be some days short. Further it is not easy to establish any local characteristics. Even if effects of possible jumps are removed before the estimations, some stations have dominant periods of more than one year. Except for annual and semi-annual periods it is not easy to identify common periods in all spectra. However a seasonal period of three months may be visualised in some series at high latitudes, e.g. **trm3** and **an3e**. Before a conclusion can be drawn, a more complete discussion

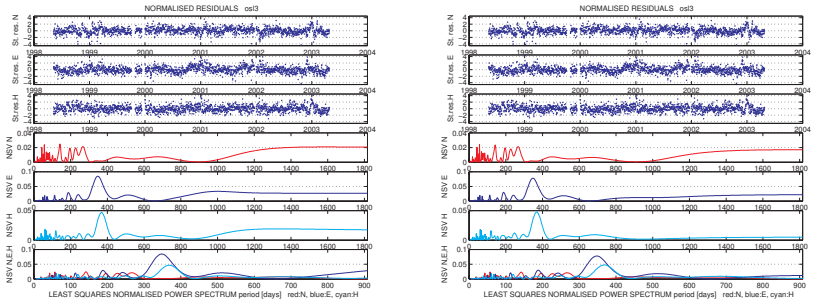

Figure 6.7: Independent Least Squares Spectrum for series: **osl3**

(left) and **osl3** corrected for the effect of jumps (right).

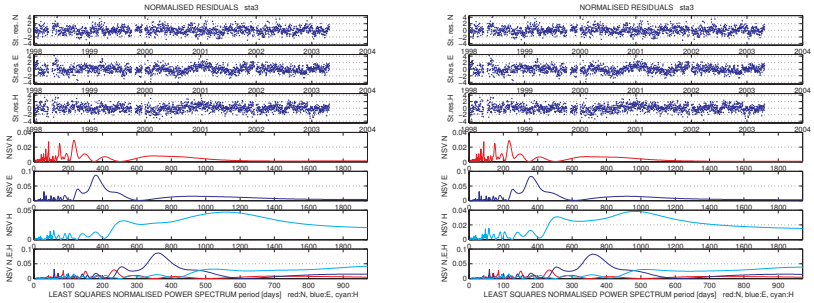

Figure 6.8: Independent Least Squares Spectrum for series: **sta3**

(left) and **sta3** corrected for the effect of jumps (right).

of these results will be carried out in the next section.

6.1.3 Final choice of univariate regression models

Using the identity from Eq. (2.18) the periodic or frequency components k can be expressed as:

$$A_k sin(2\pi f_k t + \phi_k) = \quad A_k \left[sin(2\pi f_k t) cos\phi_k + cos(2\pi f_k t) sin\phi_k \right] = \\ \beta_{a_k} cos(2\pi f_k t) + \beta_{b_k} sin(2\pi f_k t) \tag{6.1}$$

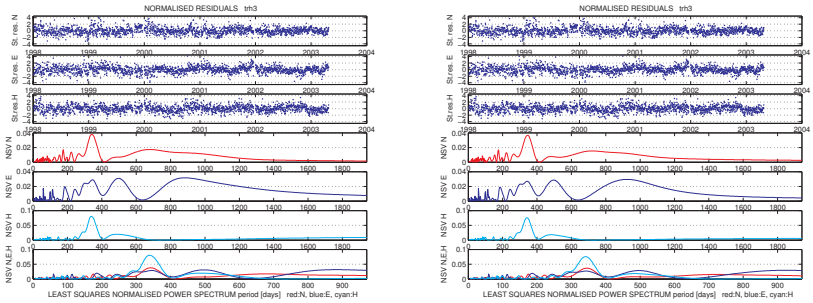

Figure 6.9: Independent Least Squares Spectrum for series: **trh3**

(left) and **trh3** corrected for the effect of jumps (right).

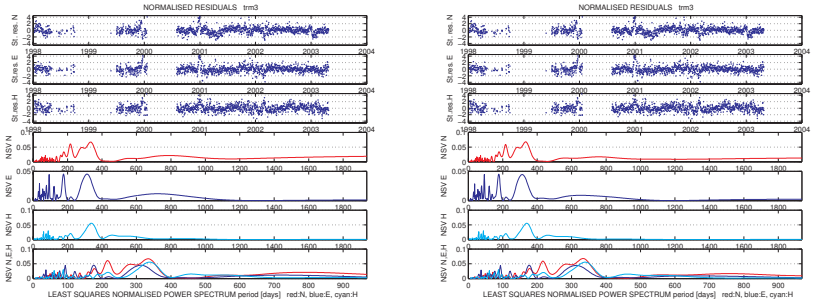

Figure 6.10: Independent Least Squares Spectrum for series: **trm3**

(left) and **trm3** corrected for the effect of jumps (right).

with the parameters $\beta_{a_k} = A_k sin\phi_k$ and $\beta_{b_k} = A_k cos\phi_k$. In addition to Eq. (3.3), a general observation y_i with the total number of q jumps and m frequency components becomes:

$$y_i = \beta_0 + \beta_1 t_i + \sum_{j=2}^{q+1} \beta_j h(t_i - t_{\beta_j}) + \sum_{k=1}^{m} [\beta_{a_k} cos(2\pi f_k t_i) + \beta_{b_k} sin(2\pi f_k t_i)] + \varepsilon_i \tag{6.2}$$

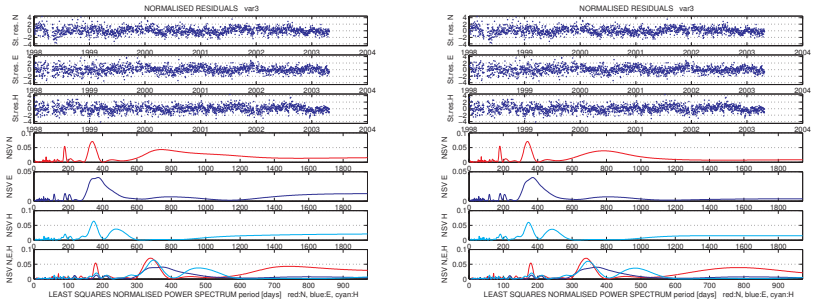

Figure 6.11: Independent Least Squares Spectrum for series: **var3**

(left) and **var3** corrected for the effect of jumps (right).

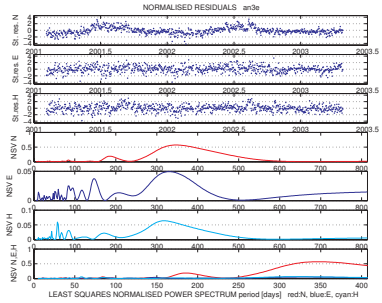

Figure 6.12: Independent Least Squares Spectrum for series: **an3e**

The choice of the periodic components may be justified from the estimated spectra as well as the expectations that at least annual and semi-annual periods must be included in the time series of CGPS residuals. The seasonal effect (3 months) might not be included in the deterministic model. However, the modelling of spectral parameters as deterministic and not stochastic parameters becomes a major advantage here. Using no a priori error properties, no constraints are made to the adjustment from these parameters. Let us assume the data do not contain a chosen frequency component, but an attempt is made to model such an effect. For very small amplitudes in data, the corresponding pair of estimated parameters $\hat{\beta}_{a_k}$ and $\hat{\beta}_{b_k}$ will both become very small ($\hat{A}_k = \sqrt{\hat{\beta}^2_{a_k} + \hat{\beta}^2_{b_k}}$). Removing this small

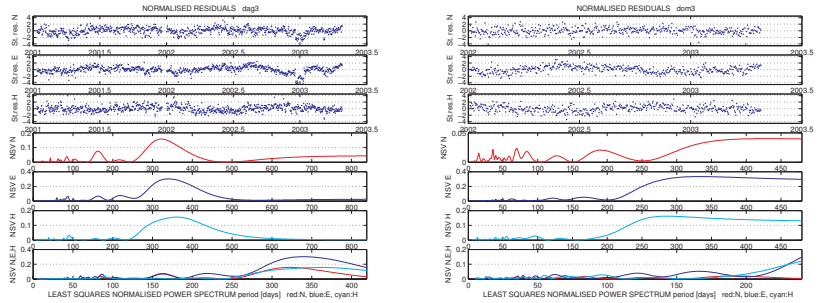

Figure 6.13: Independent Least Squares Spectrum for **dag3** and **dom3**

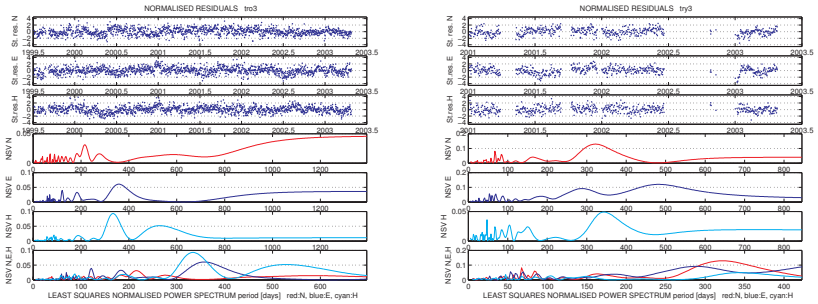

Figure 6.14: Independent Least Squares Spectrum for **tro3** and **try3**

effect from the original series will hardly make any difference to the remaining time series of residuals, that later will be used as input in a PCA for the Norwegian network.

The analysis of univariate time series of the Norwegian network chose final deterministic models including parameters for:

- β_0; constant
- β_1; linear trend
- β_2-$\beta_{(q+1)}$; The number of q offsets, and times from Table 6.2.
- β_{a_k} and β_{b_k} representing $m = 3$ periodic components of one, two and four cycles per year. $(f_1 = 1, f_2 = 2, f_3 = 4)$

to be included in a parameter vector $\boldsymbol{\beta}$ and solved in the final weighted regression matrix model:

$$\underset{(n\cdot 1)}{\boldsymbol{y}} = \underset{(n\cdot(2+q+2m))}{\boldsymbol{X}} \underset{((2+q+2m)\cdot 1)}{\boldsymbol{\beta}} + \underset{(n\cdot 1)}{\boldsymbol{\varepsilon}} \tag{6.3}$$

Results are shown in Figures 6.15 to 6.35. In the upper plots, the observations (as points) and the fitted models (curves). If jumps are estimated, they are shown as bars in the middle plots. (Software errors in a plot routine of Matlab gave confusing various width for the bars, but the interesting part, their height seems to be correct.) Lower plots show the remaining residuals after least squares model fits. After final deterministic models were chosen, the factor in the outlier-detection procedure described in Subsection 3.2.1 was set to $k = 3$.

6.1.4 Residual diagnostics for univariate time series

Goodness-of-fit tests

Adjustments with final chosen deterministic models with parameters as explained in Subsection 6.1.3 have been made. The time series of normalised residuals remaining after all known effects have been removed will be investigated. The outlier test in Subsection 3.2.1 is derived and thus valid for N(0,1)-distributed data only. To test the normality of data, the goodness-of-fit test in Subsection 3.2.2 is used. Least squares procedures, using inverse diagonal cov-matrices as in Eq. (3.11) assume the observations to be independent. To visualise the linear dependency between normalised residuals, ACFs are estimated and plotted for all series. Factor *k=3* was used in the outlier test explained in section 3.2.1. Each series of normalised residuals for each direction of the fourteen CGPS stations of the Norwegian network was used in a Goodness-of-fit test, described in subsection 3.2.2 with the Null hypothesis:

$\boldsymbol{H_0}$: Normalised residuals are observations from a N(0,1)-distribution.

At the significance level $\alpha = 0.05$, $\boldsymbol{H_0}$ will be rejected for computed sample values X^2 greater than 16.9, taken from a χ^2-table (chi-squared) with 9 degrees of freedom. Results and computed values for test statistics are listed in Table 6.3.

