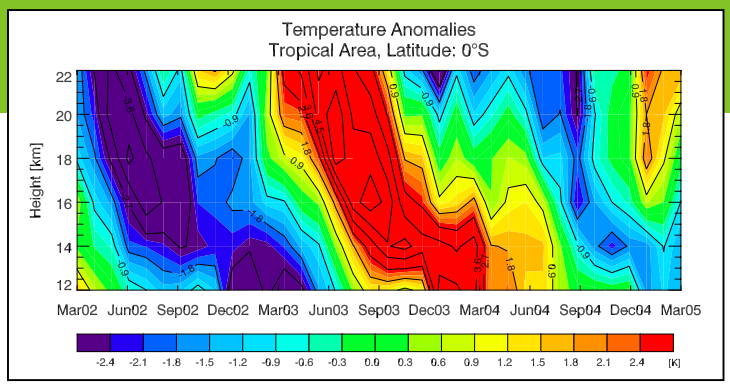
Wegener Center for Climate and Global Change
University of Graz



Scientific Report No. 7-2005

Climate Diagnostics in Radio Occultation Temperature Climatologies of CHAMP and ECMWF

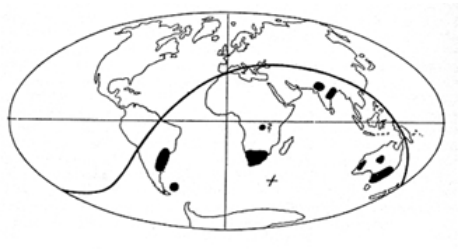
Bettina C. Lackner
Barbara Pirscher

December 2005

Wegener Center
www.wegcenter.at

The **Wegener Center for Climate and Global Change** combines as an interdisciplinary, internationally oriented research center the competences of the University of Graz in the research area „Climate, Environmental and Global Change“. It brings together, in a dedicated building close to the University central campus, research teams and scientists from fields such as geo- and climate physics, meteorology, economics, geography, and regional sciences. At the same time close links exist and are further developed with many cooperation partners, both nationally and internationally. The research interests extend from monitoring, analysis, modeling and prediction of climate and environmental change via climate impact research to the analysis of the human dimensions of these changes, i.e, the role of humans in causing and being effected by climate and environmental change as well as in adaptation and mitigation. The director of the center, hosting about 30 researchers, is the geophysicist Gottfried Kirchengast, the lead partner and deputy director is the economist Karl Steininger. (more informationen at www.wegcenter.at)

The present report is the result of a Master thesis work completed in October 2005.



Alfred Wegener (1880-1930), after whom the Wegener Center is named, was founding holder of the University of Graz Geophysics Chair (1924-1930) and was in his work in the fields of geophysics, meteorology, and climatology a brilliant, interdisciplinary thinking and acting scientist and scholar, far ahead of his time with this style. The way of his ground-breaking research on continental drift is a shining role model — his sketch on the relationship of the continents based on traces of an ice age about 300 million years ago (left) as basis for the Wegener Center Logo is thus a continuous encouragement to explore equally innovative scientific ways: *ways emerge in that we go them* (Motto of the Wegener Center).

Wegener Center Verlag • Graz, Austria

ISBN 3-9502126-4-7

December 2005

Contact: DI Mag. Bettina C. Lackner
bettina.lackner@uni-graz.at
MMMag. Barbara Pirscher
barbara.pirscher@uni-graz.at

Wegener Center for Climate and Global Change
University of Graz
Leechgasse 25
A-8010 Graz, Austria
www.wegcenter.at

Bettina C. Lackner

Barbara Pirscher

Climate Diagnostics in Radio Occultation Temperature Climatologies of CHAMP and ECMWF

Diplomarbeit

zur Erlangung des akademischen Grades einer
Magistra an der naturwissenschaftlichen Fakultät der
Karl-Franzens-Universität Graz

Betreuer:

Univ.-Prof. Mag. Dr. Gottfried Kirchengast

Mitbetreuer:

Mag. Michael Borsche

Wegener Zentrum für Klima und Globalen Wandel
und
Institutsbereich Geophysik, Astrophysik und Meteorologie,
Institut für Physik
Karl-Franzens-Universität Graz

Acknowledgments

Standing at the end of one demanding but also exciting year, we want to sincerely thank quite a large number of people for their generous assistance.

First of all, thanks to Prof. Dr. Gottfried Kirchengast, our thesis supervisor, for providing this topic we have been interested in for a long time, also for enabling to work together on one subject, so that it was possible to really go into detail, and for opening doors we did not know before. These doors did not only lead to new scientific areas but also to those of Wegener Center, including excellent infrastructure for our work.

Thanks to all members of the ARSCliSys research group for their assistance and support, especially to Mag. Michael Borsche, our co-supervisor, to Dr. Andreas Gobiet, who left us some of his computer codes, to Dr. Marc Schwärz for trouble-shooting in Fortran90, and to Thomas Kabas and Matthias Themeßl – it was a great time we spent together in the DiplomandInnenzimmer.

We also want to thank all lecturers who supported us during our studies, especially Prof. Dr. Reinhard Leitinger, who always was there to answer any questions and to give us every support possible.

We are grateful to GeoForschungsZentrum Potsdam for providing CHAMP radio occultation data, to the European Centre for Medium-Range Weather Forecasts providing ECMWF analysis data, and to U.S. National Centers for Environmental Prediction and U.S. National Centers for Atmospheric Research providing NCEP/NCAR reanalysis data.

Susi Etschmaier, thank you for proofreading the first part of our work. We were really relieved that you, as a native speaker, understood our English and returned us the work not totally "in red".

Thanks to Mag. Stefan Stangl for solving some very tricky TeX problems.

Furthermore, I, Bettina, am very grateful to the "Interuniversitary Research Centre", my employer, which enabled my study beside my work there. Thanks especially to the members of the "ecological product policy" working group, Dr. Ines Oehme, Dr. Uli Seebacher, and Dr. Manfred Klade, who agreed to postpone many of our meetings and supported me emotionally. Thanks to my friends and my family for their support and for enduring my bad moods during the last years as time was running out so often. And a big thanks to Rudi, for accepting that I preferred spending my time with my books and computer, for our discussions and his support, for washing and ironing all my clothes (I will miss this) and for all the coffees and sweets during the long nights.

I, Barbara, want to thank my parents, who still support me by finishing my third study, which I do not take for granted at all. I also want to thank Marko for his understanding in coming home that late during the last year, for his effort, when we moved to another place, especially for fixing all my shelfs and filling them with my books. Last but not least, I want to thank all my friends for accompanying me during the last years.

Abstract

Accuracy, global coverage, long term stability, and a high vertical resolution are important properties of the radio occultation technique, enabling successful measurements of atmospheric parameters. Starting from phase differences, the atmospheric temperature, as well as pressure or humidity (auxiliary data are necessary for the latter) can be calculated.

The German CHAMP satellite (CHAllenging Minisatellite Payload) carries a GPS flight receiver, which allows the acquisition of such measurements. Using several years of temperature profiles, climatologies were generated by researchers of the ARSCliSys working group at Wegener Center, University of Graz, which we in turn compared to analyses and reanalyses (supported from European and American meteorological services) as well as to two different climatological models (NRLMSISE-00 and CIRA86aQ_UoG).

One focal point was the investigation of the influence of local time at which radio occultation measurements have been taken. Due to the orbit of the satellite, a majority of the measurements within a month are made in two three-hour intervals, which are separated by a twelve hour local time lag. The investigations revealed that monthly climatologies do not show appreciable problems and that the influence of local time is generally negligible.

An analysis of CHAMP radio occultation data resulted from factor analysis (being implemented with four different calculation procedures) and principal component analysis. The techniques were compared by means of two different atmospheric fields, the search of atmospheric patterns was performed in four atmospheric domains (two on global and two on regional scale). The comparative exploration of the different pattern decomposition analysis techniques led to very useful insights into strengths and weaknesses of the methods. Concerning the identification of atmospheric pattern, the seasonal cycle was the most dominant one, which was found in all regions beyond the tropical area. After the removal of monthly means (and consequently the seasonal cycle) the causes of smaller temperature fluctuations were found, but an interpretation was not always feasible. Nevertheless, some well known atmospheric patterns, such as QBO (Quasi Biennial Oscillation) and SSW (Sudden Stratospheric Warming), could be identified.

Zusammenfassung

Genauigkeit, globale Erfassung, Langzeitstabilität und eine hohe vertikale Auflösung sind Eigenschaften der Radiookkultationstechnik, die seit einigen Jahren erfolgreich für die Messung atmosphärischer Parameter angewendet wird. Ausgehend von Phasendifferenzmessungen kann auf die Temperatur, aber auch auf den Druck oder die Feuchtigkeit (letztere jedoch nur mit zusätzlicher Hintergrundinformation) rückgeschlossen werden.

Der deutsche CHAMP-Satellit (CHAllenging Minisatellite Payload) hat einen GPS Empfänger an Bord mit welchem solche Messungen durchgeführt werden können. Am Wegener Zentrum der Uni-Graz, werden von MitarbeiterInnen der ARSCliSys-Gruppe aus den Messdaten Temperaturklimatologien erstellt, welche in dieser Arbeit mit Analysen bzw. Reanalysen (erstellt von dem europäischen und dem amerikanischen Wetterdienst), als auch mit zwei verschiedenen Klimatologie-Modellen (NRLMSISE-00 und CIRA86aQ_UoG) verglichen wurden.

Ein Schwerpunkt dabei lag auf der Untersuchung des Einflusses der Lokalzeit, zu welcher die Radiookkultationsmessungen stattgefunden haben. Aufgrund der Umlaufbahn des Satelliten erfolgt ein Großteil der Messungen innerhalb eines Monats in Zwei- bis Drei-Stunden-Intervallen, welche um zwölf Stunden zueinander verschoben sind. Es stellte sich heraus, dass der Einfluss der Lokalzeit, zu der die Messungen stattgefunden haben, im Allgemeinen vernachlässigbar ist.

Eine Analyse der CHAMP Radiookkultationsdaten erfolgte mit Hilfe der Faktorenanalyse und der Hauptkomponentenanalyse, wobei erstere durch vier unterschiedliche Verfahren implementiert wurde. Die beiden Methoden wurden anhand von zwei atmosphärischen Feldern verglichen, die Suche nach atmosphärischen Mustern wurde auf vier atmosphärische Bereiche (zwei auf globaler und zwei auf regionaler Skala) ausgeweitet. Die vergleichende Untersuchung der unterschiedlichen Techniken zum Auffinden von Mustern führte zu sehr wertvollen Einblicken in die Stärken und Schächen der Methoden. Außerhalb des tropischen Bereiches war der Jahresgang das vorherrschende Muster, welches in den Daten anzutreffen war, wurde dieser entfernt, so kamen die Ursachen für geringere Temperaturschwankungen zum Vorschein. Nicht immer war eine Interpretation der auftretenden Strukturen möglich. Dennoch konnten in vielen Fällen bekannt atmosphärische Muster wie etwa die QBO (Quasi Biennial Oscillation) oder SSW (Sudden Stratospheric Warming) beobachtet werden.

Contents

Introduction

"Global warming" is in nearly everybody's mind. In any industrialized country it is discussed if–and if so how much–the temperature raises on earth, and how the impacts will influence human (and other) lives. An intensive study examining this issue was done by IPCC (2001). They assessed that "systematic observations and reconstructions" are a "high priority area for action".

The development of meteorology and climatology as a self-standing field of science is closely linked to the invention of the thermometer, which is (mostly) attributed to Galileo Galilei (Stoehr 2004) (who lived in the 16^{th} century). Since then many technological improvements were achieved, and since 1995, the radio occultation method is established on earth to gage profiles of meteorological parameters[1]. The first experiment used to this end was the GPS Meteorology (GPS/Met) instrument, launched on board the MicroLab-1 spacecraft on April 3, 1995 (Ware et al. 1996). It provided measurements for several multi-week periods until March 1997 . These measurements were analyzed and evaluated and confirmed that the radio occultation technique is a powerful method to obtain reliable profiles of atmospheric parameters (Rocken et al. 1997) such as temperature. The technique satisfies the demands of IPCC (2001) who called to "sustain and expand the observational foundation for climate studies by providing accurate, long-term, consistent data".

After these early days of the radio occultation on earth, other scientific missions were started and because of the proceeding success still more missions will take place. Already launched missions are CHAMP (CHallenging Minisatellite Payload, see e.g., Wickert et al. (2001a)), SAC-C (Satelite de Applicaciones Cientificas, see e.g., Meehan and Hajj (2001)) and GRACE (Gravity Recovery and Climate Experiment, see e.g., Dunn et al. (2003)). Metop (e.g., Loiselet et al. (2000)) and COSMIC (Constellation Observing System for Meteorology, Ionosphere, and Climate, see e.g., Anthes et al. (2001)) are two comprehensive missions, which are planned to start in 2006.

Nowadays, the radio occultation technique only uses GPS signals (satellites launched and controlled by U.S. Department of Defense), but in the near future also signals from "Galileo" (funded by the European Space Agency, ESA, and the European Union, EU) will be available. These additional satellites will extend the possibilities of radio occultation.

This thesis wants to contribute to climate monitoring by means of CHAMP radio

[1] The method was already successfully applied at planets in the solar system before.

occultation data. Since July 2000, the German CHAMP satellite is placed in a low earth orbit and since February 2001, it provides radio occultation measurements to calculate profiles of refractivity, geopotential height, and temperature. These temperature profiles, retrieved by researchers of the ARSCliSys group under the direction of G. Kirchengast, Wegener Center, University of Graz, were used in this work to create monthly climatologies and to compare them with analyses of the ECMWF (European Centre for Medium-Range Weather Forecasts), with reanalyses of NCEP/NCAR (U.S. National Centers for Environmental Prediction/U.S. National Centers for Atmospheric Research), with NRLMSISE-00 data and with CIRA86aQ_UoG data (the latter two are climatological model data sets). These investigations and results are looked at in detail in the first part of this thesis (Part I), where Chapter 1 deals with the theoretical background of radio occultation, the calculation of temperature climatologies and the description of the applied data sets. The differences between these data sets are discussed in Chapter 2. Chapter 3 provides the comparisons amongst the data sets, providing insight into different error characteristics, in particular also regarding local time effects. Part I closes with the conclusions of the detected differences (Chapter 4).

The second part of the thesis (Part II) deals with the examination of the CHAMP RO data with focus on exploring methods of finding atmospheric patterns (data mining). In doing so, four different atmospheric fields were chosen to be investigated with the help of two different mathematical methods: principal component analysis (PCA) and factor analysis (FA). First of all, an overview of the two methods is given in Chapter 5. The theoretical background and the derivation of principal component analysis are looked at in detail in Chapter 6. Factor analysis, discussed in Chapter 7, is split up into four different methods: iterative principal factor analysis, true factor analysis, maximum likelihood factor analysis, and centroid factor analysis.

Common properties (such as the determination of the number of factors being extracted, or rotation) and the differences between the two methods (underlying mathematical models, results' dependence of the number of selected factors and explained variances) are discussed in Chapter 8. The actual examinations of the atmospheric fields (two global and two local regions) follow in Chapter 9. Besides a description of the investigated data sets, factor analysis specific problems are addressed, as well as problems occurring in regard to different dimensions of applied data matrices (coarse and detailed resolved atmospheric temperature fields were investigated) for both PCA and FA. Conclusions on the results of Part II complete the work.

Part I

Climate Diagnostics of CHAMP Radio Occultation Temperatures

1 Data Description

1.1 CHAMP Radio Occultation Data

1.1.1 The CHAMP Satellite

(Author: B. Pirscher)

In 1994 scientists of the **G**eo**F**orschungs**Z**entrum Potsdam (GFZ), under the direction of Prof. Christoph Reigber, suggested a small satellite mission to

- explore the spatial structure and temporal variability of the gravity field,
- determine the magnetic field, and
- realize a limb sounding technique to probe the earth's atmosphere using GPS[1] signals.

The satellite was launched on July 15, 2000 at 11:59:59:628 UTC[2] from Plesetsk, Russia aboard a Cosmos-3M launch vehicle.
Technical facts concerning the CHAMP Satellite (**CHA**llenging **M**inisatellite **P**ayload) are:

Total mass:	522.5 kg
Total length:	8.222 m
Length exclusive boom:	4.178 m
Width:	1.621 m
Height:	0.75 m

Figure 1.1: The CHAMP satellite, Wickert (2004).

Its initial altitude was 454 km, the nearly polar orbit is almost circular (eccentricity: 0.004, inclination: 87.2°); the period is 93.55 minutes. The intended duration of the mission is about five years (at least one orbit change maneuver is necessary because of the decrease of altitude as a result of drag). The orbit geometry leads to a global coverage of observations, with more information in low and mid latitude regions and less information in polar areas.

[1] **G**lobal **P**ositioning **S**ystem
[2] **U**niversal **T**ime **C**oordinated

The scientific instruments aboard CHAMP consist of an Electrostatic STAR Accelerometer, a GPS Receiver TRSR-2, a LASER Retro Reflector, a Fluxgate, and an Overhauser Magnetometer, an Advanced Stellar Compass, and a Digital Ion Driftmeter. A detailed description of the satellite and the instruments aboard can be found in Reigber et al. (2001).

GPS Receiver TRSR-2

The GPS receiver TRSR-2 enables high-precision orbit determination of the CHAMP satellite, accurate time and navigation information. Compared to the antenna aboard the MicroLab-I satellite (GPS/MET experiment[3]), this antenna features a higher signal to noise ratio (SNR). It operates in three measurement modes; in occultation mode it receives signals transmitted from setting GPS satellites (setting occultation). The signal of the occulting satellite is sampled at 50 Hz; the signal of a reference GPS satellite is sampled at 0.1 Hz to 1 Hz. The high gain helix antenna is mounted on the aft panel of the satellite and is inclined 20° toward nadir.

The first occultation measurements were recorded on February 11, 2001 between 19:04 UTC and 20:04 UTC.

GPS Satellites

The **G**lobal **P**ositioning **S**ystem (GPS) was conceived by the U.S. Department of Defense (DOD) in 1973 and in 1983 the DOD decided to release the GPS to civil users. In 1994 the set of satellites was completed. At least 24 satellites circuit at an altitude of 20 200 km on six different orbits (inclination approximately 55°) with a period of 11 h 58 min. Each satellite continuously transmits right circumpolar radiation at two L-Band[4] carrier frequencies, $f_1 = 1575.42$ MHz ($\lambda_1 = 0.19$ m) and $f_2 = 1227.60$ MHz ($\lambda_2 = 0.244$ m).

These two frequencies result from the fundamental frequency $f_0 = 10.23$ MHz ($\lambda_0 = 29.3$ m) by $f_1 = 154 \times f_0$ and $f_2 = 120 \times f_0$.

Using phase modulation, a binary code is modulated on the carrier frequencies containing the C/A-code ("coarse acquisition" or "clear/access"), which is modulated only on L_1. The P-code ("protected" or "precise") can be found on both the L_1 and L_2 carrier frequencies yielding the P1- and the P2-code. The basic observables used in the occultation technique are the C/A-phase and the P2-phase.

Prior to May 2, 2000 the U.S. DOD used the "Selective Availability" (S/A) mode to degrade the accuracy of satellite GPS signals.

[3]The GPS radio occultation technique was tested for the first time from April 3, 1995 to March 1997.
[4]Frequency domain between 1 GHz and 2 GHz.

1.1.2 The Retrieval Process

(Author: B. Pirscher)

Introduction

The radio occultation (RO) technique is a generally accepted method for global remote sensing. It enables the extraction of vertical profiles of atmospheric parameters such as temperature, pressure, and humidity with high accuracy.

The method is based on the limb sounding geometry using signals from GPS satellites, which are received from GPS receivers aboard **L**ow **E**arth **O**rbit (LEO) satellites. Because the propagating signals are modified by the atmosphere depending on its properties, the modification is a measure for atmospheric components and physical atmospheric characteristics.

The RO technique uses GNSS (**G**lobal **N**avigation **S**atellite **S**ystem) signals from 1 GHz to 2 GHz because the signals are only affected by the refractivity field, and other effects such as scattering, dispersion, and polarization are negligible or correctable at these wavelengths.

A short introduction of the method and its properties can be found in Kirchengast (2004) and detailed descriptions of this technique can be found in e.g., Kursinski et al. (2001), Foelsche (1999), or Steiner et al. (2001).

The Occultation Geometry

Figure 1.2 schematically illustrates the geometry arising during an occultation event. The straight line, s_0 is the distance between the transmitter T and the receiver R, while $\boldsymbol{r}_\mathrm{T}$, $\boldsymbol{r}_\mathrm{R}$, $\boldsymbol{v}_\mathrm{T}$, and $\boldsymbol{v}_\mathrm{R}$ are the radius and velocity vectors of the satellites. a is the impact parameter (specified below), r is the distance from the center of local curvature to the tangent point, and ϕ_T and ϕ_R are the angles between the ray paths and the radius vectors of the according satellites. γ is the angle between the radius vectors $\boldsymbol{r}_\mathrm{T}$ and $\boldsymbol{r}_\mathrm{R}$ at the local center of curvature.

Under accurate geometric circumstances, the ellipsoidal shape of the earth has to be applied in the geometry because it affects the spherical symmetry of the refractivity field. First errors can be eliminated by replacing the earth's radius with the local radius of curvature. The radius of curvature is defined as the radius of a sphere that is tangential to the ellipsoid at the location of the occultation event within the occultation plane (Foelsche 1999).

Definitions: Refraction, Refractivity

Because the index of refraction

$$n = \frac{c_\mathrm{vacuum}}{c_\mathrm{medium}}, \tag{1.1}$$

is close to unity in the atmosphere, it is common to use the refractivity in atmospheric studies:

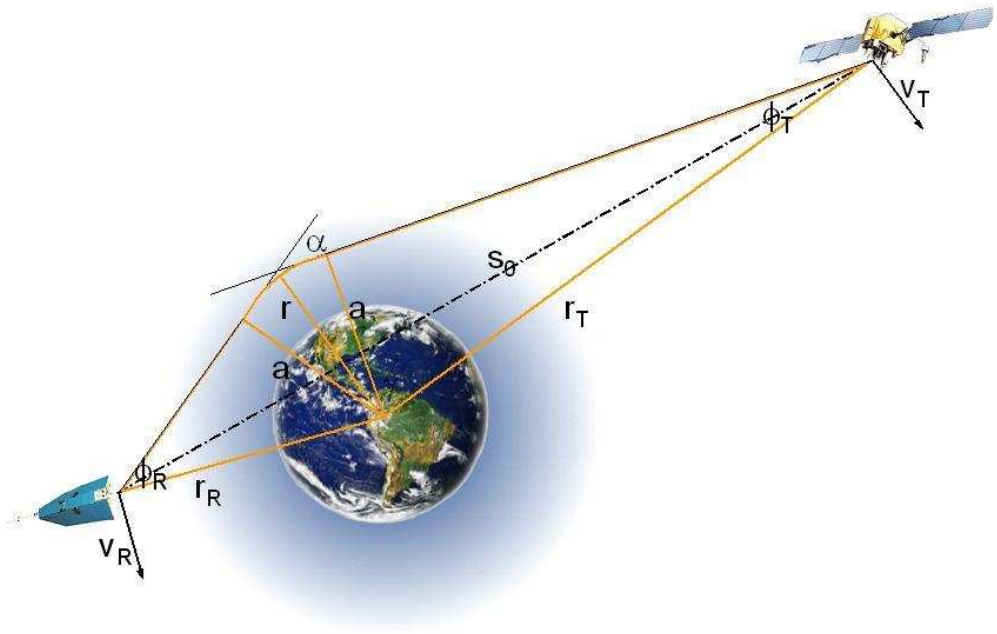


Figure 1.2: Sketch of the geometry of the occultation experiment. A GPS satellite transmits electromagnetic signals, which are received from a GPS receiver on board a LEO satellite (CHAMP).

$$N = (n - 1) \times 10^6, \tag{1.2}$$

n	index of refraction,
c_{vacuum}	speed of light in the vacuum [m/s],
c_{medium}	speed of light in the medium [m/s],
N	refractivity.

Physical Principles

On the basis of precise GPS and LEO satellite orbit data, bending angles are calculated assuming a spherical symmetric atmospheric refractive index field $n(r)$ and taking the geometric optics approximation[5]. A real refractivity index (no absorption), a monochromatic signal and small wavelengths (compared to atmospheric scales) have to be assumed.

Fermat's principle says that the actual path s between two points (from transmitter T to receiver R) taken by a beam of light (speed of light in vacuum $c_{\text{vacuum}} = c$) is the one that is traversed in the least time t:

[5]The geometric optics approximation is not valid in the lower atmosphere (< 5 km) because of high content of water vapor. The wave optics has to be applied in this region.

$$t = \int_{\mathrm{T}}^{\mathrm{R}} \frac{n(s)}{c} \mathrm{d}s = \frac{1}{c} \int_{\mathrm{T}}^{\mathrm{R}} n(s) \mathrm{d}s = \min. \tag{1.3}$$

The optical path length L of a signal propagating along the raypath T $\rightarrow$ R is defined as (Kursinski et al. 1997):

$$L = \int_{\mathrm{T}}^{\mathrm{R}} n(s) \mathrm{d}s. \tag{1.4}$$

Snell's law follows directly from the Fermat's principle. It states that the ratio of the sine of the angle of incidence ϕ_1 to the sine of the angle of refraction ϕ_2 is constant:

$$\frac{\sin \phi_1}{\sin \phi_2} = \frac{n_2}{n_1} = \frac{c_1}{c_2} = \text{const},$$

n_1 and n_2 are the index of refractivity in different medium, c_1 and c_2 are the corresponding speeds of light.

Bouguer's formula is the generalization of Snell's law. It is valid in a spherical symmetric medium ($n = n(r)$):

$$a = rn \sin \phi = \text{const}, \tag{1.5}$$

a is known as the impact parameter, r is the radius value, and ϕ is the angle between the ray vector and the local radius vector. The impact parameter is defined as the perpendicular distance between the center of local curvature and the ray path asymptote and is constant along a ray path. At the tangent point, where $\sin \phi = 1$, the impact parameter becomes $a = r_{\mathrm{t}} n(r_{\mathrm{t}})$. Bouguer's rule is generally a good approximation of the earth's atmosphere (Kursinski et al. 2001).

The Bending Angle

The change in the raypath direction accumulated along the raypath is defined as the bending angle $\alpha(a)$ (Kursinski et al. 2001).

In general, the bending angle is very small and hardly measurable, but it can be derived from accurate phase measurements because of Doppler shifted frequencies being time derivations of excess phase paths (cf., equation (1.9)). During a GPS radio occultation event the phase of the signal is the most important measurement. It varies due to the relative motion between the transmitter and the receiver (known as Doppler shift), and due to an additional propagation delay that follows from the reduction of the speed of light in the atmosphere (effect of atmospheric bending).

From the measured phase differences $\Delta\varphi_i$, it is possible to calculate the phase delays (excess phase paths) ΔL_i of the L_1 and the L_2 signal, respectively:

$$\Delta L_1 = \int_{\mathrm{T}}^{\mathrm{R}} n(s_1)\mathrm{d}s_1 - s_0 = \Delta\varphi_1\lambda_1, \tag{1.6}$$

$$\Delta L_2 = \int_{\mathrm{T}}^{\mathrm{R}} n(s_2)\mathrm{d}s_2 - s_0 = \Delta\varphi_2\lambda_2, \tag{1.7}$$

s_0 is the straight line between the transmitter T and the receiver R, $\lambda_1 = 0.19$ m and $\lambda_2 = 0.244$ m are the wavelengths of the carrier frequencies.

Due to ionospheric bending dependent on the frequency (dispersion), the signals L_1 and L_2 travel on different paths. A linear combination of equations (1.6) and (1.7) yields the elimination of the first order ionospheric impact

$$\Delta L_\mathrm{c} = \frac{f_1^2\Delta L_1 - f_2^2\Delta L_2}{f_1^2 - f_2^2}. \tag{1.8}$$

ΔL_c contains particular fractions of the neutral atmosphere (Foelsche 1999).

The atmospheric Doppler shift Δf is determined by

$$\Delta f = \frac{f}{c}\frac{\mathrm{d}(\Delta L_\mathrm{c})}{\mathrm{d}t}, \tag{1.9}$$

where f is the transmitted frequency.

A connection between the Doppler shift and the geometry of the occulting event is given by

$$\Delta f = \frac{f}{c}v_\mathrm{T}^r\cos\phi_\mathrm{T} + v_\mathrm{T}^\theta\sin\phi_\mathrm{T} + v_\mathrm{R}^r\cos\phi_\mathrm{R} - v_\mathrm{R}^\theta\sin\phi_\mathrm{R}, \tag{1.10}$$

v_T^r, v_T^θ, v_R^r, and v_R^θ are the radial and tangential components (denoted by r and θ) of the velocity vectors of the transmitter and the receiver, respectively.

Knowing these four parameters as well as Bouguer's law (1.5) with $n(\boldsymbol{r}_\mathrm{T}) = n(\boldsymbol{r}_\mathrm{R}) = 1$ and the geometry

$$(\pi - \alpha(a)) + \phi_\mathrm{R} + \phi_\mathrm{T} + \gamma = 2\pi, \tag{1.11}$$

the bending angle $\alpha(a)$ can be derived (cf., Kursinski et al. (2001)).

Another possibility deriving the ionosphere-corrected bending angle ($\alpha_\mathrm{c}(a)$) is to apply the linear combination directly to the bending (Vorob'ev and Krasil'nikova 1994)

$$\alpha_\mathrm{c}(a) = \frac{f_1^2\alpha_1(a) - f_2^2\alpha_2(a)}{f_1^2 - f_2^2}. \tag{1.12}$$

$\alpha_1(a)$ and $\alpha_2(a)$ are the uncorrected bending angles of the signals L_1 and L_2, which are obtained by equations (1.10), (1.11), and Bouguer's law.

This approach is known as "bending angle correction" and is preferred to the other approach (Steiner et al. 1999).

High Altitude Retrieval

The ionospheric correction does not remove the whole error arising from the ionospheric influence on the signal; a residual noise still remains. Beside this residual ionospheric noise other error sources can influence the retrieved bending angle, such as orbit uncertainties, local multipath errors, or receiver noise.

To minimize the errors resulting at high altitudes, background information is included in the retrieval process. By means of statistical optimization, the **b**est **l**inear **u**nbiased **e**stimator (BLUE), α_{opt}, is derived via

$$\alpha_{\mathrm{opt}} = \alpha_{\mathrm{b}} + \mathbf{B}(\mathbf{B} + \mathbf{O})^{-1}(\alpha_{\mathrm{o}} - \alpha_{\mathrm{b}}), \tag{1.13}$$

where α_{o}, α_{b} are the observed and the background bending angle profiles, $\mathbf{O}$ and $\mathbf{B}$ are the observation and background error covariance matrices, respectively.

The background information is either MSISE-90[6] climatologies or operational analyses of the ECMWF[7]. The retrieved data used in this work are obtained by implementation of ECMWF background information.

A detailed description of the high altitude retrieval can be found in Steiner et al. (2004), Gobiet and Kirchengast (2004), and an evaluation of the retrieved parameters is available from Gobiet and Kirchengast (2004) and Gobiet et al. (2005b).

The Abelian Integral Equation and the Abelian Inversion

The connection between the bending angle α, which is a function of the impact parameter a, and the index of refraction n is given by the Abelian integral equation:

$$\alpha(a) = 2a \int_{r_{\mathrm{t}}}^{\infty} \frac{1}{\sqrt{(nr)^2 - a^2}} \frac{\mathrm{d}\ln(n)}{\mathrm{d}r} \mathrm{d}r, \tag{1.14}$$

r_{t} is the radius at the tangent point.

Inverting this integral equation yields the expression of the refractivity index as a function of the radius at the tangent point. The Abelian inversion means

$$n(r) = \exp\left[\frac{1}{\pi} \int_{a_1}^{\infty} \frac{\alpha(a)}{\sqrt{a^2 - a_1^2}} \mathrm{d}a\right], \tag{1.15}$$

$a_1 = n(r)r$ is the impact parameter for the particular ray of which the tangent radius is $r = r_{\mathrm{t}}$.

It is quite evident that knowing the bending angle $\alpha(a)$, the Abelian inversion can be solved by numerical partial integration.

[6] MSISE-90 is a precursor model of the NRLMSISE-00 model, which is described in detail in Section 1.4.

[7] Details concerning analyses of the ECMWF (**E**uropean **C**entre for **M**edium-Range **W**eather **F**orecasts) are found in Section 1.2.

Derivation of Atmospheric Parameters

The knowledge of the refractivity index n enables the instant calculation of the refractivity N via equation (1.2).

At microwave wavelengths the refractivity N depends on the temperature T [K], the total pressure p [hPa], the partial pressure of water vapor e [hPa], the free electron density in the ionosphere n_e [electrons/m^3], the signal frequency f [Hz], and the mass of condensed water in the atmosphere W [g/m^3]:

$$N = 77.6\frac{p}{T} + 3.73 \times 10^5 \frac{e}{T^2} - 4.03 \times 10^7 \frac{n_e}{f^2} + 1.4W. \tag{1.16}$$

The first term of equation (1.16) represents the contribution of the dry atmosphere, which results from the polarizability of atmospheric molecules. The second term exists because of the permanent dipole moment of water vapor, which represents the so called moist term. The third and the forth term arise due to the ionospheric impact and the effect of scattering from liquid water droplets, respectively. The ionospheric correction (1.8) yields the elimination of the accompanying term, and the last term can be neglected because the content of liquid water is very small compared to the other terms in realistic dispersion.

The refractivity remains dependent on the dry and the moist term,

$$N = 77.6\frac{p}{T} + 3.73 \times 10^5 \frac{e}{T^2}. \tag{1.17}$$

This simplification is known as the Smith-Weintraub formula.

Derivation of Density: The moist part of equation (1.17) can be neglected in atmospheric regions where the specific humidity is lower than 10^{-4} kg/kg (Kursinski et al. 1997) and there the refractivity can be expressed as

$$N = 77.6\frac{p}{T}. \tag{1.18}$$

The refractivity N depends on the height h and can be combined with the density $\rho(h)$ using the state law of ideal gas:

$$\rho(h) = \frac{M}{77.6R} N(h) = \frac{M}{R}\frac{p(h)}{T(h)}. \tag{1.19}$$

$R = 8.314$ J/(K mol) is the gas constant and M is the mean molecular mass of dry air; $M = 28.964$ kg/kmol at an altitude below 80 km. The constant factor 77.6 has the unit [K/hPa]. $p(h)$ and $T(h)$ are the pressure and the dry temperature, respectively.

Derivation of Pressure: In regions below approximately 100 km (homosphere) the pressure can be derived from the equation of hydrostatic equilibrium, $\mathrm{d}p(h) = -g(h)\rho(h)\mathrm{d}h$, where g is the acceleration of gravity. Knowing $g(h)$ and $\rho(h)$, the pressure $p(h)$ can be derived

$$p(h) = \int_{h}^{\infty} g(h')\rho(h')\mathrm{d}h'. \tag{1.20}$$

Derivation of Dry Temperature: Using the ideal gas law (equation (1.19)) a second time, it is possible to derive the dry temperature profile:

$$T(h) = \frac{M}{R}\frac{p(h)}{\rho(h)}. \tag{1.21}$$

Derivation of Water Vapor: It is not possible to derive the content of water vapor in an atmospheric column without ancillary information. Auxiliary data result from independent observations and climatologies or meteorological analyses.

Characteristics of RO Measurements

Compared to conventional method of measurements, the RO technique provides several interesting properties.

1. Self-calibration: At the start of a setting event (which CHAMP measurements are) or at the end of a rising event, the unattenuated signal is measured. The self-calibration arises from the normalization of all measurements made during this event to the comparison measurement. Stability of the atmosphere during this event is required.

2. Long term stability: Due to self-calibration, no trend concerning the measurement equipment is shown. The expected drift is less than 0.1 K per decade. This property is of prime importance in climate monitoring.

3. Resolution: The vertical resolution $\Delta\zeta$ is determined by the diameter of the first Fresnel zone D_F. At GPS carrier frequency f_1, the vertical resolution is $\Delta\zeta = D_\mathrm{F} = 1.4$ km in the stratosphere. A better vertical resolving power is achieved near the earth's surface with $\Delta\zeta = D_\mathrm{F} \leq 0.5$ km.

   The horizontal resolution $\Delta\vartheta$ can be estimated knowing the vertical resolution. $\Delta\zeta = 1.4$ km yields $\Delta\vartheta = 270$ km; $\Delta\zeta = 0.5$ km corresponds to $\Delta\vartheta = 160$ km (Kursinski et al. 1997).

4. Accuracy: The accuracy of radio occultation data depends on external influences such as ionospheric conditions (depending on daytime and solar activity), properties of the sampled atmosphere (dry or wet), as well as the quality of the used instruments (e.g., signal-to-noise ratio (SNR)) and the retrieval (e.g., initialization of the Abelian integral). A detailed error analysis on radio occultation yields a retrieved temperature accuracy better than 1 K between about 4 km and 40 km at optimal conditions and better than 1 K between 10 km and 28 km height at worst conditions (Kursinski et al. 1997).

5. Global coverage: Satellites that receive GPS signals to measure phase delays are LEO satellites, which pass over the poles at an altitude of about 1 000 km in order to get global coverage. Profiles of atmospheric parameters over continental regions and above the oceans are obtained in a uniform manner, but there are more occultation events in mid and high latitude regions than in low latitude regions (cf., Section 1.1.4). More than 200 occultation events can be attained daily from the GPS receiver aboard one single satellite.

6. All-weather capability: GPS occultation observations are made at wavelengths of about 0.2 m, enabling the limb sounding of the atmosphere because these waves are almost not absorbed by clouds and aerosols. Nevertheless, tropical regions with high specific humidity cannot be scanned by the radio occultation limb sounding technique because the GPS signal breaks away.

1.1.3 Binning of Retrieved Profiles

(Author: B.C. Lackner)

To derive evenly distributed grid point temperature values ("climatologies"), the retrieved radio occultation profiles, which are without spatial regularity, are assigned to so-called "bins" – surface areas with a certain fixed dimensioning. Therefore, a fixed number of bins at all latitudes is essential, which leads to overlapping to maintain equal area. Occultation events in overlapping regions are assigned to more than one bin. This effect is stronger developed in polar regions and is also dependent on the desired bin-size (larger bins lead to more overlapping).
To generate climatologies, the bin size was chosen in a manner that at least three occultation events took place in every bin each month. For that reason the originally desired resolution of $10° \times 30°$ (latitude $\times$ longitude) was not possible, since during some months single bins with no occultation event occurred (see Figure 1.3).
Furthermore, the CHAMP profiles were interpolated to a vertical resolution of 500 m (ranging from earth's surface up to 35 km), whereby arithmetic means were computed for each altitude range of a bin.

Binning Resolutions

To compare CHAMP radio occultation profiles with other climatologies, three different forms of representations were chosen together with therefore suitable binning resolutions:

- Latitude $\times$ height plots: Consisting of four longitudinal sectors (four bins) representing the regions Eurasia-Africa (20°W – 70°E), Asia-Australia (70°E – 160°E), Pacific (160°E – 110°W), and America-Atlantic (110°W – 20°W). Each bin spreads over 10° latitude and 90° longitude (see Figure 1.4).

- Longitude $\times$ height plots: These graphs shall help to explore differences of high (60°N/S to 90°N/S), mid (30°N/S to 60°N/S), and low (30°S to 30°N) latitudes.

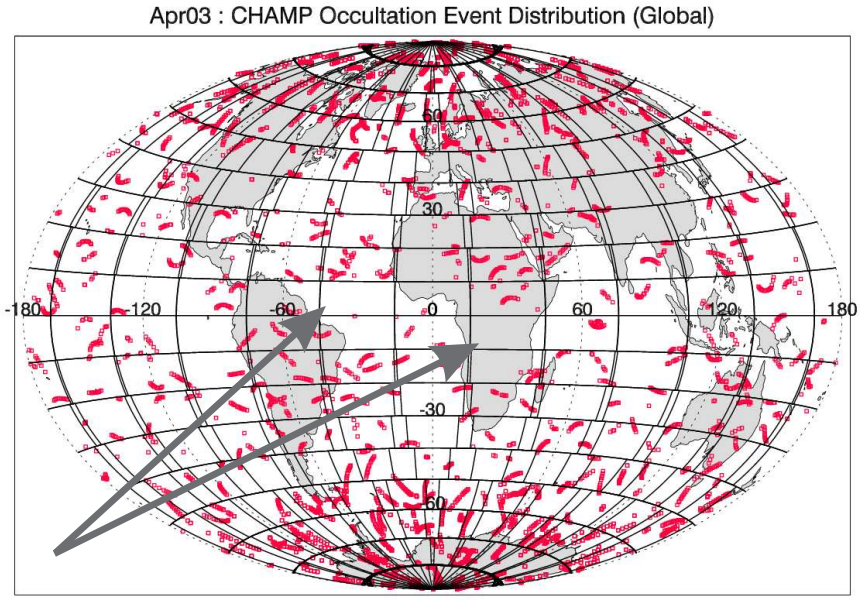


Figure 1.3: 10° × 30° binning for April 2003 – the red squares indicate the spatial distribution of radio occultation events. Two bins (marked with arrows) do not contain the required number of occultation events to calculate climatologies.

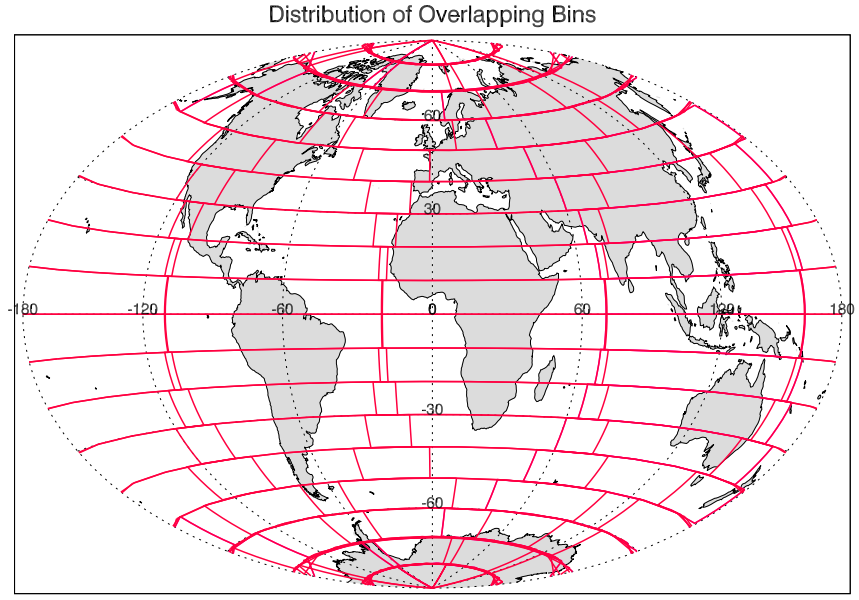


Figure 1.4: 10° × 90° binning resolution for latitude slices. The four resulting sectors contain following regions: America-Atlantic, Eurasia-Africa, Asia-Australia, and Pacific.

The bins are placed in a manner so that the Greenwich meridian is the center of one bin (see left graph in Figure 1.5) and have a resolution of 30° × 30°.

- Longitude × latitude plots: Maps are computed for four selected height levels (7 km, 15 km, 25 km, 35 km, respectively, 32 km for NCEP/NCAR data) using a more detailed binning resolution of 30° × 20° (see right graph in Figure 1.5).

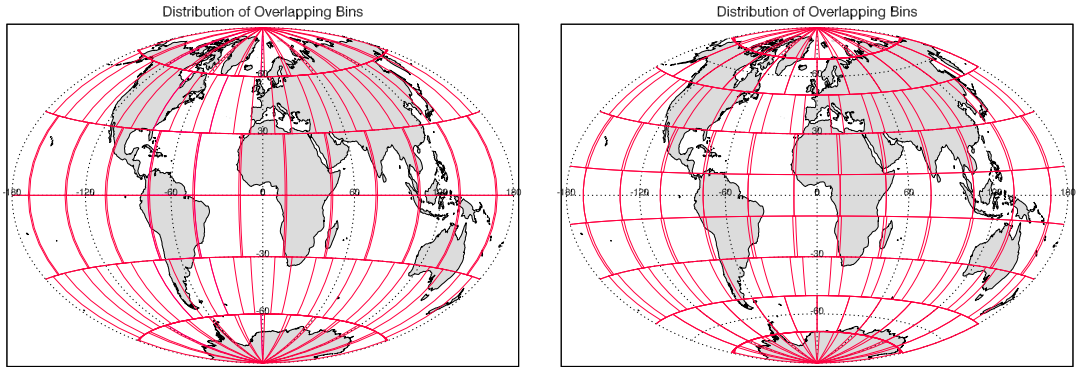


Figure 1.5: Binning resolutions. Left: 30° × 30° for longitude × height plots. Right: 30° × 20° for longitude × latitude plots.

1.1.4 Radio Occultation Data of the Selected Period

(Author: B.C. Lackner)

The total amount of retrieved RO profiles in the selected period (March 2002 to February 2004) sums up to 127 648. As the period includes 24 months, the monthly arithmetic mean of ROs is 5 319, which leads to about 175 RO profiles – on the average – per day. Anyhow, not all of these profiles can be used as different problems may occur during the retrieval process.

Radio Occultation Profiles and Their Quality Parameter

Each retrieved RO profile is signed with a quality parameter (called QF = **q**uality **f**lag), which gives information on problems encountered during the retrieval (Gobiet 2004). For further analyses, only profiles with quality flags equal to zero or two were applied.

- QF = 0: "No problems. Data may be used without restriction."
- QF = 2: "The observation error could not be estimated from data and was set conservatively to 50μrad. Use this kind of data only below 25 km!"

Table 1.1 shows the monthly amount of RO profiles, arranged according to their quality parameter.

	2002		2003		2004	
month	all	QF=0 or 2	all	QF=0 or 2	all	QF=0 or 2
Jan	–	–	5 592	4 667	5 267	4 365
Feb	–	–	5 051	4 164	4 763	3 995
Mar	4 755	3 902	5 030	4 058	–	–
Apr	5 793	4 894	5 067	4 152	–	–
May	6 029	4 944	5 869	4 839	–	–
Jun	3 938	3 188	5 906	4 763	–	–
Jul	5 366	4 466	5 713	4 542	–	–
Aug	6 628	5 443	5 420	4 350	–	–
Sep	5 938	4 856	4 794	4 028	–	–
Oct	5 629	4 417	5 324	4 462	–	–
Nov	4 302	3 345	4 819	3 928	–	–
Dec	5 322	4 176	5 333	4 257	–	–
Total	53 700	43 631	63 918	52 210	10 030	8 360

Table 1.1: Number of retrieved RO profiles (in the selected period) according to their quality parameter.

Most of the retrievals caused no problems (QF equals zero). More than 76 % of all data obtained this quality parameter wheras only 5 % of the profiles should just be used below 25 km (QF equals two).

Figure 1.6 shows the daily number of radio occultation events, which were used for the analysis (QF 0 and 2). It is worth mentioning that there are repeatedly some days with nearly no applicable events (such as in June and November 2002, March and November 2003). The noticeable high value of occultation events in the middle of November 2003 (November 15, 2003, 276 profiles) seems to be based on a failure.

Zonal Distribution of Radio Occultation Events

In regard to the zonal distribution of radio occultations, a clear relationship between the number of radio occultation events and the latitude can be determined. Due to the nearly polar orbit of CHAMP (inclination 87.2°), the larger earth surface at low latitudes is rarely covered with occultation events compared to the smaller earth surface at higher latitudes. The distribution of occultation events is symmetric with respect to the equator, local maxima appearing about every 30°, that is to say near 20°, 50°, and 80° north and south.

However, the smallest amount of events over all is found in the polar regions (north and south of 85°). Partitioning the earth in 37 latitude slices of five degree extension,

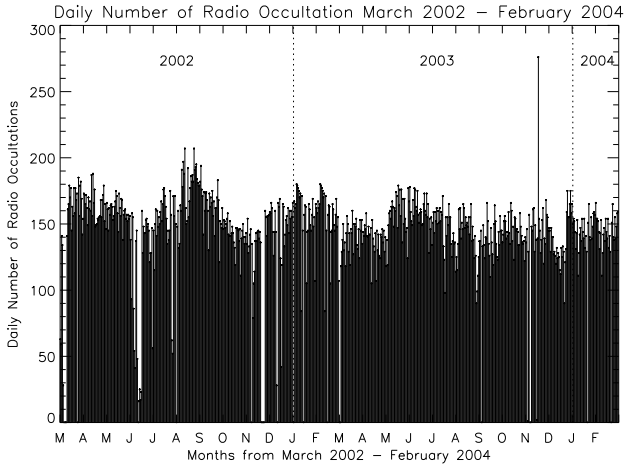


Figure 1.6: Daily number of radio occultation events from March 2002 to February 2004.

the number of RO events averages to 126 for each slice in January 2003. The "highest" northern slice (85°N – 90°N) only contains three RO events, the southern slice seven events, while the adjacent slices (80°N/S – 85°N/S) include more than hundred. The minimal number of events in the two highest slices seems to be a result of the inclination of CHAMP and the GPS-satellites and the small surface region yielding in a lower probability for an occultation event to take place.

Monthly distributions of ROs resemble those in Figure 1.7.

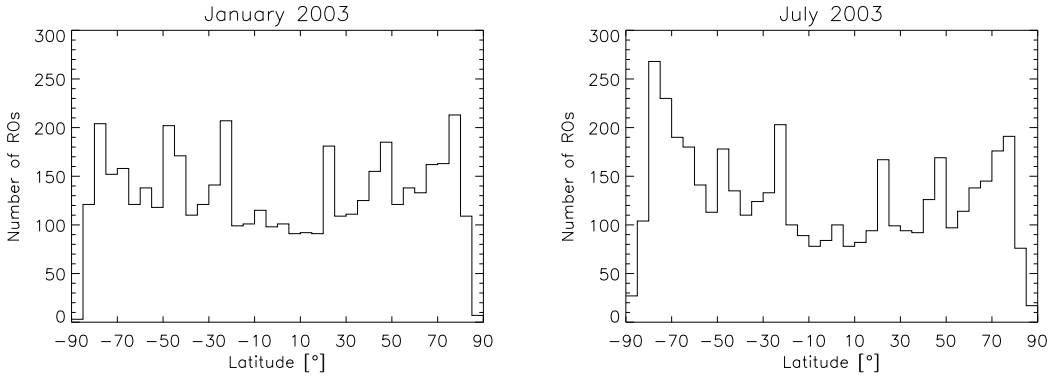


Figure 1.7: Histogram of the monthly distribution of the zonal number of ROs for two selected months (January, July 2003).

Meridional Distribution of Radio Occultation Events

As expected, there is nearly no relationship between the number of ROs and the latitude; the variance from meridional bin to bin is negligible (see Figure 1.8). A certain longitudinal period of accumulating events is barely discernible.

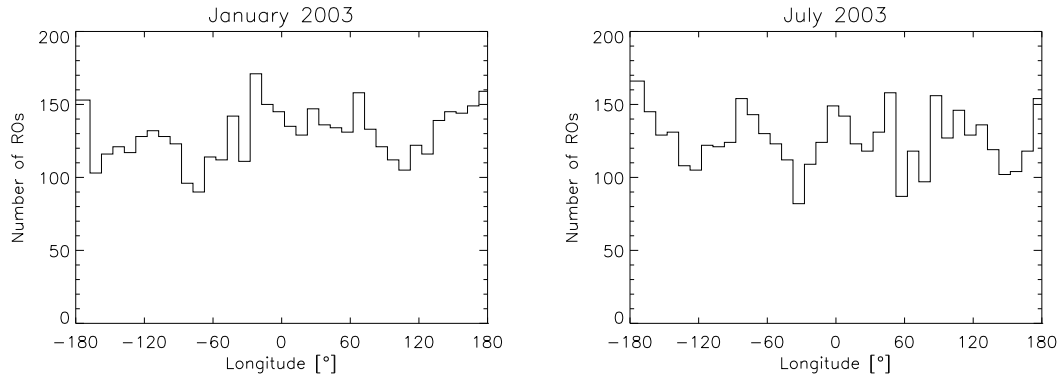


Figure 1.8: Histogram of monthly distribution of meridional number of ROs for two selected months (January, July 2003).

1.2 ECMWF Atmospheric Analysis Data

(Author: B. Pirscher)

The principal tasks of the **E**uropean **C**entre for **M**edium-Range **W**eather **F**orecasts (ECMWF) are the development of numerical methods for medium-range weather forecasting, the preparation of medium-range weather forecasts ($\leq$ 10 days) and their distribution to meteorological services, scientific and technical research directed to the improvement of these forecasts, and the collection and storage of meteorological data (Persson 2001).

Since November 25, 1997 the operational global analyses follow from the 4D-Var (**4-D**imensional **Var**iational) analysis assimilation algorithm (Bouttier and Rabier 1997/98) yielding analyses at the four main synoptic hours 00 UTC, 06 UTC, 12 UTC, 18 UTC.

1.2.1 The Data Assimilation and Analysis System

A general survey about data assimilation can be found in Bouttier and Courtier (1999), and a detailed description of the 4D-Var assimilation algorithm is available from Fischer (2001).

An analysis is an image of the best estimate of the true atmosphere's state at a given time. Because of too sparse observations, it is necessary to include background information and physical constraints to obtain an analysis close to the true state. This analysis technique is called data assimilation. The background information is an a priori estimate of the model state; it can be taken from the output of a previous analysis.

Data Availability

About 2 500 000 different observational data are used per 12-hour period. Figure 1.9 depicts some important data sources incorporated in ECMWF analyses.

- **Conventional:**
  - Surface:
    Weather stations (land, sea)
  - Profiles:
    Radiosondes,
    UHF/VHF profilers
  - Altitude: Aircrafts
- **Satellite:**
  - Imagery
  - Radiances
  - Scatterometers

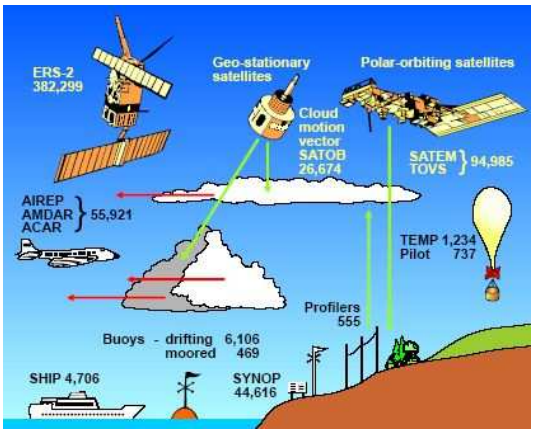


Figure 1.9: Conventional and satellite data are used in ECMWF analyses. The numbers represent the typical number of observations used to estimate the atmosphere initial conditions in a typical day (Buizza 2000).

Satellite data have become an essential part of the observing system. By means of satellite measurements, gaps in polar regions and above ocean areas can be closed.

There are polar orbiters (sun-synchronous circular orbits at an altitude of about 1 000 km), which almost pass over the poles and furnish data in high spatial resolution, and geostationary satellites (stationary with respect to one point on the earth's surface at an altitude of approximately 35 800 km), which send data in high temporal resolution.

Nearly 90 % of all assimilated data come from satellites (van der Grijn 2004).

A more detailed summary of observations used in the 4D-Var data assimilation at the ECMWF is given in Table 1.2.

Quality Control

The quality of data is of high relevance because of a significant effect on data assimilation. The quality control includes the comparison of the observations with model fields, the consistency between different data sources, and the self-consistency (Mendez 2004).

ECMWF Data Assimilation

The 4D-Var assimilation algorithm is a non-sequential (retrospective) assimilation. That means that observations made in the past and observations from the future are used. It requires, during the assimilation, to wait for the observations to be available over the whole time interval before the analysis procedure is able to start.

A schematic illustration of the 4D-Var assimilation technique is illustrated in Figure 1.10.

Observation type		Variables
SYNOP	synoptic surface observations	u, v, p_s (or z), r_h
AIREP	aircraft reports	u, v, T
SATOB	satellite cloud track winds	u, v
DRIBU	drifting buoy reports	u, v, p_s
TEMP	radiosonde soundings	u, v, T, q
PILOT	wind soundings	u, v
TOVS	satellite temperature soundings	T_b
PAOB	pseudo observations of surface pressure	p_s
SCATT	scatterometer reports	u, v

Table 1.2: Observation types used in the 4D-Var data assimilation and retrieved atmospheric parameters.

u, v:	wind components	p_s:	surface pressure	z:	geopotential height
r_h:	relative humidity	T:	temperature	q:	specific humidity
T_b:	brightness temperature				

Every six hours (00 UTC, 06 UTC, 12 UTC, 18 UTC) a 4D-Var is performed to assimilate the most recent observations using the previous forecast as background. The ECMWF data assimilation uses a 12-hour time window, from 9 hours before to 3 hours after the nominal analysis time (Persson 2001). That means that if a 12.00 UTC analysis should be computed, the earliest time observation information can enter in the analysis is 03.00 UTC, the latest time is 15.00 UTC. The analysis starts to be computed at 19.00 UTC. The algorithm is designed to find a compromise between the previous forecast at the beginning of the time window and the observed data. An analysis is optimal if it is closest in a root mean square sense to the true state.

An analysis never specifies the true state of the atmosphere. Because of errors in the background (estimation errors of the background state), observation errors (among other things instrumental errors), and analysis errors (estimation errors of the analysis state) an analysis will never be perfect.

From November 25, 1997 to September 12, 2000 the 4D-Var assimilation technique operated on a 6-hour time window (three hours before and three hours after the analysis); it has since been moved to the 12-hour cycling.

At the introduction of the 4D-Var analysis the spectral resolution of the model was T213L31[8]. After some changes in April 1998, in March 1999 (Untch and Simmons 1998/99), October 1999 (Fischer et al. 2000), and November 2000 (Buizza et al. 2001)

[8] T213L31: Horizontal resolution: T213, triangular truncation, resolving 213 waves around a great circle on the globe; vertical resolution: L31, 31 vertical levels between the earth's surface and 30 km height.

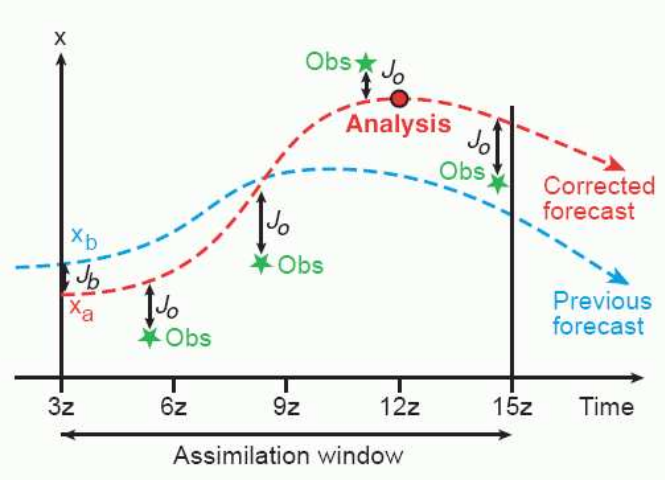


Figure 1.10: Simplified view of the 4D-Var analysis (Persson 2001). x_b denotes the background model state, x_a is the analysis model state, J_b represents the background term and J_o the observation term.

the resolution changed to $\mathrm{T_L}511\mathrm{L}60$[9].

The horizontal resolution $\mathrm{T_L}511$ is roughly equivalent to 40 km grid length at the mid-latitudes. The vertical resolution is divided into 60 height levels from the earth's surface up to a pressure of $p = 0.1$ hPa, which is about 64 km height. The levels are surfaces of constant pressure with highest resolution in the planetary boundary layer and lowest in the stratosphere and lower mesosphere.

Data Used in This Work

All data used in this work are available in the GRIB (**GRI**dded **B**inary) format. It is a bit-oriented data exchange format, which enables the transmission of large volumes of gridded data. A detailed description of this format can be found in Stackpole.

A lot of data follow from the 4D variational analysis, such as the 10 m wind components u and v, the relative and specific humidity, the temperature or cloud cover. In this work, only dry temperature data are used.

Not only because of the amount of data but also due to the similar resolution to the radio occultation data, the horizontal resolution of the ECMWF data is reduced to T42 ($\sim$ 300 km). The vertical resolution is still the same, 60 levels, up to 0.1 hPa. Because of this vertical limit (0.1 hPa corresponds approximately to 64 km), the data are extended for higher altitudes implementing the MSISE-90 model.

For each RO profile, a coinciding ECMWF profile (a so called "co-located" profile) was extracted from the analysis by spatial interpolation using the nearest time layer of the analysis.

[9] $\mathrm{T_L}511\mathrm{L}60$: Horizontal resolution: T511, triangular truncation, resolving 511 waves along the great circle on the globe; Vertical resolution: L60, 60 vertical levels between the earth's surface and 64 km height.

Monthly means of temperature data are also used in this work. ECMWF provides operational model level analysis data sets of temperature, averaged over a calendar month for each of the layers of time at 00 UTC, 06 UTC, 12 UTC and 18 UTC.

1.3 NCEP/NCAR Reanalysis Data

(Author: B.C. Lackner)

To explore the differences between the retrieved radio occultation data and analyzed data other than ECMWF analyses, NCEP/NCAR reanalysis temperature climatologies were applied.

As mentioned above, the process of data assimilation precedes the establishing of analyzed grids, which are the bases for most atmospheric research studies. However, inhomogeneities in these analyzed grids (caused by various changes over time) limit the usefulness of the data. Hence, spread over the whole world, different organizations have established reanalysis projects (e.g., ECMWF, NASA GSFC, NCAR), which aim to provide relatively "clean" sets of data for further analyses.

1.3.1 The Reanalysis Project

The "**N**ational **C**enters for **E**nvironmental **P**rediction" (NCEP) and "**N**ational **C**enter for **A**tmospheric **R**esearch" (NCAR) cooperated in the "Reanalysis Project" (Kalnay 1999) in order to reanalyze historical data using state-of-the-art models. The project started in 1989 at NCEP with the initial goal of building a "Climate Data Assimilation System", which is not affected by changes introduced by many improvements to the numerical weather prediction systems. While NCAR collects and organizes the land and marine surface data archives (including international sources such as the Japanese and European weather services JMA and ECMWF, as well as US military collections) and provides them to NCEP together with observed upper and aircraft observations, NCEP executes the processing, using a current and fixed version of the data assimilation and operational forecast model. The analyses, forecasts, quality controlled input data, etc. are again stored in NCAR's archives.

According to Kalnay (1999) the most difficult task was the assimilation of data from many different sources and formats as well as their quality control.

Initially, the reanalysis project focused on the period from 1985–1994 (Shea et al. 1994). At the moment various data sets for a huge number of atmospheric parameter are available from January 1948 up to December 2004[10]. The data sets are updated yearly.

There are two major products of the reanalysis:

[10] Temperature fields are among others obtainable as GRIB files from http://dss.ucar.edu/pub/reanalysis/rean_proj_des.html and as netCDF (**Net**work **C**ommon **D**ata **F**ormat) files from http://www.cdc.noaa.gov/cdc/reanalysis/reanalysis.shtml (February 2005).

1. Four-dimensional gridded fields of the global atmosphere (including different "re-forecasts"). Provided monthly means of gridded reanalysis fields were used for this work.

2. **B**inary **U**niversal **F**ormat **R**epresentation (BUFR) archives of the atmospheric and surface observations for the reanalysis period (including additional information to each observation as well as meta data[11]).

The gridded output variables are arranged in four classes, depending on the degree to which they are influenced by the observations and/or the model. Variables of type "A" are mainly determined by observations (upper air temperatures), type "B" variables are determined by both observations and models (variables near surface) and therefore may be improved by better models. Model-produced variables ("C" and "D"), such as surface fluxes and precipitation, should be regarded with caution.

The model used for NCEP/NCAR reanalysis is a T62 sigma coordinate system (horizontal resolution 208 km) with 28 height levels. A sigma (σ) coordinate system uses a transformed pressure coordinate for the vertical resolution. Sigma levels are defined by $\sigma = p/p_\mathrm{s}$, whereas p_s is the surface pressure defined on the model surface topography. One advantage of sigma level application is that in lower boundaries, sigma levels are approximately parallel to earth's smoothed surface. The 28 levels represent the troposphere and lower stratosphere[12]. The horizontal resolution corresponds to a $2.5° \times 2.5°$ grid with 144×73 grid points.

The assimilated observations are (Kalnay 1999):

- upper air rawinsonde[13] observations of temperature, horizontal wind, and specific humidity
- operational TOVS[14] vertical temperature soundings from NOAA polar orbiters over ocean
- temperature soundings over land only above 100 hPa
- cloud tracked winds from geostationary satellites
- aircraft observations of wind and temperature
- land surface reports of surface pressure

[11] Information about the data, which can be critical for correctly interpreting observations or derived results.

[12] Details concerning the levels are available from http://dss.ucar.edu/pub/reanalysis/model_vert_html (February 2005).

[13] According to Shea et al. (1994) a rawinsonde is a radiosonde (expendable balloon-borne instrument measuring pressure, temperature, and humidity and relaying the information to an observing station) of which the three dimensional position is measured as a function of time. Because the balloon drifts with the wind, the position and time information can be used to estimate the winds aloft. These upper air observations are referred to as "roab" data.

[14] **T**IROS–N **O**perational **V**ertical **S**ounder: The TIROS series of satellites were the first to be launched specifically for atmospheric studies.

- oceanic reports of surface pressure, temperature, horizontal wind, and specific humidity

The model includes parameterizations of all major physical processes such as convections, large scale precipitation, gravity wave drag, radiation with diurnal cycle and interaction with clouds, boundary layer physics, vertical and horizontal diffusion processes, etc. The resulting fields of the reanalysis are output every six hours, but monthly mean data are available too and are used through out this work.

1.3.2 netCDF Data Format

The reanalysis project aims to ensure the widest possible distribution of the derived products among researchers through CD-ROMs and different internet pages. Therefore the data are available as netCDF-files from NOAA's Climate Diagnostic Center[15], too.

NetCDF stands for "**net**work **C**ommon **D**ata **F**ormat" and is an interface for array-oriented data access. The software was developed at the Unidata Program Center[16] in Boulder, Colorado, USA (for more details see Rew et al. (2004)). Unidata is a National Science Foundation sponsored program aiming to make the best use of atmospheric and related data for promoting education and research. The netCDF software was intended to provide a common data access method for various Unidata applications.

The software functions as an I/O library, callable from C, FORTRAN, C++ or other languages. Likewise IDL's[17] I/O facilities allow to read scientific data formats (CDF[18], HDF[19] as well as netCDF) and were used in this context.

netCDF is "self-describing" and "portable" meaning that a data set includes information defining the data it contains and that the data in a data set are represented in a form that can be accessed by computers with different ways of storing integers, characters, and floating-point numbers. Compression of data is possible with netCDF, but it was not designed to achieve optimal compression of data. Hence, using netCDF may require more space than special-purpose archive formats that exploit knowledge of particular characteristics of specific data sets.

One of the goals of netCDF is to support efficient access to small subsets of large data sets. NetCDF uses direct access rather than sequential access. This can be much more efficient when the order in which the data are read is different from the order in which they were written, or when they must be read in different orders for different applications. As the netCDF-files from the Climate Diagnostic Center contain monthly temperature means from January 1948 to date, whereas in the context of this work only

[15] The data can be downloaded for free from http://www.cdc.noaa.gov/ → Climate Data and Resources (February 2005).

[16] The netCDF software package is available from ftp://ftp.unidata.ucar.edu/pub/netcdf/ (February 2005).

[17] **I**nteractive **D**ata **L**anguage

[18] **C**ommon **D**ata **F**ormat

[19] **H**ierarchical **D**ata **F**ormat

data from March 2002 to February 2004 were required, this feature proved to be very useful.

1.3.3 Adaption of NCEP Height Levels to RO Data Levels

NCEP/NCAR's temperature data and appurtenant geopotential heights are made available in two netCDF-files with 17 pressure levels each, namely 1000, 925, 850, 700, 600, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20, and 10 hPa (with a horizontal resolution of $2.5° \times 2.5°$ for each level), ranging from earth's surface to approximately 30 km height.

Conversion of Geopotential Heights Into Geometric Heights Above the Geoid

While the altitude profiles of the CHAMP radio occultation data are given as geometric heights (bearing on the earth's reference ellipsoid), the NCEP/NCAR altitudes are geopotential heights (measured from mean sea level – MSL). Thus, to compare these two data sets, it was necessary to convert NCEP/NCAR's geopotential altitudes into geometrical ones.

$\gamma(h, \varphi)$, the normal gravity, is a function of both geometric altitude (h) and geodetic latitude (φ).

The height dependency of gravity can be derived from the hydrostatic equation assuming spherical layers. Instead of an air column, a cone is studied of which the sections coincident with the level of R_E and $(R_\mathrm{E} + z)$ being equal to $A_0 = 1\ \mathrm{m}^2$ and A (see Figure 1.11). The horizontal extension of the cone section increases with height, whereas gravity decreases.

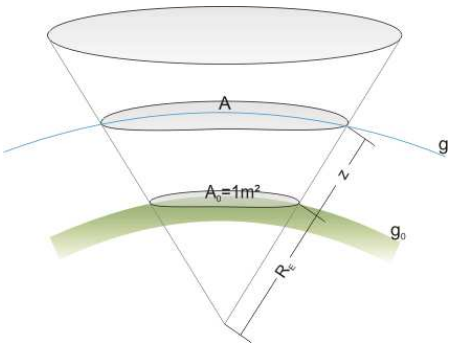


Figure 1.11: Derivation of hydrostatic equation (spherical layers).

From the similarity we conclude:

$$\frac{A}{A_0} = \frac{(R_\mathrm{E} + z)^2}{R_\mathrm{E}^2} \tag{1.22}$$

As gravity is reciprocal to the square of the distance from earth's center, the height dependency of gravity can be formulated as follows:

$$\frac{g}{g_0} = \frac{R_{\mathrm{E}}^2}{(R_{\mathrm{E}} + z)^2} \tag{1.23}$$

Geopotential heights include gravity dependence on the latitude as well. Therefore, the geopotential is weighted with the normal gravity value from 45° latitude (γ_{45}). This latitude was chosen because it was used by the World Meteorological Organization to calibrate barometers. As the earth's gravity model changed since this, a value of 9.80665 m/s^2 and a latitude of 45.542° is used.

The issue of converting between geopotential (z) and geometric (h) height can be realized by writing the expression for geopotential height in differential form:

$$\mathrm{d}z = \frac{\gamma(z)}{\gamma_{45}}\,\mathrm{d}h \tag{1.24}$$

$$\text{with } \gamma(z) = \gamma(\varphi)\,\left(\frac{R_{\mathrm{E}}}{R_{\mathrm{E}} + h}\right)^2 \tag{1.25}$$

R_{E} is the mean earth radius. Substituting $\gamma(z)$ in the differential equation results in following expression:

$$\mathrm{d}z = \frac{\gamma(\varphi)}{\gamma_{45}}\,\left(\frac{R_{\mathrm{E}}}{R_{\mathrm{E}} + h}\right)^2\,\mathrm{d}h \tag{1.26}$$

The latitude dependence of gravity weighted by the 45° latitude value is approximated with following series, where $fGrav$ stands for $\frac{\gamma(\varphi)}{\gamma_{45}}$:

$$fGrav = 0.99731 + 0.0053\,\left(\sin(\varphi)\right)^2 \tag{1.27}$$

Using differences instead of the differential form and converting the formula for geometric height, we can write:

$$\Delta h = \frac{1}{fGrav}\left(\frac{R_{\mathrm{E}} + z}{R_{\mathrm{E}}}\right)^2\,\Delta z \tag{1.28}$$

$$h(1) = \frac{1}{fGrav}\left(\frac{R_{\mathrm{E}} + 0.5\;z(1)}{R_{\mathrm{E}}}\right)^2\,z(1) \tag{1.29}$$

$$h(j+1) = h(j) + \frac{1}{fGrav}\left(\frac{R_{\mathrm{E}} + 0.5\,[z(j+1) + z(j)]}{R_{\mathrm{E}}}\right)^2\,[z(j+1) - z(j)] \tag{1.30}$$

The mean earth radius used in this iteration process is 6 371 km.

The hence received geometric heights for NCEP/NCAR temperature profiles were used in further considerations.

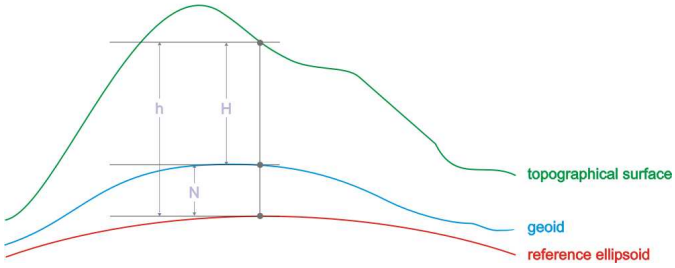


Figure 1.12: Altitude scales. Orthometric height (H), ellipsoid height (h), geoid undulation (N).

As geopotential heights bear on the geoid[20], while CHAMP heights are based on the reference ellipsoid, the undulation of the geoid is not included in this assessment. Figure 1.12 shows the connection between orthometric height (H, height above the geoid), ellipsoid height (h, height above the reference ellipsoid) and the undulation of the geoid (N), which is the difference between these two items. The undulation of the geoid rises to a magnitude of about 100 m and was not included in this respect.

Interpolation of Heights

After converting the geopotential heights into geometric ones, the NCEP/NCAR data had to be adapted to a regular height grid with 500 m steps. The left hand graph in Figure 1.13 shows the profile of one grid point composed of the 17 pressure level data (which have already been transformed from geopotential height into geometric height). The 17 levels represent the temperature progression in troposphere and lower stratosphere well, and the equidistant grid was achieved by linear interpolation between the given points (the result is depicted on the right graph in Figure 1.13, with the blue asterisks marking the given pressure level values and the red squares the interpolated equidistant grid values).

1.4 MSIS Data

(Author: B.C. Lackner)

1.4.1 Model Description

NRLMSISE-00 (for further details see Picone et al. (2002) and Picone et al. (2004)) is an empirical model of the neutral atmosphere, which enables one to calculate number densities of certain elements (He, O, N_2, O_2, Ar, H, N), total mass density, and temperature from earth's surface to thermosphere. NRL stands for "**N**aval **R**esearch **L**aboratory" (where the new model has been developed) and MSIS for "**M**ass **S**pectrometer and

[20]The geoid is the equipotential surface of the geopotential that coincides in the mean with the mean sea level.