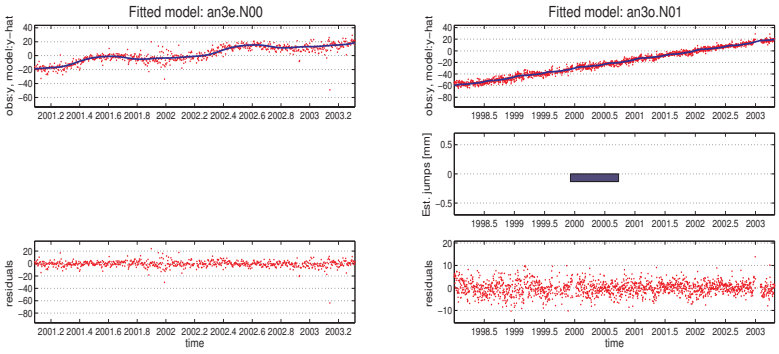

Figure 6.15: Fitted models, North component, stations: **an3e** and **an3o**

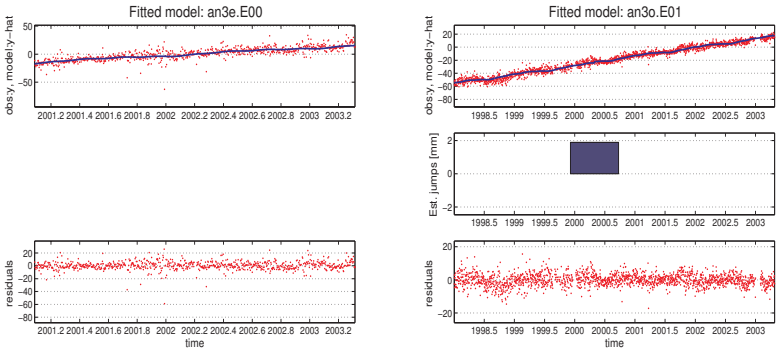

Figure 6.16: Fitted models, East component, stations: **an3e** and **an3o**

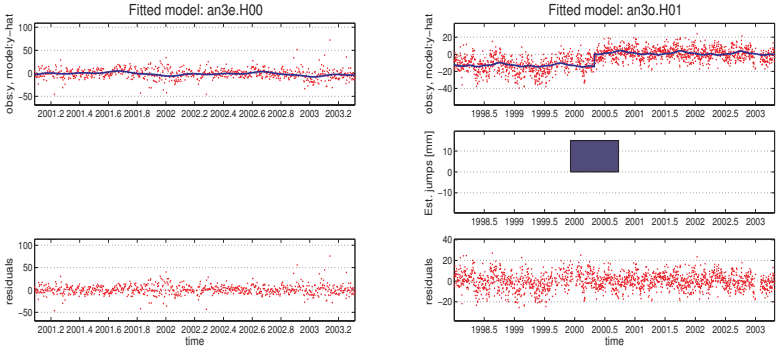

Figure 6.17: Fitted models, Height component, stations: **an3e** and **an3o**

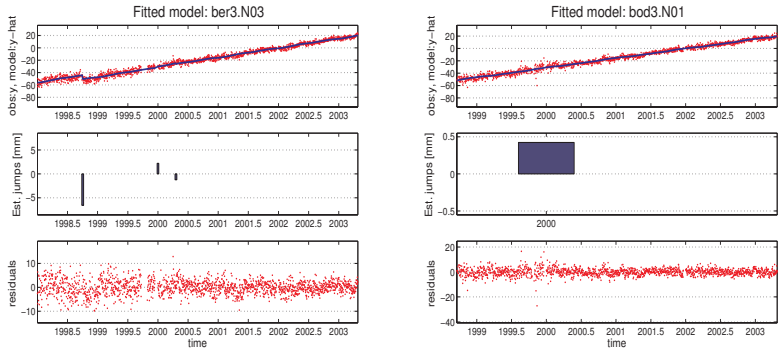

Figure 6.18: Fitted models, North component, stations: **ber3** and **bod3**

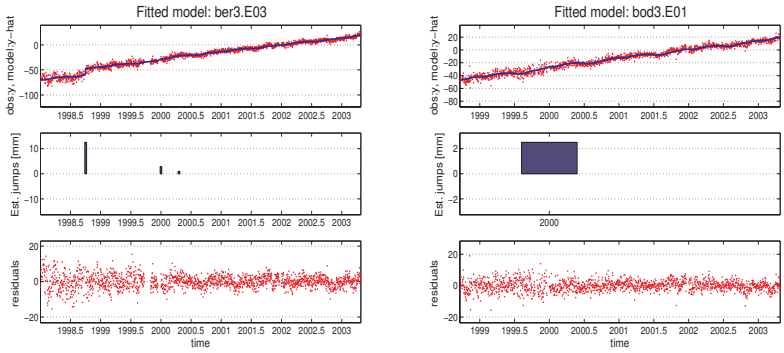

Figure 6.19: Fitted models, East component, stations: **ber3** and **bod3**

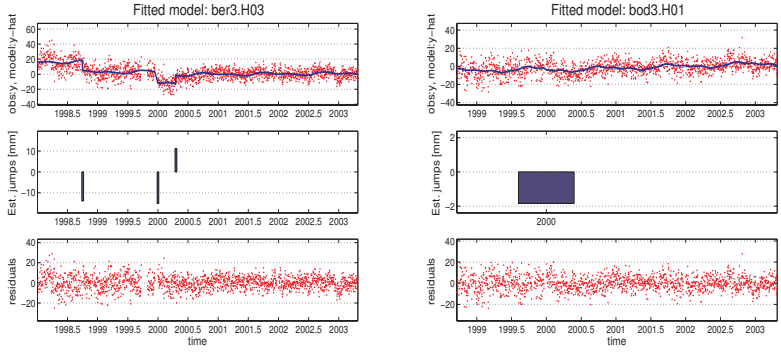

Figure 6.20: Fitted models, Height component, stations: **ber3** and **bod3**

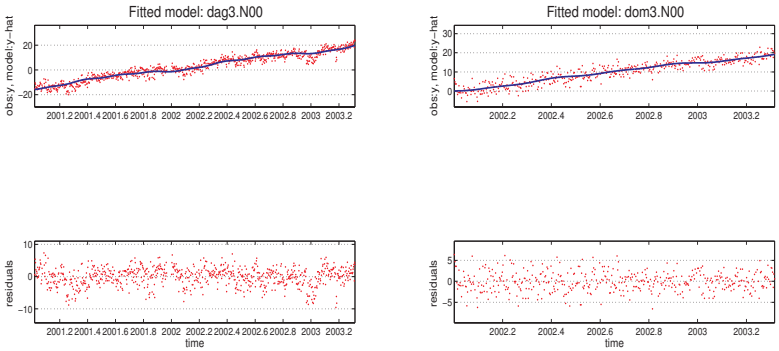

Figure 6.21: Fitted models, North component, stations: **dag3** and **dom3**

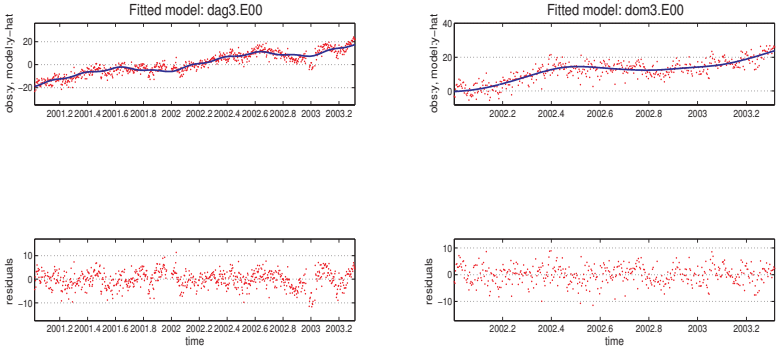

Figure 6.22: Fitted models, East component, stations: **dag3** and **dom3**

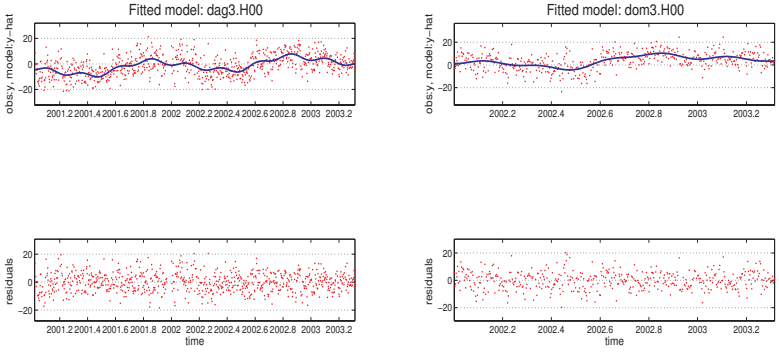

Figure 6.23: Fitted models, Height component, stations: **dag3** and **dom3**

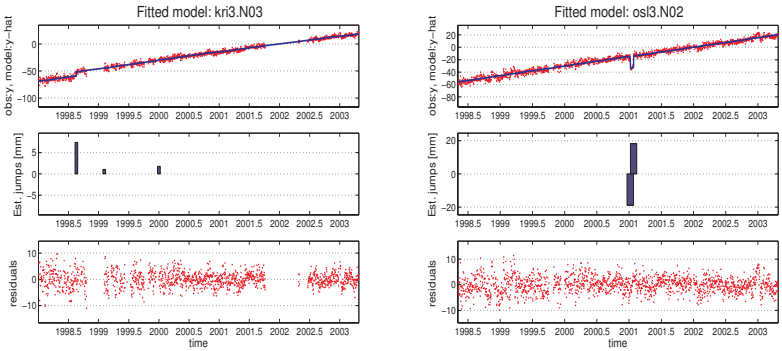

Figure 6.24: Fitted models, North component, stations: **kri3** and **osl3**

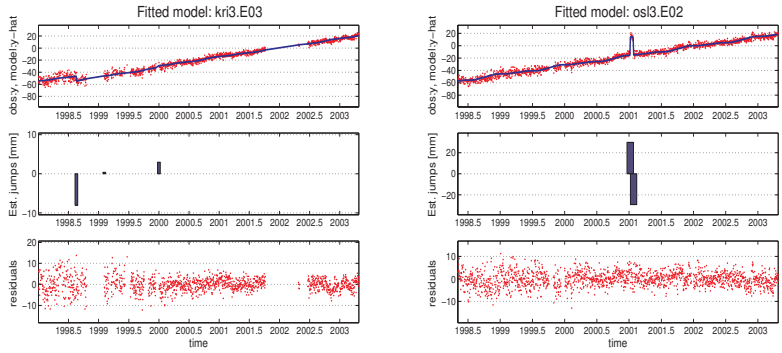

Figure 6.25: Fitted models, East component, stations: **kri3** and **osl3**

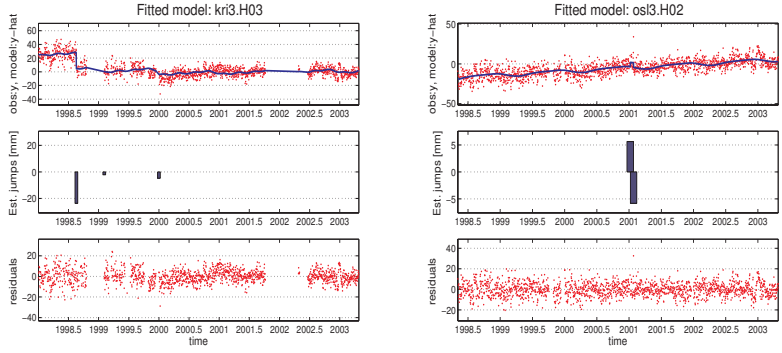

Figure 6.26: Fitted models, Height component, stations: **kri3** and **osl3**

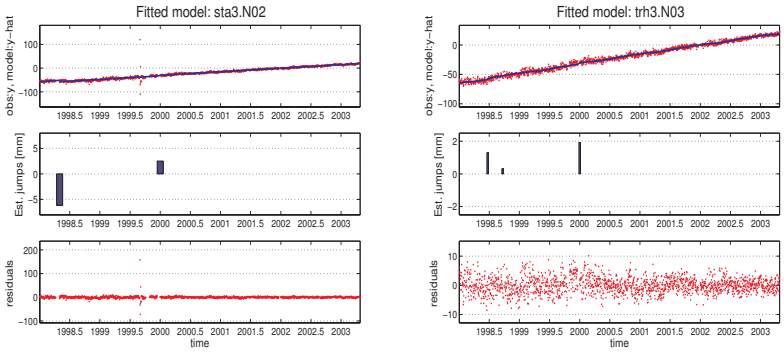

Figure 6.27: Fitted models, North component, stations: **sta3** and **trh3**

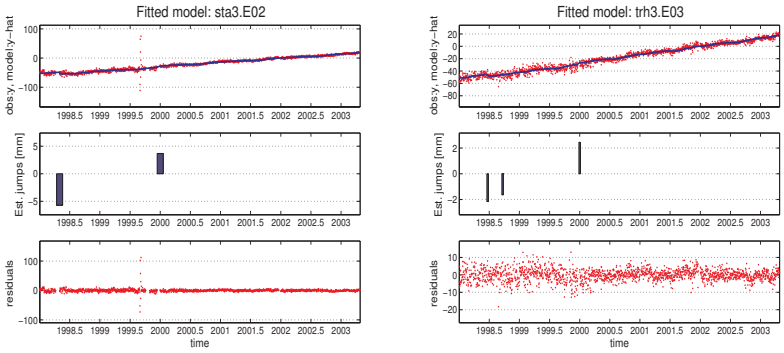

Figure 6.28: Fitted models, East component, stations: **sta3** and **trh3**

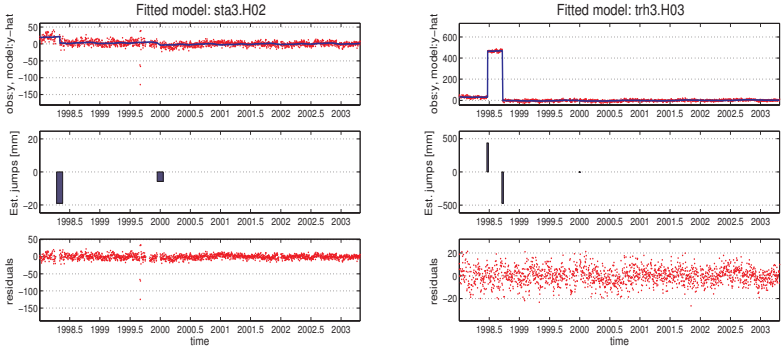

Figure 6.29: Fitted models, Height component, stations: **sta3** and **trh3**

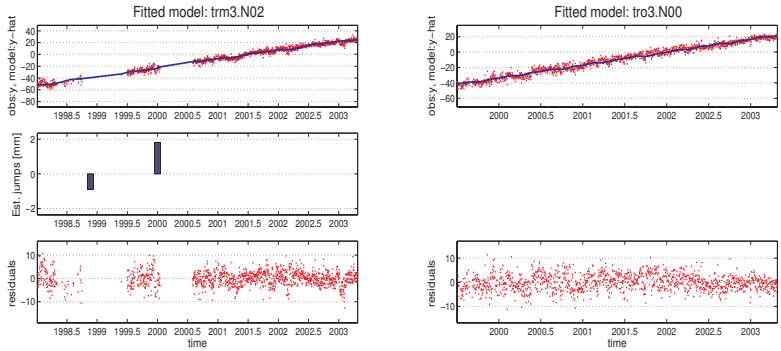

Figure 6.30: Fitted models, North component, stations: **trm3** and **tro3**

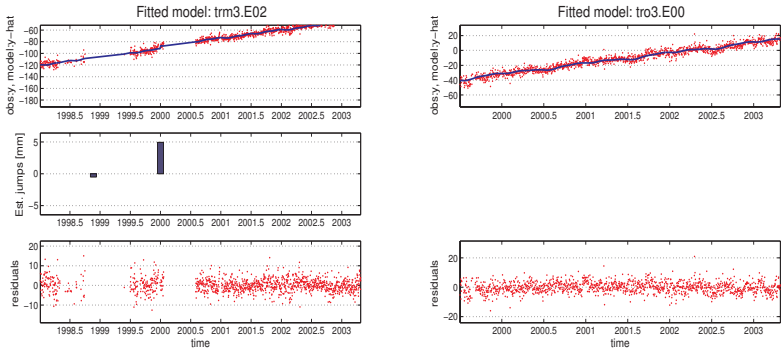

Figure 6.31: Fitted models, East component, stations: **trm3** and **tro3**

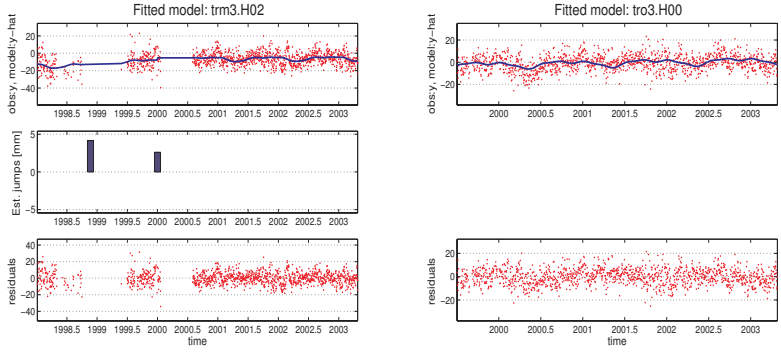

Figure 6.32: Fitted models, Height component, stations: **trm3** and **tro3**

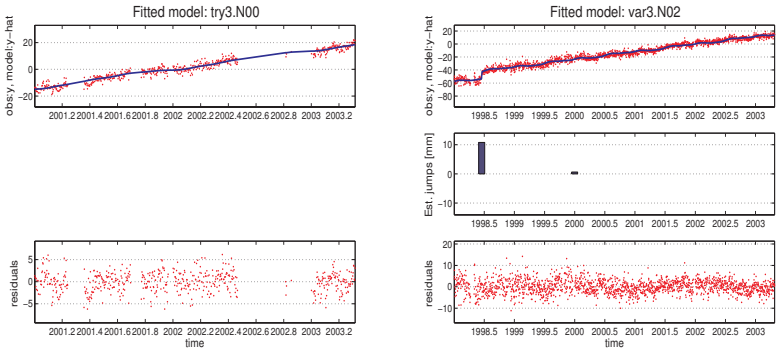

Figure 6.33: Fitted models, North component, stations: **try3** and **var3**

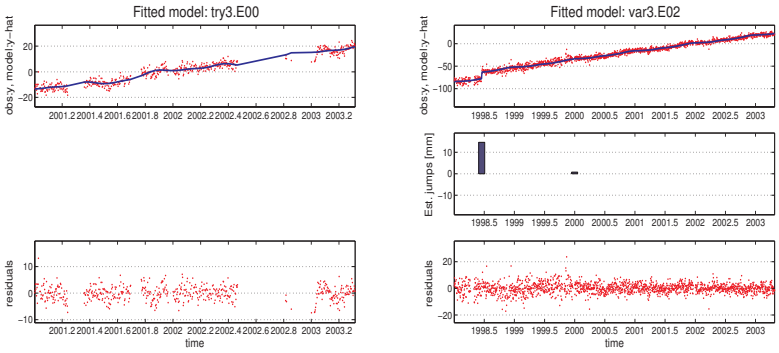

Figure 6.34: Fitted models, East component, stations: **try3** and **var3**

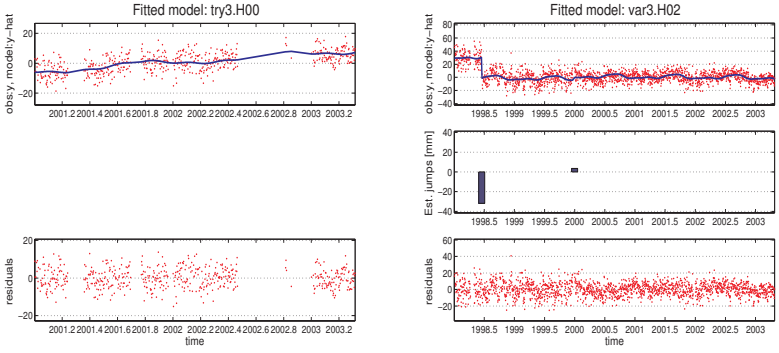

Figure 6.35: Fitted models, Height component, stations: **try3** and **var3**

Table 6.3: Goodness-of-fit tests for each station and direction of the Norwegian CGPS data. X^2 is the computed sample value for each station and direction. The test will be rejected for computed X^2 greater than 16.9, which is the χ^2-table (chi-squared) value at the significance level $(\alpha = 0.05)$.

Goodness-of-fit tests for Norwegian CGPS data, $(k = 3)$, $\boldsymbol{\alpha = 0.05}$						
Station	North	X^2	East	X^2	Height	X^2
an3e	**rejected**	33.60	Accepted	12.16	Accepted	7.27
an3o	Accepted	13.91	Accepted	9.15	Accepted	11.00
ber3	Accepted	9.27	Accepted	10.63	Accepted	15.26
bod3	Accepted	7.97	**rejected**	20.96	Accepted	11.62
dag3	**rejected**	35.96	**rejected**	20.65	Accepted	6.34
dom3	Accepted	8.09	Accepted	13.21	Accepted	4.20
kri3	Accepted	15.64	Accepted	4.72	Accepted	16.80
osl3	Accepted	15.05	Accepted	5.81	**rejected**	17.73
sta3	Accepted	8.29	Accepted	10.47	Accepted	11.49
trh3	Accepted	10.31	**rejected**	19.20	Accepted	15.22
trm3	**rejected**	27.76	Accepted	10.65	Accepted	10.94
tro3	Accepted	11.57	Accepted	9.05	Accepted	6.77
try3	**rejected**	20.88	Accepted	5.28	Accepted	6.12
var3	Accepted	5.05	Accepted	11.74	Accepted	7.21

Out of a total 42 tests, 8 of them, or 19 percent were rejected, against the expected 5 percent. However, the results show that the time series of normalised residuals mainly are, or at least are close to data taken from a $N(0, 1)$-distribution. As will be discussed in Subsection 6.1.4, this test is derived assuming observations to be independent of each other. Small estimated autocorrelations in the series make the test uncertain, but it indicates the assumption of normality to hold for the series of normalised residuals. Possible future improvements of the deterministic models will probably decrease the number of rejections in the tests above.

ACF-estimation

Three different estimation methods, the first two are described in Subsection 2.1.6, have been used to compute the ACF for the normalised residuals. The third method is the indirect one via the LST described in Figure 4.1. The methods generally gave identical results, even for data series with large gaps. As an example, three different estimated ACFs have been estimated for the height component of station Oslo from the first data set. They are all shown in Figure 6.36.

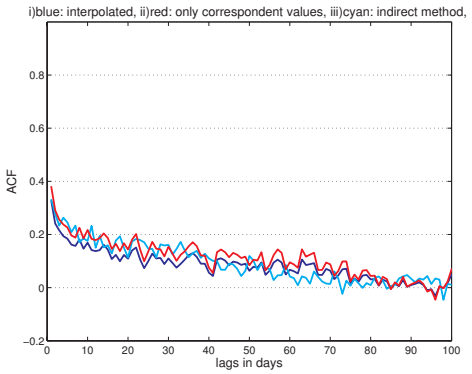

Figure 6.36: Estimation of ACF using three different methods.

The first interpolation method was chosen to estimate the ACFs for all series of normalised residuals in the Norwegian data set 3. Resulting plots for all stations and directions are shown in the Figures 6.37 to 6.50.

Comments on the results of ACF estimation for the Norwegian data

Some single lag-one auto correlations up to $ACF(1) = 0.4$ are found, but most of the lag-one auto correlations are around $ACF(1) = 0.2$. All estimated ACFs are exponential decreasing, which means that the series of normalised residuals may be interpreted as realisations from first order auto-regressive processes AR(1) with very small parameter values, also known as Markov-processes [see Wei, 1990, among others]. A brief description can be found in Appendix C. For equally weighted data, re-adjustments could be done using the known approximation to the inverse co-factor matrix from Eq. (C.8). As can be seen directly from the approximation formula, for small lag-one auto correlations, this hardly makes any difference for the estimated parameters. This is also certified by nu-

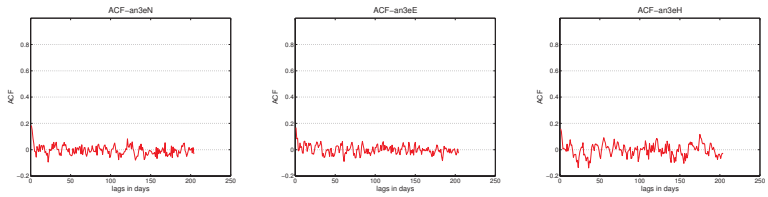

Figure 6.37: Auto correlation functions: **an3e** N,E,H

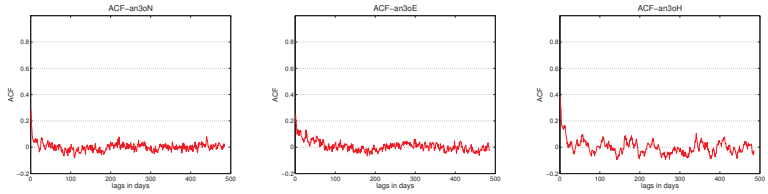

Figure 6.38: Auto correlation functions: **an3o** N,E,H

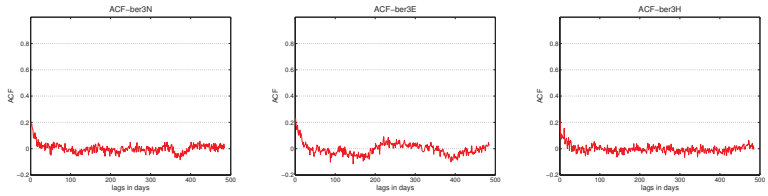

Figure 6.39: Auto correlation functions: **ber3** N,E,H

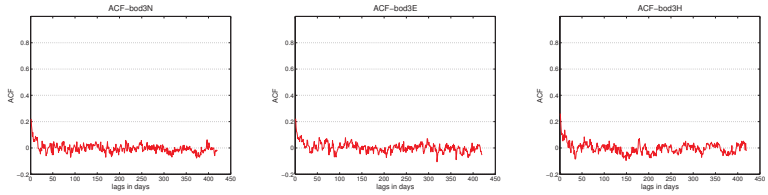

Figure 6.40: Auto correlation functions: **bod3** N,E,H

merical investigations of Craymer [1998] who's conclusion is cited in Section 4.3.

For the weighted data, a re-scaling using the variances for each observation could be done to find a new a priori covariance matrix, but the correlation structure will still be identical, so any numerical re-computations have not been performed for the final Norwegian data set. In fact the auto correlations are so small for all the

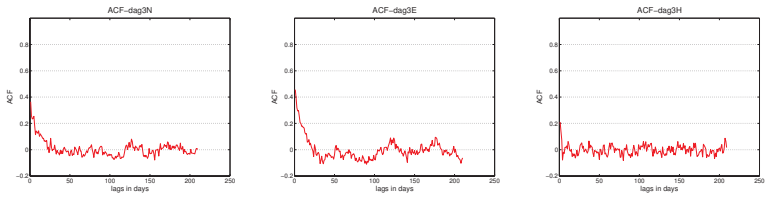

Figure 6.41: Auto correlation functions: **dag3** N,E,H

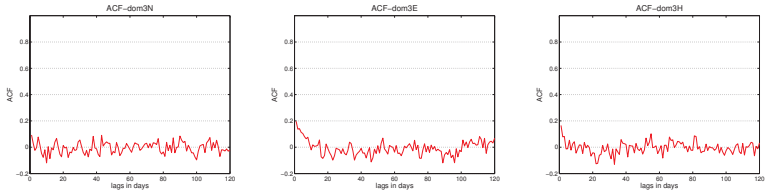

Figure 6.42: Auto correlation functions: **dom3** N,E,H

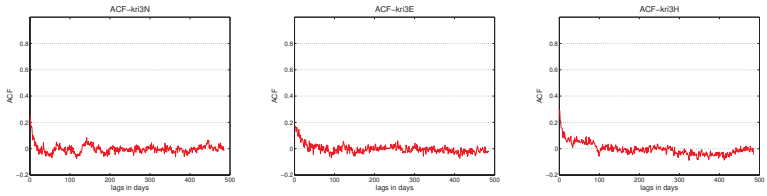

Figure 6.43: Auto correlation functions: **kri3** N,E,H

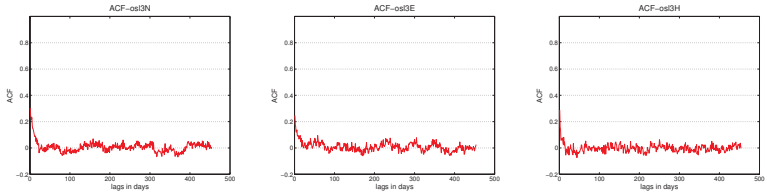

Figure 6.44: Auto correlation functions: **osl3** N,E,H

series, that it may not be a great mistake to assume the normalised residuals to be independent Gaussian white noise (taken from a $N(0,1)$-distribution) and thus the model assumptions of independent errors will hold. Together with an interpretation of estimated spectra, some of the series with largest lag-one auto correlations show possible undetected jumps. It is reasonable to believe that the lag-one auto correlations will further decrease if such jumps are determined in the future.

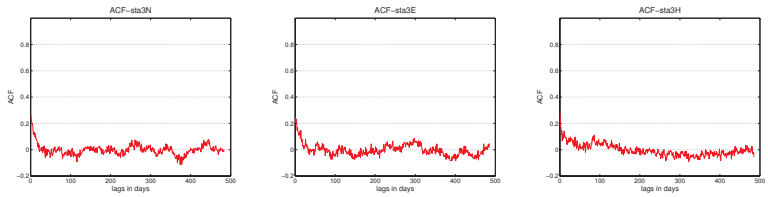

Figure 6.45: Auto correlation functions: **sta3** N,E,H

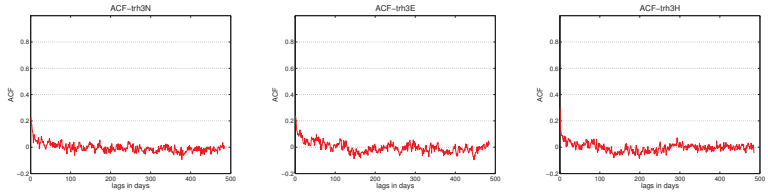

Figure 6.46: Auto correlation functions: **trh3** N,E,H

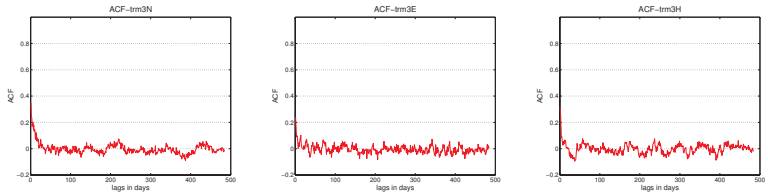

Figure 6.47: Auto correlation functions: **trm3** N,E,H

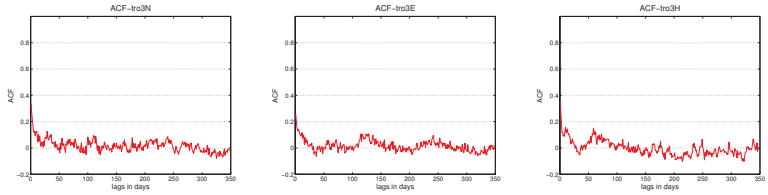

Figure 6.48: Auto correlation functions: **tro3** N,E,H

6.1.5 Likelihood ratio test for direction correlations

Goodness-of-fit tests in Section 6.1.4 have shown accordance with assumptions of the normalised residuals to be approximately $N(0,1)$-distributed. In Subsection 3.4.1 a likelihood ratio test was derived under assumptions of multi-normal distributed data. Based on time series of normalised residuals, correlation ma-

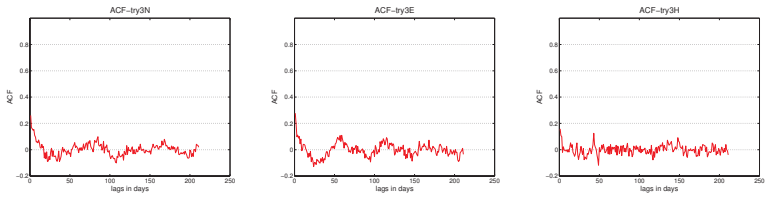

Figure 6.49: Auto correlation functions: **try3** N,E,H

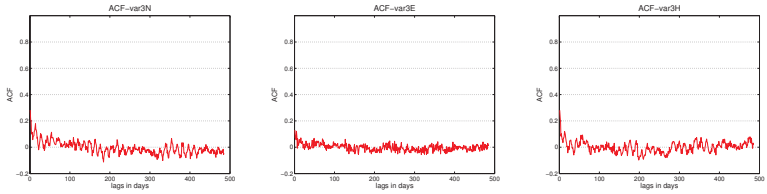

Figure 6.50: Auto correlation functions: **var3** N,E,H

trices (which are identical to covariance matrices for normalised data) have been estimated. The correlation matrices will be tested for the hypothesis from Eq. (3.48). If the null hypothesis is not rejected, there is no significant difference in an estimated $\boldsymbol{\rho}_{NEH}$-matrix, denoted $\boldsymbol{R}$, and the identity matrix $\boldsymbol{I}$. The conclusion would be that an $\boldsymbol{I}$-matrix can replace the estimated correlation matrix $\boldsymbol{R}$. In other words, correlations among the directions North, East and Height are not significant. Such a simplification makes it possible to reduce the original three dimensional multivariate problem to three univariate ones, which actually have been done all along through this study. This will particularly help the time of computation. With $p = 3$ variables (N, E, H) and a random sample j of normalised residuals denoted $\boldsymbol{x_j}^T = [n_j, e_j, h_j], \quad (j = 1, 2, ..., n)$, the data matrix:

$$\underset{(n\times 3)}{\boldsymbol{x}} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ x_{j1} & x_{j2} & x_{j3} \\ \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & x_{n3} \end{bmatrix} = \begin{bmatrix} n_1 & e_1 & h_1 \\ n_2 & e_2 & h_2 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ n_j & e_j & h_j \\ \vdots & \vdots & \vdots \\ n_n & e_n & h_n \end{bmatrix} \tag{6.4}$$

may be used as input to compute $\boldsymbol{R}$ in Eq. (3.54). It should be pointed out that the test assumes the samples j to be independent, which because of the small values of the ACFs is close to the reality for the Norwegian data. For each station, correlation matrices were estimated from the normalised residuals remaining after a final choice of deterministic models as described in the Section 6.1.3. Factor $(k = 3)$ was used in the outlier test explained in Subsection 3.2.1. No interpolation has been made for the data gaps. Only valid combinations of observations of all three directions have been used. In a special case when all off diagonal elements are equal to zero in $\boldsymbol{R}$, the determinant: $|\boldsymbol{R}| = 1$. The larger the positive correlations of off diagonal elements, the smaller becomes the determinant $|\boldsymbol{R}|$. The correlation matrices $\boldsymbol{R}_{NEH}$ for all 14 stations are listed below.

Station: **an3e**

$$\boldsymbol{R} = \begin{bmatrix} 1.0000 & 0.1177 & -0.0413 \\ 0.1177 & 1.0000 & 0.1689 \\ -0.0413 & 0.1689 & 1.0000 \end{bmatrix}$$

Station: **an3o**

$$\boldsymbol{R} = \begin{bmatrix} 1.0000 & 0.0064 & 0.0454 \\ 0.0064 & 1.0000 & 0.0496 \\ 0.0454 & 0.0496 & 1.0000 \end{bmatrix}$$

Station: **ber3**

$$\boldsymbol{R} = \begin{bmatrix} 1.0000 & 0.0174 & -0.0240 \\ 0.0174 & 1.0000 & 0.0227 \\ -0.0240 & 0.0227 & 1.0000 \end{bmatrix}$$

Station: **bod3**

$$\boldsymbol{R} = \begin{bmatrix} 1.0000 & 0.0451 & -0.0062 \\ 0.0451 & 1.0000 & 0.0396 \\ -0.0062 & 0.0396 & 1.0000 \end{bmatrix}$$

Station: **dag3**

$$\boldsymbol{R} = \begin{bmatrix} 1.0000 & 0.2246 & -0.0012 \\ 0.2246 & 1.0000 & -0.0198 \\ -0.0012 & -0.0198 & 1.0000 \end{bmatrix}$$

Station: **dom3**

$$\boldsymbol{R} = \begin{bmatrix} 1.0000 & 0.0569 & 0.0158 \\ 0.0569 & 1.0000 & -0.0214 \\ 0.0158 & -0.0214 & 1.0000 \end{bmatrix}$$

Station: **kri3**

$$\boldsymbol{R} = \begin{bmatrix} 1.0000 & -0.0190 & -0.0318 \\ -0.0190 & 1.0000 & 0.0723 \\ -0.0318 & 0.0723 & 1.0000 \end{bmatrix}$$

Station: **osl3**

$$\boldsymbol{R} = \begin{bmatrix} 1.0000 & 0.0563 & 0.0351 \\ 0.0563 & 1.0000 & -0.0123 \\ 0.0351 & -0.0123 & 1.0000 \end{bmatrix}$$

(6.5)

Station: **sta3**

$$\boldsymbol{R} = \begin{bmatrix} 1.0000 & -0.0210 & 0.0241 \\ -0.0210 & 1.0000 & 0.0678 \\ 0.0241 & 0.0678 & 1.0000 \end{bmatrix}$$

Station: **trh3**

$$\boldsymbol{R} = \begin{bmatrix} 1.0000 & 0.0173 & 0.0744 \\ 0.0173 & 1.0000 & 0.0352 \\ 0.0744 & 0.0352 & 1.0000 \end{bmatrix}$$

Station: **trm3**

$$\boldsymbol{R} = \begin{bmatrix} 1.0000 & 0.0804 & -0.0252 \\ 0.0804 & 1.0000 & 0.0360 \\ -0.0252 & 0.0360 & 1.0000 \end{bmatrix}$$

Station: **tro3**

$$\boldsymbol{R} = \begin{bmatrix} 1.0000 & 0.0526 & 0.0144 \\ 0.0526 & 1.0000 & 0.1174 \\ 0.0144 & 0.1174 & 1.0000 \end{bmatrix}$$

Station: **try3**

$$\boldsymbol{R} = \begin{bmatrix} 1.0000 & 0.0617 & -0.0237 \\ 0.0617 & 1.0000 & 0.0536 \\ -0.0237 & 0.0536 & 1.0000 \end{bmatrix}$$

Station: **var3**

$$\boldsymbol{R} = \begin{bmatrix} 1.0000 & -0.0120 & -0.0146 \\ -0.0120 & 1.0000 & -0.0150 \\ -0.0146 & -0.0150 & 1.0000 \end{bmatrix}$$

(6.6)

The largest correlation found was $r_{ne} = 0.2246$ for the series **dag3**. Except for two of the correlations for the series **an3e** and one for the series **tro3**, no correlation coefficients exceeds the value 0.1. Computed values for the test statistics $-2\ln\Lambda$ from Eq. (3.55) are compared to values taken from the χ^2-distribution table: $\chi^2_3(\alpha = 0.05) = 7.81$, and $\chi^2_3(\alpha = 0.01) = 11.34$. $\boldsymbol{H_0}$ is rejected for $-2\ln\Lambda > \chi^2_3(\alpha)$ Results and computed values for test statistics is listed in Table 6.4

At significance level $(\alpha = 0.05)$ eight out of fourteen tests were rejected, while four out of fourteen tests were rejected at significance level $(\alpha = 0.01)$. Different values of the k-factor in the outlier detection routine are tested but it does not seems to have a large effect for the test of significant cross correlations.

Dependencies other than a linear one may occur between the time series of normalised residuals, Plots are investigated for each station between series representing the directions, as is done for station **var3** in Figure 6.51. No suspicious dependencies were indicated.

Table 6.4: Likelihood ratio tests of correlation matrices estimated from time series of normalised residuals representing each of the Norwegian CGPS stations. Accepting $\boldsymbol{H_0}$ means there is no significant difference between a full or complete correlation matrix and a unity matrix.

Likelihood-ratio test, Norwegian data ($k = 3$)				$\boldsymbol{H}_0 : \boldsymbol{R_{NEH}} = \boldsymbol{I}$	
station	**determinant**	n_{obs}	$-2 \ln \Lambda$	$\boldsymbol{H}_0(\alpha = 0.05)$	$\boldsymbol{H}_0(\alpha = 0.01)$
an3e	0.9543	774	36.23	rejected	rejected
an3o	0.9955	1751	7.952	rejected	accepted
ber3	0.9986	1710	2.417	accepted	accepted
bod3	0.9963	1544	5.675	accepted	accepted
dag3	0.9492	754	39.33	rejected	rejected
dom3	0.9960	455	1.815	accepted	accepted
kri3	0.9935	1335	8.716	rejected	accepted
osl3	0.9954	1636	7.530	accepted	accepted
sta3	0.9943	1711	9.750	rejected	accepted
trh3	0.9930	1723	12.07	rejected	rejected
trm3	0.9915	1209	10.38	rejected	accepted
tro3	0.9834	1438	22.11	rejected	rejected
try3	0.9926	526	3.910	accepted	accepted
var3	0.9994	1697	1.001	accepted	accepted

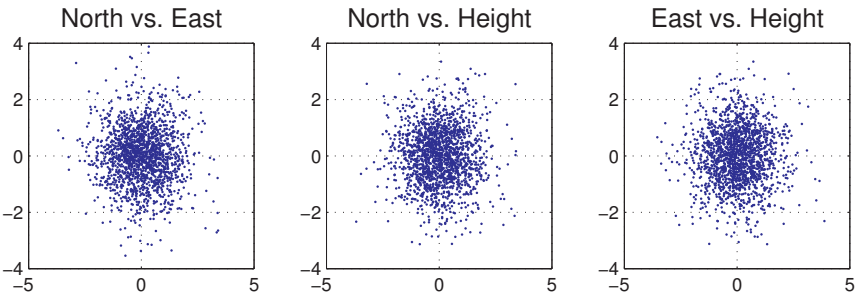

Figure 6.51: Pairs of normalised residual series representing the directions, station **var3**

A question is how possible future improved deterministic models will influence the test. Probably an improved model will decrease the estimated cross correlations and most of the resulting tests from Table 6.4 can be accepted.

6.1.6 Cross correlations

A weighted linear regression for each station and direction was performed with deterministic models including parameters described in Section 6.1.3. The factor $(k = 3)$ was used in the outlier detection procedure from Subsection 3.2.1. Because of a major portion of gaps in the series, the correlation procedure using only valid pairs of observations from the series as described in Subsection 2.1.6, has been used. Correlation matrices representing each of the three directions were estimated from the remaining time series of normalised residuals of the $p = 14$ stations, for the station vector:

$$\left[\ an3e \quad an3o \quad ber3 \quad bod3 \quad dag3 \quad dom3 \quad kri3 \quad osl3 \quad sta3 \quad trh3 \quad trm3 \quad tro3 \quad try3 \quad var3\ \right]^T$$

The matrices were computed to: for direction N:, $\boldsymbol{R_N} =$

$$\begin{bmatrix}
1.00 & 0.33 & 0.34 & 0.42 & 0.30 & 0.17 & 0.31 & 0.30 & 0.38 & 0.37 & 0.36 & 0.22 & 0.41 & 0.35 \\
0.33 & 1.00 & 0.42 & 0.60 & 0.30 & 0.33 & 0.43 & 0.43 & 0.46 & 0.52 & 0.53 & 0.42 & 0.49 & 0.52 \\
0.34 & 0.42 & 1.00 & 0.53 & 0.53 & 0.51 & 0.66 & 0.60 & 0.73 & 0.55 & 0.37 & 0.31 & 0.69 & 0.33 \\
0.42 & 0.60 & 0.53 & 1.00 & 0.38 & 0.46 & 0.55 & 0.53 & 0.59 & 0.62 & 0.50 & 0.46 & 0.62 & 0.53 \\
0.30 & 0.30 & 0.53 & 0.38 & 1.00 & 0.38 & 0.52 & 0.39 & 0.58 & 0.44 & 0.36 & 0.21 & 0.59 & 0.30 \\
0.17 & 0.33 & 0.51 & 0.46 & 0.38 & 1.00 & 0.47 & 0.42 & 0.56 & 0.51 & 0.33 & 0.25 & 0.49 & 0.30 \\
0.31 & 0.43 & 0.66 & 0.55 & 0.52 & 0.47 & 1.00 & 0.66 & 0.75 & 0.56 & 0.40 & 0.33 & 0.66 & 0.36 \\
0.30 & 0.43 & 0.60 & 0.53 & 0.39 & 0.42 & 0.66 & 1.00 & 0.65 & 0.55 & 0.33 & 0.31 & 0.57 & 0.36 \\
0.38 & 0.46 & 0.73 & 0.59 & 0.58 & 0.56 & 0.75 & 0.65 & 1.00 & 0.61 & 0.43 & 0.33 & 0.75 & 0.37 \\
0.37 & 0.52 & 0.55 & 0.62 & 0.44 & 0.51 & 0.56 & 0.55 & 0.61 & 1.00 & 0.41 & 0.35 & 0.66 & 0.42 \\
0.36 & 0.53 & 0.37 & 0.50 & 0.36 & 0.33 & 0.40 & 0.33 & 0.43 & 0.41 & 1.00 & 0.50 & 0.45 & 0.51 \\
0.22 & 0.42 & 0.31 & 0.46 & 0.21 & 0.25 & 0.33 & 0.31 & 0.33 & 0.35 & 0.50 & 1.00 & 0.30 & 0.45 \\
0.41 & 0.49 & 0.69 & 0.62 & 0.59 & 0.49 & 0.66 & 0.57 & 0.75 & 0.66 & 0.45 & 0.30 & 1.00 & 0.42 \\
0.35 & 0.52 & 0.33 & 0.53 & 0.30 & 0.30 & 0.36 & 0.36 & 0.37 & 0.42 & 0.51 & 0.45 & 0.42 & 1.00
\end{bmatrix}$$

for direction E: $\boldsymbol{R_E} =$

$$\begin{bmatrix}
1.00 & 0.39 & 0.32 & 0.41 & 0.24 & 0.28 & 0.36 & 0.31 & 0.37 & 0.36 & 0.33 & 0.20 & 0.34 & 0.31 \\
0.39 & 1.00 & 0.35 & 0.52 & 0.38 & 0.31 & 0.36 & 0.40 & 0.38 & 0.37 & 0.46 & 0.43 & 0.40 & 0.37 \\
0.32 & 0.35 & 1.00 & 0.45 & 0.52 & 0.43 & 0.45 & 0.44 & 0.54 & 0.42 & 0.30 & 0.31 & 0.54 & 0.20 \\
0.41 & 0.52 & 0.45 & 1.00 & 0.42 & 0.37 & 0.44 & 0.40 & 0.47 & 0.47 & 0.46 & 0.46 & 0.49 & 0.40 \\
0.24 & 0.38 & 0.52 & 0.42 & 1.00 & 0.40 & 0.49 & 0.29 & 0.54 & 0.49 & 0.31 & 0.28 & 0.44 & 0.24 \\
0.28 & 0.31 & 0.43 & 0.37 & 0.40 & 1.00 & 0.38 & 0.33 & 0.43 & 0.47 & 0.22 & 0.28 & 0.27 & 0.22 \\
0.36 & 0.36 & 0.45 & 0.44 & 0.49 & 0.38 & 1.00 & 0.43 & 0.55 & 0.42 & 0.30 & 0.34 & 0.53 & 0.20 \\
0.31 & 0.40 & 0.44 & 0.40 & 0.29 & 0.33 & 0.43 & 1.00 & 0.45 & 0.44 & 0.28 & 0.28 & 0.47 & 0.24 \\
0.37 & 0.38 & 0.54 & 0.47 & 0.54 & 0.43 & 0.55 & 0.45 & 1.00 & 0.45 & 0.29 & 0.36 & 0.53 & 0.21 \\
0.36 & 0.37 & 0.42 & 0.47 & 0.49 & 0.47 & 0.42 & 0.44 & 0.45 & 1.00 & 0.31 & 0.34 & 0.52 & 0.25 \\
0.33 & 0.46 & 0.30 & 0.46 & 0.31 & 0.22 & 0.30 & 0.28 & 0.29 & 0.31 & 1.00 & 0.60 & 0.34 & 0.35 \\
0.20 & 0.43 & 0.31 & 0.46 & 0.28 & 0.28 & 0.34 & 0.28 & 0.36 & 0.34 & 0.60 & 1.00 & 0.31 & 0.32 \\
0.34 & 0.40 & 0.54 & 0.49 & 0.44 & 0.27 & 0.53 & 0.47 & 0.53 & 0.52 & 0.34 & 0.31 & 1.00 & 0.32 \\
0.31 & 0.37 & 0.20 & 0.40 & 0.24 & 0.22 & 0.20 & 0.24 & 0.21 & 0.25 & 0.35 & 0.32 & 0.32 & 1.00
\end{bmatrix}$$

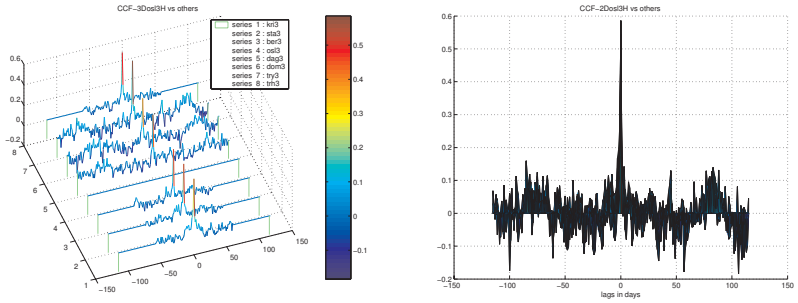

Figure 6.52: **osl3** height vs some others, 3D-plot and the 3D rotated to a 2D-plot

and for direction H: $\boldsymbol{R_H} =$

$$\begin{bmatrix}
1.00 & 0.41 & 0.30 & 0.45 & 0.20 & 0.26 & 0.25 & 0.23 & 0.28 & 0.34 & 0.36 & 0.38 & 0.30 & 0.29 \\
0.41 & 1.00 & 0.37 & 0.53 & 0.27 & 0.27 & 0.32 & 0.38 & 0.31 & 0.44 & 0.46 & 0.52 & 0.41 & 0.47 \\
0.30 & 0.37 & 1.00 & 0.48 & 0.40 & 0.45 & 0.47 & 0.46 & 0.58 & 0.46 & 0.35 & 0.40 & 0.52 & 0.26 \\
0.45 & 0.53 & 0.48 & 1.00 & 0.30 & 0.34 & 0.41 & 0.39 & 0.42 & 0.55 & 0.55 & 0.55 & 0.49 & 0.44 \\
0.20 & 0.27 & 0.40 & 0.30 & 1.00 & 0.50 & 0.34 & 0.43 & 0.36 & 0.38 & 0.24 & 0.25 & 0.51 & 0.19 \\
0.26 & 0.27 & 0.45 & 0.34 & 0.50 & 1.00 & 0.40 & 0.40 & 0.38 & 0.40 & 0.27 & 0.29 & 0.55 & 0.23 \\
0.25 & 0.32 & 0.47 & 0.41 & 0.34 & 0.40 & 1.00 & 0.55 & 0.63 & 0.44 & 0.36 & 0.34 & 0.50 & 0.29 \\
0.23 & 0.38 & 0.46 & 0.39 & 0.43 & 0.40 & 0.55 & 1.00 & 0.50 & 0.48 & 0.31 & 0.33 & 0.59 & 0.34 \\
0.28 & 0.31 & 0.58 & 0.42 & 0.36 & 0.38 & 0.63 & 0.50 & 1.00 & 0.43 & 0.33 & 0.34 & 0.48 & 0.24 \\
0.34 & 0.44 & 0.46 & 0.55 & 0.38 & 0.40 & 0.44 & 0.48 & 0.43 & 1.00 & 0.41 & 0.43 & 0.56 & 0.34 \\
0.36 & 0.46 & 0.35 & 0.55 & 0.24 & 0.27 & 0.36 & 0.31 & 0.33 & 0.41 & 1.00 & 0.76 & 0.35 & 0.45 \\
0.38 & 0.52 & 0.40 & 0.55 & 0.25 & 0.29 & 0.34 & 0.33 & 0.34 & 0.43 & 0.76 & 1.00 & 0.34 & 0.48 \\
0.30 & 0.41 & 0.52 & 0.49 & 0.51 & 0.55 & 0.50 & 0.59 & 0.48 & 0.56 & 0.35 & 0.34 & 1.00 & 0.31 \\
0.29 & 0.47 & 0.26 & 0.44 & 0.19 & 0.23 & 0.29 & 0.34 & 0.24 & 0.34 & 0.45 & 0.48 & 0.31 & 1.00
\end{bmatrix}$$

Cross correlation functions

A large number of cross correlation functions have been investigated. An example is given in Figure 6.52 for the vertical component of the normalised residual series from **osl3** among some other stations.

Correlation as a function of distance between sites

In Figures 6.53 and 6.54, the correlations are plotted as a function of the distance between different stations. The correlation decreases only slightly with the distance. A least squares fit of a linear regression model for the height components resulted in the equation: $y = -0.00017x + 0.52$, where y is the correlation, and

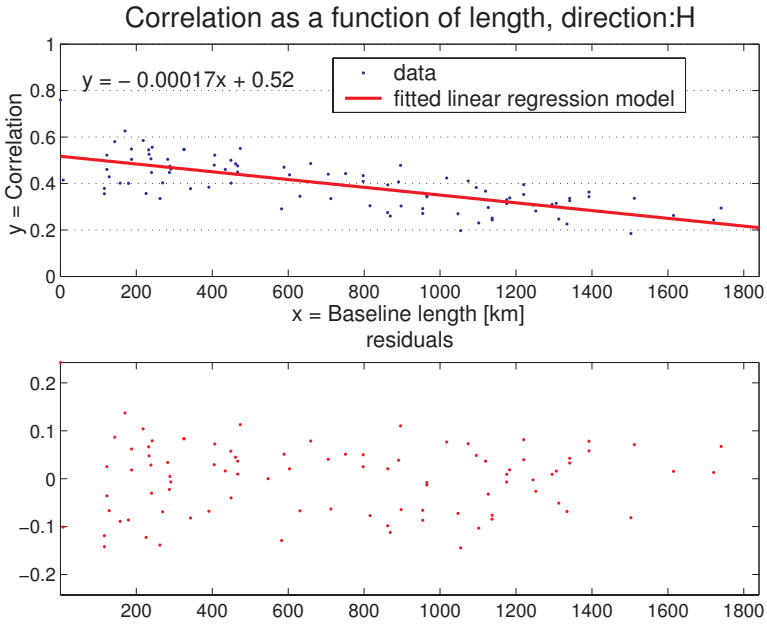

Figure 6.53: Correlation as a function of distance, H-component

x is the distance in km. Almost identical results are obtained for the other directions. Note that, this is close to the results for the height component, obtained by Johansson et al. [2002]. They computed $y = -0.0002x + 0.6$ for an almost similar investigation of the CGPS network of Sweden and Finland, and conclude that this correlation might be caused by at least one common effect in the GPS system, probably a reference frame or orbital-type effect.