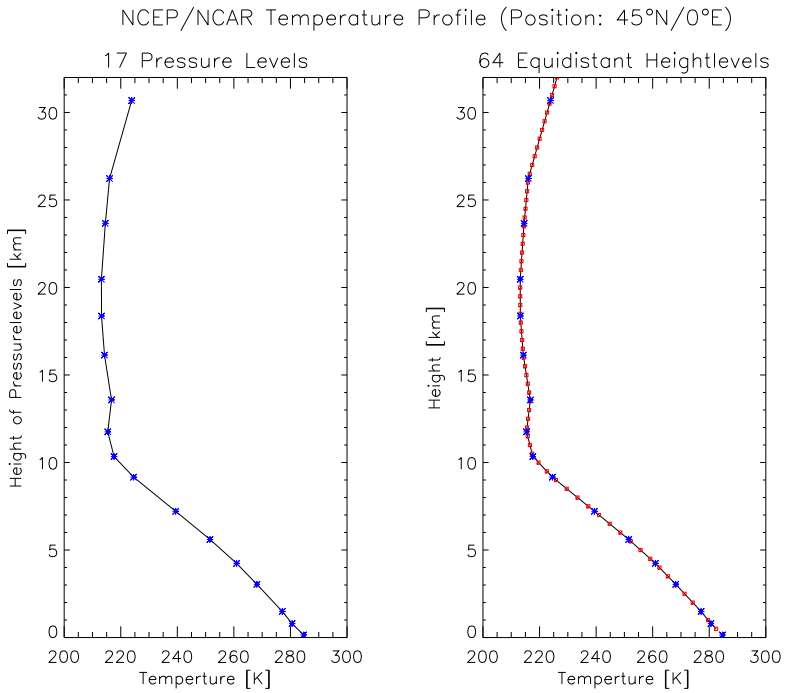


Figure 1.13: Interpolation of pressure level heights. Left: Profile of 17 pressure level data, which have already been transformed from geopotential height into geometric height (marked with blue asterisks). Right: Equidistant height grid (red squares) achieved by linear interpolation between pressure level data (blue asterisks).

Incoherent **S**catter Radar"; the primary data sources for the first model (MSIS-86). The E indicates that the model extends from ground to space (the original MSIS-86 model only covered altitudes above 90 km). NRLMSISE-00 is a major upgrade of the MSISE-90 model in the thermosphere and includes further observed values such as satellite drag data, revised O_2 and O data and an additional nonlinear solar activity term for high altitudes. The model does not depend on the calendar year.

Apart from that, the NRLMSISE-00 model complies with MSISE-90 model. Below 72.5 km, the model bases on tabulation of zonal average temperature and pressure by Barnett and Corney, which was also used for CIRA-86 (**C**OSPAR **I**nternational **R**eference **A**tmosphere; see Section 1.5.1); below 20 km additional data from the National Meteorological Center were applied. The model should be used for studies reaching across several atmospheric boundaries rather than for specialized tropospheric or stratospheric analyses.

A FORTRAN source code (driver and subroutine) for all model versions is available from the internet[21]. The database underlying the code covers several decades (1961–

[21] http://uap-www.nrl-navy.mil/models_web/msis/msis_home.htm (November 2004).

1997, exact periods depend on included data sources) and the model takes into account statistical variability while interpolating among or extrapolating the underlying data sets to estimate composition, temperature, geophysical conditions, and locations that are not directly covered by the database. Spherical harmonics are applied to represent the spatial variability of the key parameters defining temperature and number density profiles. The fundamental variables define nodes and gradients of the temperature profile for altitudes below 120 km. MSISE-90 and NRLMSISE-00 coefficients are the same below 72.5 km, since all of the "new" data only relate to the thermosphere.

The NRLMSISE-00 input variables to calculate temperatures below 80 km are:

- year (ignored in current model) and day of year (from 1 to 365)
- universal time – UT [seconds]
- local time is included as a function of UT and longitude
- altitude [km]
- geodetic longitude and latitude [degree]
- constant for solar $F10.7$ cm flux (should be set to "150" below 80 km)
- constant for daily magnetic index AP (should be set to "4" below 80 km)
- mass number ("0" for temperature)

1.4.2 Background Information About the Building of NRLMSISE-00 Climatologies

All considerations in relation to this study were based on monthly means. As the NRLMSISE-00 Fortran code calculates daily values, the annual temperature variation (see Figure 1.14) was examined at first in order to determine whether the temperature values from the middle of every month represent a good approximate value for the monthly mean, instead of calculating the monthly mean using all days of a month.

Hereafter, two ways were used to calculate monthly temperature means (for January) at one gridpoint (45°N, 0°E):

- usage of the middle (15^{th} day) of the month
- calculation of the arithmetic mean of all days for the respective month

The resulting monthly means differ in the second decimal point: whereas the arithmetic mean of all days is 250.884 K, the temperature of the 15^{th} day of the month January is 250.849 K.

Because of the minor deviation, the temperature values of the middle of every month were used for further contemplations. The NRLMSISE-00 climatologies were calculated

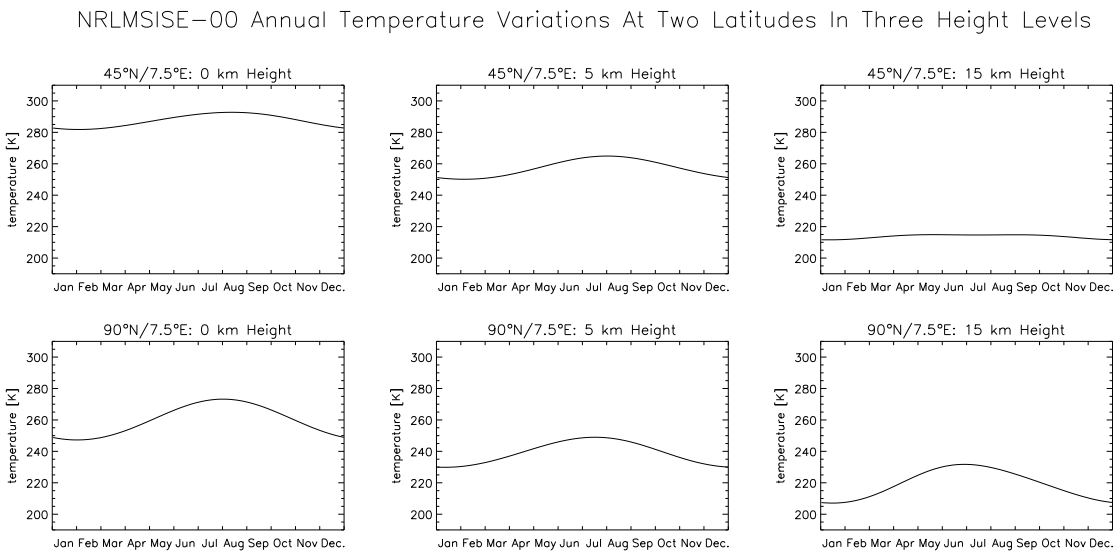


Figure 1.14: Annual temperature variations of NRLMSISE-00 at three different heights (left column: surface, middle column: 5 km altitude, right column: 15 km altitude) at mid (top) and high latitudes (bottom).

for 70 hight levels (from 0.5 km to 35 km, 0.5 km steps) with 665 grid points ($10° \times 10°$) each. After that, this data set was used to compute the desired resolutions ($10° \times 90°$, $30° \times 20°$, $30° \times 30°$) to compare NRLMSISE-00 with CHAMP RO temperature data.

To verify the correctness of the modified driver and the results of the subroutine, maps and longitude-height-graphics were compared with results from EGOPS[22]. An example of a NRLMSISE-00 map (altitude: 7 km) is shown in the left graph of Figure 1.15; a longitude $\times$ height plot (Eurasian-African sector) in the right graph of the Figure.

1.5 CIRA Data

(Author: B. Pirscher)

1.5.1 CIRA-86 Model Description

The **C**OSPAR[23] **I**nternational **R**eference **A**tmosphere, 1986, is a reference atmosphere of zonal mean temperature, geopotential height/pressure, and zonal wind. A detailed description is given in Fleming et al. (1990).

CIRA-86 is the fourth CIRA edition; former issues appeared in 1961, 1965, and 1972. These models were based on direct observational data of rockets and satellites as well as atmospheric models. They contained different kinds of atmospheric properties in

[22] **E**nd-to-End **G**NSS **O**ccultation **P**erformance **S**imulator
[23] COSPAR: **CO**mmittee on **SPA**ce **R**esearch

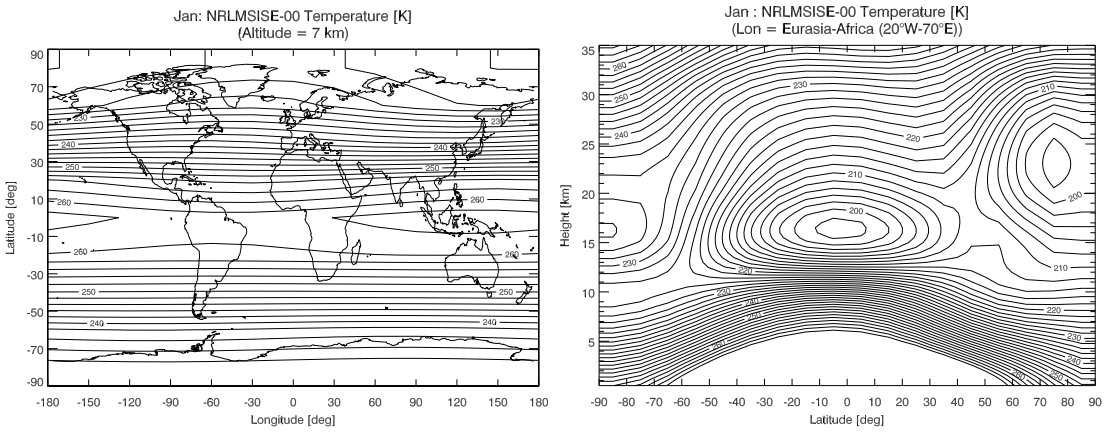


Figure 1.15: NRLMSISE-00 temperatures (January 2003). Left: Map of temperature distribution at 7 km altitude. Right: Latitude × height slice of Eurasian-African sector.

different altitude regions (altogether between 25 km and 2 000 km; mean temperature profiles were available from 30 km to 300 km).

The 12 monthly tables of CIRA-86 comprise data sets of temperature, zonal wind, and geopotential height in constant pressure coordinates and temperature, pressure, and zonal wind in constant altitude levels. The data are given from 80°S to 80°N and extend from the earth's surface (exception: pressure data set starts at 20 km) to approximately 120 km. The latitudinal resolution is 5° in pressure coordinates and 10° in altitude coordinates; the vertical resolution is $-0.25\ln(p/p_0)$ ($p_0 = 1013$ hPa) and 5 km respectively.

Three data sources are incorporated in the CIRA-86 model: the "Global Atmospheric Circulation Statistics, 1958 – 1973", compiled from Oort, the "Middle Atmosphere Reference Model Derived from Satellite Data", and the MSIS-83 empirical model.

Global Atmospheric Circulation Statistics, 1958 – 1973

The troposphere and lower stratosphere (ground to approximately 20 km) of the CIRA-86 model are taken from Oort's "Global Atmospheric Circulation Statistics, 1958 – 1973". Zonal mean temperatures and zonal wind are available in a latitudinal resolution of 5° (between 80°S and 80°N) in specific pressure levels. Data from five stations (U.S. National Meteorological Center, National Center for Atmospheric Research, Ocean Station Vessels, British Meteorological Office, and National Climatic Center) were implemented and edited.

Middle Atmosphere Reference Model Derived From Satellite Data

Because of Barnett and Corney, the publishers of the "Middle Atmosphere Reference Model Derived from Satellite Data", these data are often called "BC data". They are implemented in the CIRA-86 model in the stratosphere and mesosphere (15 km to 80 km). The data contain temperature, zonal wind, and geopotential height in constant

pressure coordinates as well as temperature, pressure, and density in constant altitude coordinates. The data are mainly derived from measurements of Nimbus 5 **S**elective **C**hopper **R**adiometer (SCR) Nadir Sounder as well as Nimbus 6 **P**ressure **M**odulator **R**adiometer (PMR) Nadir Sounder, but at lower altitudes (1000 hPa to 50 hPa) they contain data from Oort's atlas. At about 30 hPa they use data from analyses made by the Free University of Berlin.

MSIS-83 Empirical Model

Between a height of 86 km and 120 km, the CIRA-86 model bases on the MSIS-83 empirical model. Origins of these data are satellite-, rocket-, and ground based measurements. Moderate solar and low magnetic activity were supposed when implementing the values of temperature and total densities.

Combination of the Data Sets

The values of the models were merged to obtain a smooth transition.

In case of the zonal mean temperature, the values were obtained by merging the data sets of the "Middle Atmosphere Reference Model Derived from Satellite Data" and the "MSIS-83 empirical model", as mentioned above. The data of the Oort's atlas are included in the BC data set. So, up to about 0.01 hPa (approximately 80 km) only BC temperatures were incorporated in the CIRA-86 model; between 0.01 hPa and 0.002 hPa (between about 80 km and 90 km) the data were smoothed using an elemental Gaussian filter (weights: 1/4, 1/2, 1/4); above 0.002 hPa exclusively MSIS-83 empirical model data of the thermosphere were used.

1.5.2 CIRA86aQ_UoG Model Description

The name CIRA86aQ_UoG is put together from the CIRA-86 model, "aQ" meaning "and humidity" (Q is the symbol of humidity), and "UoG" standing for "University of Graz" – the point of origin of the model. A technical report of the model is given by Kirchengast et al. (1999).

Four large modifications were made in this model. The first one was the enhancement of the pressure tables from 20 km down to the ground, second the substitution of the wind tables through humidity tables, third the improvement of the resolution of the temperature and pressure tables (next to humidity tables) in regions below 15 km (from 5 km to 1 km), and forth the latitudinal dimension enhancement from 80°S and 80°N to 90°S and 90°N.

Temperature Tables of the CIRA86aQ_UoG Model

As mentioned above, the vertical resolution of the temperature below 15 km was improved to 1 km in the CIRA86aQ_UoG model. The mathematical method used was the cubic spline interpolation. The latitudinal extension was achieved by doubling the values of 80°S/80°N.

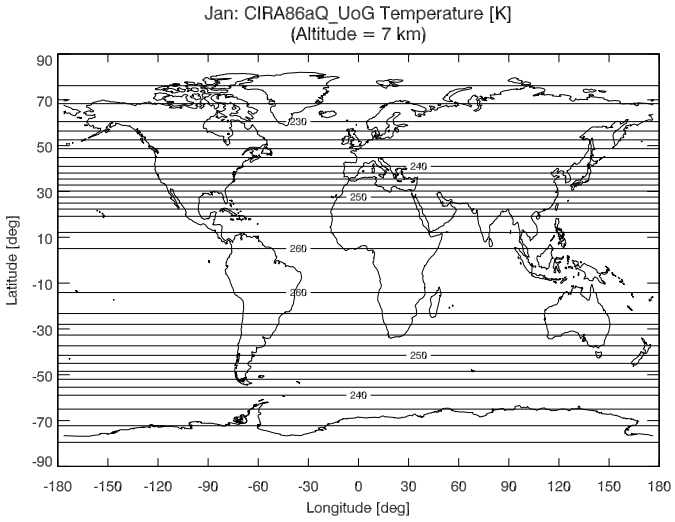


Figure 1.16: CIRA86aQ_UoG model: Temperatures at 7 km altitude.

The CIRA86aQ_UoG Fortran95 Model

The CIRA86aQ_UoG Fortran95 Model (Kirchengast et al. 1999) is implemented in the EGOPS software (details to this software can be found in Kirchengast et al. (2002)).

The model enables one to calculate temperature, specific humidity, water vapor pressure, (total) pressure, mass density, and refraction. The height-, latitude-, and longitude grid can be selected for a fixed universal time and month (January to December).

It is possible to compute a dry or a moist atmosphere and to achieve a linear or a cubic spline interpolation (both vertical and/or latitudinal), but for this it is necessary to go into the source code of EGOPS. The primary selected atmosphere (which is still used) is a moist one, the vertical interpolation type is a spline interpolation, and the latitudinal interpolation is linear.

The calculations of the CIRA86aQ_UoG temperatures are based on the following input:

- month (January to December),
- height grid: 0 km to 50 km, resolution: 0.5 km,
- latitude grid: 90°S to 90°N, resolution: 5°,
- longitude grid: $-180°$ to $+180°$, resolution: 5°,
- time: 0 UT.

Figure (1.16) shows the CIRA86aQ_UoG temperatures at 7 km height. By means of this map, the zoning of the atmospheric parameter is clearly evident.

2 Errors

2.1 Errors Comparing Observation and Reference Data

(Author: B.C. Lackner)

Nothing in our world can be measured without some error. As fas as users do not deny or ignore the error of a measurement, this shall not be a problem. Errors can arise from different sources, and understanding the types of errors may allow one to consider their effects on a measurement. In regard to CHAMP radio occultation retrieval, three errors are to be contemplated. It has to be mentioned that in this context "errors" mean the differences between the radio occultation data set relative to another selected reference data set.

2.1.1 Observational Error – Bias

A systematic deviation in a measurement is defined as a bias. It can be caused by faulty equipment, subjectivity of measurer, environmental impacts (e.g., refraction) or other "undefinable" effects. Biases are differences between the data gathered and what the data are thought to present. A bias is a tendency of the data to fall more to one side of the average than the other. Hence this systematic error cannot be reduced by enlarging the sample – when data are biased, gathering bigger samples means that the average of the data is certain to differ from the expected value.

In the case of our investigation, the differences between the retrieved CHAMP temperature profiles and the co-located "true" ECMWF ("model") temperature profiles for each bin result in the bias (the principal is illustrated in Figure 2.1). In equation (2.1) N stands for the number of profiles contributing to the temperature value in the analyzed bin.

$$\Delta T^{\text{bias}} = \frac{1}{N} \sum_{i=1}^{N} \left(T_i^{\text{retr}} - T_i^{\text{true}} \right) \tag{2.1}$$

The total observational error (root-mean-square error of bias), including systematic and random errors, results from the root of the sum of squared bias and standard deviation. In the calculation routine, the number of occultation events in each bin was checked, and only for a minimum of three events, the bias was computed by forming differences between CHAMP and co-located ECMWF profiles and then averaging all the differences in the bin.

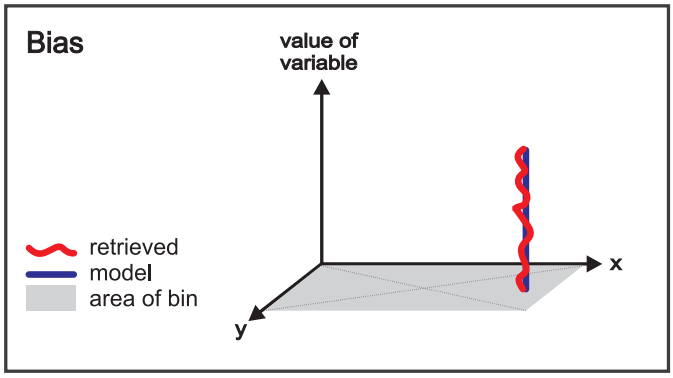


Figure 2.1: Calculation of bias between CHAMP radio occultation profiles ("retrieved") and ECMWF temperature profiles ("model") for a selected bin.

2.1.2 Sampling Error

In general, a sample is a part of the total; sampling then is the selection of a subset from a larger whole. The sampling error is associated with an estimate due to sampling. As more observations are gathered, the sampling error tends to balance itself out (whereas the bias persists).

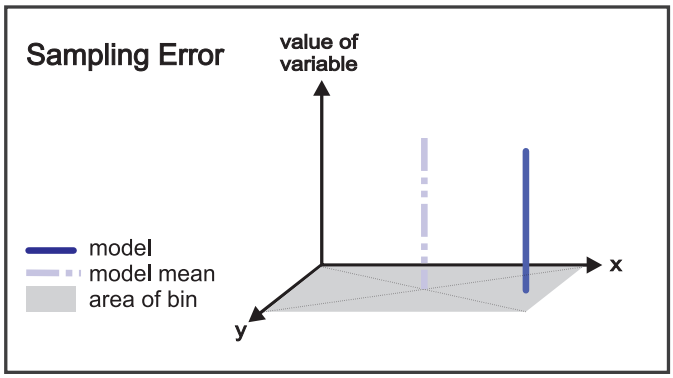


Figure 2.2: Sampling error: Difference between "true" ECMWF mean temperature field and mean field of CHAMP profiles in a selected bin.

Since the scanning of the atmosphere with radio occultation measurements is always discrete, variations in both time and location occur. The sampling error of the radio occultation profiles is determined by comparing the "true" (ECMWF) mean temperature field at the location and date of the occultation event of a bin and the mean field obtained from "true" profiles ("model mean" $\overline{T^{\text{true}}}$) in this bin (see Figure 2.2).

$$\Delta T^{\text{sampling}} = \frac{1}{N} \sum_{i=1}^{N} \left(T_i^{\text{true}} - \overline{T^{\text{true}}} \right) \tag{2.2}$$

Owing to the nearly polar orbit of CHAMP, occultation events are (referring to the same surface area) rarer at lower latitudes. But since tropical temperature variations are rather humble, the sampling error in these regions is not that big.
The influence of the local time of radio occultation events on the sampling error was also examined and will be dealt with in detail later on.

2.1.3 Total Error

Finally, the total climatological error is a combination of both observational error (systematic and random errors) and sampling error. For every bin it is the difference between the mean CHAMP and mean "model" fields (see Figure 2.3).

$$\Delta T^{\text{total}} = \sqrt{(\Delta T^{\text{observational}})^2 + (\Delta T^{\text{sampling}})^2} \tag{2.3}$$

The total error was not only computed for the ECMWF-"model" but also for CIRA86aQ-UoG, NRLMSISE-00, and NCEP/NCAR climatologies; as for the latter three models just mean fields were available (and not co-located profiles, which are needed to calculate bias and sampling error).

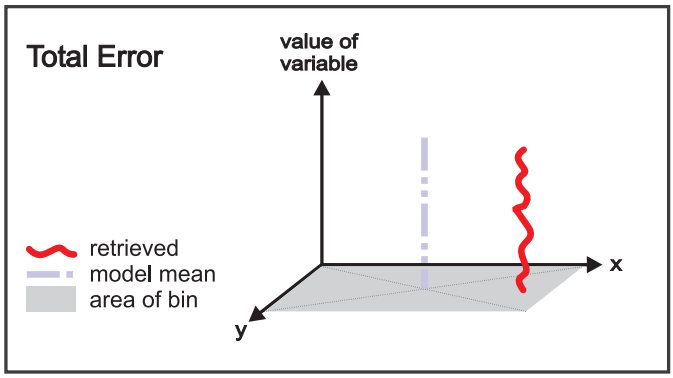


Figure 2.3: Total error: Difference between mean CHAMP and mean model field.

2.2 Local Time Considerations

(Author: B.C. Lackner)

The lack of continuity in the coverage, leading to sampling error, is a characteristic problem of (low earth-orbiting) satellite data. In terms of temperature data retrievals, the local time of the occultation events plays an essential role because of distinct daily temperature variations. A monthly shift of the local time of a certain (meridional) sector's occultation events could dupe a temperature trend without physical relevance – simply caused by an inappropriate sampling interval. To explore the retrieved data behavior, the local time for each event was calculated ($\text{LocalTime}_{\text{event}} = \text{UTC}_{\text{event}} + \lambda_{\text{event}} \cdot \frac{24}{360}$).

2.2.1 Investigation of Monthly Local Time Distribution

To get a first impression about the local time distribution of RO events, their local time was plotted depending on the longitude of the event (see upper two graphs in Figure 2.4). The graphs show that the events are not uniformly distributed in time but tend to accumulate "twofold" during a month with a time lag of roughly twelve hours in between. The time range of the event accumulation varies about three hours from month to month, as can be seen clearly from the histograms in Figure 2.4. While in May 2003 the peaks of the bimodal distribution of the histogram occur in the early morning (between 3 a.m. and 6 a.m.) and afternoon (3 p.m. to 6 p.m.), while one month later in June 2003 the peaks move to midday (12 a.m. to 3 p.m.) and midnight (12 p.m. to 3 a.m.). This scheme applies to the remaining months as well.

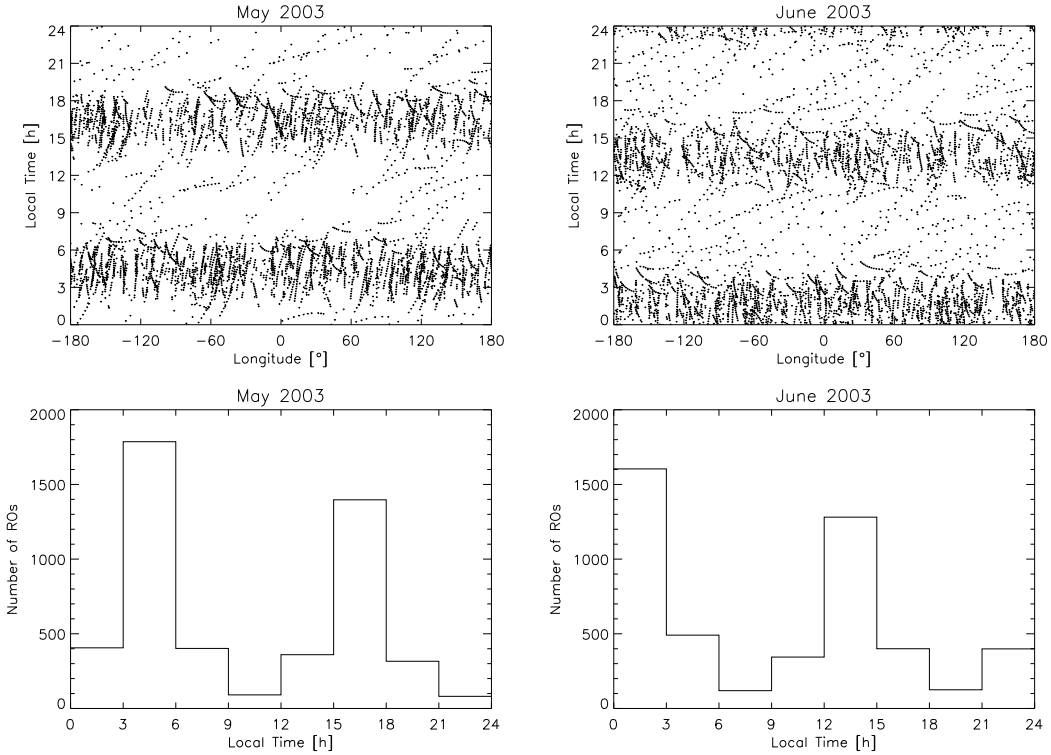


Figure 2.4: Local time distribution of RO events May and June 2003. Top: Local time of radio occultation events as a function of longitude. Bottom: Histogram of number of radio occultation events with three hour time-steps. A time-shift of approximately three hours per month is clearly visible.

2.2.2 Investigation of Seasonal Local Time Distribution

As expected, the local time influence fades when seasons are considered instead of months. The graphs seem to turnabout. In the histograms, two narrow gaps remain, while two peaks were formed in monthly considerations, with a twelve hour time lag in between. The better (although not yet "perfect") distribution is shown in the upper graphs of Figure 2.5 as well.

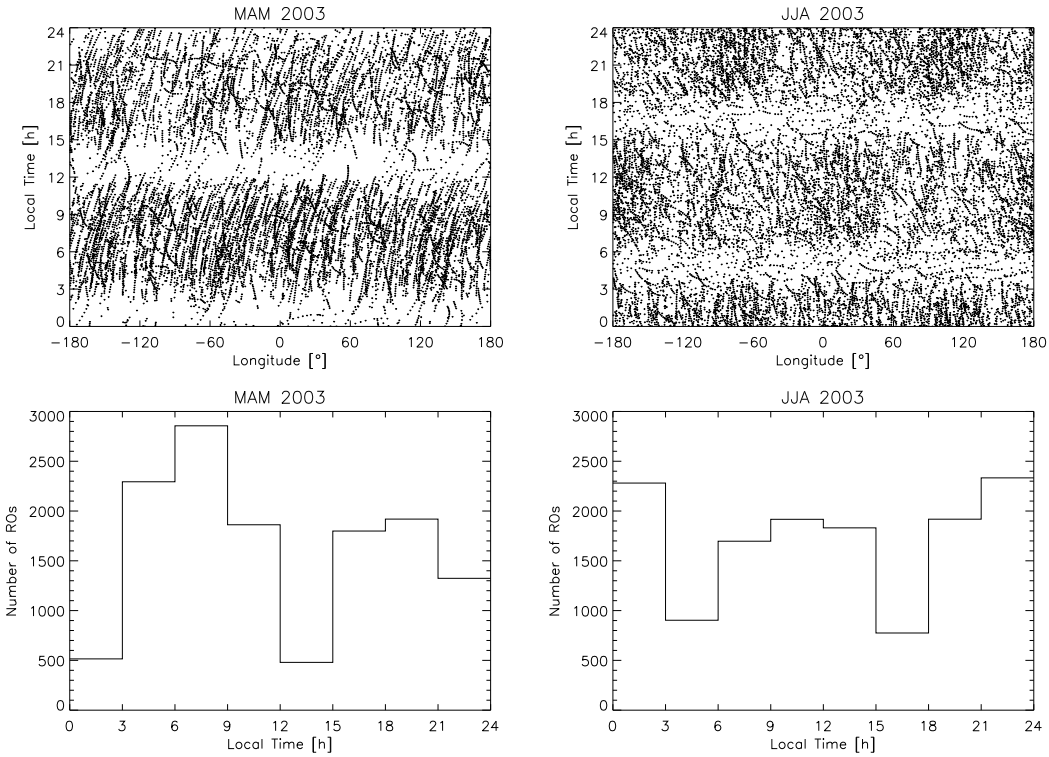


Figure 2.5: Local time distribution of RO events during two seasons (March, April, May – MAM and June, July, August – JJA) in 2003. Top: Local time of radio occultation events as a function of longitude. Bottom: Histogram of number of radio occultation events with three hour time-steps.

3 Comparison of Data

3.1 CHAMP Radio Occultation Data and ECMWF Analysis Data

(Author: B. Pirscher)

On account of the availability of co-located ECMWF profiles, the comparison of CHAMP RO climatologies and ECMWF analyses is divided in the appraisal of the bias, the sampling error, and the total error. The results below 5 km height will not be analyzed because of the standard geometric optics approach utilized in the retrieval (the same applies to the analyses of the other data as well).

3.1.1 Bias

The mean deviation between CHAMP climatologies and ECMWF analyses is referred to the bias, resulting either from the CHAMP RO measurement and the corresponding retrieval process or from the ECMWF analysis and the appendant data assimilation system.

Generally, the bias is marginal, but in some regions larger differences can be found. They are situated at the low latitude tropopause, in the tropical region between a height of 25 km and 30 km, at high latitudes in the southern hemisphere (in winter) as well as at all latitudes above approximately 29 km height. These features are shown in Figure 3.1.

Even though ECMWF analyses are incorporated in the retrieval process at high altitudes, the CHAMP RO climatologies are not dominated by the background information by 35 km. Interestingly, a positive bias, which constitutes up to +2 K, encounters above 31 km anyway. A detailed investigation of the bias arising in the middle and high stratosphere was done by Gobiet et al. (2005b).

Bias at Low Latitudes

The bias, arising during the observation period from March 2002 to February 2004 along the prime meridian in low latitude regions, is depicted in Figure 3.2 (top).

The CHAMP RO measurements are systematically warmer than ECMWF analyses at the tropopause level between 15 km and 18 km in low latitude regions. The positive bias is visible in all months and constitutes up to +2 K. In similar studies the same result was found by Wickert (2004), Steiner et al., and Gorbunov and Kornblueh (2003). Gorbunov

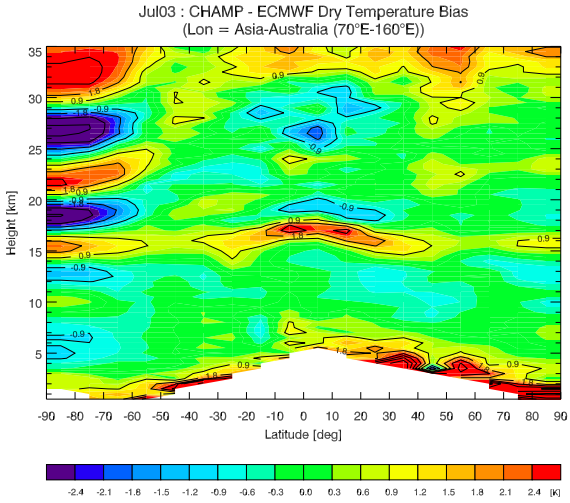


Figure 3.1: Bias of CHAMP RO climatology and ECMWF analysis in July 2003. An oscillating structure can be noticed at high southern latitudes, a positive bias larger than +0.9 K can be found in the tropopause between approximately 40°S and 40°N, a negative deviation can be recognized, in the low latitude region (20°S to 20°N) between 25 km and 30 km height, and a positive bias occurs above 31 km height at all latitudes.

and Kornblueh (2003) attribute the bias to the lower vertical resolution compared to RO measurements (RO resolution amounts to about 1 km at that altitude).

But when analyzing single difference profiles, a more complicated situation was found by Gobiet et al. (2005a). They noticed that ECMWF profiles cannot be thought to be smoothed versions of CHAMP radio occultation profiles.

A negative bias can be observed at tropical latitudes (10°S to 10°N) between a height of 25 km and 30 km. As can be seen in Figure 3.2 (top) it can be realized during the whole observation period with variable intensity and changing extension, decreasing from 2002 to 2004. It averages −1 K; maximum values can be observed in August 2002 (over the Pacific region, 190°E to 110°W) when it constitutes more than −2 K. The cause for that deviation is unidentified and needs further investigation.

Bias at High Latitudes

In the southern hemisphere winter, an interesting feature emerges at high southern latitudes. The bias gets a wavelike characteristic, oscillating between positive values up to +3 K and negative values down to −2.7 K. Starting at about 11 km height this anomaly ranges up to 35 km (Figure 3.1). It is clearly observable from May 2002 to August 2002 and from March 2003 to September 2003 (Figure 3.2, bottom). Maximum values can be noticed in July 2002 and in July and August 2003, whereas in 2002 the wavelike bias is less pronounced than in 2003.

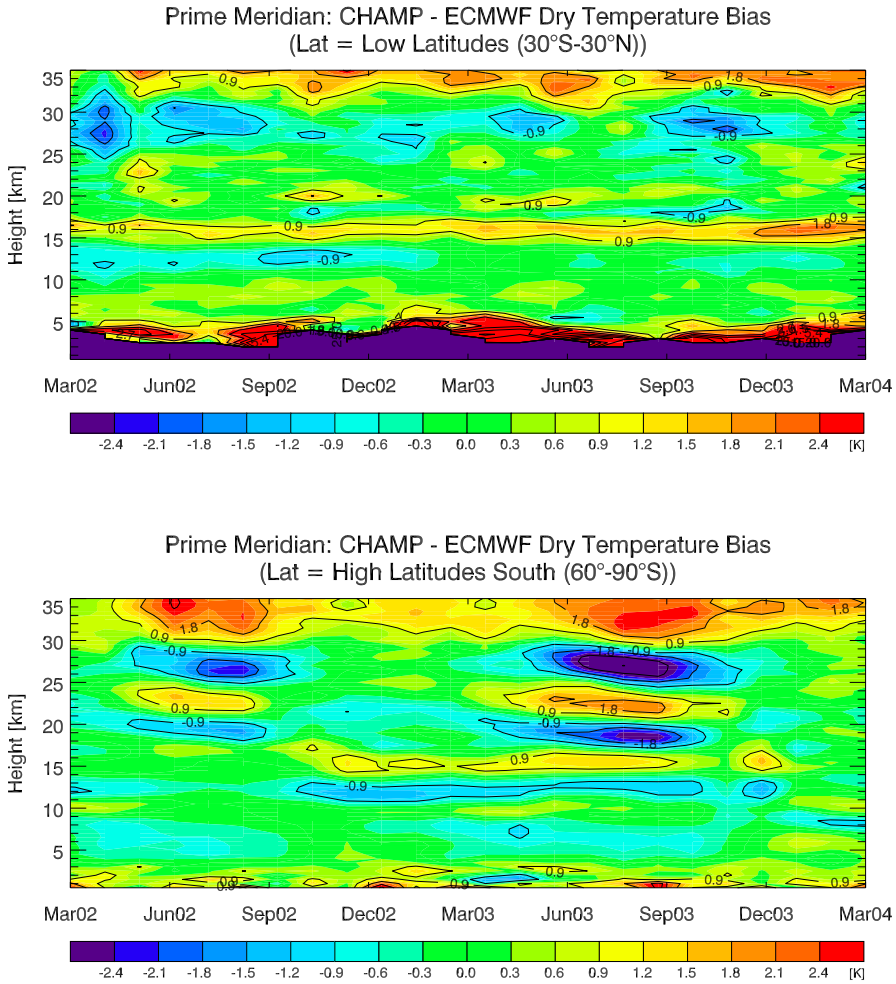


Figure 3.2: Bias along the prime meridian at low latitudes (top) and at high southern latitudes (bottom). Top: The positive bias between 15 km and 18 km height can be found during the whole observation period; the negative bias between 25 km and 30 km altitude varies in intensity during the months. A positive bias occurs at about 31 km height. Bottom: An oscillating structure arises from May to August 2002 and from March to September 2003. It is more pronounced in 2003.

The bias indicates that the ECMWF analysis does not exactly represent the polar vortex in this region. The large magnitude of the bias, which is larger than errors resulting from radio occultation measurements, and that there are no resolution-induced effects nor sampling errors argue for the CHAMP RO data (Gobiet et al. 2005a).

A similar structure cannot be observed in the high latitudinal northern hemisphere, where the bias rarely exceeds ±0.6 K.

3.1.2 Sampling Error

Due to discrete spatial and temporal sampling through occultation events, the sampling error affects climatologies arising from radio occultation measurements.

Each month more than 4 000 events are registered, and, due to the high inclination of the satellite, the majority is recorded in mid latitude regions, less in polar regions. As we will see, the sampling error will be higher at high latitudes than at low latitudes.

Foelsche et al. (2003) analyzed the sampling error depending on the number of occultation events. When enlarging the number of events, they determined the sampling error being reduced, but, in consequence of the same spatial and temporal distribution of all additional events, the error reduction was minimal.

Since we know about the inhomogeneous distribution of the occultation measurements relating to the local time (cf., Section 1.1.4), the impact of the temporal sampling will be considered later in this section.

Comparison Between Low, Mid, and High Latitudes

In general, the sampling error is lower than ±0.6 K, but considerably larger differences can be noticed at temperate and polar latitudes. There is no contradiction to the latitudinal distribution of the occultation events because of smaller temperature variations in low latitude regions compared to polar regions.

Low Latitudes: The low sampling rate at low latitude regions (30°S to 30°N) and the small variability of temperature yield a sampling error mostly lower than ±0.3 K during the whole observation period above 7 km altitude, with only some small fluctuations up to ±0.9 K.

Mid Latitudes: The sampling error observed at southern and northern mid latitudes (30°S to 60°S and 30°N to 60°N) does not show a uniform pattern. Large deviations can be noticed in very different ways.

In the mid latitudinal southern hemisphere, a large negative sampling error (up to −5 K) at an altitude of 10 km attracts attention, from 12 km to about 23 km altitude, a positive anomaly arises, both features can be observed during almost all months. In March 2003, when the positive sampling error is most prominent, it reaches +3.5 K; otherwise it is about +1 K to +2 K, but it cannot be noticed in August, September, and October 2003. Between a height of 25 km and 35 km, negative deviations can be found in April and May 2002 and from April to July 2003. During the southern summer

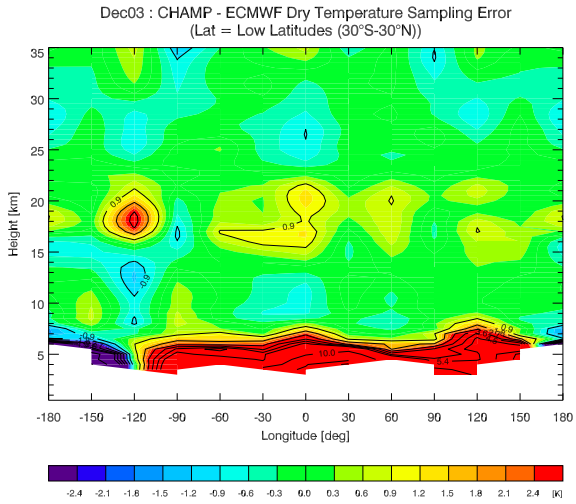


Figure 3.3: Sampling error at low latitudes in December 2003. Generally, the sampling error is very small, but some irregular positive and negative deviations can be found.

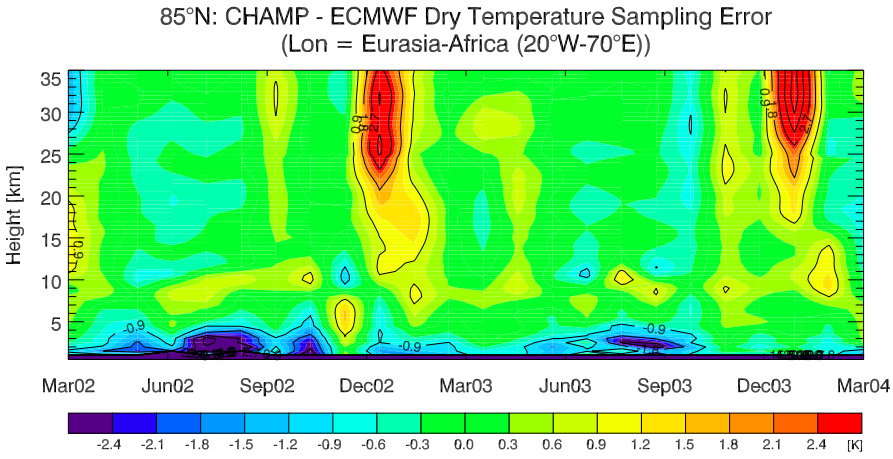


Figure 3.4: Sampling error between the CHAMP RO climatology and the ECMWF analysis from March 2002 to February 2004 at high northern latitudes (80°N to 90°N), Eurasian-African sector. In general, the deviation is very small, but in December 2002 and 2003 a large positive deviation can be found.

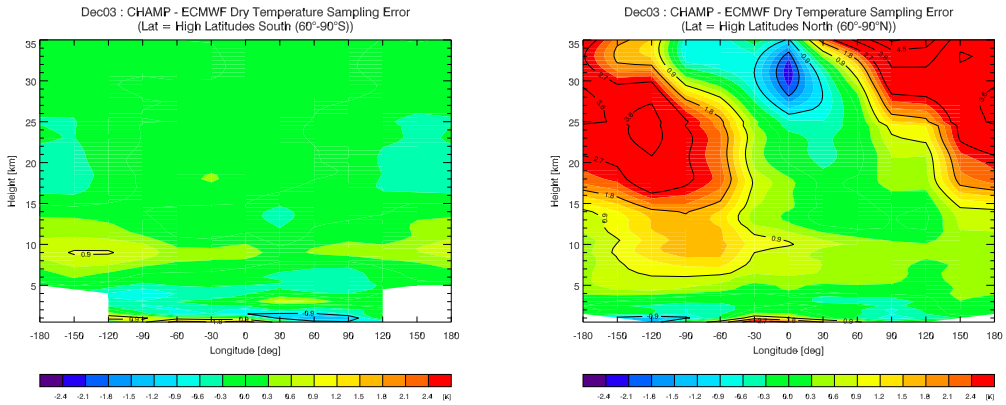


Figure 3.5: Sampling error at high latitudes in December 2003. Left: Southern hemisphere. No sampling error occurs in this region. Right: Northern hemisphere. During the polar winter the polar vortex appears in that region and has a positive impact to the sampling error. It amounts up to +4 K and is symmetric to the prime meridian. There, a small negative sampling error can be found.

months, December to February 2002 and 2003, it is less than ±0.6 K. Throughout the other months, no particular pattern can be recognized.

In the same line, the northern mid latitudes can be described but in a more regular way. The negative deviation, which can be found at an altitude of 10 km, and the positive deviations arising between 12 km and 23 km (most prominent in June 2002 and April 2003) are more pronounced compared to the mid latitudinal southern hemisphere. Above a height of 23 km no characteristic features are observable, mostly no sampling error arises at that height, only in some cases it is negative (November 2002) and sometimes it is positive (December 2002).

High Latitudes: Figure 3.4 depicts the sampling error arising between March 2002 and February 2004 at high northern latitudes (80°N to 90°N) in the Eurasian-African sector. Generally, the deviation is within ±0.3 K, but in northern winter months (especially in December), a high positive sampling error occurs above a height of 10 km in December 2002 and above 17 km in December 2003.

Figure 3.5 shows the comparison at high northern and high southern latitudes recognized at the same time (December 2003). During southern summer time, almost no sampling error can be found at high southern latitudes, whereas in northern polar regions (where it is winter), a positive deviation can be observed.

The opposite can be noticed in southern winter time when no deviations arise at high northern latitudes, but higher positive differences (considerably smaller compared to the northern hemisphere) in southern polar regions.

The observed sampling errors arising at high southern latitudes constitute less than ±0.3 K in December, January, and February 2002 and 2003, while a small positive

deviation (+1 K) can be found in springtime 2002 and 2003 (exception: March 2002 when a small negative deviation occurs), and a stronger positive sampling error can be realized from June to September with maximum values in August 2002 and September 2003 (about +4 K). The pattern arising in August 2002 mirrors to the sampling error arising in the high northern latitude regions in December 2002 and 2003.

Analyzing the sampling error at high northern latitudes, deviations, generally less than ±0.3 K, can be noticed during June, July, and August 2002 and June, July, August, and September 2003. Between a height of 10 km and 25 km, some stronger negative deviations can be found. From September 2002 to December 2002 and from October 2003 to December 2003, an interesting feature develops, which can be seen in Figure 3.5, right. The sampling error appears in an asymmetric shape with maximal values (approximately +4 K) between 60°E and 180°E and between 60°W and 180°W and minimal deviations around the prime meridian. In January 2003 the break down of the feature can be noticed, while in January 2004 it cannot be observed. During springtime, no characteristic trait can be found.

Analyse of Local Time Issues

Considering the dependence of the occultation events on the local time as presented in Section 1.1.4, it might be interesting to select the deviation arising from this fact.

To perform this examination, co-located ECMWF profiles were selected at randomized time. Figure 3.6 shows this "new" uniform distribution of the radio occultation events in time.

New artificial sampling error estimates were made and compared to actually measured ones. Figure 3.7 depicts differences arising from the sampling error measured in December 2003 at high northern latitudes (see Figure 3.5, right) and the sampling error, resulting from co-located ECMWF profiles, at randomized time. While the actually measured sampling error in December 2003 was relatively large, the difference plot between the actually measured and the "artificial" sampling error does not show large deviations. Investigating all the other months from March 2002 to February 2004 (the time series in the Eurasian-African sector at high northern latitudes is depicted as an example in Figure 3.8), it can be noticed that the deviations generally remain smaller than ±0.3 K.

The conclusion that the discrete sampling of the CHAMP satellite in time and the resulting bimodal distribution (cf., Figure 2.4) in the local time of the radio occultation measurements has no essentially influence on the sampling error can be drawn.

3.1.3 Total Error

The total error is a result of the observational error and the sampling error. Because of the relationship between the observational error and the bias all features recognized in the total error analysis can also be found in the above mentioned deviations. Depending on the season (month) and the considered latitudinal range, it is possible to determine the provenance of the observed deviation.

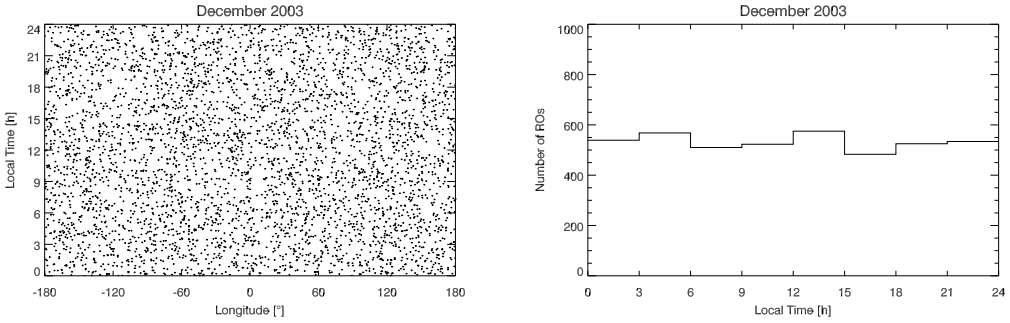


Figure 3.6: Distribution of RO events at uniformly randomized time in December 2003. Left: Randomized local time of radio occultation events as a function of longitude. Right: Histogram of the number of radio occultation events with three hour time-steps. The events are uniformly distributed in time.

The main characteristics of the bias – the positive anomaly in the tropopause between 30°S and 30°N on average, the negative deviation between a height of 25 km and 30 km in an narrower band at the equator and the wavelike structure at high southern latitudes, which is outstanding in southern latitude winter – are never covered from the sampling error; they can always be recognized in the total error.

The same is valid for the sampling error. When large sampling errors can be observed (e.g., in high northern latitude winter), they are reflected in the total error.

When both deviations arise at the same time in a large magnitude (e.g., in high southern latitude winter, a large bias and a large sampling error can be observed), it is not possible to distinguish between these errors without examining them separately. That is very important to know, because while examining the total error from CHAMP RO measurements and NCEP/NCAR reanalyses, NRLMSISE-00-, and CIRA86aQ_UoG climatologies, it is not possible to determine the derivation of the error. The measured deviation always arises from all incorporated sources of error.

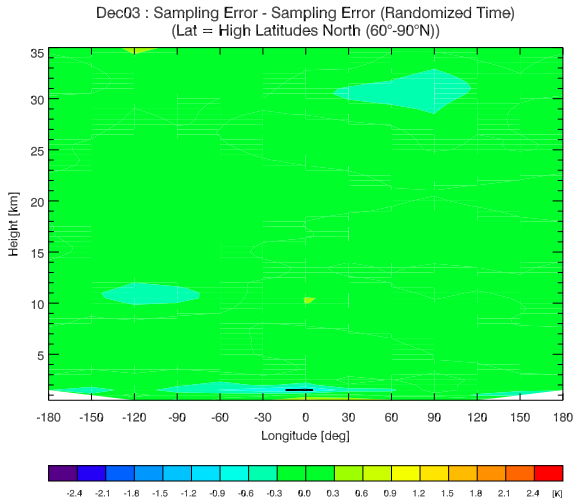


Figure 3.7: Difference between the actually measured sampling error and the sampling error calculated at randomized time at high northern latitudes in December 2003. Since the local time has no bearing on the sampling error, no deviation can be observed.

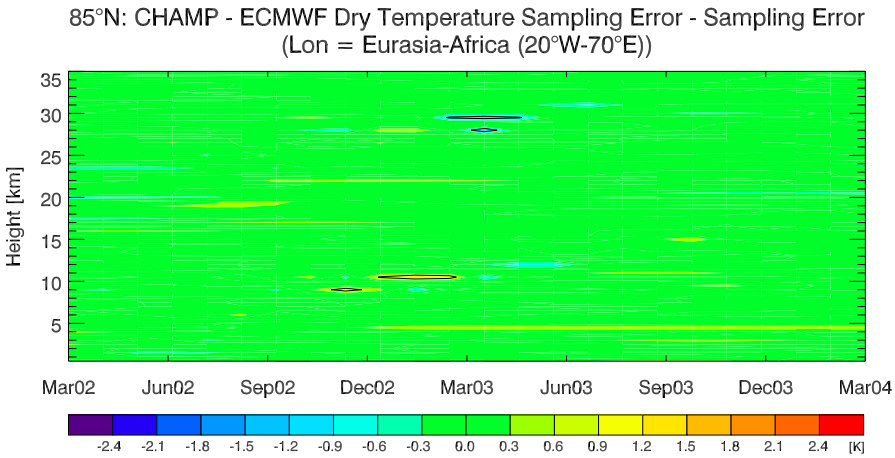


Figure 3.8: Time series (from March 2002 to February 2004, at high northern latitudes (80°N to 90°N), Eurasian-African sector) of the difference between the actually measured sampling error and the sampling error calculated at randomized time.

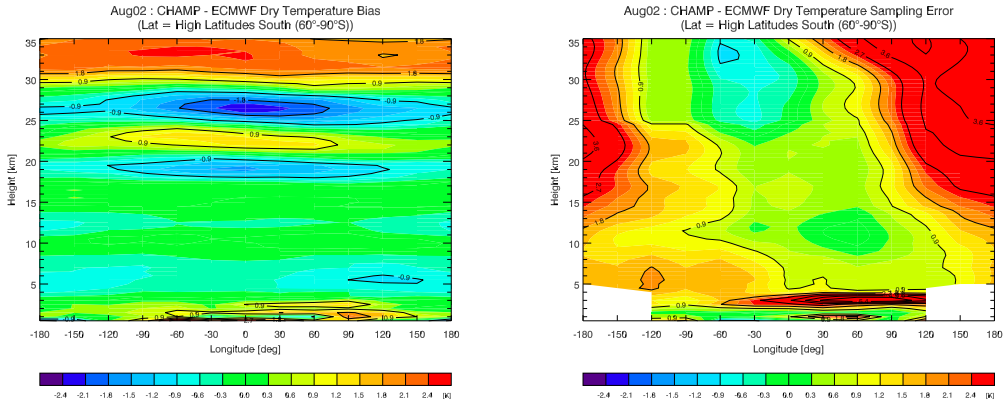


Figure 3.9: High southern latitude bias (left) and sampling error (right) in August 2002. The bias exhibits the well-known vertical structure; the sampling error shows a large positive deviation nearly everywhere.

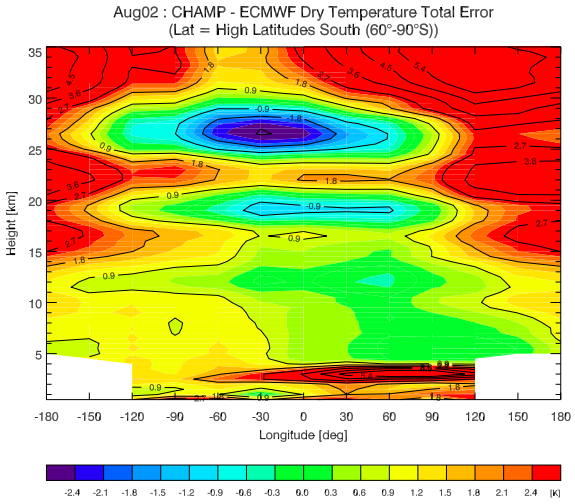


Figure 3.10: The composition of the bias and the sampling error is observable in the total error. The wavelike structure results from the bias; the positive deviation arising from 180°W to 60°W and from 90°E to 180°E at all altitudes can be attributed to the sampling error.

3.2 CHAMP Radio Occultation Data and NCEP Reanalysis Data

(Author: B.C. Lackner)

Whilst co-located profiles of ECMWF data are available for comparison with the results of the CHAMP retrieval, only climatologies of monthly means of the NCEP/NCAR reanalyses were used. Thus neither bias nor sampling error was computed to compare these data with RO data. The total error was calculated by subtracting NCEP/NCAR reanalysis data from RO data after averaging the NCEP/NCAR climatologies according to used binning resolutions.
Latitude × height, latitude × longitude, and longitude × height plots as well as time series and movies of the temporal variations served to visualize and helped in analyzing the results.

3.2.1 General Remarks

The differences between the retrieved CHAMP climatologies and NCEP/NCAR reanalyses are, as expected, larger compared to ECMWF data, which are included as a priori information in the retrieval process. Some structures remain all along the monthly timescale, such as lower tropospherical CHAMP RO temperatures.

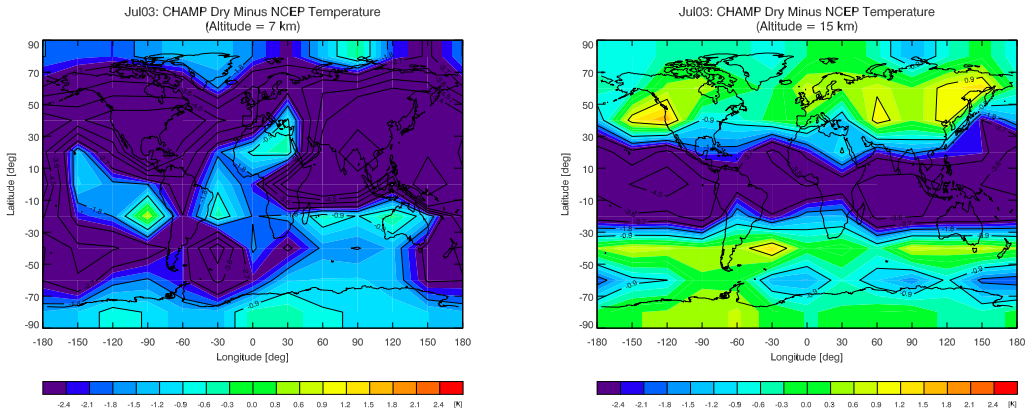


Figure 3.11: Left: Big temperature differences between CHAMP RO data and NCEP/NCAR reanalysis data can be found in the troposphere at an altitude of 7 km (e.g., July 2003). The dark blue areas indicate differences up to more than −5 K. Right: Outside the tropics and subtropics (north and south of 30°) and above the upper boundary of the tropopause (15 km height level is depicted), CHAMP and NCEP/NCAR-data show quite good agreement (green color). Graphs of other months are in general quite similar to those of July 2003.

Beside the larger differences in the troposphere, which are also results of comparing “dry” CHAMP RO temperatures with “real” temperatures, including troposphere

humidity, mentionable varieties mainly occur at higher latitudes during polar winter seasons. Details will be discussed in the next section. The best temperature agreement between the two "models" is thus given at mid latitudes at altitudes above the tropopause (see Figure 3.11).

One reason for the differences between CHAMP RO and NCEP/NCAR temperature climatologies can be found in the data used for (re-)analyses and their characteristics. Radiosondes (respectively rawinsondes in NCEP/NCAR's case) observations are not uniformly distributed. There are much more data available over quite densely populated northern hemisphere areas than over oceans, which also cover of course the largest part of the southern hemisphere (Lindzen 1990). Even though the vertical resolution of these data is fairly high, limitations are given due to required inter- and extrapolation. Additional data from satellites, such as TOVS, do not lead to "exact" values, as they only provide a low vertical resolution of the order of several kilometers (Schoellhammer et al. 2003) and therefore are not able to produce detailed information, e.g., under strongly perturbed conditions, which can be mainly seen at high northern latitudes during winter time.

3.2.2 Seasonal Considerations

Spring (March, April May) [1] March patterns are mainly influenced from the just ending northern polar winter, whereas in southern polar areas the differences between the two climatologies are not considerably distinct.

North of about 60° the differences are mainly negative (even up to −10 K in the Pacific region) and extend from the upper boundary of the troposphere to an altitude of 30 km. An exception is the Eurasian-African sector where positive deviations appear, even though the patterns are quite similar in their structure (see Figure 3.12).

The temperature differences in April look much like those in March, but the intensity of the contrasts wanes.

May is marked by little differences, the four sectors show similar deviations. Strongest discrepancies are formulated at higher southern latitudes.

Summer (June, July, August) June, the beginning of the northern summer, resembles May; large differences do appear. In general, the varieties between CHAMP RO and NCEP/NCAR data migrate from the northern hemisphere to the southern, where the polar winter is announcing itself. During the three northern summer months (June, July, August), there is quite a good agreement between the two "models" above the troposphere in the northern hemisphere. Main differences of less than −2 K are only found at high altitudes (around 25 km to 30 km) north of 60° latitude, whereas at mid latitudes the varieties range around ±1 K.

Bigger differences occur in the southern hemisphere outside the tropics and subtropics above a height of 20 km with deviations of more than −3 K. At the high southern

[1] As long as it is not specifically stated, northern hemisphere seasons are referred to.

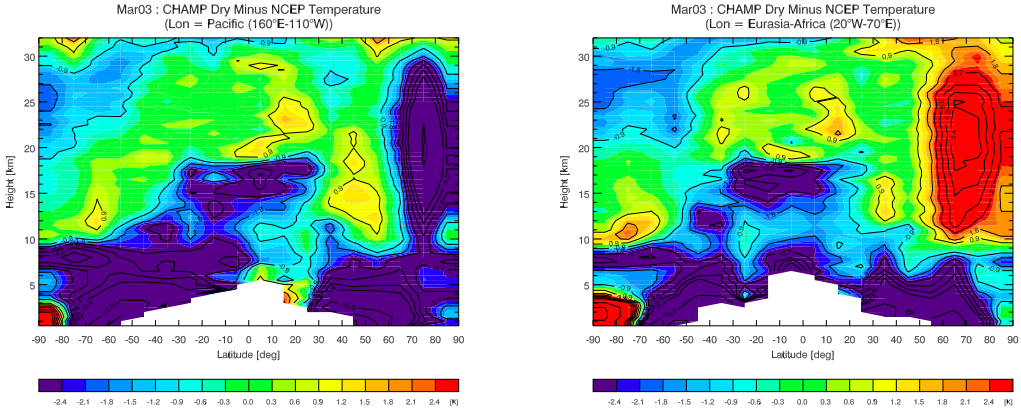


Figure 3.12: Differences between two selected sectors in March 2003. Left: Pacific region. North of 60° huge negative differences between CHAMP and NCEP/NCAR (more than −10 K) appear. Left: Eurasian-African region: The same area is signed by positive deviations of more than +5 K.

latitudes (higher than 60°), a wavelike structure evolves with an amplitude of more than −3 K at a height of 25 km and about +3 K at 32 km (see lower left graph in Figure 3.13). In contrast to high polar altitudes, remarkably good agreement between RO and NCEP/NCAR-data is found at lower height levels between 5 km and 20 km (less than ±1 K).

While in June and July there are nearly no differences in the structures in the four different sectors (Pacific, America-Atlantic, Eurasia-Africa, and Asia-Australia), in August the American-Atlantic and Asian-Australian regions show again deviations in opposite directions, which increase during the following months.

Autumn (September, October, November) As the sun drifts again to the southern hemisphere during autumn, a rather symmetrical distribution of temperature differences of the models is found in October (see Figure 3.14), even though there are differences in the direction of the temperature variations according to the four sectors considered.

While September is still influenced by the typical structures of the southern winter, which are even more pronounced in some regions than it was during summer, the big differences migrate again toward higher latitudes in the northern hemisphere during October and November.

Furthermore, during autumn the differences between the four sectors are noteworthy. Whereas CHAMP RO temperatures are up to +10 K higher than NCEP/NCAR values in the American-Atlantic sector south of around 50°S (above 12 km) in September and October, the NCEP/NCAR temperatures surmount the CHAMP RO values by more than +5 K in the same region in the Asian-Australian sector. During November, these differences fade again.

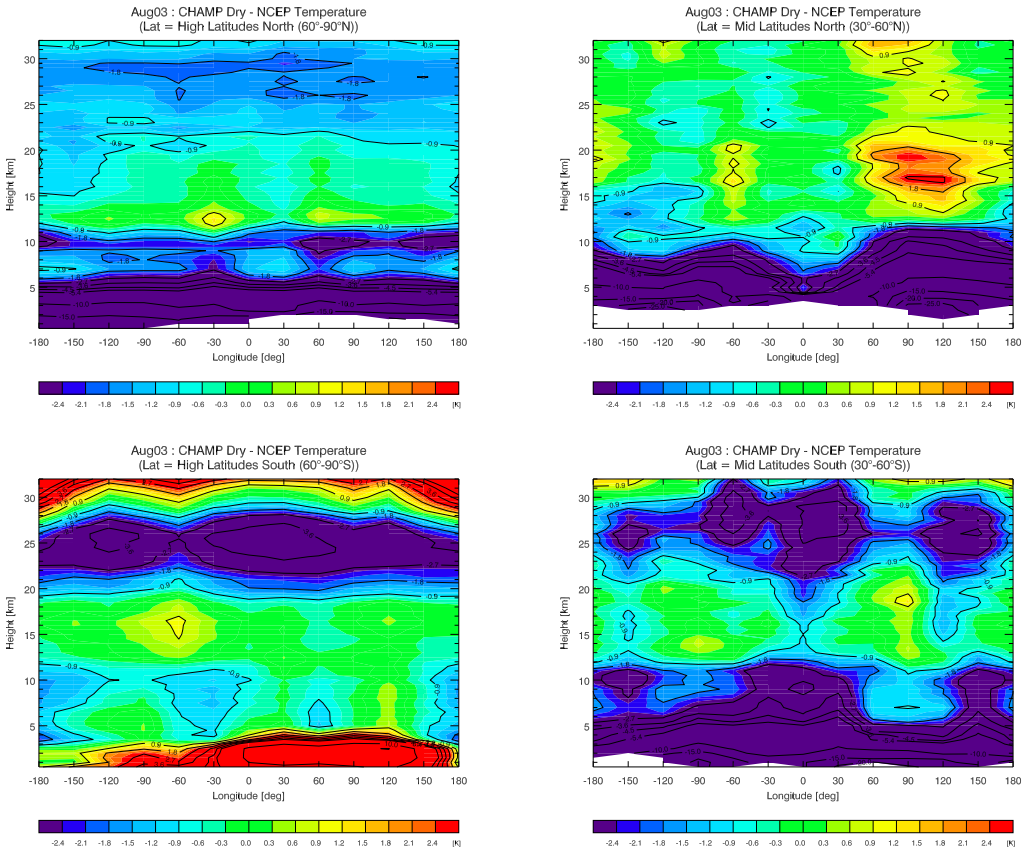


Figure 3.13: Top: Temperature differences between CHAMP RO and NCEP/NCAR data at high (left graph) and mid northern latitudes (right graph). In summer, the stratospherical differences between the two models are weakly developed in the northern hemisphere, focusing on altitudes above 20 km at the high latitudes. The spatial limited positive deviation in the right graph is geographically situated above eastern Asia. Bottom: In the southern hemisphere the temperatures match quite well below 20 km, but the differences increase a lot at higher altitudes.

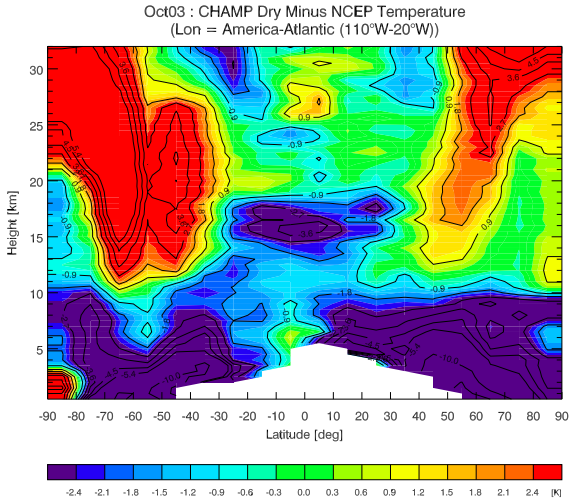


Figure 3.14: American-Atlantic sector in October 2003: CHAMP RO and NCEP/NCAR temperature differences are arranged quite symmetrically according to the equator.

Winter (December, January, February) The polar northern winter leads to larger temperature deviations between the two models in the northern hemisphere, as it was the case during the southern winter in the southern hemisphere. The structures resemble those shown in Figure 3.12 (March 2003) with the differences between the sectors emerging again. This phenomenon is illustrated in Figure 3.15 as well.

In contrast to the southern polar winter, in which CHAMP RO and NCEP/NCAR data show good agreement at altitudes between 5 km and 20 km (cf., Figure 3.13), this is not the case in the northern polar winter. The structures at higher altitudes show more changes in the northern hemisphere, as if the change of land and ocean surface also influence high levels.

3.3 CHAMP Radio Occultation Data and MSIS Data

(Author: B.C. Lackner)

As described in Section 1.4.1, NRLMSISE-00 is an empirical model based on zonal average temperature and pressure tabulation below 72.5 km and additional data from the U.S. National Meteorological Center below 20 km. In contrast to analysis (ECMWF) or reanalysis (NCEP/NCAR) data, which are based on a variety of real-time measurements, NRLMSISE-00 climatologies do not alter for different years, as they are the result of averages over a certain period (depending on the data sources used to implement the model, mainly originating from 1961 to 1997 in this case). Furthermore, the model aims to be used for studies reaching across several atmospheric boundaries (up to thermosphere), and not for troposphere or a part of stratosphere as used in this context.

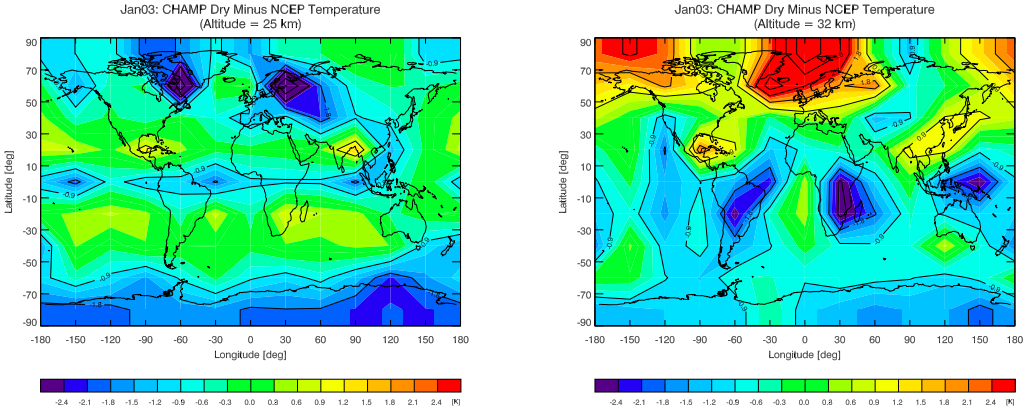


Figure 3.15: Maps of CHAMP RO and NCEP/NCAR reanalysis temperature differences at two different heights for January 2003. Left: 25 km altitude. Whereas in the northern hemisphere the deviations seem to depend on sectors of some longitudes, they are more or less uniformly distributed in the southern hemisphere. Right: 32 km altitude. A similar picture as at 25 km. In the northern polar areas, the differences between the models tend to show in opposite directions for different longitude sectors (ranging from +4 K to −1 K).

3.3.1 General Remarks

Due to above mentioned reasons, NRLMSISE-00 climatologies differ much more from CHAMP RO climatologies than from ECMWF or NCEP/NCAR. Deviations of much more than ±2 K predominate most areas. However, the most critical points are again found in the tropical tropopause and the polar winter regions. Unlike NCEP/NCAR climatologies, nearly no differences appear according to the four different sectors of longitude that were considered.

Below about 20 km, a wavelike structure is taking shape at tropical latitudes with positive deviations (indicating higher CHAMP RO than NRLMSISE-00 temperatures) between about 10 km and 15 km and negative differences above and below this level. Some seasons show a further positive deviation in higher levels (see Figure 3.16).

The advantage of a model based on average temperatures – such as NRLMSISE-00 – is that the independence of climatologies of the years helps to identify years, seasons or months with stronger abnormalities than average, which can also be seen in some cases in our examination period.

Varying deviations of CHAMP values from the longtime mean temperatures (NRLMSISE-00) during southern winter seasons are evident regarding Figure 3.17. Lower stratosphere (between 12 km to 25 km) temperatures were much higher than average over the Antarctic region in southern winter and spring 2002. In September and October 2002, the south polar anomalies were larger than any in the last 20 years (Angell et al. 2002). In contrast to the warm southern winter/spring in 2002, in 2003 (August to

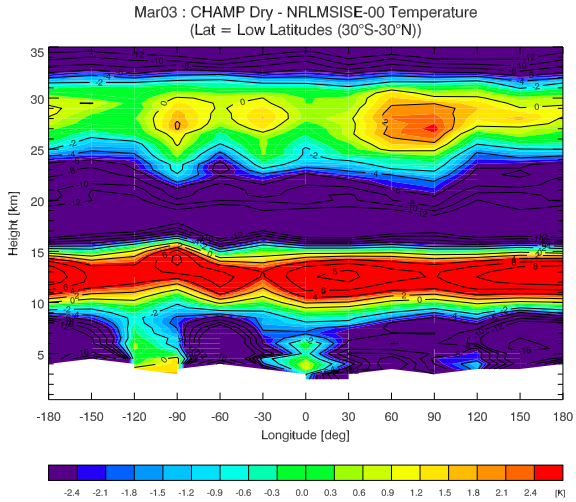


Figure 3.16: March 2003 differences between CHAMP RO and NRLMSISE-00 climatologies. Wavelike patterns occur in tropical stratosphere.

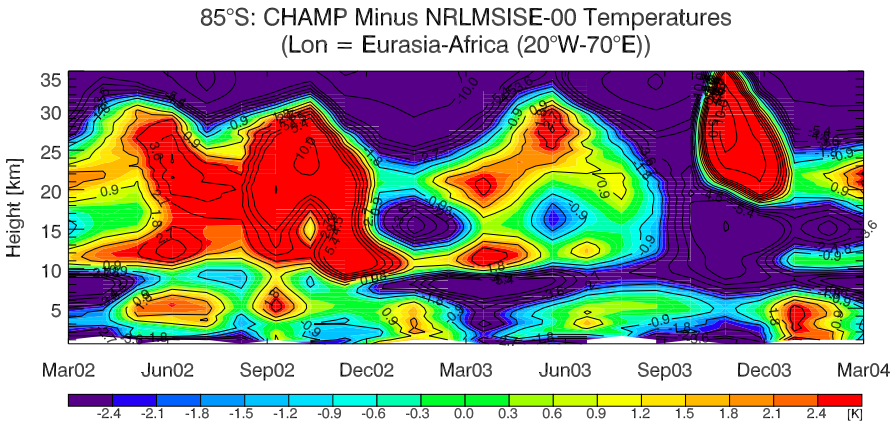


Figure 3.17: Differences between CHAMP and NRLMSISE-00 temperatures at high southern latitudes (Eurasian-Asian sector from 75°S to 90°S) from March 2002 to February 2004. While the retrieved values were much higher than average during August 2002 to October 2002, one year later, negative deviations appear during the same time in 2003.

October) for most of the southern polar region, minimum temperatures were below long term average minimum values, in fact near record low temperatures (Angell et al. 2003b), resulting in a strong polar vortex.

3.3.2 Seasonal Considerations

Spring (March, April May) Like NCEP/NCAR climatologies, differences in March to CHAMP RO data are first found at northern mid and high latitudes above 20 km altitude, decreasing during April and May (more than −10 K in March, about +4 K in April and May). April 2003 shows, over all longitudes, positive deviations at high altitudes (around 30 km to 35 km) over the northern polar region, which are contrary to the negative differences in March and May and seem to be an exceptional case for this year.

Best temperature agreement can be located at an altitude around 15 km to 30 km mainly in the southern hemisphere during March, migrating to the northern hemisphere by April and May. On the other hand, rather large deviations are found during the whole season at altitudes of 35 km (negative deviations of about −6 K over the tropics and the northern Atlantic, positive deviations of approximately +4 K and more over central Asia and parts of the southern ocean (see Figure 3.18).

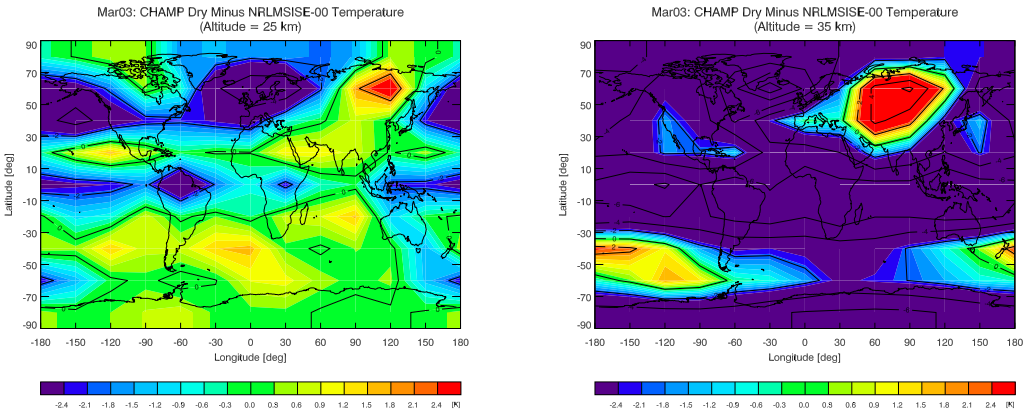


Figure 3.18: Left: During March 2003 the best temperature agreement between CHAMP RO and NRLMSISE-00 climatologies is located around 25 km in most parts of the southern hemisphere. Right: At altitudes of 35 km, huge deviations between the two models occur, which is evident from dark blue and red colors in this graph, standing for mean temperature differences of ±4 K.