6.1.7 Principal Component Analysis

The applied theory for this subsection is described in Chapter 5. The cross correlation matrices estimated in Subsection 6.1.6 show some large correlation coefficients between most sites, and for all directions, and makes an PCA of current interest. The CGPS data series from the Norwegian network include several intervals of missing data, and they are of fairly different lengths. This introduces a scale factor problem for the estimation of variances and covariances. To avoid this

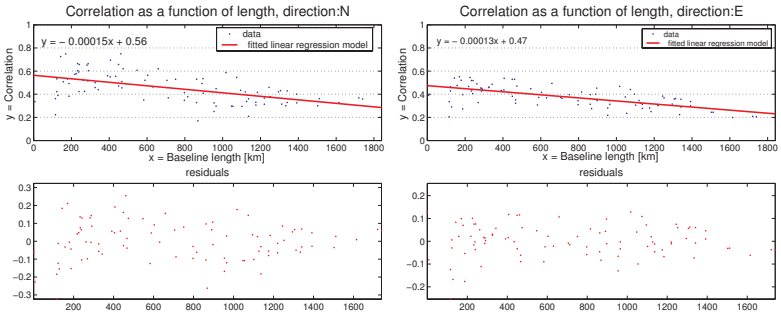

Figure 6.54: Correlation as a function of distance, N and E-component

error, the normalised residuals have been chosen as input to the three separately performed PCAs for the directions North, East and Height. For normalised data, the estimated covariance matrix becomes a correlation matrix, and thus the matrices from Subsection 6.1.6 can be used. Using correlation matrices which have diagonal elements one, ensures that the stations become equally weighted in the PCA. On the other hand, this will disregard the different variability of each station series and thus introduce another scale problem. Compared with Root Mean Square (RMS) values of each single series, the latter problem is considered to give a smaller influence on the PCAs and will therefore be chosen for the Norwegian data. For the purpose using CGPS data to identify geophysical processes, another question would be if larger variability at some sites is really a consequence of lager magnitudes of surface displacements or is it also a result of poor GPS-satellite geometry and thus lower precision of the CGPS observations?

Results

Eigenvalues sorted by size and the accumulated part of the total sample variance obtained by each PC are listed in Table 6.5 and visualised in Figure 6.55. Figure 6.55 shows that more than 50 percent of the total variance for the north direction are explained by the first PC or statistical mode, and more than 60 percent by the two first PC's and so on. For the other directions, east and height, the overall variance explanation by the two first PCs is a bit smaller, but they also show a very significant first PC or mode. This indicates that at least one common mode is dominating the data series from all stations and directions. The contribution from each

Table 6.5: Eigenvalues for the North, East and Height component of the Norwegian CGPS data. j refers to the number of eigenvalues which is sorted in descending order, λ_j is the eigenvalue and accum. the fraction of variance explained by the first j modes

	North		**East**		**Height**	
j	λ_j	accum.	λ_j	accum.	λ_j	accum.
1	7.044	0.503	5.987	0.428	6.153	0.439
2	1.407	0.604	1.328	0.522	1.609	0.554
3	0.840	0.664	0.868	0.584	0.826	0.613
4	0.694	0.713	0.784	0.640	0.783	0.669
5	0.633	0.758	0.704	0.691	0.693	0.719
6	0.559	0.798	0.685	0.740	0.599	0.762
7	0.497	0.834	0.586	0.782	0.543	0.800
8	0.461	0.867	0.568	0.822	0.521	0.838
9	0.392	0.895	0.547	0.861	0.473	0.871
10	0.361	0.921	0.482	0.896	0.445	0.903
11	0.337	0.945	0.434	0.927	0.395	0.931
12	0.309	0.967	0.413	0.956	0.369	0.958
13	0.260	0.985	0.357	0.982	0.340	0.982
14	0.205	1.000	0.258	1.000	0.252	1.000

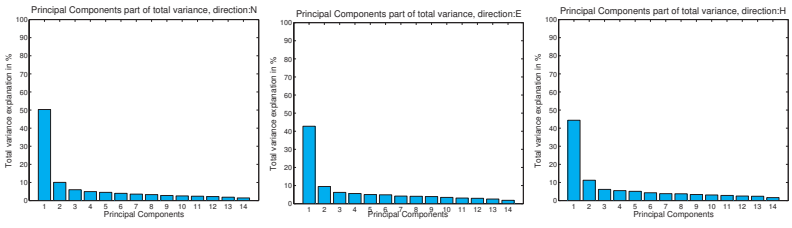

Figure 6.55: Overall variance explanation from the PC's.

station to the overall variance obtained by each mode is given as the magnitudes of the eigenvector elements, or their scaled version, the factor loadings. To show this spatial pattern, the first three loadings are illustrated as arrows for all directions in Figure 6.56. These values, as well as numerical values of the normalised eigenvectors to the corresponding six largest eigenvalues are listed in Tables D.1 to D.6 in Appendix D. All stations score positive at the first loadings for all directions. This confirms that one mode is common for the entire network of Norway.

The second factor loadings differ for the stations in the northern and southern parts. Identical patterns are shown for all directions, a dip of the country around an North-East to South-West axis north of Trondheim. From the third and higher order PCs, it is more difficult to see a pattern. However, the third loadings for the height component may indicate a possible local effect in the mountain region of southern Norway. Only a few stations are located in mountain regions, hence this effect is more uncertain. An interpretation of the modes and their causing processes is more difficult. The contributions to each mode may also be caused by one or more interacting effects. Moreover, the Norwegian CGPS data have shown to be of variable quality. To confirm the results from the PCA of Norwegian data, it was decided to work out a new PCA for a larger area based on another, independently pre-processed, data set from an European network of CGPS stations. Results and conclusions from the Norwegian PCA will be drawn out and related to the European PCA analysis performed in Section 6.2. More results from the PCA of the Norwegian data set can be found in [Haakonsen et al., 2004a].

6.2 PCA of European CGPS data

Another approach is used for the PCA of the European CGPS data. It is expected that most harmonic constituents visible in the station series are of a regional or global spatial scale. Such signals are expected to be expressed in the first few modes, while local effects are assumed to appear in the higher modes from a PCA. Station dependent local effects will probably not affect the results if only the modes of lower order are treated.

A preliminary description of the four years (2000.0-2004.0) CGPS data from the European network is made in Subsection 1.4.2. A further selection of stations was made on the basis of a pre-processing of the coordinate time series. For that, only the time series of vertical components was de-trended and a test for outliers was carried out. Outliers were removed from the time series. Weighted linear regressions were performed for each station, using a priori diagonal observation weight matrices created from the inverted output variances of the PPP. Epochs of known jumps in the time series were provided by Heflin (personal communication 2004) and included in the linear regression as Heaveside step functions. No periodic

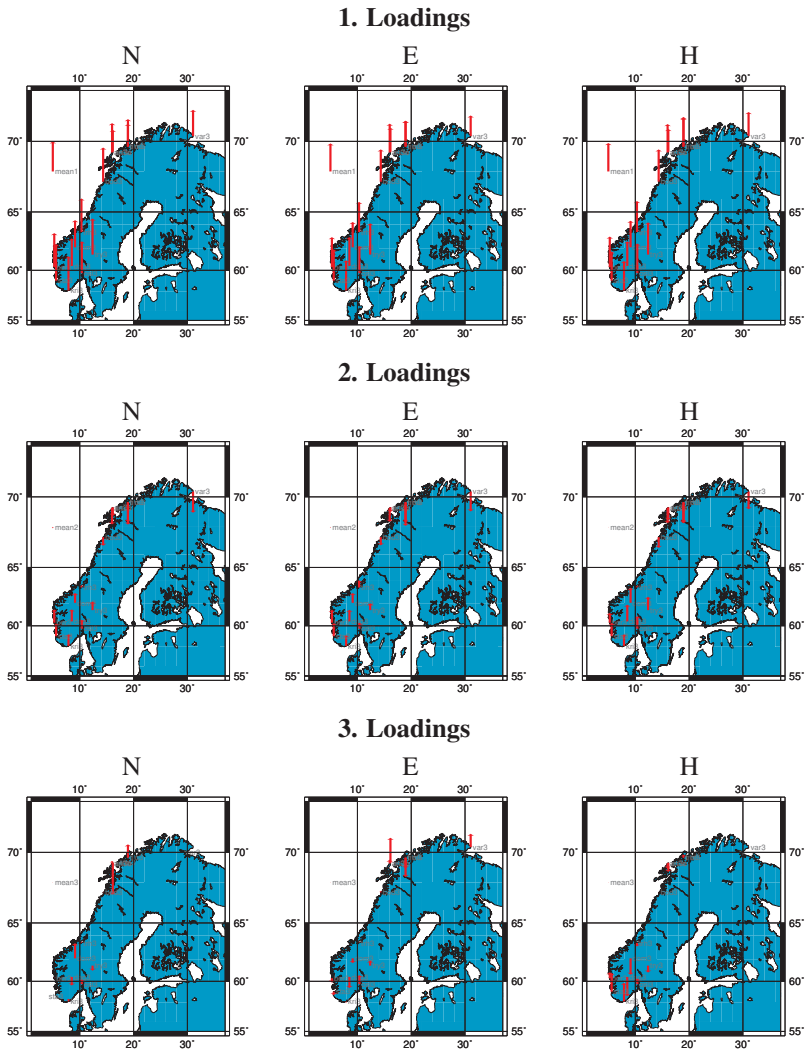

Figure 6.56: PCA of the Norwegian CGPS data, Loadings for the North, East and Height components of the first three modes. mean1,mean2 and mean3 are the mean values of first, second and third loadings, respectively.

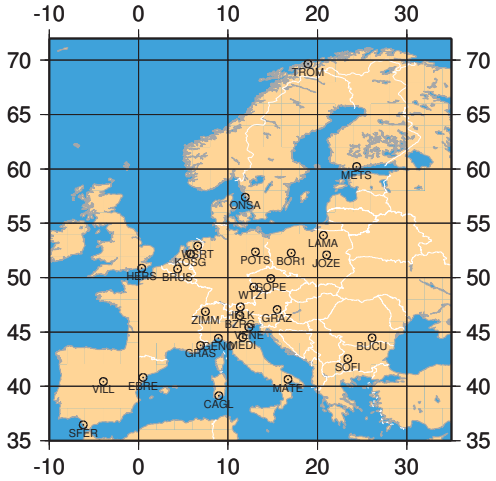

Figure 6.57: Location of the 28 European CGPS sites.

components were included in the regression models. Subsequently, time series of residuals were computed by removing a constant, a linear trend, and any jumps as determined in the regressions. For the further analysis, the zero mean time series of residuals remaining from the weighted regressions are used, so no weights are used in the subsequent analyses. The intension was to use more than the 28 finally selected European stations. Some series e.g. MADR, MAD2 (both in Spain) and PENC (Hungary) showed very low computed cross correlation coefficient considered to the other station series. It was assumed that these low correlations were due to local problems at these sites (e.g. undetected offsets) and therefore, these stations were excluded from the analysis. Furthermore, the station KIRU (Kiruna, Sweden) revealed problems due to snow cover on the antenna in late winter, and this station was also excluded. The final selection resulted in 28 European IGS stations covering most of the chosen time interval. Their locations are shown in Figure 6.57. The average amount of data gaps for all series remaining after outlier

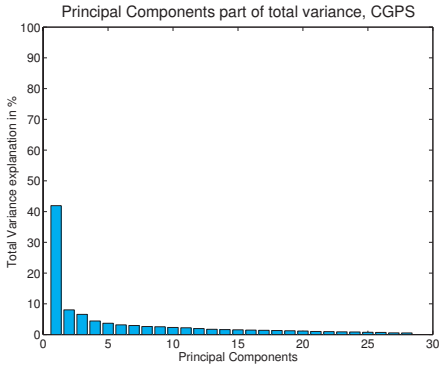

Figure 6.58: The fraction of variance explained by each mode of the European CGPS data

detections is 7.8 percent. Most of the stations are located in the middle and south of Europe, with only a few stations in Scandinavia. Thus, the spatial coverage is not optimal. The eigenvalues for observed vertical land motion from the unweighted residual time series of European CGPS data are shown in Table 6.6. The fraction of overall variance associated with each mode is shown in Figure 6.58. Normalised eigenvector elements (or normalised loading factors) for the first six modes found from the de-trended European CGPS data are shown in Figure 6.59. Temporal variations, illustrated by the time series of amplitudes representing each of the modes are shown in the left-hand side and the Least Squares or the Variance Spectra of these series are shown in the right-hand side of Figure 6.60.

Results

Very similar to the PCA of the Norwegian data, the European data also reveals the existence of a few common modes. The first mode accounts for 42 percent, the second mode 8 percent and the third mode explains 7 percent of the overall sample variance. This is almost identical to the results from the vertical component of the Norwegian CGPS data shown in Figure 6.55. As shown in Table 6.6 all subsequent modes have non-zero variance associated with them. The loadings shown in Figure 6.59 illustrate the geographical pattern for the first six modes. As with the Norwegian data, the first mode corresponds to a common up and

Table 6.6: Eigenvalues from the unweighted residual time series of European CGPS data. j refers to the number of the PC or mode. λ_j is the eigenvalue, Fract. the fraction of variance associated with this mode, and Accum. the fraction of variance explained by the first q modes.

j	λ_j	Fract.	Accum.
1	8.301	0.419	0.419
2	1.592	0.080	0.500
3	1.299	0.066	0.565
4	0.875	0.044	0.609
5	0.731	0.037	0.646
6	0.627	0.032	0.678
7	0.585	0.030	0.708
8	0.518	0.026	0.734
9	0.502	0.025	0.759
10	0.455	0.023	0.782
11	0.438	0.022	0.804
12	0.388	0.020	0.824
13	0.340	0.017	0.841
14	0.319	0.016	0.857
15	0.308	0.016	0.873
16	0.290	0.015	0.887
17	0.275	0.014	0.901
18	0.261	0.013	0.914
19	0.247	0.012	0.927
20	0.225	0.011	0.938
21	0.199	0.010	0.948
22	0.190	0.010	0.958
23	0.174	0.009	0.967
24	0.164	0.008	0.975
25	0.145	0.007	0.982
26	0.137	0.007	0.989
27	0.108	0.005	0.995
28	0.106	0.005	1.000
Sum	19.800		

down motion of the entire network. The second mode describes a tilting of the network over an axis approximately running from NW to SE through the centre

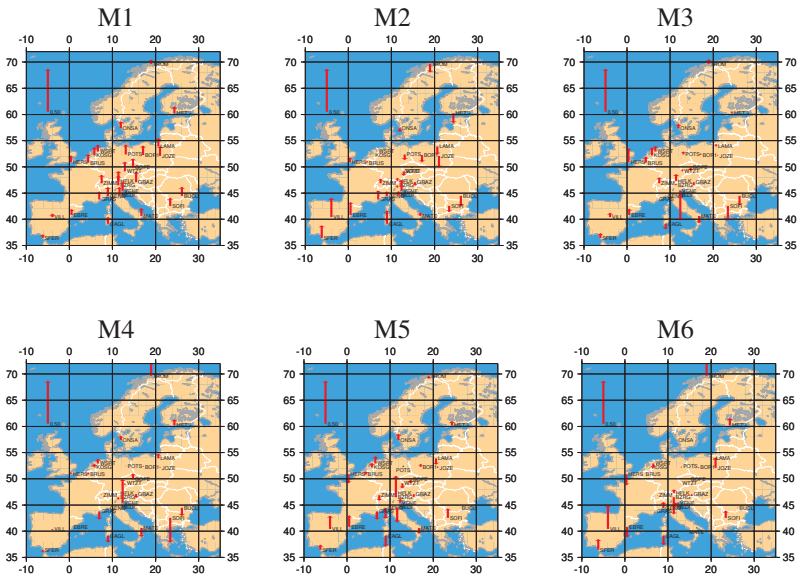

Figure 6.59: Eigenvector elements (normalised loading factors) for the first six common modes in de-trended European CGPS data.

of the Network, while the third mode is a similar motion over an axis running from SW to NE. Concerning the temporal variations of the modes in Figure 6.60, only a few frequencies are dominating, while some of them are found in several modes. For the first mode, which is the most important one considering a variance reduction, a twin peak is found around frequencies representing an annual and a fifteen month cycle. These are closer than the separation limit for the LSS shown in Eq.(4.42) and make a precise determination uncertain. Nevertheless, the fifteen month period is close to the Chandler wobble period in polar motion. However, displacements induced by polar motion are taken into account in the EOP series that is used when processing the CGPS time series. Therefore, the 15 month signal is tentatively associated here with the Fourteen to Sixteen months Oscillation (FSO) identified by Plag [1997] in air pressure and further described by Plag [2004]. A non-linear iterative fit routine as discussed in Subsection 4.2.2 and described in Plag [1988] has been used. Initial values are taken for four fre-

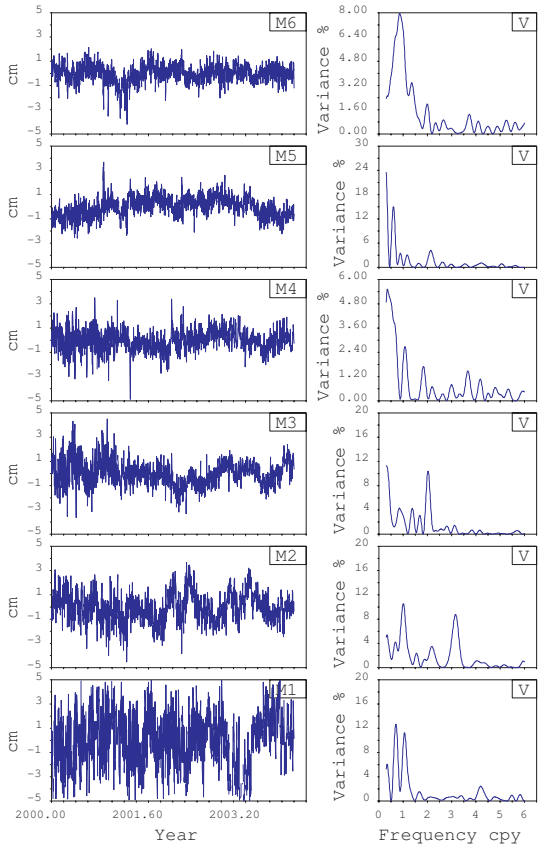

Figure 6.60: The time series of amplitudes representing the first six modes (M1-M6) of the de-trended equally weighted European CGPS data and their Least Squares or Variance Spectra (V).

quencies from the first mode spectra in Figure 6.60. Final results after this fit give frequencies corresponding to periods of 446, 382, 218 and 87 days for the first mode. This confirms the results from the spectral analysis of univariate series of the Norwegian CGPS data in Subsection 6.1.7 as well as for the data set used in Haakonsen et al. [2003], which revealed such periods for several sites. Domi-

nant frequencies are found using the same iteration technique for the other modes. They are presented in Table 6.12 together with the possible variances that can be achieved if the parameters were included in the deterministic model used in the LSSA for the first mode. The spectrum of the second mode shows dominant peaks for frequencies around one, two and three cycles per year which have also been found for the Norwegian data.

An interpretation of the magnitudes of the amplitude series is not performed. As will be shown in Table 6.15, the sample variance for a single series is within the interval 0.5-1.4 cm^2 which corresponds to standard deviations in the interval 0.7-1.2 cm. The first PC or mode is found along the specific chosen direction of maximal variance in the 28-dimensional space with orthogonal axis related to the 28-dimensional points from the original station series. For strong correlations between the 28 variables represented by the station series, the variance of the first mode in a 28-D space is summed up and can be very large. For the European data, the maximum variance is 42 percent out of 19.8 cm^2, respectively. This gives a standard deviation of 2.9 cm. Assuming normally distributed data, this means that one third of all amplitudes will be situated outside the limits of $\pm$2.9 cm as shown for the first mode in 6.60. It may be more interesting is to find out which amplitudes this summed signal will contribute to each station if it is reconstituted to the original time domain. With sample variances for the station series in the interval 0.5-1.4 cm^2, and the first mode contribution of 42 percent, this implies that two thirds of the values in the station series in average will be limited by standard deviations within the interval $\pm$(0.4-0.8) cm. If one assumes that the common mode signal is a kind of average value, this means that most of the reconstituted first mode series have two thirds of their values within the interval: $\pm$0.6 cm. Two examples are plotted in Figure 6.61, the time series with respectively the lowest and largest sample variance (WSRT and VENE) together with a signal composed from the first mode of CGPS data and reconstituted to the the original time domain. In order to visualise the spatial dependency, inter-station correlation coefficients y plotted as a function of distance x are shown together with fitted linear regression models in Figure 6.62. Input series for the correlation estimation are: (1) de-trended European CGPS series (left-hand side), (2) de-trended European CGPS series where the effect from the first mode has been removed (in the

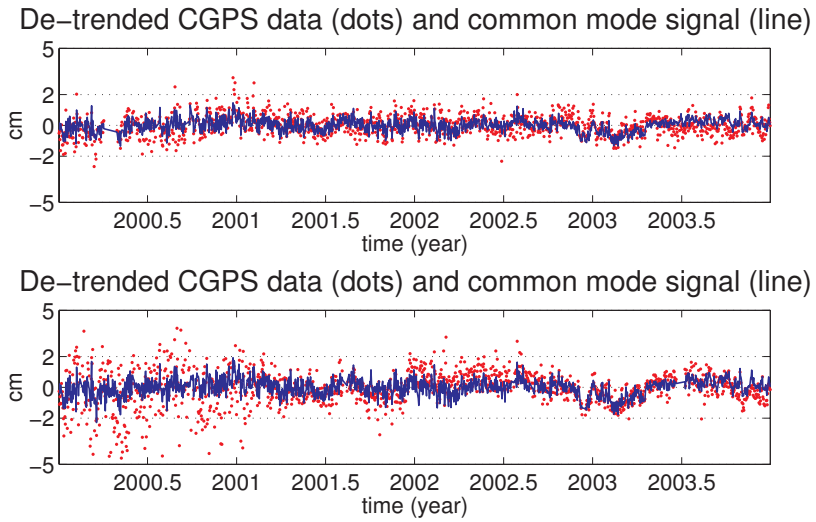

Figure 6.61: De-trended series from CGPS data (dots) and a signal composed from the first common mode (lines) for the stations WSRT (upper plot) and VENE (lower plot)

middle) and (3) de-trended European CGPS series where the effects of the first and second modes have been removed (right-hand side). The fit in the first plot may not be comparable to the results from the Norwegian CGPS data in Figure 6.53 or from the results from Johansson et al. [2002] because the effect of some periodic components were subtracted in the de-trending of these data. However, the fitted regression models happen to be almost identical. In the second plot of Figure 6.62, the effect of the first mode has been removed before the correlation estimation. It shows that most of the spatial correlation is vanished. A further decrease is found removing the effect from the first two modes as in the third plot. To summarise: *If one or more effects are determined and found to explain most of the variation in the first mode from CGPS data, correction for these effects applied on the CGPS time series can considerably decrease their spatial correlation.* Atmospheric surface loading predictions will be investigated as one possible contributing effect. Some more results from the PCA of the European CGPS data set can be found in [Haakonsen et al., 2004b] or [Haakonsen et al., 2004c]

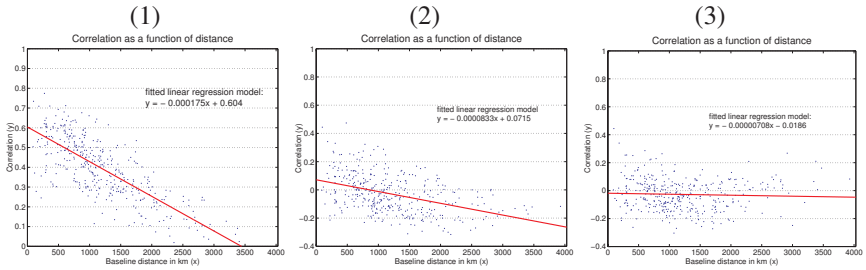

Figure 6.62: Inter-station correlation coefficients y plotted as a function of distance x and fitted linear regression models. The series used for correlation estimation are: (1) de-trended European CGPS series, (2) de-trended European CGPS series where the effect from the first mode has been removed, and (3) de-trended European CGPS series where the effects of the first and second modes have been removed. Note that different ranges are used for the y-axes in second and third plot.

6.3 PCA of Atmospheric Loading Predictions

Five data sets of atmospheric loading predictions of surface displacements have been investigated. A preliminary description is given in Subsection 1.4.3. The five data sets of predictions are denoted as P1 to P5, see Table 6.7. In order to ensure consistency with the residual time series of European CGPS data, atmospheric loading time series of six hour samples were averaged to daily (24h) solutions, for the time interval 2000.0-2004.0, and tested for trends. No significant trends were found in the predictions over the four year interval considered. However, in order to ensure constant zero mean, and for the later comparison with CGPS data, simple regressions have also been performed for these series and the de-trended residual time series are used in the further analyses. It is pointed out that no weights are used in the subsequent analyses, thus the a posteriori variance information from the linear regressions is not used to weight the individual time series.

PCA of atmospheric loading predictions P3

This Section will now only present the results from the PCA using data set P3. The eigenvalues from the PCA of de-trended P3 data are shown in Table 6.8 and

Table 6.7: Models and input data set used for predictions of displacements induced by atmospheric loading. The predictions are denoted as P1 to P5. Sources for air pressure fields are European Center for Medium Range Weather Forecast (ECMWF) and National Center for Environmental Prediction (NCEP). Origin of the reference frames are the Center of mass of the solid Earth (CE) and the Center of Mass of the earth system (CM), which in this case means the solid Earth and the atmosphere. Earth models are PREM [Dziewonski and Anderson, 1981] and G+B [Gutenberg - Bullen, see Farrell, 1972]. Computation methods are SHE: Summation of spherical Harmonic Expansion, and CGF: Convolution of Green's Function and load anomaly. Authors are PG: Pascal Gegout and TvD: Tonie van Dam. Data source is http://www.sbl.statkart.no [cited: 30 October 2004].

Prediction	Input	Ref.Fr.	Earth M.	Comp.	Author
P1	ECWMF	CE	PREM	SHE	PG
P2	ECWMF	CM	PREM	SHE	PG
P3	NCEP	CE	PREM	SHE	PG
P4	NCEP	CM	PREM	SHE	PG
P5	NCEP	CE	G+B	CGF	TvD

illustrated in Figure 6.63. Eigenvalues for modes of order higher than 14 are below 10^{-4} and therefore omitted. Normalised eigenvector elements (or normalised loading factors) for the first six modes are shown in Figure 6.64. The temporal variation, illustrated by the time series of amplitudes representing each of the modes are shown in the left-hand side, and their Least Squares or Variance Spectra are shown in the right-hand side of Figure 6.65. Inter-station correlation coefficients y plotted as a function of distance x are shown in Figure 6.66.

Results

As can be seen in Table 6.8 and Figure 6.63, the PCA for the atmospheric loading predictions only reveal a few dominant modes. The first mode accounts for 76 percent of the overall sample variance, the second for 13 percent and the third 6 percent. Together these three modes explains 95 percent of the total variance, while a signal extracted from the six lowest modes will account for more than 99

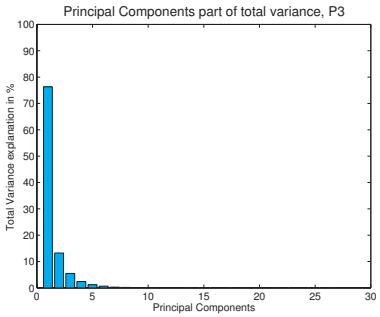

Figure 6.63: The fraction of variance explained by each PC or mode for the atmospheric loading predictions of data set P3.

Table 6.8: Eigenvalues for predicted vertical land motion. Values are for P3 predictions from Table 6.7. j refers to the number of PC or mode. λ_j is the eigenvalue, Fract. the fraction of variance associated with this mode, and Accum. the fraction of variance explained by the first j modes. Note that all eigenvalues for modes with $j > 14$ are below 10^{-4} and therefore omitted.

j	λ_j	Fract.	Accum.
1	2.050	0.763	0.763
2	0.356	0.132	0.895
3	0.147	0.055	0.950
4	0.064	0.024	0.974
5	0.032	0.012	0.986
6	0.018	0.007	0.992
7	0.007	0.003	0.995
8	0.005	0.002	0.997
9	0.003	0.001	0.998
10	0.002	0.001	0.999
11	0.001	0.000	0.999
12	0.001	0.000	0.999
13	0.001	0.000	0.999
14	0.001	0.000	1.000
⋮	⋮	⋮	⋮
Sum	2.686		

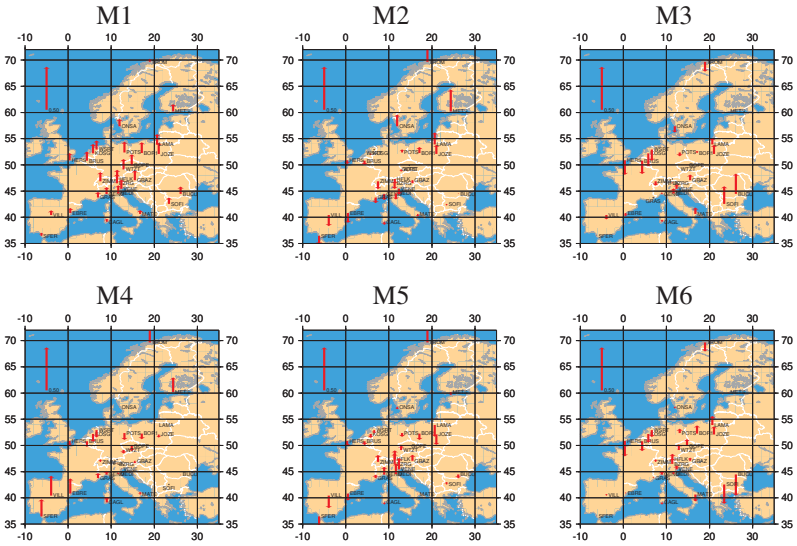

Figure 6.64: Eigenvector elements (normalised loading factors) for the first six common modes in atmospheric loading predictions, P3 in Table 6.7.

percent. Only a small variance information will be lost if the original 28 series are represented by a few modes. The geographical patterns of the loadings for the first six modes are illustrated in Figure 6.64. As for the European CGPS data, the first mode shows a common up and down motion of the whole network. The second mode describes a tilting along a line approximately parallel to the 52°N meridian. The third mode has a geographical pattern which is again very similar to the one found for the third mode in European CGPS data, a tilting of the network along an axis crossing the Alps from SW to NE. The fourth mode shows differences for the stations in Central Europe and the surroundings. The temporal variation of the P3 data illustrated in Figure 6.65 shows a few dominating frequency components for the lower order modes. The spectrum computed from the amplitude time series of the first mode are dominated by periodic components of 541, 402, 149, 115 and 84 days. The largest variance reduction to be achieved modelling one periodic component only is up to 5 percent using the last one. The spectrum of the second mode shows mainly periodic components of 724, 183, 119

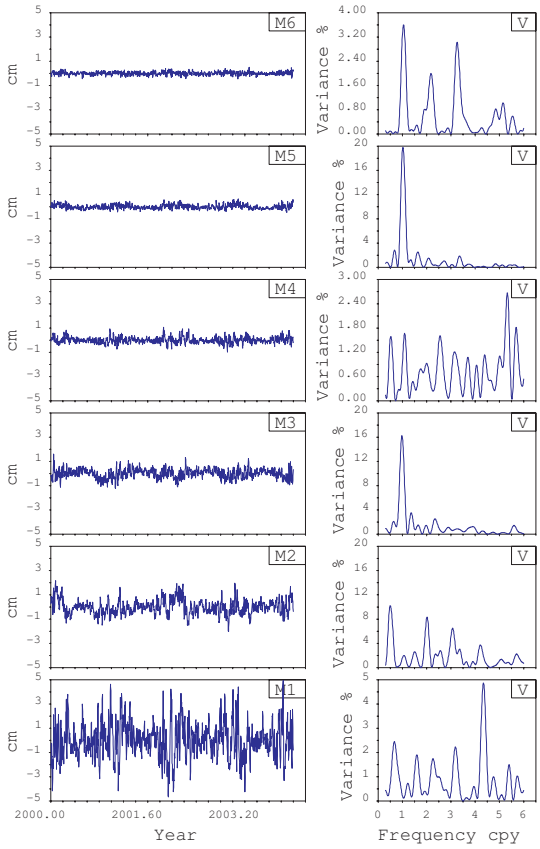

Figure 6.65: The time series of amplitudes representing the first six modes (M1-M6) of the atmospheric loading predictions P3 and their Least Squares or the Variance Spectra.

and 85 days that can account for up to ten percent of the variance part explained by this mode. Surprisingly an annual period was found to be small for the second mode. However, a very large annual peak is visible in the third mode. Detected periods and possible variance reductions if the harmonic constituents are modelled are shown in Table 6.12. The inter-station correlation for the atmospheric loading

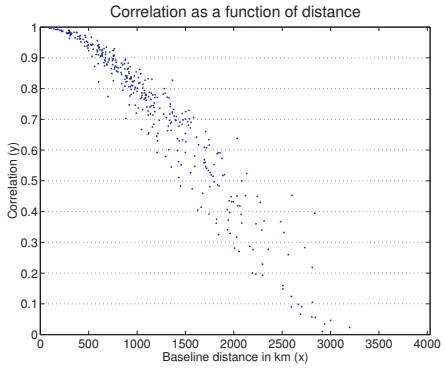

Figure 6.66: Inter-station correlation coefficients y plotted as a function of distance x for the atmospheric loading predictions P3.

predictions P3 is shown in Figure 6.66. These correlations are even higher than for the CGPS data with values of more than 0.3 for distances up to 2000 km. The correlations do not decrease linearly as a function of distance, as the case of CGPS data. A cosine-formed correlation function would be a better choice to describe the spatial correlation of the pressure loading effect and therefore, no linear regression model is performed here.

6.4 Relations between CGPS-data and atmospheric loading predictions

A search will now be made for the data set of predictions that shows most conformity to the de-trended CGPS data. Five separate PCAs were computed for the five sets of atmospheric loading predictions described in Table 6.7. All figures as given for the P3 data in last section will not be presented. An attempt to determine similarities will be made, based on the patterns of spatial correlations and an extraction of a few lower modes that can be reconstituted to the station series and compared with corresponding series from CGPS data. The latter will be performed through regression and correlation analyses of these pairs of station series.

Table 6.9: Correlation of spatial pattern of the common modes in European CGPS data and the five sets of atmospheric loading predictions given in Table 6.7. The correlation coefficients are computed for pairs of eigenvector elements to corresponding modes of European CGPS data and one of the loading predictions. The direction of eigenvectors are found in a maximising procedure of which one of two arbitrary opposite directions can be chosen. The negative signs, at least for the first three modes can thus be neglected. The coefficients are given for the first six modes, only. Coefficients for the other modes are not significant.

No.	P1	P2	P3	P4	P5
1	0.782	0.777	0.788	0.792	0.735
2	-0.707	-0.682	-0.758	-0.734	-0.725
3	0.664	-0.617	0.663	-0.638	0.655
4	-0.085	-0.113	-0.041	-0.070	-0.112
5	-0.180	-0.181	0.222	-0.212	-0.298
6	-0.053	0.057	-0.014	-0.115	0.070

Normalised loadings which are the magnitudes of the eigenvector elements for the European CGPS and atmospheric loading predictions P3 are shown in Figures 6.59 and 6.64. Similar plots could be used to illustrate the geographical pattern of the modes for the data sets P1, P2, P4 and P5. Instead, correlation coefficients have been computed using the eigenvector elements from the first six eigenvectors of the P1-P5 data sets and the first six eigenvectors of CGPS data. The results are shown in Table 6.9. Correlation and regression coefficients as described in Eq.(5.33) have been estimated to relate the different station series of CGPS with atmospheric loading predictions. First the original series have been used. (These series can be seen as a special case where a choice of all $q = 28$ modes are extracted and reconstituted to the stations, see Eq.(5.31)). Station dependent regression coefficients computed between each of the de-trended atmospheric loading data sets P1 to P5 (predictors) and the de-trended CGPS data (observations) are listed in Table 6.10. Computed correlation coefficients for these series are given in Table 6.11. It is not obvious that the second mode of atmospheric loading predictions should look more similar to the second mode of CGPS than for example the third mode would do, or vice versa. However, this seems to be the case. This

Table 6.10: Station dependent regression coefficients for de-trended atmospheric loading predictions (predictors) for the five data sets given in Table 6.7 and European CGPS data (observations). The regression coefficients are dimensionless.

Regression coefficients						
No.	Station	P1	P2	P3	P4	P5
1	BOR1	0.861	0.676	0.943	0.731	0.821
2	BRUS	0.754	0.290	0.842	0.642	0.675
3	BUCU	1.166	0.864	1.191	0.879	1.094
4	BZRG	0.970	0.708	0.879	0.627	0.803
5	CAGL	0.326	0.292	0.538	0.453	0.263
6	EBRE	0.828	0.583	0.795	0.604	0.648
7	GENO	0.627	0.426	0.695	0.492	0.539
8	GOPE	0.817	0.656	0.855	0.705	0.752
9	GRAS	0.424	0.267	0.537	0.341	0.464
10	GRAZ	0.806	0.651	0.826	0.673	0.723
11	HERS	0.644	0.504	0.779	0.619	0.558
12	HFLK	0.630	0.350	0.733	0.438	0.615
13	JOZE	0.928	0.682	0.955	0.707	0.820
14	KOSG	0.799	0.618	0.878	0.678	0.701
15	LAMA	0.675	0.550	0.721	0.587	0.671
16	MATE	0.541	0.381	0.844	0.492	0.728
17	MEDI	0.517	0.345	0.593	0.432	0.530
18	METS	0.579	0.478	0.674	0.525	0.531
19	ONSA	0.607	0.469	0.791	0.548	0.568
20	POTS	0.844	0.672	0.952	0.739	0.825
21	SFER	0.753	0.493	0.672	0.484	0.606
22	SOFI	0.941	0.707	0.922	0.686	0.894
23	TROM	0.816	0.547	0.940	0.596	0.683
24	VENE	0.794	0.718	0.714	0.684	0.679
25	VILL	0.496	0.332	0.494	0.362	0.446
26	WSRT	0.749	0.541	0.890	0.628	0.674
27	WTZT	0.800	0.623	0.845	0.663	0.718
28	ZIMM	0.553	0.393	0.581	0.442	0.570
	MEAN	0.723	0.529	0.788	0.588	0.664

can be seen in the spatial patterns for the first three modes of CGPS data shown in Figure 6.59 and atmospheric loading predictions (represented by data set P3

Table 6.11: Station dependent correlation coefficients for observed and predicted vertical land motion. The atmospheric loading predictions are as given in Table 6.7. The correlation coefficients are computed for pairs of vertical land motion observed by CGPS and one of the five different atmospheric loading predictions. The correlation coefficients are dimensionless.

Correlation coefficients						
No.	Station	P1	P2	P3	P4	P5
1	BOR1	0.464	0.423	0.468	0.420	0.424
2	BRUS	0.356	0.178	0.363	0.309	0.337
3	BUCU	0.464	0.430	0.438	0.400	0.402
4	BZRG	0.340	0.291	0.294	0.243	0.277
5	CAGL	0.071	0.087	0.069	0.094	0.066
6	EBRE	0.227	0.196	0.221	0.198	0.184
7	GENO	0.215	0.179	0.210	0.181	0.192
8	GOPE	0.358	0.337	0.362	0.343	0.328
9	GRAS	0.135	0.105	0.135	0.109	0.137
10	GRAZ	0.334	0.320	0.327	0.310	0.294
11	HERS	0.260	0.228	0.263	0.236	0.254
12	HFLK	0.266	0.174	0.291	0.202	0.263
13	JOZE	0.389	0.336	0.391	0.334	0.339
14	KOSG	0.415	0.364	0.426	0.369	0.387
15	LAMA	0.325	0.309	0.326	0.308	0.315
16	MATE	0.159	0.149	0.167	0.143	0.179
17	MEDI	0.190	0.156	0.208	0.181	0.196
18	METS	0.383	0.362	0.378	0.343	0.357
19	ONSA	0.367	0.325	0.361	0.297	0.350
20	POTS	0.450	0.412	0.449	0.401	0.417
21	SFER	0.199	0.171	0.189	0.170	0.174
22	SOFI	0.270	0.263	0.248	0.235	0.250
23	TROM	0.369	0.289	0.368	0.276	0.332
24	VENE	0.214	0.234	0.191	0.214	0.184
25	VILL	0.143	0.110	0.149	0.123	0.129
26	WSRT	0.416	0.343	0.432	0.348	0.402
27	WTZT	0.340	0.309	0.335	0.304	0.306
28	ZIMM	0.241	0.200	0.251	0.218	0.249
	MEAN	0.299	0.260	0.297	0.261	0.276

in Figure 6.64). This is also confirmed to be the case for all five sets of atmospheric loading predictions because of the large absolute values of the correlation coefficients in Table 6.9. Based on similarities in the spatial patterns for the first three modes, more station dependent regression coefficients are computed. The results are presented in Appendix E. Station series used for these estimations are constructed from signals extracted from each of the first three modes and then reconstituted to the stations, Tables E.2 to E.4 . Finally a signal constructed from a combination of the first three modes is used in Table E.5. Mainly because most of the signal in atmospheric loading predictions is attributed to the first mode, some strange regression coefficients are found combining the station series of the second and third modes. This is probably because all values in some station series are very close to zero. However, the correlation coefficients computed for both sets of station series composed from the extracted first modes in Table E.1 are larger than the coefficients in Table 6.9. This is likely to be due to the lower amount of noise in these CGPS series. Interpreting extracted modes as a signal in CGPS, these series may be considered as filtered versions of the original series with lower noise level, and the correlation coefficients become smaller. In this way the regression coefficient may be a better tool for data comparison than the relative correlation coefficient. The results from these investigations confirm that the best fit between the European CGPS data and the atmospheric loading predictions are achieved using the complete series without extracting any of the modes. It may not be a greate surprise that the atmospheric loading prediction data set P3 seems to give the highest conformity with the European CGPS data, which explains why the results from a PCA of these data are selected in Section 6.3. The choice of P3 can also be argued with the regression coefficients that are closest to one in the third column in Table 6.10. Spectra of the different modes representing both data sets are dominated by a few frequencies summarised for the first three modes in Table 6.12. The frequencies were found in a non-linear fit [see Plag, 1988] of a start model determined from the spectra shown in the Figures 6.60 and 6.65. Some frequency components are found in several modes. The annual constituent is found in most modes both in the CGPS series and the predictions. In general, there appears to be good agreement between the harmonic constituents found in the amplitude functions of modes from both P3 predictions and CGPS data. However, the amplitudes for the constituents in the modes from P3 predic-

Table 6.12: Main harmonic constituents determined in the spectra from time varying amplitudes of European CGPS data in Figure 6.60 and the atmospheric loading predictions P3 in Figure 6.65 representing the first three modes. The period is given in years, variance is the possible variance reduction in percentage that can be achieved if the harmonic constituent is included in the deterministic model.

	CGPS		**P3**	
Mode no.	Period [year]	Variance [%]	Period [year]	Variance [%]
M1	1.222	8.3	1.482	2.4
	1.047	6.6	1.100	1.8
	0.596	1.1	0.407	1.0
	0.238	2.4	0.315	2.2
			0.230	4.9
M2	1.395	3.7	1.982	10.1
	0.986	10.5	1.015	1.4
	0.455	3.3	0.501	8.1
	0.318	8.6	0.326	6.4
			0.234	3.7
M3	1.052	3.5	1.401	1.4
	0.722	3.9	1.019	16.3
	0.579	2.3	0.506	1.4
	0.498	10.1	0.319	0.7
			0.178	1.4

tions are much smaller than in the modes from CGPS, indicating that there must be additional processes contributing to the CGPS data. Moreover, some of the constituents appear to be present in a certain mode of the the P3 predictions but in a different mode of the CGPS data. Thus, there seems to be a kind of cross-mode leakage of harmonic constituents between the P3 predictions and the CGPS data, which again may be due to other, partly correlated processes contributing to the CGPS observations, which are not accounted for in the P3 predictions.