Summer (June, July, August) During summer, the structures of the differences remain more or less the same in the northern hemisphere. A relatively narrow band of positive deviations is located near 13 km altitude at high northern latitudes and 15 km at mid latitudes, respectively, whereas at higher and lower altitudes negative deviations prevail (see upper two graphs in Figure 3.19).

Large differences between CHAMP RO and NRLMSISE-00 data are found in the southern hemisphere (see lower two graphs in Figure 3.19), whereat the high positive deviations (around 50°S) in 2002 are much more pronounced than in 2003 (in August 2003 nearly no positive deviations occur at high southern latitudes). At certain height levels at mid and high southern latitudes, overall differences (positive and negative) of more than 20 K occur, while negative deviations of more than −16 K appear above 30 km.

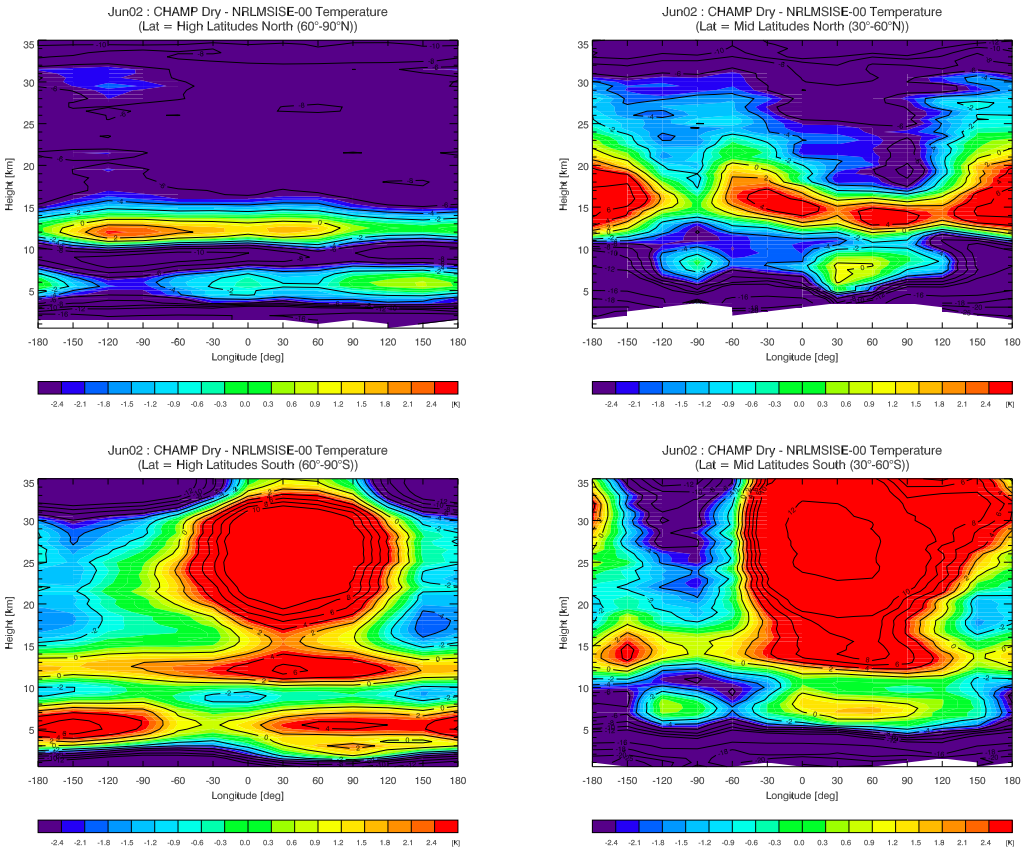


Figure 3.19: Top: Temperature differences between CHAMP RO and NRLMSISE-00 climatologies in high (left graph) and mid (right graph) latitudes in June 2003. The only positive deviations, standing for higher CHAMP RO temperatures, are found in a relatively narrow band around 13 km (high latitudes) and 15 km (mid latitudes). Bottom: Same picture for southern hemisphere, where huge (mainly) positive deviations characterize the southern winter season.

Autumn (September, October, November) While in September large differences between the two models mainly are focused on the southern hemisphere, during October and November they migrate again slowly to the northern hemisphere, where winter

is approaching. A striking structure evolves during this season at polar and subpolar southern latitudes around a height of 20 km to 30 km, where quite big positive differences appear until November. In December the structures seem to be "mirrored" by the equator, intensifying during the following months (see Figure 3.20).

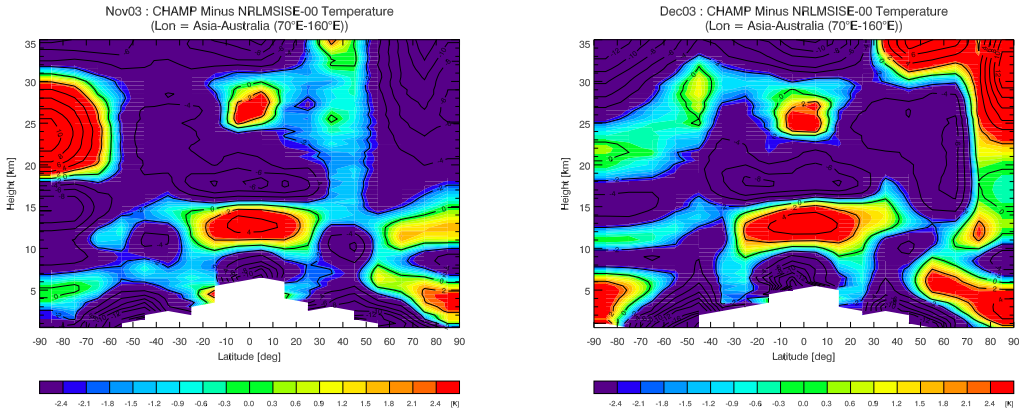


Figure 3.20: While in November, positive deviations of more than +8 K at heights of 20 km to 30 km occur in polar and subpolar southern regions (left graph), one month later, in December, the structure seems to be "mirrored" by the equator and is now situated over northern high latitudes with deviations of more than +10 K (right graph).

Winter (December, January, February) During these three months, the largest differences between CHAMP RO and NRLMSISE-00 data are located above the northern polar area. The structures shown in Figure 3.20 increase in January with CHAMP RO temperatures exceeding NRLMSISE-00 values by more than +20 K in some places. In general, mid and high northern latitudes are characterized by huge differences between the two models at all levels in January.

In the southern hemisphere, to a large extent, negative deviations occur, which top at high altitudes (around 30 km and higher) as well. In the course of February, the pronounced northern structures fade again but still remain stronger than the southern ones.

3.4 CHAMP Radio Occultation Data and CIRA86aQ_UoG Data

(Author: B. Pirscher)

While computing the total errors between the CHAMP RO and the CIRA86aQ_UoG climatologies from January to December 2003, some interesting features can be found.

In general, the CIRA86aQ_UoG climatology is significantly colder compared to the CHAMP RO climatology. Deviations up to −6 K dominate the structure arising from the comparison between both data sets (CHAMP minus CIRA86aQ_UoG), but partly emerging positive total errors show very large dimensions as well. In December 2003, for example, at high north latitudes above 10 km height, the deviation is larger than +20 K.

Longitudinal Behavior of the Total Error

Since CIRA86aQ_UoG is comprised a model of zonal mean temperature (as well as zonal mean geopotential height/pressure and zonal wind), the land-see distribution is disregarded. This effects the positive total error, which is less pronounced above the Pacific (160°W to 110°E) and the American-Atlantic region (110°W to 20°W) compared to regions above big landmasses (sectors above Eurasia-Africa (20°W to 70°E) and Asia-Australia (70°E to 160°E)), where the deviation remains higher. Figure 3.21 depicts this situation in January 2003. While the deviations in the southern hemisphere and at the northern low latitudes are quite similar in the Pacific and Eurasia-Africa sectors, above the maritime sector a smaller extension of the positive total error can be noticed.

A converse situation between the sectors arises in September 2003 (not shown), when a big positive deviation up to +10 K can be found between 40°S and 70°S from a height of 12 km to 30 km in the Eurasian-African sector, and a big negative deviation (−10 K) can be observed at the same latitudinal range from a height of 15 km to 30 km. The situation above Asia and Australia still refers (but less distinctive, about +4 K) to the situation realized above Eurasia-Africa. In the American-Atlantic sector, the deviations are rather negative, but they do not show that significant pattern realized in the Pacific region.

A similar situation occurs in February and March 2003 at mid northern latitudes between 40°N and 70°N, but the highest total error remains within the limits of ±5 K.

Latitudinal Behavior of the Total Error

Depending on the season and the considered latitudinal range, some characteristic features can be found.

Because of the constant solar radiation in low latitude regions, hardly any fluctuations arise during one year. The CHAMP RO climatology and the CIRA86aQ_UoG climatology represent this region in a similar way, so that the total error remains the same.

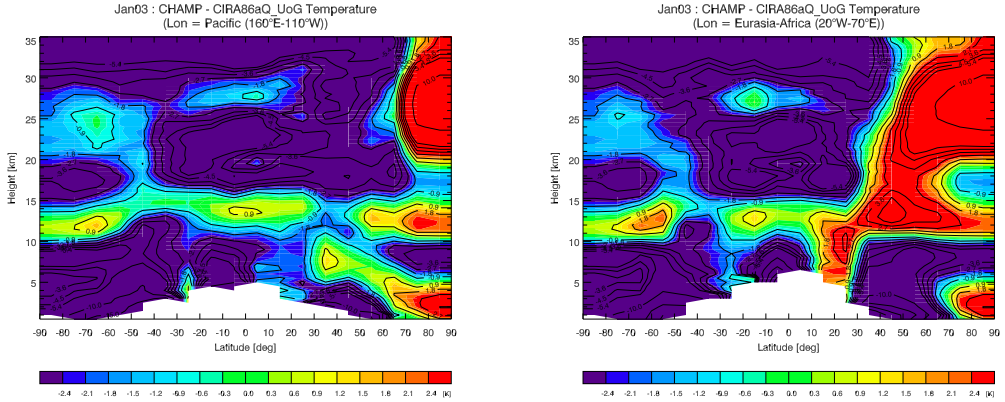


Figure 3.21: Comparison between one maritime sector (Pacific, left) and one sector situated above a big landmass (Eurasia-Africa, right). Whereas the total error is overall about the same in the southern hemisphere and at low latitudes in the northern hemisphere, the positive deviation is smaller in the Pacific region at mid and high latitudes compared to the Eurasian-African area.

At mid and high latitudes the solar radiation has a higher impact to the arising climatologies, and some seasonal features emerge in the total error examination.

Low Latitudes: Focusing on the low latitudes, a negative total error between -2 K and -6 K predominates the arising structures. Only small regions with minor deviations can be found.

During the whole year of 2003, the total error remains smaller than ± 2 K at the low latitudes from 10 km to 15 km, and from an altitude of 25 km to 30 km. From June to September 2003, another band with a "small" total error occurs between 17 km and 20 km height. In this connection, "small" means within the limits of ± 2 K.

Mid and High Latitudes: Because of the deviation being affected by seasonal circumstances, the mid and high latitudes will be discussed in that context. Figure 3.22 depicts four months (January, April, July, and October) representing the four seasons (winter, spring, summer, and fall).

During the northern winter months November to January (Figure 3.22, top left), a strong positive total error can be noticed at the northern latitudes northwards of 40°N. Starting at a height of 10 km, it reaches more than +10 K from 24 km to 30 km north of 75°N. At southern latitudes small deviations with different signs can be found between a height of 10 km and 15 km, and from 19 km to 28 km a small negative deviation occurs at the same time.

Due to the weakening of the formation in February and March, a vertical structure results in April (Figure 3.22, top right), when positive deviations can be observed at high northern and high southern latitudes from a height of 10 km to 15 km and from about 25 km, in between a negative deviation up to -3.6 K arises. At the mid and

low latitudes, some regions characterized by small total errors can be realized between 10 km and 15 km and between approximately 25 km and 30 km. Larger dimensions of this deviation are prevented because of a large region (40°S to 40°N between 15 km and 25 km height) exhibiting a negative total error (about −5.4 K).

From April to July (Figure 3.22, bottom left) the positive total error at high northern latitudes (above 25 km) disappears. The deviation stays positive from 10 km to 15 km and it becomes negative from 19 km to 27 km. The structure is similar to that arising in January 2003 at high southern latitudes. The counterpart to the feature arising in January at the northern latitudes does not occur in July, but emerges in September and is most prominent in October (Figure 3.22, bottom right), when in the northern hemisphere the positive total error grows.

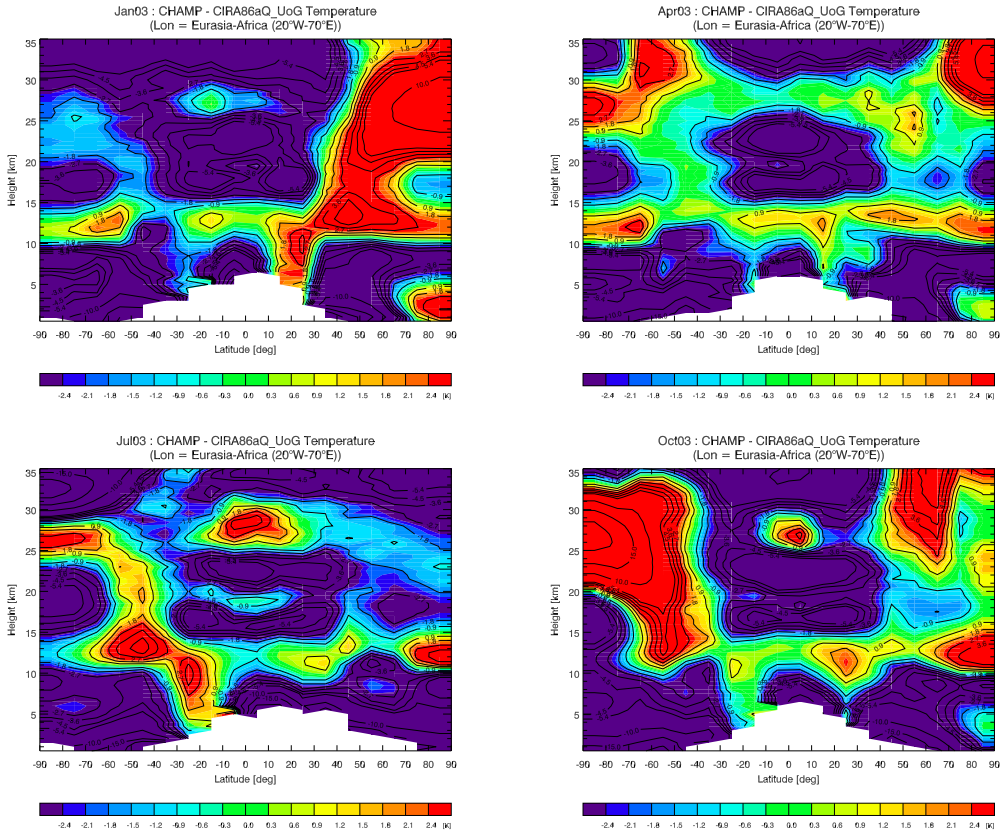


Figure 3.22: Seasonal circumstances arising in the total error between CHAMP RO climatologies and the CIRA86aQ_UoG model. In general, the total error is negative, but in January a big positive deviation can be found at high and mid northern latitudes above a heighz of 10 km. In April and in July a relatively symmetric pattern can be observed. In October, when the positive total error is building up at the northern high latitudes, a large positive deviation occurs at high southern latitudes.

3.5 Further Comparisons

3.5.1 EGOPS MSISE-90 Data in Comparison With NRLMSISE-00 Temperature Data

(Author: B.C. Lackner)

To examine differences between MSISE-90 data used in EGOPS and NRLMSISE-00 data, the latter were subtracted from the EGOPS temperatures at each grid point and height. The MSISE-90 model used in EGOPS is a parametrization of the MSISE-90 model using Chebyshef polynomials and spherical harmonics and is described by Høeg et al. (1998).

Because of the minor differences between the two "MSIS-models", the colors in the plots rang, different from previous graphs, from −1 K (dark blue) to +1 K (red).

While the data sets match quite well below about 20 km, the differences increase with higher altitudes (see Figure 3.23).

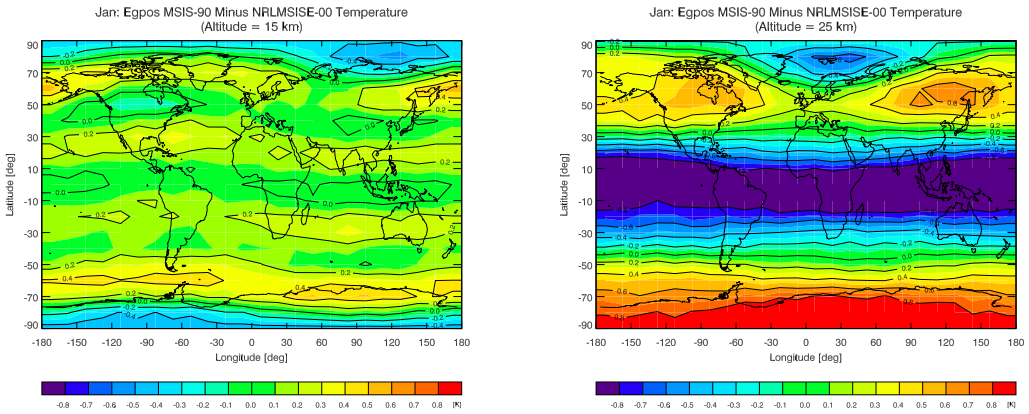


Figure 3.23: Maps of differences between EGOPS MSISE-90 and NRLMSISE-00 temperature values in January. Left: Below 20 km altitude (map shows 15 km height level) the differences are minor, ranging mainly between ±0.2 K (apart from polar and subpolar regions). Right: At higher altitudes (map shows 25 km height level) the deviations increase and rise to their highest values over tropical and polar regions.

Over tropical/subtropical regions, wavelike structures emerge with "fixed" wave crests at heights of about 21 km (positive deviations of +0.4 K to +0.6 K, standing for lower NRLMSISE-00 values), 26 km (negative deviations up to −0.8 K, mainly during northern autumn and winter) and 31 km (again positive deviations of more than +0.8 K during the whole year), that is to say every 5 km, which seems to be caused by the parametrization of the MSISE-90 model. The amplitude of this pattern increases with height. While at mid latitudes the differences between the models are pronounced to a lesser extent, the wavelike structures turn up again over polar regions (see Figure 3.24), but there they

change their position at altitude during the year.

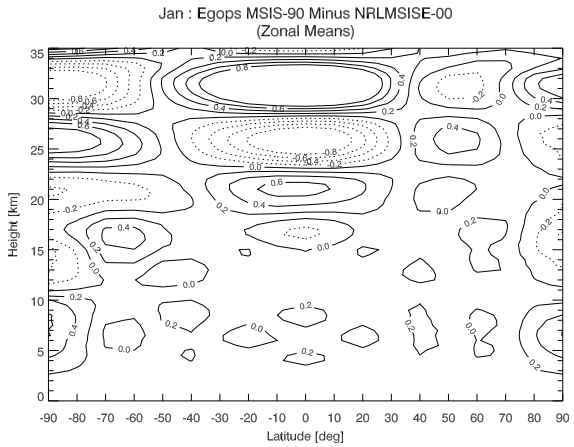


Figure 3.24: Differences between EGOPS MSISE-90 and NRLMSISE-00 in January. Below 20 km nearly no differences appear (apart from polar areas), whereas at higher altitudes wavelike structures emerge above all over tropical and subtropical regions.

The absolutely highest deviations between EGOPS MSISE-90 and NRLMSISE-00 temperatures are found at high altitudes. The maximal difference between the two models below 35 km appears at a height of 31.5 km ($\varphi = 0°$, $\lambda = 112.5°$).

3.5.2 CIRA86aQ_UoG Data in Comparison With NRLMSISE-00 Temperature Data

(Author: B. Pirscher)

CIRA86aQ_UoG and NRLMSISE-00 temperature data result from atmospheric investigations lasting many years. They differ in the general kind of the model as well as in incorporated data. Whereas CIRA86aQ_UoG is a zonal model, NRLMSISE-00 exhibits small latitudinal variations in temperature.

The total error analysis is performed for the months January to December (independent of the year) because the CIRA86aQ_UoG model does not exist for different years (the same applies to NRLMSISE-00 model as well).

The total error resulting from the comparison between CIRA86aQ_UoG and NRLMSISE-00 climatologies reaches its highest values at the northern high latitudes. The deviations (CIRA86aQ_UoG minus NRLMSISE-00) achieve -12 K in December and grow up to $+16$ K in January.

The deviation will be analyzed in longitudinal slices (Pacific, Eurasia-Africa, Asia-Australia, and America-Atlantic) and in latitudinal regions (low latitudes, mid latitudes, and high latitudes).

Longitudinal Behavior of the Total Error

Concerning the total error variations at longitude, some large fields of positive and negative deviations can be found, similar to the total errors found between CHAMP RO climatologies and the CIRA86aQ_UoG model. From January to April and from November to December, a high negative total error can be noticed in the Pacific sector and in the Asian-Australian region (less pronounced) above 11 km altitude from 40°N to approximately 85°N. At the same time and location (above an altitude of 15 km), a positive deviation arises over the Eurasian-African area and the American-Atlantic sector (less stable but with larger maximum values). Figure 3.25 depicts this situation in January between 160°E and 110°W (Pacific) and between 20°W to 70°E (Eurasia-Africa).

In September and October a high negative total error (−6 K) can be found at high and mid southern latitudes in the Pacific and Asian-Australian sectors from 50°S to the polar region, whereas the other two sectors show an oscillating structure with more positive proportion in September and a much smaller negative pattern in October.

Apart from those features, no oppositional behavior can be realized in the four sectors.

It seems that the total error deviations still result from the zoning performed in the CIRA86aQ_UoG model.

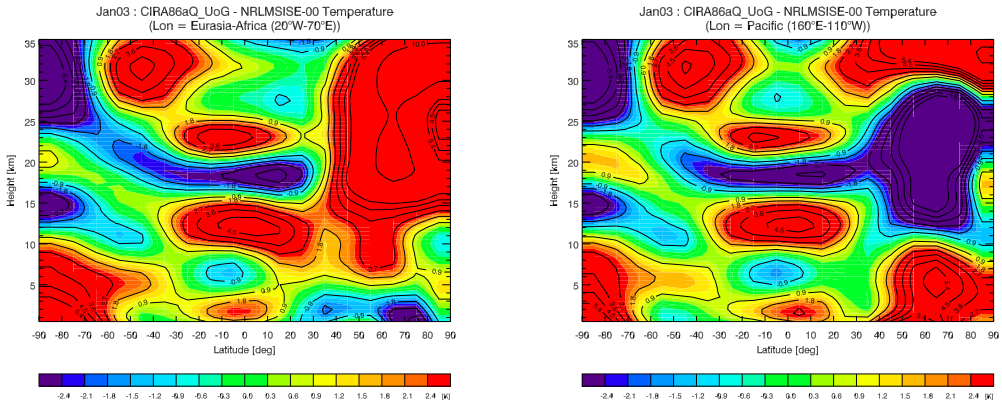


Figure 3.25: Total error arising in the Pacific and in the Eurasian-African sector. When above the ocean, a negative deviation can be noticed between 40°N and 85°N from a height of 11 km to 29 km above Eurasia and Africa, and a bigger positive deviation already arises around 6 km height.

Latitudinal Behavior of the Total Error

When large positive/negative deviations can be found at high northern or southern latitudes (60°N/S to 90°N/S), they also extend in mid latitude regions (30°N/S to 60°N/S), and only low latitudes remain unaffected. This behavior of total error follows from the general differences between both models.

A wavelike structure predominates in the total error arising at low latitudes during all months. At the earth's surface, up to about 5 km altitude, a positive deviation can be found at any time. This positive anomaly extends (+3 K) to a height of 15 km, from May to October, unless a small negative deviation breaks this band. Every five kilometer height the sign turns around, and the deviations amount between −2 K and +4 K.

This oscillating structure results from the vertical interpolation type (cubic spline interpolation) chosen to generate the CIRA86aQ_UoG climatologies.

A wavelike pattern can also be noticed at high and mid latitudes at the time when no significant differences between both climatologies occur (essentially between May and September at high northern latitudes and between January and August as well as in December at high southern latitudes). The zonal structure is less stable and less pronounced compared to the oscillating feature arising in low latitude regions.

4 Conclusions

(Authors: B.C. Lackner, B. Pirscher)

CHAMP RO data are characterized by being discrete in space and time. In our work, and in general as well, climatologies are used to analyze and map data sets. The received maps are not exact data; they include a substantial amount of interpolation and therefore should better be addressed as "analyzed data" (Lindzen 1990). The same applies to ECMWF and NCEP/NCAR (re)analyses as well as NRLMSISE-00 and CIRA86aQ_UoG data. Nevertheless, always having this fact in mind, large differences between the four investigated data sets can be determined. We generally termed these differences "errors" in this work, but it is important to note that this is meant purely in a relative sense ("error" of some data set relative to the selected reference data set).

All in all, ECMWF analyses data proved to be the best data set compared to CHAMP retrieved temperatures. Over most areas the deviations (total error) remained within ± 1 K between 7 km and 29 km altitude. Larger deviations at lower heights could not be interpreted because of the geometric optics approximation. The reasons for the warm bias above 29 km have to be looked at in detail in future investigations.

The warm bias (standing for higher CHAMP than ECMWF temperatures) at the low latitude tropopause, which is present during the whole observation period (cf., Figure 3.2, top), is worth mentioning. Compared with this, NCEP/NCAR reanalysis data tend to overestimate tropopause temperatures at low latitudes (see left graph of Figure 4.1). Randel et al. (2002) also came to this conclusion in a study focusing on comparing climatological data sets for the middle atmosphere, which are used in the research community.

In contrast, at high southern latitudes during winter time (May to September 2002 and May to October 2003), wavelike patterns form at altitudes above tropopause when comparing CHAMP RO and ECMWF data (cf., Figure 3.2, bottom). These features have been discussed in detail in Section 3.1.1. The oscillating structure cannot be found in error analyses between CHAMP RO data and NCEP/NCAR reanalysis (see Figure 4.1, right).

While NCEP/NCAR and CHAMP temperature profiles go more or less together below heights of 20 km at high southern latitudes during July, the wavelike vertical appearance of ECMWF profiles are clearly evident in the right graph of Figure 4.1.

Basically, deviations between CHAMP and NCEP/NCAR temperatures are slightly bigger compared to ECMWF deviations. Analyses (ECMWF) and reanalyses (NCEP/-NCAR) climatologies are mainly based on radiosondes and satellite data at higher al-

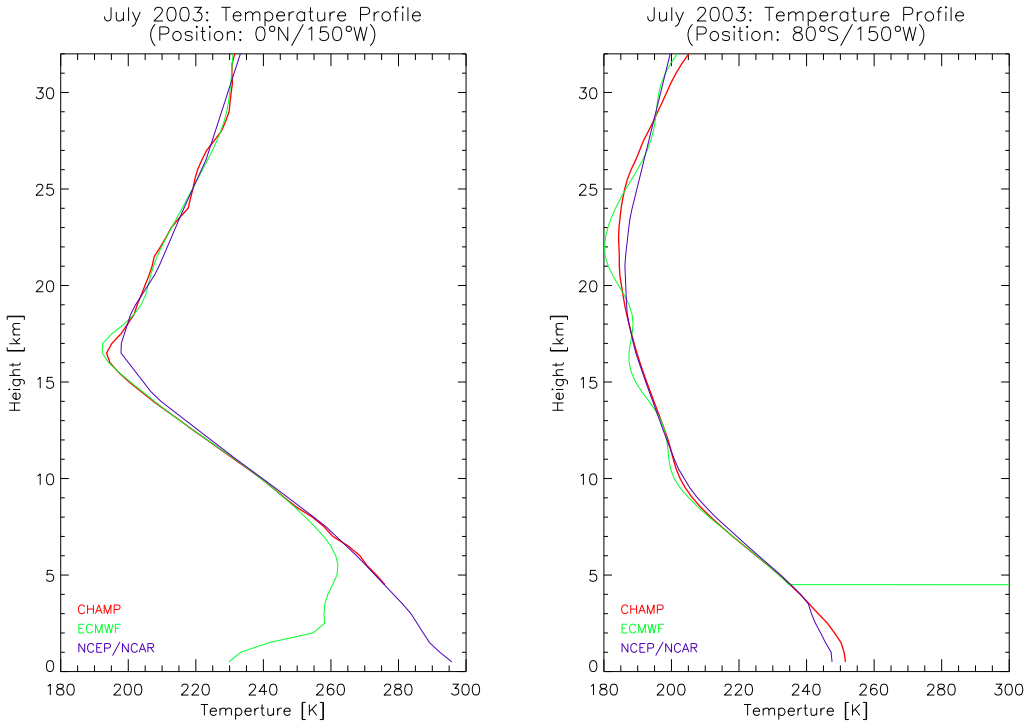


Figure 4.1: Profiles of CHAMP RO (red), ECMWF (green), and NCEP/NCAR (dark blue) temperatures in July 2003 at the equator and at 80°S and 150°W (pacific region). Left: At the equatorial tropopause, NCEP/NCAR temperatures are clearly higher compared to CHAMP RO and ECMWF, whereas the ECMWF temperature profile shows colder values (resulting in a warm bias CHAMP RO minus ECMWF) between 15 km and 18 km altitude. Right: Up to 20 km, CHAMP and NCEP/NCAR climatologies agree quite well at high southern latitudes, while an oscillating structure of the ECMWF profile is clearly visible. Above 20 km a cold total error can be noticed between CHAMP and NCEP/NCAR (derived from CHAMP minus NCEP/NCAR), while the ECMWF profile keeps its wavelike structure.

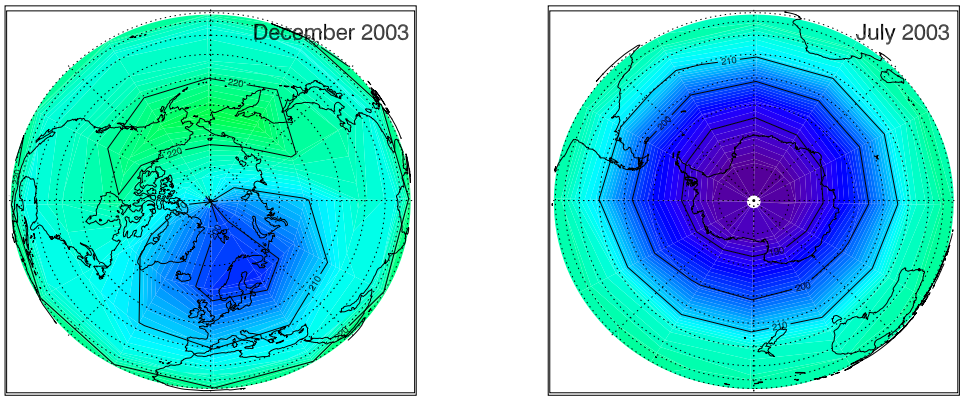


Figure 4.2: CHAMP RO temperatures around the north and south pole during northern (December 2003) and southern (July 2003) winter time at 25 km altitude. Left: Due to the topography, the polar vortex in the northern hemisphere is less pronounced and shows a bipolar structure with higher temperatures over East Asia and Alaska and lower values centered over Europe. Right: In southern winter a strong, symmetrical polar vortex is situated in the antarctic region.

titudes. Since the character of the data assimilation strongly influences the received data, certain facts have to be considered. These include the spatially discrete ascent of radiosondes (mainly over densely populated northern hemisphere areas) and the low vertical resolution of temperature data derived from satellite soundings (TOVS). This results in the inability to produce detailed information under strongly perturbed conditions, which can mainly be seen at high northern latitudes during winter time. This in turn results in the fact that some atmospheric patterns do not clearly appear in the models.

It seems that the polar vortex influences atmospheric structures at high latitudes, above all during winter time. In the winter months, the polar vortex results from a radiative cooling of the air over the poles. Due to the topography, the polar vortex is significantly more pronounced and stable in the southern hemisphere (see Figure 4.2, left) than it is in the northern hemisphere, where land and ocean surfaces repeatedly alternate (cf., Figure 4.2, right).

Concerning CHAMP temperatures and ECMWF analyses, the big sampling error arising in December 2002 and 2003 at high northern latitudes as well as in September 2003 at high southern latitudes, seems to result from sampling taking place outside of the polar vortex (where temperatures are higher) and therefore shows higher temperatures than expected.

Compared to ECMWF and NCEP/NCAR (re)analyses, NRLMSISE-00 as well as CIRA86aQ_UoG temperatures arise from long-time average atmospheric parameters.

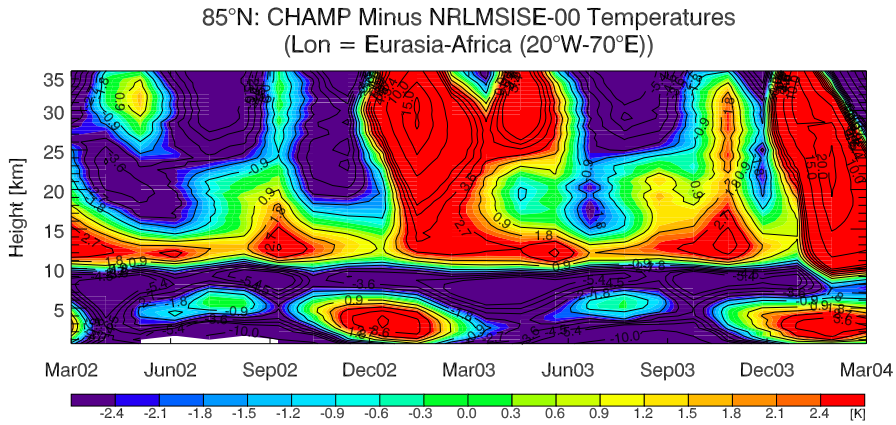


Figure 4.3: Time series (March 2002 to February 2004) of differences between CHAMP RO and NRLMSISE-00 climatologies at 85°N (80°N to 90°N) in the Eurasian-Asian region. Sudden stratospheric warmings took place between November and March. The biggest temperature deviations between CHAMP and NRLMSISE-00 data topped about +15 K in winter 2002/03, and during winter time 2003/04 even more than +20 K occurred in January and February around 20 km.

That is why they do not reflect the actual atmospheric state but represent "average" atmosphere conditions and can, for example, be used to investigate the variations of certain parameters across a few atmospheric boundaries.

Because of the zonal structure of CIRA86aQ_UoG climatologies and the older issue date, NRLMSISE-00 data were used to investigate temporal variations of temperature during the period considered (cf., Section 3.3).

Big differences between the models in northern winter at high latitudes attracted attention, seeming to be caused by "**S**udden **S**tratospheric **W**arming" (SSW) events[1]. SSWs are characterized by short-term (within a few days) stratospheric temperature increase by up to 50 K accompanied by a weakening of the polar vortex. They are initiated by the propagation of planetary wave disturbances from the troposphere into the stratosphere. SSWs occur in the northern hemisphere because of orography and land-sea temperature contrasts.

In the northern hemisphere, stratospheric midwinter warmings have been observed between December and March.

In order to track SSWs in northern stratosphere winters, CHAMP RO climatologies were compared with NRLMSISE-00 data.

Winter 2002/2003 In winter 2002/2003 stratospheric temperatures were extremely low in November and December.

[1] They are also called "Stratospheric Midwinter Warmings". Major (including stratospheric warming and a total change of circulation at 10 hPa) and minor (can be intense too, but they do not result in a reversal of the circulation at the 10 hPa level) midwinter warmings are distinguished.

In November, the monthly mean temperature at 22 km altitude was the lowest observed since 1964 (Vintersol 2003). At the end of December a strong stratospheric warming developed. Further warmings were observed around mid January propagating downwards to tropopause level and then again in February and March. The biggest temperature deviations between CHAMP and NRLMSISE-00 data topped about +15 K (see Figure 4.3).

Winter 2003/2004 Minimum temperatures observed during winter 2003/04 in the lower stratosphere were remarkably above average (cf., Manney et al. (2005) and Labitzke and Naujokat (2004)), and clearly higher than in winter 2002/03. Differences of even more than +20 K occurred in January and February around 20 km (cf., Figure 4.3).

Figure 4.4 depicts the temporal and vertical progression of a major midwinter warming in winter 2003/04 (November 2003 to February 2004) at three different altitude levels, namely 35 km (top row), 25 km (middle row), and 15 km (bottom row).

As can be seen, stratospheric temperatures rose dramatically in December (first at higher levels) with subsequent significantly warming throughout the arctic stratosphere and downward progression in January and February. The following negative temperatures starting at high levels in January propagate downwards during February as well. A very strong polar vortex is re-established in February above 30 km altitude.

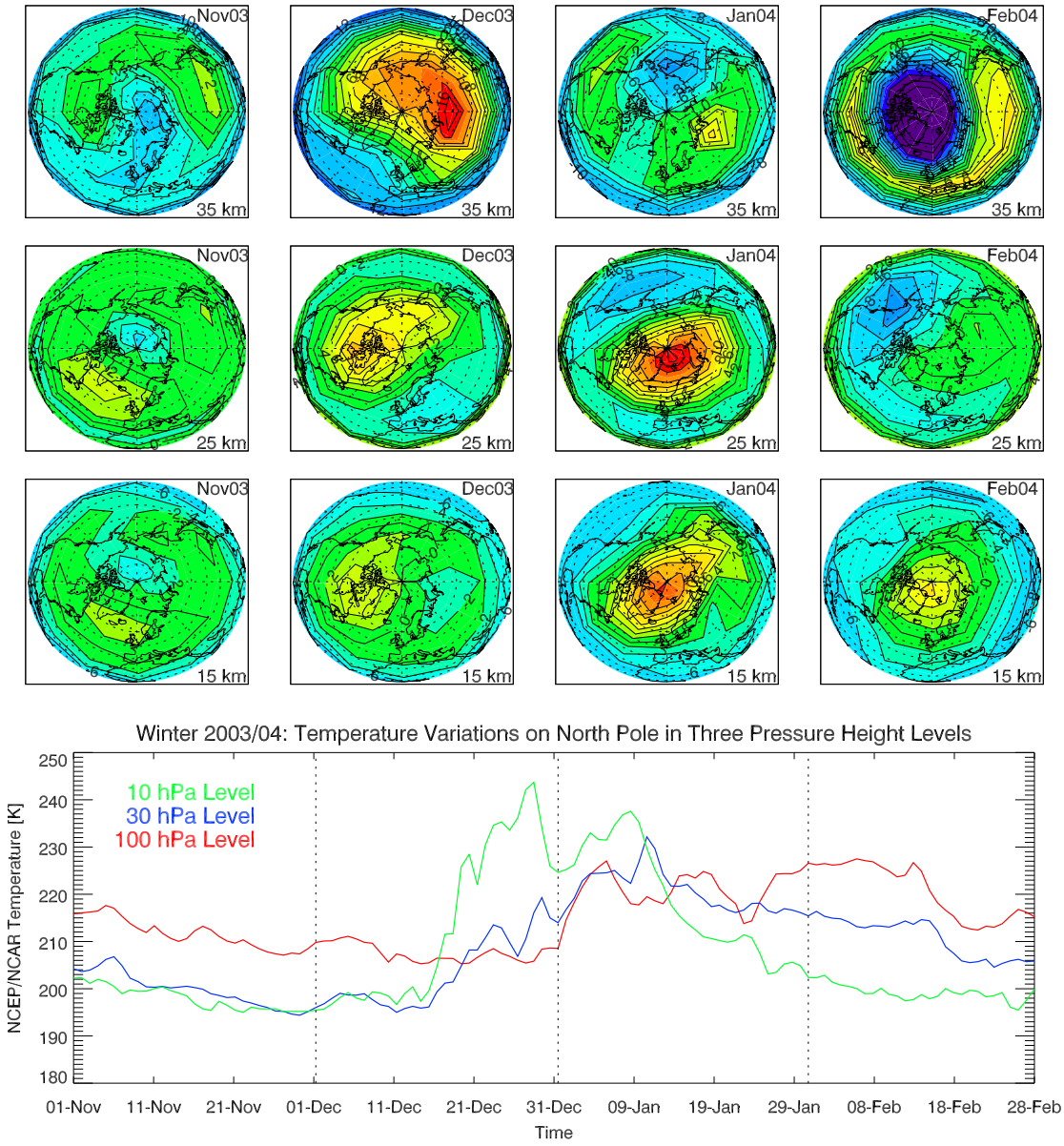


Figure 4.4: Top: Temporal and vertical progression of a major midwinter warming in winter 2003/04 (November 2003 to February 2004) at three different altitude levels, namely 35 km (top row), 25 km (middle row), and 15 km (bottom row). Stratospheric temperatures rose in December (first in higher levels) with subsequent significant warming throughout the arctic stratosphere and downward progression in January and February. The following negative temperatures starting at high levels in January propagate downwards during February. A very strong polar vortex is re-established in February above 30 km altitude. Bottom: NCEP/NCAR daily mean reanalysis temperatures from November 1, 2003 to February 28, 2005, at three pressure height levels. While at high altitudes (green line, about 30 km height) the increase of temperature starts in mid December, there is a time-lag at lower altitudes.

Part II

Factor Analysis and Principal Component Analysis

5 Introduction to Component and Factor Analysis

5.1 General Considerations

(Author: B.C. Lackner)

In this part of the work **P**rincipal **C**omponent **A**nalysis (PCA) and **F**actor **A**nalysis (FA) will be looked at in detail. These are the two most commonly applied types of the great family of methods for multivariate (statistical) analysis. The term "factor analysis" is used twofold in this context. One the one hand, it stands for the variety of different mathematical models applied to uncover latent structures in data sets (such as principal component analysis, cluster analysis etc.), on the other hand, it is an own mathematical model, which will be discussed later on. Getting into the subject, Garson (2005) found a very cute non-technical approach to explain the goal of factor analysis:

> *"A mother sees various bumps and shapes under a blanket at the bottom of a bed. When one shape moves toward the top of the bed, all the other bumps and shapes move toward the top also, so the mother concludes that what is under the blanket is a single thing, most likely her child. Similarly, factor analysis takes as input a number of measures and tests, analogous to the bumps and shapes. Those that move together are considered a single thing, which it labels a factor. That is, in factor analysis the researcher is assuming that there is a "child" out there in the form of an underlying factor, and he or she takes simultaneous movement (correlation) as evidence of its existence."* (Garson 2005)

Factor analysis techniques are multivariate methods. They deal with data containing observations on two or more variables each measured on a set of objects. The measured values of one variable on all objects and the measured values of all variables on one object respectively, are the items of vectors. It is characteristic for factor analysis to look at these vector items in their entirety. Only their common distribution as well as their relationship among each other are of interest.

Even though factor analysis had its beginning in psychology, it is no longer restricted to a certain discipline. Depending on the purpose of the study and the disciplines, there are two main applications:

- Data reduction: To reduce a large number of variables to a smaller number of "factors" for modeling purposes is the most known application of factor analytic techniques.

- Undercover the underlying structure: Factor analysis tries to unveil patterns of relationship among many dependent variables. A smaller subset of variables is selected from the entire data set, based on the highest correlations of the original variables.

Following these main applications, two key concepts can be formulated. **Exploratory Factor Analysis** is the most common form of factor analysis techniques. There is no prior theory and one seeks to uncover the underlying structures of a relatively large set of variables, assuming that any indicator may be associated with any factor.

On the other hand, **Confirmatory Factor Analysis** tries to determine if the number of selected factors meet the assumptions on the basis of a pre-established theory. In other words, the purpose of this kind of analysis is to confirm a hypothesized factor structure. It plays an important role in social sciences where researchers, for instant, try to determine, if measures created to represent a latent variable really belong together (Garson 2005).

Dependent on which variances or correlations are of interest, several "factor analytical data modes" can be distinguished. A clear description of Cattell's Data Cube, which is depicted in Figure 5.1, present Reyment and Jöreskog (1993). Cattell was the first who described phenomenons in terms of three fundamental properties or dimensions, namely objects, variables, and occasions (the latter standing for time). This idea of

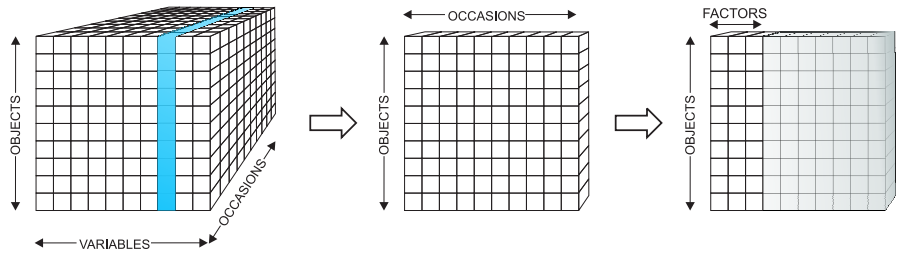


Figure 5.1: Left: Cattell's Data Cube consists of objects, variables, and occasions. According to Reyment and Jöreskog (1993). Middle: For S-mode analysis, which was performed on selected CHAMP RO data, an object-by-occasion slab was cut out of the data cube. Right: A S-mode factor analysis results in a factor matrix as depicted. The gray accentuated columns signify factors, which can be neglected as they do not contribute much to the total variance of the data set.

data classification, originating from psychology, can be transferred to nearly any study, where such kind of data are investigated. A strategy to analyze data is to consider pairs of two dimensions, leaving the third one fixed. This is equivalent to cutting out any one slab of the data cube. There are six possible pairs, leading to six possible analysis modes, which are referred to as O-, P-, Q-, R-, S-, and T-mode analyses.

- O- and P-mode analyses examine variable-by-occasion slabs. The data matrix shows the variation of characteristics of an object over a period of time and therefore it is a form of time series analysis.

- Q- and R-mode analyses investigate object-by-variable slabs. R-mode analysis is most common in social sciences and often not labeled as such, as it is assumed to be the "normal" kind of analyses. Q-mode factor analysis is also called "inverse factor analysis" as it seeks to cluster the cases rather than the variables at a given point of time.

- In S- and T-mode analyses, the data matrix builds up of an object-by-occasion slab. From this it follows that one variable, measured on a number of objects located at various points of time, is analyzed.

On each of these slabs, some form of factor analysis can be performed. In atmospheric sciences, different arrangements of meteorological variables, grid points, and time series take the place of Cattell's dimensions. The most frequently used mode in atmospheric sciences is S-mode (Jolliffe 2002). There, one fixed meteorological variable is measured at a certain number of grid points for several times, which results in an object-by-occasion slab. Such kind of slabs of Cattell's cube (blue accentuated area in left Figure 5.1) were used for our investigations in this part of the work, which were carried out with selected CHAMP RO temperature fields and will be discussed in detail, subsequent to the theoretical description of principal component and factor analysis.

As PCA and FA are mostly treated in the form of R-mode in literature, the notation of an object-by-variable data set will also be used throughout this work, where the spatial locations are the variables, the different times are the objects, and the meteorological variable (temperature) is fixed (cf., Section 5.2).

Factor analysis as an own method and principal component analysis seem to resemble each other, as both can be used to analyze the structure of covariance or correlation matrices with the goal to either reduce the dimensionality of variables or to estimate latent variables. But the aims of the two methods are not the same.

Principal component analysis, short PCA, which will be described in detail in Chapter 6, transforms a set of p variables linearly and orthogonally into a number of new hypothetical variates, called components, which are uncorrelated. To define these new variables, the latent roots (arranged in descending order) and respective vectors of a covariance or correlation matrix are used. The new variables are chosen such that the first variable accounts for the maximum variance of the data, the second for the maximum residual variance and so on. In most cases, few new variables explain a large proportion of the total variance found in the data.

Factor analysis aims in explaining the covariances of the variables in terms of a much smaller number of hypothetical variables, called factors (the theory of factor analysis will be discussed in Chapter 7). In general, the correlation matrix is used and the

first question to be answered is whether there are any correlations between the variables or not. If there are correlations, factor analysis seeks for a new variable such that all partial correlation coefficients between the original variables are zero after eliminating the effect of this new variable. If not, two new variables are postulated and so on (Lawley and Maxwell 1971).

Thus, for the beginning we just want to record that factor analysis is covariance- or correlation oriented, whereas principal component analysis is variance-oriented.

5.2 The Factor Model

(Author: B. Pirscher)

5.2.1 Data Matrix

Statistical methods are an essential part in the investigation of large samples. Compared to the past, larger amounts of data are collected, and the rate will continue to accelerate in the next decades.

Meteorological data are recorded at different times at diverse locations; afterward they are organized and reproduced in a data matrix. A feasible data matrix $\boldsymbol{X}$ is depicted in Table 5.1; it is composed of n objects and p variables, resulting in a $(n \times p)$-matrix.

	Variable 1	Variable 2		Variable j		Variable p
Object 1	x_{11}	x_{12}	$\dots$	x_{1j}	$\dots$	x_{1p}
Object 2	x_{21}	x_{22}	$\dots$	x_{2j}	$\dots$	x_{2p}
$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\dots$	$\vdots$
Object i	x_{i1}	x_{i2}	$\dots$	x_{ij}	$\dots$	x_{ip}
$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\dots$	$\vdots$
Object n	x_{n1}	x_{n2}	$\dots$	x_{nj}	$\dots$	x_{np}

Table 5.1: Structure of a data matrix. Each variable is assigned to one column, each object (observation) to one row.

In case of the CHAMP RO temperatures, the grid point information (latitude, longitude, height) is arranged in the columns of the data matrix (each grid point represents one variable), whereas the temporal information of the time series can be found in the rows. The spatial resolution depends on the analyzed atmospheric field, two global areas (Eurasian-African latitude $\times$ height slice and a map at an altitude of 15 km) as well as two regional fields (situated at the high southern latitudes and at the equator) were investigated. The time series always stayed the same and contained 36 monthly means (from March 2002 to February 2005).

5.2.2 Description of the Model

The derivation of the model follows to a large extend the one of Reyment and Jöreskog (1993).

The data matrix $\boldsymbol{X}$ can be analytically made up of the two matrices $\boldsymbol{F}$ and $\boldsymbol{A}$ with

$$\boldsymbol{X}_{(n\times p)} = \boldsymbol{F}_{(n\times p)}\boldsymbol{A}'_{(p\times p)}. \tag{5.1}$$

The matrix $\boldsymbol{F}$ is called matrix of factor scores (principal components); each column vector of $\boldsymbol{F}$ represents a hypothetical variable. The n elements in one column reflect the amount of variance of this factor. The coefficients of the linear combination, the elements of $\boldsymbol{A}$, are known as factor loadings.

To achieve the aim of data reduction, it is possible to single out the first k hypothetical variables from $\boldsymbol{F}$. Then the observed data matrix $\boldsymbol{X}$ is portioned into a "systematic" part $\boldsymbol{FA'}$ and an "error" part $\boldsymbol{E}$. The systematic part is proposed to be educible as a linear combination of a few factor variables that describe or explain the interdependence of a set of variables in terms of the factors (Anderson 1984). The factor loading matrix $\boldsymbol{A}$ combined with the matrix $\boldsymbol{F}$ results in an estimate of the data set $\boldsymbol{X}$.

$$\boldsymbol{X}_{(n\times p)} = \boldsymbol{F}_{(n\times k)}\boldsymbol{A}'_{(k\times p)} + \boldsymbol{E}_{(n\times p)} \tag{5.2}$$

The matrix of residuals $\boldsymbol{E}$ contains the error terms; k specifies the number of factors being used, necessarily $k \leq p$ (cf., Section 8.1).

Transforming equation (5.2) in scalar notation results in

$$x_{ij} = \sum_{l=1}^{k} f_{il}a_{jl} + e_{ij}. \tag{5.3}$$

Any particular row, $\boldsymbol{x}'$ of $\boldsymbol{X}$ can be written as

$$\boldsymbol{x}' = \boldsymbol{f}'\boldsymbol{A}' + \boldsymbol{e}' \tag{5.4}$$

and equation (5.2) yields, written in vector algebra,

$$\boldsymbol{x} = \boldsymbol{Af} + \boldsymbol{e}. \tag{5.5}$$

$\boldsymbol{x}' = (x_1, \ldots, x_p)$ is one of the selected objects of the data matrix, $\boldsymbol{f}' = (f_1, \ldots, f_k)$ is the vector of the matrix $\boldsymbol{F}$ yielding, as the product with $\boldsymbol{A}$, an optimal estimate of $\boldsymbol{X}$, and $\boldsymbol{e}' = (e_1, \ldots, e_p)$ is the corresponding vector of residuals.

Equation (5.5) is generally known as the fundamental model equation for factor analytic techniques. It reveals that each observed variable is composed of a weighted sum of factors plus an error term $\boldsymbol{e}$, which results from the difference between the actually observed vector $\boldsymbol{x}$ and the estimated state vector $\boldsymbol{Af}$ (Reyment and Jöreskog 1993).

The model equations of the variables are

$$\begin{aligned} x_1 &= a_{11}f_1 + a_{12}f_2 + \cdots + a_{1k}f_k + e_1 \\ x_2 &= a_{21}f_1 + a_{22}f_2 + \cdots + a_{2k}f_k + e_2 \\ \vdots & \quad \vdots \\ x_p &= a_{p1}f_1 + a_{p2}f_2 + \cdots + a_{pk}f_k + e_p, \end{aligned} \tag{5.6}$$

whereas the best estimate for $(x_1, \ldots, x_p)$ is

$$\begin{aligned} c_1 &= a_{11}f_1 + a_{12}f_2 + \cdots + a_{1k}f_k \\ c_2 &= a_{21}f_1 + a_{22}f_2 + \cdots + a_{2k}f_k \\ \vdots & \quad \vdots \\ c_p &= a_{p1}f_1 + a_{p2}f_2 + \cdots + a_{pk}f_k. \end{aligned} \tag{5.7}$$

5.3 Geometrical Considerations

(Author: B.C. Lackner)

5.3.1 Variable, Object, and Factor Space

The vectors of the variables can be interpreted as coordinate axis of a p-dimensional space in such way that each column of the data matrix $\boldsymbol{X}$ forms one coordinate axis. In this **variable space**, the n objects are depicted by n points and their coordinates are given by a p-tuple of (variable) values.

On the other hand, objects can be as well understood as coordinate axis. In this **object space**, the coordinate axis correspond to the p column vectors of the data matrix $\boldsymbol{X}$ and each variable is depicted by a radial vector with n items whose starting point is the origin of the coordinate system.

The length of a radial vector $\boldsymbol{x}_i$ is the distance from the end-point to the origin of the coordinate system ($\sqrt{\sum (x_{ij} - 0)^2} = \sqrt{\sum (x_{ij})^2} = |\boldsymbol{x}_i|$). Hence, in a rectangular coordinate system, the magnitude of a vector is its geometrical length.

By analogy with the variable space, a **factor space** can be created by taking the p columns of the factor matrix $\boldsymbol{F}$ as coordinate axis. Due to the orthogonality of the factors, a rectangular coordinate system arises. Usually, only k factors are selected, which form a k-dimensional subspace of the factor space.

According to the relation $\boldsymbol{X} = \boldsymbol{F}\boldsymbol{A}'$, the variable vectors of $\boldsymbol{X}$ can be mapped in the factor space by the factor matrix $\boldsymbol{F}$. For every variable vector $\boldsymbol{x}_i$ exists a radial vector $\boldsymbol{f}_i$ in the factor space. The end-point coordinates of $\boldsymbol{x}_i$ are the factor loadings in the factor space.

5.3.2 Eigenvalues and Eigenvectors

The following considerations are based on Reyment and Jöreskog (1993). To understand the geometrical role of eigenvalues and eigenvectors, it is advisable to consider a data set with two variables. On condition that the variables are standardized and normally distributed, a bivariate scatter diagram will show the contour of an ellipse (uncorrelated variables will form a circle, perfectly correlated variables a line). If more than two variables are considered, a p-dimensional hyper ellipsoid will be formed, which can be described by $\boldsymbol{w}'\boldsymbol{R}^{-1}\boldsymbol{w} = \text{const}$. In this equation $\boldsymbol{R}$ is the correlation or the covariance matrix and $\boldsymbol{w}$ is a vector containing the coordinates of the points on the hyper ellipsoid. Now, the axis of the (hyper) ellipsoid can be formulated:

- The major axis is determined by the points furthest away from the centroid. The result of this maximization problem (find the points on the ellipsoid which squared distance to the origin $\boldsymbol{w}'\boldsymbol{w}$ is maximized) leads to the characteristic equation $\boldsymbol{R}\boldsymbol{w} = \lambda\boldsymbol{w}$, with $\boldsymbol{w}'\boldsymbol{w} = \lambda$. One may conclude that $\boldsymbol{w}$ has to be an eigenvector of $\boldsymbol{R}$, corresponding to the largest eigenvalue λ. Geometrically, the eigenvector defines the direction of the major axis of the (hyper)ellipsoid, whereas the length of the axis is given by the root of the eigenvalue.
- In a similar way, the second axis is constructed with the second largest eigenvalue and its corresponding eigenvector and so on.

In summary it may be said that the eigenvectors point out the directions of the maximum variances. The eigenvector of the largest eigenvalue indicates the direction of the maximum variance, the eigenvector of the second largest eigenvalue indicates, as the eigenvectors are linearly independent, the direction of the maximum variance orthogonal to the first one and so on. In the same way it is obvious that, as the root of an eigenvalue stands for the length of an axis, an eigenvalue of zero indicates a corresponding axis of zero length, suggesting that the dimensionality of the space containing the data points is less than the original space.

6 Principal Component Analysis

(Author: B. Pirscher)

6.1 Introduction to Principal Component Analysis

Principal **C**omponent **A**nalysis (PCA) was originated by Pearson (1901) who developed the mathematical foundation of PCA; Hotelling (1933) derived and formalized the solution of the principal components.

The method has become one of the best known techniques of exploratory multivariate data analysis and has found applications in many areas of scientific research. The technique has been applied for example in astronomy by Kanbur and Mariani (2004) who investigated light curves of RR Lyrae, in medicine by Dutta et al. (2005) who examined artherosclerotic human coronary arteries, in geology by Thy and Esbensen (1993) who looked for lava and dike compositions, in social science by Scherer and Avellaneda (2001) or Brockett et al. (2002), in physics by Natraj et al. (2005), and, as in that case, in climatology.

The Principal component analysis is a multivariate statistical method, which aims at

- reducing the number of variables in a data set and
- detecting some structures (dominant patterns of variation) in the relationship between the variables.

An orthogonal transformation of variables generates a new set of uncorrelated hypothetical variables, the factor scores $\boldsymbol{f}$ whose synonym is "**p**rincipal **c**omponents" (PC) in PCA. The new variables are sorted into descending order according to their amount of accounted variance. So, the first principal component represents most of the variance, the second one accounts for a maximum of the remaining variance. Each further PC accounts for a maximum amount of residual variance. Cumulatively, all the new variables account for 100 % of the intrinsic variation. With the view of getting a smaller dimension only few new variables should account for a huge amount of variation, so most of the intrinsic information will be conserved.

The mathematical technique used in PCA is called eigenanalysis. That means that a square symmetric matrix (the covariance matrix or the correlation matrix) will be solved for their eigenvalues and eigenvectors. The eigenvector associated with the largest eigenvalue has the same direction as the first principal component. The eigenvector associated with the second largest eigenvalue determines the direction of the second principal component and it is perpendicular to the first one.

6.2 Definition of Principal Components

To define the principal components (PC) it is necessary to fall back on the mathematical model given in equation (5.2) resulting in

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{f} \tag{6.1}$$

in vector notation.

Suppose that $\boldsymbol{x}' = (x_1, \ldots, x_p)$ is a vector of one object consisting of p variables being analyzed. Then, the principal components $\boldsymbol{f}' = (f_1, \ldots, f_p)$ are calculated by a linear combinations of $\boldsymbol{x}$.

$$\begin{array}{rclcll} f_1 & = & \boldsymbol{a}_1'\boldsymbol{x} & = & a_{11}x_1 + a_{21}x_2 + \cdots + a_{p1}x_p & = & \sum_{i=1}^{p} a_{i1}x_i \\ f_2 & = & \boldsymbol{a}_2'\boldsymbol{x} & = & a_{12}x_1 + a_{22}x_2 + \cdots + a_{p2}x_p & = & \sum_{i=1}^{p} a_{i2}x_i \\ \vdots & & \vdots & & \vdots & & \\ f_p & = & \boldsymbol{a}_p'\boldsymbol{x} & = & a_{1p}x_1 + a_{2p}x_2 + \cdots + a_{pp}x_p & = & \sum_{i=1}^{p} a_{ip}x_i \end{array} \tag{6.2}$$

and, in general,

$$f_j = \boldsymbol{a}_j'\boldsymbol{x} = a_{1j}x_1 + a_{2j}x_2 + \cdots + a_{pj}x_p = \sum_{i=1}^{p} a_{ij}x_i, \qquad j = 1, \ldots, p \tag{6.3}$$

with $\boldsymbol{a}_j' = (a_{1j}, a_{2j}, \ldots, a_{pj})$ being a vector of constants. It is normalized to unit length, so that $\boldsymbol{a}_j'\,\boldsymbol{a}_j = \sum_{i=1}^{p} a_{ij}^2 = 1$, which denotes an orthogonal transformation.

As mentioned above the variance of the original data set is portioned to the principal components $f_1, \ldots, f_p$, whereas the first few f_j will express most of the variation and the last ones can be considered to be constant.

6.3 Derivation of Principal Components

The calculation of the principal components $\boldsymbol{f}$ and the parameters $\boldsymbol{a}$ are based on the calculation of the covariance matrix $\boldsymbol{\Sigma}$ or the correlation matrix $\boldsymbol{P}$; the following implementation will be based on the covariance matrix, details concerning the correlation matrix will be discussed afterward.

The first principal component f_1 has to fulfill the requirements that

1. $\mathrm{Var}(f_1) = \mathrm{Var}(\boldsymbol{a}_1'\boldsymbol{x}) \rightarrow$ Maximum and
2. $\boldsymbol{a}_1'\boldsymbol{a}_1 = 1$.

Because of

$$\begin{array}{rcl} \mathrm{Var}(f_1) & = & \mathrm{Var}(\boldsymbol{a}_1'\boldsymbol{x}) \\ & = & \mathrm{E}[\boldsymbol{a}_1'\boldsymbol{x} - \mathrm{E}(\boldsymbol{a}_1'\boldsymbol{x})]^2 \\ & = & \mathrm{E}[\boldsymbol{a}_1'[\boldsymbol{x} - \mathrm{E}(\boldsymbol{x})]]^2 \\ & = & \boldsymbol{a}_1'\mathrm{E}[(\boldsymbol{x} - \mathrm{E}(\boldsymbol{x}))'(\boldsymbol{x} - \mathrm{E}(\boldsymbol{x}))]\boldsymbol{a}_1 \\ & = & \boldsymbol{a}_1'\boldsymbol{\Sigma}\boldsymbol{a}_1 \end{array} \tag{6.4}$$

with $\boldsymbol{\Sigma}$ being the covariance matrix, the first condition leads to the maximization of the product $\boldsymbol{a}_1'\boldsymbol{\Sigma}\boldsymbol{a}_1$ under the constraint of $\boldsymbol{a}_1'\boldsymbol{a}_1 = 1$.

Applying the technique of the lagrange multipliers (cf., Appendix B.1) to the function $\boldsymbol{a}_1'\boldsymbol{\Sigma}\boldsymbol{a}_1$ yields

$$L(\boldsymbol{a}_1, \lambda) = \boldsymbol{a}_1'\boldsymbol{\Sigma}\boldsymbol{a}_1 - \lambda(\boldsymbol{a}_1'\boldsymbol{a}_1 - 1). \tag{6.5}$$

To get an extremum it is necessary to set the derivation of (6.5) to zero:

$$\frac{\partial(\boldsymbol{a}_1'\boldsymbol{\Sigma}\boldsymbol{a}_1)}{\partial\boldsymbol{a}_1} = 2\boldsymbol{\Sigma}\boldsymbol{a}_1 \qquad \text{and} \qquad \frac{\partial(\lambda\boldsymbol{a}_1'\boldsymbol{a}_1)}{\partial\boldsymbol{a}_1} = 2\lambda\boldsymbol{a}_1, \tag{6.6}$$

getting

$$\frac{\partial L}{\partial\boldsymbol{a}_1} = 2\boldsymbol{\Sigma}\boldsymbol{a}_1 - 2\lambda\boldsymbol{a}_1 = 0 \tag{6.7}$$

and

$$(\boldsymbol{\Sigma} - \lambda\boldsymbol{I}_p)\boldsymbol{a}_1 = 0 \tag{6.8}$$

with $\boldsymbol{I}_p$ being the $(p \times p)$-identity matrix.

Since equation (6.8) is a homogeneous system of equations, a nontrivial solution only exists if the matrix $(\boldsymbol{\Sigma} - \lambda\boldsymbol{I}_p)$ is singular, and

$$\det(\boldsymbol{\Sigma} - \lambda\boldsymbol{I}_p) = 0.$$

This equation is equal to the characteristic equation (cf., Appendix A.1.10), thus λ is an eigenvalue of the covariance matrix $\boldsymbol{\Sigma}$. To decide which eigenvalue is suitable to the first principal component, look at

$$\begin{aligned} \mathrm{Var}(\boldsymbol{a}_1'\boldsymbol{x}) &= \boldsymbol{a}_1'\boldsymbol{\Sigma}\boldsymbol{a}_1 \\ &= \boldsymbol{a}_1'\lambda\boldsymbol{I}_p\boldsymbol{a}_1 \\ &= \lambda\boldsymbol{a}_1'\boldsymbol{I}_p\boldsymbol{a}_1 \\ &= \lambda\boldsymbol{a}_1'\boldsymbol{a}_1 \\ &= \lambda \end{aligned} \tag{6.9}$$

and notice that the largest (first) eigenvalue is the appropriate one. The first principal component $f_1 = \boldsymbol{a}_1'\boldsymbol{x}$ is composed of $\boldsymbol{a}_1$ being the eigenvector of $\boldsymbol{\Sigma}$ associated with the largest eigenvalue λ_1 and the data vector $\boldsymbol{x}$.

The derivation of the second principal component is similar to the first one. On condition that the second principal component has to explain maximum residual variance, $\boldsymbol{a}_2'\boldsymbol{a}_2 = 1$ and $\boldsymbol{a}_2'\boldsymbol{a}_1 = 0$, it follows that

$$(\boldsymbol{\Sigma} - \lambda\boldsymbol{I}_p)\boldsymbol{a}_2 = 0 \tag{6.10}$$

and the eigenvector $\boldsymbol{a}_2$ associated with the second largest eigenvalue λ_2 of $\boldsymbol{\Sigma}$ belongs to the second principal component $f_2 = \boldsymbol{a}_2'\boldsymbol{x}$.

In general, the j^{th} principal component f_j is $\boldsymbol{a}_j'\boldsymbol{x}$ and its variance is $\text{Var}(\boldsymbol{a}_j'\boldsymbol{x}) = \lambda_j$, with λ_j being the j^{th} largest eigenvalue of the covariance matrix $\boldsymbol{\Sigma}$ and $\boldsymbol{a}_j$ being the corresponding eigenvector.

Writing each eigenvector $\boldsymbol{a}_j,\ j = 1, \ldots, p$ in one of the columns of the matrix $\boldsymbol{A}_{(p\times p)}$, equation (6.2) can be written as

$$\boldsymbol{f} = \boldsymbol{A}'\boldsymbol{x}. \tag{6.11}$$

Because $\boldsymbol{A}_{(p\times p)}$ is an orthogonal matrix, the principal components $\boldsymbol{f} = (f_1, \ldots, f_p)$ are defined by an orthogonal linear transformation of $\boldsymbol{x}$.

Leaving vector notation for a short moment and going over to matrix algebra yields

$$\boldsymbol{F} = \boldsymbol{X}\boldsymbol{A} \tag{6.12}$$

with $\boldsymbol{F}$ being the matrix of principal components, $\boldsymbol{X}$ the data matrix, and $\boldsymbol{A}$ the matrix of eigenvectors.

6.4 Properties of Principal Components

6.4.1 Covariance Matrix of Principal Components

Calculating the covariance matrix of $\boldsymbol{f}$, it can be noticed that it results in a diagonal matrix whose diagonal elements are the eigenvalues of $\boldsymbol{\Sigma}$, hence it is named $\boldsymbol{\Lambda}$

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_p \end{pmatrix}. \tag{6.13}$$

The diagonal structure results from the mutual uncorrelation (independence) of the principal components. In this connexion it can be found that

$$\boldsymbol{\Lambda} = \text{Var}(\boldsymbol{f}) = \text{Var}(\boldsymbol{A}'\boldsymbol{x}) = \boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A} \tag{6.14}$$

and (because of $\boldsymbol{A}$ being an orthogonal matrix with $\boldsymbol{A}\boldsymbol{A}' = \boldsymbol{I}_p$)

$$\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{\Lambda}\boldsymbol{A}'. \tag{6.15}$$

Equations (6.14) and (6.15) describe the relationship between the covariance matrix of the principal components and the covariance matrix of the original data set.

6.4.2 Variance of Principal Components

It is actually known that the eigenvalues of the covariance matrix can be interpreted as variances of the principal components. Because of

$$\begin{aligned}\sum_{l=1}^{p} \mathrm{Var}(f_l) &= \sum_{l=1}^{p} \lambda_l \\ &= \mathrm{tr}(\boldsymbol{\Lambda}) \\ &= \mathrm{tr}(\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}) \\ &= \mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{A}') \\ &= \mathrm{tr}(\boldsymbol{\Sigma}) \\ &= \sum_{l=1}^{p} \mathrm{Var}(x_l),\end{aligned} \tag{6.16}$$

it can be recognized that the sum of variances of the original data set is equal to the sum of variances of the principal components, provided that $\mathrm{tr}(\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{A}')$ which is true if $\boldsymbol{A}$ and $\boldsymbol{\Sigma}$ are square matrices[1]. That means that the total variance of the original data set $\boldsymbol{X}$ is broken up into p components, whereas each of these components is obtained by projecting $\boldsymbol{X}$ onto one of the eigenvectors $\boldsymbol{a}_j$.

The j^{th} principal component f_j accounts for a particular proportion of total variation. In relative terms it contains

$$\text{fraction of total variation}(f_j) = \frac{\lambda_j}{\sum_{l=1}^{p} \lambda_l} \cdot 100 \tag{6.17}$$

and the first k principal components explain

$$\text{fraction of total variation}\left(\sum_{l=1}^{k} f_l\right) = \frac{\sum_{l=1}^{k} \lambda_l}{\sum_{l=1}^{p} \lambda_l} \cdot 100, \qquad k \leq p, \tag{6.18}$$

of the total variation; both proportions are given as percentage.

6.4.3 Use of the Correlation Matrix

Using the correlation matrix $\boldsymbol{P}_{(p\times p)}$ to derive principal components results in

$$\boldsymbol{f} = \boldsymbol{A}'\boldsymbol{z}, \tag{6.19}$$

where the columns of $\boldsymbol{A}$ (the rows of $\boldsymbol{A}'$, respectively) contain the eigenvectors of the correlation matrix and $\boldsymbol{z}$ are the standardized variables of $\boldsymbol{x}$. The determination of the correlation matrix (more precisely the sample correlation matrix) can be found in Section 6.4.5.

[1] The trace of a product of two square matrices is independent of the order of the multiplication.

The derivation of the principal components basing on the correlation matrix is the same as that basing on the covariance matrix.

Applying equation (6.16) to the correlation matrix yields that the sum of variances of the principal components (being equal to the sum of variances of the standardized variables) is equal to p, the number of variables. That is because the correlation matrix $\boldsymbol{P}$ only consists of ones in the principal diagonal so that the trace of the correlation matrix is p.

Furthermore, these sums equal to the sum of eigenvalues of the correlation matrix

$$\sum_{l=1}^{p} \lambda_l = p. \tag{6.20}$$

The proportion of total variance accounted for the j^{th} principal component is

$$\text{fraction of total variation}(f_j) = \frac{\lambda_j}{p}. \tag{6.21}$$

The eigenvalues and eigenvectors of the covariance and the correlation matrix are different and cannot be transfered into each other by a simple mathematical formula. Even if the principal components derived from the correlation matrix are rescaled by multiplication with the standard deviation, they will be different from the principal components calculated from the covariance matrix. The reason is that principal components are invariant under orthogonal transformations of $\boldsymbol{x}$ but not under oblique transformations, and the transformation from $\boldsymbol{x}$ to $\boldsymbol{z}$ (standardized variable) is counted among the latter (Jolliffe 2002).

The advantage of the correlation matrix over the covariance matrix is that the variables are standardized and their variances equal to one so that all variables can be weighted equally. That is useful, if the variables are measured in different units because they will be more directly comparable. Jolliffe (2002) mentions two disadvantages of the correlation matrix PCs namely that it is more difficult to base statistical inference and that they are less easy to be interpreted.

The covariance matrix is appropriate, if all variables have the same unit because the variables having a large variance will dominate the first principal components and that is the favored property calculating PCs. Nevertheless, the correlation matrix can also be successfully used in this case.

Jolliffe (2002) shows, considering an example, that the correlation matrix is more suitable for calculating PCs than the covariance matrix if the variances (thus the standard deviations) of the individual variables are widely scattered. The reason is that the first few "covariance based" principal components mainly contain information about the relative sizes of variances.

6.4.4 Principal Components With Small Variances

If there are linear dependencies between original variables, some eigenvalues of the covariance matrix will be zero. Supposing that m eigenvalues are zero, the rank of $\boldsymbol{\Sigma}$ is

$(p - m)$ rather than p. Any principal component with zero variance defines an exact constant linear relationship between some of the elements of $\boldsymbol{x}$, which results from redundant information of the original data set. Knowing these relationships allows to reduce of the number of variables in the data without loosing any information.

The number of zero eigenvalues arising from the covariance matrix is equal to the number of zero eigenvalues arising from the correlation matrix.

6.4.5 Sample Principal Components and Matrix Notation

Jolliffe (2002) mentions some different ways of centering a data matrix. One possibility is the centering by either medians or modes but it is also possible to center the data about the mean for each variable, or to refer them to the mean of each variable and to each observation.

The second possibility, the reference of the variables to their means $\bar{x}_j$ will be applied when calculating the sample covariance matrix.

The mean of one variable, which is defined as

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}, \qquad j = 1, \ldots, p \tag{6.22}$$

is used when calculating the deviation scores of the variables

$$y_{ij} = x_{ij} - \bar{x}_j. \tag{6.23}$$

x_{ij} and y_{ij} can be thought to be the ij^{th} element of the matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ containing the original data and the deviation scores, respectively. The anomalies y_{ij} are the values of the j^{th} variable (x_j) measured around its mean $\bar{x}_j$ for any i observation.