An attempt has been made to examine time varations of computed correlation coefficients between the station series of CGPS and P3 predictions. For a time window of one year, correlation and regression coefficients have been computed in steps of two months. The average correlation coefficients for all stations are

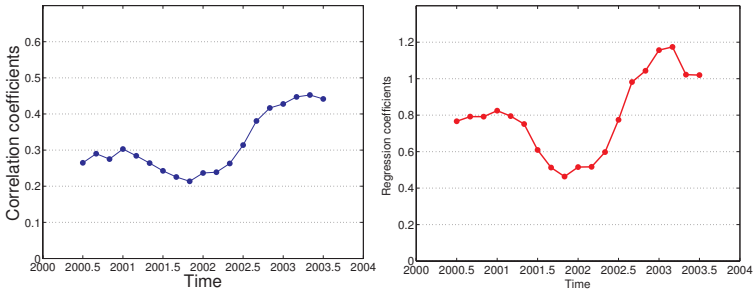

Figure 6.67: Time variations of average correlation (left) and regression (right) coefficients computed from the station series of CGPS and atmospheric loading predictions P3. For a time window of one year, average correlation and regression coefficients are computed in steps of two months.

shown in Figure 6.67. The graph shows an unexpected wave formed increase through time. As can be seen from the time series of amplitudes representing the first mode of CGPS data in Figure 6.60, the pattern of the last year of first the mode is different from the first part. The winter 2003/2004 seems to be rather dominant. This may introduce biases to the correlation estimation. Another explanation can be the improved CGPS data precision over time. The variance is mainly lower for the latest part of the series, which could contribute to an increase in the computed correlation coefficients.

6.5 CGPS corrected for atmospheric loading

Finally PCAs have been performed with de-trended CGPS data corrected for various combinations of atmospheric loading predictions P3. Based on the argument stated in the end of Section 6.2, two assumptions will be :

1. Assuming that an unmodelled part of CGPS data mostly is due to atmospheric surface loading effects, the overall sample variance in such PCA will significantly decrease, compared to the overall sample variance from a

Table 6.13: Various combinations of de-trended European data corrected for atmospheric loading predictions P3

Data	Data description
set 1	De-trended CGPS data.
set 2	De-trended CGPS minus P3 loadings.
set 3	De-trended CGPS minus P3 loadings scaled with the station dependent regression coefficients γ from Table 6.10.
set 4	De-trended CGPS minus the first mode (M1) of P3 loadings scaled with station dependent regression coefficients $\gamma_{(M1)}$ from Table E.2.
set 5	De-trended CGPS minus first and second mode (M1+M2) of P3 loading data scaled with station dependent regression coefficients $\gamma_{(M1+M2)}$.
set 6	De-trended CGPS minus first, second and third mode (M1+M2+M3) of P3 loading data scaled with station dependent regression coefficients $\gamma_{(M1+M2+M3)}$ from Table E.5

PCA with CGPS data only.

2. A successful modelling of contributing effects to the CGPS data will significantly reduce the amount of variance from 42 percent for the first common mode, as the remaining residuals after a correction are closer to independent white noise. In a theoretical view, when all contributing effects and their interactions are identified in the CGPS time series, only independent white noise will appear in the de-trended series. Thus very small correlations between these series can be found and the magnitudes of the eigenvalues from a PCA probably becomes almost equal in size $(\approx 1/28)$.

The different de-trended European CGPS data, corrected for various combinations of atmospheric loading predictions P3 for the time interval 2000.0-2004.0 for this investigations are given in Table 6.13. The results from the PCAs are given in Table 6.14. Computed sample variances for each station series and the total variance reduction compared to data set 1 are shown in Table 6.15. Finally plots illustrating the inter-station correlation for data sets 2 and 3 are shown in Figure 6.68. The largest variance reduction is achieved in the data set corrected with the original P3 data and scaled with the regression coefficients relating them to CGPS data. A variance reduction up to ten percent is possible. Correcting the CGPS data

Table 6.14: PCAs separately performed for the six different data sets given in Table 6.13. Columns for each data set are: Eigenvalues (eig.val), variance explanation (var.) and accumulated part of total variance (acc.) explained by the first six modes. Var.red. is the variance reduction compared to the sum of eigenvalues in the first data set

Data ⇒	set 1			set 2			set 3		
No	eig.val	var.	acc.	eig.val	var.	acc.	eig.val	var.	acc.
1	8.301	0.419	0.419	7.024	0.388	0.388	6.947	0.387	0.387
2	1.592	0.080	0.500	1.510	0.083	0.471	1.461	0.081	0.468
3	1.299	0.066	0.565	1.204	0.067	0.538	1.208	0.067	0.536
4	0.875	0.044	0.609	0.839	0.046	0.584	0.824	0.046	0.581
5	0.731	0.037	0.646	0.725	0.040	0.624	0.718	0.040	0.622
6	0.627	0.032	0.678	0.608	0.034	0.658	0.607	0.034	0.655
:	:			:			:		
Sum	19.800			18.103			17.939		
Var.red.				8.6%			9.4 %		
Data ⇒	set 4			set 5			set 6		
No	eig.val	var.	acc.	eig.val	var.	acc.	eig.val	var.	acc.
1	6.942	0.377	0.377	6.947	0.380	0.380	6.971	0.384	0.384
2	1.581	0.086	0.463	1.461	0.080	0.460	1.485	0.082	0.466
3	1.298	0.071	0.533	1.302	0.071	0.531	1.213	0.067	0.532
4	0.876	0.048	0.581	0.882	0.048	0.579	0.838	0.046	0.579
5	0.729	0.040	0.621	0.732	0.040	0.619	0.728	0.040	0.619
6	0.623	0.034	0.654	0.610	0.033	0.653	0.612	0.034	0.652
:	:			:			:		
Sum	18.413			18.290			18.162		
Var.red.	7.0			7.6%			8.3%		

with atmospheric loading predictions will only slightly reduce the correlation as a function of distance. For the atmospheric loading predictions P3 a reduction from 0.604 to 0.555 for the constant term of the linear regression equation (the correlation coefficient at zero distance) was achieved. As argued in the end of Section 6.4, the large amount of variance attributed to the first mode of CGPS data after the corrections for atmospheric loading predictions, probably means that other unknown contributing processes and their interactions are present in the

Table 6.15: Sample variances for station series composed from CGPS and five different sets of mode combinations with atmospheric loadings (P3) data given in Table 6.13. tot.sum is the column sums, var.red. is total variance reductions compared to the first column.

	Data⇒	set 1	set 2	set 3	set 4	set 5	set 6
No.	Station	Var1	Var2	Var3	Var4	Var5	Var6
1	BOR1	0.574	0.449	0.448	0.468	0.461	0.464
2	BRUS	0.602	0.525	0.522	0.560	0.550	0.534
3	BUCU	0.720	0.585	0.581	0.635	0.630	0.609
4	BZRG	0.947	0.867	0.864	0.863	0.858	0.859
5	CAGL	0.792	0.791	0.788	0.810	0.799	0.795
6	EBRE	0.754	0.720	0.717	0.724	0.717	0.713
7	GENO	0.682	0.658	0.652	0.643	0.649	0.650
8	GOPE	0.746	0.651	0.648	0.658	0.658	0.659
9	GRAS	0.682	0.678	0.669	0.672	0.672	0.672
10	GRAZ	0.721	0.648	0.644	0.641	0.643	0.643
11	HERS	0.730	0.683	0.679	0.715	0.701	0.687
12	HFLK	0.709	0.657	0.648	0.654	0.652	0.650
13	JOZE	0.982	0.832	0.832	0.875	0.843	0.851
14	KOSG	0.498	0.410	0.407	0.443	0.440	0.423
15	LAMA	0.757	0.689	0.676	0.682	0.676	0.679
16	MATE	0.617	0.600	0.599	0.614	0.614	0.604
17	MEDI	0.673	0.657	0.643	0.637	0.644	0.644
18	METS	0.500	0.445	0.428	0.448	0.431	0.432
19	ONSA	0.442	0.389	0.385	0.377	0.398	0.394
20	POTS	0.586	0.468	0.468	0.493	0.487	0.481
21	SFER	0.488	0.475	0.471	0.495	0.482	0.481
22	SOFI	0.995	0.934	0.933	0.944	0.947	0.931
23	TROM	0.681	0.589	0.589	0.662	0.643	0.638
24	VENE	1.397	1.354	1.345	1.339	1.352	1.348
25	VILL	0.778	0.780	0.761	0.786	0.770	0.770
26	WSRT	0.428	0.349	0.348	0.377	0.379	0.363
27	WTZT	0.751	0.670	0.666	0.664	0.664	0.662
28	ZIMM	0.565	0.548	0.529	0.535	0.530	0.526
	tot.sum	19.800	18.103	17.939	18.413	18.290	18.162
	var.red.		8.6%	9.4%	7.0%	7.6%	8.3%

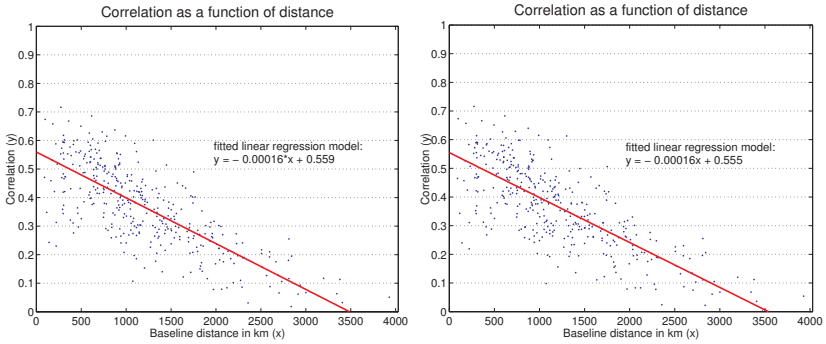

Figure 6.68: Inter-station correlations, CGPS corrected for P3 data only (left), and CGPS corrected for P3 data scaled with regression coefficients from the P3 column of Table 6.10 (right)

CGPS data. Another question would be whetter these biases are really the result of geophysical processes or effects due to satellite orbit modelling ?

6.6 Variance explained by periodicity in CGPS data

Determination of the amplitudes and sample variance reductions, caused by the effect of the annual and fifteen months period in the European CGPS data will be performed. As can be seen in Figure 6.60, the first mode of de-trended CGPS data are dominated by two frequencies, determined to the periods 382.4 and 446.3 days. Parameters representing these frequencies will be highly correlated, and their estimates rather uncertain, using a least squares fit with a time series of four years as observations. 382.4 days may thus be the annual period of 365.25 days that is expected for the CGPS data. 446.3 is quite close to the Chandler period of 435 days that should have be corrected for in the pre-processing stage of CGPS time series. Nevertheless a similar period is visible in the data. In order to make a picture of their order of magnitude on each station, spectral parameters representing the pairs 382.4 and 446.3 days, and 365.25 and 435 days respectively have been included in the weighted regressions used for a de-trending of each station series of CGPS data. Amplitudes are computed as in Eq.(2.22). For the remaining residual time series after the regressions, the weight information is omitted and

sample variances have been estimated for each station. The results are shown in Table 6.16. The columns are (1): station number, (2): station name, (3): sample variances for the series of unweighted residuals when parameters for constant, linear trend and jumps are included in the deterministic model. (4): sample variances for the series of unweighted residuals using parameters as in column (3) and two periodic components of 446.3 and 382.4 days. (5) and (6): Computed amplitudes in cm for the periodic components of 446.3 and 382.4 days, respectively from a weighted regression with parameters as in (4). (7): sample variances for the series of unweighted residuals using parameters as in column 3 and two periodic components of 435.0 and 365.25 days. (8) and (9): Computed amplitudes in cm for the periodic components of respectively 435.0 and 365.25 days from a weighted regression with parameters as in (7). tot.sum = Total sums of sample variances for all series. var.red. = Total variance reduction compared to the first column. The results show that a variance reduction of 12-13 percent is possible including these periodic components in the regressions. The amplitudes are found to be significant for most of the sites. Magnitudes up to 0.8 cm appear for some stations. Finally, it should be pointed out that the 446 days period should not be interpreted as the Chandler period. It inserted in Table 6.16 to show the order of magnitude for amplitudes for an interpretation in this frequency band.

Table 6.16: Sample variances (unit: cm^2) and computed amplitudes (unit: cm) for various de-trended European CGPS station series when spectral parameters for two periodic components are included in the weighted regressions.

No.	Station	var.	var.	$A_{446.3}$	$A_{382.4}$	var.	$A_{435.0}$	$A_{365.25}$
1	BOR1	0.574	0.475	0.46	0.53	0.486	0.31	0.48
2	BRUS	0.602	0.518	0.35	0.48	0.537	0.19	0.39
3	BUCU	0.720	0.482	0.68	0.80	0.501	0.47	0.72
4	BZRG	0.947	0.817	0.67	0.43	0.825	0.52	0.45
5	CAGL	0.792	0.758	0.33	0.19	0.761	0.31	0.17
6	EBRE	0.754	0.741	0.13	0.13	0.743	0.12	0.10
7	GENO	0.682	0.605	0.44	0.47	0.615	0.31	0.42
8	GOPE	0.746	0.638	0.65	0.28	0.635	0.61	0.32
9	GRAS	0.682	0.620	0.44	0.31	0.625	0.37	0.32
10	GRAZ	0.721	0.536	0.65	0.71	0.559	0.47	0.62
11	HERS	0.730	0.669	0.38	0.19	0.679	0.34	0.12
12	HFLK	0.709	0.567	0.36	0.71	0.585	0.15	0.55
13	JOZE	0.982	0.768	0.64	0.81	0.762	0.44	0.77
14	KOSG	0.498	0.432	0.29	0.44	0.435	0.16	0.40
15	LAMA	0.757	0.655	0.48	0.49	0.652	0.46	0.44
16	MATE	0.617	0.545	0.43	0.35	0.546	0.37	0.33
17	MEDI	0.673	0.610	0.47	0.43	0.622	0.34	0.38
18	METS	0.500	0.410	0.28	0.50	0.407	0.18	0.47
19	ONSA	0.442	0.406	0.26	0.33	0.408	0.16	0.30
20	POTS	0.586	0.456	0.53	0.56	0.468	0.40	0.50
21	SFER	0.488	0.472	0.12	0.21	0.472	0.06	0.19
22	SOFI	0.995	0.857	0.46	0.65	0.853	0.31	0.60
23	TROM	0.681	0.588	0.39	0.51	0.584	0.30	0.47
24	VENE	1.397	1.280	0.72	0.40	1.290	0.64	0.40
25	VILL	0.778	0.761	0.16	0.09	0.760	0.17	0.10
26	WSRT	0.428	0.411	0.22	0.18	0.415	0.16	0.15
27	WTZT	0.751	0.615	0.58	0.59	0.626	0.46	0.53
28	ZIMM	0.565	0.504	0.43	0.40	0.512	0.32	0.36
	tot.var.	19.800	17.196			17.363		
	var.red		13.2%			12.3%		

Chapter 7

Conclusions and recommendations

7.1 Conclusions

During past few years, the computation of CGPS-time series of daily solutions has been customary in most European countries that have participated in the project within the EUREF Permanent Network (EPN). The exploration of the contributing processes to CGPS-time series is still incomplete. Up to now, most investigations have started by analysing univariate series with the purpose of characterising local or site dependent effects. However, the influences and interactions from other effects make this characterisation difficult.

The main purpose of this thesis was to study preliminary CGPS time series in order to determine the possible contributory influence from this data, especially that of a periodic kind. In the first part of numerical investigations, univariate series from the Norwegian network have been carefully examined. An attempt to remove detected and interpreted effects have been made. After the de-trending the remaining series have been normalised to compose an input multivariate data set to perform a Principal Component Analysis (PCA). In the second part of the investigations, time series from a European network of CGPS stations have been analysed. Another approach has been used for these data. The goal of the PCA is to find a transformation in the orthogonal multi-dimensional space, to another orthogonal set of axis, that identifies the directions of maximal variances. The components representing these directions are denoted as the principal components

or modes. Through this orthogonal transformation, the original time series can be exactly represented by a new set of series, namely the time series of amplitudes representing the modes. A complete reconstruction from amplitudes to original series is possible with the inverse orthogonal transformation. It is assumed that the global or regional effects give the largest contributions to the overall variance of all series. Selecting only a few lower order modes for a reconstitution of the series to the original system of orthogonal axis will bring out these effects. The investigations of the European network of CGPS data has concentrated on the regional effects found in the lower order modes. Local effects are assumed to appear in the higher order modes.

7.1.1 Univariate analyses of Norwegian CGPS series

The study of the Norwegian CGPS data has been performed in two steps. First, univariate time series of CGPS data, representing each site and direction (N, E, H-components) have been investigated with a focus on the spectral analysis and jump detection techniques. In a second stage, the remaining de-trended series of residuals have been normalised. These series are then combined to multivariate series, and used as input to a PCA. Simulations as well as investigations of the univariate CGPS-time series, have shown that the detection of unknown jumps is a great challenge. Jump or offset estimates may be largely correlated with other parameter estimates. Wrong decisions in the jump detection phase, may affect other parameter estimates and are of vital importance for correlation function and spectrum estimation. Three different methods of ACF-estimation have been investigated. They gave almost identical results for the investigated time series. Estimated auto correlation functions for normalised residuals remaining after weighted linear regressions with final models, show small values, also for lower lags. Thus, CGPS residual time series may be interpreted as realisations from first order auto regressive processes with small parameter values which are also known as Markov-processes and are not very different from white noise processes. Tests of normality show that the normalised residuals from each station series show similarity with normally distributed data with zero mean and variance one. Cross correlation coefficients between time series of normalised residuals have been computed and tested. The investigations reveal very small or insignifi-

cant correlation coefficients between the components for each site. This probably arises from the fact that the CGPS data are daily (24h) samples, covering almost exactly two complete cycles for the GPS satellites. Thus all possible satellite geometries have been repeated twice, which may explain that high correlations between components for shorter GPS series are more related to the geometry. This statement gives an opportunity to treat the components of each station individually, which is a major advantage for the time consuming computations. On the other hand, corresponding components for different stations show significant spatial correlation up to 0.7 for nearby stations. This correlation is only slightly decreasing with distance and indicates one or more common effects for the entire network in Norway. Considering these facts, it was decided to make a multivariate analyses of CGPS data.

7.1.2 Multivariate analyses

The time series of residuals remaining after de-trending, using weighted regressions, have been combined in two different ways for the computation of sample covariance matrices required for the PCAs of the Norwegian and European networks.

The CGPS data series from the Norwegian network include several intervals of missing data, and they are of fairly different lengths. This introduces a scale factor problem for the estimation of variances and covariances that makes a consistent mutivariate analysis problematic. In order to avoid this possible error, the residuals, normalised with a posteriori standard deviations from the weighted regressions have been used as input to separate PCAs for the three directions of the Norwegian data. Using normalised data, the covariance matrix becomes identical to the correlation matrix. Correlation matrices have all diagonal elements equal to one which ensures that the stations become equally weighted in a PCA. This disregards the different variability of each station series and introduces another scale problem. After a comparison of RMS-values of each single series, the latter solution was considered to give a smaller influence and was chosen for the Norwegian data. The results from the PCAs of Norwegian data show that more than 50 percent of the total variance is explained by the first mode for the N-direction. For

the E- and H-directions, respectively, the first mode explains 43 and 44 percent. More than 60 percent of the total variance is explained by the two first modes for the N-direction and almost the same for the other components. This indicates at least one or two dominating common effects or factors that contribute to the series for all stations and directions. The first loadings show a common variation for the entire network of Norway that is almost uniform for all directions. The second loadings show one or more local effects which are different for the southern and northern parts for all directions. An axis from north-west to south-east north of Trondheim separates the regions. The third loadings indicate a possible local effect for stations in the mountain areas. This is visible in the height direction, but the few number of stations in the mountain area makes this interpretation uncertain. It was decided to analyse an independently pre-processed data set from a European network to confirm the results from the analysis of Norwegian data.

In the PCA for European CGPS data, only the vertical component has been used. Unlike the Norwegian case, the a posteriori variance information from the weighted linear regressions has not been used to weight the de-trended series in subsequent analyses. A sample covariance matrix from the station series of residuals in the common time interval (2000.0-2004.0) has been computed. This analysis confirms the results from the Norwegian network. The first, second and third mode respectively, represent 42, 8 and 7 percent of the overall sample variance. The spatial pattern of normalised loadings representing the first mode shows a common up and down motion for the entire network. The spectrum of time varying amplitudes of the first mode shows approximately dominant peaks for an annual period and a period close to the Chandler period. The annual variation is expected and is attributed to seasonal variations. On the other hand, the period close to the Chandler period was unexpected, as this period should be taken into account in the pre-processing stage of the CGPS time series. The investigation has shown that including parameters for these periods in the regression models make it possible to reduce the average sample variance for the stations by 12-13 percent. Magnitudes of amplitudes computed from these periodic components found in a simultaneously estimation are up to 8 mm for the annual and 6 mm for the Chandler-like period, respectively. The spatial pattern of the second mode shows a tilting over a NW to SE axis through the centre of the Network, while the pattern represent-

ing the third mode is a similar motion over an axis running from SW to NE. The spectrum of time varying amplitudes of second mode is dominated by frequency components approximately of 1, 2 and 3 cycles per year, while spectra representing almost all the other modes includes semi-annual and annual periods. The computation of inter-station correlation coefficients shows that the spatial correlation in CGPS data only decreases slightly as a function of distance. This confirms earlier results from similar investigations of the height component for the Swedish and Finnish network performed by Johansson et al. [2002]. This correlation will vanish after a removal of the effect from first mode. A determination of possible effects contributing to each mode has proved to be difficult as several effects and interactions may appear in different modes. However, an attempt has been made to relate CGPS data to five different data sets of atmospheric surface loading predictions.

A comparison has been made of station series as well as extracted modes from separate PCAs for the CGPS and the atmospheric loading predictions. The spatial patterns of the first three modes are quite similar. The spectra of time varying amplitudes representing the modes also look quite similar, except for a dominant double peak in the first mode of the CGPS data. The variance explained by the first three modes for one selected data set of predictions, data set P3 is 76, 13 and 6 percent, thus the first three modes account for 95 percent of the overall variance. Model relations for series of CGPS and atmospheric loading predictions are searched using regression and correlation techniques. The best fitted data set of predictions was selected for further examination. Different corrections for this atmospheric loading effect on CGPS data using the derived models have been performed. Investigations show that a total variance reduction of CGPS data up to 10 percent is possible when correcting for the effect of atmospheric loading. Some investigations of the time variability for computed correlation and regression coefficients have used a “sliding” one-year time window in two-month steps through the series. The precision of CGPS data is better for the last part of the time interval, thus a correlation coefficient is assumed to increase with time. Results also confirm this, but it also shows an unexpected wave form with a minimum value at the turn of the year 2001/2002.

7.2 Further investigations

Using longer (future) series it would be of interest to find out more about the time varying relation between estimated correlation and regression coefficients. Here, one must ask if this really is a wave-formed relation in time, or if the results in this study are related to the present pre-processing technique of CGPS-time series and the handling of the reference frame shift from ITRF97 to ITRF2000. However, the results confirm that a discussion of the homogeneity of CGPS time series should continue. In this context the time series of varying amplitudes for the first mode of CGPS should be mentioned as it shows a very significant effect of the winter 2003/2004 in the first mode.

Most data sets from CGPS used in this investigation show an unexpected signal close to the Chandler period. EOP time series are used in the pre-processing stage of CGPS time series to account for the effect of the Chandler period. Further investigations of this signal should be made. The spatial correlation of CGPS data is mainly attributed to the first mode. An explanation of the contributing effects to the first mode becomes essential in the near future. Until these effects are determined, the PCA has proved to be a good alternative when used as a filtering technique for improvement of the data series in a region.

In the final part of this study, PCA is used as a tool to separately determine the common modes in the CGPS data and the atmospheric loading predictions that are valid for the particular Euro-Asian techtonic plate. The effect of atmospheric loadings may probably appear different for other continents. Newertheless, the determination of corresponding relations between common modes in CGPS time series and atmospheric or other loading effects for other techtonic plates, would be of valuable interest. So far the PCA method has not shown any limitations in use even for larger scales. Future studies may focus on PCA for both hemispheres or even for the entire globe. The goal of modelling all displacements on the Earth's surface with proper accuracy in near real-time is being sought for several purposes. Improved monitoring of displacements related to the Earth's surface will contribute to a better determination of precise reference frames. As such work would also cover the most unstable areas, the prediction of natural dis-

asters such as earthquakes and volcanic activity would be most valuable. For such applications it can be relevant to use the PCA filtering technique as a tool in the monitoring of the Earth. PCA can be used to extract a common mode signal for a region. Reconstituted to each CGPS station time series this can be seen as a filtered version of the original CGPS station series. Reconstructing the signal from the lowest modes gives global or regional effects. Equivalently, a signal reconstructed from a sum of the highest modes would probably represent local effects. In a combination with a prediction and filtering approach as in Koch [1999], the predicted values in near future as well as their confidence interval limits can be estimated. Later observed values outside these limits will indicate an unusual local motion and might give a warning in possible earthquake areas.

References

M. Abbasi. Comparison of Fourier, Least Squares and Wavelet Spectral Analysis Methods, Tested on Persian Gulf Tidal Data. Master's thesis, Surveying Engineering Department, Civil Engineering Faculty, K.N.Toosi University of Technology, Teheran, Iran, 1999.

Z. Altamimi, P. Sillard, and C. Boucher. A new release of the International Terrestrial Reference Frame for earth science applications. *J. Geophys. Res.*, page 2214, 2002. doi: 10.1029/2001JB000561.

P.H. Andersen. Realization of terrestrial and celestial reference systems using space geodedic observations. *Nordic space activities*, pages 6–9, 2002. ISSN:0805-7397.

G. Blewitt. Self-consistency in reference frames, geocenter definition, and surface loading on the solid earth. *J. Geophys. Res.*, 2002. 10.1029/2002JB002082.

George E.P. Box, Gwilym M. Jenkins, and Gregory C. Reinsel. *Time series analysis : forecasting and control*. Englewood Cliffs, N.J. : Prentice Hall, third edition, 1994. 0-13-060774-6.

Michael R. Craymer. *The Least Squares Spectrum, its Inverse Transform and Auto correlation Function: Theory and some Applications in Geodesy*. PhD thesis, Department of Civil Engeneering, University of Toronto, Canada, 1998.

A. M. Dziewonski and D. L. Anderson. Preliminary reference Earth model. *Phys. Earth Planet. Int.*, 25:297–356, 1981.

William J. Emery and Richard E. Thomson. *Data analysis methods in physical*

oceanography. Amsterdam : Elsevier, 2 edition, 2001. 0-444-50757-4, 0-444-50756-6.

W. E. Farrell. Deformation of the Earth by surface loads. *Rev. Geophys. Space Phys.*, 10:761–797, 1972.

T.A. Haakonsen and H. Nahavandchi. Error analysis of GPS/SATREF data. Presentation at Geodesi og Hydrografidagene, Hønefoss, Norway, 2003.

T.A. Haakonsen and H. Nahavandchi. Correlation and spectral analysis of the Norwegian permanent GPS stations using a Least-squares procedure. Presentation in European Geosciences Union, Nice, France, 2004.

T.A. Haakonsen, H. Nahavandchi, and H-P. Plag. Determination of the Error-Spectrum of Continuous GPS Observations. Presentation in EGS-AGU-EUG Joint Assembly, Nice, France, 2003.

T.A. Haakonsen, H. Nahavandchi, and H-P. Plag. Principal Component Analysis of the Norwegian permanent GPS stations. Presentation in European Geosciences Union, Nice, France, 2004a.

T.A. Haakonsen, H-P. Plag, and H.P. Kierulf. Common modes in GPS-determined vertical land motion in Europe and their relation to atmospheric loading. *J. Geophys. Res.*, 2004b. Submitted.

T.A. Haakonsen, H-P. Plag, H.P. Kierulf, and G. Blewitt. Improving reference frame stability by modelling common modes of surface displacements using Empirical Orthogonal Functions. Presentation AGU Fall Meeting, San Francisco, USA, 2004c.

M. B. Heflin, D. Argus, F. Jefferson, F. Webb, and J.F. Zumberge. Comparison of a GPS-defined global reference frame with ITRF2000. *GPS Solutions*, 6 No 1-2:72–75, 2002.

M. B. Heflin, W. I. Bertiger, G. Blewitt, A. P. Freedman, K.J. Hurt, S.M. Lichten, U. J. Linquister, Y. Vigue, F. H. Webb, T.P. Yunck, and J.F. Zumberge. Global geodesy using GPS without fiducial sites. *J. Geophys. Res.*, 19:131–134, 1992.

M.B. Heflin. GPS time series. http://sideshow.jpl.nasa.gov/mbh/series.html, 2004.

K. B. Howell. *Principles of Fourier analysis*. Chapman and Hall/CRC, 2001. 0-8493-8275-0.

J.M. Johansson, J.L. Davis, H.-G. Scherneck, G.A. Milne, M. Vermeer, J.X. Mitrovica, R.A. Bennett, B. Jonsson, G. Elgered, P. Elosegui, H. Koivula, M. Poutanen, B.O. Ronnang, and I.I. Shapiro. Continuous GPS measurements of postglacial adjustment in Fennoscandia. *J. Geophys. Res.*, 2002.

Richard A. Johnson and Dean W. Wichern. *Applied Multivariate Statistical Analysis, Fourth edition*. Prentice Hall, Upper Saddle River, New Jersey, 1998.

Benjamin Kedem and Konstantinos Fokianos. *Regression models for time series analysis*. Wiley series in probability and statistics. Hoboken, N.J. : Wiley-Interscience, 2000. 0-471-36355-3.

K.-R. Koch. *Parameter estimation and hypothesis testing in linear models*. Springer-Verlag, Berlin, 1999. 2nd, updated and enl. ed.

Erwin Kreyszig. *Advanced engineering mathematics / Erwin Kreyszig. - 8th ed.* New York : Wiley, 1999. ISBN:0-471-33328-x (h.), 0-471-15496-2 (ib.).

A. Leick. *GPS satellite surveying*. Hoboken, N.J. : Wiley, 2004. 0-471-05930-7.

E.N. Lorenz. Empirical orthogonal functions and statistical weather prediction. Sci. rep. no. 1, statistical forecasting project., MIT, Cambridge, MA, USA., 1956.

D. D. McCarthy and G. Petit. Iers conventions (2003). IERS Technical Note 32, International Earth Rotation and Reference Systems Service, Central Bureau, Bundesamt fur Kartographie und Geodasie, 2004. ISBN:3-89888-884-3.

D.E. Newland. *Random vibrations, spectral and wavelet analysis*. Prentice Hall, 1993. Third edition.

H.-P. Plag. A regional study of Norwegian coastal long-period sea-level variations and their causes with special emphasis on the Pole Tide. *Berl. Geowiss. Abhandl. Reihe A*, 14:1–175, 1988.

H.-P. Plag. Chandler wobble and pole tide in relation to interannual atmosphere-ocean dynamics. In H. Wilhelm, W. Zürn, and H.-G. Wenzel, editors, *Tidal Phenomena*, number 66 in Lecture Notes in Earth Sciences, pages 183–218. Springer, 1997.

H-P. Plag. The IGGOS as the backbone for global observing and local monitoring: a user driven perspective. *J. Geodynamics*, 2004. in press.

A. J. Pope. The statistics of residuals and the detection of outliers. NOAA Technical Report NOS 65 NGS 1 65, U.S. Department of Commerce, Rockville, USA., 1976.

W. H. Prescott. Will a continuous GPS array for L. A. help earthquake hazard assessment? *EOS, Trans. Am. Geophys. Union*, 77:417, 1996a.

W. H. Prescott. Yes: the L. A. array will radically improve seismic risk assessment. *EOS, Trans. Am. Geophys. Union*, 77:419 & 427, 1996b.

William H. Press, William T. Vetterling, Saul A. Teukolsky, and Brian P. Flannery. *Numerical recipes in FORTRAN*. Cambridge University Press, New York, USA, 1992. ISBN:0-521-43064-X.

M.B. Priestley. *Spectral Analysis and Time Series*, volume 1. Academic Press Inc., London, 1981. 0-12-564901-0.

J. C. Savage. No: the L. A. array is not ready for prime time. *EOS, Trans. Am. Geophys. Union*, 77:419 & 427, 1996.

H.-G. Scherneck, J. M. Johansson, and R. Haas. BIFROST project: Studies of variations of absolute sea level in conjunction with postglacial rebound in Fennoscandia. In R. Rummel, H. Drewes, W. Bosch, and H. Hornik, editors, *Towards an Integrated Global Geodetic Observing System (IGGOS)*, volume 120 of *International Association of Geodesy Symposia*, pages 241–244. Springer, Heidelberg, 2000.

H.-G. Scherneck, J. M. Johansson, J. M. Mitrovica, and J. L. Davis. The BIFROST project: GPS detremined 3-D displacements in Fennoscandia from 800 days of continous observations in the SWEPOS network. *Tectonophysics*, 294:305–321, 1998.

Robert H. Shumway and David S. Stoffer. *Time series analysis and its applications*. Springer texts in statistics. Springer, New York, 2000. 0-387-98950-1.

K.D. Smith, D. von Seggern, G. Blewitt, L. Preston, J.G. Anderson, B.P. Wernicke, and Davis J.L. Evidence for deep magma injection beneath Lake Tahoe, Nevada-California. *Science*, 2004.

P.J.G. Teunissen. *Adjustment theory*. Delft University Press, 2000. 90-407-1974-8.

T. van Dam. Effects to topographic variability on estimated atmospheric pressure loading effects. *Journal of Geodesy*, 2004, in review.

T. van Dam, H-P. Plag, O. Francis, and P. Gegout. GGFC, Special Bureau for Loading: current status and plans. In B.Richter, W.Schwegman, and R.Dick, editors. *Proceedings of the IERS Global Geophysical Fluid Center Workshop, Munich, November 20-21, number 30 in International Earth Rotation and Reference System Service, IERS Technical Note, pages 180-198*, 2003.

Petr. Vanicek and Edward J. Krakiwsky. *Geodesy, second edition*. Elsevier Science Publishers B.V., Amsterdam, The Netherlands, 1986. ISBN:0-444-87775-4.

R.E. Walpole and R.H. Myers. *Probability and Statistics for Engineers and Scientists*. Prentice-Hall International, Inc., New Jersey, USA, $5^{textrmth}$ edition, 1993. ISBN:0-13-244070-9.

William W. S. Wei. *Time series analysis*. Addison-Wesley Publishing Company, USA, 1990.

D. Wells, P. Vanícek, and S. Pagiatakis. *Least-squares spectral analysis revisited*. Fredericton, N.B. : Department of Surveying Engineering,University of New Brunswick, 1985.

G. Barrie Wetherill. *Regression Analysis with Applications*. Chapman and Hall, London, 1986.

J. F. Zumberge, M. B. Heflin, D. C. Jefferson, and M. M. Watkins. Precise point positioning for the efficient and robust analysis of GPS data from large networks. *J. Geophys. Res.*, 102:5005–5018, 1997.

Appendix A

Law of covariances

The following linear equations have been defined to explain error propagation

$$\boldsymbol{y_1} = \boldsymbol{A_1}\boldsymbol{x_1} + \boldsymbol{b_1} \qquad\qquad \boldsymbol{y_2} = \boldsymbol{A_2}\boldsymbol{x_2} + \boldsymbol{b_2} \tag{A.1}$$

The elements in the $\boldsymbol{y_1}$-vector is linear combinations of the elements in the $\boldsymbol{x_1}$-vector. Hence the $\boldsymbol{A_1}$-matrix is a coefficient matrix. $\boldsymbol{b_1}$ is a vector of constants. The elements in the $\boldsymbol{y_2}$-vector is linear combinations of the elements in the $\boldsymbol{x_2}$-vector. Hence the $\boldsymbol{A_2}$-matrix is a coefficient matrix. $\boldsymbol{b_2}$ is a vector of constants. The expectation of the vectors $\boldsymbol{y_1}$ and $\boldsymbol{y_2}$ can be written as

$$E(\boldsymbol{y_1}) = \boldsymbol{A_1}E(\boldsymbol{x_1}) + \boldsymbol{b_1} \qquad\qquad E(\boldsymbol{y_2}) = \boldsymbol{A_2}E(\boldsymbol{x_2}) + \boldsymbol{b_2} \tag{A.2}$$

and the covariance of the two vectors follows as:

$$\begin{aligned}
cov(\boldsymbol{y_1}, \boldsymbol{y_2}) =& E[(\boldsymbol{y_1} - E(\boldsymbol{y_1}))(\boldsymbol{y_2} - E(\boldsymbol{y_2}))^T] \\
=& E[((\boldsymbol{A_1}\boldsymbol{x_1} + \boldsymbol{b_1}) - (\boldsymbol{A_1}E(\boldsymbol{x_1}) + \boldsymbol{b_1})) \\
& ((\boldsymbol{A_2}\boldsymbol{x_2} + \boldsymbol{b_2}) - (\boldsymbol{A_2}E(\boldsymbol{x_2}) + \boldsymbol{b_2}))^T] \\
=& E[\boldsymbol{A_1}(\boldsymbol{x_1} - E(\boldsymbol{x_1}))((\boldsymbol{A_2}(\boldsymbol{x_2} - E(\boldsymbol{x_2})))^T] \\
=& E[\boldsymbol{A_1}(\boldsymbol{x_1} - E(\boldsymbol{x_1}))(\boldsymbol{x_2} - E(\boldsymbol{x_2}))^T \boldsymbol{A_2}^T] \\
=& \boldsymbol{A_1}E[(\boldsymbol{x_1} - E(\boldsymbol{x_1}))(\boldsymbol{x_2} - E(\boldsymbol{x_2}))^T]\boldsymbol{A_2}^T \\
=& \boldsymbol{A_1}cov(\boldsymbol{x_1}, \boldsymbol{x_2})\boldsymbol{A_2}^T
\end{aligned}$$

Therefore the covariance law can be defined as

$$cov(\boldsymbol{y_1}, \boldsymbol{y_1}) = \boldsymbol{A_1} cov(\boldsymbol{x_1}, \boldsymbol{x_1}) \boldsymbol{A_1}^T \tag{A.3}$$

or more often written as:

$$cov(\boldsymbol{y_1}) = \boldsymbol{A_1} cov(\boldsymbol{x_1}) \boldsymbol{A_1}^T \tag{A.4}$$

Appendix B

The inverse of a regular quadratic block matrix

Let $\boldsymbol{N}$ be a regular quadratic block matrix of dimensions $(d+s)\cdot(d+s)$ and $\boldsymbol{M}$ its inverse

$$\boldsymbol{N}=\begin{bmatrix}\boldsymbol{N_{dd}} & \vdots & \boldsymbol{N_{ds}}\\ \dots & & \dots\\ \boldsymbol{N_{sd}} & \vdots & \boldsymbol{N_{ss}}\end{bmatrix}\Leftrightarrow\boldsymbol{N^{-1}}=\boldsymbol{M}=\begin{bmatrix}\boldsymbol{M_{dd}} & \vdots & \boldsymbol{M_{ds}}\\ \dots & & \dots\\ \boldsymbol{M_{sd}} & \vdots & \boldsymbol{M_{ss}}\end{bmatrix} \tag{B.1}$$

Because of $\boldsymbol{NN^{-1}}=\boldsymbol{I}$ it is possible to write:

$$\begin{bmatrix}\boldsymbol{N_{dd}} & \vdots & \boldsymbol{N_{ds}}\\ \dots & & \dots\\ \boldsymbol{N_{sd}} & \vdots & \boldsymbol{N_{ss}}\end{bmatrix}\cdot\begin{bmatrix}\boldsymbol{M_{dd}} & \vdots & \boldsymbol{M_{ds}}\\ \dots & & \dots\\ \boldsymbol{M_{sd}} & \vdots & \boldsymbol{M_{ss}}\end{bmatrix}=\begin{bmatrix}\boldsymbol{I_{dd}} & \vdots & \boldsymbol{0_{ds}}\\ \dots & & \dots\\ \boldsymbol{0_{sd}} & \vdots & \boldsymbol{I_{ss}}\end{bmatrix} \tag{B.2}$$

Multiplying the sub-matrices gives the equations:

$$\boldsymbol{N_{dd}M_{dd}}+\boldsymbol{N_{ds}M_{sd}}=\boldsymbol{I_{dd}} \tag{B.3}$$
$$\boldsymbol{N_{dd}M_{ds}}+\boldsymbol{N_{ds}M_{ss}}=\boldsymbol{0_{ds}} \tag{B.4}$$
$$\boldsymbol{N_{sd}M_{dd}}+\boldsymbol{N_{ss}M_{sd}}=\boldsymbol{0_{sd}} \tag{B.5}$$
$$\boldsymbol{N_{sd}M_{ds}}+\boldsymbol{N_{ss}M_{ss}}=\boldsymbol{I_{ss}} \tag{B.6}$$

Inserting $\boldsymbol{M_{dd} = N_{dd}^{-1} - N_{dd}^{-1} N_{ds} M_{sd}}$ from Eq.(B.3) into Eq.(B.5) gives

$$\boldsymbol{N_{sd} N_{dd}^{-1} - N_{sd} N_{dd}^{-1} N_{ds} M_{sd} + N_{ss} M_{sd} = 0_{sd}} \Leftrightarrow$$
$$\boldsymbol{M_{sd} = -(N_{ss} - N_{sd} N_{dd}^{-1} N_{ds})^{-1} N_{sd} N_{dd}^{-1}}$$

and $\boldsymbol{M_{dd}}$ may now be rewritten as an expression of $\boldsymbol{N}$-sub matrices only:

$$\boldsymbol{M_{dd} = N_{dd}^{-1} + N_{dd}^{-1} N_{ds} (N_{ss} - N_{sd} N_{dd}^{-1} N_{ds})^{-1} N_{sd} N_{dd}^{-1}} \tag{B.7}$$

Inserting Eq.(B.4) $\boldsymbol{M_{ds} = -N_{dd}^{-1} N_{ds} M_{ss}}$ in Eq.(B.6), one finds

$$\boldsymbol{-N_{sd} N_{dd}^{-1} N_{ds} M_{ss} + N_{ss} M_{ss} = I_{ss}} \Leftrightarrow$$
$$\boldsymbol{(N_{ss} - N_{sd} N_{dd}^{-1} N_{ds}) M_{ss} = I_{ss}} \Leftrightarrow$$
$$\boldsymbol{M_{ss} = (N_{ss} - N_{sd} N_{dd}^{-1} N_{ds})^{-1}}$$

Finally, the inverse of the regular quadratic block matrix $\boldsymbol{N}$ is

$$\boldsymbol{N^{-1} = M = \begin{bmatrix} M_{dd} & \vdots & M_{ds} \\ \dots & & \dots \\ M_{sd} & \vdots & M_{ss} \end{bmatrix}} \tag{B.8}$$

with sub matrices given by:

$$\boldsymbol{M_{dd} = N_{dd}^{-1} + N_{dd}^{-1} N_{ds} (N_{ss} - N_{sd} N_{dd}^{-1} N_{ds})^{-1} N_{sd} N_{dd}^{-1}} \tag{B.9}$$
$$\boldsymbol{M_{ds} = N_{dd}^{-1} N_{ds} (N_{ss} - N_{sd} N_{dd}^{-1} N_{ds})^{-1}} \tag{B.10}$$
$$\boldsymbol{M_{sd} = -(N_{ss} - N_{sd} N_{dd}^{-1} N_{ds})^{-1} N_{sd} N_{dd}^{-1}} \tag{B.11}$$
$$\boldsymbol{M_{ss} = (N_{ss} - N_{sd} N_{dd}^{-1} N_{ds})^{-1}} \tag{B.12}$$

Appendix C

Models with autocorrelated errors

The following is a short summary of ARIMA time series models and the relation between the specific Auto Regressive (AR)- and the more traditional regression model. For a more detailed discussion, the reader is referred to Shumway and Stoffer [2000], Wei [1990] and Wetherill [1986]. A regression model of assumed equally weighted observations has the form:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{C.1}$$

where $\boldsymbol{y}$ is an $(n \times 1)$ observed response vector and $\boldsymbol{X}$ the $(n \times (r+1))$ design matrix of corresponding predictor variables. $\boldsymbol{\varepsilon}$ is the $(n \times 1)$ vector of errors assumed to have following properties:

$$\boldsymbol{E}(\boldsymbol{\varepsilon}) = \mathbf{0} \qquad , \qquad \boldsymbol{cov}(\boldsymbol{y}) = \boldsymbol{cov}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{\Sigma_{yy}} \tag{C.2}$$

where $\mathbf{0}$ is a zero vector and $\boldsymbol{\Sigma_{yy}}$ a $(n \times n)$ a priori known cofactor matrix. In a special case of independent errors of equal weights, a unity matrix $\boldsymbol{I}$ can be used as the cofactor matrix $\boldsymbol{\Sigma_{yy}}$.