Because the covariance matrix $\boldsymbol{\Sigma}$ is never known exactly, the deviation scores can be used to calculate the sample covariance by

$$s_{jj'} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)(x_{ij'} - \bar{x}_{j'}) \tag{6.24}$$

and it is convenient to define the sample covariance matrix by

$$\boldsymbol{S} = \frac{1}{n-1}\boldsymbol{Y'Y}. \tag{6.25}$$

$\boldsymbol{Y}_{(n\times p)}$ is the data matrix in deviated form; it is composed of n rows (number of objects) and p columns (number of variables).

If the data form a random sample from a multivariate distribution with finite second moments, then $\boldsymbol{S}$ is an unbiased estimate of the true covariance matrix.

The calculation of the principal components by the sample covariance matrix is equal to the calculation mentioned above. Strictly speaking, the eigenvalues and the eigenvectors of the sample covariance matrix have to be denoted by $\hat{\lambda}_1, \ldots, \hat{\lambda}_p$ and $\hat{\boldsymbol{a}}_1, \ldots, \hat{\boldsymbol{a}}_p$, respectively.

Following that notation and using simple matrix algebra, the principal components can be calculated by

$$\boldsymbol{F} = \boldsymbol{Y}\boldsymbol{A}. \tag{6.26}$$

Each column of the matrix $\boldsymbol{A}$ contains one eigenvector of the sample covariance matrix.

Because of the adjustment of the variables by a constant value, the principal components will have zero mean values instead of $\bar{f}_1, \ldots, \bar{f}_p$.

The sample correlation matrix $\boldsymbol{R}$ consists of the elements

$$r_{jj'} = \frac{1}{n-1} \sum_{i=1}^{n} \frac{(x_{ij} - \bar{x}_j)}{s_j} \frac{(x_{ij'} - \bar{x}_{j'})}{s_{j'}} \tag{6.27}$$

with $r_{jj'}$ being the correlation coefficient between the variables j and $j^{'}$. $\bar{x}_j$ and $\bar{x}_{j'}$ are the actually known sample means of x_j and $x_{j'}$, s_j and $s_{j'}$ are the corresponding standard deviations,

$$s_j = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}, \qquad i = 1, \ldots, p. \tag{6.28}$$

Because of

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \tag{6.29}$$

being known as z-standardized data the sample correlation matrix follows from

$$\boldsymbol{R} = \frac{1}{n-1} \boldsymbol{Z}'\boldsymbol{Z} \tag{6.30}$$

and the PCs are calculated by

$$\boldsymbol{F} = \boldsymbol{Z}\boldsymbol{A} \tag{6.31}$$

with $\boldsymbol{A}$ containing the eigenvectors of the correlation matrix.

6.4.6 Normalization of Principal Components

Instead of the common normalization $\boldsymbol{a}_j'\boldsymbol{a}_j = 1$, it is possible to use other kinds of normalization techniques. When using the command PCOMP in IDL the normalization results from

$$\widetilde{f_j} = \sqrt{\lambda_j} f_j \tag{6.32}$$

and, interestingly,

$$\widetilde{\boldsymbol{a}}_j = \sqrt{\lambda_j} \boldsymbol{a}_j. \tag{6.33}$$

In that case, the principal components and the coefficients should be renormalized.

Another common practice (especially in meteorological and climatological applications) is the renormalization by

$$\widetilde{f}_j = \frac{1}{\sqrt{\lambda_j}} f_j \tag{6.34}$$

and

$$\widetilde{\boldsymbol{a}}_j = \sqrt{\lambda_j}\boldsymbol{a}_j \tag{6.35}$$

so that $\mathrm{Var}(\widetilde{f}_j) = 1$.

Since

$$\widetilde{\boldsymbol{a}}_j'\widetilde{\boldsymbol{a}}_j = \lambda_j \tag{6.36}$$

the lengths of the eigenvectors are proportional to their respective eigenvalues.

Using matrix notation equation (6.35) yields

$$\widetilde{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{\Lambda}^{1/2} \tag{6.37}$$

and the renormalized eigenvectors can be found in the columns of the matrix $\widetilde{\boldsymbol{A}}$.

The covariance matrix restates (cf., equation (6.15))

$$\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{\Lambda}\boldsymbol{A}' = \widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{A}}' \tag{6.38}$$

and analogical the correlation matrix can be calculated by

$$\boldsymbol{P} = \widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{A}}'. \tag{6.39}$$

knowing that the matrix $\widetilde{\boldsymbol{A}}$ differs if it is calculated from the covariance or the correlation matrix because of $\boldsymbol{A}$ and $\boldsymbol{\Lambda}$ being different.

By means of the matrix $\widetilde{\boldsymbol{A}}$ the fraction of total variation accounted for by a factor can be calculated by

$$\text{fraction of total variation}(\widetilde{f}_j) = \frac{\sum_{i=1}^{p} \widetilde{a}_{ij}^2}{\sum_{i=1}^{p} \lambda_i} \cdot 100, \tag{6.40}$$

which is the same as equation (6.17).

In atmospheric science the eigenvectors defining the principal components are often referred to **e**mpirical **o**rthogonal **f**unctions (EOFs). The term "coefficient" or "loading" applies to the eigenvectors or to the renormalized eigenvectors, calculated by equation (6.35).

Jolliffe (2002) notes that the normalization used in equations (6.34) and (6.35) has the disadvantage that the components and coefficients are less easy to interpret and to compare because each set has a different normalization on its coefficients.

von Storch and Zwiers (2003) mention that the units, which are normally carried by the principal components will be transferred to the renormalized eigenvectors.

6.4.7 Transformation From Principal Components to Original Data

The interrelation between the principal components $\boldsymbol{F}$ and the deviated original data set $\boldsymbol{Y} = (\boldsymbol{X} - \bar{\boldsymbol{X}})$ is given by

$$\boldsymbol{F} = \boldsymbol{Y}\boldsymbol{A} = (\boldsymbol{X} - \bar{\boldsymbol{X}})\boldsymbol{A}, \tag{6.41}$$

where the j^{th} column of $\bar{\boldsymbol{X}}$ consists of the j^{th} column mean $\bar{\boldsymbol{x}}_j$.

Moreover it is already known that

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{F}\boldsymbol{A}' \\ (\boldsymbol{X} - \bar{\boldsymbol{X}}) &= \boldsymbol{F}\boldsymbol{A}' \\ \boldsymbol{X} &= \boldsymbol{F}\boldsymbol{A}' + \bar{\boldsymbol{X}}. \end{aligned} \tag{6.42}$$

If k principal components account for a large amount of variation present in the original data and only k principal components will be taken into following considerations it is possible to calculate the error, arising from the reduction of the data set by

$$\boldsymbol{E}_{(n\times p)} = \boldsymbol{X}_{(n\times p)} - \boldsymbol{X}^*_{(n\times p)} \tag{6.43}$$

with

$$\boldsymbol{X}^*_{(n\times p)} = \boldsymbol{F}^*_{(n\times k)}\boldsymbol{A}^{*'}_{(k\times p)} + \bar{\boldsymbol{X}}_{(n\times p)}. \tag{6.44}$$

$\boldsymbol{F}^*_{(n\times k)}$ only contains the first k principal components and $\boldsymbol{A}^{*'}_{(k\times p)}$ contains the first k eigenvectors of the covariance/correlation matrix[2].

6.5 Summary

The main steps performing a PCA are summarized in Figure 6.1. Following the arrows the way from the original data matrix results in the principal components and factor loadings as well as in the visualization and interpretation of the results.

As can be seen, the technique is mostly based upon the eigenvalues and eigenvectors of the covariance or the correlation matrix, hence the name "eigentechnique" is really appropriate.

The principal component analysis of some atmospheric fields was done in IDL. To check the correctness of the code, some tests, which were derived from the theoretical background were made. The essential formulas were already mentioned in the sections above, but they will be litest below again. The notation is still the same, but in the enumeration one to four, $\boldsymbol{S}$ denotes the covariance matrix as well as the correlation matrix. $\widetilde{\boldsymbol{F}}$ and $\widetilde{\boldsymbol{A}}$ follow from the renormalization by equation (6.34) and equation (6.35), respectively.

[2]Some selection rules determining the first k principal components will be discussed in Section 8.1.

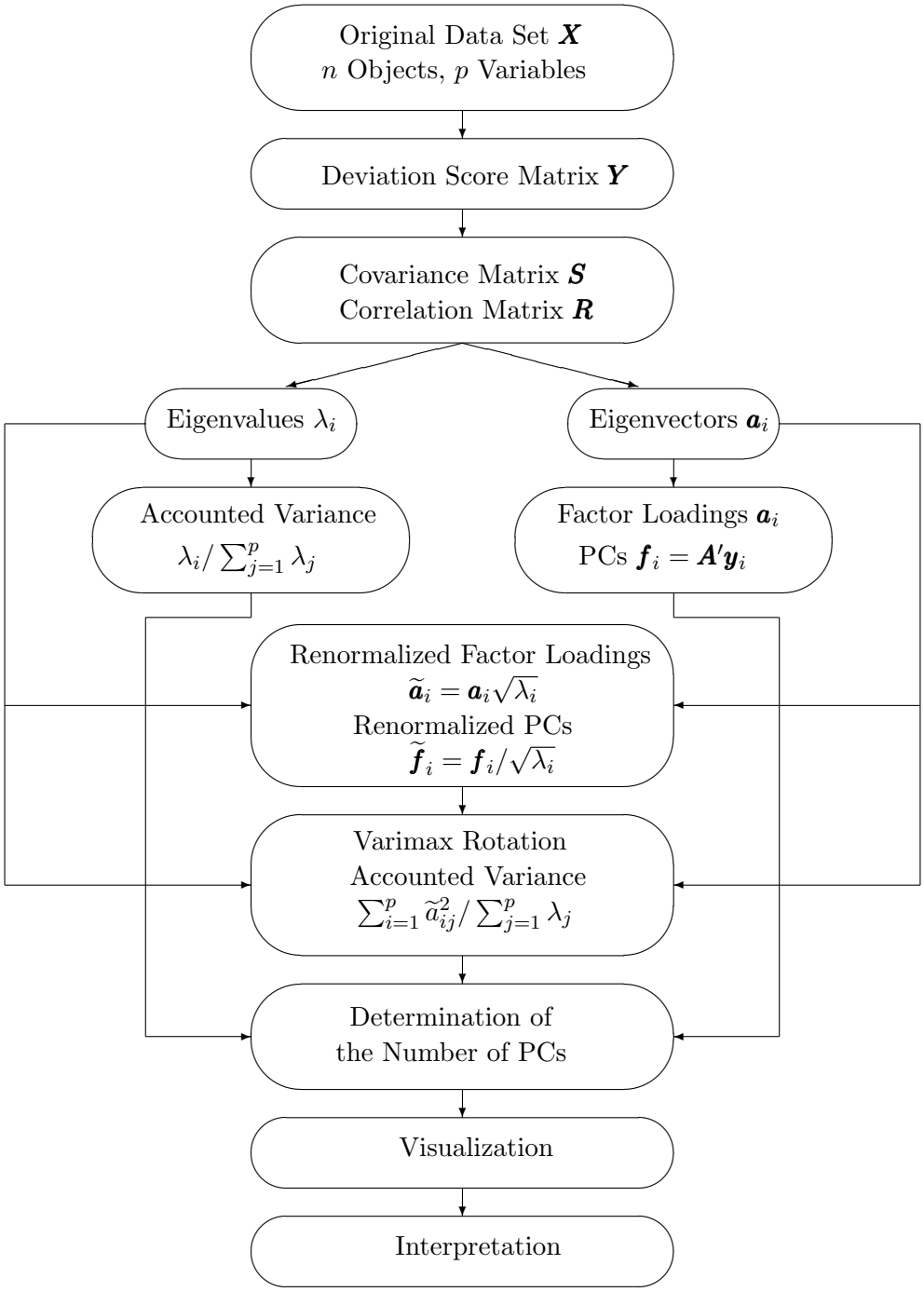


Figure 6.1: The way of calculation processes when performing a PCA. Following the arrows it is possible to understand the way of the principal component analysis. The dotted boxes denote possibilities of calculation, which do not have to be necessarily done. The determination of the number of PCs and the varimax rotation are described in detail in Sections 8.1 and 8.2, respectively.

1. $\boldsymbol{Y} = \boldsymbol{F}\boldsymbol{A}' = \widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{A}}'$,
2. $\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I}$ and $\widetilde{\boldsymbol{A}}'\widetilde{\boldsymbol{A}} = \boldsymbol{\Lambda}$,
3. $\boldsymbol{A}\boldsymbol{A}' = \boldsymbol{I}$ and $\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{A}}' = \boldsymbol{S}$,
4. $\boldsymbol{A}'\boldsymbol{S}\boldsymbol{A} = \boldsymbol{\Lambda}$ and $\widetilde{\boldsymbol{A}}'\boldsymbol{S}\widetilde{\boldsymbol{A}} = \boldsymbol{\Lambda}^2$,

in case of covariance matrices:

5. $(\boldsymbol{F}'\boldsymbol{F})/(n-1) = \boldsymbol{\Lambda}$ and $(\widetilde{\boldsymbol{F}}'\widetilde{\boldsymbol{F}})/(n-1) = \boldsymbol{I}$,
6. $\mathrm{tr}(\boldsymbol{S}) = \mathrm{tr}(\boldsymbol{\Lambda}) = \mathrm{tr}[(\boldsymbol{F}'\boldsymbol{F})/(n-1)]$ and $\mathrm{tr}[(\widetilde{\boldsymbol{F}}'\widetilde{\boldsymbol{F}})/(n-1)] = p$,
7. $\boldsymbol{A}[(\boldsymbol{F}'\boldsymbol{F})/(n-1)]\boldsymbol{A}' = \boldsymbol{S}$ and $\widetilde{\boldsymbol{A}}[(\widetilde{\boldsymbol{F}}'\widetilde{\boldsymbol{F}})/(n-1)]\widetilde{\boldsymbol{A}}' = \boldsymbol{S}$.

6.6 Numerical Results of a Short Example

A short example should demonstrate the results calculated with the principal component analysis.

The numerical outputs calculated with the PCA serve to get a better idea of the method, to check one's own code and to compare the results to the four implemented factor analysis techniques. The eigenvalues, eigenvectors, renormalized coefficients, and the accounted amount of variance will be given from calculations of the covariance matrix as well as from the correlation matrix. A comparison between the results following from these different matrices will be drawn. Afterward the results calculated with varimax rotation will be presented (anticipating the theoretical background of the varimax procedure, which will be discussed in detail in Section 8.2). Later on, the example will be expanded to the selection rules.

The data set stems from Mardia et al. (1979) and represents marks in open-book and closed-book examinations.

Five variables,

- Mechanics (closed-book),
- Vectors (closed-book),
- Algebra (open-book),
- Analysis (open-book), and
- Statistics (open-book)

were collected in 88 observations, hence the analyzed data were given in a (88×5)-matrix.

Before doing the principal component analysis the mean of each variable was subtracted from the original data set.

Sample Covariance Matrix and Sample Correlation Matrix: The calculation of the sample covariance matrix yields

$$\boldsymbol{S} = \begin{pmatrix} 305.768 & 127.223 & 101.579 & 106.273 & 117.405 \\ 127.223 & 172.842 & 85.157 & 94.673 & 99.012 \\ 101.579 & 85.157 & 112.886 & 112.113 & 121.871 \\ 106.273 & 94.673 & 112.113 & 220.380 & 155.536 \\ 117.405 & 99.012 & 121.871 & 155.536 & 297.755 \end{pmatrix}$$

and the sample correlation matrix amounts

$$\boldsymbol{R} = \begin{pmatrix} 1.000 & 0.553 & 0.547 & 0.409 & 0.389 \\ 0.553 & 1.000 & 0.610 & 0.485 & 0.436 \\ 0.547 & 0.610 & 1.000 & 0.711 & 0.665 \\ 0.409 & 0.485 & 0.711 & 1.000 & 0.607 \\ 0.389 & 0.436 & 0.665 & 0.607 & 1.000 \end{pmatrix}.$$

Eigenvalues: As it is shown in Table 6.1, the eigenvalues calculated from the sample covariance matrix $\boldsymbol{S}$ are considerably larger compared to the eigenvalues, which are based on the sample correlation matrix $\boldsymbol{R}$. That is because the correlation matrix follows from standardized variables whose variance is one, hence the sum of the eigenvalues, which is always given by the trace of the matrix equals to the number of variables, which is five in this case.

	λ_1	λ_2	λ_3	λ_4	λ_5
$\boldsymbol{S}$	686.989	202.111	103.747	84.630	32.153
$\boldsymbol{R}$	3.181	0.740	0.445	0.388	0.247

Table 6.1: Eigenvalues of the covariance matrix and the correlation matrix.

Not Rotated Eigenvectors: Comparing the eigenvectors derived from the covariance matrix and the correlation matrix (shown in Table 6.2) it can be noticed that the values of the loadings differ a little bit, but all in all they are similar.

Differences concerning the signs of the eigenvectors can be found in the fifth eigenvector where all elements show opposite signs. The reason is that the calculation of eigenvectors is independent of their direction and that results in an arbitrary sign. When the eigenvectors are of opposite signs, the corresponding principal components have to be of opposite signs too, because the composition of the two parameters will yield an approximation of the actually measured data.

Concerning the fourth eigenvector it can be recognized that the sign of the third element also differs but in both cases it is the element with the lowest loading.

	$a_1^{\text{no rot}}$	$a_2^{\text{no rot}}$	$a_3^{\text{no rot}}$	$a_4^{\text{no rot}}$	$a_5^{\text{no rot}}$
$\boldsymbol{S}$	0.505	0.749	0.300	0.296	0.079
	0.368	0.207	−0.416	−0.783	0.189
	0.346	−0.076	−0.145	−0.003	−0.924
	0.451	−0.301	−0.597	−0.518	0.286
	0.535	−0.548	0.600	−0.176	0.151
$\boldsymbol{R}$	0.400	0.645	0.621	0.146	−0.131
	0.431	0.442	−0.705	−0.298	−0.182
	0.503	−0.129	−0.037	0.109	0.847
	0.457	−0.388	−0.136	0.666	−0.422
	0.438	−0.470	0.313	−0.659	−0.234

Table 6.2: Not rotated eigenvectors of the covariance matrix and the correlation matrix.

The sum of squared elements of each column is always one because of one is the length of all eigenvectors. The sum of squared elements of each row, the communality, is also always one, because, if $k = p$, 100 % of total variation is accounted for.

Figure 6.2 left, displays the five variables diagrammed with respect to their first two factor loadings (eigenvectors) calculated with the covariance matrix. Because of the similarity, the results calculated with the correlation matrix are not shown. Variables marked with straight letters denote not rotated eigenvectors, italic letters display varimax rotated eigenvectors being examined later on in detail.

Because of the first few loadings representing a large amount of intrinsic variation, only a few of them will be interpreted. The number of "most important" eigenvectors can be determined by some selection rules being discussed in Section 8.1. Assuming that two eigenvectors reflect the most prominent patterns only two of them will be interpreted. Mardia et al. (1979) attribute the first eigenvector representing an "average" because all variables have a medium positive weight. The second eigenvector distinguishes between the first two variables, which have positive loadings and the three other variables, which have negative loadings. Remembering the meaning of the variables it is evident that it differentiates between open-book and closed-book examinations.

Principal Components (Factor Scores): The principal components calculated with the covariance matrix and the correlation matrix are given in a (88×5)-matrix, respectively, and the results will not be given in numerical values.

Figure 6.2 right, shows a scatter plot in which all 88 individuals are drawn with respect to their first two principal components, calculated by means of the covariance matrix. Because of the results are nearly identical, those of the correlation matrix are not shown.

Analyzing the forming depicted in the scatter plot it can be seen that the first principal component accounts for more variance than the second principal component.

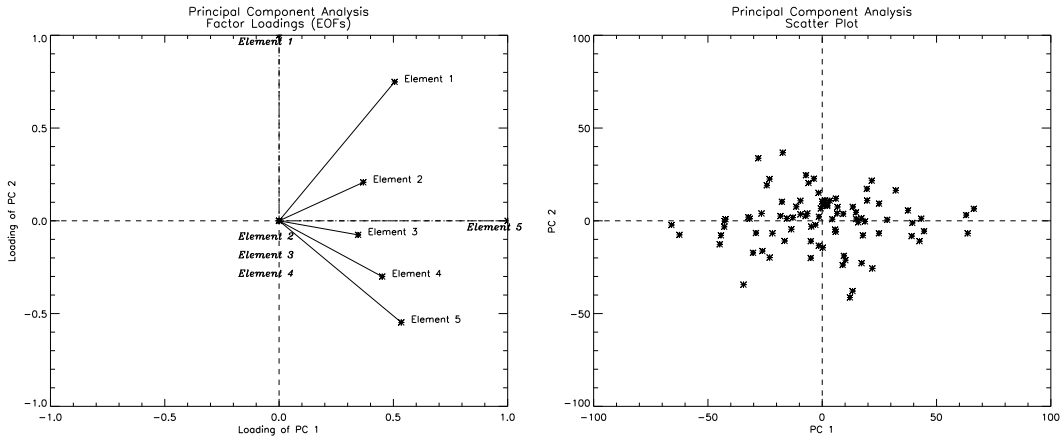


Figure 6.2: Left: Variables depicted with respect to their first and second factor loadings (EOFs) calculated by means of the covariance matrix. Straight letters denote not rotated EOFs, italicized characters denote varimax rotated EOFs. Right: Scatter plot of 88 individuals with respect to their first two principal components also calculated from the covariance matrix.

Renormalized Coefficients: The renormalized coefficients are calculated with equation (6.35), which means that each eigenvector is normalized by the square root of its corresponding eigenvalue. Because of the differences in the magnitude of the eigenvalues, the resulting renormalized coefficients shown in Table 6.3, differ if they are calculated with the covariance matrix or the correlation matrix; only the signs will remain the same.

It can be seen that, due to the renormalization, the magnitude of the columns can not be compared to each other. The interpretation of the first eigenvector as "average" cannot be done with the renormalized coefficients, but the splitting of the first two variables and the last three ones can be seen in the second renormalized coefficient of both matrices.

Percentage of Explained Variance: The percentage of variance explained by each principal component can be calculated from the respective eigenvalue divided by the sum of all eigenvalues or, which is equivalent, by means of the renormalized coefficients whose squared sum of the elements situated in the corresponding column have to be divided by the sum of all eigenvalues (cf., equations (6.17) and (6.40), respectively). The results of the calculations can be seen in Table 6.4.

Comparing the amount of explained variance calculated with the covariance matrix and the correlation matrix it can be noticed that the distribution is similar to each other. The first principal component accounts for a little bit more than 60 % in both cases, the second one about 15 %. Cumulatively the first two PCs explain 80 % (covariance matrix) and 78 % (correlation matrix).

Varimax Rotated Eigenvectors: Anticipating the theoretical background of the varimax rotation the numerical results applied to the example will be given. The only

	$\widetilde{a}_1^{\text{no rot}}$	$\widetilde{a}_2^{\text{no rot}}$	$\widetilde{a}_3^{\text{no rot}}$	$\widetilde{a}_4^{\text{no rot}}$	$\widetilde{a}_5^{\text{no rot}}$
$\boldsymbol{S}$	13.248	10.645	3.054	2.725	0.450
	9.655	2.949	−4.233	−7.202	1.071
	9.060	−1.079	−1.480	−0.030	−5.239
	11.824	−4.278	−6.077	4.767	1.619
	14.013	−7.788	6.114	−1.617	0.858
$\boldsymbol{R}$	0.713	0.555	0.414	0.091	−0.065
	0.769	0.380	−0.470	−0.186	−0.090
	0.898	−0.111	−0.025	0.068	0.420
	0.815	−0.334	−0.091	0.415	−0.210
	0.782	−0.405	0.208	−0.410	−0.116

Table 6.3: Not rotated renormalized coefficients of the covariance matrix and the correlation matrix.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
$\boldsymbol{S}$	61.9115 %	18.2142 %	9.34971 %	7.62689 %	2.89765 %
$\boldsymbol{R}$	63.6196 %	14.7914 %	8.89930 %	7.75785 %	4.93181 %

Table 6.4: Explained variance of the principal components, calculated with the covariance matrix and the correlation matrix.

mention to the technique is that the method contributes to facilitate the interpretation of the intrinsic results. It can be applied to the eigenvectors as well as to the renormalized coefficients whereas in the first case it generates as much near zero values and only a few larger values (lower than one) for each variable.

Table 6.5 represents the varimax rotated eigenvalues if all eigenvectors have been rotated, and if the rotation is only applied to the first and second eigenvector.

Comparing the results yielded from the calculation of the covariance matrix and the correlation matrix it can be noticed that the first two eigenvectors are practically identical, the third and the fourth are interchanged.

Applying the varimax rotation only to the first and second eigenvector changes the results; the clearness of the results is clouded because the varimax rotation is dependent on the number of eigenvectors being rotated. That is one of the disadvantages of rotation techniques. Nevertheless, as can be seen in Table 6.5 for each calculated value always the same elements are the largest. The rotated first eigenvector yields the fifth element being the most important one and the second eigenvector yields the first element being the most pronounced one; independent of which matrix is used.

	$a_1^{\text{var rot}}$	$a_2^{\text{var rot}}$	$a_3^{\text{var rot}}$	$a_4^{\text{var rot}}$	$a_5^{\text{var rot}}$
$\boldsymbol{S}$	8.813e–08	1.000	3.347e–09	−3.424e–09	−6.487e–09
	−3.325e–09	3.424e–09	2.610e–08	−1.000	−2.212e–09
	−2.151e–09	6.487e–09	−1.343e–08	−2.177e–09	−1.000
	−1.219e–08	−3.402e–09	−1.000	2.610e–08	−1.245e–08
	1.000	−9.284e–08	1.229e–08	3.326e–09	2.151e–09
$\boldsymbol{S}$	−0.0960	0.898			
	0.148	0.396			
	0.313	0.165			
	0.539	0.061			
	0.762	−0.074			
$\boldsymbol{R}$	1.619e–07	1.000	6.356e–08	−8.641e–09	7.477e–10
	3.251e–09	−6.348e–08	−1.000	2.807e–09	−2.532e–08
	1.903e–09	7.478e–10	9.373e–09	−3.002e–09	1.000
	−3.607e–09	−8.641e–09	−2.807e–09	1.000	−3.260e–09
	1.000	−1.794e–07	−2.858e–09	−3.608e–09	1.901e–09
$\boldsymbol{R}$	−0.075	0.755			
	0.075	0.613			
	0.478	0.203			
	0.599	−0.031			
	0.634	−0.108			

Table 6.5: Varimax rotated eigenvectors of the covariance matrix and the correlation matrix. Varimax rotation is operated to all five eigenvectors as well as to only the first two eigenvectors.

Varimax Rotated Renormalized Coefficients: Because the coefficients are not normalized to unit length but to the lengths of the respective eigenvalues the rotation causes a shifting to values other than zero or one, depending on the magnitude of the eigenvalues. Nevertheless, the rotation will yield distinctive differences between the values of each variable.

Table 6.6 shows the varimax rotated renormalized coefficients for all as well as only two rotated coefficients. Again, the dependency of the number of rotated elements on the varimax rotation can be seen.

Percentage of Explained Variance of Rotated Eigenvectors/Renormalized Coefficients: The percentage of variance accounted for by rotated eigenvectors can only be calculated by means of the rotated renormalized coefficients. Again the squared sum of one column will be divided by the sum of the eigenvalues.

As can be seen in Table 6.7 the amount of variance is more evenly distributed on

	$\widetilde{a}_1^{\text{var rot}}$	$\widetilde{a}_2^{\text{var rot}}$	$\widetilde{a}_3^{\text{var rot}}$	$\widetilde{a}_4^{\text{var rot}}$	$\widetilde{a}_5^{\text{var rot}}$
$\boldsymbol{S}$	2.504	16.293	−2.535	−4.286	−3.040
	2.178	3.447	−2.467	−11.960	−2.662
	3.547	2.745	−3.814	−3.049	−8.303
	4.170	2.351	−13.210	−2.906	−3.810
	15.689	2.593	−4.591	−2.845	−3.966
$\boldsymbol{S}$	3.268	16.678			
	5.481	8.477			
	7.623	5.014			
	11.797	4.351			
	15.734	3.078			
$\boldsymbol{R}$	0.143	0.932	−0.245	0.145	0.174
	0.166	0.262	−0.910	0.188	0.202
	0.334	0.258	−0.287	0.359	0.781
	0.281	0.158	−0.196	0.890	0.257
	0.909	0.150	−0.165	0.266	0.230
$\boldsymbol{R}$	0.194	0.882			
	0.350	0.784			
	0.762	0.488			
	0.840	0.264			
	0.860	0.188			

Table 6.6: Varimax rotated renormalized coefficients of the covariance matrix and the correlation matrix. Varimax rotation is applied to all five renormalized coefficients and to the first and second renormalized coefficient.

the five principal components as were the unrotated principal components. The rotated variables together account for the same amount of intrinsic variation, but no longer gradually account for the maximum possible variation.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
$\boldsymbol{S}$	25.875 %	26.777 %	20.064 %	16.875 %	10.409 %
$\boldsymbol{R}$	21.298 %	21.027 %	20.711 %	20.954 %	16.011 %

Table 6.7: Explained variance of the varimax rotated eigenvectors, calculated with the covariance matrix and the correlation matrix.

7 Factor Analysis

(Author: B.C. Lackner)

One year ago, in 2004, the 100^{th} anniversary of factor analysis was celebrated on the conference "Factor Analysis at 100".

Factor analysis (short FA) as such was "invented" by a psychologist named Charles Spearman in 1904. Until the middle of 1930s, the mathematical technique caused no problems, since only few factors with a priory known structure were assumed and used for interpretations (Kaplunovsky 2004). During the 1930s, Thurstone came up with a "center of gravity" method for the estimation of loadings, which led in the end to the centroid method, which will be presented in Section 7.2.4. Less than ten years later, factor analysis was first put on statistical footing.

Nearly at the same time, principal component analysis was introduced by Hotelling, competing in a certain kind with the "older" factor analysis, as the differences and aims of both methods were not clearly defined and the discourse about this topic lasted for several years (sometimes it seems, it is continuing to this day). In the 1960s, Lawley and Maxwell (1971) presented a common factor model as a statistical model and until today, quite a variety of methods, being based on different hypothesis and mathematical models, are available.

This chapter will deal with the mathematical model "factor analysis". As PCA, factor analysis tries to explain the correlation between a large set of variables in terms of a small number of underlying *factors*, which cannot be observed directly. This implies that when a researcher obtains a measurement for a certain variable, this measurement is the result of the influences of underlying factors, which are assumed to be linear. According to Tucker and MacCallum (1997), the factor model implies two parts of factors:

- Common factor: This part of the factors affects more than one of the observed variables.
- Specific factor: This is a factor on which only one variable has an impact.

In addition to these two types of factors, the observed variables are as well influence by an error of measurement, which is included in the model as a supplementary factor. The components of this "unobserved" error part are assumed to be uncorrelated.

There is a direct relation between the errors of a measurement and the reliability of the variables: Low measurement errors will result in a high reliability. Measurement error and specific factors are united to "unique factors". The common and specific factors together stand for the "reliable" ones.

The distinction of different types of factors (common, specific and error) helps to explain, how the factors account for the variance of the observed variables, which is as well given by three parts.

The part of variance, which is, in general, of highest interest, is given by the variance of a variable shared with common factors and is named "common variance" or communality. A second part arises from the specific factors and is therefore labeled "specific variance" or specificity. The third and last part of the variance is connected with the errors of measurement and is termed "error variance". The latter is frequently combined with the specific variance and together they form the so called "unique variance" or uniqueness (compare Figure 7.1). As the common variance is given by the communality h^2, the uniqueness of a variable is $(1-h^2)$, as it is the proportion of a variables' variance that is not shared with a factor structure.

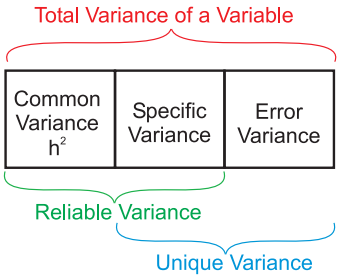


Figure 7.1: Schematic representation of different variances (according to Pohlmann (2005)).

Thus, factor analysis can be considered as a method explaining the covariation between observed variables. Usually, observed variables tend to be correlated with each other to varying degrees and according to factor analysis, these correlations are given due to the influence of common factors as different observed variables may be effected by the same factor (Tucker and MacCallum 1997). In other words, only the common factors account for the correlations in the observed variables, while specific factors and measurement errors are able to influence nothing but one single variable.

7.1 The Mathematical Model

To put the preceding consideration into a mathematical model, we will start from a data vector $\boldsymbol{x}$ with mean μ and covariance $\boldsymbol{\Sigma}$. Then the factor model is defined as follows:

$$\boldsymbol{x} - \mu = \boldsymbol{A}\boldsymbol{f} + \boldsymbol{u}. \tag{7.1}$$

In this equation, $\boldsymbol{A}$ is the $(k \times p)$ matrix of factor weights, called loadings, $\boldsymbol{f}$, the column vector of one factor (with k components), stands for the common factors, and the p elements of the column vector $\boldsymbol{u}$, the unique factors, contain specific and error factors. All factors are uncorrelated among each other and the common factors are each standardized (mean of zero and variance of one).

The **factor loadings** $\boldsymbol{A}$ are the analogue to the renormalized coefficients in principal component analysis. The items of the matrix are the correlation coefficients between the variables and the factors. The variance in all variables accounted for by each factor is

given by summing up the squared factor loadings for one factor and dividing the result by the number of variables (p). These variances show the relative importance of the different factors in explaining the variance of the given data.

It is possible for a loading a_{ii} to exceed the value one. Then, ψ_{ii} will be negative, which is an unacceptable solution, as ψ_{ii} is a variance.

The sum of squared factor loadings for all factors results in the variance of one variable accounted for by all factors, the **communality** h^2. These squared multiple correlations may be interpreted as the reliability of the indicator, but their values must always be considered in relation to the interpretability of the factors. Furthermore, a communality larger than one stands for an uncorrect solution, which may be based on a too small sample size or too many/few selected factors.

The covariance matrix of $\boldsymbol{u}$ is the diagonal matrix $\boldsymbol{\Psi} = \text{diag}(\psi_{11}, \ldots, \psi_{pp})$. The covariance matrix $\boldsymbol{\Sigma}$ of the variables $\boldsymbol{x}_i$ is defined as $E(\boldsymbol{x}\boldsymbol{x}')$ and contains the covariances σ_{ij} of $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, and the variances σ_{ii}. The variances consist of two parts (see equation (7.2)), where the first part of the variance, $\sum_{j=1}^{k}(a_{ij})^2$, stands for the communality h_i^2, the second part, ψ_{ii}, is the unique variance as described in the introduction to this chapter.

$$\sigma_{ii} = \sum_{j=1}^{k}(a_{ij})^2 + \psi_{ii} \tag{7.2}$$

Combining $\boldsymbol{\Sigma} = E(\boldsymbol{x}\boldsymbol{x}')$ with equation (7.1) yields:

$$E(\boldsymbol{x}\boldsymbol{x}') = E\left[(\boldsymbol{A}\boldsymbol{f} + \boldsymbol{u})(\boldsymbol{A}\boldsymbol{f} + \boldsymbol{u})'\right] \tag{7.3}$$

$$E(\boldsymbol{x}\boldsymbol{x}') = E\left[\left(\boldsymbol{A}\boldsymbol{f}\boldsymbol{f}'\boldsymbol{A}' + \boldsymbol{A}\boldsymbol{f}\boldsymbol{u}' + \boldsymbol{u}\boldsymbol{f}'\boldsymbol{A}' + \boldsymbol{u}\boldsymbol{u}'\right)\right]. \tag{7.4}$$

As common and unique factors are assumed to be uncorrelated, the expected value $E(\boldsymbol{f}\boldsymbol{u}') = 0$ (then also $E(\boldsymbol{u}\boldsymbol{f}') = 0$), the second, and third term of equation (7.4) will become zero. Taking additionally into consideration that $E(\boldsymbol{f}\boldsymbol{f}') = \boldsymbol{I}_p$ and $E(\boldsymbol{u}\boldsymbol{u}') = \boldsymbol{\Psi}$, equation (7.4) results in

$$\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}' + \boldsymbol{\Psi}. \tag{7.5}$$

In factor analysis $\boldsymbol{A}$ and $\boldsymbol{\Psi}$ are unknown and have to be estimated from the data $\boldsymbol{X}$. FA' task is to find out, whether for a specified value of $k < p$ an unique $\boldsymbol{\Psi}$ with positive diagonal values and an unique $(p \times p)$ matrix $\boldsymbol{A}$, which satisfies equation (7.5), can be defined (Lawley and Maxwell 1971).

To determine, whether equation (7.5) is fulfilled for a given k, the difference s between the number of distinct elements of $\boldsymbol{\Sigma}$, $(\frac{1}{2}p(p+1))$, and the number of free parameters of $\boldsymbol{A}$ and $\boldsymbol{\Psi}$, $(p + pk - \frac{1}{2}k(k-1))$, is considered (under the condition that $k > 1$ and

$\boldsymbol{A}'\boldsymbol{\Psi}^{-1}\boldsymbol{A} = \text{diagonal}$):

$$s = \frac{1}{2}p(p+1) - \left[p + pk - \frac{1}{2}k(k-1)\right] \tag{7.6}$$

$$= \frac{1}{2}(p-k)^2 - \frac{1}{2}(p+k). \tag{7.7}$$

Depending on the value of s, three cases can occur (Mardia et al. 1979):

1. $s < 0$: In this case, the model contains more parameters than equations and an infinity of exact solutions for A and $\boldsymbol{\Psi}$ can be found, resulting in the fact that the model is not well-defined.

2. $s = 0$: The factor model contains as many parameters as $\boldsymbol{\Sigma}$ and offers no simplification.

3. $s > 0$: There are more equations than parameters and therefore it is not possible to solve the model exactly. So, the aim is to look for an approximate solution to receive simpler explanations for the data by the factor model. This is usually the case.

7.1.1 Factor Scores

The different factor analysis techniques, which will be discussed in Section 7.2, result in an estimation for $\boldsymbol{A}$ and $\boldsymbol{\Psi}$. In finding a solution for $\boldsymbol{A}$, many mathematical constraints are taken so that the factor matrix $\boldsymbol{F}$, containing the "new", hypothesized variables, drops out of the equation (cf., equations (7.1) and (7.5)).

In practice, $\boldsymbol{A}$, $\boldsymbol{\Psi}$, and μ, are not known in advance but estimated from the same data the factor scores are in demand. Even though it would be attractive to estimate the factor scores, loadings, and unique variances at the same time from the data, this is not possible as there are too many parameters (Mardia et al. 1979).

In literature many methods to estimate factor scores can be found. Here, only two of them will be presented in the following. Factor scores are the factor analysis analogue to PCA' principal components.

Bartlett's Weighted Least Square Estimator Method

The unknown factor scores are treated as parameters to be estimated. Starting from the factor model (equation (7.1)), $\boldsymbol{x}$ is supposed to be an observation from a distribution with mean $\boldsymbol{Af}$, covariance matrix $\boldsymbol{\Psi}$; $\boldsymbol{A}$ and $\boldsymbol{\Psi}$ are known (Anderson 1984). Furthermore, $\mu = 0$. The log likelihood of this distribution is given by

$$L(\boldsymbol{x};\boldsymbol{f}) = -\frac{1}{2}\left(\boldsymbol{x} - \boldsymbol{Af}\right)'\boldsymbol{\Psi}^{-1}\left(\boldsymbol{x} - \boldsymbol{Af}\right) - \frac{1}{2}\log\left|2\pi\boldsymbol{\Psi}\right|. \tag{7.8}$$

The first derivative of L with respect to $\boldsymbol{f}$ is set equal to zero and reshaped to obtain an estimate $\hat{\boldsymbol{f}}$ (Mardia et al. 1979).

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{f}} &= 0 && (7.9)\\ \boldsymbol{A}'\boldsymbol{\Psi}^{-1}(\boldsymbol{x}-\boldsymbol{A}\boldsymbol{f}) &= 0 && (7.10)\\ \boldsymbol{A}'\boldsymbol{\Psi}^{-1}\boldsymbol{x}-\boldsymbol{A}'\boldsymbol{\Psi}^{-1}\boldsymbol{A}\hat{\boldsymbol{f}} &= 0 && (7.11)\\ \boldsymbol{A}'\boldsymbol{\Psi}^{-1}\boldsymbol{A}\hat{\boldsymbol{f}} &= \boldsymbol{A}'\boldsymbol{\Psi}^{-1}\boldsymbol{x} && (7.12)\\ \hat{\boldsymbol{f}} &= \left(\boldsymbol{A}'\boldsymbol{\Psi}^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}'\boldsymbol{\Psi}^{-1}\boldsymbol{x} && (7.13) \end{aligned}$$

In equation (7.13) the term $\left(\boldsymbol{A}'\boldsymbol{\Psi}^{-1}\boldsymbol{A}\right)$ is not necessarily diagonal. According to Mardia et al. (1979), Bartlett's factor scores have the favorable property of being an unbiased estimate. However, the estimated factor scores are not necessarily uncorrelated, even if the factor loadings are orthogonally rotated (Hu n.y.).

Knowing the factor scores and the factor loadings, the specific factor scores $\hat{\boldsymbol{u}}$ can be estimated by $\hat{\boldsymbol{u}} = \boldsymbol{x} - \boldsymbol{A}\boldsymbol{f}$.

Anderson-Rubin Method

Anderson and Rubin modified the routine to calculate Bartlett's factor scores in a way that the derived factors are uncorrelated (Hu n.y.). The estimates for the factors $\boldsymbol{f}^*$ are given through:

$$\boldsymbol{f}^* = \left[\left(\boldsymbol{A}'\boldsymbol{\Psi}^{-1}\boldsymbol{A}\right)\left(\boldsymbol{I}+\boldsymbol{A}'\boldsymbol{\Psi}^{-1}\boldsymbol{A}\right)\right]^{-1/2}\boldsymbol{A}'\boldsymbol{\Psi}^{-1}(\boldsymbol{x}-\mu) \tag{7.14}$$

7.2 Description of the Four Implemented FA-Techniques

Factor analysis, unlike PCA, is unaffected by rescaling of variables. Since generally in practice rather the relationship between the variables is of interest than their scaling, data are usually summarized by the sample correlation matrix $\boldsymbol{R}$ and not by the (estimated) covariance matrix $\boldsymbol{S}$ (Mardia et al. 1979). The model (cf., equation (7.5)) then becomes:

$$\boldsymbol{R} = \boldsymbol{A}\boldsymbol{A}' + \boldsymbol{\Psi}. \tag{7.15}$$

In this section, a description of four different factor analysis techniques for extracting factors from a data set is given. Two of them are algebraic methods, one is a statistical and one a geometrical one.

7.2.1 Iterative Principal Factor Analysis According to Mardia (PFA)

Principal factor analysis, also known as "principal axis factoring" or "common factor analysis", is one of the most commonly applied techniques of factor analysis. It proceeds very much like principal component analysis, even though there are subtle differences.

A detailed treatise of the calculation procedure described in the following can be found by Mardia et al. (1979).

Principal factor analysis can be applied for a k-factor model under the constraint that s (cf., equation (7.6)) is positive. At the outset, the correlation matrix $\boldsymbol{R}$ is needed to estimate $\boldsymbol{A}$ and $\boldsymbol{\Psi}$ (thought for standardized variables) in equation (7.15).

In a first step, the communalities $\hat{h}_i^2, (i = 0, \ldots, p)$ are estimated. For this, two methods are described in literature:

1. To derive the i^{th} communality, the square of the multiple correlation coefficient of the i^{th} variable with all other variables is calculated.

2. The first estimate for the communality is given by the largest correlation coefficient between the i^{th} variable and one of the other variables: $\hat{h}_i^2 = \max_{i \neq j} |r_{ij}|$. This estimation method for $\hat{h}_i^2$ was used within the scope of this work.

As in principal component analysis, where the eigenvalues and eigenvectors of a covariance or correlation matrix are calculated, a similar procedure is used in principal factor analysis but, in contrast to the covariance or correlation matrix, a so called **reduced correlation matrix**, given by $\boldsymbol{R}^{\text{reduced}} = \boldsymbol{R}^{\text{original}} - \boldsymbol{\Psi}$, is employed. The ones in the main diagonal of $\boldsymbol{R}^{\text{original}}$ are replaced by the estimated communalities $\hat{h}_i^2$, as $\hat{h}_i^2 = 1 - \hat{\psi}_{ii}$. In the following, the eigenvectors $\boldsymbol{u}_i$ and the corresponding eigenvalues λ_i (which have to be sorted according to their magnitude) of the reduced correlation matrix $\boldsymbol{R}^{\text{reduced}}$ are used to estimate the factor loading matrix $\hat{\boldsymbol{A}}$ applying the spectral decomposition theorem:

$$\hat{\boldsymbol{A}} = \boldsymbol{U}\Lambda^{1/2}, \tag{7.16}$$

where the matrix $\boldsymbol{U}$ contains the eigenvectors of $\boldsymbol{R}^{\text{reduced}}$ (each eigenvector stands in one column) and $\Lambda^{1/2}$ is a diagonal matrix with the roots of the eigenvalues in the main diagonal.

Iterative Process

The estimates obtained in this way are systematically biased (Reyment and Jöreskog 1993). Therefore, to improve the results of principal factor analysis as described above, an iterative process is recommended. Starting with the first results for factor loadings (cf., equation (7.16)), improved values for the communalities are gained by

$$\hat{\psi}_{ii} = 1 - \sum_{j=1}^{k} \hat{a}_{ij}^2, \qquad (i = 0, \ldots, p) \tag{7.17}$$

$$\hat{h}_i^2 = 1 - \hat{\psi}_{ii}, \tag{7.18}$$

leading to a new reduced correlation matrix, which can again be decomposed. These steps are repeated until stable results or a desired accuracy is given.

A principal factor analysis solution is permissible, if all $\hat{\psi}_{ii}$ are non-negative (a negative ψ_{ii} is an unacceptable solution, for ψ_{ii} is a variance). Negative ψ_{ii} can be caused by

a too small sample size or too many/few selected factors. Since the sample size is generally a given quantity, only the number of selected factors can be changed to achieve a permissible solution.

7.2.2 True Factor Analysis According to Jöreskog (True FA)

In the 1960s, Jöreskog succeeded in developing a remarkable factor model (see Reyment and Jöreskog (1993)). The derived equation system of this model, even though it is founded on statistics, yields a direct solution without having to apply an iterative routine. There is as well no need to estimate starting values so that, in general, the results are reached quicker than with other methods. Jöreskog formulated several factor analysis solutions, but in the context of this work, only a scale-free method, named "true factor analysis", is considered.

Starting again from equation (7.15), one supposes that $\boldsymbol{\Psi}$ is known, so that

$$\boldsymbol{R} - \boldsymbol{\Psi} \approx \boldsymbol{A}\boldsymbol{A}'. \tag{7.19}$$

Then, the columns of $\boldsymbol{A}$ are chosen as eigenvectors of $\boldsymbol{R} - \boldsymbol{\Psi}$ (corresponding to the k largest eigenvalues), so that the sum of squares in each column equals the corresponding eigenvalue (Reyment and Jöreskog 1993).

As discussed in Section 7.2.1, there are several methods to estimate the communalities $\hat{h}_i^2$, which also lead to $\boldsymbol{\Psi}$. Reyment and Jöreskog (1993) state that it has been shown that the squared multiple correlation coefficient provides the best possible lower bound for the communalities, and that such an estimate of $\boldsymbol{\Psi}$ results from

$$\hat{\boldsymbol{\Psi}} = \frac{1}{(\text{diag } \boldsymbol{R}^{-1})}. \tag{7.20}$$

As the estimate of $\boldsymbol{\Psi}$ in equation (7.20) is systematically biased (too large), Jöreskog proposed to multiply it with a scalar θ, which is less than one and has to be estimated from the data as well.

$$\hat{\boldsymbol{\Psi}} = \theta \ \frac{1}{(\text{diag } \boldsymbol{R}^{-1})} \tag{7.21}$$

To obtain Jöreskog's scale free estimation of $\boldsymbol{A}$ and $\boldsymbol{\Psi}$, equation (7.19) is pre- and postmultiplied by $\boldsymbol{\Psi}^{-1/2}$:

$$\boldsymbol{\Psi}^{-1/2}\boldsymbol{R}\boldsymbol{\Psi}^{-1/2} - \boldsymbol{I}_p \ \approx \ \boldsymbol{\Psi}^{-1/2}\boldsymbol{A}\boldsymbol{A}'\boldsymbol{\Psi}^{-1/2}. \tag{7.22}$$

The eigenvalues and eigenvectors of the left hand side of the equation lead to an estimate for $\boldsymbol{A}$. Using as well equation (7.20), the eigenvalues and eigenvectors of

$$(\text{diag } \boldsymbol{R}^{-1})^{1/2} \ \boldsymbol{R} \ (\text{diag } \boldsymbol{R}^{-1})^{1/2} - \boldsymbol{I}_p \tag{7.23}$$

have to be found.

Hence, to estimate θ and the factor loadings $\boldsymbol{A}$, the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ and the corresponding eigenvectors, which are summarized in the $(p \times p)$ matrix $\boldsymbol{U}$, of the auxiliary matrix $\boldsymbol{R}^*$ are calculated. This matrix is scale-invariant, so it does not matter whether to employ the covariance or correlation matrix.

$$\boldsymbol{R}^* = (\text{diag } \boldsymbol{R}^{-1})^{1/2} \, \boldsymbol{R} \, (\text{diag } \boldsymbol{R}^{-1})^{1/2} \tag{7.24}$$

Selecting k factors, the least square estimate of $\hat{\theta}$ is the average of the $(p-k)$ smallest eigenvalues of this auxiliary matrix $\boldsymbol{R}^*$. To determine the number of factors, on the one hand the eigenvalues of $\boldsymbol{R}^*$ can be enlisted (or visualized with a scree plot), on the other hand it has to be ensured that $\hat{\theta}$ is less than one.

The least square estimates of the factor loadings $\boldsymbol{A}$ are given by means of the sample correlation matrix, the eigenvectors (in $\boldsymbol{U}_k$), the eigenvalues (in Λ_k), and the weighting factor $\hat{\theta}$; see equation (7.26).

$$\hat{\theta} = \frac{1}{p-k} \sum_{m=k+1}^{p} \lambda_m \tag{7.25}$$

$$\hat{\boldsymbol{A}} = (\text{diag } \boldsymbol{R}^{-1})^{1/2} \, \boldsymbol{U}_k \, \left(\boldsymbol{\Lambda}_k - \hat{\theta} \boldsymbol{I}_k \right)^{1/2} \tag{7.26}$$

7.2.3 Maximum Likelihood Factor Analysis (ML-FA)

In the 1940s, factor analysis was put on "statistical footing" by Lawley, implementing maximum likelihood estimations. However, as late as the mid 1960s, there was no good way for computing the estimates. The advantage of this method is that there are tests available to check, whether the model fits the data well. The drawback of the method shows up in a slowly converging iterative process, so that it is recommended to use the results of another method as starting values (in the context of this work, the results of principal factor analysis were used). Anyhow, a multitude of algorithms can be found in literature to improve the calculation procedure.

Unlike principal, true, and centroid factor analysis, which have no distributional assumptions, maximum likelihood factor analysis expects the data to be multi-normally distributed (Weber 1974). A normal, or Gaussian, distribution is characterized by two parameters, the mean of the variables and the variance. This probably most known distribution is unimodal and symmetric about the mean. The multivariate normal, or spherical, distribution, is a generalization of the normal. The first two moments (mean and variance) are complemented by two further moments, namely skewness and kurtosis, which jointly describe the multi-normal distribution. Furthermore, it should be noted that the multivariate normal distribution is not a mere composite of univariate normal distributions and that even if every variable in a set is normally distributed, it is still possible that the combined distribution is not multivariate normal (Rigdon 1996).

To estimate the factor loadings $\boldsymbol{A}$, one starts from the covariance matrix $\boldsymbol{\Sigma}$, using the information given by the covariance matrix $\boldsymbol{S}$ of the sample $\boldsymbol{X}$. As the sample is

expected to be a part of a multi-normal population, the distribution function for the elements of the covariance matrix can be formulated. This distribution was derived by Wishart and therefore is named "Wishart distribution". The Wishart distribution is the multivariate analogue to the chi-square distribution[1] and is related to the multivariate normal in the same way the chi-square is related to the univariate normal (Rigdon 1996). The Wishart distribution is given through:

$$f(\boldsymbol{S}) = c\,|\boldsymbol{\Sigma}|^{-\frac{1}{2}n}\;|\boldsymbol{S}|^{\frac{1}{2}(n-p-1)}\;\exp\left[-\frac{n}{2}\,\mathrm{trace}\left(\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\right)\right], \tag{7.27}$$

with c as a constant depending on the sample size.

The likelihood of a given covariance matrix $\boldsymbol{S}$ is taken equal to the value of the density function for that observed covariance matrix. Hence, the likelihood, named with L of an observed $\boldsymbol{S}$ is given by $L = f(\boldsymbol{S})$.

For further considerations, a logarithmic conversion, the so called "log-likelihood", is commonly investigated.

$$\ln(L) = \ln(c) - \frac{1}{2}n\ln|\boldsymbol{\Sigma}| + \frac{1}{2}(n-p-1)\ln|\boldsymbol{S}| - \frac{n}{2}\,\mathrm{trace}\left(\boldsymbol{S}\,\boldsymbol{\Sigma}^{-1}\right) \tag{7.28}$$

This expression is to be maximized according to the loadings a and unique variances ψ (given by $\boldsymbol{\Sigma}$ and $\boldsymbol{S}$). As the log-likelihood is an increasing monotonic function of the likelihood L, the maximum of $\ln(L)$ occurs with the maximum of L. Hence, maximizing the log-likelihood maximizes the likelihood (Tucker and MacCallum 1997). For that, the first derivative has to be equated to zero.

While early computing procedures were made up of alternating iterations between solutions for the factor loading matrix $\boldsymbol{A}$ and the unique variance matrix $\boldsymbol{\Psi}$, more effective methods, like the "gradient" or "Newton-Raphson" method are used nowadays, to achieve a maximum likelihood solution.

Expectation Maximization Algorithm According to Nielsen

The EM (expectation-maximization) algorithm is a possible computational device for maximum likelihood estimations. The idea is to treat the unobservable common factors $\boldsymbol{f}$ as missing data under the assumption that $\boldsymbol{f}$ and $\boldsymbol{\Psi}$ have a joint normal distribution (Anderson 1984). The advantage of the EM algorithm is, among other things, that it converges fast.

The EM algorithm used in this work is taken by Nielsen (2004). In the E-step of the algorithm, the expectation of the covariances is calculated on the basis of trial values of the parameters $\boldsymbol{A}$ and $\boldsymbol{\Psi}$. For the first step, the results for $\boldsymbol{A}$ and $\boldsymbol{\Psi}$ from principal factor

[1] The chi-square (χ^2) distribution is an univariate distribution resulting when univariate-normal variables are squared and possibly summed. The distribution is squared to the right and has a minimum of zero, the mean is equal to its degrees of freedom and the variance is equal to twice the degrees of freedom, on condition that the mean of the original variables is zero (Rigdon 1996).

analysis were employed. Three equations have to be calculated within the E-step:

$$\boldsymbol{\Sigma_F} = \left(\boldsymbol{I}_k + \boldsymbol{A}' \boldsymbol{\Psi}^{-1} \boldsymbol{A}\right)^{-1} \tag{7.29}$$

$$E(\boldsymbol{F}) = \boldsymbol{\Sigma_F} \boldsymbol{A}' \boldsymbol{\Psi}^{-1} \boldsymbol{Z} \tag{7.30}$$

$$E\left(\boldsymbol{F}\boldsymbol{F}'\right) = n\boldsymbol{\Sigma_F} + E(\boldsymbol{F})\ E\left(\boldsymbol{F}'\right) \tag{7.31}$$

$\boldsymbol{\Sigma_F}$ is an estimate for a covariance matrix of the factor scores; the matrix $\boldsymbol{Z}$, first mentioned in equation (7.30), stands for the standardized variables of $\boldsymbol{X}$; n is the number of objects (months in case of the investigated atmospheric data sets).

During the M-step, as the name already indicates, the likelihood function is maximized on the basis of the results of the E-step. Updated values of the parameters $\boldsymbol{A}$ and $\boldsymbol{\Psi}$ are derived. The two equations belonging to this step are:

$$\boldsymbol{A} = \boldsymbol{Z}\, E\left(\boldsymbol{F}'\right)\ E\left(\boldsymbol{F}\,\boldsymbol{F}'\right)^{-1} \tag{7.32}$$

$$\boldsymbol{\Psi} = \frac{1}{n} \operatorname{diag}\left(\boldsymbol{Z}\boldsymbol{Z}' - \boldsymbol{A}\, E(\boldsymbol{F})\ \boldsymbol{Z}'\right). \tag{7.33}$$

The two steps alternate and the procedure usually converges to the maximum likelihood estimators. In the example discussed in Section 7.3 the algorithm was continued until the maximal difference of any element of $\boldsymbol{A}$ or $\boldsymbol{\Psi}$ from one M-step to another was less than 0.05.

7.2.4 Centroid Factor Analysis (Centroid-FA)

The centroid factor analysis is the oldest method of factor analysis and goes back to Thurstone, who introduced it in the 1930s. Hence, this method was developed prior to the advent of large scale computers and at those days, it was very successful, because it is based on simple summations, which can be easily carried out with a desk calculator. Nevertheless, the method shows quite a few weak spots, like a certain arbitrariness (which shows up in the signs of the loadings) and it depends also, in contrast to the other methods described, on the scale of the measured values (using the covariance matrix leads to other results than using the correlation matrix). The here presented calculation procedure follows the description of Weber (1974).

The centroid method developed by Thurstone is a geometric method. The p variables can be presented as radial vectors starting from the origin of an orthogonal coordinate system. The idea of the geometric formulation is, to define a new coordinate system with the same origin as the primary one, whose abscissa passes through the radial vectors end points' center of gravity and therefore is called center of gravity axis. The standardized vector along this axis corresponds to the first factor (which is also named first "centroid" in the case of centroid factor analysis).

For an algebraic solution, the factors are extracted one by one, always starting again with a new residual correlation matrix to calculate the next factor. The procedure continues, until the final residuals are small enough so that the resulting factors only have

small absolute factor weights.

The pure algebraic solution of Thurstone's geometrical model is based on summation and some other basic arithmetical operations as described in the following.

Extraction of the First Factor

In a first step, as in principal factor analysis, the ones in the main diagonal of the correlation matrix have to be replaced by the estimated communalities $\hat{h}_i^2$, which are the absolutely highest correlation coefficients of the corresponding row/column of the correlation matrix. This reduced correlation matrix is the starting point for further calculations.

To achieve the values of the factor loadings for the first factor, the sum t_i of each column (or row) of this reduced matrix is calculated.

$$t_i = \sum_{j=1}^{p} r_{ji}^{\text{reduced}}, \qquad i = 1, \ldots, p \tag{7.34}$$

To continue, one has to derive the root of the total of these sums, called T.

$$T = \sqrt{\sum_{i=1}^{p} t_i} \tag{7.35}$$

Out of these sums, the factor loadings a_{i1} for the first factor are achieved by

$$\boldsymbol{a}_1 = \frac{t_i}{T}, \qquad i = 1, \ldots, p. \tag{7.36}$$

To continue, a $(p \times p)$ matrix, given by $\boldsymbol{aa}'$ leads to the first residual correlation matrix $\boldsymbol{R}_1^{\text{reduced}} = \boldsymbol{R} - \boldsymbol{a}_1\boldsymbol{a}_1'$, which is the starting point for the extraction of the second factor.

Extraction of the Following Factors

The sum of the rows and columns of the residual correlation matrix are approximately equal to zero, which is a result of the construction of the centroid, being the origin of the new coordinate system. As from this residual matrix, the loadings of the second factor cannot be extracted in the same way as for the first factor, a further step has to be taken before continuing: The signs of one or more variates are changed (plus is turned to minus and contrariwise), which is equivalent to turning the sign of one or more columns and corresponding rows. Only the diagonal elements always stay positive.

The goal is, to achieve a residual correlation matrix with as many positive values as possible, to be able to create a new centroid. The process of changing signs in columns and rows is called "mirroring" (only the sign of one coordinate of a point in a two-dimensional space is changed) or "reflection".

If many variables are employed, a multitude of different possibilities leads to an acceptable solution and because of that, the centroid factor analysis looses its objectivity.

After the mirroring procedure, the loadings of the second factor are calculated by the same summations and mathematical operations as it was done for the first factor. Before that, the new communalities of the residual correlation matrix have to be determined as well and are again set in the main diagonal. Following to the calculation procedure, the loadings of the variates, whose signs were changed, have to get back their initial signs. If negative values occur during the summation of columns, the signs of these columns have to be turned as well, before beginning the factor extraction process.

These steps are continued, until the residual correlation matrix mainly contains very small values, which means that the correlations are already exhausted.

According to Lawley and Maxwell (1971), the results can be improved by comparing the first estimates of the communalities with the calculated values. If the values differ significantly, the loadings should be extracted once more by using the calculated communalities instead of the highest (absolute) correlation coefficients. In the example mentioned below (Section 7.3), the loadings were recalculated until the differences between the estimated and calculated values for the communalities were less than 0.01 (two steps were needed to achieve this required accuracy), whereas for the atmospheric data sets, a difference value of 0.1 was considered as sufficient.

7.3 Differences in the Results of the 4 Methods Presented on an Example

To compare differences in the results of the four different implemented factor analysis methods, a small data set of Mardia et al. (1979), composed of $n = 88$ observations (students) and $p = 5$ variables (examination marks on different topics with open or closed books) was employed (cf., Section 6.6).

The correlation matrix, which is the starting point for all implemented factor analysis methods, for this sample is again given by:

$$R = \begin{pmatrix} 1.000 & 0.553 & 0.547 & 0.409 & 0.389 \\ 0.553 & 1.000 & 0.610 & 0.485 & 0.436 \\ 0.547 & 0.610 & 1.000 & 0.711 & 0.665 \\ 0.409 & 0.485 & 0.711 & 1.000 & 0.607 \\ 0.389 & 0.436 & 0.665 & 0.607 & 1.000 \end{pmatrix}$$

If not quoted differently, for each method, $k = 2$ factors were chosen.

7.3.1 Examined Matrices and the Eigenvalues

As described in Section 7.2, each of the methods makes use of another matrix to extract the factors, which should explain the data's variances. The cornerstones of each method

are summarized below.

Iterative PFA: A reduced correlation matrix $\boldsymbol{R}^{\text{reduced}} = \boldsymbol{R} - \boldsymbol{\Psi}$ is employed. For the first iteration step, the ones in the diagonal of the correlation matrix are replaced by a first estimate for the communalities, which is the largest correlation between the i^{th} and another variable. During the following iteration steps, the recalculated values for the communalties are used to define the actual reduced matrix.

True FA: An auxiliary matrix $\mathbf{R}^* = (\text{diag}\boldsymbol{R}^{-1})^{1/2}\,\boldsymbol{R}\,(\text{diag}\boldsymbol{R}^{-1})^{1/2}$, with $\boldsymbol{R}^{-1}$ being the inverse of the correlation matrix of the data is applied, to achieve the eigenvalues and eigenvectors for further estimations.

ML-FA: Iterative PFA is used before calculating maximum likelihood factors using an EM-algorithm, during which no inherent matrix is examined.

Centroid-FA: The same reduced correlation matrix as in iterative PFA is used in the beginning to calculate loadings of the first factor. To obtain the further loadings, a so-called "residual correlation matrix" is derived, using the loadings and the reduced correlation matrix of the previous factor.

While in iterative principal factor analysis and in true factor analysis, the eigenvalues and eigenvectors are calculated by the decomposition of a certain matrix, in maximum likelihood and centroid factor analysis, the corresponding eigenvalues are given by the sum of the squares of the factor loadings of each factor. Table 7.1 shows the different eigenvalues of the example.

While the first two eigenvalues of principal, ML-, and centroid factor analysis are very similar, the ones of true factor analysis vary in size. The eigenvalues of the auxiliary matrix $\boldsymbol{R}^*$ in this method seem to be very high, compared to the other methods. But it has to be taken into account that the values of this auxiliary matrix are weighted to achieve the required values of $\boldsymbol{A}$ and $\boldsymbol{\Psi}$. The eigenvalues, which are derived from the factor loadings are again similar to those of the other methods.

Slight differences between the methods occur in the decrease of the eigenvalues' amount from the first to the fifth eigenvalue. In iterative principal factor analysis, the decrease of the eigenvalues is taking shape strongest, resulting even in one negative eigenvalue, whereas in centroid factor analysis the last three eigenvalues are larger than those of the other methods considered (see Figure 7.2).

How Many Different Factors are Needed?

A major question of a typical factor analysis is how many different factors are needed to explain the variances of an original data set in an appropriate way. There is a variety of selection rules and some of them will be explained in detail in Section 8.1.

Most of these rules originate from principal component analysis, even though they are recommended for factor analysis. Anyhow, for atmospheric data sets like those, which will be discussed in Section 9.2, these rules do not seem to be appropriate. Factor analysis is much more sensitive to the number of selected factors than PCA. Following the

Method	EV1	EV2	EV3	EV4	EV5	Comment
Iterative PFA	2.840	0.357	0.083	0.022	–0.053	3 steps until desired precision, i.e., for each element of $\boldsymbol{\Psi}$: $\max \lvert \boldsymbol{\Psi}_{\text{step } i} - \Psi_{\text{step } (i-1)} \rvert < 0.05$
	2.833	0.338	0.050	0.009	–0.033	after 3 iteration steps
True FA	7.005	1.334	0.812	0.741	0.655	Eigenvalues of $\boldsymbol{R}^*$
	2.862	0.379	0.080	0.049	$*$	Eigenvalues derived from factor loadings extracting 4 factors
ML-FA	2.831	0.354	0.087	0.021	$*$	Results from PFA as initial guess; Eigenvalues derived from factor loadings extracting 4 factors
Centroid-FA	2.839	0.356	0.093	0.065	0.037	Eigenvalues derived from factor loadings extracting 5 factors

$*$ ... calculation not possible due to mathematical constraints

Table 7.1: Differences of the eigenvalues according to the four implemented factor analysis methods.

requirements of the selection rules, in most cases too many factors would be extracted, resulting in some negative values for the ψ_{ii} and therefore in a non permissible solution (cf., Section 7.1). The problems of the different methods arising when using atmospheric data sets will be discussed in Section 9.2. At this place it just should be noted that the results of a scree plot (cf., Section 8.1), as shown in Figure 7.2, are a good starting point for the selection of the number of factors. In the case of our example, two factors were chosen.

7.3.2 Factor Loadings and Unique Matrix $\boldsymbol{\Psi}$

As discussed in detail in Section 7.1, the goal of factor analysis is to find a solution for the model equation $\boldsymbol{R} = \boldsymbol{A}\,\boldsymbol{A}' + \boldsymbol{\Psi}$, where $\boldsymbol{A}$ is the matrix of the factor loadings and $\boldsymbol{\Psi}$ a diagonal matrix containing the unique variances.

The four implemented methods of factor analysis led to slightly different results for the factor loadings, which are presented in Table 7.2 as well as the varimax rotated factor loadings.

The aim of factor rotation is to make the output more understandable and to facilitate the interpretation. The sum of the eigenvalues is not affected by rotation, but it alters the eigenvalues of particular factors and changes the factor loadings (Garson 2005). A detailed description of factor rotation and the varimax rotation method is found in Sec-

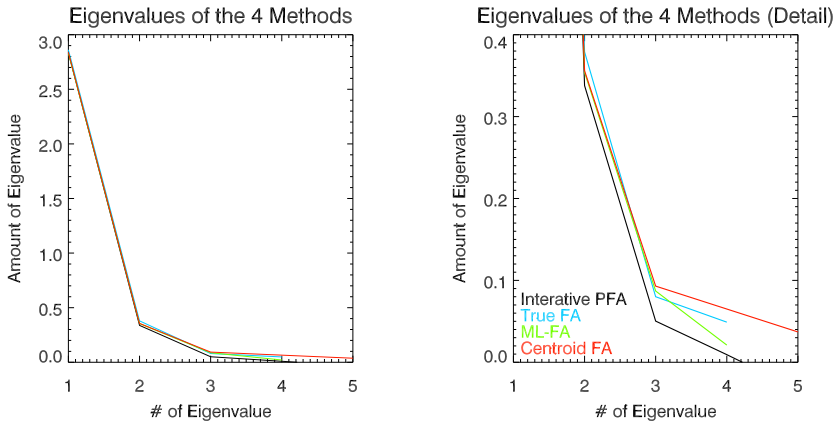


Figure 7.2: Differences in the eigenvalues (derived from the factor loadings) of the four implemented methods. While the first two eigenvalues of all methods are quite similar, the detail graphic on the right side shows the fanning out of the eigenvalues number three to five.

tion 8.2.

In general, the differences between the factors calculated with the four methods are minor and range, considering one variable, between 7.8 % to 8.7 % with the exception of factor two. Here, the loadings of the third variable differ about 67 % between true (factor loading = −0.037) and centroid (factor loading = −0.113) factor analysis, with reference to the absolutely higher value. Large differences from 10.4 % to 22 % are also found for variables one, two and four regarding these two methods. Anyhow, these differences disappear after rotating the factors, where again a maximum difference of around 8 % is achieved. If this fact is valid for data sets in general, varimax rotation of factor loadings seems to reduce differences in the loadings resulting from the four factor analysis methods.

The effect of factor rotation is clearly visible in this example. As expected, the five different variables seem to be characterized by two underlaying factors (marks depend rather on open- and closed books than on the five topics). While the first two variables (standing for closed books) of the rotated loadings of the first factor have rather small values, the three remaining variables (standing for open books) are strongly pronounced. The second factor is given the other way round: Here, the first two variables are highly loaded, whereas the last two variables are rather humble.

The matrix $\mathbf{\Psi}$ contains the variable-specific, unique variances, which cannot be explained by common factors. In other words, the values of $\mathbf{\Psi}$ explain, how much purely random or unique variance each observed variable includes. Following values were achieved with the four different factor analysis techniques for the diagonal of $\mathbf{\Psi}$:

Iterative PFA: 0.4640 0.4062 0.2276 0.3350 0.3962

Method	Factor Loadings					$\Sigma_{i=1}^{p} a_i^2$	Tot. Var.
Factor 1							
Iter.PFA	0.645	0.711	0.877	0.779	0.733	2.833	56.66 %
True FA	0.632	0.700	0.879	0.785	0.740	2.826	56.52 %
ML-FA	0.692	0.753	0.865	0.750	0.698	2.843	56.85 %
Cent.FA	0.648	0.715	0.890	0.769	0.718	2.830	56.59 %
Factor 2							
Iter.PFA	0.347	0.296	–0.065	–0.242	–0.259	0.338	6.76 %
True FA	0.371	0.317	–0.037	–0.205	–0.229	0.334	6.68 %
ML-FA	0.358	0.309	–0.045	–0.231	–0.254	0.343	6.87 %
Cent.FA	0.315	0.284	–0.113	–0.263	–0.253	0.326	6.53 %
Varimax Rotated Factor 1							
Iter.PFA	0.265	0.349	0.709	0.749	0.725	1.781	35.63 %
True FA	0.266	0.353	0.713	0.744	0.723	1.780	35.60 %
ML-FA	0.286	0.364	0.682	0.717	0.693	1.673	33.47 %
Cent.FA	0.274	0.345	0.740	0.749	0.704	1.798	35.96 %
Varimax Rotated Factor 2							
Iter.PFA	0.682	0.687	0.520	0.322	0.279	1.390	27.79 %
True FA	0.683	0.682	0.515	0.325	0.278	1.380	27.60 %
ML-FA	0.724	0.728	0.535	0.318	0.267	1.513	30.25 %
Cent.FA	0.666	0.688	0.508	0.315	0.289	1.358	27.16 %

Table 7.2: The (rotated) factor loadings according to the four implemented factor analysis methods.

True FA:	0.4589	0.4084	0.2419	0.3379	0.3832
ML-FA:	0.4675	0.4080	0.2314	0.3211	0.3830
Centroid-FA:	0.4720	0.4136	0.2009	0.3470	0.4175

In contrast to the loadings, the values only differ minor regarding the four methods, even though a slight increase of the deviation can be noticed from the third value on, which is mainly given due to diverging values of the centroid method. Apart from the third value (deviation of about 17 % between true and centroid factor analysis), the differences between the methods stay below 8.2 % (as reference, the highest value of each variable was considered).

Total Variance Explained by the Factors

While the first factor pattern delineates the largest pattern of relationship in the data, the second the next largest pattern and so on, the amount of variation in the data described by each factor decreases successively with each factor. The ratio of the sum of

the values of the squared loadings (communality h^2) to the number of variables (p), multiplied by 100, equals to the percentage of total variation in the data that is patterned. It is a measure for the order, uniformity, or regularity in the data.

Variation explained by ...	Iterat. PFA	True FA	ML-FA	Centroid-FA
Factor 1	56.66 %	56.52 %	56.85 %	56.56 %
Factor 2	6.76 %	6.68 %	6.87 %	6.42 %
Varimax rotated Factor 1	35.63 %	35.60 %	33.47 %	35.96 %
Varimax rotated Factor 2	27.79 %	27.60 %	30.25 %	27.16 %
Factor 1+2	63.42 %	63.20 %	63.72 %	62.98 %

Table 7.3: Variance of the exemplary data set explained by two selected factors.

Table 7.3 shows the amount of the variances explained by the selected two factors. As the factor loadings are only influenced by the common factors, the unique variance is excluded from this consideration. All in all, about 63 % of the variance of the observed data can be explained by the selected two factors. While in the unrotated case, the largest part of the variance is explained by the first factor, rotation leads to a more equal distribution of the explained variances, anyhow, the total variance stays the same.

There are nearly no differences in the variances regarding the four methods for unrotated factor loadings. For varimax rotated factor loadings, the maximum likelihood method leads to an even more uniform distribution of the variances than the other three methods.

How Well do the Hypothesized Factors Explain the Observed Data?

Having extracted the common and unique variances, the original correlation matrix can be reproduced making use of the estimates for $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{\Psi}}$ in equation (7.15). A residual correlation matrix,

$$\boldsymbol{R}_{\text{residual}} = \boldsymbol{R}_{\text{original}} - \hat{\boldsymbol{A}}\,\hat{\boldsymbol{A}}' - \hat{\boldsymbol{\Psi}}, \tag{7.37}$$

can be investigated to check, whether the selected factors succeed in reproducing the original correlations, which is the fact, when the elements of the residual correlation matrix are small. Table 7.4 shows the maximal absolute value of any item of the residual correlation matrix for the four methods, as well as the root mean square (RMS) of all elements of the matrix, which is defined by

$$\text{RMS} = \sqrt{\frac{\sum_{i,j} r_{ij}^2}{p^2}}. \tag{7.38}$$

	Iterat. PFA	True FA	ML-FA	Centroid-FA
Largest absolute value of $\boldsymbol{R}_{\text{residual}}$	0.0260	0.0208	0.0780	0.0112
RMS of all elements of $\boldsymbol{R}_{\text{residual}}$	0.0054	0.0021	0.0241	0.0009

Table 7.4: Selected values of the four residual correlation matrices, indicating how well the number of selected factors reproduce the original correlation matrix.

Generally, the residual correlation matrices of all four methods just show small values, nevertheless, some interesting facts occur in this example. Even though centroid factor analysis is the oldest and, in the age of the computer no longer used, method, the most accurate results were achieved with it. Anyhow, because of the arbitrariness of the method (cf., Section 7.2.4), which will also be shown in the atmospheric data set, it is not given preference to it.