For data taken serially in time the error terms in $\boldsymbol{\varepsilon}$ are nearly always correlated. If we assume the errors to fulfil second order weak stationarity, that means correlation between following errors ε_i and ε_{i+1} to be equal, correlations between ε_i and ε_{i+2}, and so on correlations between ε_i and ε_{i+j}, for all possible i and j in the time series, then the $\boldsymbol{\Sigma_{yy}}$-matrix will take the form:

$$\boldsymbol{\Sigma_{yy}} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 & \dots \\ \rho_1 & 1 & \rho_1 & \rho_2 & \dots \\ \rho_2 & \rho_1 & 1 & \rho_1 & \\ \rho_3 & \rho_2 & \rho_1 & 1 & \ddots \\ \vdots & \vdots & & \ddots & \ddots \end{bmatrix} \tag{C.3}$$

where $\rho(k) = \rho_k$, $(k = 1, 2, ..., n-1)$ are values from the ACF for the time series of errors. Inverting such a cofactor matrix might be a problem. It would be easier if the correlation structure of the errors was known. Knowledge about the correlation structure helps to classify the kind of process that might have generated the errors. Simpler approximation formulas to compute the inverse of a cofactor matrix, given the kind of process have been derived long time ago. A way of classifying stochastic processes is using ARIMA-modelling [e.g. Wei, 1990].

Classification of stochastic processes (ARIMA)

Different kinds of classification can be done for stochastic processes. They can be grouped as broad- or narrow-band processes based on their properties in frequency domain. Because the normalised spectrum and the ACF contain identical information, it is possible to classify the kind of process analysing the ACF or ACvF of the process in time domain. Based on the estimated ACF and the Partial Auto Correlation Function (PACF), [see e.g. Wei, 1990], a classification of stochastic process can be done. Some use the terms Auto Regressive- (AR) or moving average (MA) processes. Combinations and integrated (I) types of these may also be used. The notation ARIMA is very common.

Auto Regressive processes of order p, $\boldsymbol{AR}(p)$ and
Moving Average processes of order q, $\boldsymbol{MA}(q)$ or a combination of them
Auto Regressive Moving Average processes of order p, q, $\boldsymbol{ARMA}(p, q)$

AR(p)-process

Assume that ε_t is a zero mean process, and the term ϵ_t is purely random or white noise. An AR(p) model (with p parameters) can then be written as:

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \ldots + \phi_p \varepsilon_{t-p} + \epsilon_t \tag{C.4}$$

This represents a regression or prediction of the current value ε_t of a time series as a function of the past p values of the series, and hence the term Auto regression is suggested for this kind of models.

AR(1)-process

An AR(1) model can be written as:

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \epsilon_t \tag{C.5}$$

As can be seen, the observation at time t is a function of the observation at the time $t-1$, and are thus also influenced by earlier errors. For Gaussian distributed data, the AR(1)-process is also known as a Markov process [see e.g. Wei, 1990]. In an AR(1)-process, the first coefficient of the PACF is ϕ_1, the linear relation between ε_t and ε_{t-1}. For coefficients $|\phi_1| < 1$, the process is stationary. For estimation of the PACF see Wei [1990]. The ACF of an AR(1) process has the known form:

$$\rho_k = \phi_1 \rho_{k-1} = \phi_1^k, \qquad k \geq 0 \quad \text{and} \quad \rho_0 = 1 \tag{C.6}$$

Characteristics for an AR(1) process are:

ACF	Exponentially decreasing as can be seen from Eq. (C.6)
PACF	A spike at lag one $(= \phi_1)$, then the PACF cuts off.

If it is reasonable that a time series of residuals arises from an AR(1) process, the $\boldsymbol{\Sigma_{yy}}$-matrix in Eq. (C.3), becomes:

$$\boldsymbol{\Sigma_{yy}} = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \ldots \\ \rho & 1 & \rho & \rho^2 & \ldots \\ \rho^2 & \rho & 1 & \rho & \\ \rho^3 & \rho^2 & \rho & 1 & \ddots \\ \vdots & \vdots & & \ddots & \ddots \end{bmatrix} \tag{C.7}$$

where all the ρ 's should have index 1, and therefore are skipped. It can be shown that its inverse can be approximated with [see e.g. Wetherill, 1986]:

$$\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{-\mathbf{1}} \approx \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho & 0 & 0 & \dots \\ -\rho & (1+\rho^2) & -\rho & 0 & \dots \\ 0 & -\rho & \ddots & \ddots & \\ \vdots & \vdots & \ddots & (1+\rho^2) & -\rho \\ 0 & 0 & & -\rho & 1 \end{bmatrix} \tag{C.8}$$

Remark:
If the parameter $\phi_1 = \rho_1$ in an AR(1)-process is small, the cofactor matrix $\boldsymbol{\Sigma}_{\boldsymbol{yy}}$ will be almost like the identity-matrix as in the independent case.

Appendix D

More PCA results with Norwegian data

Table D.1: Columns of normalised eigenvectors to the corresponding six largest eigenvalues in the North component in the Norwegian CGPS data.

Coloumns of normalised eigenvectors.						
No.	1.	2.	3.	4.	5.	6.
1	0.194	-0.168	-0.804	0.158	-0.036	0.448
2	0.257	-0.317	0.067	0.250	-0.104	-0.461
3	0.296	0.259	0.018	-0.032	0.148	0.098
4	0.299	-0.166	0.024	0.266	-0.130	0.012
5	0.241	0.235	-0.220	-0.666	-0.055	-0.273
6	0.238	0.190	0.384	-0.052	-0.634	0.429
7	0.299	0.229	0.061	0.024	0.328	-0.043
8	0.277	0.177	0.119	0.356	0.403	-0.054
9	0.319	0.244	-0.002	-0.036	0.110	0.054
10	0.293	0.039	0.031	0.249	-0.263	-0.022
11	0.241	-0.384	0.000	-0.384	-0.021	-0.006
12	0.199	-0.438	0.342	-0.230	0.413	0.458
13	0.316	0.169	-0.129	-0.041	-0.066	-0.139
14	0.231	-0.416	0.001	-0.033	-0.134	-0.283

Table D.2: Columns of station number, station name and 1st, 2nd and 3rd factor loadings for the North component in the Norwegian CGPS data.

Factor loadings.				
No	Station	1.load	2.load	3.load
1	an3e	0.514	-0.200	-0.737
2	an3o	0.683	-0.376	0.061
3	ber3	0.786	0.307	0.017
4	bod3	0.793	-0.196	0.022
5	dag3	0.640	0.279	-0.202
6	dom3	0.632	0.225	0.352
7	kri3	0.795	0.271	0.056
8	osl3	0.735	0.211	0.109
9	sta3	0.847	0.289	-0.002
10	trh3	0.777	0.047	0.029
11	trm3	0.640	-0.455	0.000
12	tro3	0.527	-0.519	0.313
13	try3	0.838	0.201	-0.118
14	var3	0.613	-0.493	0.001
	sum loadings	9.820	-0.409	-0.098

Table D.3: Columns of normalised eigenvectors to the corresponding six largest eigenvalues for the East component in the Norwegian CGPS data.

Coloumns of normalised eigenvectors.						
No.	1.	2.	3.	4.	5.	6.
1	0.230	-0.097	0.622	-0.079	-0.229	-0.584
2	0.273	-0.274	0.099	0.004	-0.093	-0.005
3	0.285	0.261	-0.107	0.096	0.097	-0.007
4	0.305	-0.162	0.041	-0.010	0.060	-0.019
5	0.274	0.236	-0.280	-0.137	0.464	-0.189
6	0.240	0.202	-0.114	-0.722	-0.283	0.101
7	0.284	0.219	-0.030	0.218	0.014	-0.219
8	0.259	0.120	0.230	0.285	-0.406	0.575
9	0.300	0.261	-0.084	0.102	0.022	-0.156
10	0.285	0.146	0.026	-0.249	-0.131	0.188
11	0.245	-0.478	-0.269	0.133	-0.121	-0.182
12	0.243	-0.405	-0.487	0.059	-0.208	0.034
13	0.297	0.135	0.127	0.387	0.241	0.157
14	0.197	-0.402	0.334	-0.268	0.575	0.340

Table D.4: Columns of station number, station name and 1st, 2nd and 3rd factor loadings for the East component in the Norwegian CGPS data.

Factor loadings.				
No	Station	1.load	2.load	3.load
1	an3e	0.562	-0.112	0.580
2	an3o	0.669	-0.315	0.092
3	ber3	0.698	0.301	-0.100
4	bod3	0.746	-0.187	0.038
5	dag3	0.671	0.272	-0.261
6	dom3	0.588	0.233	-0.106
7	kri3	0.695	0.252	-0.028
8	osl3	0.633	0.138	0.214
9	sta3	0.735	0.301	-0.078
10	trh3	0.698	0.168	0.024
11	trm3	0.599	-0.551	-0.251
12	tro3	0.595	-0.467	-0.454
13	try3	0.726	0.156	0.118
14	var3	0.483	-0.464	0.311
	sum loadings	9.097	-0.273	0.100

Table D.5: Columns of normalised eigenvectors to the corresponding six largest eigenvalues for the Height component in the Norwegian CGPS data.

Coloumns of normalised eigenvectors.						
No.	1.	2.	3.	4.	5.	6.
1	0.214	-0.240	-0.076	-0.740	-0.223	0.334
2	0.269	-0.282	-0.000	0.006	-0.312	0.006
3	0.280	0.207	0.176	-0.233	0.238	0.052
4	0.299	-0.205	0.041	-0.156	-0.066	-0.309
5	0.234	0.291	-0.534	0.106	0.105	0.146
6	0.251	0.257	-0.495	-0.042	0.130	0.202
7	0.273	0.240	0.409	0.143	0.027	0.105
8	0.279	0.217	0.143	0.298	-0.266	0.048
9	0.270	0.260	0.464	-0.150	0.179	0.180
10	0.289	0.037	-0.071	-0.067	-0.235	-0.692
11	0.265	-0.374	0.017	0.159	0.513	-0.072
12	0.271	-0.390	-0.075	0.135	0.408	-0.006
13	0.308	0.226	-0.131	0.066	-0.151	-0.142
14	0.222	-0.334	0.016	0.420	-0.390	0.422

Table D.6: Columns of station number, station name 1st, 2nd and 3rd factor loadings for the Height component in the Norwegian CGPS data.

Factor loadings.				
No	Station	1.load	2.load	3.load
1	an3e	0.530	-0.304	-0.070
2	an3o	0.668	-0.357	-0.000
3	ber3	0.696	0.262	0.160
4	bod3	0.741	-0.260	0.038
5	dag3	0.581	0.369	-0.485
6	dom3	0.622	0.326	-0.450
7	kri3	0.677	0.305	0.372
8	osl3	0.691	0.275	0.130
9	sta3	0.670	0.330	0.422
10	trh3	0.716	0.046	-0.065
11	trm3	0.657	-0.474	0.015
12	tro3	0.673	-0.494	-0.069
13	try3	0.764	0.287	-0.119
14	var3	0.550	-0.423	0.014
	sum loadings	9.236	-0.112	-0.105

Appendix E

More results with European data

Table E.1: Station dependent correlation coefficients for de-trended atmospheric loading predictions (predictors) for the five data sets given in Table 6.7 and European CGPS (observations) station series extracted from the first modes.

Correlation coefficients M1 series						
1	BOR1	0.413	0.360	0.411	0.356	0.387
2	BRUS	0.419	0.363	0.419	0.361	0.392
3	BUCU	0.418	0.363	0.417	0.359	0.390
4	BZRG	0.431	0.372	0.432	0.372	0.404
5	CAGL	0.414	0.359	0.412	0.355	0.393
6	EBRE	0.410	0.357	0.410	0.354	0.383
7	GENO	0.421	0.367	0.421	0.365	0.396
8	GOPE	0.407	0.351	0.405	0.346	0.380
9	GRAS	0.409	0.354	0.406	0.348	0.387
10	GRAZ	0.419	0.368	0.417	0.364	0.393
11	HERS	0.415	0.370	0.413	0.365	0.388
12	HFLK	0.343	0.279	0.348	0.282	0.328
13	JOZE	0.407	0.351	0.405	0.346	0.380
14	KOSG	0.414	0.360	0.412	0.356	0.387
15	LAMA	0.409	0.357	0.407	0.351	0.379
16	MATE	0.416	0.363	0.415	0.358	0.390
17	MEDI	0.418	0.363	0.417	0.360	0.390
18	METS	0.412	0.357	0.411	0.353	0.386
19	ONSA	0.412	0.355	0.411	0.351	0.385
20	POTS	0.416	0.361	0.415	0.357	0.390
21	SFER	0.422	0.374	0.420	0.370	0.393
22	SOFI	0.405	0.356	0.401	0.350	0.384
23	TROM	0.430	0.371	0.431	0.370	0.401
24	VENE	0.407	0.349	0.406	0.347	0.380
25	VILL	0.427	0.384	0.424	0.375	0.400
26	WSRT	0.422	0.366	0.422	0.363	0.395
27	WTZT	0.424	0.374	0.423	0.369	0.398
28	ZIMM	0.420	0.367	0.420	0.365	0.394
	MEAN	0.414	0.360	0.413	0.356	0.388

Table E.2: Station dependent regression coefficients for de-trended atmospheric loading predictions (predictors) for the five data sets given in Table 6.7 and European CGPS (observations) station series extracted from the first modes.

Regression coefficients M1 series						
No	Station	P1	P2	P3	P4	P5
1	BOR1	0.600	0.445	0.652	0.483	0.591
2	BRUS	0.670	0.413	0.732	0.555	0.590
3	BUCU	0.886	0.547	0.955	0.599	0.904
4	BZRG	0.830	0.602	0.851	0.630	0.776
5	CAGL	1.175	0.679	1.835	0.925	0.967
6	EBRE	0.884	0.554	0.886	0.572	0.798
7	GENO	0.917	0.635	1.032	0.715	0.833
8	GOPE	0.591	0.436	0.611	0.455	0.558
9	GRAS	0.892	0.606	1.106	0.721	0.904
10	GRAZ	0.804	0.592	0.834	0.623	0.771
11	HERS	0.612	0.477	0.724	0.543	0.507
12	HFLK	0.635	0.440	0.674	0.470	0.590
13	JOZE	0.661	0.471	0.683	0.491	0.637
14	KOSG	0.616	0.470	0.657	0.501	0.544
15	LAMA	0.574	0.412	0.619	0.444	0.554
16	MATE	1.066	0.641	1.455	0.820	1.154
17	MEDI	0.812	0.560	0.843	0.594	0.754
18	METS	0.575	0.374	0.713	0.446	0.548
19	ONSA	0.536	0.382	0.733	0.492	0.496
20	POTS	0.631	0.476	0.711	0.531	0.622
21	SFER	0.954	0.458	0.867	0.454	0.786
22	SOFI	0.920	0.558	0.964	0.599	0.915
23	TROM	0.838	0.415	1.073	0.499	0.819
24	VENE	0.760	0.531	0.754	0.546	0.706
25	VILL	0.641	0.402	0.609	0.396	0.586
26	WSRT	0.523	0.395	0.599	0.445	0.457
27	WTZT	0.711	0.537	0.760	0.575	0.665
28	ZIMM	0.733	0.542	0.733	0.552	0.680
	MEAN	0.752	0.502	0.845	0.560	0.704

Table E.3: Station dependent regression coefficients for de-trended atmospheric loading predictions (predictors) for the five data sets given in Table 6.7 and European CGPS (observations) station series extracted from the second modes.

Regression coefficients M2 series						
No	Station	P1	P2	P3	P4	P5
1	BOR1	0.553	0.595	0.561	0.576	0.495
2	BRUS	0.225	-0.152	0.204	0.199	0.154
3	BUCU	-2.466	-1.409	7.399	65.820	-11.613
4	BZRG	-0.154	-0.137	-0.190	-0.165	-0.140
5	CAGL	0.921	0.833	2.066	2.135	0.726
6	EBRE	0.642	0.604	0.625	0.597	0.547
7	GENO	0.281	0.255	0.350	0.326	0.238
8	GOPE	0.118	0.069	-0.770	0.226	0.215
9	GRAS	0.555	0.515	0.773	0.749	0.570
10	GRAZ	-0.326	-0.276	-0.469	-0.384	-0.338
11	HERS	0.684	0.794	0.664	0.725	0.398
12	HFLK	0.189	0.169	0.226	0.200	0.170
13	JOZE	0.713	0.781	0.621	0.634	0.594
14	KOSG	0.621	2.175	0.158	0.161	0.175
15	LAMA	0.334	0.347	0.328	0.322	0.290
16	MATE	0.247	0.212	0.613	0.591	0.328
17	MEDI	0.287	0.258	0.355	0.323	0.272
18	METS	0.185	0.183	0.223	0.205	0.159
19	ONSA	0.123	0.118	0.161	0.146	0.108
20	POTS	0.671	0.699	0.837	0.863	0.703
21	SFER	0.753	0.683	0.700	0.670	0.612
22	SOFI	-0.797	-0.606	-2.047	-1.502	-0.996
23	TROM	0.227	0.214	0.286	0.247	0.190
24	VENE	0.713	0.636	0.855	0.758	0.656
25	VILL	0.908	0.865	0.852	0.818	0.806
26	WSRT	-0.432	-0.354	-3.641	-1.611	-1.485
27	WTZT	-0.737	-0.638	-1.047	-0.859	-0.804
28	ZIMM	0.273	0.258	0.290	0.267	0.253
	MEAN	0.190	0.275	0.392	2.608	-0.240

Table E.4: Station dependent regression coefficients for de-trended atmospheric loading predictions (predictors) for the five data sets given in Table 6.7 and European CGPS (observations) station series extracted from the third modes.

Regression coefficients M3 series						
No	Station	P1	P2	P3	P4	P5
1	BOR1	-0.543	-11.515	-0.687	-5.727	-0.334
2	BRUS	0.385	-39.707	0.416	0.376	0.328
3	BUCU	0.422	0.426	0.400	0.391	0.380
4	BZRG	-1.014	2.713	-1.034	-3.829	-0.648
5	CAGL	1.374	0.898	2.221	0.853	0.748
6	EBRE	1.761	3.588	1.757	10.378	2.421
7	GENO	-0.187	-0.307	-0.190	-0.132	-0.091
8	GOPE	3.107	-1.468	3.846	-1.915	1.240
9	GRAS	4.929	5.317	-10.845	-0.463	-1.134
10	GRAZ	0.483	0.884	0.468	0.653	0.366
11	HERS	0.757	0.718	0.870	0.869	0.584
12	HFLK	7.338	2.062	53.959	4.801	-9.056
13	JOZE	-0.098	-0.128	-0.106	-0.133	-0.077
14	KOSG	0.528	0.458	0.528	0.469	0.414
15	LAMA	-0.283	-0.359	-0.332	-0.395	-0.225
16	MATE	0.717	0.654	1.007	0.744	0.688
17	MEDI	0.186	0.247	0.168	0.174	0.137
18	METS	-0.413	-0.261	3.446	-0.783	-0.365
19	ONSA	0.585	0.564	0.567	0.689	0.422
20	POTS	0.656	0.360	0.662	0.422	0.766
21	SFER	22.698	-1.403	-4.433	-0.844	-2.753
22	SOFI	0.779	0.756	0.681	0.640	0.605
23	TROM	0.799	-4.476	0.456	1.119	0.508
24	VENE	5.643	8.543	5.681	6.999	4.202
25	VILL	0.956	1.948	0.931	4.329	1.345
26	WSRT	0.421	0.363	0.445	0.403	0.324
27	WTZT	1.257	0.452	1.976	0.695	5.255
28	ZIMM	1.042	0.796	1.204	1.012	1.227
	MEAN	1.939	-0.996	2.288	0.778	0.260

Table E.5: Station dependent regression coefficients for de-trended atmospheric loading predictions (predictors) for the five data sets given in Table 6.7 and European CGPS (observations) station series extracted from the first three modes.

Regression coefficients M1+M2+M3 series						
No	Station	P1	P2	P3	P4	P5
1	BOR1	0.686	0.517	0.737	0.544	0.665
2	BRUS	0.619	0.333	0.688	0.544	0.534
3	BUCU	0.808	0.584	0.826	0.591	0.836
4	BZRG	0.821	0.608	0.846	0.638	0.775
5	CAGL	0.695	0.466	1.161	0.677	0.582
6	EBRE	0.601	0.345	0.636	0.428	0.576
7	GENO	0.739	0.529	0.845	0.613	0.684
8	GOPE	0.587	0.453	0.604	0.468	0.550
9	GRAS	0.629	0.431	0.804	0.545	0.666
10	GRAZ	0.799	0.611	0.824	0.631	0.767
11	HERS	0.590	0.428	0.718	0.516	0.489
12	HFLK	0.617	0.415	0.670	0.458	0.595
13	JOZE	0.762	0.561	0.770	0.557	0.720
14	KOSG	0.615	0.471	0.669	0.514	0.533
15	LAMA	0.609	0.451	0.640	0.461	0.581
16	MATE	0.752	0.529	1.122	0.711	0.870
17	MEDI	0.661	0.472	0.695	0.511	0.629
18	METS	0.415	0.327	0.495	0.366	0.366
19	ONSA	0.462	0.355	0.603	0.427	0.412
20	POTS	0.714	0.547	0.806	0.600	0.695
21	SFER	0.562	0.185	0.570	0.272	0.502
22	SOFI	0.915	0.659	0.929	0.672	0.890
23	TROM	0.363	0.299	0.427	0.332	0.278
24	VENE	0.590	0.562	0.526	0.532	0.491
25	VILL	0.399	0.212	0.428	0.291	0.402
26	WSRT	0.488	0.366	0.572	0.423	0.420
27	WTZT	0.774	0.589	0.824	0.622	0.721
28	ZIMM	0.637	0.464	0.651	0.494	0.607
	MEAN	0.640	0.456	0.717	0.516	0.601