Just as centroid factor analysis surprises with good results in this example, maximum-likelihood factor analysis leads to the worst. Unfortunately, it is beyond the scope of this work to analyze the reasons for this fact.

Iterative and principal factor analysis achieve about the same accuracy, with root mean square deviations between the "true" and calculated correlation matrix of less than 1 %, which surely can be considered as a "good" result.

7.3.3 Factor Scores

In Section 7.1.1, two different methods to calculate the factor scores, knowing the factor loadings, the unique variance matrix $\boldsymbol{\Psi}$ and the original data, namely according to Bartlett and Anderson-Rubin, were presented.

To get a feeling for the two factor scores, factor pattern coefficients ($\boldsymbol{Af}$), not including the unique factors $\boldsymbol{u}$, were calculated and subtracted from the original data (cf., equation (7.1)). Since the unique factors are achieved by estimation as well (employing the original data, the factor scores and factor loadings), it would be possible to include them.

But, on the assumption that the unique factors $\boldsymbol{u}$ are more or less the same for all methods (as the factor loadings and unique matrix $\boldsymbol{\Psi}$ do not differ much), the differences between the original data and the calculated factor pattern coefficient (as described above) served for the comparison of Bartlett's and Anderson-Rubin's factor scores.

In Table 7.5, "min" and "max" stands for the minimal and maximal value of the formed differences between original data and factor patterns; "RMS" is the root mean square of all elements of the difference matrix. The rather high values may go back to the fact that the unique factors are not included in the calculation.

Very similar to the factor loadings and $\boldsymbol{\Psi}$, there are no striking differences between

	Not Rotated Results			Varimax Rotated Results		
Method	min	max	RMS	min	max	RMS
Bartlett's Factor Scores						
Iter. PFA	−26.47	23.94	9.65	−7.16	81.66	39.39
True FA	−22.77	24.79	7.91	−7.33	82.42	39.44
ML-FA	−22.04	24.26	7.98	−6.04	80.93	39.74
Cent.FA	−33.17	28.80	14.23	−6.35	82.12	39.69
Anderson-Rubin's Factor Scores						
Iter. PFA	−27.61	20.98	9.79	−10.46	67.87	31.65
True FA	−26.18	21.16	8.71	−10.67	68.32	31.55
ML-FA	−25.60	21.60	8.67	−9.31	70.76	32.09
Cent.FA	−32.26	24.12	12.28	−9.67	68.77	32.15

Table 7.5: Minimal and maximal differences as well as root mean square between original data and reproduced data employing factor pattern coefficients excluding the unique factors.

the four factor analysis techniques, which was expected, as the factor scores depend on the factor loadings and $\mathbf{\Psi}$. But it also seems that there are only minor differences in the results, no matter whether Bartlett's or Anderson-Rubin's method is used to calculate the factor scores.

As iterative principal factor analysis proved to be the most appropriate technique for atmospheric data sets (at least for those, which were investigated in the context of this work, cf., Section 9.2), Bartlett's factor scores were used for further considerations.

8 Common Properties and Differences of PCA and FA

(Authors: B.C. Lackner and B. Pirscher)

8.1 Determination of the Number of Factors

Principal component analysis and factor analysis aim to reduce the number of variables and thereby to reduce the dimensionality of the original data matrix. It appears to be reasonable to "cut" the factors at one well-defined value, which is determined by one particular selection rule. There are a lot of different methods extracting each another number of variables. It depends on the PCA'/FA' goal, which method emerges to be the appropriate one.

8.1.1 Cumulative Percentage of Total Variation

Reducing the number of factors entails that the total amount of variation of the original data cannot be reproduced exactly because

$$\sum_{l=1}^{p} \mathrm{Var}(x_l) = \sum_{l=1}^{p} \mathrm{Var}(f_l) > \sum_{l=1}^{k} \mathrm{Var}(f_l) \tag{8.1}$$

with $k < p$.

The total percentage of variation accounted for by the first k factor scores t_k is defined as

$$t_k = \frac{\sum_{l=1}^{k} \mathrm{Var}(f_l)}{\sum_{l=1}^{p} \mathrm{Var}(f_l)} \cdot 100 = \frac{\sum_{l=1}^{k} \lambda_l}{\sum_{l=1}^{p} \lambda_l} \cdot 100. \tag{8.2}$$

One possibility of reducing the number of variables is to select a fixed cumulative percentage of intrinsic total variation, which should be accounted for by the factor scores. Then the favored number k is set by the smallest value for which the desired percentage is exceeded

$$\frac{\sum_{l=1}^{k} \lambda_l}{\sum_{l=1}^{p} \lambda_l} \cdot 100 > t^*. \tag{8.3}$$

The chosen value of accounted variation t^* is arbitrary, it is located between 70 % and 90 %, sometimes up to 99 %, depending on the data set. If a small number of PCs are very dominant and some sources of variation, which are arranged behind are wanted to be analyzed, the cut-off value should be selected higher than 90 % but if there are a lot of variables contributing comparably little of the total variation, it is justifiable to use a cut-off value even smaller than 70 %.

8.1.2 Kaiser's Rule

Calculating the factor scores by means of the correlation matrix, the Kaiser's rule states that the number of selected factor scores should be equal to the number of eigenvalues λ larger than one[1]. The reason is that only factor scores with variances larger than one contain more information than the original variables. All variables associated with eigenvalues smaller than one can be neglected because of their small content of information. According to Jolliffe (2002), the influence of sampling variation can be inhibited if the limit is chosen more tolerant, he proposes $\lambda > 0.7$.

If factor scores of the covariance matrix are given the rule of Kaiser has to be modified because of their non standardized variances. Jolliffe (2002) mentions that the average value of the eigenvalues $\bar{\lambda}$ or rather $0.7\bar{\lambda}$ could be chosen as cut-off value, but problems may arise if covariance matrices have widely differing variances because then too few factor scores will satisfy the criterion.

8.1.3 Scree Test

The scree graph is a graphical representation of the eigenvalues, which are arranged according to their size and plotted on the ordinate against their index on the abscissa. A break from a steep to a shallow slope in the graph represents the last principal component, which should be taken into further consideration. The mathematical point of view says that the criterion evaluates the difference between three adjoined eigenvalues λ_{j-1}, λ_j, and λ_{j+1}. Factors corresponding to eigenvalues whose differences are fairly constant can be neglected. The test is still a subjective criterion because no explicit formula can be specified.

Alternatively, in atmospheric sciences the logarithm of the eigenvalue ($\log \lambda$) can be plotted against the index; this is known as log-eigenvalue (LEV) diagram. The number of principal components, whose logarithmic eigenvalues are connected to a straight line, should be discarded.

[1] In case of FA, the eigenvalues are derived by summing up the squared factor loadings of the respective factor.

8.1.4 Application to the Example

As the results of FA depend on the number of extracted factors (cf., Section 8.3), only PCA' results are to be examined at this point.

Cumulative Percentage of Total Variation Table 8.1 shows the cumulative percentage of information explained by the factors of the covariance matrix and the correlation matrix. Dependent on the chosen cut-off value, two or more factors could pass the criterion. Choosing the usual value of 90 %, four factors could be retained in further calculations, independent of the matrix used.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
$\boldsymbol{S}$	63.620 %	78.411 %	87.310 %	95.068 %	100.000 %
$\boldsymbol{R}$	61.912 %	80.126 %	89.476 %	97.103 %	100.001 %

Table 8.1: Cumulative percentage of information explained by the factors of the covariance matrix and the correlation matrix.

Kaiser's Rule Recalling the eigenvalues calculated from the covariance matrix and the correlation matrix (depicted in Table 6.1) of the example discussed in Section 6.6, the number of selected factors can be specified by means of the Kaiser's rule.

Looking at the covariance matrix it can be recognized that two eigenvalues are larger than the average of all eigenvalues ($\bar{\lambda} = 221.93$) meaning that two factors will pass the criterion. Being more tolerant and using the cut-off value at $0.7\bar{\lambda} = 155.35$, the number of selected factors will be the same.

If factor scores of the correlation matrix are used the determining value is one, also being the average of the eigenvalues. Because of only one eigenvalue being larger than one, only one factor satisfies this criterion; if the criterion is expanded to 0.7 the second factor can be added.

Scree Test and LEV-Test Figure 8.1 shows the scree plot (left) and the LEV diagram (right) for the open/closed book example. The "elbow" in the scree plot clearly occurs at the second eigenvalue, so only two principal components could be taken into following considerations. The same is obtained looking at the LEV diagram. All eigenvalues from the second one upward can be connected with an imaginary straight line.

The results obtained from calculations with the covariance matrix are similar to the plots shown in Figure 8.1, except for the dimensions of the eigenvalues and the logarithmic eigenvalues. Furthermore, the LEV diagram obtained from the correlation matrix is easier to be interpreted, because of the logarithmic eigenvalues of the covariance matrix decreasing more even and the roughly straight line connecting practically all logarithmic eigenvalues.

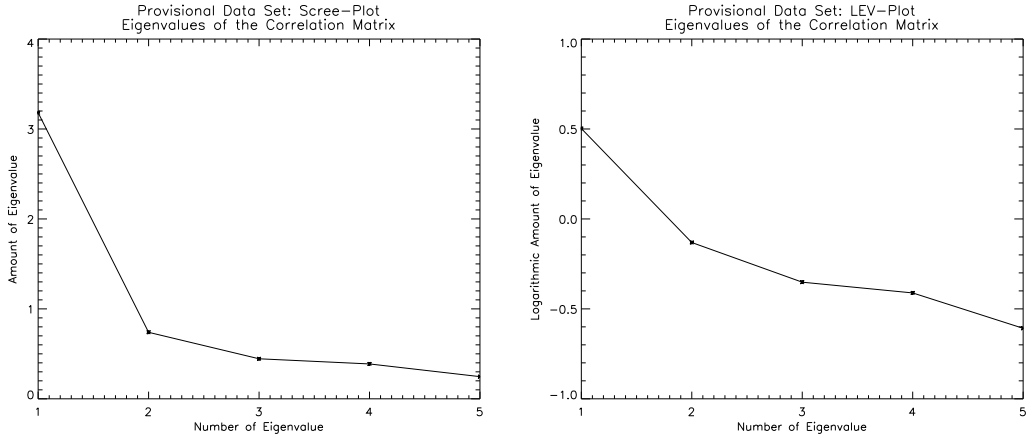


Figure 8.1: Left: Scree graph created for the open/closed book example, generated with the correlation matrix. Right: LEV diagram obtained from the same data set applied to the same matrix.

Summary Table 8.2 summarizes the number of factors being retained by several selection rules discussed above. The cut-off value of the cumulative percentage of explained variance is chosen to be 90 %. The values given in parentheses are the number of factors being retained by the enhanced Kaiser's rule.

It can be seen that the number of the most important factors is very similar, most methods yield two factors to be selected. In most cases, the behavior of the selection rules is not as uniform.

	Cum. Variance	Kaiser's rule	Scree-Test	LEV-Plot
$\boldsymbol{S}$	4	2 (2)	2	2
$\boldsymbol{R}$	4	1 (2)	2	2

Table 8.2: Number of factors passing the examination with different rules.

8.2 Rotation of Factor Loadings

A main interest in doing multivariate statistical analysis is to simplify the variables of an abstract data field and to find typical patterns hidden behind the data. Sometimes a lot of variables contribute to the first factors and the interpretation of the result is very difficult. Rotating the factors facilitates the interpretation of the results.

According to von Storch and Zwiers (2003), some arguments pro and contra the application of rotation of the factor loading matrix have to be considered.

1. Advantages:

- The technique produces compact patterns, which can be used for further investigations;
- Rotated factor loadings are less sensitive to the distribution of observing locations than not rotated factor loadings;
- Rotated factor loadings are often statistically more stable than not rotated factor loadings;

2. Disadvantages:

- There is no strong rotation criterion, it can be chosen arbitrarily;
- The results are sensitive to the normalization of the factor loadings;
- If the number of selected factors has changed, the calculation has to be repeated;
- There may be a loss of information about the dominant sources of variation in the data.

Performing a rotation technique means changing the factor loadings/coefficients, which can be interpreted as projections of the variables onto the factors. The configuration of the variables remains the same. An oblique rotation yields factors without the constraint of orthogonality, which results in measurable correlations between them. These correlations can be examined. In orthogonal rotation techniques the axes remain perpendicular to each other and the factors stay uncorrelated.

All rotation methods' purpose a "simple structure", which means that the number of variables, which are loading on the factors should be minimized to get a better interpretability. Variables, which are associated with more than one factor in the same magnitude, should be reassigned to one particular factor.

Reyment and Jöreskog (1993) cite Thurstone (1935), who defined some criteria to establish a simple structure. These criteria are that:

- Each factor loading matrix/matrix of coefficients $\boldsymbol{A}$ should contain at least one zero in each row;
- Each factor loading matrix/matrix of coefficients $\boldsymbol{A}$ should contain at least k zeros in each column (k is the number of extracted factors);
- Every pair of factors should have some variables being "high loaded" on one factor and "low loaded" on the other;
- Every pair of factors should have some variables being low loaded on both factors;
- Every pair of factors should have only a few variables, which have non vanishing loadings on both.

Following these criteria the new factor loading matrix $\boldsymbol{B}$ should have a lot of near-zero values and some high loadings and the new factors are independent on most of the variables.

Generally, the matrix of rotated factor loadings (or rotated empirical orthogonal functions) $\boldsymbol{B}$ can be calculated by

$$\boldsymbol{B} = \boldsymbol{AT}, \tag{8.4}$$

where $\boldsymbol{A}$ is the matrix of unrotated factor loadings/coefficients and $\boldsymbol{T}$ is the transformation matrix, which turns out to be

$$\boldsymbol{T} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}, \tag{8.5}$$

with the angle φ being the quantity of interest.

Bortz (1989) mentions the following rotation techniques:

* Binormamin
* Biquartimin
* Covarimin
* Equimax
* Maxplane
* Oblimax
* Oblimin
* Parsimax
* Promax
* Quartimax
* Quartimin
* Tandem
* Varimax
* Varisim

Most of them realize oblique rotations, and some orthogonal rotations.

8.2.1 The Varimax Procedure

The most popular orthogonal rotation technique is the varimax procedure, which has been developed by Kaiser (1958). The technique aims at maximizing the variance of the squared loadings in each factor; factors with medium loadings on a variable will be amplified or damped.

The variance of squared loadings of the factor j is defined as

$$s_j^2 = \frac{1}{p}\sum_{i=1}^{p}(a_{ij}^2)^2 - \frac{1}{p^2}\left(\sum_{i=1}^{p} a_{ij}^2\right)^2, \tag{8.6}$$

with p being the number of variables and a_{ij} being the elements of the unrotated factor loading matrix.

Simplicity is achieved if the accumulated variance of the selected k factors is maximized

$$Q = \sum_{j=1}^{k} s_j^2 \rightarrow \text{ Maximum.} \tag{8.7}$$

All pairs of factors j and j' are rotated, one after the other, so that the respective sum $s_j^2 + s_{j'}^2$ will be a maximum

$$s_j^2 + s_{j'}^2 = \left[\frac{1}{p}\sum_{i=1}^{p}(a_{ij}^2)^2 - \frac{1}{p^2}\left(\sum_{i=1}^{p}a_{ij}^2\right)^2\right] + \left[\frac{1}{p}\sum_{i=1}^{p}(a_{ij'}^2)^2 - \frac{1}{p^2}\left(\sum_{i=1}^{p}a_{ij'}^2\right)^2\right]. \quad (8.8)$$

Q results from

$$Q = \frac{1}{p}\sum_{j=1}^{k}\sum_{i=1}^{p}(a_{ij}^2)^2 - \frac{1}{p^2}\sum_{j=1}^{k}\left(\sum_{i=1}^{p}a_{ij}^2\right)^2. \quad (8.9)$$

According to Bortz (1989), the conditional equation of the determination of the angle φ following from the rotation of the factors j and j' is

$$C = \frac{2 \cdot \left[p\sum_{i=1}^{p}(a_{ij}^2 - a_{ij'}^2)\cdot(2a_{ij}\cdot a_{ij'}) - \sum_{i=1}^{p}(a_{ij}^2 - a_{ij'}^2)\cdot\sum_{i=1}^{p}(2a_{ij}a_{ij'})\right]}{p\cdot\left[\sum_{i=1}^{p}\left((a_{ij}^2 - a_{ij'}^2)^2 - (2a_{ij}a_{ij'})^2\right)\right] - \left[\left(\sum_{i=1}^{p}(a_{ij}^2 - a_{ij'}^2)\right)^2 - \left(\sum_{i=1}^{p}(2a_{ij}a_{ij'})\right)^2\right]}, \quad (8.10)$$

where the absolute value $|C|$ is equal to

$$|C| = \tan(4\cdot\varphi). \quad (8.11)$$

The last step in the calculation of the angle of rotation, is to determine the proper quadrant, in which the angle is situated.

- If the enumerator and dominator are positive, the transformation matrix is

$$\boldsymbol{T} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}. \quad (8.12)$$

- If the enumerator is positive and the dominator negative, the transformation matrix is

$$\boldsymbol{T} = \begin{pmatrix} \cos(45^\circ - \varphi) & -\sin(45^\circ - \varphi) \\ \sin(45^\circ - \varphi) & \cos(45^\circ - \varphi) \end{pmatrix}. \quad (8.13)$$

- If the enumerator is negative and the dominator positive, the transformation matrix is

$$\boldsymbol{T} = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}. \quad (8.14)$$

- If the enumerator and dominator are negative, the transformation matrix is

$$\boldsymbol{T} = \begin{pmatrix} \cos(45^\circ - \varphi) & \sin(45^\circ - \varphi) \\ -\sin(45^\circ - \varphi) & \cos(45^\circ - \varphi) \end{pmatrix}. \quad (8.15)$$

Each rotation between two factors j and $j^{'}$ delivers one transformation matrix $\boldsymbol{T}_{jj'}$ and, because of k factors being rotated, there will be $k(k-1)/2$ transformation matrices.

The resulting transformation matrix after one cycle of rotation is

$$\boldsymbol{T}^{*} = \boldsymbol{T}_{12} \cdot \boldsymbol{T}_{13} \cdot \ldots \boldsymbol{T}_{jj'} \ldots \tag{8.16}$$

and the matrix of rotated factor loadings $\boldsymbol{B}$ can be calculated by

$$\boldsymbol{B} = \boldsymbol{A}\boldsymbol{T}^{*}. \tag{8.17}$$

After the calculation of $\mathbf{B}$, the accumulated variance of the squared loadings of each factor will be computed once more, with (b_{ij}) being incorporated in the calculation of Q. As long as the old and the new values of Q differ to a certain amount, the rotation cycle will be continued and iterated until Q reaches a maximum.

If all variables nearly have the same communality, it is possible to normalize the rows of the factor loading matrix to unit length by dividing each element in the row by the square root of the associated communality. After rotation, the elements have to be transformed back to their original lengths.

A further aspect is that the fraction of total variance accounted for by a factor changes performing a varimax rotation, the variance will be more evenly distributed among the factors. The accounted variance of each factor can be calculated by equation (6.40), with the rotated renormalized loadings being the input variable $\tilde{a}$.

8.3 Differences between PCA and FA

In the Chapters 6 and 7, an example of Mardia et al. (1979) was employed, to demonstrate the properties of PCA and the four FA techniques, respectively. Now, the differences between the two methods should be investigated by means of the same example.

To understand the differences, let's call once more the models to mind:

$$\text{PCA}: \quad \boldsymbol{x} = \boldsymbol{A}\boldsymbol{f}(+\boldsymbol{e}) \tag{8.18}$$

$$\text{FA}: \quad \boldsymbol{x} = \boldsymbol{A}\boldsymbol{f} + \boldsymbol{u}. \tag{8.19}$$

The error term $\boldsymbol{e}$ in equation (8.18) is put in parenthesis because its amount depends on the number of extracted factors. The more components are used to explain the data's variance, the smaller is the error term; if all principal components are selected, $\boldsymbol{e}$ is equal to zero. As the number of extracted factors is determined within the calculation process, the same goes for the amount of the error term.

In contrast, the uniqueness $\boldsymbol{u}$ (see equation (8.19)) always plays an important role in FA. To be able to compare PCA and FA results, the renormalized PCA coefficients and principal components have to be used. Then, the PCA coefficients and FA factor loadings, which are stored in $\boldsymbol{A}$, have the same meaning; they are weighting values for the principal components/factor scores $\boldsymbol{f}$.

The differences between the two methods can be tracked in the coefficients/factor loadings $\boldsymbol{A}$, the principal components/factor scores $\boldsymbol{f}$, and the uniqueness $\boldsymbol{u}$. While $\boldsymbol{A}$ and $\boldsymbol{f}$ only differ in their amount, $\boldsymbol{u}$ is a key player. In Chapter 5 we mentioned that PCA is variance oriented, whereas FA is covariance oriented, which also leads us to the methods' differences.

PCA does not distinguish between common (affected by more than one variable) and unique (only affected by one variable) variances and both can be found in the principal components. Compared with this, the goal of FA is to separate common and unique variances by putting the common variances in the matrix $\boldsymbol{A}$ and the unique variances in the covariance matrix of $\boldsymbol{u}$ that is $\boldsymbol{\Psi}$. This allocation of variances implies that, as in FA only the factor loadings are used to calculate the amount of total variance in the data (the unique factors $\boldsymbol{u}$ or the unique variances $\boldsymbol{\psi}_{ii}$ are not included), the total explained variance with FA is generally less than with PCA (the meaning of common and unique variance in regard to the differences between PCA and FA will be dealt with later on).

Concerning atmospheric data sets this means that FA includes for example the possibility to spot those grid points and areas, which are strongly determined by unique variances and therefore are not covered in large-scale patterns. Even though PCA does not include this possibility, preference may be given to this method due to shorter calculation process and the independence on the number of selected factors, which will be discussed below.

Actually, in PCA the term "factor loadings" corresponds to the eigenvectors as well as to the renormalized coefficients calculated from the correlation matrix or from the covariance matrix. Because the four implemented factor analysis techniques refer to the correlation matrix, the comparison is only performed on this type of matrix.

Table 8.3 shows that the first and the second factor loadings generally point to the same direction (same sign). The first factor loadings always offer a positive sign and the second ones are positive in the first and the second element and negative from the third to the fifth element.

The differences between the loadings calculated with principal component analysis and the four implemented factor analysis techniques are depicted in Table 8.4.

As can be seen, there are systematical differences between the loadings. The first coefficients of PCA always take values, which are higher than those calculated with FA; the mean differences range between 0.0424 (iterative PFA) and 0.0482 (true FA) with variances between 0.00042 (iterative PFA) and 0.00069 (maximum likelihood FA). The smallest mean differences occur, when the factor analysis calculations are performed with the iterative principal factor analysis; similar differences to PCA are obtained with maximum likelihood factor analysis.

The differences of the second factors are more pronounced compared to the first ones. The mean differences range between -0.0264 (true FA) and 0.023 (centroid FA), with definitely stronger variances (compared to the first factor loadings) between 0.01644 (iterative PFA) and 0.01846 (centroid FA). The first and the second elements are always higher if the calculations are performed with PCA but elements three to five (where all factor loadings were negative) are lower if calculated with PCA. So, the differences

Factor Loading 1					
PCA	0.713	0.769	0.898	0.815	0.782
Iter.PFA	0.645	0.711	0.877	0.779	0.733
True FA	0.632	0.700	0.879	0.785	0.740
ML-FA	0.692	0.753	0.865	0.750	0.698
Cent.FA	0.648	0.715	0.890	0.769	0.718
Factor Loading 2					
PCA	0.555	0.380	−0.111	−0.334	−0.405
Iter.PFA	0.347	0.296	−0.065	−0.242	−0.259
True FA	0.371	0.317	−0.037	−0.205	−0.229
ML-FA	0.358	0.309	−0.045	−0.231	−0.254
Cent.FA	0.315	0.284	−0.113	−0.263	−0.253

Table 8.3: Differences of the factor loadings according to the PCA and to the four implemented factor analysis methods.

between the elements are stronger in case of PCA and less in case of FA.

In principal component analysis, different choices of k principal components do not affect the components or coefficients, that is to say that no matter how many components are extracted, the principal components, the explained variances, and the coefficients of the first, the second, and so on factor will always stay the same. This is not the case in factor analysis, which shows a large sensitivity to the number of k extracted factors. Figure 8.2 shows the changes of the first factors' loadings (left side graphs) and of the unique variances, which are found in the diagonal of the matrix $\mathbf{\Psi}$ (right side graphs), according to the number of k selected factors (different line colors in the graphs) and to three factor analysis techniques (maximum likelihood factor analysis results are not depicted as they do not differ much from iterative principal factor analysis, which was used to achieve the starting values for this technique). Due to mathematical constraints (cf., Section 7.3 and Section 9.2), a maximum of only two factors could be extracted for iterative principal factor analysis, four for true factor analysis whereas all five factors could be selected for centroid factor analysis.

The differences are better pronounced for the unique variances: The larger the number of extracted factors, the smaller are the unique variances. Furthermore, for principal and centroid factor analysis, the absolute differences between the unique variances for each variable are decreased by selecting more factors. For only one extracted factor, the highest unique variance is given for the first, second, fourth, and fifth variable, whereas the lowest for the third variable. In general, the largest decrease of the unique variances is given by selecting two factors instead of only one factor. A further increase of extracted factors results in an additional decrease of the unique variances, which however is less pronounced.

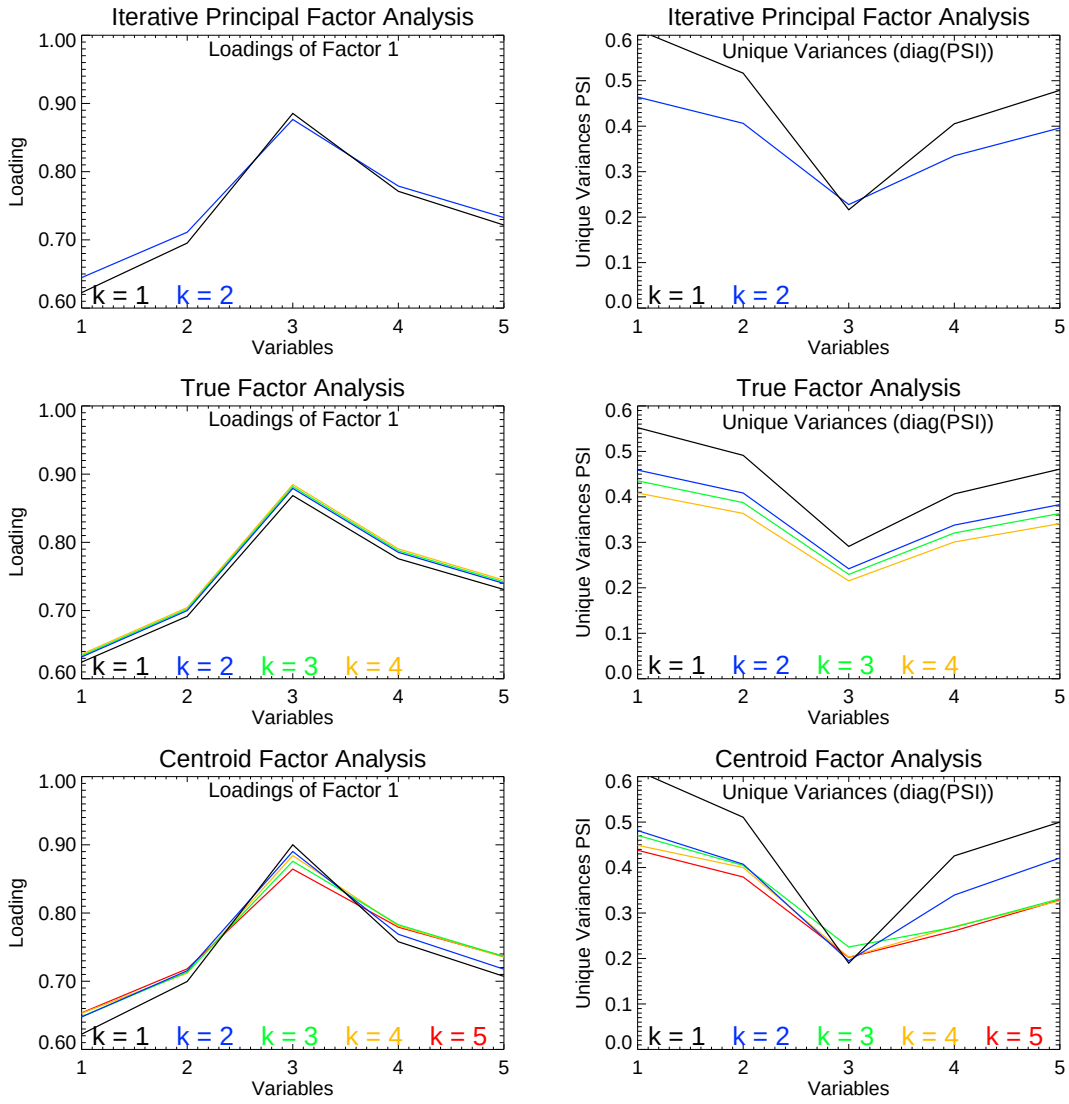


Figure 8.2: Factor Analysis: Dependence of common and unique variances on the number of selected factors.

	Difference between Factor Loading 1					mean
PCA – Iter.PFA	0.068	0.058	0.021	0.016	0.049	0.0424
PCA – True FA	0.081	0.069	0.019	0.030	0.042	0.0482
PCA – ML-FA	0.021	0.016	0.033	0.065	0.084	0.0438
PCA – Cent.FA	0.065	0.054	0.008	0.046	0.064	0.0474
	Difference between Factor Loading 2					mean
PCA – Iter.PFA	0.208	0.084	−0.046	−0.092	−0.146	0.0016
PCA – True FA	0.184	0.063	−0.074	−0.129	−0.176	−0.0264
PCA – ML-FA	0.197	0.071	−0.066	−0.103	−0.151	−0.0104
PCA – Cent.FA	0.240	0.096	0.002	−0.071	−0.152	0.0230

Table 8.4: Differences between the loadings calculated with the PCA ("coefficients") and the loadings calculated with the four implemented factor analysis techniques, as well as mean of differences.

A similar, but inverse and less pronounced, performance is shown in regard to the factor loadings. The more factors are extracted, the larger become the variables' loadings. For principal and centroid factor analysis the differences between the loadings of one factor are again diminished by increasing the number of selected factors.

As the common variance explained by a factor is given by the sum of the squared loadings for this factor (divided by the number of variables), different values are achieved according to the number of extracted factors.

Table 8.5 shows the explained variances of PCA and centroid FA according to the number of extracted factors. As the results of the four different factor analysis techniques are quite similar for Mardia's example, centroid factor analysis was taken, because with this technique up to five factors could be extracted without mathematical constraints. The differences between PCA and FA are clearly visible: PCA is not dependent on the number of extracted factors; each factor always contributes the same amount to the explained variance and all five factors together explain 100 % of the data's variance. FA results, by contrast, change with a varying "k". The total variances increase and the unique variances decrease with a rising number of extracted factors. Hence, also the allocation of common and unique variances on the two matrices depends on k, as the total of common and unique variances always amount to 100 %.

While an increasing k leads to the explanation of a larger part of the total variance in PCA (where 64 % are explained by one factor and 100 % by all five factors), only a comparatively small increase of the "total" (which is in fact the common) variance explained can be noticed in FA (55 % are explained by one factor and 68 % by all five factors), because the resting variance can be found in the unique factors.

# of selected factors	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
Principal Component Analysis					
Factor 1	63.62 %	63.62 %	63.62 %	63.62 %	63.62 %
Factor 2	– %	14.79 %	14.79 %	14.79 %	14.79 %
Factor 3	– %	– %	8.90 %	8.90 %	8.90 %
Factor 4	– %	– %	– %	7.76 %	7.76 %
Factor 5	– %	– %	– %	– %	4.93 %
Total	**63.62 %**	**78.41 %**	**87.31 %**	**95.07 %**	**100.00 %**
Centroid Factor Analysis					
Factor 1	55.23 %	56.59 %	57.03 %	57.31 %	56.79 %
Factor 2	– %	6.53 %	7.04 %	7.21 %	7.12 %
Factor 3	– %	– %	1.90 %	1.56 %	1.86 %
Factor 4	– %	– %	– %	0.94 %	1.30 %
Factor 5	– %	– %	– %	– %	0.74 %
Total	**55.23 %**	**63.12 %**	**65.98 %**	**67.02 %**	**67.81 %**
Unique Variances of Centroid Factor Analysis					
ψ_{11}	61.27 %	48.13 %	47.08 %	44.88 %	43.81 %
ψ_{22}	51.06 %	40.73 %	40.47 %	40.00 %	37.92 %
ψ_{33}	18.98 %	19.48 %	22.53 %	20.28 %	20.24 %
ψ_{44}	42.57 %	33.95 %	26.88 %	27.04 %	26.07 %
ψ_{55}	49.99 %	42.10 %	33.15 %	32.70 %	32.90 %
Mean Unique Variance	**44.77 %**	**36.88 %**	**34.02 %**	**32.98 %**	**32.19 %**

Table 8.5: Explained variances of PCA and centroid FA according to the number of extracted factors.

9 PCA and FA – Application to Atmospheric Data

9.1 Data Sets

(Author: B. Pirscher)

By means of principal component analysis and iterative principal factor analysis, some CHAMP retrieved atmospheric temperature data were investigated. The goal was the locating of some atmospheric patterns in four different regions arising in a time period of 3 years (between March 1, 2002 until February 28, 2005) yielding a 36 months observation period. Two global fields and two regional areas were analyzed for this purpose. To test the stability of principal component and factor analysis, each field was investigated in regard to two different resolutions (horizontal or vertical), one coarser and one finer, making an impact on the number of grid points.

1. The first investigated temperature field is located in the longitudinal sector of Eurasia-Africa (20°W to 70°E). The latitudinal resolution is 30° (yielding mean temperature values for low, mid, and high latitudes in the northern and southern hemisphere); the vertical resolution is 2 km and 5 km (from ground to a height of 34 km/35 km). Regarding the higher vertical resolution, the retrieved data span from a height of 2 km to 34 km resulting in 17 vertical levels, the lower vertical resolution includes 7 vertical levels from 5 km to 35 km. Altogether that makes 102 and 42 grid points, respectively.

2. The second global field comprises the temperature at an altitude of 15 km. The horizontal resolutions are $15° \times 45°$ and $30° \times 45°$ (latitude $\times$ longitude), arising in 12 and 6 latitude bins and 8 longitude bins, respectively. The first case yields 96 grid points, the second one only 48.

3. Regarding to the high southern latitudes, temperature data were analyzed between 57.5°S and 87.5°S. Building the zonal mean in 5° intervals (yielding 6 latitude bands), the vertical resolution was chosen to be 2 km and 5 km, resulting in 17 and 7 vertical levels. Therefore, the temperature data are given in 102 and 42 grid points.

4. The second regional field is set at the low latitudes to investigate the tropical tropopause. Zonal means for every 5° latitude between 17.5°S and 17.5°N were calculated, yielding 7 latitudinal regions; the vertical area is located between 12 km

and 22 km; the vertical resolution is once 1 km and once 2 km, resulting in 11 and 6 vertical layers and 77 and 42 grid points, respectively.

9.1.1 Pre-Treatment of Data

Multivariate statistical methods require the data being mean corrected. Generally, the mean of each variable is subtracted from the original data so that the variables are centered to zero. Then, the calculated factors $\boldsymbol{F}$ also have zero means. In the context of atmospheric data, every grid point contributes to one variable.

Investigating monthly means of atmospheric data, seasonal impacts are expected to dominate the first factors. Following Feeney and Hester (1967) (cited in Jolliffe (2002)), who removed a linear trend from some stock market prices before doing a principal component analysis, which yielded the first PC being similar to the second PC resulted from a previous analysis without removed trend, it should be possible to analyze monthly mean atmospheric data by diminishing the seasonal impact by calculating and subtracting the respective mean for each month (e.g., arithmetic mean of all Januaries).

Other possibilities adjusting the data to a mean would be the calculation of the mean from all objects or from all objects and all variables, but that does not seem to make sense.

Therefore, two different ways were chosen for the pre-treatment of the data:

1. The centering of the grid point data to the mean of all three years (abbreviated by "3-year mean");

2. The centering of the data to their twelve monthly means (abbreviated by "monthly mean").

9.1.2 Details on the PCA/FA of Atmospheric Fields

Figure 9.1 summarizes the different calculation procedures before applying the PCA/FA to atmospheric fields.

The examination of each atmospheric field was split in two sections, in the investigation of the results after the elimination of the 3-year mean and in the analysis of the results following from monthly mean corrected data.

While for PCA in each case the calculation was done by means of the sample correlation matrix $\boldsymbol{R}$ and the sample covariance matrix $\boldsymbol{S}$, for FA only the correlation matrix $\boldsymbol{R}$ was applied (cf., Section 9.2).

Because the PCA is a technique, which is based on the decomposition of the covariance matrix or the correlation matrix (each is positive semi definite) in its eigenvalues and eigenvectors, it is always possible to perform the calculation procedure (even though the results need not make sense in all cases). For both methods Reyment and Jöreskog (1993) require the number of variables being larger than the number of objects, but the investigated atmospheric fields never met this demand.

Furthermore, Reyment and Jöreskog (1993) state that "the number of objects must be sufficiently large to bring about stability in the variances and covariances (or cor-

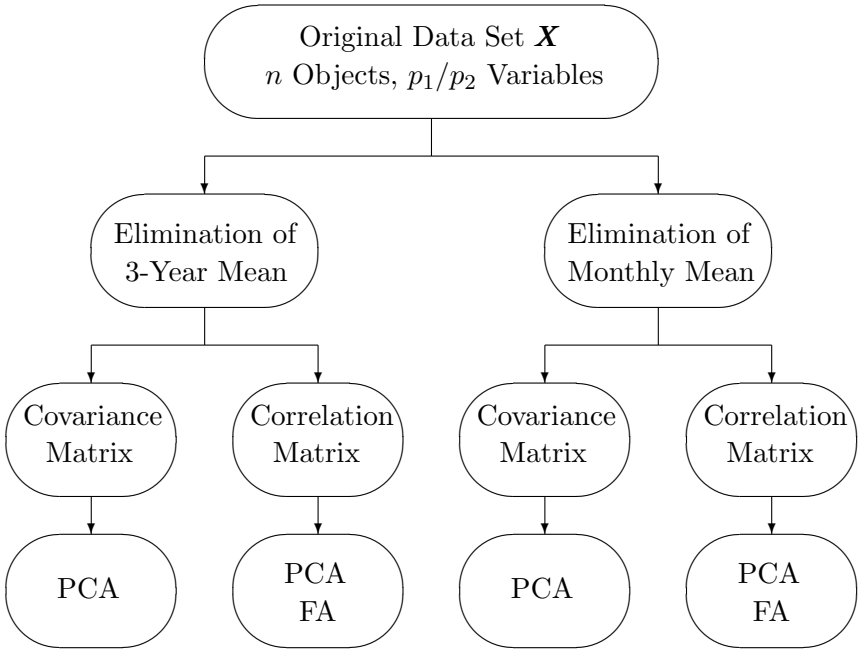


Figure 9.1: Pre-treatment of atmospheric data sets before performing PCA and FA calculations.

relations)". The stability of the PCA/FA was looked at by analyzing the principal components/factor scores, the coefficients/loadings and, in particular, the reconstruction of the data fields by means of the PCs/factor scores and the coefficients/loadings. The reconstruction is not possible if the data matrix is not suitable to be investigated by a PCA/FA.

For some applications, a reduction of the number of variables for further calculations is desired. This can be achieved by different criteria all yielding different numbers. When analyzing a data set by looking for atmospheric patterns, most of the criteria yield a too large number of extracted factors, but only a few of them can be physically interpreted. According to von Storch and Zwiers (2003), the physical interpretation is often limited to the first factor, because of the constraint of orthogonality, which has to be applied to them: Each column of the matrix $\boldsymbol{A}$ (which contains the coefficients/factor loadings) has to be orthogonal to all the others. The number of the "most important factors" will be shown in a tabular scheme, but the interpretation will only be based on the first few factors.

9.2 Factor Analysis Specific Problems With CHAMP RO Temperatures

(Author: B.C. Lackner)

While the results of the four different factor analysis techniques in the example of Mardia et al. (1979), which was discussed in Section 7.3, differ only insignificantly, applying the "real" atmospheric data set of CHAMP RO temperatures to the four selected areas (see Section 9.1), larger differences and difficulties, up to the impossibility to achieve results, showed up in the calculation process.

In contrast to PCA, only the results derived from the correlation matrices were considered, because employing the covariance matrices in the calculation process failed in as good as all cases. In addition, factor analysis as such is said to be invariant in regard to the type of the used matrix (even though the practice showed that this seems not to be valid for the calculation routines).

Table 9.1 presents if for the four factor analysis techniques implemented a problem occurred during the calculation process dependent on

- the selected atmospheric data set,
- the resolution (coarse or detailed grid),
- and the deviation score correction (3-year mean stands for centering the data to the mean of the variables and monthly mean stands for centering to the twelve monthly means).

Even a quick glance to the table shows that there are just very few "+", which indicate no problems during the calculation process. The symbol turning up most frequently is the "⊗", standing for a variety of problems during the calculation procedure, even though results could be achieved. These problems will be discussed more detailed in the following section. In general, they can be addressed as:

1. Problems with the variances, e.g., the variance of the first extracted factor is not the largest one as it is supposed to be.
2. The total variance explained by the selected number of factors is greater than 100 %.
3. Some values of the matrix $\mathbf{\Psi}$ are negative, which is an unacceptable solution since the items of this matrix are variances.
4. To achieve a mathematical meaningful result, only one factor could be extracted, which means that a rotation cannot be carried out.
5. Certainly, there is no conspicuousness in the results, but the plotted factors show strange properties.

Tropical Region	Detailed	Monthly Mean	+	⊖	⊗	⊗
		3-Year Mean	+	⊖	⊗	⊗
	Coarse	Monthly Mean	+	⊗	⊗	+
		3-Year Mean	+	⊗	⊗	+
South Polar Region	Detailed	Monthly Mean	⊗	⊖	⊗	⊗
		3-Year Mean	⊗	⊖	⊗	⊗
	Coarse	Monthly Mean	+	⊗	⊗	+
		3-Year Mean	+	⊗	⊗	⊗
Global Maps	Detailed	Monthly Mean	+	⊖	⊗	+
		3-Year Mean	+	⊖	⊖	⊗
	Coarse	Monthly Mean	+	⊖	⊗	⊗
		3-Year Mean	+	⊖	⊗	⊗
Global Lat×Height	Detailed	Monthly Mean	⊗	⊖	⊗	⊗
		3-Year Mean	⊗	⊖	⊗	⊗
	Coarse	Monthly Mean	+	⊗	⊗	⊗
		3-Year Mean	+	⊗	⊗	⊗
Resolution		Deviation Score Matrix	Iterative PFA	True FA	ML-FA	Centroid FA

+ ... no problems occurred during the calculation procedure
⊗ ... various problems occurred during the calculation procedure, but calculation was possible
⊖ ... impossibility to carry out calculation

Table 9.1: Problems of the four implemented factor analysis techniques occurring during the calculation process with the four selected atmospheric data sets of CHAMP RO temperatures.

In general, worse results appeared when a data set with a detailed resolution was investigated (this goes for all techniques apart from centroid factor analysis, where in one case, namely the monthly mean centered global maps with detailed resolution, no problems showed up during the calculation process, whereas for the coarse resolution problems occurred). Iterative principal factor analysis shows two striking examples for this fact, namely the global latitude $\times$ height field of the Eurasian-African sector and the south polar region. The problems arising there affected the factors and resulted either in a jagged appearance (Eurasian-African longitude slice) or in singular peaks (south polar region) of them (cf., Figure 9.2), even though the loadings of the coarse and detailed data sets just differed insignificantly.

For all detailed resolved data sets and also for the coarse resolution of global maps it was impossible to carry out the calculation with respect to true factor analysis. This problem resulted from negative values of the inverse of the correlation matrix and will be discussed particularly in Section 9.2.2.

The following sections will deal with special features, which have to be considered when applying atmospheric data sets such as CHAMP RO temperatures to the different factor analysis techniques as well as method specific problems with these data. All tables (apart from the one for iterative principal factor analysis, which caused no problems for the remaining coarse resolved data sets at all) are divided in two parts, one containing "general and method specific values", like the maximal number of factors that may be extracted to get a mathematically acceptable result or the number of computational steps until the required accuracy was given, the other demonstrating "method specific problems" and, in the case of true factor analysis, possible solutions.

As the most usable results were derived from iterative principal factor analysis of coarse resolved data sets, the results (only of data sets with coarse resolution) of this technique were used for the interpretation in the Sections 9.4 to 9.7.

9.2.1 Iterative Principal Factor Analysis and CHAMP RO Data

This most commonly applied factor analysis technique proved to be a useful method for atmospheric data sets, regarding the computational effort, even though several facts have to be considered. Table 9.2 shows some method specific values and further values used for the selected data sets.

The number of factors that should be extracted, was at first estimated according to the amount of eigenvalues of the decomposed reduced correlation matrix (cf., Section 7.2.1). It happened that, if too many factors were selected, negative ψ_{ii} values turned up. For the selected atmospheric data sets, a maximum of 4 to 17 factors could be extracted so that all variances ψ_{ii} stayed positive. The total variance explained by these values ranged between 72.2 % and 99.6 %, which is quite high.

The number of factors that was selected in the following, arose on the one hand from the maximal possible number as described above, and on the other hand from scree plots. Mostly, a relatively small number of extracted factors already explained quite a

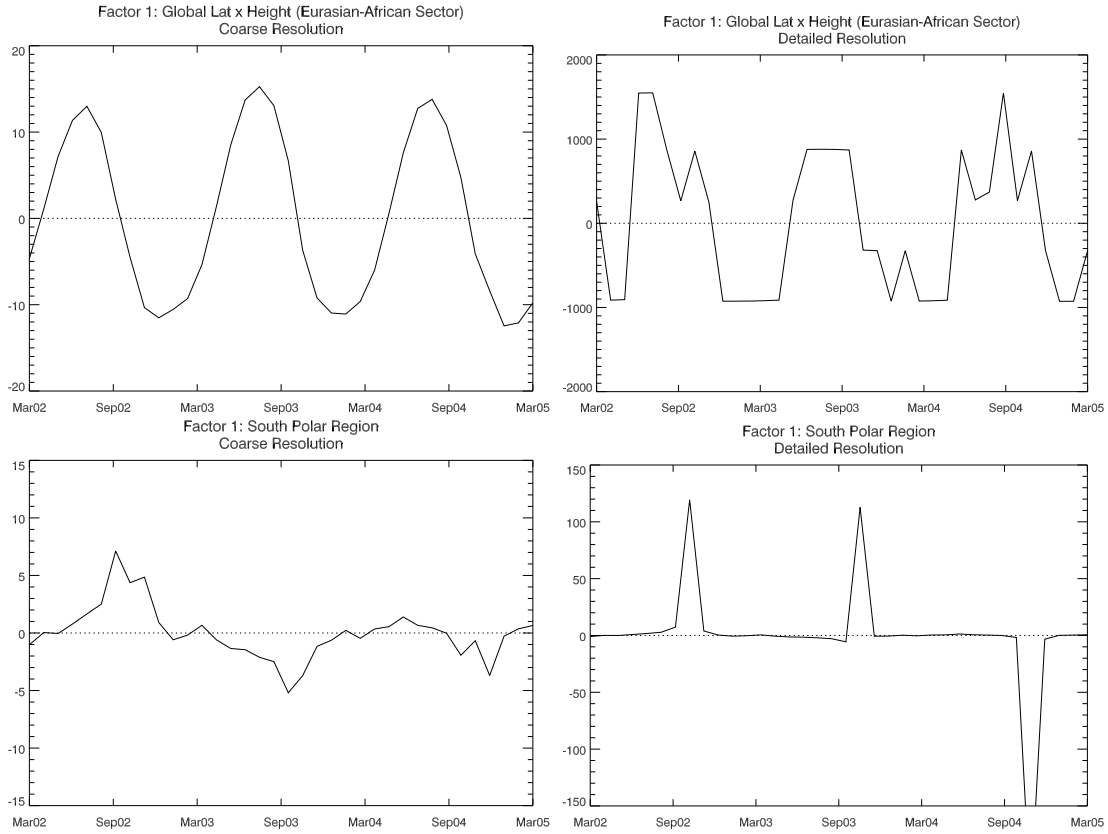


Figure 9.2: Please note in both rows the different y-ranges between the plots resulting from data sets with coarse (left) and detailed (right) resolution.
Top: First factor (according to Bartlett) of the Eurasian-African latitude $\times$ height slice, which was centered to the 3-year mean. While a smooth annual cycle (3 years) is given in the left side graph, which is the outcome of the data set with the coarse resolution, the investigation of the detailed data set resulted in a jagged appearance for the same factor. Anyhow, the annual cycle is still discernible.
Bottom: First factor (according to Bartlett) of the South Polar region, which was centered to monthly means. A similar jagged appearance of the factor resulted from the detailed resolution. The second peak in the right side graph even follows no more the corresponding one of the coarse resolution in the left side graph.

Iterative Principal Factor Analysis								
	Global Lat×Height		Global Maps		South Polar Region		Tropical Region	
Deviation Score Matrix	3-Year Mean	Monthly Mean	3-Year Mean	Monthly Mean	3-Year Mean	Monthly Mean	3-Year Mean	Monthly Mean
General and Method Specific Values								
maximal # of factors (until all ψ_{ii} are positive)	4	8	6	7	11	9	17	13
total variance of max. # of factors given in %	85.14	76.09	81.26	72.71	99.61	93.95	97.97	97.20
# of selected factors	**3**	**6**	**3**	**5**	**2**	**6**	**4**	**5**
total variance of selected # of factors given in %	82.74	68.51	70.83	63.65	95.93	89.20	91.69	92.51
steps until required accuracy* is given	3	3	3	3	3	3	3	3
s (degrees of freedom: $s > 0$)	738	624	987	898	778	624	699	661
* For each element $\max \lvert \boldsymbol{\Psi}_{\text{step i}} - \boldsymbol{\Psi}_{\text{step (i−1)}} \rvert < 0.05$ and $\max \lvert \boldsymbol{A}_{\text{step i}} - \boldsymbol{A}_{\text{step (i−1)}} \rvert < 0.05$								

Table 9.2: Problems of iterative principal factor analysis technique occurring during the calculation process with the four selected atmospheric data sets of CHAMP RO temperatures.

huge amount of the total variance.

Nevertheless, this technique showed differences between the two kinds of data centering. For monthly mean corrected data sets, two or three times as many factors had to be extracted to explain approximately the same amount of total variance as it was given by 3-year mean corrected data sets. Still taking into account this fact, the deviation score matrices, which were adjusted by the 3-year mean, then explained a total variance of more than 70.8 % in any case while the deviation score matrices, which were adjusted by the monthly means, only came up to more than 63.6 %. These monthly means corrected matrices did no longer contain the annual cycle of temperature variation (seasons), which dominated the first factors of 3-year mean corrected matrices. Thus, unique variances were stronger pronounced for monthly mean centered data than for 3-year mean centered data. An exception to this rule is given by the tropical data set, where the factor analysis technique succeeds for both deviation score matrices in explaining about the same amount of total variance. Anyhow, this makes sense, as the annual cycle of temperature variation is not very pronounced in the tropics.

Beside the restriction concerning the maximal number of factors that can be extracted

with the iterative principal factor analysis technique so that all ψ_{ii} are positive, constraints are also given by the number of degrees of freedom, dependent on the number of variables p and the number of selected factors k. This context is given by equation (7.6) in Section 7.1. The degree of freedom s is required to be positive, so that there are more equations than parameters. The theoretical limit of the requirement $s > 0$ is approached when $k < 34$ (if $k \geq 34$, then $s < 0$). Since in the investigated data sets the maximal $k = 17$, this requirement is always fulfilled.

The last value quoted in Table 9.2 is the number of steps during the iterative process that were necessary to obtain the claimed accuracy. After every single iteration step, the difference between each element of the "old" and recalculated loading matrix $\boldsymbol{A}$ and the specific variance matrix $\boldsymbol{\Psi}$ was formed, to determine the then achieved accuracy. The iteration was repeated until the largest difference at all was less than 0.05, which was generally achieved after only three iteration steps.

9.2.2 True Factor Analysis and CHAMP RO Data

As iterative principal factor analysis demands a certain maximal number of extracted factors so that all ψ_{ii} are greater equal to zero, true factor analysis needs a minimal number of extracted factors to fulfill mathematical requirements.

Here, the estimate of $\boldsymbol{\Psi}$, which is systematically biased (Reyment and Jöreskog 1993) has to be multiplied by the scalar θ. This value is the average of the $(p - k)$ smallest eigenvalues of the decomposed auxiliary matrix $\boldsymbol{R}^*$ (cf., equation (7.25)) and, according to Reyment and Jöreskog (1993), it has to be less than one, which goes hand in hand with the must to select a certain number of k factors. Surprisingly, in some of the selected cases the number of necessarily extracted factors to fulfill this requirement is quite high (see first row in Table 9.3). In the average, twice as many factors had to be extracted for monthly mean centered data sets, topping in $k = 24$ factors to be selected for the south polar and the tropical region that is more than the half of the $p = 42$ objects of the matrix investigated. Furthermore, the method did not succeed in allocation common and unique variances. Caused by this fact, the total variances explained approached 100 % for the south polar and tropical region (as well for 3-year mean centered data matrices) and after all 90 % for the Eurasian-African latitude $\times$ heigth slice, respectively (meaning that nearly no unique variances remainded). Furthermore, the 24 required factors for monthly mean centered south polar and tropical region data lead to quite a small (around 0.20) weighting factor $\hat{\theta}$, while in all other cases, $\hat{\theta}$ ranges around 0.90.

To estimate the factor loadings, true factor analysis makes use of an auxiliary matrix $\boldsymbol{R}^*$ (cf., equation (7.24) in Section 7.2.2). To obtain $\boldsymbol{R}^*$, the square root of the inverted correlation matrix has to be extracted, which turned up to be the root cause of all the troubles with this technique, with the outcome of the impossibility to apply true factor analysis to the global maps data set.

The reason for this problem can be found in the properties of the correlation matrix $\boldsymbol{R}$ of the investigated data set. True factor analysis is based on generalized least square

True Factor Analysis								
	Global Lat×Height		Global Maps		South Polar Region		Tropical Region	
Deviation Score Matrix	3-Year Mean	Monthly Mean	3-Year Mean	Monthly Mean	3-Year Mean	Monthly Mean	3-Year Mean	Monthly Mean
General and Method Specific Values								
minimal # of factors that $\hat{\theta} < 1$ (# of selected factors)	**6**	**12**			**11**	**24**	**13**	**24**
total variance of selected factors given in %	89.12	88.22			99.66	99.99	98.07	100
weighting factor $\hat{\theta}$ ($\hat{\theta} < 1$)	0.921	0.903			0.959	0.152	0.924	0.191
Method Specific Problems								
Calculation successful	⊗	⊗	⊖	⊖	⊗	⊗	⊗	⊗
eigenvalues of $\boldsymbol{R}$ < quoted value set to this value	0.05	0.05	0.5	0.5	6E−04	E−06	0.04	E−10
# of eigenvalues of $\boldsymbol{R}$ set to this value	23	21	36	33	17	18	12	15
maximal difference between $\boldsymbol{R}_{\text{original}}$ and $\boldsymbol{R}_{\text{rebuilt}}$	0.029	0.026	0.322	0.285	3E−04	5E−07	0.001	4E−08
# of negative eigenvalues of diag($\boldsymbol{R}^{-1}$) before setting selected eigenvalues to quoted values	16	17	5	28	11	24	30	26
# of negative eigenvalues of diag($\boldsymbol{R}^{-1}$) after setting selected eigenvalues to quoted value	0	0	5	2	0	0	0	0
⊗ ... problem occurred during the calculation procedure ⊖ ... impossibility to carry out calculation								

Table 9.3: Problems of true factor analysis technique occurring during the calculation process with the four selected atmospheric data sets of CHAMP RO temperatures.

estimations, which require the analyzed correlation matrix to be positive definite (Rigdon 1997). Among other things, this is ensured as long as all eigenvalues of the matrix are positive and the determinant is greater than zero.

This requirement is not met by the atmospheric data sets from which the determinant of the correlation matrix is equal to zero. Thus, the matrix is positive **semi**definite.

Rigdon (1997) reports on experiments of Ed Cook to cure those bad matrices. On that occasion, negative eigenvalues of a decomposed ill-natured correlation matrix were set to a small positive value (0.05). By means of these new eigenvalues (summarized in the diagonal matrix $\boldsymbol{\Lambda}_{\text{new}}$) and the eigenvectors (summarized in the matrix $\boldsymbol{U}$) of the original correlation matrix, a "new" well-behaving correlation matrix is built up using the relation

$$\boldsymbol{R}_{\text{rebuilt}} = \boldsymbol{U}\ \boldsymbol{\Lambda}_{\text{new}}\ \boldsymbol{U}'. \tag{9.1}$$

In the case of the investigated atmospheric data sets, the eigenvalues of $\boldsymbol{R}$ were positive, but the values were with the exception of the first few ones mainly very small. Following Cooks experiment, these small eigenvalues were set to a quite arbitrarily fixed larger threshold value.

At first, for each data set a threshold value of 0.05 was taken and all eigenvalues less than 0.05 were set to this value. In the following, the inverse of the matrix $\boldsymbol{R}_{\text{new}}$ was calculated and examined with regard to negative values in the main diagonal. If these values passed the test (that is to say that they all were positive), a smaller threshold value was taken and the procedure repeated, until negative values in the main diagonal of the inverse of $\boldsymbol{R}_{\text{new}}$ appeared. The different threshold values can be found in Table 9.3 as well as the number of eigenvalues being affected by this value, which range between 12 and 36 (out of 42).

To check, whether the differences between $\boldsymbol{R}_{\text{original}}$ and $\boldsymbol{R}_{\text{new}}$ do not surmount a tolerable dimension, the largest deviation of any element between the two matrices was quoted (mostly, the differences were far less than 0.03, see "method specific problems" in Table 9.3).

Except for the global maps data set, this trick turned out to be very successful. Whereas between 11 and 30 items of diag $\left(\boldsymbol{R}^{-1}_{\text{original}}\right)$ showed negative values, all of them were positive after applying Cook's trick (cf., last two rows of Table 9.3).

Unfortunately, the data sets of the global maps turned up to be resistant to this trick and remained ill-natured even though the threshold value was increased. That's why it was not possible to apply true factor analysis to these data sets.

9.2.3 Maximum Likelihood Factor Analysis and CHAMP RO Data

In contrast to the other three techniques, maximum likelihood factor analysis requires the investigated data to be normally (Mardia et al. 1979), in the opinion of Weber (1974) multi-normally, distributed.

According to von Storch and Zwiers (2003), temperature is approximately normally distributed, particularly if averaged over a certain period, which holds true of the investigated CHAMP RO data.

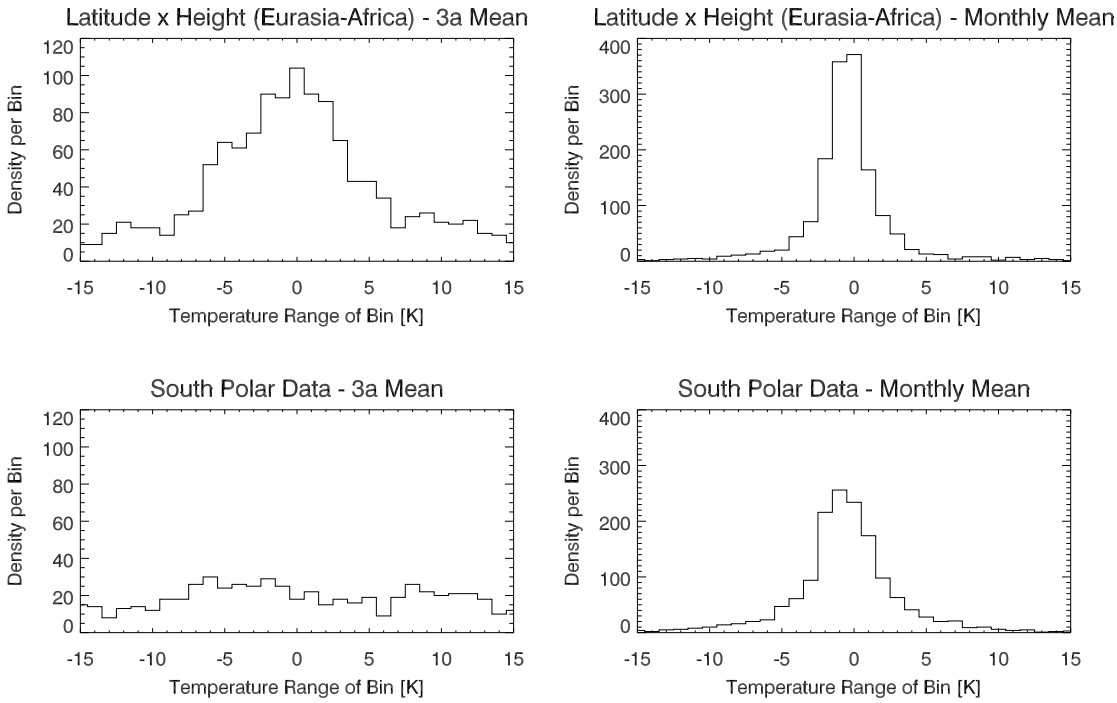


Figure 9.3: Frequency distribution of two selected fields. With the exception of the 3-year centered south polar data set, the investigated data are more or less normally distributed. Top: Eurasian-African data set 3-year (left) and monthly mean (right) centered. Bottom: South polar data set 3-year (left) and monthly mean (right) centered. The 3-year mean centered south polar data set resulted in quite a uniform distribution, while the monthly mean centered data remained normally distributed.

The distributions of the four investigated data sets show a more or less normal distribution, which is better pronounced by monthly mean centered data, even though they tend to fall to the left side (slightly positive skewness). Furthermore, the kurtosis of monthly mean centered data is greater than the one of 3-year mean centered data (cf., Figure 9.3).

The minor decrease of the latter data sets may be given due to larger annual temperature variation, where higher- or lower-than-average temperatures strongly influence the distribution (by decreasing the kurtosis). At first glance, the rather uniform distribution of the 3-year centered south polar data set causes amazement. But as the seasonal temperature variations in polar areas are quite pronounced due to the change of polar summer and winter, larger density of absolutely higher deviations seem to make sense.

Even though the theoretically required normal distributions of the investigated data seemed to be given, problems (see Table 9.4) occurred during the calculation procedure.

As the results of iterative principal factor analysis were used as starting values for this factor analysis technique, the same number of factors was taken for the beginning (see first row of Table 9.4). This yielded senseless values of more than 100 % (ranging from

Maximum Likelihood Factor Analysis								
	Global Lat×Height		Global Maps		South Polar Region		Tropical Region	
Deviation Score Matrix	3-Year Mean	Monthly Mean	3-Year Mean	Monthly Mean	3-Year Mean	Monthly Mean	3-Year Mean	Monthly Mean
General and Method Specific Values								
maximal # of factors as used for iterative principal FA	3	6	3	5	2	6	4	5
total variance of that # of factors given in %	208.9	100.7	122.5	3044	109.3	105.6	155.0	214.0
# of selected factors (so that total variance $< 100\,\%$)	**2**	**4**	**2**	**4**	**1**	**3**	**2**	**2**
total variance of selected # of factors given in %	74.38	90.67	93.09	71.56	77.41	80.5	87.5	74.28
# of steps, until required accuracy ($\max \lvert \boldsymbol{\Psi}_{\text{step i}} - \boldsymbol{\Psi}_{\text{step (i}-1)} \rvert < 0.05$ and $\max \lvert \boldsymbol{A}_{\text{step i}} - \boldsymbol{A}_{\text{step (i}-1)} \rvert < 0.05$)	6	62	15	41	2	27	12	6
Method Specific Problems								
tot. variance $> 100\,\%$ in case that # of eigenvalues equal to iterat. PFA		⊗	⊗		⊗	⊗		
1^{st} factor explains $> 100\,\%$ of tot. variance (when # of eigenvalues = iterat. PFA)	⊗			⊗			⊗	⊗
1^{st} factor does not explain largest part of tot. variance		⊗						
communality $h^2 > 1$ for single variables	⊗	⊗	⊗	⊗		⊗	⊗	⊗
⊗ ... problem occurred during the calculation procedure								

Table 9.4: Problems of maximum likelihood factor analysis technique occurring during the calculation process with the four selected atmospheric data sets of CHAMP RO temperatures.

101 % to 3 044 %) for the total variances explained in all cases. To obtain meaningful results, the number of extracted factors was decreased in the following until the total variance explained remained under 100 %, which was mostly achieved by bisecting the number used for iterative principal factor analysis. In one case, namely the 3-year mean centered south polar data set, only one factor could be extracted, so that the total variance remained below 100 % (with this factor, 77 % were explained).

In general, only few factors (between one and four) were needed to explain 72 % to 93 % of the total variance (see row three and four in Table 9.4).

Such as iterative principal factor analysis, maximum likelihood factor analysis makes use of an iterative procedure (EM-algorithm, cf., Section 7.2.3) to achieve the final results. The required accuracy was defined similarly to iterative principal factor analysis. While just few iteration steps were necessary in iterative principal factor analysis (namely three), the implemented EM-algorithm for maximum likelihood needed much more of them (between two for one selected factor up to 62 for four selected factors). The number of iteration steps to achieve the required accuracy increased considerably with a larger number of selected factors (see Table 9.5).

# of Factors Extracted	# of Iteration Steps of EM-Algorithm
1	2
2	6; 6; 12; 15
3	27
4	41; 62

Table 9.5: Number of necessary iterative steps in maximum likelihood factor analysis to achieve the required accuracy.

Beside the problems discussed above, maximum likelihood showed one more peculiarity: In one case, namely the monthly mean centered Eurasian-African latitude × height slice, the first of the four extracted factors did not explain the largest part of the variance as it is expected in general (this was as well valid for the rotated factor). Furthermore, the fourth extracted factor again explained more variance than the third one. In fact, the first factor explained 19 %, the second factor 40 %, the third factor 12 %, and the fourth factor again 20 %. And finally, the communalities for single variables surmounted one (standing again for more than 100 %) for most of the data sets (see last row of Table 9.4).

Summing up, maximum likelihood factor analysis (basing on the described EM-algorithm) turned out to be not a suitable technique for the selected investigated atmospheric data fields.

9.2.4 Centroid Factor Analysis and CHAMP RO Data

Because of the arbitrariness of centroid factor analysis, which was discussed in Section 7.2.4, no great hopes were placed in this technique. Nevertheless, some interesting results

should be mentioned.

Let us start again with the number of extracted factors, which are shown in the first row of Table 9.6. Like in iterative principal factor analysis, the maximal number of factors that can be extracted, depends on the ψ_{ii}, which have to be positive. Due to this fact, a maximum of three factors may be selected for the monthly mean centered global maps, while for the remaining data sets only one or two factors could be extracted at a time. But in contrast to iterative principal and maximum likelihood factor analysis, where even a small number of extracted factors explained quite a large amount of the total variance, centroid factor analysis did not succeed in this.

Above all, for the two global data sets, only less than 30 % (in one case less than 8 %) of the total variance could be explained. From these small values it follows that most of the variance given in the data can be found in $\mathbf{\Psi}$, which contains the unique variances (those variances, which are solely influenced by one variable that is to say by one grid point in the case of the atmospheric data sets).

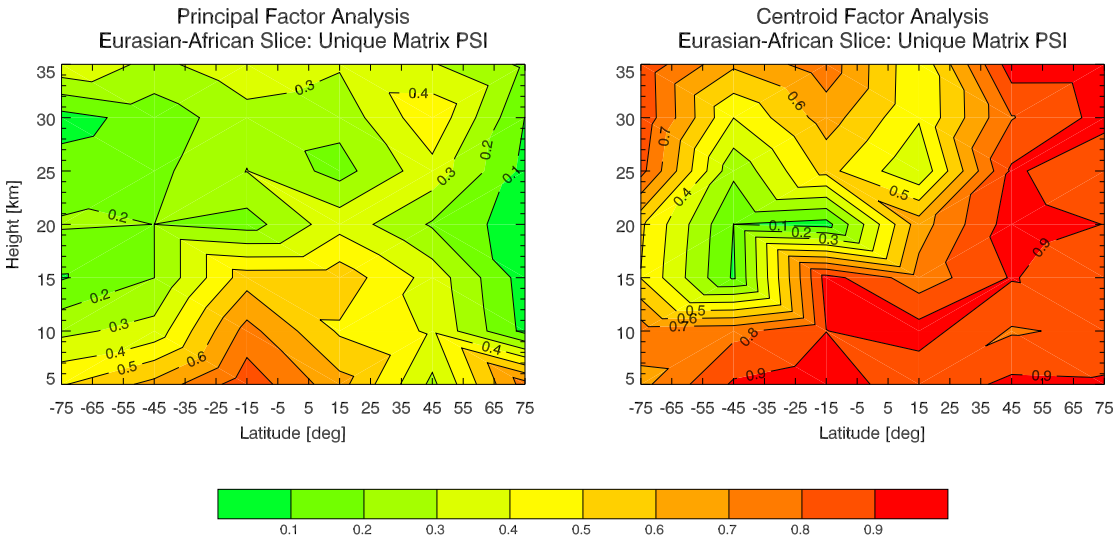


Figure 9.4: Differences in the unique matrix $\mathbf{\Psi}$ of the monthly mean centered Eurasian-African data set: While applying iterative principal factor analysis, most of the variance is explained by the specific factors so that nearly no variance is left to be explained by the unique factors of $\mathbf{\Psi}$ (left side graph), centroid factor analysis packs most of the variance in the matrix $\mathbf{\Psi}$ (right side graph).

Figure 9.4 shows the unique variance matrices $\mathbf{\Psi}$ derived from the monthly mean centered Eurasian-African data set for both, iterative principal factor analysis and centroid factor analysis. While the iterative principal factor analysis derived matrix is dominated by green colors, standing for small values of $\mathbf{\Psi}$, in the centroid factor analysis results red hues, standing for large values of $\mathbf{\Psi}$. Larger values of $\mathbf{\Psi}$ were found by centroid factor analysis in general for both global data sets (cf., last row of Table 9.6).

The two regional data sets (south polar and tropical area) do not show this feature. Their unique variance matrices are again limited pronounced and the total variance ex-

plained by the factors ranges between 60 % and 78 %, which is comparable to the results of the other three applied factor analysis methods.

Like for iterative principal FA and maximum likelihood FA, an iterative calculation process was implemented to improve the results (cf., Section 7.2.4). The calculation was repeated until the communalities approached quite a stable value (that is to say that the largest difference of any value of the communalities was less than 0.1 from one step to another. This value was fixed by a trial-and-error technique). This goal was reached after eight steps at the latest (mostly three steps sufficed).

For one data set, the monthly mean centered Eurasian-African slice, centroid factor analysis, like maximum likelihood factor analysis, did not succeed in extracting the highest variance with the first factor. The three extracted factors explained 5 %, 17 %, and 8 %, respectively.

Furthermore, no factor rotation was possible for three data sets (3-year mean centered Eurasian-African slice, global maps, and south polar region), as only one factor at all was extractable to fulfill the mathematical requirement of positive ψ_{ii} (see Table 9.6).

Centroid Factor Analysis								
	Global Lat×Height		Global Maps		South Polar Region		Tropical Region	
Deviation Score Matrix	3-Year Mean	Monthly Mean	3-Year Mean	Monthly Mean	3-Year Mean	Monthly Mean	3-Year Mean	Monthly Mean
General and Method Specific Values								
max. # of factors that all ψ_{ii} are positive (selecetd # of factors)	**1**	**3**	**1**	**3**	**1**	**2**	**2**	**2**
total variance of selected factors given in %	7.99	29.43	14.0	15.03	78.27	59.87	77.70	67.86
steps until required accuracy	4	8	3	4	3	3	3	3
maximal Δh^2 between last two steps	0.006	0.031	0.030	0.082	0.015	0.069	0.096	0.072
Method Specific Problems								
1^{st} factor does not explain largest part of tot. variance		⊗						
only 1 factor selectable (no rotation possible)	⊗		⊗		⊗			
all values $\psi_{ii} > 0.6$ (most values even around 0.9	⊗	⊗	⊗	⊗				
⊗ ... problem occurred during the calculation procedure								

Table 9.6: Problems of centroid factor analysis technique occurring during the calculation process with the four selected atmospheric data sets of CHAMP RO temperatures.

9.3 Differences Between Coarse and Detailed Resolutions

(Authors: B.C. Lackner and B. Pirscher)

The stability of the FA and PCA was investigated by means of the results yielded from different resolutions of the four analyzed temperature fields. Therefore, the factors/principal components, the loadings/coefficients, and the data reconstructions were compared separately for the sample correlation matrix $\boldsymbol{R}$ and the sample covariance matrix $\boldsymbol{S}$ (only PCA).

Table 9.7 shows the results of the calculations carried out for the atmospheric fields investigated in this work. As can be seen, sometimes the stability of the FA/PCA breaks down; a closer examination shows that the problems always occur for the more detailed resolved temperature fields. There, the number of variables (grid points) is considerably larger than the number of objects and the methods cannot resolve the true temporal character of the temperature anomalies at the grid points. A detailed analysis of the problems, which appeared in the calculation processes, are given in the following.

Eurasian-African Slice: To recall the different height grids in the Eurasian-African sector, the matrix dimensions of the temperature fields should be given once more: The temperatures located in the longitudinal sector of Eurasia-Africa (20°W to 70°E) are given in a latitudinal resolution of 30° (6 latitudinal bins) and a vertical resolution of 2 km and 5 km from ground to a height of 34 km/35 km yielding 102 grid points in case of the finer vertical resolution and 42 grid points in case of the coarser vertical resolution. Due to the observation period of 36 months, the data are given in a (36×102)-matrix and in a (36×42)-matrix, respectively.

Table 9.7 shows that in the case of the finer resolved temperature fields, FA and PCA cannot be always performed without difficulties, whereas the coarser ones never show a problem.

According to that, the stability of iterative principal FA gets lost in case of the detailed resolution of 3-year as well as of monthly mean centered data. Surprisingly, the factor loadings do not show striking deviations between the two different resolutions. However, the loadings contain variances (correlations) and therefore, only may vary between -1 and $+1$ due to mathematical constraints, whereas the factor scores carry the temperature information, which is given by the data set.

Figure 9.5 shows the differences between the results of the two resolutions for the 3-year mean centered data for August 2004 (all other months yield similar pictures, as well as PCA does). As mentioned above, the factor loadings of the coarse and detailed resolution, shown in the third row of the Figure, are very similar and also explain about the same amount of variance for the first factor. The far-reaching difference is caused by the factor scores, which are depicted in the second row of the Figure. The coarse resolved data (see Figure 9.5 second row left side) result in a very smooth variation of the factor scores, where the annual cycle of the temperature is clearly visible, ranging between ± 15 units. Even though the annual temperature cycle is still discernible in the factor scores of the detailed resolved data (see Figure 9.5 second row right side), the jagged

	Global Lat×Height				Global Maps				South Polar Region				Tropical Region			
Resolution	Coarse		Detailed		Coarse		Detailed		Coarse		Detailed		Coarse		Detailed	
Deviation score matrix	3-Year Mean	Monthly Mean	3-Year Mean	Monthly Mean	3-Year Mean	Monthly Mean	3-Year Mean	Monthly Mean	3-Year Mean	Monthly Mean	3-Year Mean	Monthly Mean	3-Year Mean	Monthly Mean	3-Year Mean	Monthly Mean
Sample Correlation Matrix $\boldsymbol{R}$																
Iterative Principal Factor Analysis																
Reconstruction	+	+	⊖	⊖	+	+	+	+	+	+	⊖	⊖	+	+	+	+
Factors	+	+	⊖	⊖	+	+	+	+	+	+	⊖	⊖	+	+	+	+
Loadings	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Principal Component Analysis																
Reconstruction	+	+	⊖	⊖	+	+	+	+	+	+	⊖	⊖	+	+	+	+
PCs	+	+	⊖	⊖	+	+	+	+	+	+	⊖	⊖	+	+	+	+
Coefficients	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Sample Covariance Matrix $\boldsymbol{S}$																
Principal Component Analysis																
Reconstruction	+	+	⊖	⊖	+	+	+	+	+	+	⊖	⊖	+	+	+	+
PCs	+	+	⊖	⊖	+	+	+	+	+	+	⊖	⊖	+	+	+	+
Coefficients	+	+	+	⊖	+	+	+	+	+	+	+	⊖	+	+	+	+
+ ... no problems occurred during the calculation procedure ⊖ ... problems occurred																

Table 9.7: Problems of PCA and the iterative principal FA occurring during the calculation process with the four selected atmospheric data sets of CHAMP RO temperatures in regard to the two different resolutions.

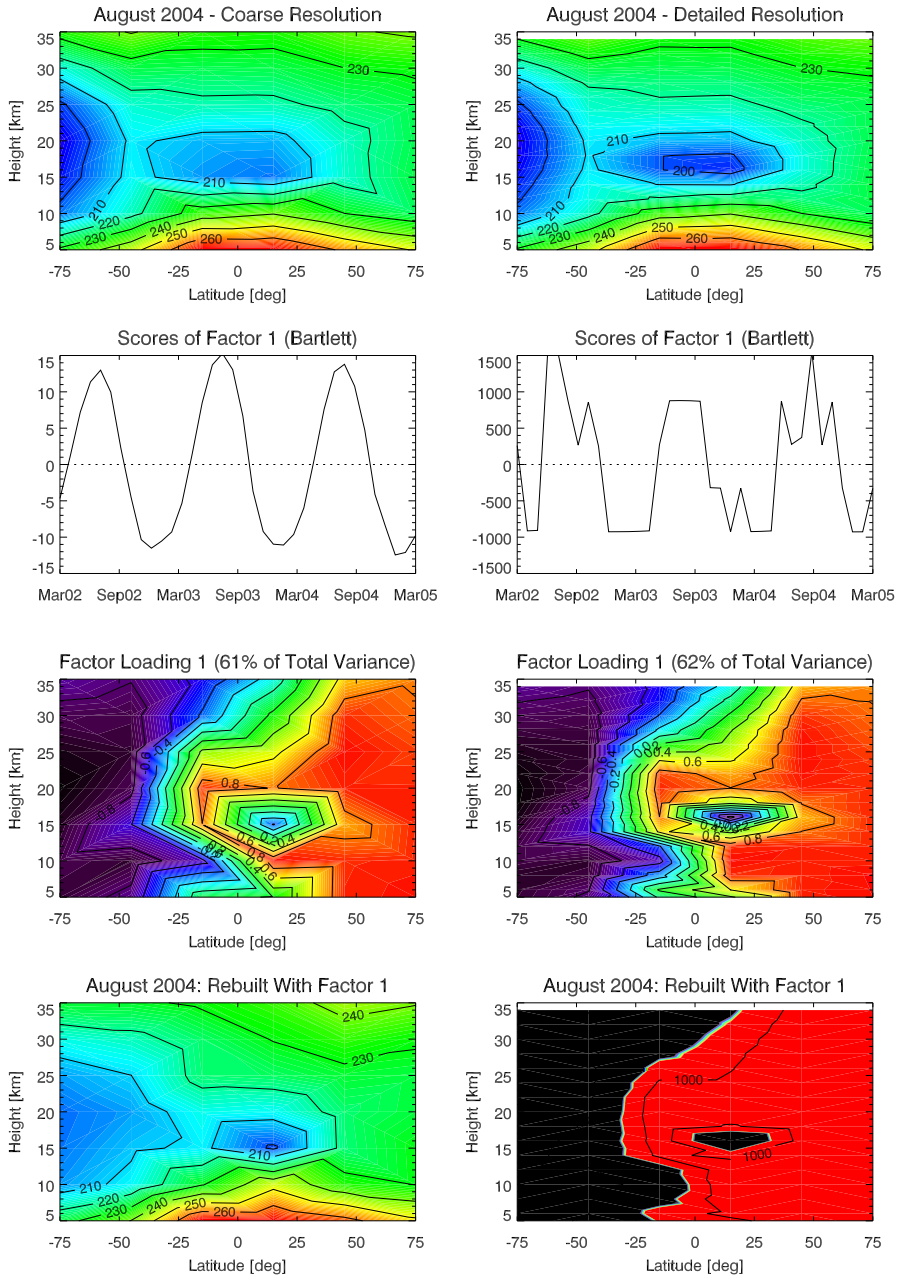


Figure 9.5: Iterative principal FA: Differences between coarse (left column) and fine (right column) resolution. The original temperature data [K] are depicted in the first row, the factor scores of the first factor in the second row, the loadings in the third row, and the reconstructed temperature data in the last row. Please mind the different y-axis range (±15 units in case of the coarse and ±1500 units in case of the fine resolution) for the factor score graphs in the second row. A detailed description of the pictures is given in the text.

appearance and the y-range of ±1 500 units, which is by two decimal places larger than for the coarse resolution, yield senseless temperature values for the reconstructed data, which are shown in the last row (right hand side). There, the vector multiplication of factor scores and loadings ($\boldsymbol{Af}$) results in temperatures between −1 320 K (black color) and +526 902 K (red color) instead of the required +200 K to +270 K, which result again from the factor scores and loadings of the coarse resolution. Comparing only the first row containing the original data and the last row with the reconstructed data, the loss of stability regarding the finer resolved data set is obvious. Concerning the coarse resolution it should be mentioned that even though only the first factor, which explains about 60 % of the total variance, was used to reconstruct the original data, a very good result was achieved.

The characteristics of factor scores, loadings, and rebuild data derived from monthly mean centered data are similar to those described above and therefore are not further discussed.

Similar to iterative principal FA, the PCA data reconstructions (generated from 30 PCs/coefficients) of the 3-year mean centered temperatures, calculated with the correlation matrix are very poor for the detailed resolution, because of the principal components. The first PC, for example, should be a sinusoidal curve representing the seasons, but it is depicted as jagged line, such as in FA. Furthermore, the amplitudes of these principal components are very large; the values of the first PC are larger than $\pm 1 \times 10^5$ units, compared to about ±50 units in case of the coarse resolution. The same result is obtained with monthly mean centered data. Again, the amplitudes of the principal components are too large ($\pm 8 \times 10^4$) units in case of the 2 km vertical resolution involving the impossibility of data reconstruction. The coefficients calculated with the correlation matrix $\boldsymbol{R}$ discover similar patterns independent of the matrix dimensions (but of course dependent on the temperature anomalies).

If the PCA results are obtained by means of the covariance matrix, the reconstructed data (again generated with 30 PCs/coefficients) of the 3-year mean centered temperatures are similar for both vertical resolutions, but if the reconstruction is restricted to the first PC/coefficients the finer vertical resolution does not show a good agreement with the original data set in some months (March 2002, September 2002, November 2002, May 2003, October 2003, November 2003, January 2004, June 2004, September 2004, November 2004, February 2005). Having a look at the first principal components it can be noticed that they take values of near to zero in these months, which should not be the case. Beside, the temporal resolution given from the principal components yields a useless magnitude of the amplitudes, which are again too large (between -5×10^5 units and 1×10^6 units).

Centering the detailed data to the annual cycle yields principal components and coefficients, which show no resemblances to the results obtained from the coarser grid. The principal components are again jagged with large amplitudes (-5×10^5 units to $+8 \times 10^5$ units), and the respective reconstruction is bad at least until the third PC/coefficients.

Global Maps: The temperatures at an altitude of 15 km are given in a latitudinal-longitudinal resolution of ($15° \times 45°$) as well as ($30° \times 45°$). The different latitudinal resolutions yield 12 and 6 latitudinal bins, respectively; the longitudinal resolution is constantly kept at 8 longitudinal bins, yielding altogether 96 grid points for the finer and 48 grid points for the coarser latitudinal resolution. Thus, the temperature anomalies are given in a (36×96)-matrix and a (36×48)-matrix arising from the 36 monthly means being analyzed.

Independently of which mean value is subtracted from the real temperatures, the FA and the PCA can be performed successfully for both resolutions.

Even though the data rebuilt with iterative principal factor analysis' derived factor scores and loadings show a good agreement with the original data, some slight differences are given according to the two resolutions, mainly for 3-year mean centered data. The total variance explained by the second factors differ by 5 %, which results that also the factor scores show small deviations. In case of the first factors, the total variances explained are as well not the same for the coarse (53 %) and detailed (47 %) resolution, but here, their factor scores nearly match perfectly.

The first factors' loadings and scores of the monthly mean centered data of the coarse resolution correspond to those of the detailed resolution and also the variances explained by the first factor are nearly the same (23 % and 21 %). Differences of a factor four appear for the factor scores of the second extracted factor, but they vanish again for the third extracted factor.

Looking at the PCA reconstructed data fields it can be realized that they also nearly look like the same when they are calculated by means of the coarser and the finer resolution. The conclusion that the individual elements (the PCs and the coefficients) are quite similar can be drawn. The theory can be corroborated by comparing the respective principal components and their coefficients and, in fact, they are very similar in all cases. So, principal component analysis is robust to the enlargement of the matrix dimension (from a (36×48))-matrix to a (36×96)-matrix).

The solely small differences between the different resolved temperature fields can be found in the amplitudes of the principal components, and in the fact that the smaller resolved field yields better resolved coefficients. The amplitudes of the principal components calculated with the correlation matrix differ for example between ± 20 units (first PC of the fine resolved field centered to the annual cycle) and ± 15 units (first PC of the coarse resolved field centered to the annual cycle). Larger amplitudes always resulted from the finer resolved temperature field.

South Polar Region: Let us call to our mind that the south polar temperature data are given as zonal mean temperatures with a 5° latitudinal resolution (from 57.5°S to 87.5°S) yielding 6 zonal bands, and a vertical resolution of 2 km (17 height levels) and 5 km (7 height levels). The finer resolved temperature field is given by a (36×102)-matrix, the coarser grid results in a (36×42)-matrix.

Table 9.7 shows that the detailed grid is too small (implying a too large data matrix) and therefore causes problems of the methods.

Similar to the Eurasian-African Slice, the problems of iterative principal factor analysis with the detailed data sets originate from faulty factor scores, which are responsible for the impossibility of a correct reconstruction of the data. An exception are the factor scores of the first factor of 3-year mean centered data, which are nearly identical for the coarse and detailed resolution and reflect the annual temperature cycle in the data (the amplitudes of the scores range between ±20 units). Only a very small deviation (of roughly 3 units) in October 2004 drops a hint that there may be a problem.

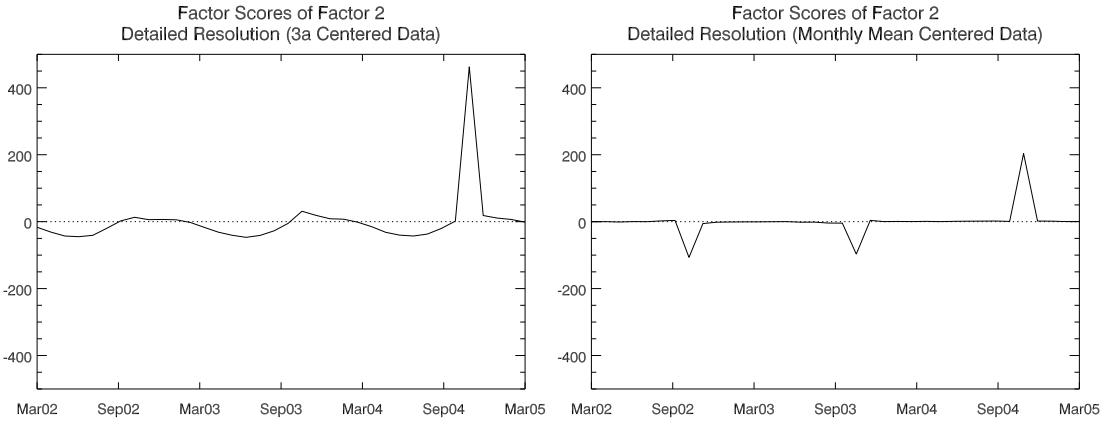


Figure 9.6: Iterative principal factor analysis problems with factor scores (of factor two) in regard to detailed resolution in the south polar region.

The left graph in Figure 9.6 shows that there really is a problem with October 2004. While up until that point of time, the scores of factor two are as well very similar to those of the coarse resolved data set, with an amplitude between about ±50 units, a singular peak, reaching more than +400 units, appears for this single unstable month and influences the reconstruction of the data in a negative way. A similar picture is given by the factor scores of the third factor with the only difference that they are mirrored relative to the zero line.

While the factor loadings of the coarse resolution still resemble those of the detailed one, the factor scores of the monthly mean centered detailed data set seem to be heavily disturbed (cf., right graph in Figure 9.6). For each October of the observation period, a peak appears, whereas the factor scores are barely pronounced for the time in between. The sign of the October 2004 peak is always opposite to the peaks of October 2002 and 2003 and reaches an amplitude of more than +200 units for the third extracted factor. Even though monthly mean centered coarse resolved data are as well mainly influenced by larger temperature anomalies in Septembers and Octobers, the factor score features of the detailed resolution prevent a successful reconstruction of the data and therefore cause the instability of the factor analysis for this atmospheric data set.

Applying PCA, the 3-year mean centered data in the south polar region yield good coefficients from the 2 km and from the 5 km vertical resolution. Both resolutions are able to detect the most important pattern arising in this region. Analyzing the respective principal components, it can as well be seen that the PCs of the fine resolved

correlation matrix are defective and cannot be compared to the PCs calculated by means of the coarser resolved temperature field. In the first case, the first three PCs are composed of a nearly constant value with a peak arising in October 2004 (approximated magnitude PC 1: 5000 units; PC 2: 9×10^4 units; PC 3: -9×10^4) units; it cannot be explained by a rational reason but has to be attributed to an instability of the PCA. The principal components calculated from the less detailed data matrix are by all means plausible (approximated amplitudes: ± 100 units). Since the principal components do not resolve the true temporal behavior of the data field in case of the fine vertical grid, the reconstruction of the real temperature data fails, like it was the case for FA.

The same problem occurs in regard to the monthly mean subtracted temperature data. In that case, the principal components exhibit a peak in October each year (most pronounced in 2004), whereas it is nearly constant in all the other months. The magnitudes of the peaks take values in the same range as when calculated from the 3-year mean centered temperatures. Equal to factor analysis, the reconstruction of the data field breaks down due to the principal components. The instability of the PCA at the fine resolved data field cannot be observed when looking at the respective coefficients.

An interesting feature can be found analyzing the 3-year centered coefficients and the respective principal components, which are based on the covariance matrix. The first PCs/coefficients of the finer resolved field seem to result from a mathematical instability of the PCA. The PCs look like the PCs calculated from the correlation matrix. They are constant until October 2004 when a huge peak (magnitude $\sim 1 \times 10^6$ units) arises and are constant again after October 2004. The respective coefficients show a pattern never seen before. Strong structures can be found from about 8 km to 18 km (especially pronounced from 67°S to 85°S) and above 20 km height (in all latitude regions) with different signs. Consequently, the reconstruction of the first pattern is useless in case of the 2 km vertical resolved data field. Comparing the second and third PCs/coefficients of the fine resolved grid to the first and second PCs/coefficients of the coarse resolved field, it can be realized that they are practically equal, except for their different signs. The opposite sign can be explained by the reverse direction of the eigenvectors, which can be chosen arbitrarily.

The investigation of the differences between the two resolutions of the monthly mean centered temperatures yields the same result as the examination of the 3-year mean corrected temperature field. The principal components and coefficients of the two different resolutions yield the same result when neglecting the first principal component and its corresponding coefficients of the fine resolved field. That means that the second fine resolved PCs/coefficients belong again to the first coarse resolved PCs/coefficients and the third fine resolved PCs/coefficients remember to the second rough resolved PCs/coefficients.

Tropical Region: The stability of the statistical methods FA and PCA in the tropical region was investigated comparing the results from a (36×77)-matrix and from a (36×42)-matrix. The matrix dimensions came from the regional temperature field, which was set at the low latitudes between 17.5°S and 17.5°N (5° zonal means yielding 7 latitudinal

regions) from a height between 12 km and 22 km (different vertical resolutions, 1 km: 11 height levels, 77 grid points and 2 km: 6 height levels, 42 grid points).

A glance at Table 9.7 shows that no problems constricted the calculation procedures, neither for FA nor for PCA.

Hence, iterative principal FA succeeded for both resolutions in extracting appropriate factor scores and loadings to rebuild the original data, even though slight differences occurred between the coarse and detailed resolved factor scores and loadings. The factor scores' amplitudes of the 3-year mean and the monthly mean centered data sets always stay within ± 10 units, no outliers appear. A general difference can be observed between the coarse and detailed resolution according to the total variance explained by each factor. While the loadings of the first extracted factor of the coarser resolved data set succeed in explaining only 59 % (3-year mean centered) and 42 % (monthly mean centered), those of the detailed data sets manage to account for 63 % and 57 %, respectively. Conversely, the loadings of the second and third extracted factors of the coarse data set explain a bit more (about 3 %) of the remaining variance than those of the detailed data set. These small differences in factor loadings and scores yield nearly identical and correct reconstruced data sets.

The comparison of the principal components and the coefficients from the two vertical resolutions in case of the correlation matrix yields that the results look like the same, regardless of which mean value was subtracted. As a result, the reconstructed time series are proper in all cases (smaller/coarser vertical resolution, 3-year mean/monthly mean subtraction). So, the factor analysis and the principal component analysis are robust to the enlargement of the data matrix in this region.

Furthermore, it attracts attention that the form of the principal components and the shape of the coefficients are similar if they are calculated with the 3-year mean or with the monthly mean subtracted data field. The only difference is that the PCs are a little bit more jagged in the first case.

The PCA calculated with the covariance matrix is also stable when enlarging the dimension of the matrix from 42 to 77 grid points. The principal components, their coefficients, and the reconstructed time series are practically indistinguishable when using the finer and the coarser vertical resolution, independent of which mean was subtracted before doing the PCA.

In the following, the PCA and FA results of the four CHAMP RO temperature fields are looked at in detail. As one goal of this work was to compare the peculiarities of the two methods according to atmospheric data sets like the ones investigated, for two selected fields, namely the Eurasian-African slice (see Section 9.4) and the south polar region (see Section 9.6) an in-depth analyze was performed, to find out whether there are differences and if so, what they look like. If not stated differently, PCA and iterative principal FA were compared. To come to the point, PCA and iterative principal FA yielded quite similar outcomes and therefore mainly the PCA results were used to verify the interpretations with plots. Provided that deviations between PCA and iterative principal FA were given, this is separately stated.

9.4 Temperature Data in the Eurasian-African Sector

(Authors: B.C. Lackner and B. Pirscher)

9.4.1 PCA/FA of 3-Year Mean Subtracted Temperature Anomalies in the Eurasian-African Sector

The investigation of the data set follows the left branch of Figure 9.1 that implies that the elimination of the 3-year mean is done for the temperature fields before doing the PCA/FA. The statistical methods were applied to the sample correlation matrix and, in case of PCA, to the sample covariance matrix.

Number of Factors Extracted: Table 9.8 shows the number of factors given from different selection rules applied to the sample correlation matrix $\boldsymbol{R}$ and the sample covariance matrix $\boldsymbol{S}$ as well as the results of iterative principal factor analysis, which are given due to mathematical constraints (cf., Section 9.2.1).

Method	Cum. Var.>90 %	Kaiser's rule	Scree Test	LEV-Test	FA
$\boldsymbol{R}$, 3-Year Mean	6	5 (6)	4	4	3 (4)
$\boldsymbol{S}$, 3-Year Mean	2	3 (3)	3	5	–

Table 9.8: Number k of factors following from different selection rules applied to PCA as well as the results of iterative principal factor analysis.

The first column represents the number of factors given from the cumulative percentage of total variance, which was selected to be greater than 90 %. The second column shows the results following from the Kaiser's rule ($\#\{k \in \mathbb{N} | \lambda_k > 1\}$ in case of the correlation matrix and $\#\{k \in \mathbb{N} | \lambda_k > \bar{\lambda}\}$ for the covariance matrix) and the modified Kaiser's rule ($\#\{k \in \mathbb{N} | \lambda_k > 0.7\}$ and $\#\{k \in \mathbb{N} | \lambda_k > 0.7\bar{\lambda}\}$, respectively) given in the parenthesis. The third column represents the number of factors estimated from the scree plot and the fourth column the results evaluated by the LEV-diagram. In the last column, the number of factors that were extracted with iterative principal FA are given. The value given in the parenthesis is the theoretically maximal amount of factors that can be extracted to achieve a mathematically correct solution.

The number of factors estimated from the scree-test and the LEV-test are quite subjective. Figure 9.7 depicts the scree-plot (left) and the LEV-plot (right) obtained from the correlation matrix.

The "broken stick" in the scree-plot arises at eigenvalue number 4. The determination of the cut-off value by means of the LEV-plot is pretty difficult because most of the eigenvalues can be connected by a straight line, but it seems to be acceptable to take $k = 4$. The eigenvalues and their respective logarithmic eigenvalues depicted in Figure 9.7 were calculated in IDL with the subroutine "SVDC" and not with the subroutine "EIGENQL". Both procedures enable the calculation of the eigenvalues of a matrix but the results differ a little bit concerning the last few eigenvalues, whereas they are

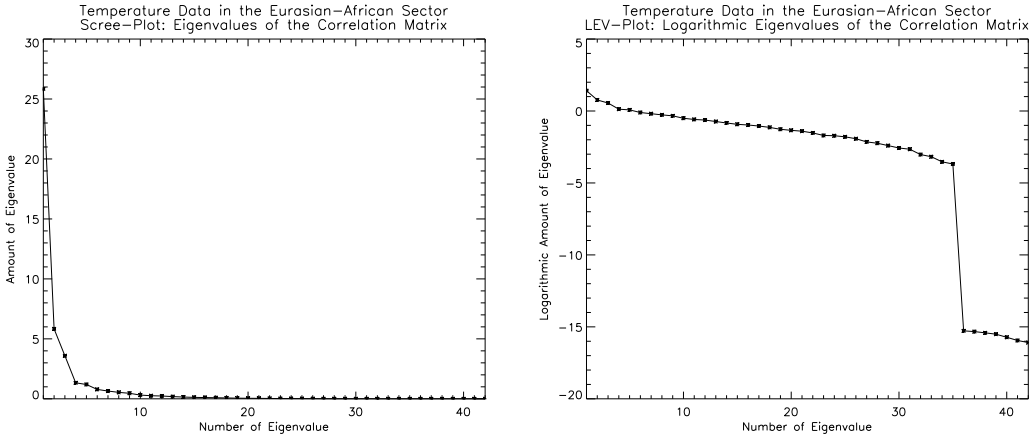


Figure 9.7: Scree-Plot (left) and LEV-Diagram (right) of the 3-year mean subtracted sample correlation matrix.

identical for the first ones. If they were calculated with the help of "EIGENQL", they became slightly negative, which is not true for positive semi definite matrices, and both, correlation matrices and covariance matrices are assigned to this kind of matrices. So, the last eigenvalues obtained from "EIGENQL" are incorrect.

Eigenvalues: The first three eigenvalues of the correlation matrix $\boldsymbol{R}$ and the covariance matrix $\boldsymbol{S}$ are shown in Table 9.9. The FA eigenvalues of $\boldsymbol{R}$ are not given by matrix decomposition but by summing up the squared factor loadings for each factor. Nevertheless, they just differ insignificantly from those of PCA, which can be put down to the fact that the values of the unique variances ψ_{ii} are very small for this data set and that therefore, the common variances in the loadings correspond with the PCA renormalized coefficients.

Matrix	Method	λ_1	λ_2	λ_3
$\boldsymbol{R}$, 3-Year Mean	PCA	25.83	5.83	3.59
	FA	25.71	5.65	3.39
$\boldsymbol{S}$, 3-Year Mean	PCA	1786.46	230.30	73.44

Table 9.9: Eigenvalues of the sample correlation matrix and eigenvalues derived from squared factor loadings for FA, respectively as well as eigenvalues of the sample covariance matrix.

Because of the renormalization with the square root of the eigenvalues, the renormalized coefficients take a larger range of values compared to the eigenvectors itself. The different magnitudes of the correlation and the covariance matrix based eigenvalues yield different pronounced patterns, which are visible in the renormalized coefficients ($\widetilde{\boldsymbol{a}}_j = \sqrt{\lambda_j}\boldsymbol{a}_j$) and in the renormalized principal components ($\widetilde{\boldsymbol{f}}_j = \boldsymbol{f}_j/\sqrt{\lambda_j}$). First ones are more prominent in case of the covariance matrix, second ones are more distinctive

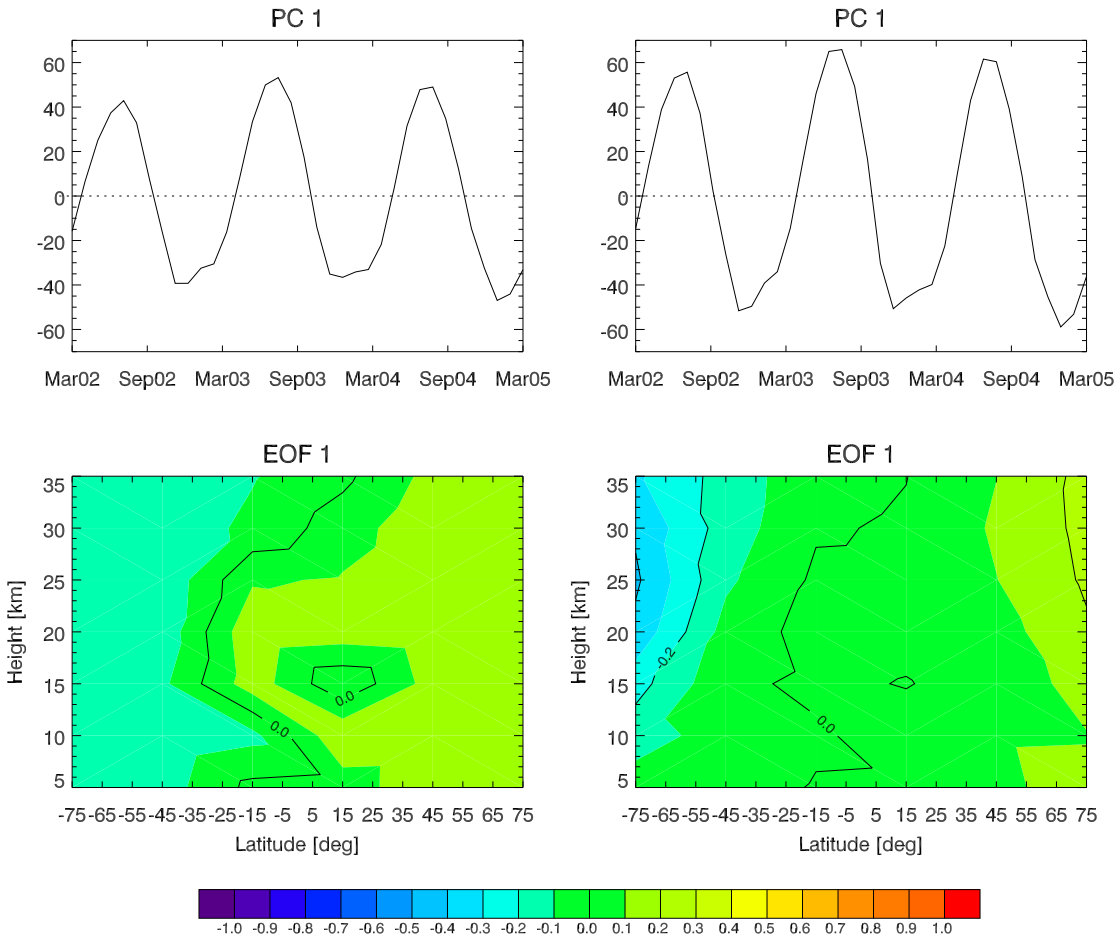


Figure 9.8: First principal component (top) and corresponding coefficient (bottom), calculated after the elimination of the 3-year mean by means of the correlation matrix (left) and the covariance matrix (right).

in case of the correlation matrix.

A comparison between the not normalized coefficients (eigenvectors) and the renormalized coefficients can be drawn when looking at Figures 9.8 and 9.9. The first one depicts the not normalized and the second one the renormalized results. The range of y-axis in regard to the principal components and the range covered by the color bar are the same, which results in very small magnitudes of the renormalized principal components (particularly in case of the covariance matrix) and in a very uniform picture of the eigenvectors.

PCA Differences When Applying the Sample Correlation and Covariance Matrix

An objective comparison between the results given from the correlation matrix and from the covariance matrix is only valid if the results are not normalized. The outcomes of

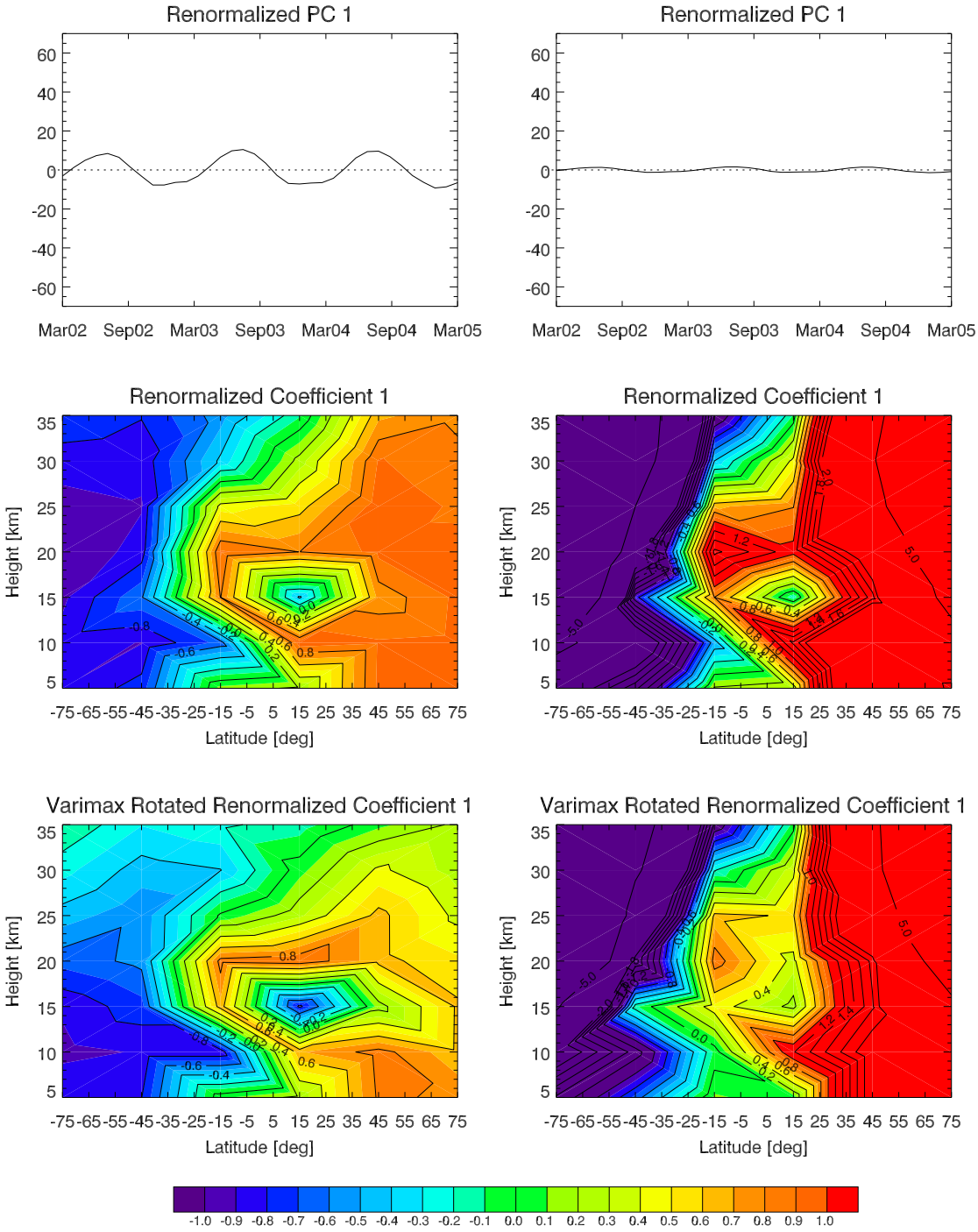


Figure 9.9: First renormalized principal component (top), the corresponding coefficient (middle), and the varimax rotated coefficient (bottom), calculated after the elimination of the 3-year mean by means of the correlation matrix (left) and the covariance matrix (right).

the respective calculation processes are shown in Figure 9.8 (left: correlation matrix, right: covariance matrix; top: PCs, bottom: eigenvectors). Comparing the principal components and the coefficients, it can be noticed that generally both show the same shape; the principal components exhibit a sinusoidal cycle with maxima and minima approximately arising at the same time and the coefficients feature a bipolar structure, where the southern and the northern hemisphere are of opposite signs.

To perform a closer examination, differences between the respective principal components and between the respective coefficients were calculated, the results are depicted in Figure 9.10 (principal components, left, coefficients, right). The comparison shows that the amplitude of the correlation matrix based PC is a little bit smaller compared to the covariance matrix based PC, the difference (correlation matrix minus covariance matrix) amounts up to ±15 units. The difference-plot concerning the eigenvectors depicts that the pattern found from the correlation matrix is stronger at high southern latitudes above a height of 16 km and at the low southern and northern latitudes between a height of 10 km and 25 km. The coefficients calculated by means of the covariance matrix are stronger at high and mid southern latitudes, particularly at lower heights. Generally, the deviations are relatively small, the maximum absolute difference contributes to $\max(|\boldsymbol{a}_1^{\text{corr}} - \boldsymbol{a}_1^{\text{covar}}|) = 0.23$. The locations of these most pronounced anomalies are situated at the high southern latitudes between a height of 25 km and 30 km (positive) and at high and mid southern latitudes between a height of 5 km and 15 km (negative). Some positive deviations, which cannot be noticed because of the large range of the colorbar are situated at the high northern latitudes where they always remain smaller than 0.1.

It can be concluded that both calculation processes detect the same pattern, which can only be distinguished in regard to the amount of their values.

Comparison Between PCA and FA Applying the Sample Correlation Matrix

To examine, how well the PCA results actually correspond with those of FA, difference plots for the first two components/factors and the coefficients/loadings where created for this atmospheric data set. The PCA results were compared with iterative principal FA and true FA, as these two techniques yielded mathematical correct solutions.

Figure 9.11 shows the differences between the renormalized principal components/-iterative principal FA factors scores and the renormalized principal coefficients/iterative principal FA loadings in regard to the first and second extracted factor. While the PCA coefficients of the first factor (upper left graph in Figure 9.11) range between −12 and +15 units, the principal FA factor scores vary between ±10 units, but both reproduce the seasonal temperature variation very well. The largest deviations (of around 5 units) between the two methods are found at the peak values of the amplitudes. A similar picture is given by the second factor, where the absolute values of the PCA coefficients again surmount those of the principal FA factor scores. In addition, a slight temporal offset between PCA components and FA factor scores (which lag behind the PCA components) is given. The differences between the PCA coefficients and principal FA

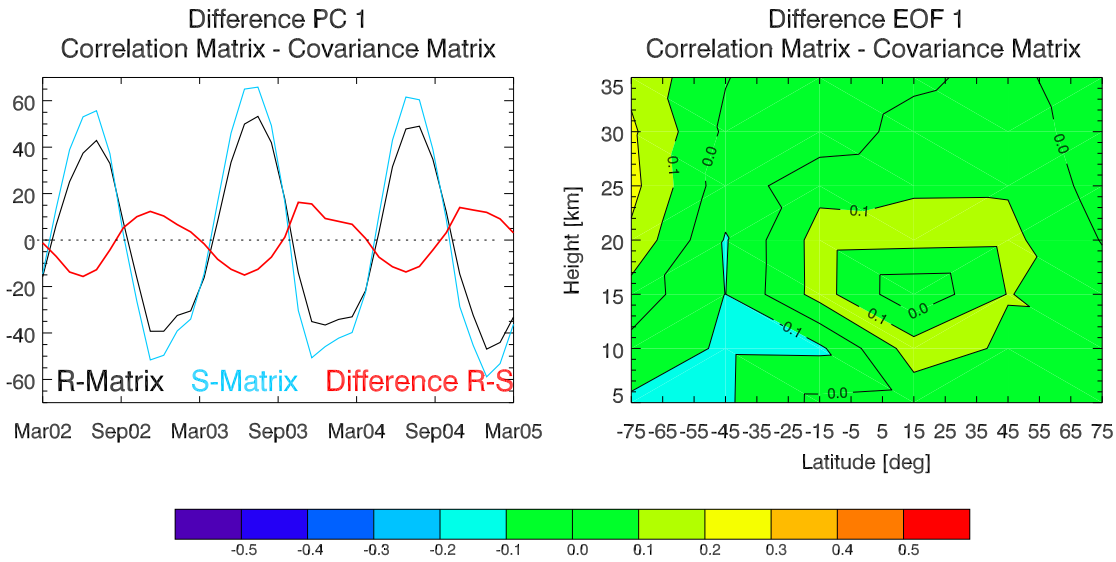


Figure 9.10: Differences between the first PC and the first EOF if they are calculated with the aid of the correlation matrix and the covariance matrix.

loadings are depicted in the second row of Figure 9.11. As the deviations between the two methods are extremely small, the color bar range was changed to ±0.04 units. As good as no differences are given for the first factors (green color), small ones in lower hight levels between the equator and 50°N, but they even do not surmount 0.04 units. Furthermore, as the first factor contributes to more than 60 % of the total variance explained, whereas the second factor only achieves around 13 %, these differences seem to be neglectible. In general, it must be kept in mind that the loadings and factor scores of FA are only influenced by the common variances, whereas the unique variances are stored in the matrix $\mathbf{\Psi}$, which is not included in this considerations. So, the smaller factor scores, which are the weights for the loadings, can be caused by the remaining unique variances.

As true factor analysis provided as well a mathematically correct solution for the Eurasian-African data set, the differences to this technique should be shortly addressed. According to the components/factor scores quantity, the deviations between PCA and true FA increase considerably, so that the differences nearly reach the same quantity as the PCA principal components (the PCA renormalized principal components range between ±10 units for the first two factors, whereas the true FA factor scores vary between ±1.5 units). In contrast to the differences of principal components/factor scores, the renormalized coefficients/factor loadings of these two methods are again very similar with differences below ±0.02 units (with the exception of a small area below 8 km height extending from 5°S to 35°N).

Since the results of iterative principal factor analysis resemble to the PCA results

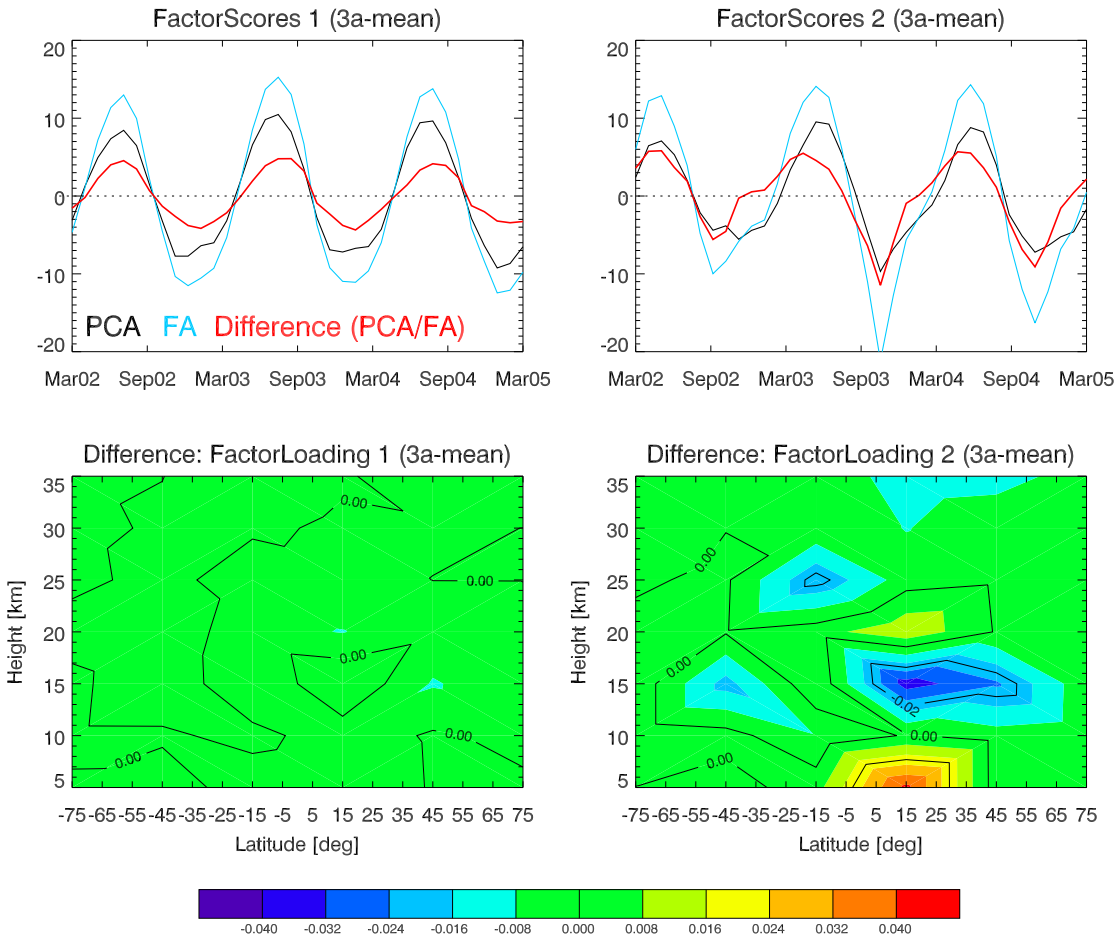


Figure 9.11: Eurasian-African sector: Differences between PCA and iterative principal FA for the first two extracted factors.

(renormalized principal components and coefficients), only principal components and their corresponding coefficients are used for the following interpretation to simplify matters (that applies also to the three remaining data sets).

Interpretation

First Principal Component and First Coefficients: Figure 9.9 depicts the first renormalized principal components and coefficients calculated by the correlation matrix (left) and the covariance matrix (right). The figures are quite similar; the southern hemisphere and the northern hemisphere show a strong variability of opposite signs.

The rotation of both coefficients changes the picture. In case of the correlation matrix the structure arising between 5°N and 25°N at an altitude between 13 km and 17 km is even more pronounced, whereas the strong structure found at the mid and high latitudes in the northern and southern hemisphere nearly disappears. Rotating the coefficients,

which are based on the covariance matrix yields a structure emphasizing the latitudinal variation, the pattern at the low latitudes vanishes.

The not rotated first coefficients account for 61.51 % (correlation matrix) and 81.30 % (covariance matrix) of the total variation, the varimax rotated first coefficients account for 34.70 % and 61.65 %, respectively.

The seasonal impact is clearly visible in the times series of the principal components, which show a sinusoidal cycle, as well as in the shape of the coefficients. Reconstructing the data with the first PC/coefficients and performing a time series at high and mid latitudes affirms the domination of the seasonal impact in the data. During one year, the temperature anomaly is positive between May and October and negative from November to April in northern polar regions, at the high southern latitudes it is of opposite sign. The magnitude of the not normalized principal components is about ± 50 units in both cases. According to von Storch and Zwiers (2003), the not normalized principal components carry the units of the data set, but in this case the components fluctuate between ± 50 K, which certainly is too much.

Interpreting the structure, which arises at the low latitudes between a height of 13 km and 17 km, a discussion of seasonal impacts on equatorial regions (from 10°S to 10°N) performed by Seidel et al. (2001) should be mentioned. They found a semiannual temperature anomaly in the troposphere with maxima arising in the equinoxes, which did not spread out in the tropopause region, where an annual cycle with maximum in August was found. The tropopause temperature was coldest at highest altitude during northern hemisphere winter. The northern hemisphere tropopause temperature remained colder during northern winter than the southern hemisphere tropopause temperature during southern winter. The tropopause height changed within the seasons, it was always higher in the winter hemisphere.

Following these results of Seidel et al. (2001), a closer examination of the pattern was done by means of an analysis of the time series of the african region between the equator and 30°N. The time series of measured temperature anomalies and the time series of the first reconstructed coefficients (calculated with the correlation matrix) are shown in Figure 9.12, top. It can be noticed that the area at an altitude of about 15 km shows hardly any variation, whereas the reconstructed first pattern shows a reversal of the signs compared to the heights below and above.

The bad vertical resolution in the region causes a blurred picture, but it is possible that the change of the tropopause height within the seasons and the cold temperature, especially in the northern tropical tropopause, are responsible for the pattern at the low northern latitudes primarily found in the first coefficients.

The low southern latitudes show another height dependent behavior; the respective time series are depicted in Figure 9.12, bottom. Again, the reconstructed patterns were calculated with the principal components and coefficients of the correlation matrix. It can be seen that above about 12 km height the temperatures follow the seasonal cycle arising in the northern hemisphere, below the temperatures exhibit a reversed sign and the anomalies agree with the cycle emerging in the southern hemisphere.

Second Principal Component and Second Coefficients: The second (renormalized) principal component and the respective second coefficients (shown in Figure 9.13) seem also to be seasonal influenced. Similar to the first coefficients, a latitudinal structure can be noticed in particular in the rotated coefficients, which points at the seasonal cycle. The principal components again show a sinusoidal cycle, the amplitudes of the not normalized PCs range between ±25 units (correlation matrix) and from −30 units to +20 units (covariance matrix). The magnitude of renormalized PCs is shown in the range between ±15 units to allow a better investigation.

The principal components and the coefficients, which arise from the covariance matrix show opposite sign compared to the structures in the first PC/coefficients and the second PC/coefficients of the correlation matrix. The explanation is that the eigenvectors and therefore the coefficients are independent of sign (due to the arbitrariness of the direction of the eigenvectors). Because of the PCs are also dependent on the direction of the eigenvectors, the reconstruction of the data field (which incorporates both PCs and EOFs) again yields the same sign.

Another already known structure can be found in the rotated coefficients arising from the covariance matrix, namely the pattern, which occurs at the low (especially northern) latitudes.

Generating time series of the reconstructed data (only with the second PC/coefficients) it becomes clear that this pattern is responsible for the temporal adjustment of the temperature anomalies. This result implicates that the patterns detected from the first and the second coefficients are not orthogonal to each other, which should be true in case of the principal component analysis. So, it disagrees to the theoretical background of PCA.

Regarding the accounted variances, the second coefficients accounts for 13.88 % of the total variance in case of the correlation matrix and 10.48 % in case of the covariance matrix. The rotated coefficients contribute 36.01 % and 26.50 %, respectively.

Third Principal Component and Third Coefficients: Figure 9.14 depicts the third renormalized principal components and the respective coefficients; the rotated renormalized coefficients are not shown because they do not facilitate the interpretation. Looking at the principal components preserves the impression that they are still based on the seasonal cycle but that there is an additional frequency present. Analyzing the corresponding coefficients, the classical seasonal structure (southern and northern high latitudes being of opposite signs) cannot be noticed.

Instead of that, an interesting feature arises at the low and mid latitudes (30°N to 30°S) above 20 km height. A similar structure is found in the data set if the seasonal cycle is eliminated. This pattern will be interpreted later on.

The variance of the third coefficients calculated by means of the correlation matrix makes 8.55 %, whereas the coefficients calculated by the covariance matrix contribute 3.34 %. The respective rotated coefficients account for 10.71 % and 3.79 % of the total variance.

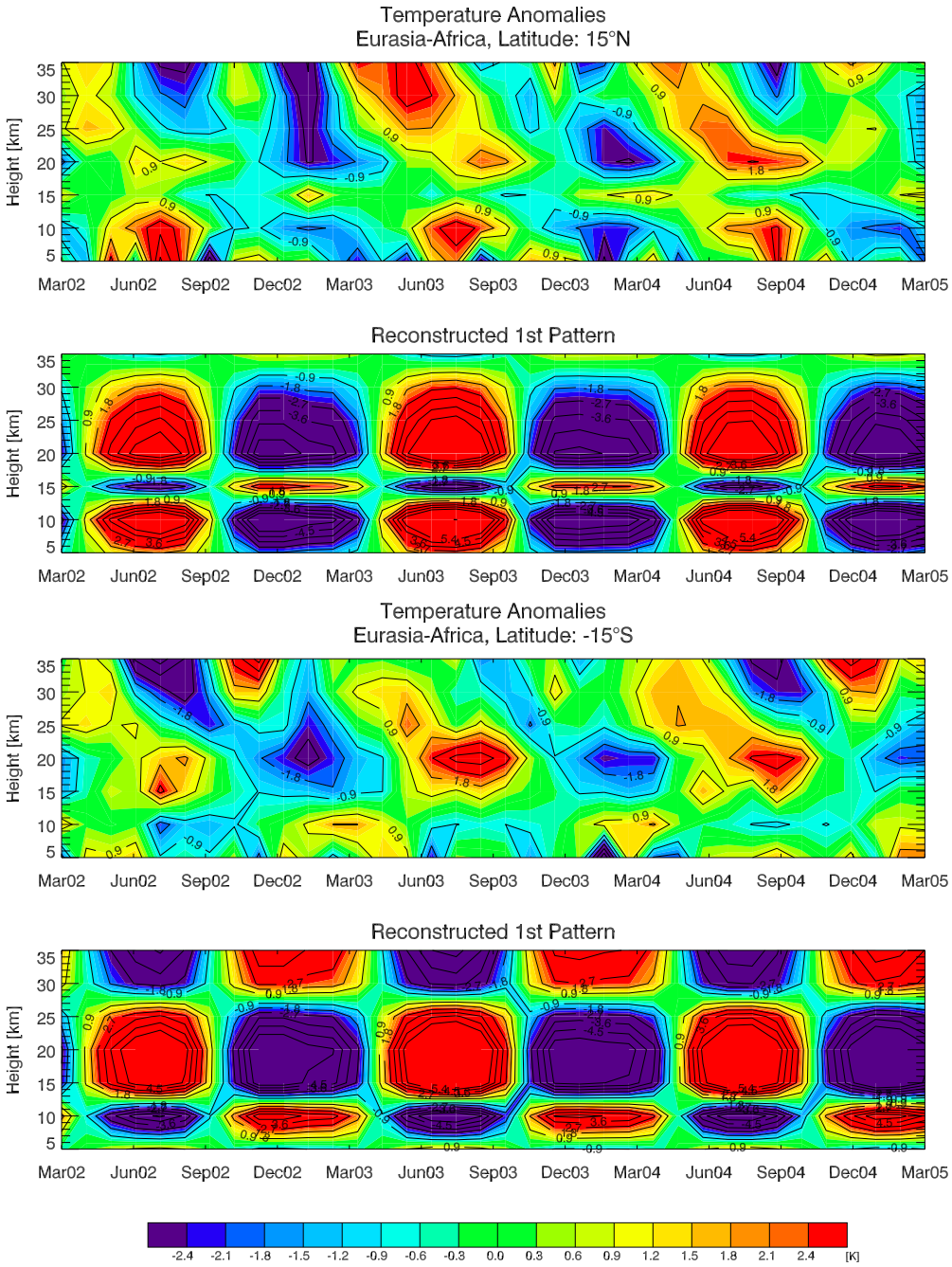


Figure 9.12: Measured temperature anomalies and reconstruction of the data set (with the first PC/coefficient of the correlation matrix) at 15°N (equator to 30°N), top, and 15°S (equator to 30°S), bottom.

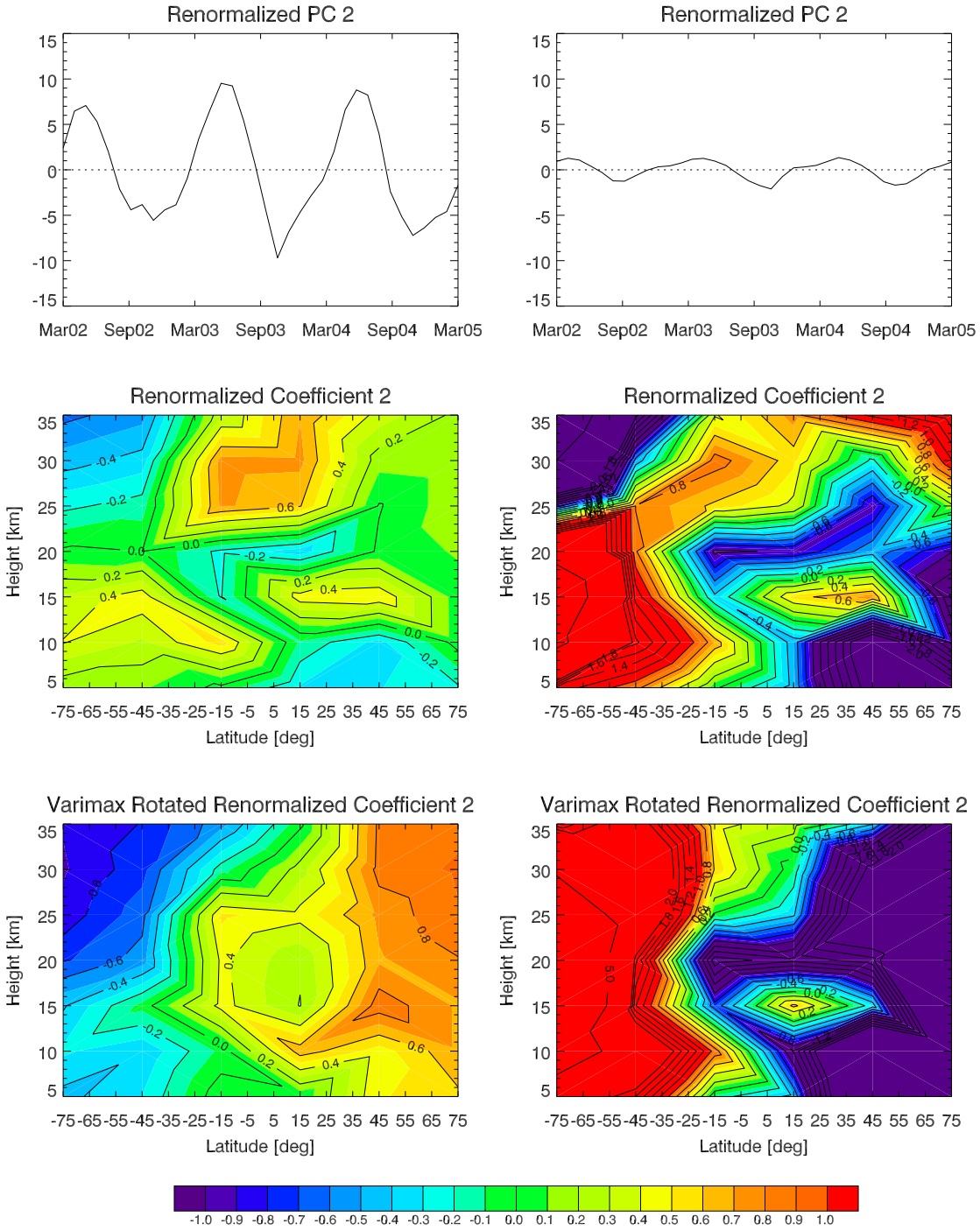


Figure 9.13: Second renormalized principal component (top), the corresponding coefficient (middle), and the varimax rotated coefficient (bottom), calculated after the elimination of the 3-year mean by means of the correlation matrix (left) and the covariance matrix (right).

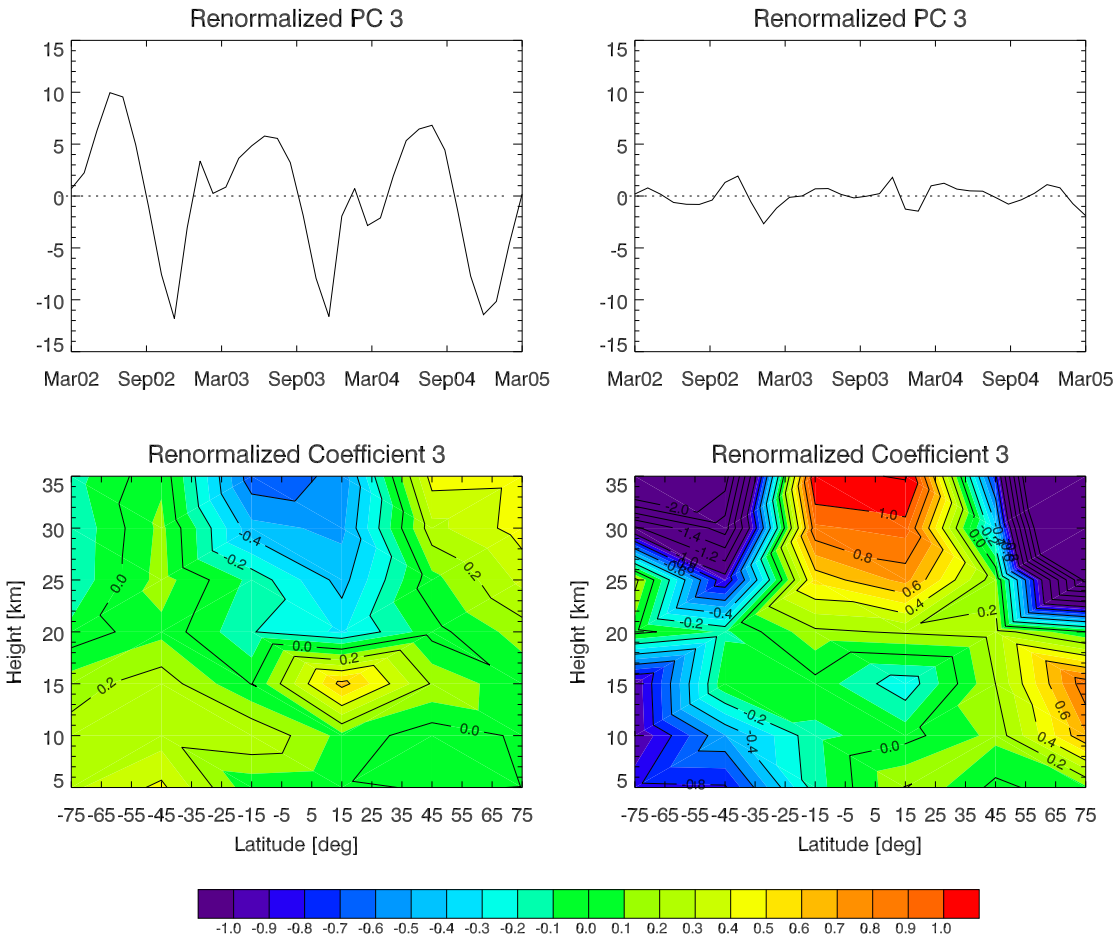


Figure 9.14: Third renormalized principal component (top) and corresponding coefficient (bottom), calculated after the elimination of the 3-year mean by means of the correlation matrix (left) and the covariance matrix (right).

Accounted Variance: A summary of the contribution on the amount of total variance of not rotated and varimax rotated coefficients/loadings is shown in Table 9.10 for both, PCA and iterative principal FA. The minor differences between PCA and iterative principal FA are reflected by the values. As can be recognized, the accounted variances of the correlation matrix and the covariance matrix are considerably larger before doing a varimax rotation.

Method	Not Rotated Coefficients/Loadings		Varimax Rotated Coefficients/Loadings	
	PCA	FA	PCA	FA
$\boldsymbol{R}$, 3-Year Mean				
$\tilde{\boldsymbol{a}}_1$	61.51 %	61.22 %	34.70 %	34.55 %
$\tilde{\boldsymbol{a}}_2$	13.88 %	13.46 %	36.01 %	37.14 %
$\tilde{\boldsymbol{a}}_3$	8.55 %	8.07 %	10.71 %	11.05 %
$\boldsymbol{S}$, 3-Year Mean				
$\tilde{\boldsymbol{a}}_1$	81.30 %		61.65 %	
$\tilde{\boldsymbol{a}}_2$	10.48 %		26.50 %	
$\tilde{\boldsymbol{a}}_3$	3.34 %		3.79 %	

Table 9.10: Accounted variances of the first three not rotated and varimax rotated coefficients/loadings.

9.4.2 PCA/FA of Monthly Mean Subtracted Temperature Anomalies in the Eurasian-African Sector

As seen in Section 9.4.1, the seasonal cycle is very dominant in the first principal components/factor scores if the mean temperature of each grid point is eliminated from the original values. To detect some structures with lower intensity, the dominant pattern was removed and the temperatures were centered to their monthly means (right branch of Figure 9.1).

Number of Factors Extracted: Table 9.11 shows the number of factors, which are "most important". The numbers are derived from different selection rules applied to the sample correlation matrix $\boldsymbol{R}$ and the sample covariance matrix $\boldsymbol{S}$. The value of the last column is again the selected one for iterative principal FA and in parenthesis the maximal possible number of k to achieve a mathematical correct solution.

Method	Cum. Var.>90 %	Kaiser's rule	Scree Test	LEV-Test	FA
$\boldsymbol{R}$, Monthly Mean	11	11 (12)	7	?	6 (8)
$\boldsymbol{S}$, Monthly Mean	7	7 (8)	6	6	

Table 9.11: Number of extracted factors k yielded from the cumulative percentage of variance being greater than 90 % (first column), the Kaiser's rule, and the modified Kaiser's rule in parenthesis (second column), the scree test (third column), the LEV diagram (fourth column), and for iterative principal FA (fifth column, the maximal number for k due to mathematic contraints is again given in parenthesis).

The application of the selection rules is the same as mentioned in Section 9.4.1. The determination of the number of factors, which are extracted with the help of the LEV

diagram was not possible for the correlation matrix, because not even a small break occurs within the first 24 logarithmic eigenvalues; they are connected by a straight line.

Comparing the maximal number of factors that can be used by iterative principal factor analysis, it can be noticed that, like for the 3-year mean corrected data set, it keeps at the lower bound of the values according to the selection rules.

Eigenvalues of the Matrices: The first three eigenvalues of the sample correlation matrix (and those derived from the factor loadings in the case of iterative principal FA) and the sample covariance matrix are shown in Table 9.12.

Matrix	Method	λ_1	λ_2	λ_3
$\boldsymbol{R}$, Monthly Mean	PCA	8.77	6.99	5.28
	FA	8.49	6.77	4.98
$\boldsymbol{S}$, Monthly Mean	PCA	35.33	30.05	29.09

Table 9.12: Eigenvalues of the sample correlation matrix and sample covariance matrix.

As can be seen, the eigenvalues of the monthly mean subtracted data set are considerably smaller compared to the eigenvalues calculated from the 3-year mean subtracted data. Because of that, the magnitude of renormalized coefficients will not be as different as in case of the 3-year mean centered coefficients.

PCA Differences When Applying the Sample Correlation and Covariance Matrix

As can be seen, in Figure 9.18 the first renormalized coefficients of the correlation matrix and the covariance matrix of the monthly mean corrected data detect different patterns in the atmosphere above Eurasia and Africa. Even though the eigenvectors actually enable a comparison between both matrices, they are not shown in this context. Instead of that, difference plots (Figure 9.15) of the first principal components (left) as well as of the first coefficients (right) are shown to point at the arising anomalies.

Because of the different appearance found in the renormalized coefficients, it is not surprising that both the principal components and the eigenvectors show strong deviations.

Generally, the amplitude of the principal component, which is based on the covariance matrix is larger compared to the correlation matrix based PC. The shapes of the principal components are also not consistent. Strong deviations can be found in spring 2003, 2004, and 2005 when the peaks of the curves show in opposite directions, but in the summer and autumn months of 2002, 2003, and 2004 the PCs show similar behavior. Apparently, the PCs are of different origin.

The difference plot (correlation matrix minus covariance matrix) of the coefficients confirms that assumption. Strong positive differences dominate the picture given in Figure 9.15, right, negative differences only occur at mid and low latitudes above a height of 20 km and sometimes, less pronounced (not visible in the Figure), between an altitude of 5 km and 15 km. The maximum deviations, which amount up to +0.63, can be found

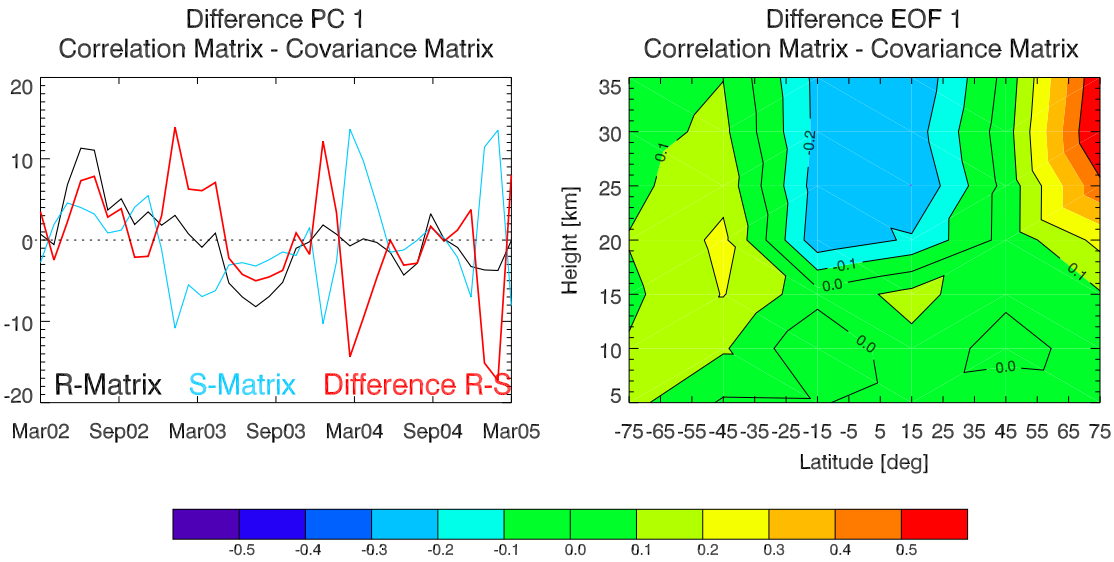


Figure 9.15: Differences between the first PC and the first EOF if they are calculated with the aid of the correlation matrix and the covariance matrix.

at the high northern latitudes above 25 km height, those, which arise in the southern hemisphere seldom excess +0.10. The most pronounced negative anomaly at mid and low latitudes amounts −0.30. Compared to the maximum absolute deviations found in the 3-year mean corrected data set, these values are pretty high and the assumption that the patterns underly different origins is confirmed. The different patterns found from both matrices will be discussed later on in detail.

Comparison Between PCA and FA Applying the Sample Correlation Matrix

For the monthly mean centered Eurasian-African data set, the differences between PCA and iterative principal FA as well as true FA were investigated.

Looking at the components/factor scores of PCA and iterative principal FA, which are depicted in the upper two graphs of Figure 9.16, the arbitrariness of the direction of the eigenvectors, which is fixed during the calculation process, outcrops.

While for the first extracted factor, PCA components and iterative principal FA factor scores have the same sign, they are of opposite signs for the second extracted factor (cf., upper right graph in Figure 9.16). In contrast to the 3-year mean centered data set, they do not vary much concerning their amount so that the difference does not exceed ±1 units (the components and factor scores mainly vary between ±3 units, only two peaks reach a bit more than ±5 units). The opposite signs of the components/factor scores in case of the second extracted factor can be ignored, because the coefficients/loadings are also of opposite signs and therefore, the reconstructed patterns, which are achieved by matrix multiplication of components/coefficients (PCA) and factor scores/loadings (FA), respectively, yield the same result. To represent the real differences between the two

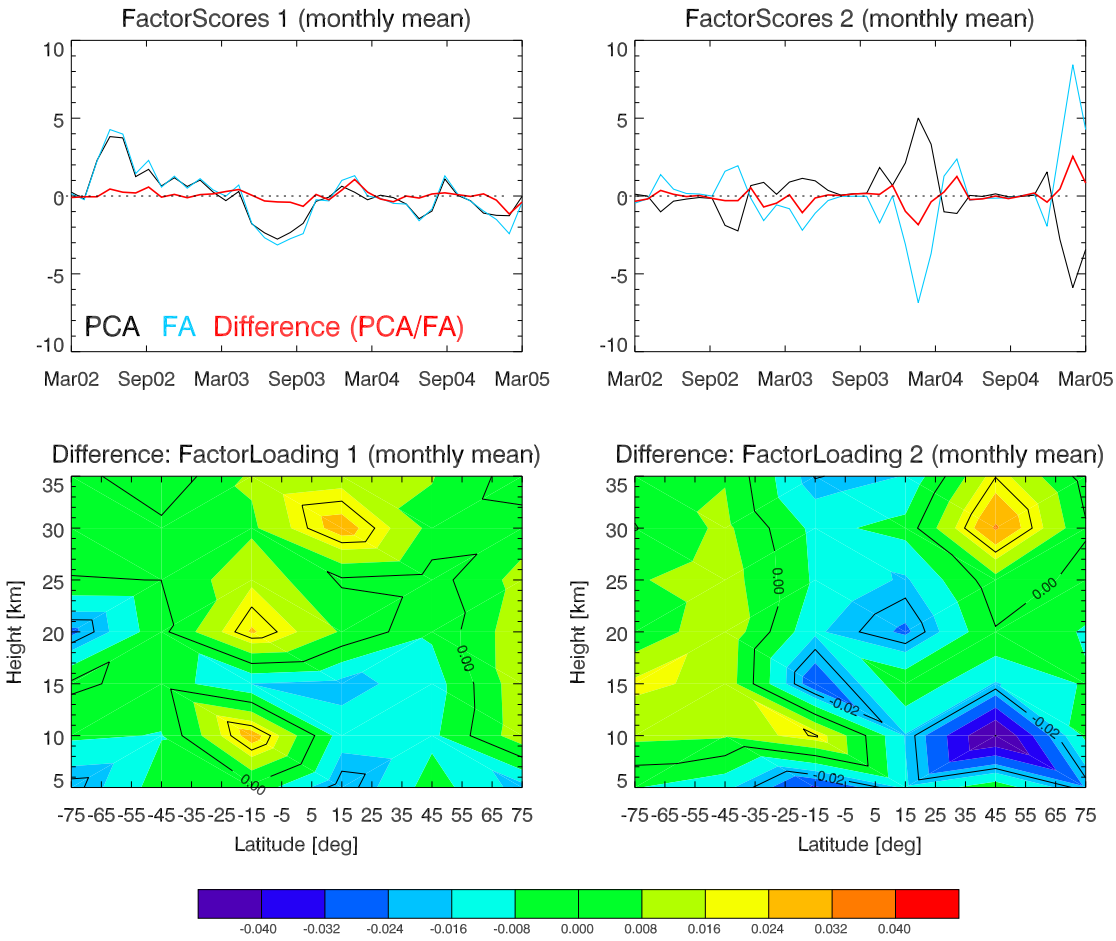


Figure 9.16: Differences between PCA and iterative principal FA for the first two selected factors.

methods, the PCA results were subtracted from the FA results for the first extracted factor and added in case of the second extracted factor, where they showed opposite signs. The coefficients/loadings of the two methods are again very similar with locally restricted deviations of about ± 0.03 units.

The differences in regard to true factor analysis are more pronounced. Certainly, the components/factor scores correspond quite well (even though larger deviations are given, where the amplitudes achieve a maximum, which is in this case generally stronger for PCA than for true FA), but therefore the differences between coefficients and loadings strengthen, so that for wide areas a variation of more than ± 0.05 units is achieved.

The differences between the two FA methods are possibly given by the unique variances of the matrix $\mathbf{\Psi}$, which vary essentially from each other, mainly at higher latitudes. There, true FA seems to have problems in allocating the unique and common variances

to certain grid points (cf., Figure 9.17), which may be reflected by comparing true FA results with PCA.

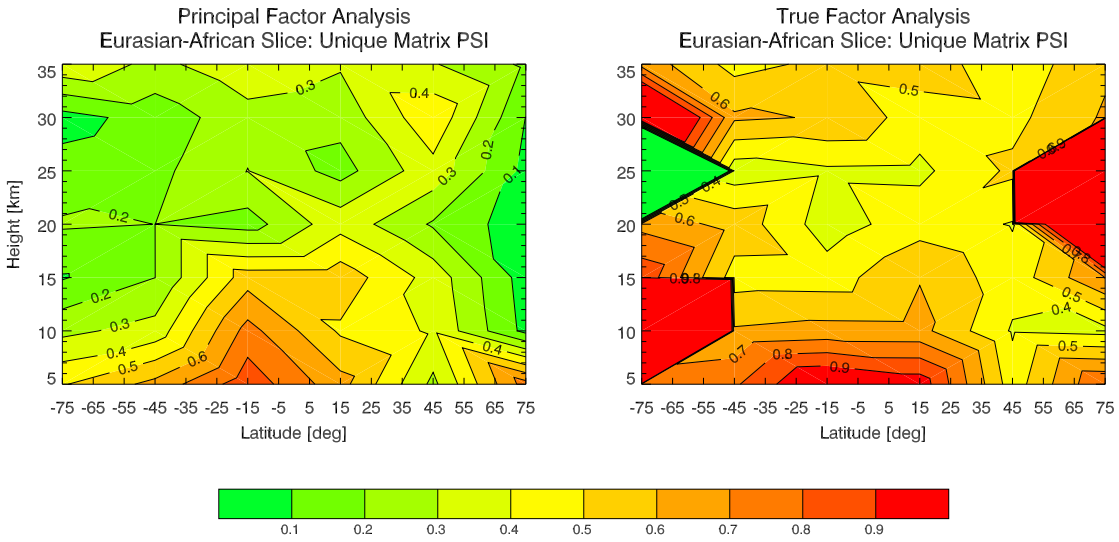


Figure 9.17: Unique variance matrices $\mathbf{\Psi}$ of monthly mean centered Eurasian-African temperatures calculated with iterative principal FA (left) and true FA (right). While the unique variances derived with iterative principal factor analysis mainly range between 20 % and 40 %, true factor analysis seems to fail in seperating unique and common variances correctly at higher latitudes.

Interpretation

First Principal Component and First Coefficients: The magnitude of the not normalized principal components varies between −8 units and +12 units in case of the correlation matrix and between −11 units and +14 units in case of the covariance matrix. Figure 9.18 (top) shows the renormalized PCs (left: correlation matrix, right: covariance matrix); it is remarkable that the structures of both PCs are not as similar as the PCs calculated from the 3-year mean subtracted data set.

The first renormalized coefficients calculated with the aid of the correlation matrix and the covariance matrix have the structure in common, which arises at the mid southern latitudes. The pattern emerging at the high northern latitudes can only be found in the covariance matrix based coefficients, whereas the features arising at low and mid latitudes (up to 35°N/S) as well as that at the high southern latitudes can only be noticed in the coefficients, which are based on the correlation matrix.

As mentioned above, the renormalized coefficients shown in Figures 9.14 (third 3-year mean eliminated coefficients) and 9.18 (first monthly mean eliminated coefficients) resemble each other in case of the correlation matrix. A conspicuous structure arises above 20 km height between the latitudinal range of 35°S and 35°N. Comparing respective reconstructed time series it becomes clear that the third coefficient still contains a strong

seasonal impact but this influence is removed in the first monthly mean coefficients. So, the reconstructed patterns are not similar and the coefficients must have another origin.

To investigate the noticed patterns in the first renormalized coefficients of the sample correlation matrix and the sample covariance matrix, time series of the corresponding regions were made.

The original monthly mean centered temperature anomaly at 75°S (top), the respective reconstructed patterns found by means of the correlation matrix (middle) and by means of the covariance matrix (bottom) are shown in Figure 9.19. Comparing these time series yields the correlation matrix pattern being more pronounced and more convenient to detect and describe the effective temperature anomaly arising at the high southern latitudes. The feature describes the southern polar vortex, which can be found at high latitudes. It causes particularly low temperatures and was relatively weak in 2002 (it even split in two in September 2002), but powerful in 2003, and average in 2004. Even though the real structure extends throughout all height levels, the detected pattern (of the correlation matrix) rests at lower altitudes. An examination of the temperature anomalies in that region are also discussed in Section 9.6.2.

The structure emerging at the low and mid northern and southern latitudes is also better resolved in case of the correlation matrix, the respective time series (measured temperature anomalies and reconstructed data) at 15°S are depicted in Figure 9.20. The pattern calculated on basis of the correlation matrix locates the structures, which emerges in that region, whereas the one recalculated with the first coefficients derived by means of the covariance matrix neglects that structure. The time series of the reconstructed first principal component/coefficient shows a formation arising in June every year; it remains until September in 2002 and 2003, in 2004 it is less pronounced. Since the structure comes from averaging the temperatures between the equator and 30°S, it is possible that it is due to the QBO (quasi-biennial oscillation). The QBO, which arises in equatorial regions and can influence the average temperatures between 0° and 30°S, will be discussed in more detail in Section 9.7.1. The same reconstructed time series develops in the northern hemisphere between 0° and 30°N.

The third remarkable structure occurs at the high northern latitudes, but it is only detected from the coefficients calculated with the covariance matrix. This pattern is completely disregarded by the first coefficients, which are based on the correlation matrix. Respective time series of that region and reconstructed patterns are shown in Figure 9.21. The temperature anomalies both are positive and negative, arise first at high altitudes, propagating downwards; they occur between November and April, the northern hemisphere winter. According to literature (Manney et al. 2005; Angell et al. 2003a, 2004a, 2005), the pattern can be attributed to sudden stratospheric warming events (cf., Chapter 4). Thus, a relatively cold December 2002 was followed by a strong warming in mid January; in the early winter 2003/2004, temperatures were above average, they rose in December and remained above average until end of February, but for March 2004 the temperature anomalies became negative; record low temperatures were observed in winter 2004/2005 until February, when a sudden stratospheric warming was on its way at high altitudes. Mostly, the SSW events propagate down to a height of around 10 km. Anyway, the reconstruction of the covariance matrix based pattern does not succeed in

resolving the real height structure and it cuts off at heights of approximately 20 km.

The first PCA coefficients, calculated by elimination of seasonal influences by means of the correlation matrix, contribute 20.87 % and the ones, which are based on the covariance matrix contribute 61.51 % to the total variance; the respective varimax rotated PCA coefficients account for 9.97 % and 34.70 %, respectively.

Second Principal Component and Second Coefficients: The second coefficients (Figure 9.22) calculated by means of the correlation and the covariance matrix completes the picture given from the analysis of the first coefficients. The second coefficients derived from the correlation matrix focus on the high northern latitudes and the respective coefficients, which are based on the covariance matrix, on the high southern latitudes as well as on the low southern and northern latitudes, which are not described in the respective first coefficients.

The combination of the first and the second coefficients yields an acceptable reconstruction of the actually temperature field. It can be concluded that the two matrices focus on different zonal regions and that the spatial and temporal location of variation cannot be resolved from one single factor.

Accounted Variance: The contribution to the amount of total variance of not rotated and varimax rotated coefficients/loadings calculated after the elimination of monthly means is shown in Table 9.13. Comparing the explained variances of Table 9.10 and Table 9.13 it can be stated that the accounted variances of the correlation matrix and the covariance matrix are considerably larger, if the 3-year mean of each variable is eliminated before calculating the principal components, than it is the case if the seasonal impact is removed.

While there are nearly no differences between PCA and iterative principal FA derived variances for the "conventional" coefficients/loadings, larger deviations are given by the varimax rotated coefficients/loadings. There, iterative principal FA distributes the variances of the first two factors quite equally (about 15 %), wheras PCA favors the second factor (also 15 %) above the first one (around 10 %). All in all, regarding the first three extracted factors, a larger total variance is explained by the varimax rotated loadings (40 %) than by the varimax rotated coefficients (31 %).

Method	Not Rotated Coefficients/Loadings		Varimax Rotated Coefficients/Loadings	
	PCA	FA	PCA	FA
$\boldsymbol{R}$, Monthly Mean				
$\tilde{\boldsymbol{a}}_1$	20.87 %	20.21 %	9.97 %	14.73 %
$\tilde{\boldsymbol{a}}_2$	16.66 %	16.11 %	15.15 %	15.72 %
$\tilde{\boldsymbol{a}}_3$	12.56 %	11.86 %	6.32 %	9.84 %
$\boldsymbol{S}$, Monthly Mean				
$\tilde{\boldsymbol{a}}_1$	25.69 %		24.39 %	
$\tilde{\boldsymbol{a}}_2$	21.85 %		16.29 %	
$\tilde{\boldsymbol{a}}_3$	21.16 %		19.74 %	

Table 9.13: Accounted variances of the first three monthly mean centered not rotated and varimax rotated coefficients/loadings.

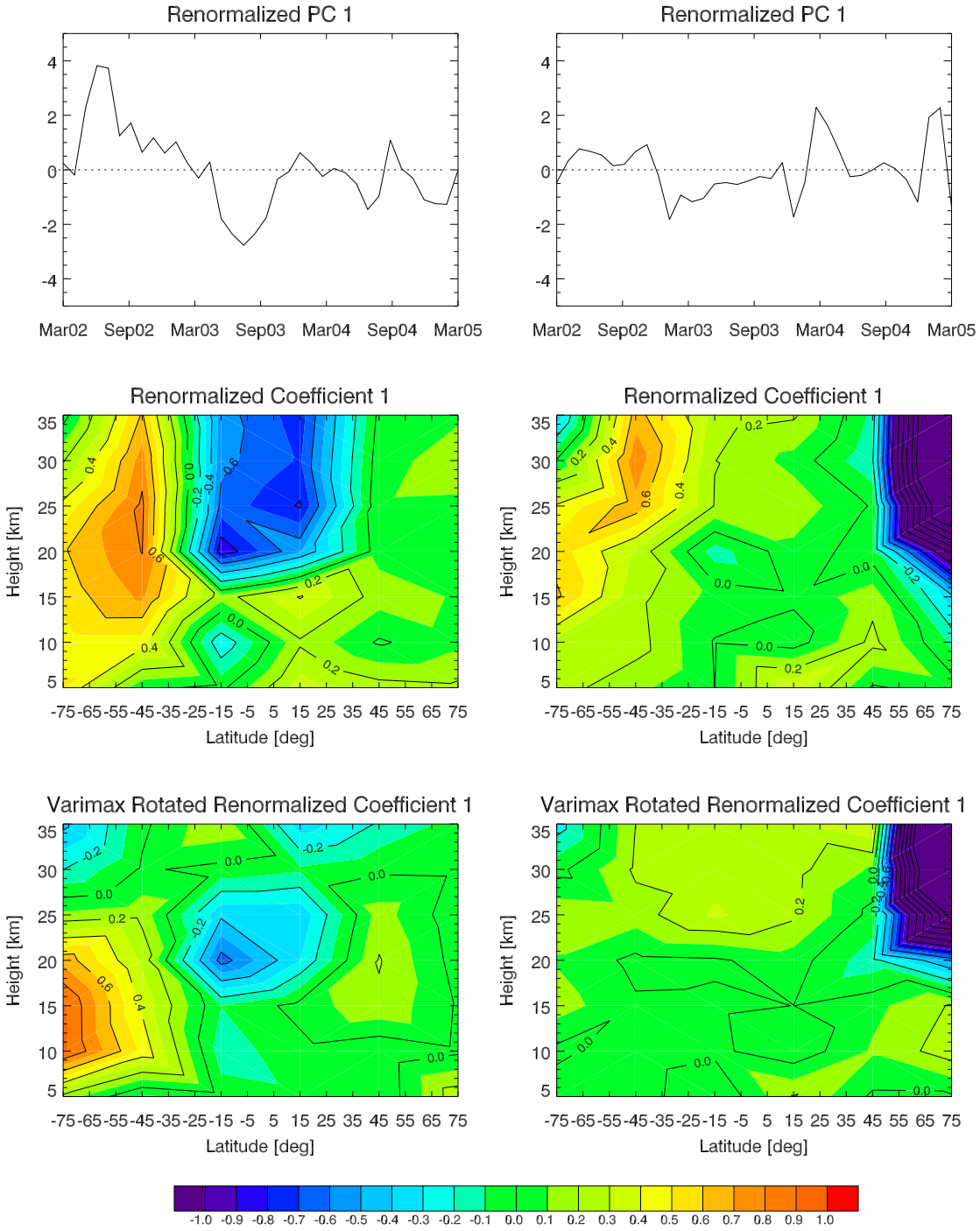


Figure 9.18: First renormalized principal component (top), the corresponding coefficient (middle), and the varimax rotated coefficient (bottom), calculated after the elimination of the monthly mean by means of the correlation matrix (left) and the covariance matrix (right).

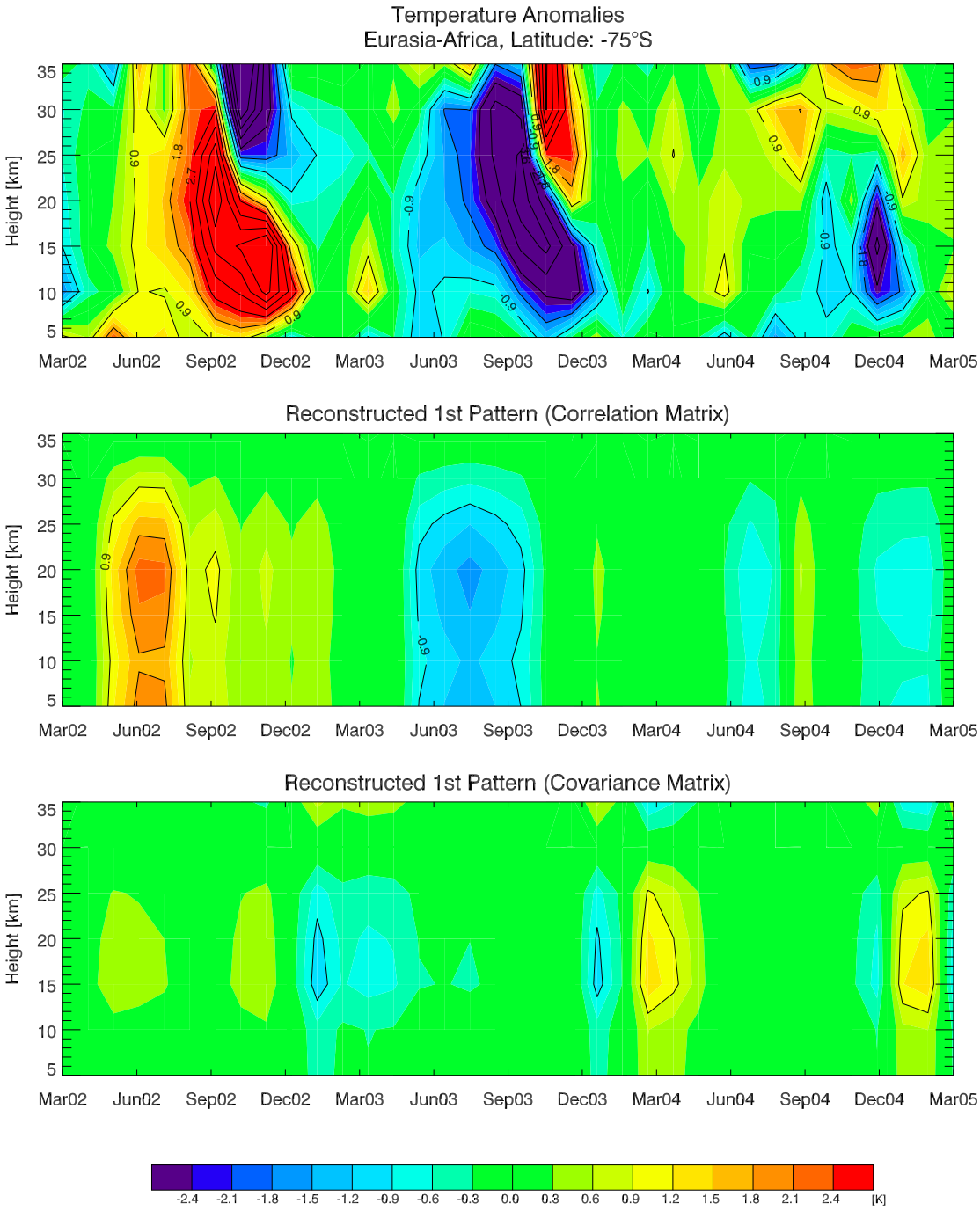


Figure 9.19: Comparison of the monthly mean eliminated time series (top) and the reconstruction of the first PC/coefficient calculated by means of the correlation matrix (middle), and the covariance matrix (bottom) at 75°S (60°S to 90°S).

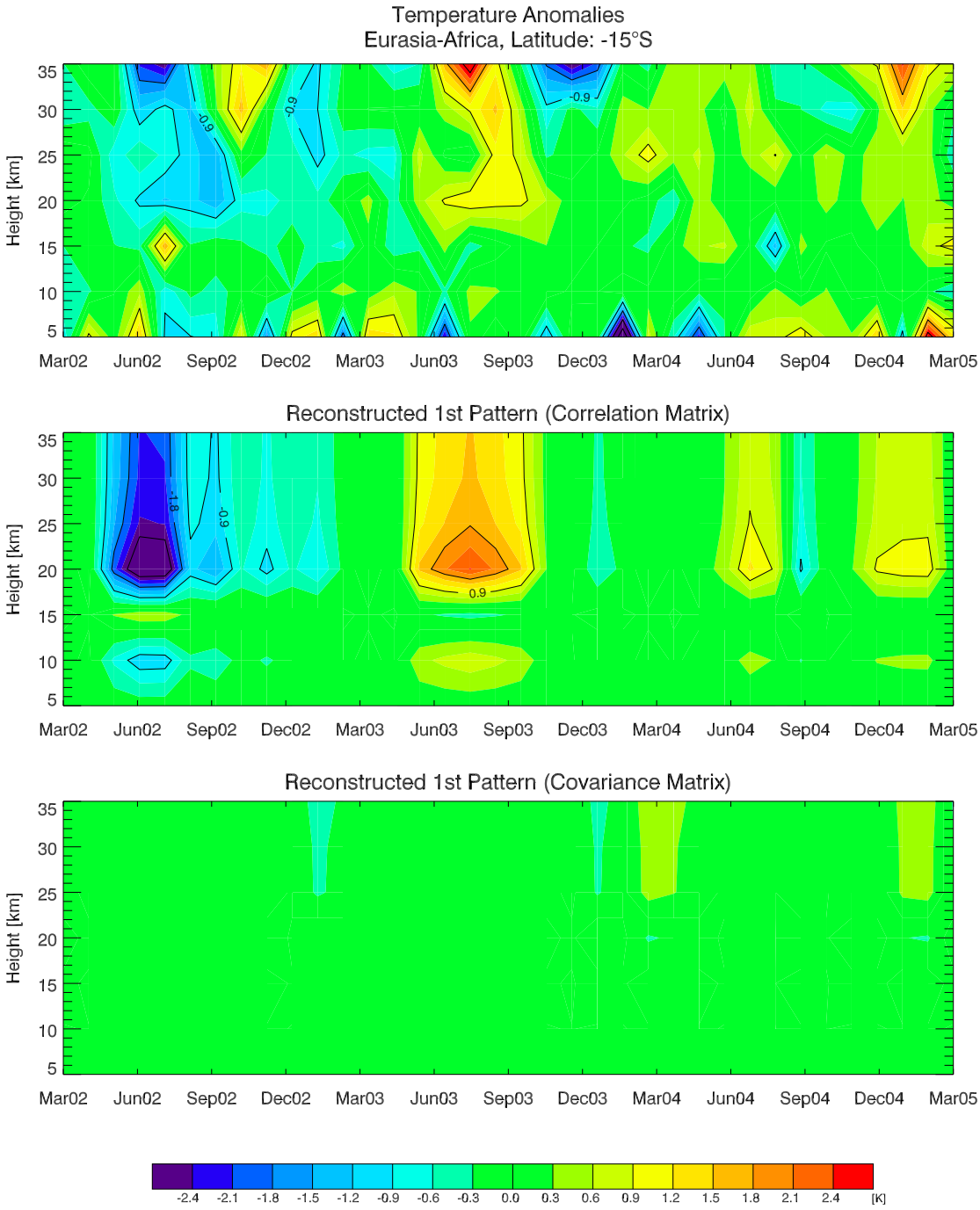


Figure 9.20: Comparison of the monthly mean eliminated time series (top) and the reconstruction of the first PC/coefficient calculated by means of the correlation matrix (middle), and the covariance matrix (bottom) at 15°S (0°S to 30°S).

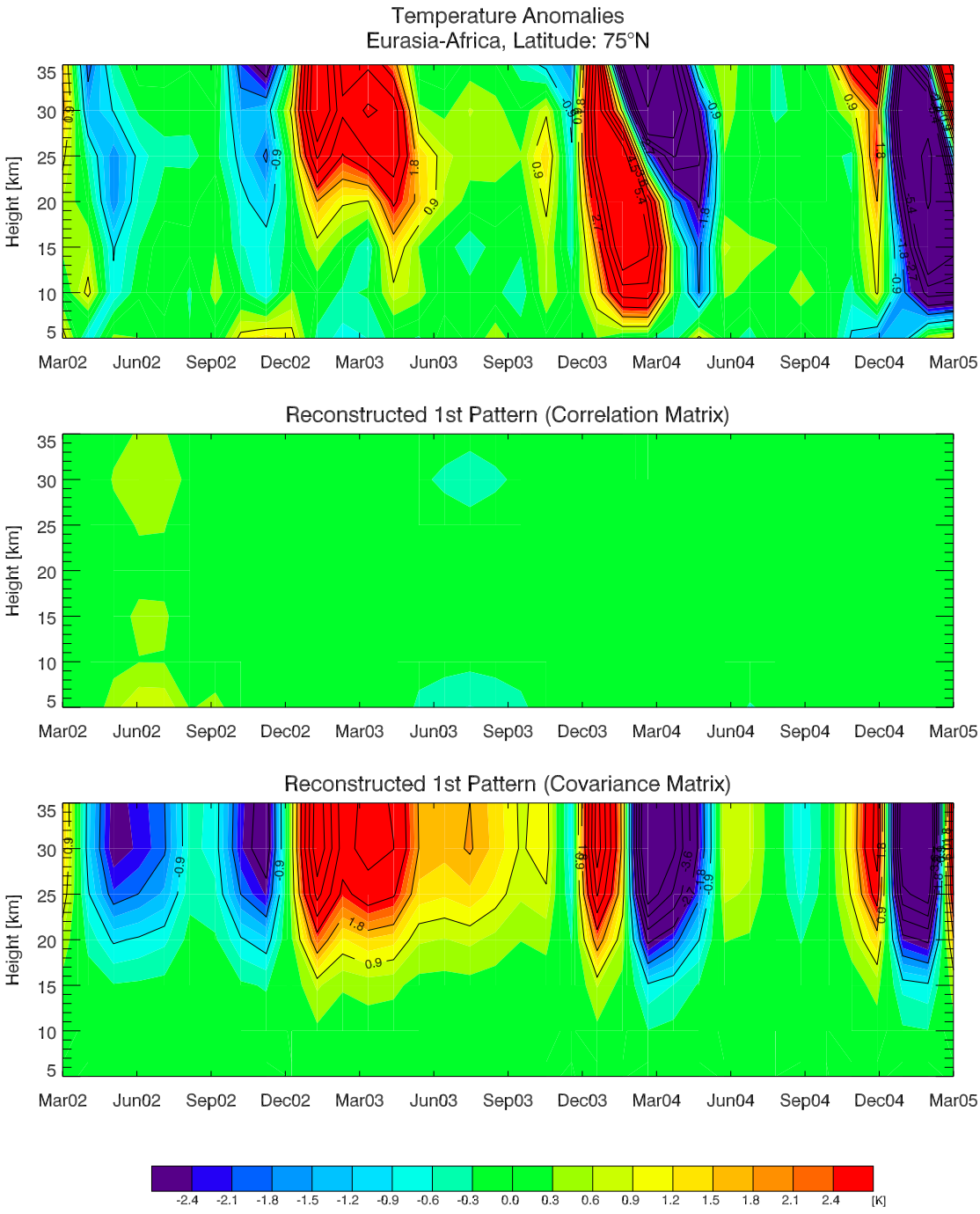


Figure 9.21: Comparison of the monthly mean eliminated time series (top) and the reconstruction of the first PC/coefficient calculated by means of the correlation matrix (middle), and the covariance matrix (bottom) at 75°N (60°N to 90°N).

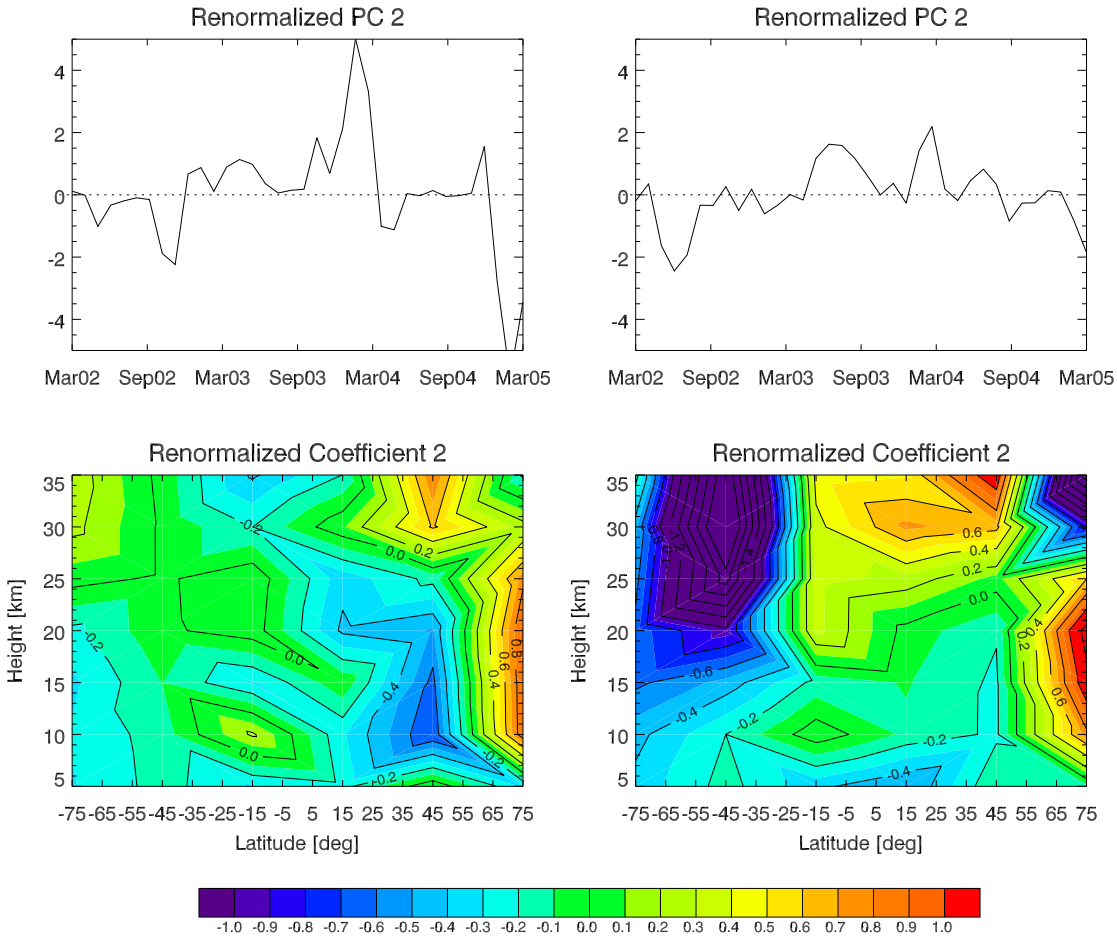


Figure 9.22: Second renormalized principal component (top) and the corresponding coefficient (bottom), calculated after the elimination of the monthly mean by means of the correlation matrix (left) and the covariance matrix (right).

9.5 Temperature Data at 15 km Height

(Authors: B.C. Lackner and B. Pirscher)

The temperatures at an altitude of 15 km are given in a latitudinal-longitudinal resolution of $(30° \times 45°)$.

9.5.1 PCA/FA of 3-Year Mean Subtracted Temperature Anomalies at 15 km Height

Number of Factors Extracted: Table 9.14 shows the number of factors given from different selection rules applied to the sample correlation matrix and the sample covariance matrix. The last column contains the selected number of factors for iterative principal FA and in parenthesis the maximal possible number of k to achieve a mathematical correct solution.

Method	Cum. Var.>90 %	Kaiser's rule	Scree Test	LEV-Test	FA
$\boldsymbol{R}$, 3-Year Mean	9	7 (10)	4	3	3 (6)
$\boldsymbol{S}$, 3-Year Mean	2	2 (2)	3	3	

Table 9.14: Number k of factors following from different selection rules for PCA and mathematical constraints for iterative principal FA.

The number of extracted factors mostly varies between two and four, except for the correlation matrix, if it is investigated by means of the cumulative variance and the Kaiser's rule, where nine and seven factors should be retained for further calculations, respectively. The number of selected factors for factor analysis as well as the maximal possible number of factors correspond quite well with the requirements according to Kaiser's rule, scree and LEV-test in this case.

Eigenvalues: The first three eigenvalues of the 3-year mean corrected temperature data calculated with the correlation matrix $\boldsymbol{R}$ and the covariance matrix $\boldsymbol{S}$ as well as the eigenvalues derived from the factor loadings are shown in Table 9.15.

Matrix	Method	λ_1	λ_2	λ_3
$\boldsymbol{R}$, 3-Year Mean	PCA	22.77	9.18	2.87
	FA	22.62	8.84	2.54
$\boldsymbol{S}$, 3-Year Mean	PCA	1547.92	91.73	23.37

Table 9.15: Eigenvalues of the sample correlation matrix and the sample covariance matrix as well as eigenvalues derived from the factor loadings in case of iterative principal FA.

The eigenvalues of both, PCA and iterative principal FA, again show a very good agreement. The fact that the eigenvalues calculated from the correlation matrix are

much smaller compared to the eigenvalues of the covariance matrix attracts once more attention. The effect of the renormalization (with the square root of the eigenvalues) of the principal components and the coefficients is the same as discussed in Section 9.4.1. The renormalized principal components calculated from the correlation matrix are larger compared to those calculated with the covariance matrix, whereas the renormalized coefficients are more pronounced in case of the covariance matrix.

First Principal Component and First Coefficients: The first renormalized principal components and first renormalized coefficients calculated from the correlation matrix and the covariance matrix are depicted in Figure 9.23. Results, which are based on the correlation matrix are shown on the left side, those of the covariance matrix on the right hand side. Because the varimax rotation changes the coefficients only a little bit, so that the patterns nearly stay the same, the varimax rotated coefficients are not shown.

The amplitude of the not normalized principal components is about ± 50 units in case of the correlation matrix and about ± 60 units in case of the covariance matrix. So, the amplitude of the principal component based on the covariance matrix is again a little bit larger compared to the one, which are based on the correlation matrix. The sinusoidal character of the curves is contained both times, it can also be noticed (with a smaller magnitude due to the renormalization) in Figure 9.23, top.

The patterns, which arise in the coefficients are similar in case of the correlation matrix and the covariance matrix. The high northern and the high southern hemisphere show a strong variability of opposite signs, which can be interpreted as the seasonal influence. The low latitudes show smaller temporal temperature variations. But two more features can be found in the maps: The variability at the low and mid northern latitudes is more pronounced in the western hemisphere over the Pacific and shows the same deviation as that found in the southern hemisphere at high latitudes. The second conspicuous pattern is centered at the low southern latitudes (between the equator and about 25°S) above Africa; it extends to South America in the west and to Australia in the east. Looking to the finer ($15° \times 45°$) resolution, the same anomaly results even in a small band around the globe with maximum width over South Africa.

Recalling the first coefficients of the Eurasian-African data set, some remarkable structures at the low latitudes (cf., Figure 9.9) at 15 km height were noticed. It seems that the patterns noticed there have the same origin as the features mentioned above.

Again, some time series were plotted to investigate these deviations. The time series of the 3-year mean centered temperature anomalies (top) and the reconstructed time series of the correlation matrix (middle), and the covariance matrix (bottom) can be seen in Figure 9.24. The temperature deviations, which occur at the low southern latitudes (being oppositional to the remaining southern hemisphere pattern) can be clearly seen in the top time series. The same structure is strongly overrated in case of the correlation matrix but relatively good resolved in case of the covariance matrix.

The feature arising in the low (and mid) northern hemisphere cannot be resolved in the time series of the temperature anomalies at 45°E, but it can already be found at the Greenwich meridian. Generating a time series along 90°W (where it is strongest) warm

and cold air masses can be clearly seen migrating at 15 km height from south to north with the annual cycle.

Second Principal Component and Second Coefficients: The sinusoidal shape of the principal components is visible in both the correlation matrix and the covariance matrix based PCs. The second PC is temporal adjusted to the first one (the extrema of the first PC are in July/August (maxima) and in December (minimum), whereas they are in October/November (maxima) and in May/June/July (minima) in the second PC). The pattern covers the whole northern hemisphere (correlation matrix) and the mid and high northern and southern hemisphere (covariance matrix). A small area of an opposite deviation can be found south-eastern Australia (both matrices).

The reconstructed time series (not shown) yield temperature anomalies, which simultaneously start to get warmer in the northern hemisphere and are still warm in the southern hemisphere; contrariwise they start to get colder in the northern hemisphere and are still cold in the southern hemisphere. So, the feature causes a small period of overlapping of warm and cold temperature anomalies in both hemispheres. Together with the first pattern a temporal movement of the temperature anomalies is caused. An exception is the longitudinal sector situated at the date line, where the pattern in the northern hemisphere and that in the southern hemisphere are of opposite signs. This feature can also be found in the real temperatures at that longitudes.

The covariance matrix resolves that pattern in both hemispheres, whereas the correlation matrix concentrates on the northern hemisphere and neglects the pattern in the southern hemisphere (exception: pattern at the date line).

Accounted Variance: The contribution to the amount of total variance of not rotated and varimax rotated coefficients/loadings is shown in Table 9.16.

Method	Not Rotated Coefficients/Loadings		Varimax Rotated Coefficients/Loadings	
	PCA	FA	PCA	FA
$\boldsymbol{R}$, 3-Year Mean				
$\tilde{\boldsymbol{a}}_1$	47.43 %	47.12 %	41.92 %	36.98 %
$\tilde{\boldsymbol{a}}_2$	19.12 %	18.42 %	4.26 %	20.99 %
$\tilde{\boldsymbol{a}}_3$	5.98 %	5.30 %	3.21 %	12.86 %
$\boldsymbol{S}$, 3-Year Mean				
$\tilde{\boldsymbol{a}}_1$	90.75 %		78.50 %	
$\tilde{\boldsymbol{a}}_2$	5.38 %		0.21 %	
$\tilde{\boldsymbol{a}}_3$	1.37 %		1.64 %	

Table 9.16: Accounted variances of the first three not rotated and varimax rotated coefficients/-loadings.

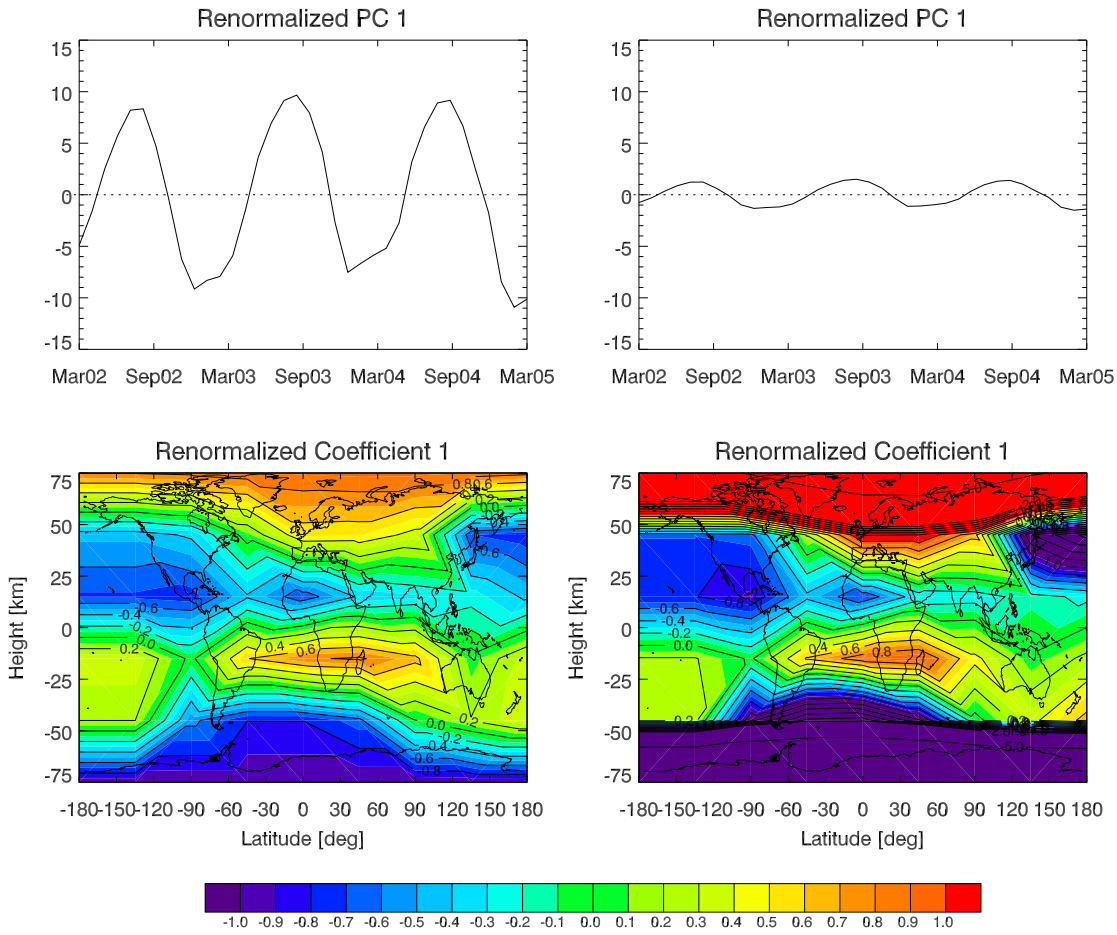


Figure 9.23: First renormalized principal component (top) and corresponding coefficient (bottom), calculated after the elimination of the 3-year mean by means of the correlation matrix (left) and the covariance matrix (right).

9.5.2 PCA/FA of Monthly Mean Subtracted Temperature Anomalies at 15 km Height

Number of Factors Extracted: Table 9.17 shows the number of factors, which should be retained when using them for additional calculations.

Compared to the number of extracted factors calculated after the elimination of the 3-year mean, k now is larger.

Eigenvalues of the Matrices: The first three eigenvalues of both matrices (the data are centered to the monthly mean) are shown in Table 9.18. The FA eigenvalues are again derived from the factor loadings.

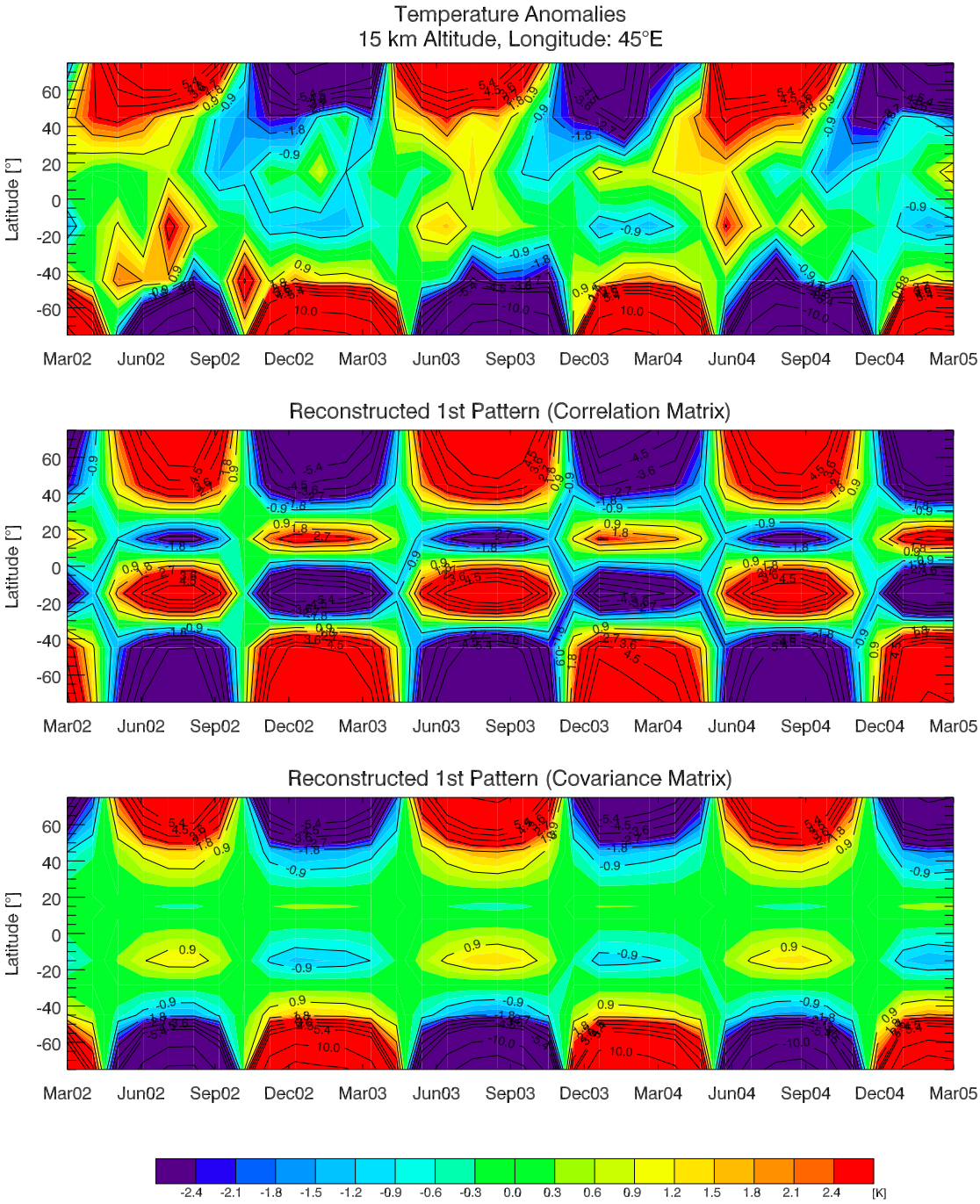


Figure 9.24: Measured temperature anomalies (top) and reconstruction of the data set by means of the correlation matrix (middle) and the covariance matrix (bottom) at 45°E (22.5°E to 67.5°E).

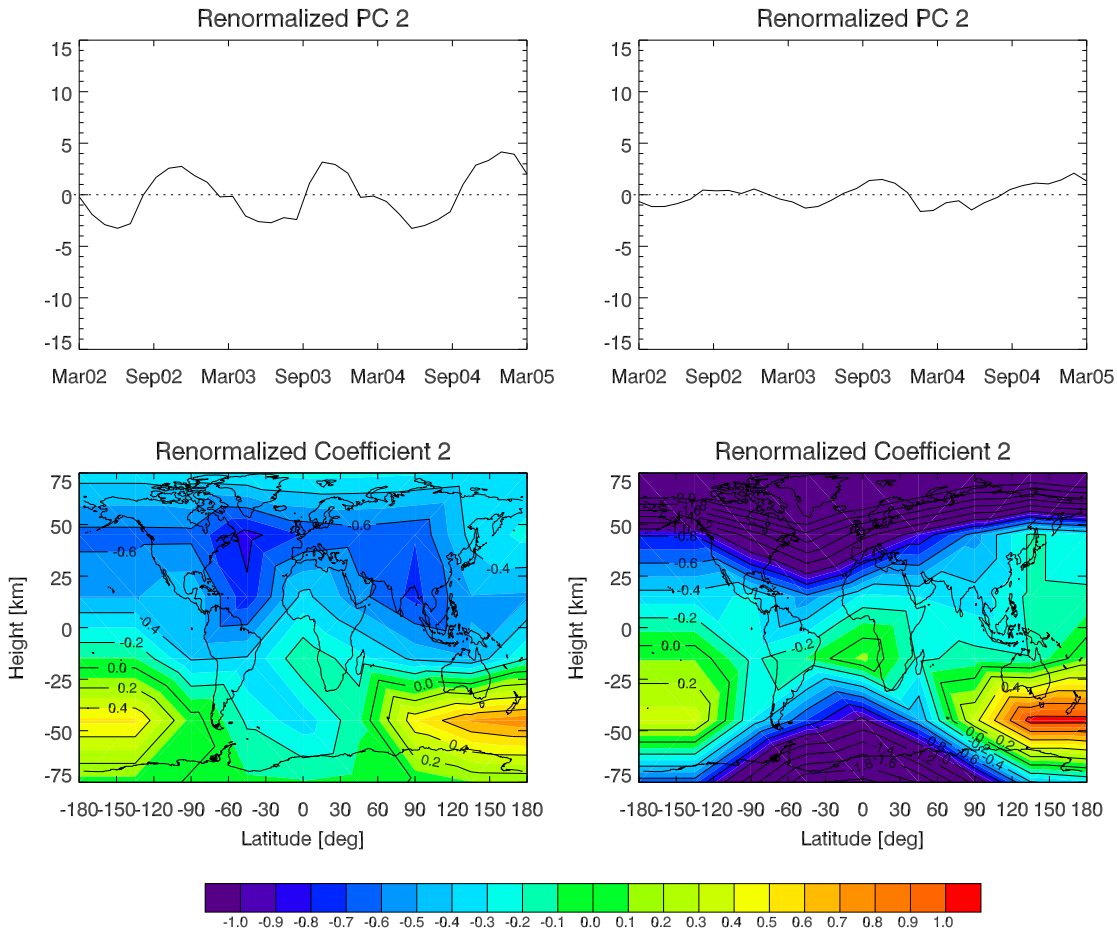


Figure 9.25: Second renormalized principal component (top) and corresponding coefficient (bottom), calculated after the elimination of the 3-year mean by means of the correlation matrix (left) and the covariance matrix (right).

First Principal Component and First Coefficients: The renormalized principal components of the correlation matrix and the covariance matrix are depicted in Figure 9.26 together with their respective renormalized coefficients. In general, the PCs as well as the coefficients are similar, if they are calculated by means of the correlation matrix and the covariance matrix; the amplitudes of the PCs range between ± 20 units in case of the covariance matrix and are a little bit smaller in case of the correlation matrix. The coefficients show a strong anomaly arising in high latitude regions, north and south, with opposite signs. Another remarkable feature can be found between about 25°N and 55°N in a longitudinal area spanning from 60°E to the data line and it shows the same sign as the structure at high southern latitudes. Furthermore, the correlation matrix based coefficients show some small structures at low and mid latitudes, but they are all only little pronounced.

The interpretation of the pattern found in the first coefficients is again facilitated by

Method	Cum. Var.> 90 %	Kaiser's rule	Scree Test	LEV-Test	FA
$\boldsymbol{R}$, Monthly Mean	12	11 (13)	6	3	5 (7)
$\boldsymbol{S}$, Monthly Mean	5	6 (7)	5	3	

Table 9.17: Number k of factors following from different selection rules and according to the requirements of iterative principal FA (last column).

Matrix	Method	λ_1	λ_2	λ_3
$\boldsymbol{R}$, Monthly Mean	PCA	11.14	8.89	4.75
	FA	10.94	8.71	4.34
$\boldsymbol{S}$, Monthly Mean	PCA	53.61	42.11	6.91

Table 9.18: Eigenvalues of the correlation matrix and covariance matrix calculated after elimination of the monthly mean.

generating respective time series. Comparing these times series (measured temperatures and reconstructed data) at different longitudes it can be noticed that there are only few differences. Only the pattern over eastern Asia, which is relatively weak compared to the pattern at high latitudes, influences the latitudinal range between 25°N and 55°N. The temperature anomalies along 90°W and the recalculated data (correlation matrix) are depicted in Figure 9.27.

The times series show a strong variability in the southern hemisphere (south of 30°S) from June 2002 to January 2003 (positive), from June 2003 to December 2003 (negative) and from December 2004 to January 2005 (negative). These structures can be attributed to the southern polar vortex. In 2002, the antarctic polar vortex was relatively warm, due to an intense stratospheric warming it even split in two in September 2002; in 2003 it was very strong, and in 2004 it was less pronounced than in 2003 but much stronger compared to 2002 (Angell et al. 2002, 2003b, 2004b).

Prominent patterns arising in the northern hemisphere can be found from October 2002 to January 2003 (negative), from February to May 2003 (positive), from December 2003 to March 2004 (strong positive), and from January 2005 to the end of the time series (end of February 2005). These patterns can be attributed to arctic stratospheric features. Manney et al. (2005) report on the one hand on an unusually cold early arctic winter 2002/2003, which was interrupted from a major warming in late January and further warmings mid-February and early March, on the other hand on a particular warm winter 2003/2004 in the arctic stratosphere. According to Angell et al. (2005) extremely cold temperatures (lower than -78 °C) in the lower stratosphere dominated over portions in the arctic region in the winter 2004/2005 from December to February (cf., Chapter 4).

These structures are present in the first reconstructed time series of the first coefficients of the correlation matrix as well as of the covariance matrix.

Second Principal Component and Second Coefficients: The second principal component of the correlation matrix and the covariance matrix can be characterized as a curve without any regularities. Minima occur in autumn 2002 and in late winter (February) 2004 (more pronounced in case of the correlation matrix), maxima can be found in autumn 2003 and in later winter 2005. The amplitudes of the not normalized PCs amount to about ±15 units. The coefficients form a symmetric structure with deviations on both poles and with some smaller anomalies in between with opposite signs.

The contribution of the second principal component and the second coefficients to the reconstruction of the intrinsic data is negligible, because the most important patterns are already included in the first ones. So, it is not possible to attribute the structures of the second factor to another atmospheric pattern.

Accounted Variance: The contribution to the amount of total variance of not rotated and varimax rotated coefficients/loadings is summarized in Table 9.19. It can be noticed that the quoted variance is spread over the first three coefficients more regular compared to the 3-year mean eliminated coefficients. Nevertheless, the variation accounted for by the first factors of the covariance matrix is noticeable larger than the variation of the first factors of the correlation matrix.

Method	Not Rotated Coefficients/Loadings		Varimax Rotated Coefficients/Loadings	
	PCA	FA	PCA	FA
$\boldsymbol{R}$, Monthly Mean				
$\tilde{\boldsymbol{a}}_1$	23.20 %	22.79 %	20.27 %	21.48 %
$\tilde{\boldsymbol{a}}_2$	18.51 %	18.14 %	18.11 %	18.51 %
$\tilde{\boldsymbol{a}}_3$	9.89 %	9.05 %	5.60 %	8.73 %
$\boldsymbol{S}$, Monthly Mean				
$\tilde{\boldsymbol{a}}_1$	43.55 %		38.81 %	
$\tilde{\boldsymbol{a}}_2$	34.21 %		35.37 %	
$\tilde{\boldsymbol{a}}_3$	5.61 %		3.54 %	

Table 9.19: Accounted variances of the first three not rotated and varimax rotated coefficients/-loadings.

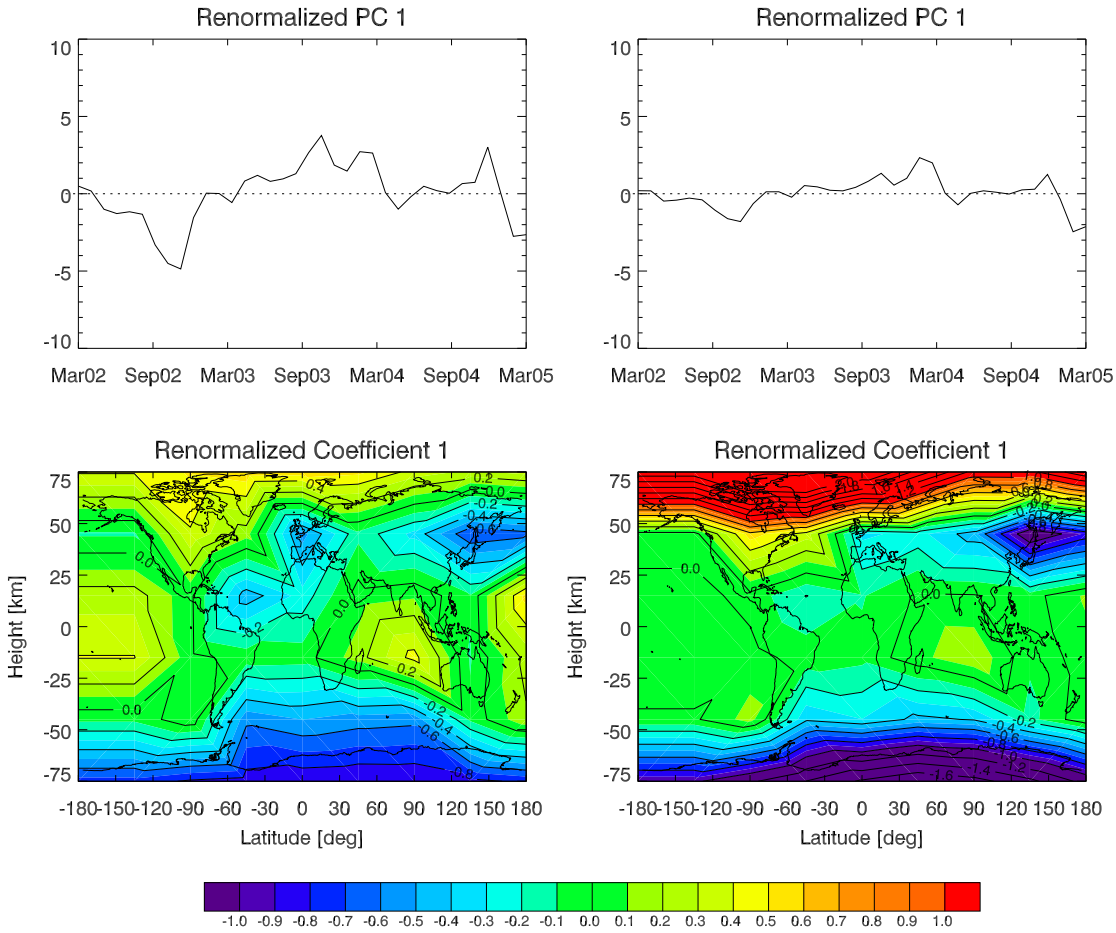


Figure 9.26: First renormalized principal component (top) and corresponding coefficient (bottom), calculated after the elimination of the monthly mean by means of the correlation matrix (left) and the covariance matrix (right).

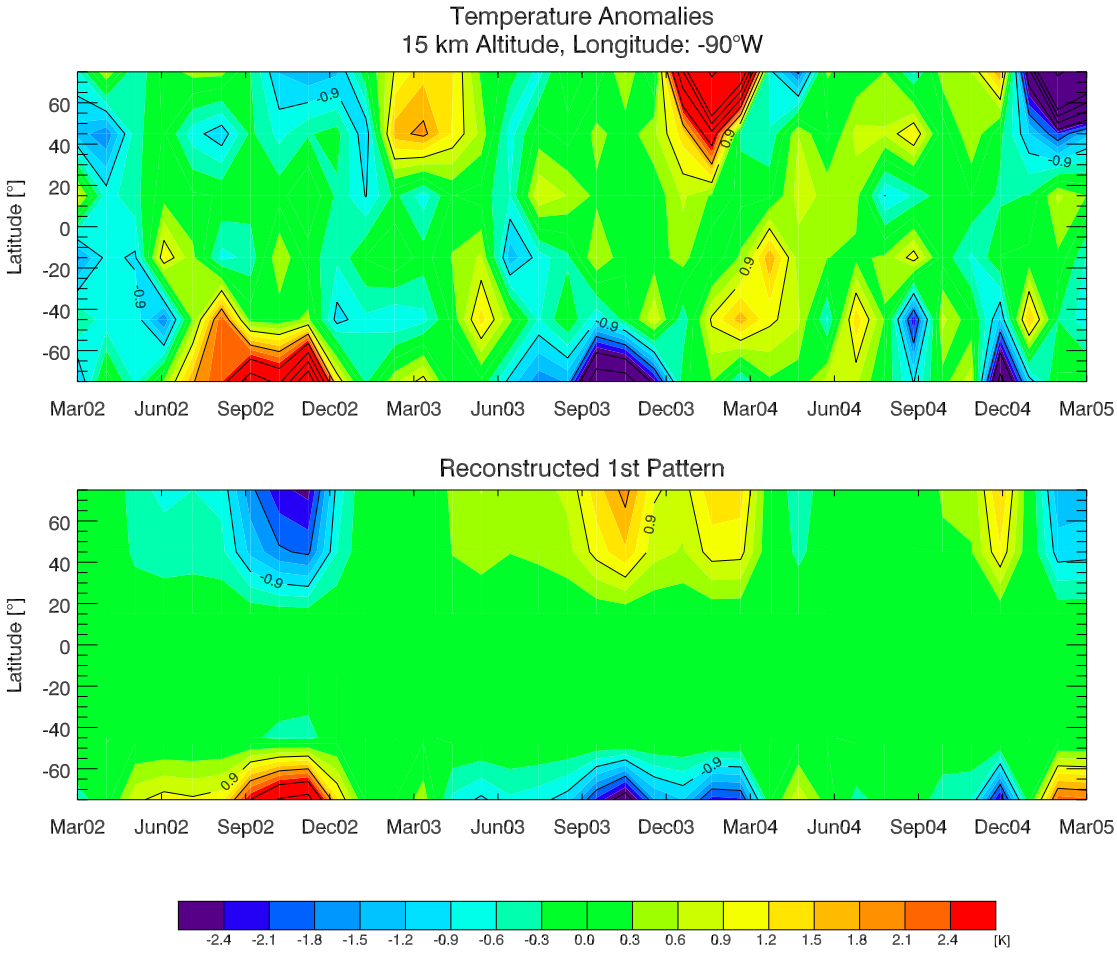


Figure 9.27: Measured temperature anomalies (top) and reconstruction of the data set (by means of the correlation matrix) at 90°W (67.5°W to 112.5°W).

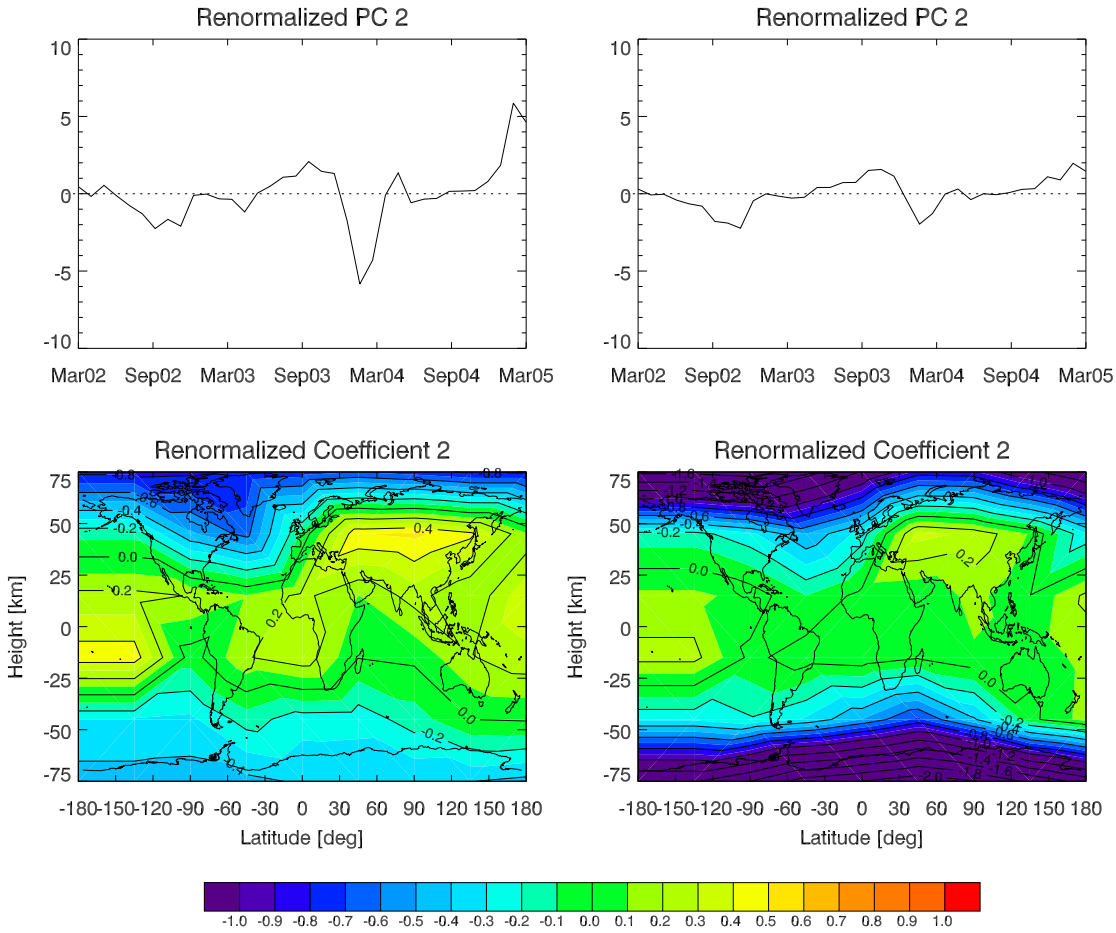


Figure 9.28: Second renormalized principal component (top) and corresponding coefficient (bottom), calculated after the elimination of the monthly mean by means of the correlation matrix (left) and the covariance matrix (right).

9.6 Temperature Data in the South Polar Area

(Authors: B.C. Lackner and B. Pirscher)

The temperature data in the south polar region are given as a zonal mean temperature field with a 5° latitudinal resolution (from 57.5°S to 87.5°S) yielding 6 zonal bands, and a vertical resolution of 5 km (7 height levels).

9.6.1 PCA/FA of 3-Year Mean Subtracted Temperature Anomalies in the South Polar Area

Number of Factors Extracted: Table 9.20 shows the number of factors given from different selection rules applied to the sample correlation matrix and to the sample covariance matrix as well as the number of factors extracted with iterative principal FA with the maximal number of extractable factors (due to mathematical constraints) in parenthesis.

Method	Cum. Var.>90 %	Kaiser's rule	Scree Test	LEV-Test	FA
$\boldsymbol{R}$, 3-Year Mean	2	2 (3)	3	5	2 (11)
$\boldsymbol{S}$, 3-Year Mean	2	2 (2)	3	4	

Table 9.20: Number k of factors being estimated from different selection rules and according to mathematical constraints in case of iterative principal FA.

The application of the selection rules is the same as that discussed in Section 9.4.1.

The number of factors given from the different criteria is relatively similar around two or three so that further considerations will be restricted to the first and to the second principal components and their coefficients. The same applies to the selected factors for iterative principal FA, even though in this case $k = 11$ factors could have been extracted. But as the first two factors already explained more than 95 % of the total variance, which was the highest amount achieved by iterative principal FA in regard to the four different atmospheric data sets, these two were considered to be sufficient.

Eigenvalues: The first three eigenvalues of the 3-year mean centered temperature data calculated with the sample correlation matrix and the sample covariance matrix, as well as the eigenvalues derived from the factor loadings of iterative principal factor analysis are shown in Table 9.21.

As only two factors were extracted with iterative principal FA (see above), no value for the third one can be presented.

Similarly to the temperature data given in the Eurasian-African sector, the eigenvalues of the 3-year mean centered data of the covariance matrix are comparatively high, which results in huge renormalized coefficients.

Matrix	Method	λ_1	λ_2	λ_3
$\boldsymbol{R}$, 3-Year Mean	PCA	33.07	7.29	0.83
	FA	33.03	7.26	
$\boldsymbol{S}$, 3-Year Mean	PCA	6946.66	1171.26	101.45

Table 9.21: Eigenvalues of the sample correlation matrix and sample covariance matrix.

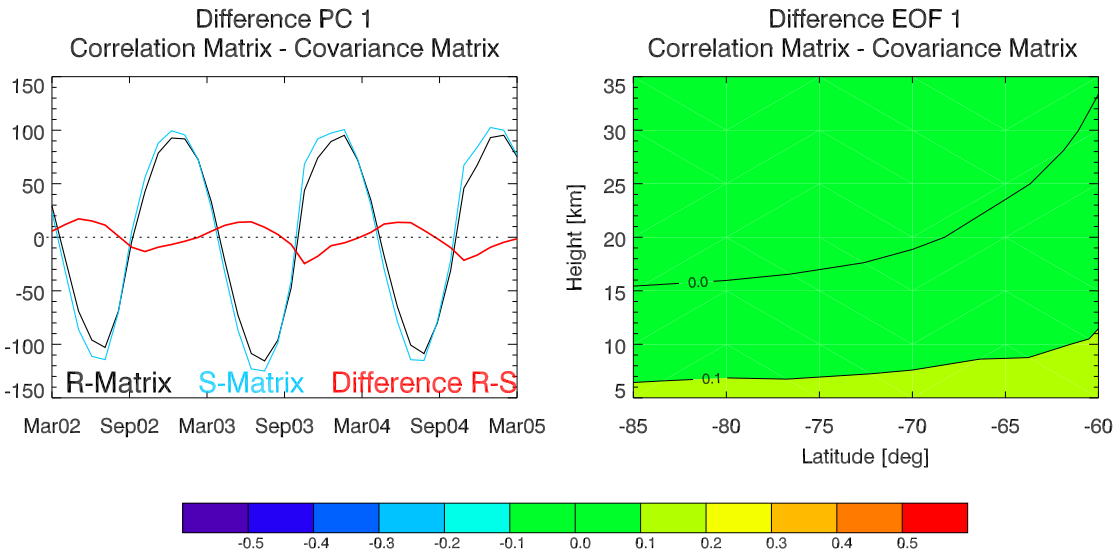


Figure 9.29: Differences between the first PC and the first EOF if they are calculated with the aid of the correlation matrix and the covariance matrix.

Comparison Between Sample Correlation Matrix and Sample Covariance Matrix

As mentioned in the discussion of the Eurasian-African data set, the renormalization of the data inhibits a correct comparison between the results given from the correlation matrix and from the covariance matrix. Nevertheless, it can be noticed that the detected patterns are similar to each other and no large differences occur. Both principal components show a sinusoidal cycle and the appearance of both coefficients turns out to be uniformly colored. A difference plot between the principal components and the coefficients calculated by means of the correlation matrix and the covariance matrix depicts that there are small deviations in the detected patterns (shown in Figure 9.29).

Concerning the principal components the difference amounts up to ± 20 units. Generally, the PC, which is based on the covariance matrix is the stronger one. The deviations (correlation matrix minus covariance matrix), which result in the coefficients are not noticeable. The differences are very small, they always remain smaller than ± 0.13, whereas small positive deviations can be found below 8 km height.

Generally, the results given from the correlation matrix and from the covariance matrix are in a good agreement. Therefore, if the interpretation of that data field is the priority objective of principal component analysis, it does not matter, which matrix is used.

Comparison Between PCA and FA Applying the Sample Correlation Matrix

The south polar region was chosen as the second data set to be investigated in detail according to differences between PCA and FA in general (the Eurasian-African slice, discussed in Section 9.4 represented a global data set, whereas the south polar area stands for a regional one). For the 3-year mean centered south polar temperature fields, both iterative principal FA and true FA yielded correct solutions (cf., Section 9.2). As maximum likelihood FA and centroid FA only allowed one factor to be extracted to achieve a meaningful result (total variance less than 100 %), these two methods were not included in the comparison.

The results of PCA correspond nearly perfectly to those of iterative principal FA (cf., Figure 9.30). While the coefficients and factor loadings are nearly identical (lower two graphs), small variations are given regarding the coefficients/factor scores. At the amplitudes maxima, the iterative principal FA values exceed the PCA results by 3 units (5 units in regard to the second factor) on average. A glance at the unique variance matrix $\mathbf{\Psi}$ of iterative principal FA shows that for this data set, only a very small part of the total variances were attributed to the unique ones. For the largest part in the south polar region, only 10 % to 20 % of the explained variances were put into $\mathbf{\Psi}$, larger values are solely found at low altitudes (5 km to 8 km). In other words, iterative principal FA did not succeed in this case in splitting up the existing variance in a common and unique part. Theoretically this stands for the fact that, provided that the values of $\mathbf{\Psi}$ are only very small, the mathematical model of FA approaches the PCA model and it becomes the same, when $\mathbf{\Psi}$ is equal to zero. If this was the case the results of both methods would be the same.

True factor analysis performs the separation of common and unique variance even much worse. Nevertheless, larger differences to PCA occur than it was the case regarding iterative principal FA. Certainly, the PCA components correspond to the true FA factor scores, but the coefficients and factor loadings vary by around 2 %. Anyhow, these small differences may be caused by the mathematical transformations, which were necessary to be capable of computing the auxiliary matrix $\boldsymbol{R}^*$ (cf., Section 9.2) and therefore to achieve true FA results.

Interpretation

First Principal Component and First Coefficients: The sinusoidal curve of the first principal component depicted in Figure 9.31 top, clearly reflects the seasonal cycle. The maximum temperature anomaly arises in December and January each year, lowest values can be found in July and August.

Compared to the results arising at the high southern latitudes yielded from the investigation of the temperature field in the Eurasian-African sector, the first principal component and the corresponding coefficients show opposite signs, but recomposing them results in the same summer and winter dependent temperature cycle.

The structure arising in the not rotated coefficients is homogeneous across all latitudes and all heights. The pattern changes a little bit after the varimax rotation, then the coefficients show a faint height-dependent formation.

Second Principal Component and Second Coefficients: The second principal component depicted in Figure 9.32 top, also shows the seasonal cycle with the phase being shifted with regard to the temporal variation of the first PC. The maxima emerge in October and November, minima occur in June and July.

The respective coefficients show a height dependent pattern, altitudes above 21 km display an inverted structure to heights below 21 km. Reconstructing the data from the first and the second PCs and coefficients, reproduces the real situation very well. The reconstructed fraction of the second factor is primarily responsible for the earlier temporal warming and cooling of the higher altitudes compared to near surface regions.

Because the third principal component on the one hand only accounts for 1.99 % of the total variability and on the other hand also shows a seasonal cycle, it will not be discussed.

Accounted Variance: The contribution to the amount of total variance of not rotated and varimax rotated coefficients/loadings is shown in Table 9.22. It can be noticed that already the first and the second extracted factor account together for more than 95 % of the total variation. Also for that reason, only two factors were extracted with iterative principal FA, so that no value can be given for the third one.

Method	Not Rotated Coefficients/Loadings		Varimax Rotated Coefficients/Loadings	
	PCA	FA	PCA	FA
$\boldsymbol{R}$, 3-Year Mean				
$\tilde{\boldsymbol{a}}_1$	78.73 %	78.64 %	50.53 %	51.30 %
$\tilde{\boldsymbol{a}}_2$	17.37 %	17.29 %	44.80 %	44.63 %
$\tilde{\boldsymbol{a}}_3$	1.99 %		1.72 %	
$\boldsymbol{S}$, 3-Year Mean				
$\tilde{\boldsymbol{a}}_1$	83.82 %		60.52 %	
$\tilde{\boldsymbol{a}}_2$	14.13 %		36.09 %	
$\tilde{\boldsymbol{a}}_3$	1.22 %		2.25 %	

Table 9.22: Accounted variances of the first three not rotated and varimax rotated coefficients/-loadings.

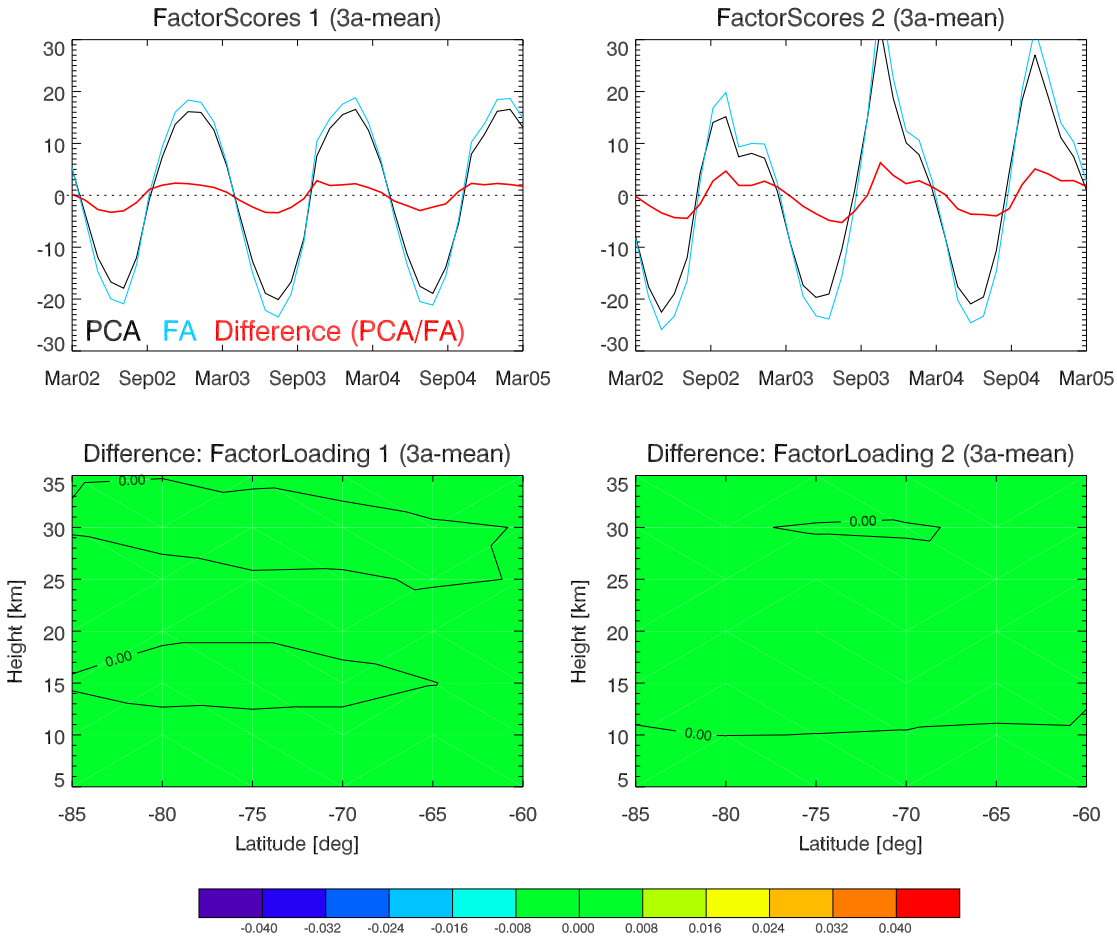


Figure 9.30: Differences between PCA and iterative principal FA in the south polar region. The upper two graphs show the PCA renormalized coefficients (black line), the iterative principal FA factor scores (blue line) and the differences between the two (red line) for the first (left) and second (right) extracted factors. The lower two graphs represent the differences between the weighting coefficients of the two methods (matrix $\boldsymbol{A}$), again for the first (left) and second (right) factor.

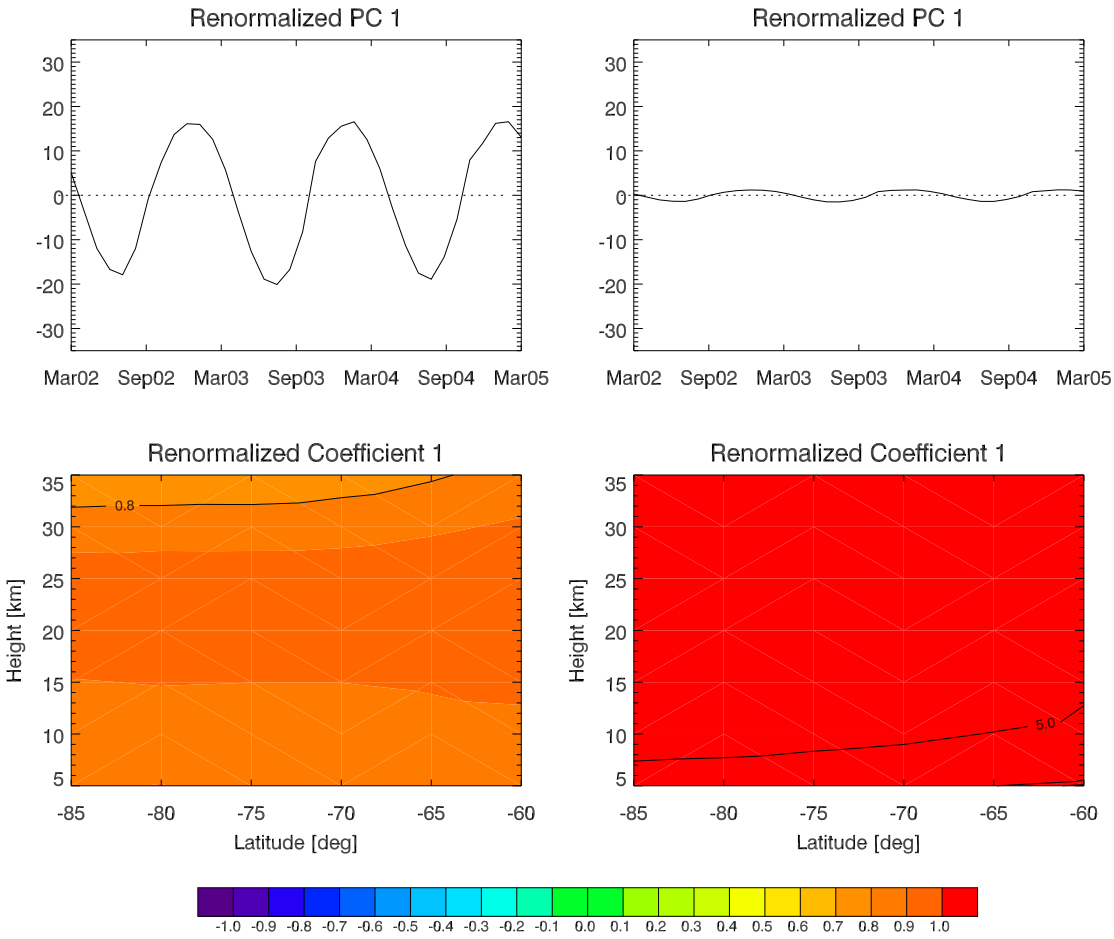


Figure 9.31: First renormalized principal component (top) and the corresponding coefficient, calculated after the elimination of 3-year mean by means of the correlation matrix (left) and the covariance matrix (right).

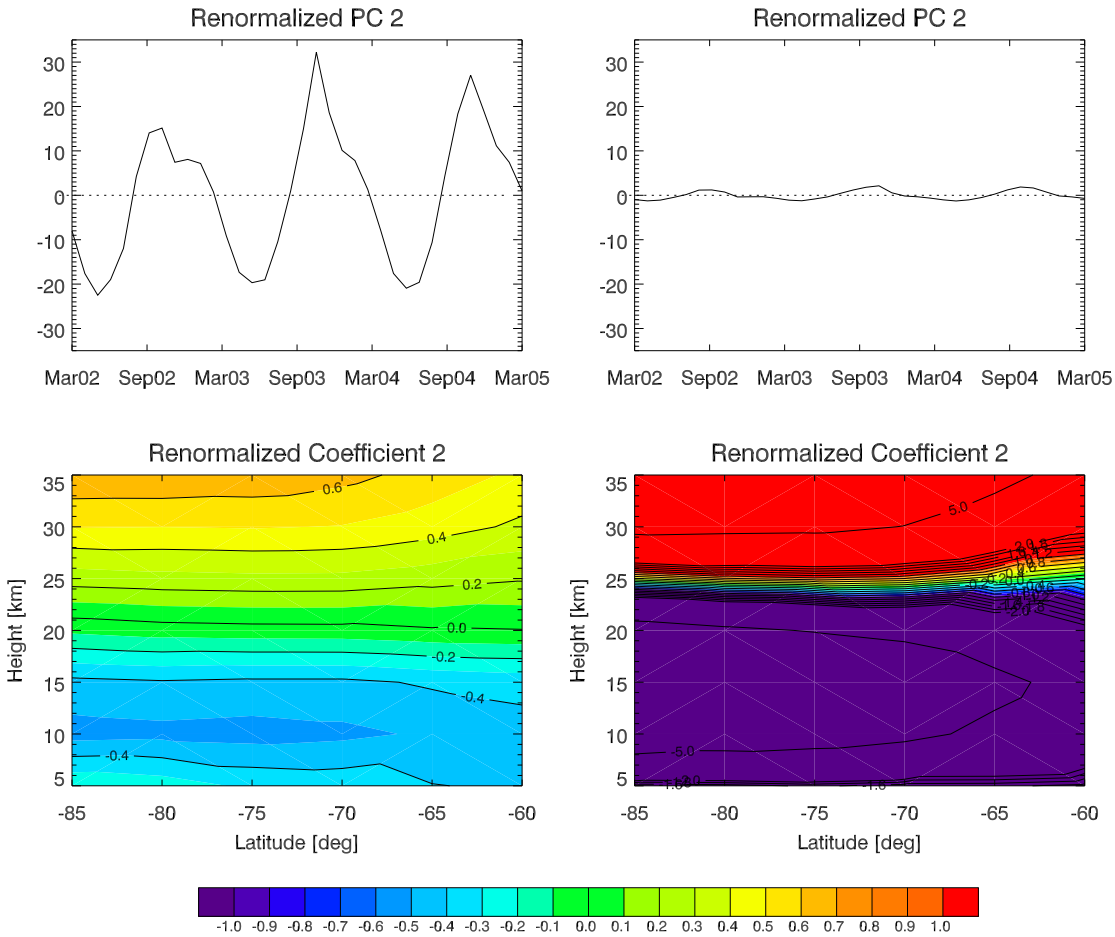


Figure 9.32: Second renormalized principal component (top) and the corresponding coefficient, calculated after the elimination of 3-year mean by means of the correlation matrix (left) and the covariance matrix (right).

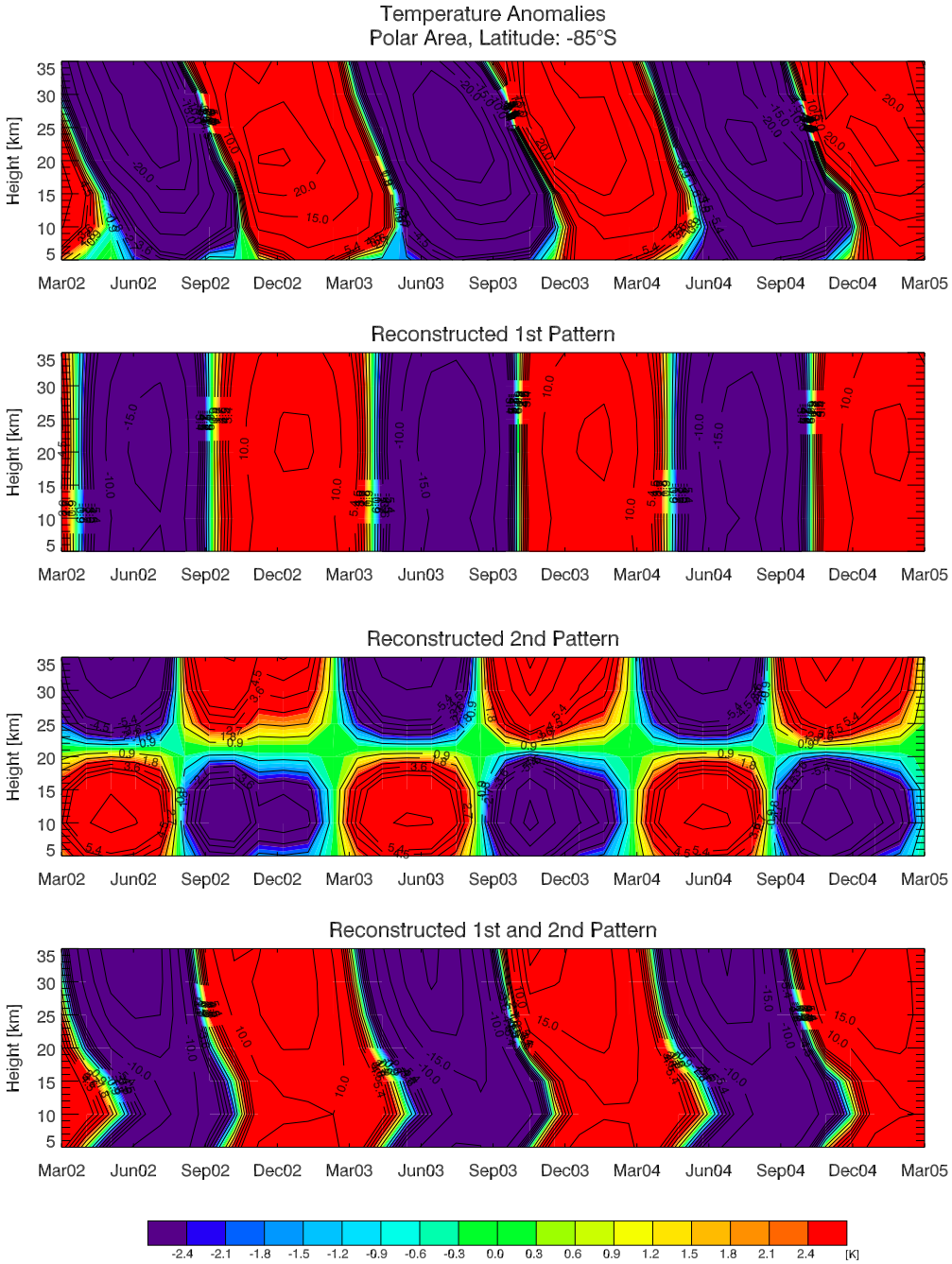


Figure 9.33: Temperature anomalies (corrected for 3-year mean), top, and reconstructed time series of the first, the second, and the combination of the first and the second principal component/coefficient at 85°S (82.5°S to 87.5°S).

9.6.2 PCA/FA of Monthly Mean Subtracted Temperature Anomalies in the South Polar Area

As seen so far, the seasonal cycle is particularly strong pronounced in high latitude regions. To look for atmospheric patterns being hidden behind the seasonal cycle, the monthly mean was subtracted from the data field in each grid point.

Number of Factors Extracted: Table 9.23 shows the number of factors, which should be retained when using them for additional calculations.

Method	Cum. Var.>90 %	Kaiser's rule	Scree Test	LEV-Test	FA
$\boldsymbol{R}$, Monthly Mean	6	6 (8)	7	3	6 (9)
$\boldsymbol{S}$, Monthly Mean	5	5 (6)	6	3	

Table 9.23: Number of extracted factors k, estimated from the cumulative variance limited to 90 %, the Kaiser's rule, and the modified Kaiser's rule, the scree-test, the LEV-test as well as the number of factors used for iterative principal FA.

Compared to the number of extracted factors calculated after the elimination of 3-year mean, the number k now is larger, independent of the kind of test and of the underlying matrix. Nevertheless, the interpretation will be limited to the first two PCs/coefficients.

Eigenvalues: The first three eigenvalues of both matrices are shown in Table 9.24.

Matrix	Method	λ_1	λ_2	λ_3
$\boldsymbol{R}$, Monthly Mean	PCA	15.58	11.22	4.81
	FA	15.50	11.12	4.64
$\boldsymbol{S}$, Monthly Mean	PCA	76.94	53.65	14.28

Table 9.24: Eigenvalues of the sample correlation matrix and sample covariance matrix calculated after elimination of the monthly means.

Comparison Between Sample Correlation Matrix and Sample Covariance Matrix

Similar to the results obtained from the 3-year mean corrected data set in the south polar region, the renormalized coefficients of the monthly mean centered data (Figure 9.38) show patterns, which resemble each other. Both coefficients form a height dependent structure, which is of opposite sign below and above approximately 25 km height. The intensity of the feature decreases when moving from the south pole to 60°S.

Exactly this decrease of intensity represents the difference between the results of both matrices. As can be seen in Figure 9.34, right, the difference concerning the coefficients is negative at highest latitudes and positive at lower latitudes between a height of 10 km

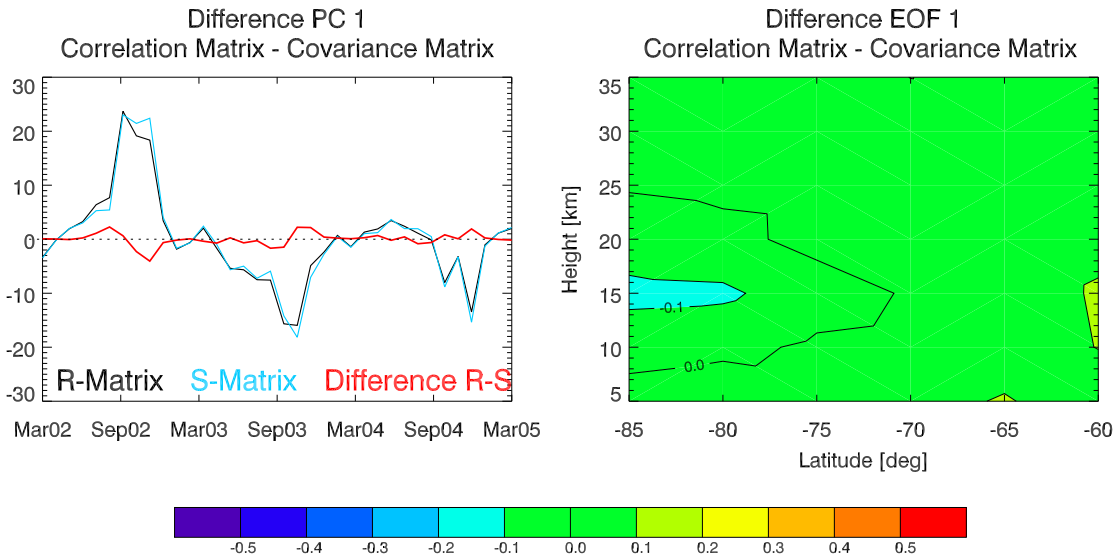


Figure 9.34: Differences between the first PC and the first EOF if they are calculated with the aid of the correlation matrix and the covariance matrix.

and 15 km. That means that the pattern detected from the covariance matrix is stronger at the south pole, whereas it is smaller at 60°S.

Looking at the shapes of the principal components as well as at the difference between them (shown in Figure 9.34, left) it can be recognized that the structures found in the coefficients (correlation matrix and covariance matrix) represent the same pattern. The cycle of the PCs look very similar and the difference always remains smaller than ±5 units.

Comparison Between PCA and FA Applying the Sample Correlation Matrix

The monthly mean centered south polar temperature field was an exception in regard to the four selected atmospheric data sets, as three FA techniques succeeded in attaining a mathematically correct solution, namely iterative principal FA, true FA, and centroid FA. The differences of these techniques to PCA should be briefly addressed.

Similar to the 3-year mean corrected south polar data set, nearly no differences are given now between PCA and iterative principal FA. In fact, the both methods agree still better: The maximal deviations between the principal components and the factor scores are even in the extreme values of the curves within the scope of less than ±1 unit (September, October 2002 and 2003), and mainly they stay very close to zero. The same goes for the coefficients/factor loadings differences, which are more or less not existent (for the coefficients/loadings, the graphs look similar to those in Figure 9.30, below).

Coming to true FA, the contrast to PCA is more clearly visible. Certainly, the components/factor scores are again quite the same, but concerning the coefficients/-

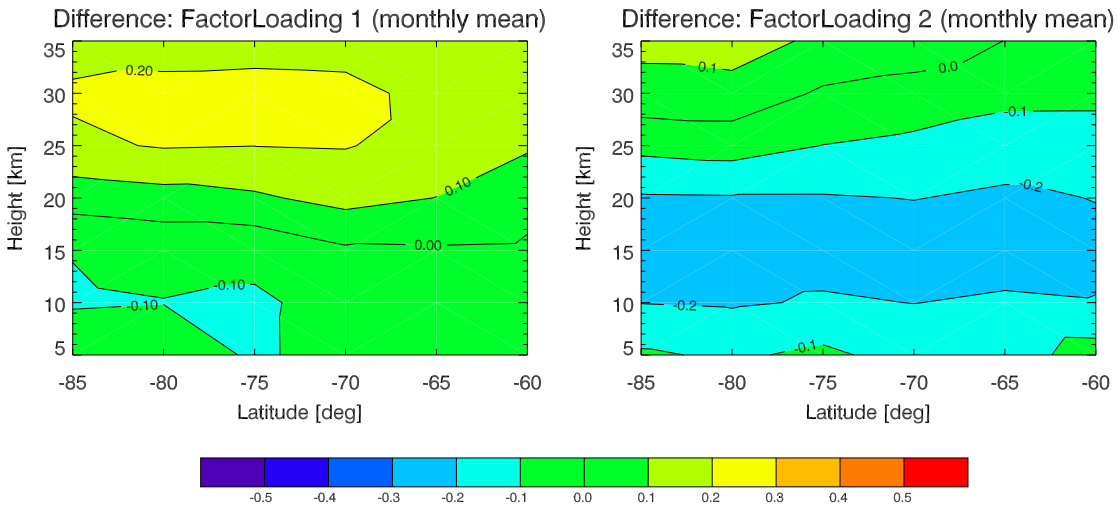


Figure 9.35: Differences between PCA coefficients and true FA factor loadings. Due to the fact that the differences between PCA and true FA as well as centroid FA are stronger pronounced than compared to iterative principal FA, the color bar range was extended to ±0.5 to achieve expressive plots.

loadings, relatively large deviations occur. Figure 9.35 shows these differences. Because of their larger amount and to enable the direct comparison to the results of the differences between PCA and centroid FA, the color bar range was extended to ±0.5 units (instead of ±0.05 units, which was taken for the comparison between the similar results of PCA and iterative principal FA) to achieve expressive plots. While for the first extracted factor (left graph) more pronounced deviations of around +0.2 units (standing for 20 % differences of the common variances) appear at higher altitudes (25 km to 34 km), differences of the same amount, but of opposite signs, are found at lower heights (10 km to 20 km) in regard to the second extracted factor (right graph).

Even larger deviations are given between PCA and centroid FA and furthermore, they are not restricted to the coefficients/loadings. Concerning the principal components and the factor scores, the arbitrariness of the sense for the eigenvectors appears again, so that PCA components/coefficients and centroid FA factor scores/loadings show opposite signs. Anyhow, as explained earlier, this fact does not influence the results as such. The differences between PCA components and centroid FA factor scores are mainly given by a temporal lag of the curves, which mainly occurs for centroid FA. The factor scores of this method seem to react more slowly to temporal changes than the components of PCA, so that the most pronounced deviations (achieving nearly the same amplitude as the components/factor scores themselves) occur simultaneously to the largest amplitudes (during September, October of each year). Nevertheless, beyond the extrema the curve match quite well. The actual differences between PCA and centroid FA turn up in the coefficients/loadings. As the signs of the two methods' results were inverse, the PCA coefficients were added to the FA factor loadings. To avoid getting misled, this has to

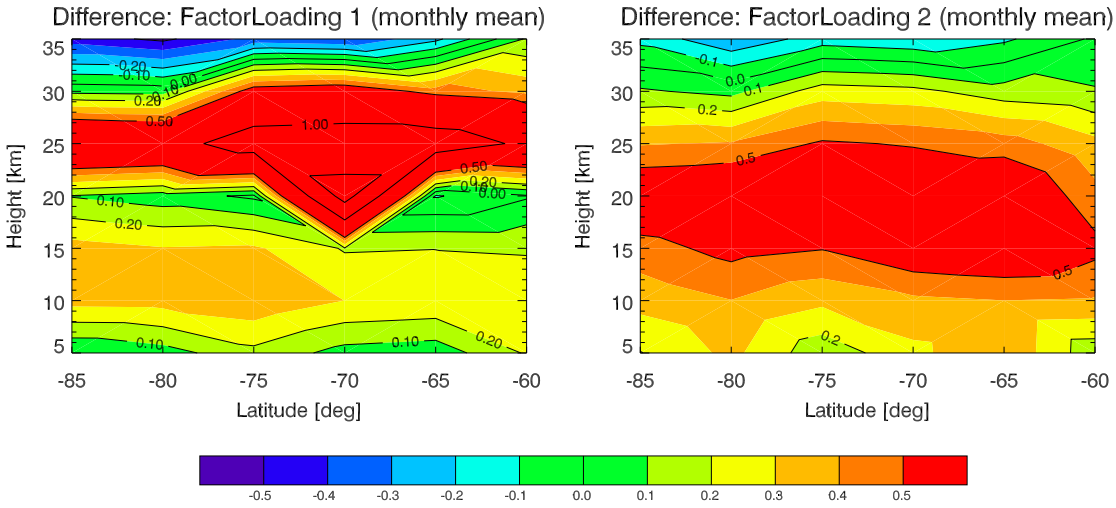


Figure 9.36: Differences between PCA coefficients and centroid FA factor loadings.

be kept in mind when interpreting Figure 9.36. Even if for both extracted factors the deviations between the two methods are mainly positive, centroid FA achieved similar structures (of opposite signs) for its factor loadings in regard to PCA. The differences at the whole tend to become positive, on the one hand because positive structures were more pronounced in PCA than the respective negative structures in centroid FA, and on the other hand because negative structures in PCA were less pronounced than the respective positive ones in centroid FA (so the sum of the coefficients and loadings was rather positive).

The most striking difference in regard to the first extracted factor (see left graph of Figure 9.36) is given by a triangle-shaped structure between a height of 15 km and 27 km. This strange structure also exists in the first factor loadings of centroid FA and it points to a problem of the method concerning the spatial allocation of common variances. Furthermore, the pronounced positive PCA structure below 20 km is less developed in centroid FA' loadings, where it is rather concentrated between 18 km and 20 km.

Compared to the first extracted factor, the coefficients/loadings differences between PCA and centroid FA are less intensive for the second extracted factor, but they still surmount the differences between PCA and the other two FA techniques essentially.

The question is, whether these larger deviations between PCA and true FA as well as centroid FA could be caused by the fact that these two techniques are more successful in separating the total variance of a data set into common and unique one, than iterative principal FA.

A glance at Figure 9.37 shows that the unique matrices $\mathbf{\Psi}$ do not contain the hoped-for explanations. Only centroid FA (right graph) seems to manage the separation of the unique variances, whereas iterative principal FA (left graph) and above all true FA (in

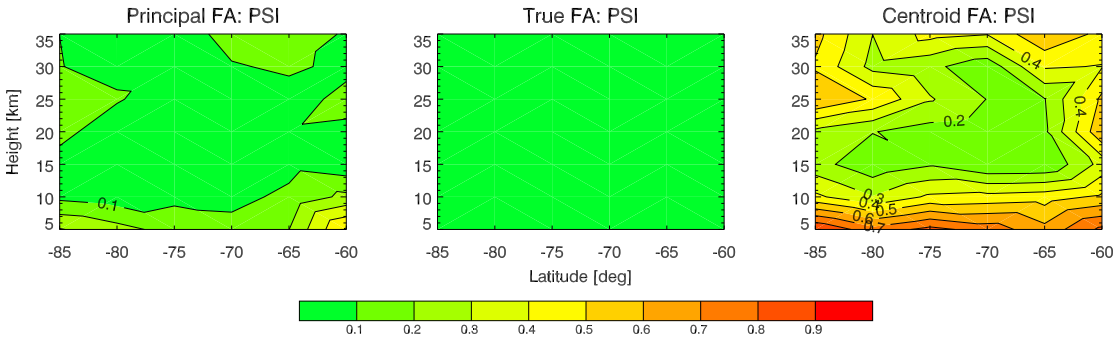


Figure 9.37: Unique variance matrices $\mathbf{\Psi}$ of iterative principal FA (left), true FA (middle), and centroid FA (right). The green color stands for small, the red for large unique variances.

the middle) attribute just a small amount of the total variance to $\mathbf{\Psi}$.

This is probably caused by the technique specific different number of extracted factors: Only two sectors could be extracted for centroid FA (conditional on the requirement that all ψ_{ii} have to be positive), so that the technique had more problems to extract the common variances, as only the correlations of two factors could be considered. The common variance explained by these two factors achieved barely 60 %, so that an appreciable part of the total variance rested for the unique variances (20 % to 50 % for the largest part in the considered region).

In contrast, 24 factors had to be extracted applying true FA, where the mathematical constraint requires the weighting factor $\hat{\theta}$ to be less than one (cf., Section 9.2.2). This large number of extracted factors resulted in 99.99 % of the variance explained by the 24 common factor loadings, so that as good as no variance was left for the unique variances in $\mathbf{\Psi}$, which can be seen clearly in the middle graph of Figure 9.37.

The unique variances resulting from iterative principal FA are as well rather small (mostly less than 20 %), which is the result of the high value of explained common variance (nearly 90 %), which was achieved by six extracted factors.

Even though the differences between the FA techniques seem to be caused by the varying number of extracted factors, which results in different unique variance matrices $\mathbf{\Psi}$, the deviations to PCA results cannot be explained by this fact in their entirety. The heterology of PCA and FA techniques probably has to be tracked down in the different mathematical calculation procedures, but this was out of the scope of this work.

Interpretation

First Principal Component and First Coefficients: The first principal component shows three characteristics arising in September and October each year. In 2002 the feature is of opposite sign compared to 2003 and 2004. Investigating the first coefficients, it can be noticed that the structure shows a height dependent behavior with the formation at the lower heights being more pronounced compared to higher altitudes. The formations are again of opposite signs, the parting line is at an altitude of about

28 km. The latitudinal dependency is insignificant.

The reconstruction of the first PC and the respective coefficients yields a pattern from a height of 5 km to 27 km appearing first in May 2002 and 2003 and lasting until December 2002 and January 2004, respectively. In 2002, a strong positive temperature anomaly is developed, in 2003 it is a negative temperature pattern. In 2004, the distinct structure only arises from September until December. The pattern between a height of 5 km and 27 km coincidents with a small counterpart that can be found above 30 km altitude. The same structure also has been found in the monthly mean corrected temperature fields in the Eurasian-African sector as well as the 15 km height map, where it was attributed to the southern polar vortex (cf., Sections 9.4.2 and 9.5.2). The correctness of the detected structures can be verified by several papers (Angell et al. 2002, 2003b, 2004b; Gobiet et al. 2005a).

Second Principal Component and Second Coefficients: The second principal component shows a spike in August and September 2002 and another in October and November the same year of opposite sign. The same feature can be noticed in 2003, being opposed to the structures of 2002, whereas in 2004 only a small "unipolar" peak can be discovered. The second coefficients resolve a structure occurring above 20 km height, below an inverse pattern is less pronounced.

The reconstructed time series shows the features emerging in the months detected from the principal components, an alternating pattern arising in August/September and October/November 2002 and 2003. The year 2004 gets out line, the discussed pattern cannot be observed there.

Accounted Variance: The contribution to the amount of total variance of not rotated and varimax rotated coefficients/loadings is summarized in Table 9.25. It can be noticed that the stated variance is spread over the first three coefficients more regularly compared to the 3-year mean eliminated coefficients/loadings. That is caused by the absence of the most dominating pattern, namely the seasons. The extremely similar results of PCA and iterative principal FA are here also very well reflected.

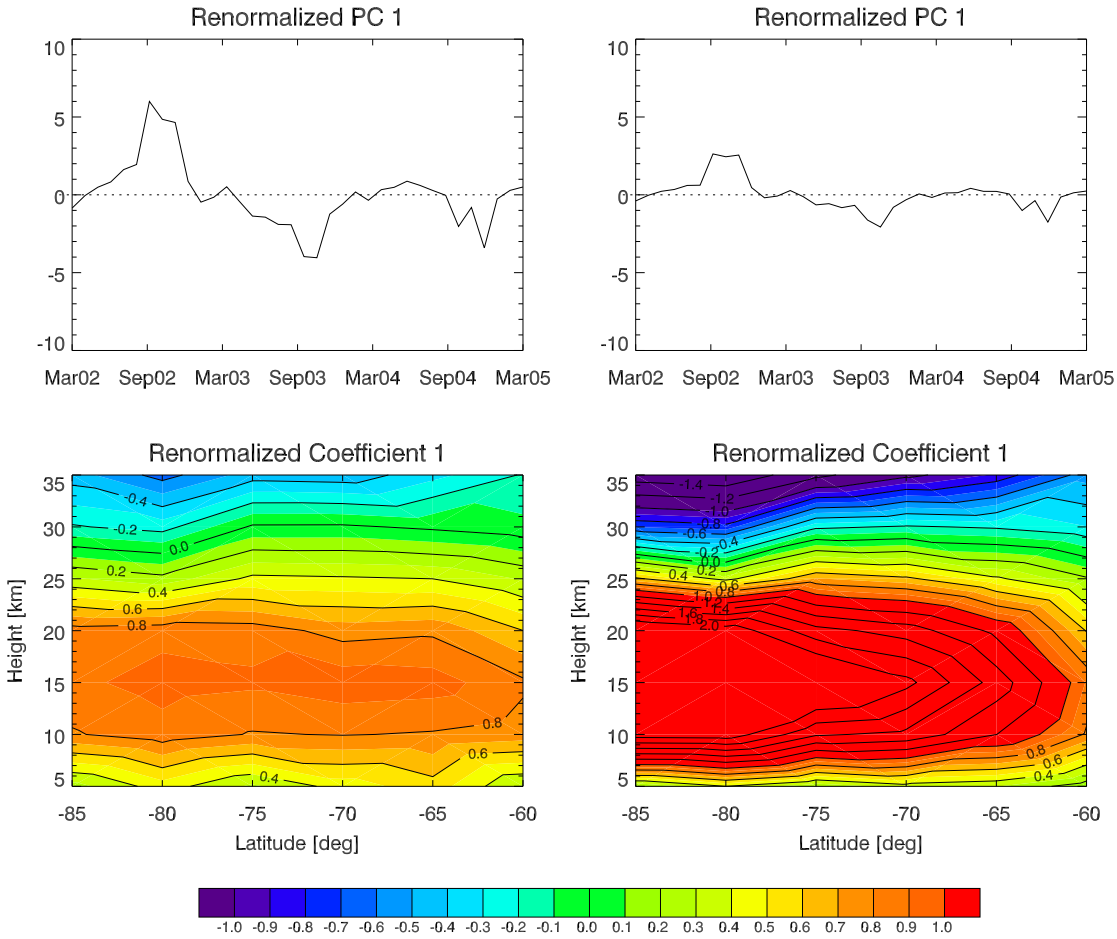


Figure 9.38: First renormalized principal component (top) and the corresponding coefficient, calculated after the elimination of the seasonal mean by means of the correlation matrix (left) and the covariance matrix (right).

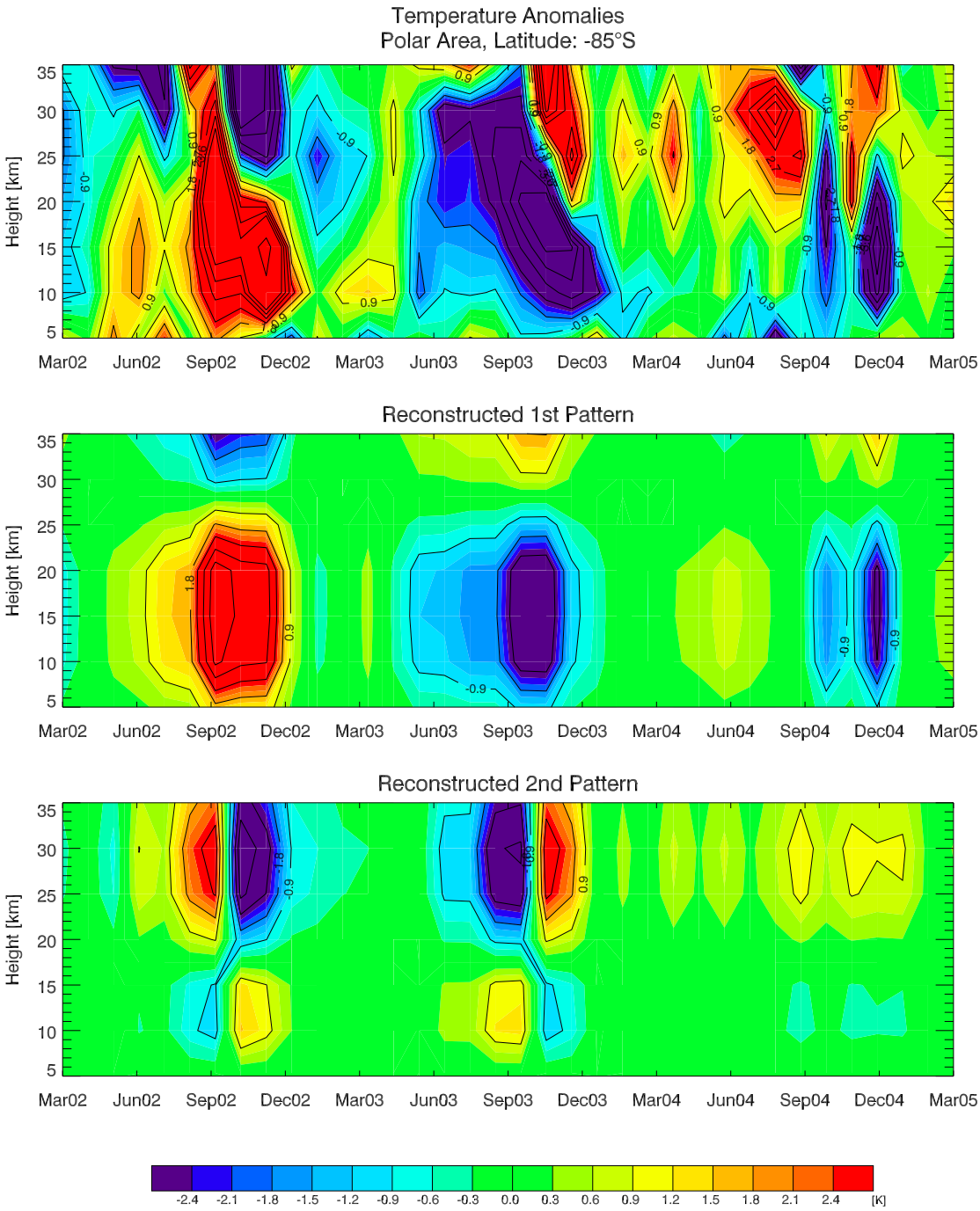


Figure 9.39: Temperature anomalies (corrected for monthly mean), top, and reconstructed time series of the first and second principal component (middle and beneath) at 85°S (82.5°S to 87.5°S).

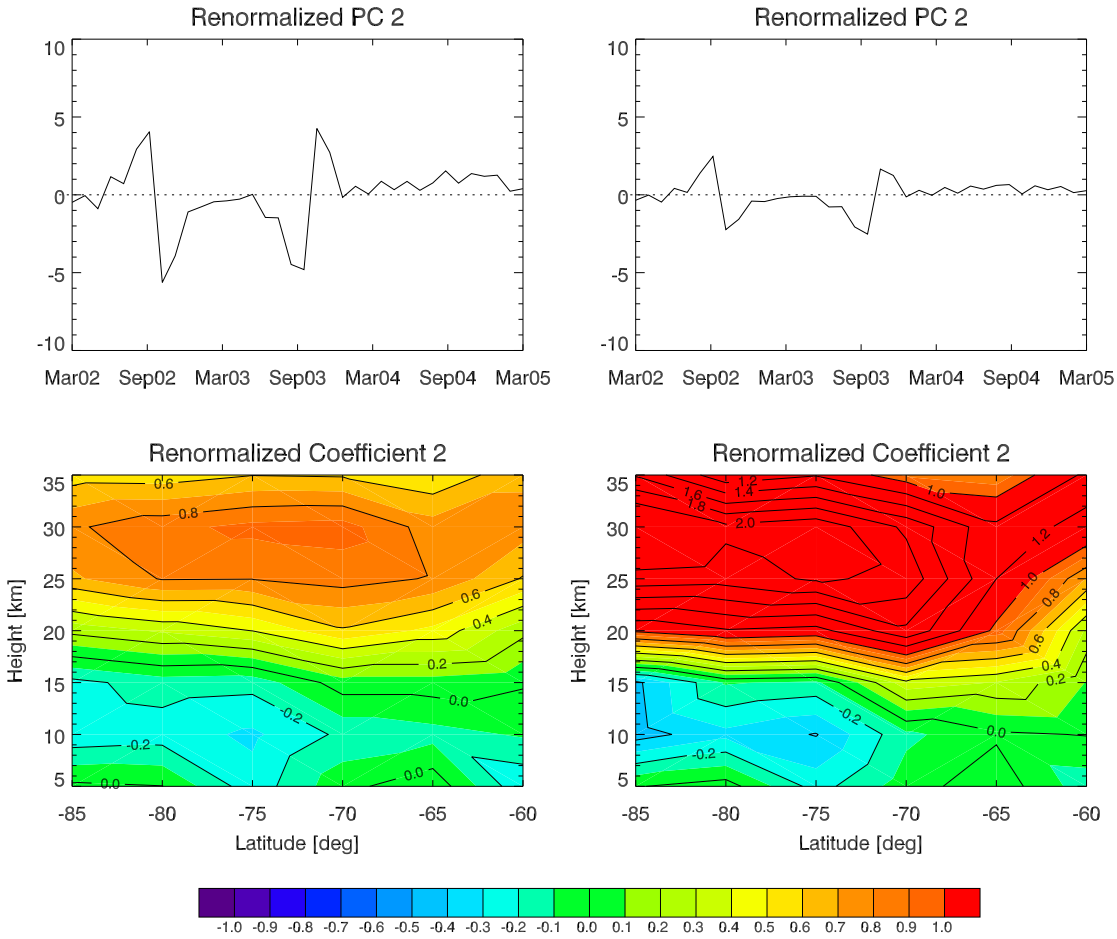


Figure 9.40: Second renormalized principal component (top) and the corresponding coefficient, calculated after the elimination of the seasonal mean by means of the correlation matrix (left) and the covariance matrix (right).

Method	Not Rotated Coefficients/Loadings		Varimax Rotated Coefficients/Loadings	
	PCA	FA	PCA	FA
$\boldsymbol{R}$, Monthly Mean				
$\tilde{\boldsymbol{a}}_1$	37.09 %	36.90 %	28.23 %	28.21 %
$\tilde{\boldsymbol{a}}_2$	26.69 %	26.48 %	20.89 %	20.79 %
$\tilde{\boldsymbol{a}}_3$	7.27 %	11.05 %	10.85 %	12.03 %
$\boldsymbol{S}$, Monthly Mean				
$\tilde{\boldsymbol{a}}_1$	43.23 %		34.71 %	
$\tilde{\boldsymbol{a}}_2$	30.14 %		25.30 %	
$\tilde{\boldsymbol{a}}_3$	8.02 %		11.06 %	

Table 9.25: Accounted variances of the first three not rotated and varimax rotated coefficients/-loadings.

9.7 Temperature Data Near the Tropical Tropopause

(Authors: B.C. Lackner and B. Pirscher)

The second analyzed regional field is located at the low latitudes between 17.5°S and 17.5°N (5° zonal means yielding 7 latitudinal regions) to investigate the tropical atmosphere between 12 km and 22 km.

Generally, the seasonal fluctuations are less pronounced at the low latitudes compared to high latitudinal regions. Hence, the results of PCA and FA of the 3-year mean centered tropical temperature field are similar to the output of the monthly mean subtracted temperatures.

9.7.1 PCA/FA of Temperature Anomalies Near the Tropical Tropopause

Number of Factors Extracted: Table 9.26 shows the number k of extracted factors estimated with different selection rules applied to the correlation matrix and the covariance matrix as well as the number of factors extracted with iterative principal FA with the maximal number of extractable factors (due to mathematical constraints) in parenthesis. The results are shown for both data, being 3-year mean centered and monthly mean centered.

Method	Cum. Var.>90 %	Kaiser's rule	Scree Test	LEV-Test	FA
$\boldsymbol{R}$, 3-Year Mean	4	4 (6)	5	9	4 (17)
$\boldsymbol{S}$, 3-Year Mean	3	4 (4)	5	5	
$\boldsymbol{R}$, Monthly Mean	4	5 (6)	4	7	5 (13)
$\boldsymbol{S}$, Monthly Mean	3	3 (4)	4	5	

Table 9.26: Number of extracted factors k determined with different selection rules and according to mathematical constraints in case of iterative principal FA.

As in the previous sections, the selection rules yield a number of factors being greater than the number of factors being interpreted. The tropical data set takes a special position for iterative principal FA in regard to the maximal number of extractable factors, which is higher than for all other analyzed atmospheric data sets (given in parenthesis in Table 9.26). Nevertheless, a much smaller number of extracted factors (namely four and five) satisfied to explain around 92 % of the total variance in the data set.

Eigenvalues: The first three eigenvalues of the sample correlation matrix and the sample covariance matrix calculated after the removal of 3-year mean and monthly mean are shown in Table 9.27. The iterative principal FA quoted values are again derived from the factor loadings of the first three extracted factors.

In both cases, the PCA results correspond very well with the iterative principal FA results, independently if rotated or not.

Matrix	Method	λ_1	λ_2	λ_3
$\boldsymbol{R}$, 3-Year Mean	PCA	24.68	8.85	3.38
	FA	24.61	8.79	3.28
$\boldsymbol{S}$, 3-Year Mean	PCA	115.08	43.42	11.13
$\boldsymbol{R}$, Monthly Mean	PCA	22.02	9.24	5.19
	FA	21.96	9.20	5.07
$\boldsymbol{S}$, Monthly Mean	PCA	64.13	31.52	6.26

Table 9.27: Eigenvalues of the sample correlation matrix $\boldsymbol{R}$ and the sample covariance matrix $\boldsymbol{S}$ once without the mean of each variable (3-year mean), the other after elimination of the monthly mean.

Furthermore, it can be noticed that the 3-year mean eliminated eigenvalues are always a bit larger than the eigenvalues calculated after the subtraction of monthly means. The reason is the remaining variance being larger in the first case. Besides, the eigenvalues of the correlation matrix are smaller than the eigenvalues of the covariance matrix, an aspect that was mentioned already several times as this is also valid for the other atmospheric data sets.

First Principal Component and First Coefficients: First of all, the common properties of the 3-year mean and the monthly mean centered renormalized principal components and coefficients are discussed.

As mentioned above, the low latitudes do not show the same temporal behavior as the high northern and southern latitudes, because the seasons are less pronounced. The comparison of the temperature anomalies confirms the low seasonal impact in the tropic region, especially at the equator. Beyond the equatorial region, the picture changes slowly and the seasons gain influence on the temperature data.

The first principal component and the first coefficients of the temperature anomalies calculated after the removal of 3-year mean and after the elimination of monthly mean are depicted in Figures 9.41 and 9.42.

The varimax rotated coefficients are not shown because they are similar to the not rotated coefficients and do not facilitate the interpretation. Comparing the monthly mean centered principal components and the 3-year mean centered PCs calculated from both matrices, it seems that the first ones are a smoothed version of the second ones. Generally, they show the same drift, but the 3-year mean centered temperatures express a stronger fluctuating structure. The corresponding coefficients are also very similar and no essential differences can be noticed. This impression is confirmed by comparing the reconstructed time series.

Looking at the results of the correlation and covariance matrix based PCs/coefficients yields that the principal components are equal apart from the magnitude. The correlation matrix based coefficients show a slight bipolar structure, whereas an unipolar pattern can be found in the covariance matrix based coefficients over the equator at a

height of approximately 18 km (in case of the monthly mean elimination).

The monthly mean centered time series of temperature anomalies at the equator (Figure 9.43, top) shows alternating anomalies (negative from March 2002 to June 2003 and from December 2003 to March 2005; positive–overlapping–in between) propagating downwards, which suggests that the detected pattern stems from the quasi-biennial oscillation (QBO).

This atmospheric pattern dominates the variability of the equatorial stratosphere between a height of approximately 15 km to 50 km. It is symmetrical in regard to the equator and extends from maximal 15°S to 15°N. Easterly and westerly winds (20 m/s to 30 m/s) alternate in an averaged period of 28 months, influencing the temperature variability. Easterly winds are connected with low temperatures; westerly winds with comparatively high temperatures. An extra-tropical counterpart is situated in the polar regions. During equatorial easterly winds, the polar vortex is warmer and more disturbed, than during westerly winds (Salby 1996; Hupfer and Kuttler 1998; Baldwin et al. 2001).

The QBO influence on the polar vortex is also visible in the time series of CHAMP RO data. The disturbances of the polar vortex in southern hemisphere winter 2002 (cf., Figures 9.19 and 9.39) and in northern hemisphere winter 2002/2003 (cf., Figure 9.21), which were interpreted as sudden stratospheric warmings, coincide with the strong negative temperature anomalies in tropical regions.

Schmidt et al. (2005) already detected the QBO pattern in the CHAMP data set. They also analyzed monthly mean corrected temperatures and created plots similar to Figure 9.43.

Moving from the equator toward the subtropics, the influence of the QBO decreases and the seasons gain influence.

The conspicuous pattern located between a height of 13 km and 17 km (found above Eurasia-Africa and described in detail in Section 9.4.1) cannot be noticed in the zonal mean temperature anomalies between a height of 12 km and 22 km.

Concerning the amount of accounted variance of the first principal component, Table 9.28 shows that it is comparatively small compared to the first PCs in case of the south polar region or the global fields. Dependent on the examined data matrix (correlation or covariance matrix, 3-year mean or monthly mean subtracted), the accounted amount of variance ranges between 52 % and 62 %.

Second Principal Component and Second Coefficients: The renormalized principal component, calculated after the removal of the 3-year mean, again corresponds to the monthly mean centered PC. The same goes for the coefficients.

Another aspect is that the second coefficients calculated by means of the correlation matrix have the same structure as the second covariance matrix based coefficients.

Figure 9.44 depicts the second principal components and its respective coefficients calculated by means of the correlation matrix (left) and the covariance matrix (right) after elimination of the monthly mean. A height dependent pattern can be recognized both times with opposite signs. Despite the short time series, the principal components

show a sinusoidal character with a frequency of more or less two years, which is consistent with the QBO theory.

As can be seen in Figure 9.43, the second PC/coefficients is responsible for the resolution of the height dependent temporal behavior of the QBO. Anyhow, it always must be considered that even though the factors generally are assumed to be orthogonal (and therefore, the coefficients should be based on different physical patterns), this is not the case in the investigated atmospheric data fields.

Accounted Variance: A summary of the contribution to the amount of total variance of not rotated and varimax rotated coefficients/loadings is shown in Table 9.28. It can be recognized that the accounted variances of the correlation matrix and the covariance matrix are considerably larger if the mean of each variable is eliminated before calculating the principal components/loadings compared to the case of the removal of the seasonal impact.

Method	Not Rotated Coefficients/Loadings		Varimax Rotated Coefficients/Loadings	
	PCA	FA	PCA	FA
$\boldsymbol{R}$, 3-Year Mean				
$\tilde{\boldsymbol{a}}_1$	58.77 %	58.59 %	38.68 %	38.73 %
$\tilde{\boldsymbol{a}}_2$	21.08 %	20.93 %	28.05 %	27.41 %
$\tilde{\boldsymbol{a}}_3$	8.05 %	7.81 %	15.65 %	15.99 %
$\boldsymbol{S}$, 3-Year Mean				
$\tilde{\boldsymbol{a}}_1$	61.47 %		48.04 %	
$\tilde{\boldsymbol{a}}_2$	23.19 %		28.92 %	
$\tilde{\boldsymbol{a}}_3$	5.94 %		9.05 %	
$\boldsymbol{R}$, Monthly Mean				
$\tilde{\boldsymbol{a}}_1$	52.43 %	52.28 %	37.41 %	37.50 %
$\tilde{\boldsymbol{a}}_2$	22.01 %	21.91 %	30.09 %	29.58 %
$\tilde{\boldsymbol{a}}_3$	12.35 %	12.07 %	16.67 %	17.21 %
$\boldsymbol{S}$, Monthly Mean				
$\tilde{\boldsymbol{a}}_1$	58.34 %		47.62 %	
$\tilde{\boldsymbol{a}}_2$	28.68 %		36.04 %	
$\tilde{\boldsymbol{a}}_3$	5.69 %		6.81 %	

Table 9.28: Accounted variances of the first three not rotated and varimax rotated coefficients/-loadings.

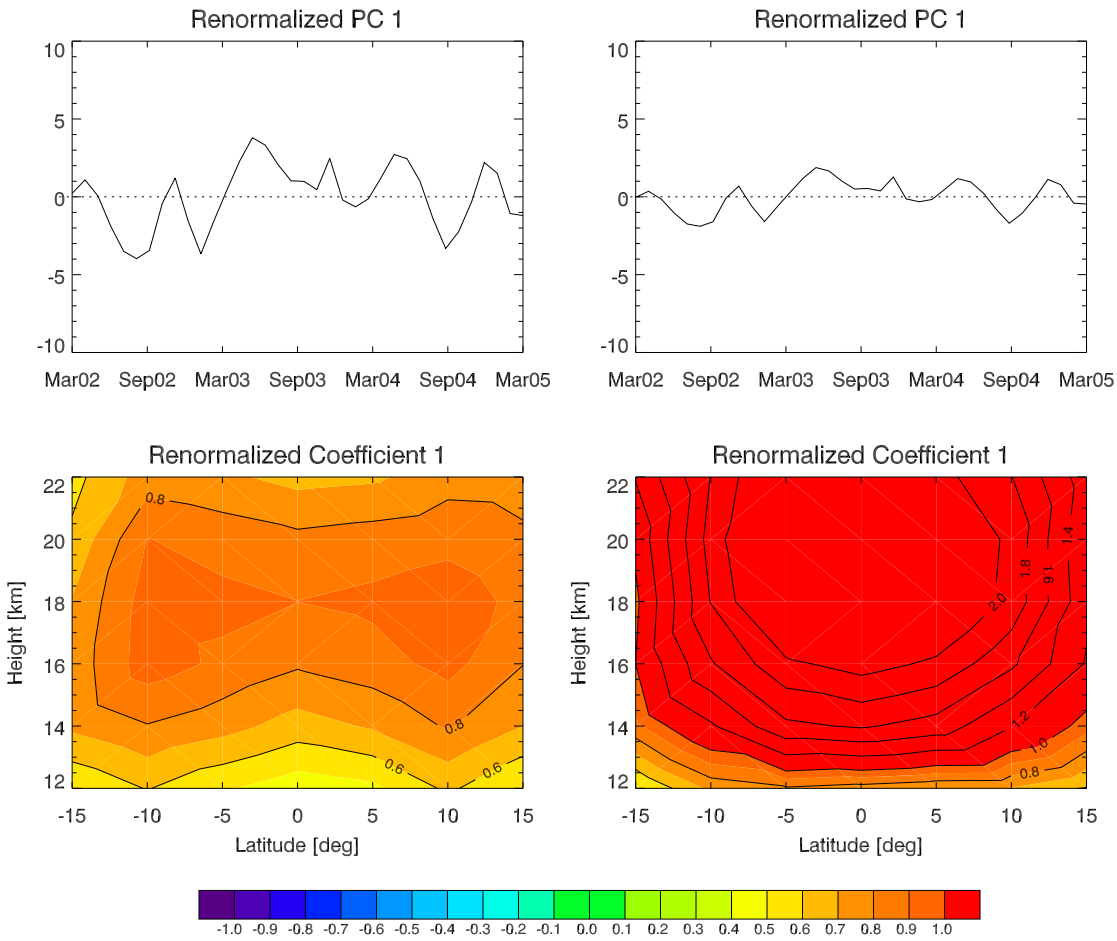


Figure 9.41: First renormalized principal components (top) and the corresponding coefficients (bottom), calculated after the elimination of 3-year mean by means of the correlation matrix (left) and the covariance matrix (right).

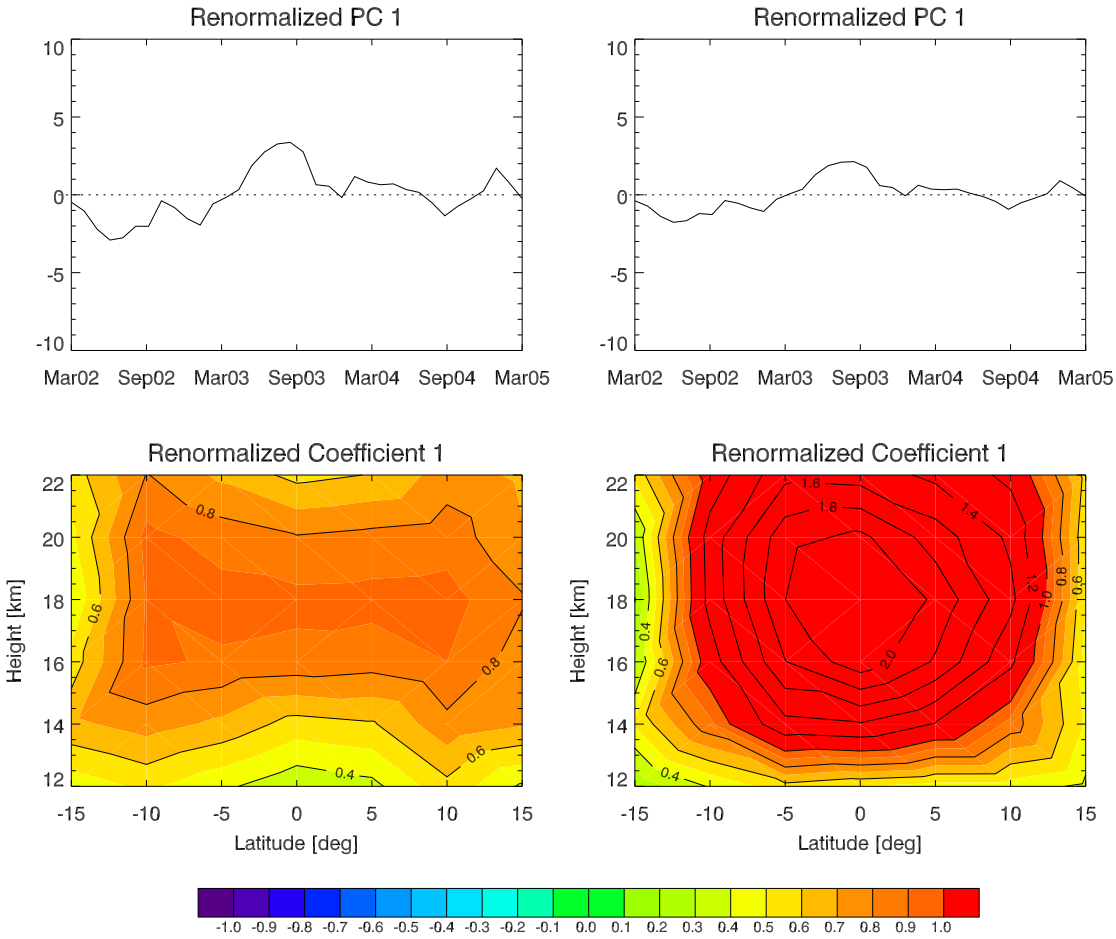


Figure 9.42: First renormalized principal components (top) and the corresponding coefficients, calculated after the elimination of the monthly means by means of the correlation matrix (left) and the covariance matrix (right).

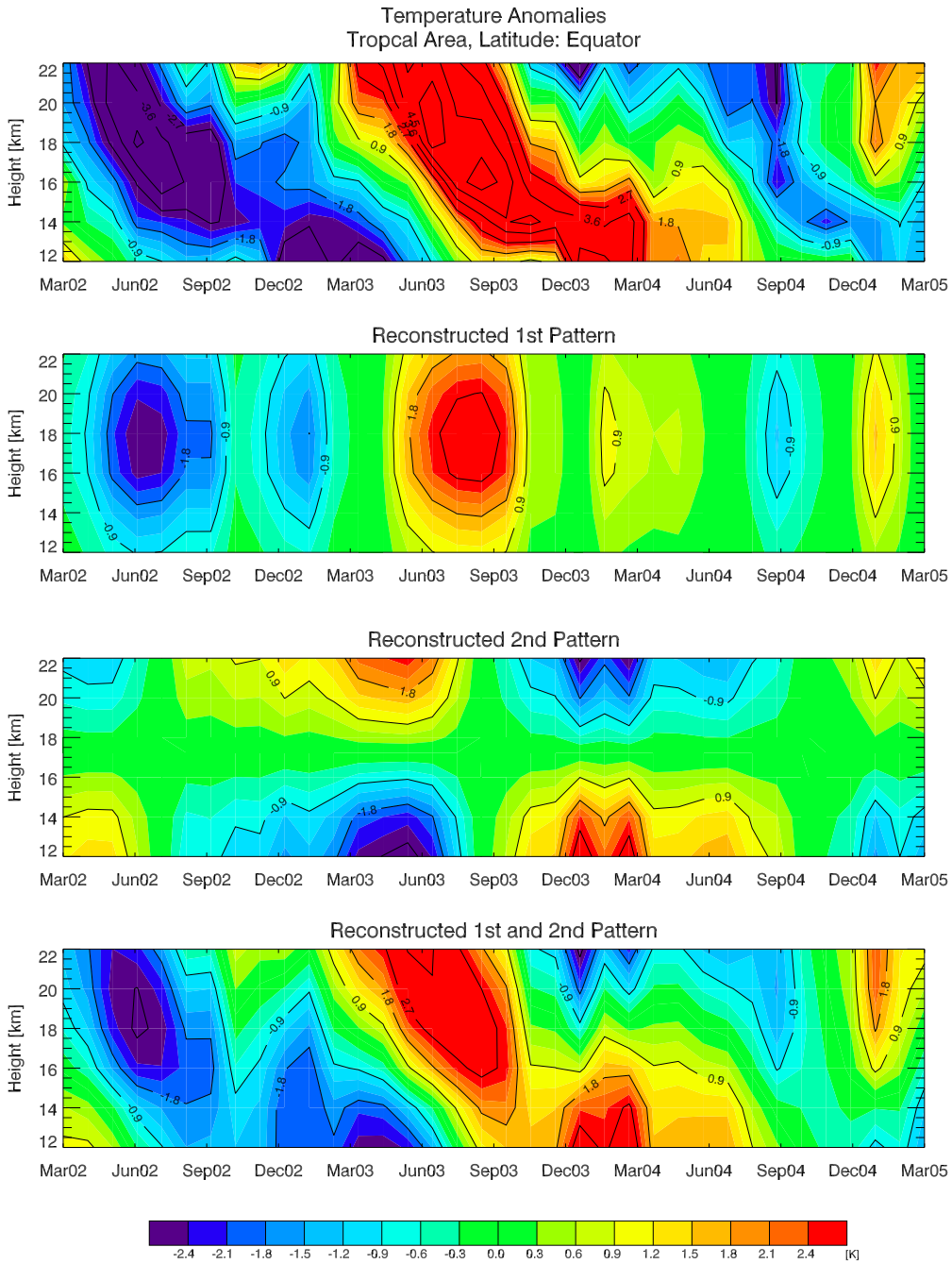


Figure 9.43: Temperature anomalies (corrected for the annual cycle), top, and reconstructed time series of the first and second principal component (middle and beneath), as well as a combination of the first and second factors at the equator (2.5°S to 2.5°N).

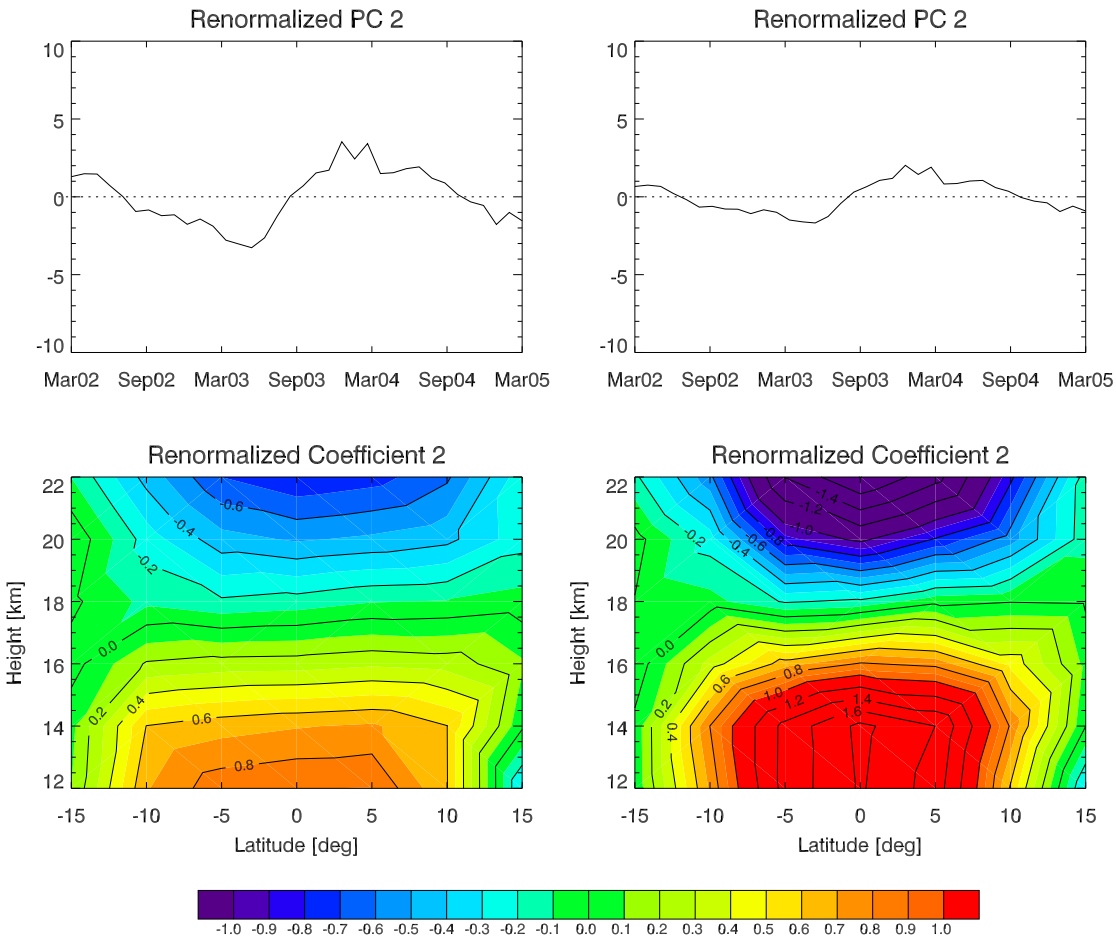


Figure 9.44: Second renormalized principal components (top) and the corresponding coefficients, calculated after the elimination of the monthly means by means of the correlation matrix (left) and the covariance matrix (right).

10 Conclusions

(Authors: B.C. Lackner and B. Pirscher)

Thousands of years ago, humankind thought that it owed the seasons to Demeter, the goddess of the earth. Her daughter Persephone was retained in the underworld by Hades, the brother of Zeus. So, Demeter induced a terrible continuous snowstorm to extort from Zeus to speak up for Persephone by his brother Hades. Because they could not reach an agreement, which is still often the case in todays' politics, they came up with a compromise: Persephone had to stay with Hades in the underworld for four months each year, while the rest of the time of the year she spent on the earth's surface. So, the seasons were born, by the permanent change of Demeter's mood caused by her joy to see Persephone again and by her sadness to loose the daughter once more for four months.

Nowadays, we do no longer believe in stories like this, but still we search for connections and explanations for atmospheric processes. Both methods used in this work, principal component analysis and factor analysis, proved to be a good tool for such purposes, even though they show several specific advantages as well as disadvantages, which are based on the underlying models. The PCA model is given by $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{f}(+\boldsymbol{e})$, the FA one by $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{f} + \boldsymbol{u}$. The PCA error term $\boldsymbol{e}$ is put in parenthesis because its amount depends on the number of extracted components. The more components are used to explain the data's variance, the smaller is the error term; if all principal components are selected, $\boldsymbol{e}$ is equal to zero. In contrast, the uniqueness term $\boldsymbol{u}$ in FA always plays an important role.

The differences between the models are caused by the fact that PCA does not distinguish between common (affected by more than one variable) and unique (only affected by one variable) variances and both can be found in the principal components. In contrast, the goal of FA is to separate common and unique variances. This allocation of variances to two different matrices implies, as in FA only the factor loadings are used to calculate the amount of total variance in the data (the unique factors $\boldsymbol{u}$ are not included) that the total explained variance with FA is generally less than with PCA.

Furthermore, PCA is not dependent on the number of extracted factors; each factor always contributes the same amount to the explained variances and all factors together explain 100 % of the data's variance. FA results by contrast change with a varying number of factors "k". The total variances increase and the unique variances decrease with a rising number of extracted factors. Hence, also the allocation of common and unique variances to the two matrices depends on k, but the total of common and unique variances always amounts to 100 %.

Concerning atmospheric data sets, this means that FA includes the possibility to spot

those grid points and areas, which are strongly determined by unique variances and therefore are not covered in large-scale patterns. Even though PCA does not include this possibility, preference may be given to this method due to shorter calculation process and the independence on the number of selected factors, which was discussed above.

Having these differences always in mind, both methods are able on the one hand to find new atmospheric patterns and on the other hand to validate known processes in investigated data fields. Therefore, a good basic knowledge of the used data is indispensable. This basic knowledge was acquired in the first part of this work.

Nevertheless, applying PCA and FA to real data sets, the following characteristics have to be kept in mind, in order not to stumble into technical pitfalls:

- **Matrix Dimension:** Both methods require a certain dimension of the data matrix. In the usual case, there are more objects than variables and therefore it is not possible to solve the model exactly. Otherwise, problems may occur during the calculation process. The atmospheric data fields investigated within this work did not comply with this requirement, nevertheless, successful calculations were possible for a part of these data sets. Due to the short available time series of solely 36 months (temperature means), this issue called for special attention. Investigating data sets, the methods' stability is easily exceeded using a too detailed spatial resolution. Because of that, the selected resolution was well considered. To find out the limits of a suitable resolution, each atmospheric field was investigated in two different resolutions, a coarse and a detailed one. The sensibility of the models showed that for the fine-resolved data fields, the stability of both methods was exceeded in two cases, namely in the Eurasian-African (latitude $\times$ height) slice and in the south polar region. For these two data fields, the ratio between the number objects and variables was particularly bad.

- **Atmospheric Data Fields are Different:** To verify the program codes of PCA and FA, a well known example of literature was engaged, were both methods were applied successfully. Nevertheless, problems occurred partly with the investigated atmospheric data fields. The loss of the methods' stability often did not show in the coefficients/loadings, even though the principal components/factor scores were disturbed. Therefore, it is a must to check the correctness of all results. If any problem can be found within a calculation process, it can be helpful to check the matrix dimensions.

- **Arbitrariness of the Factors:** Since the results of PCA and FA are based on the decomposition of matrices in eigenvectors, it has to be kept in mind that the direction of the eigenvectors is arbitrarily defined. In order to avoid misinterpretation, both the coefficients/loadings and the principal components/factor scores have to be considered, because the patterns themselves are composed by matrix multiplication of the respective results. Neglecting this fact, negative values of

coefficients/loadings, for instance, could easily be interpreted as global cooling, even if the rebuilt pattern would show a warming.

- **Similar Coefficients/Loadings Rely on Different Patterns:** The analysis of the Eurasian-African slice showed that the coefficients/loadings of the first extracted factor of the monthly mean centered data were nearly identical to the coefficients/loadings of the third extracted factor of 3-year mean corrected data. The combination of the principal components/factor scores with the coefficients/-loadings yielded different atmospheric patterns, so that again in each of the two resulting matrices not enough information was included to derive correct interpretations. The real origin of the patterns can only be detected by considering both coefficients/loadings and principal components/factor scores.

- **What About Orthogonality?** According to the theory, the coefficients/loadings and principal components/factor scores of every extracted factor have to be orthogonal to each other. This would imply that the first detected pattern cannot be found again in the second factor. Nevertheless, the annual temperature cycle remained at least in the first three principal components/factor scores of 3-year mean centered data sets. So, it seems that in the investigated cases, the methods did not succeed completely to extract the whole signal with one factor. This was presumably caused by the strength of the seasonal signal in temperature data fields.

- **Correlation Matrix Equal to Covariance Matrix?** Even though the correlation matrix is a kind of normalized covariance matrix, one would expect the results to be similar for both matrices. This was not the case for the monthly mean centered Eurasian-African slice, where completely different patterns were detected from the PCA coefficients. Nevertheless, the reconstruction of the data by means of the first two coefficients and principal components showed that these differences nearly vanish in reconstruction.

Only part of the information comprised in the data sets could be detected in the four atmospheric data fields. Surprisingly much information is still hidden in the data set, but unfortunately searching all these patterns would blast the scope of this work. Furthermore, the reasons for the failure of factor analysis techniques in certain selected data sets was not considered in detail. Anyway, the goal was to get familiar with the methods and to find out their weak and strong points.

From our point of view, we succeeded.

Bibliography

T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley, New York, 2nd edition, 1984.

J. K. Angell, L. T. Flynn, M. E. Gelman, D. Hofmann, C. S. Long, A. J. Miller, S. Oltmans, and S. Zhou. Southern Hemisphere Winter Summary 2002. Available from: `www.cpc.ncep.noaa.gov/products/stratosphere/winter_bulletins/` (April 2005), 2002. Issued by NOAA Climate Prediction Center.

J. K. Angell, L. T. Flynn, M. E. Gelman, D. Hofmann, C. S. Long, A. J. Miller, S. Oltmans, and S. Zhou. Northern Hemisphere Winter Summary 2002–2003. Available from: `http://www.cpc.ncep.noaa.gov/products/stratosphere/winter_bulletins/nh_02-03/` (September 2005), 2003a. Issued by NOAA Climate Prediction Center.

J. K. Angell, L. T. Flynn, M. E. Gelman, D. Hofmann, C. S. Long, A. J. Miller, S. Oltmans, and S. Zhou. Southern Hemisphere Winter Summary 2003. Available from: `www.cpc.ncep.noaa.gov/products/stratosphere/winter_bulletins/` (April 2005), 2003b. Issued by NOAA Climate Prediction Center.

J. K. Angell, L. T. Flynn, M. E. Gelman, D. Hofmann, C. S. Long, A. J. Miller, S. Oltmans, and S. Zhou. Northern Hemisphere Winter Summary 2003–2004. Available from: `http://www.cpc.ncep.noaa.gov/products/stratosphere/winter_bulletins/nh_03-04/` (September 2005), 2004a. Issued by NOAA Climate Prediction Center.

J. K. Angell, L. T. Flynn, M. E. Gelman, D. Hofmann, C. S. Long, A. J. Miller, S. Oltmans, and S. Zhou. Southern Hemisphere Winter Summary 2004. Available from: `www.cpc.ncep.noaa.gov/products/stratosphere/winter_bulletins/` (September 2005), 2004b. Issued by NOAA Climate Prediction Center.

J. K. Angell, L. T. Flynn, M. E. Gelman, D. Hofmann, C. S. Long, A. J. Miller, S. Oltmans, and S. Zhou. Northern Hemisphere Winter Summary 2004–2005. Available from: `http://www.cpc.ncep.noaa.gov/products/stratosphere/winter_bulletins/nh_04-05/` (September 2005), 2005. Issued by NOAA Climate Prediction Center.

R. A. Anthes, C. Rocken, and Y.-H. Kuo. *Applications of COSMIC to Meteorology and Climate*, pages 115–156. Springer, Hong Kong, 2001.

M. P. Baldwin, L. J. Gray, T. J. Dunkerton, K. Hamilton, P. H. Haynes, W. J. Randel, J. R. Holton, M. J. Alexander, I. Hirota, T. Horinouchi, D. B. A. Jones, J. S. Kinnersley, C. Marquardt, K. Sato, and M. Takahashi. The Quasi-biennial Oscillation. *Reviews of Geophysics*, 39:179–229, 2001.

J. Bortz. *Statistik für Sozialwissenschaftler*. Springer-Verlag, Berlin, Heidelberg, 3nd edition, 1989.

F. Bouttier and P. Courtier. Data assimilation concepts and methods. *Meteorological Training Course Lecture Series (Printed 9 January 2001)*, March 1999. Available from: `http://twister.caps.ou.edu/OBAN2004/Assim_concepts.pdf`, (January 2005).

F. Bouttier and F. Rabier. The operational implementation of 4D-Var. *ECMWF Newsletter*, (78):2–5, 1997/98.

P. L. Brockett, R. A. Derrig, L. L. Golden, A. Levine, and M. Alpert. Fraud Classification Using Principal Component Analysis of RIDITs. *Journal of Risk & Insurance*, 69:341–371, 2002.

R. Buizza. Chaos and weather prediction. *Meteorological Training Course Lecture Series, ECMWF 2002*, January 2000. Available from: `http://www.gi.alaska.edu/~bhatt/Teaching/ATM693.Climate.JC/climate.papers/Chaos.pdf` (January 2005).

R. Buizza, D. S. Richardson, and T. N. Palmer. The new 80-km High-Resolution ECMWF EPS. *ECMWF Newsletter*, (90):1–8, 2001.

C. Dunn, W. Bertiger, Y. Bar-Sever, S. Desai, D. Haines, B. Kuang, G. Franklin, I. Harris, G. Kruizinga, T. Meehan, S. Nandi, D. Nguyen, T. Rogstand, J. T. Brooks, J. Tien, L. Romans, M. Watkins, S.-C. Wu, and J. Kim. Instruments of GRACE – GPS Augments Gravity Measurements. *GPS World*, 2003.

R. K. Dutta, R. B. Roijers, P. H. A. Mutsaers, J. J. M. de Goeij, and G. J. van der Vusse. Principal component analysis of elements in atherosclerotic human coronary arteries. *Nuclear Instruments and Methods in Physics Research B*, 231:245–250, apr 2005.

G. J. Feeney and D. D. Hester. *Stock market indices: A principal components analysis*, pages 110–138. Risk aversion and portfolio choice. Wiley, New York, 1967.

M. Fischer. Assimilation Techniques (4): 4dVar. *Meteorological Training Course Lecture Series, ECMWF 2002*, April 2001. Available from: `http://www.ecmwf.int/newsevents/training/rcourse_notes/pdf_files/Assim_techniques_4dVar.pdf` (January 2005).

M. Fischer, E. Gerard, A. Ghelli, P. Janssen, G. Kelly, A. P. McNally, M. Miller, A. Simmons, J. Teixeira, and P. Viterbo. The IFS cycle CY21r4 made operational in October 1999. *ECMWF Newsletter*, (87):2–9, 2000.

E. L. Fleming, S. Chandra, J. J. Barnett, and M. Corney. Zonal Mean Temperature, Pressure, Zonal Wind, and Geopotential Height as Functions of Latitude. *Advances in Space Research, COSPAR International Reference Atmosphere: 1986, Part II: Middle Atmosphere Models*, 10(12):11–59, 1990.

U. Foelsche. *Tropospheric water vapor imaging by combination of ground-based and spaceborne GNSS sounding data*. PhD thesis, Institut für Meterologie und Geophysik, Universität Graz, Graz, 1999.

U. Foelsche, G. Kirchengast, and A. K. Steiner. *Global Climate Monitoring based on CHAMP/GPS Radio Occultation Data*, pages 397–407. First CHAMP Mission Results for Gravity, Magnetic and Atmospheric Studies. Springer, Berlin, 2003.

G. Garson. Factor Analysis. Available from: `http://www2.chass.ncsu.edu/garson/pa756/factor.htm`, 2005. North Carolina State University.

A. Gobiet. CCR v2 Quality-Parameter Description. Private Communication, 2004. Institut für Geophysik, Astrophysik und Meteorologie, University of Graz.

A. Gobiet, U. Foelsche, A. K. Steiner, M. Borsche, G. Kirchengast, and J. Wickert. Climatological validation of stratospheric temperatures in ECMWF operational analyses with CHAMP radio occultation data. *Geophysical Research Letters*, 32(L12806), 2005a. doi:10.1029/2005GL022617.

A. Gobiet and G. Kirchengast. *Advancement of GNSS radio occultation retrieval in the upper stratosphere*, pages 137–148. Occultations for Probing Atmosphere and Climate. Springer, 2004.

A. Gobiet, G. Kirchengast, J. Wickert, C. Retscher, D. Y. Wang, and H. A. *Evaluation of Stratospheric Radio Occultation Retrieval Using Data from CHAMP, MIPAS, GOMOS, and ECMWF Analysis Fields*. Springer, Berlin-Heidelberg-New York, 2005b.

M. E. Gorbunov and L. Kornblueh. Analysis and validation of Challenging Minisatellite Payload (CHAMP) radio occultation data. *Journal of Geophysical Research*, (108), 2003. D18, 4584, doi:10.1029/2002JD003175.

P. Høeg, G. B. Larsen, H. Benzon, J. Grove-Rasmussen, S. Syndergaard, M. D. Mortensen, J. Christensen, and K. Schultz. GPS Atmosphere Profiling Methods and Error Assessments. Technical Report 98-7, Danish Meteorological Institute, Atmosphere Ionosphere Remote Sensing Division, Lyngbyvej 100, DK-2100 Copenhagen, Denmark, 1998.

H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, (24):417–441, 1933.

X. Hu. Factor Analysis, Principal Component Analysis, Non-central Distributions, Spectral Analysis, Analysis of Missing Values, Multiple Imputation, Monte Carlo Markov Chain, EM Algorithm, Hot Deck Procedure, Projection-Pursuit Algorithm, Median

Absolute Deviation, Robust Statistics. Available from: `http://www.math.ucla.edu/~xhu/pca-fa.pdf` (July 2005), n.y.

P. Hupfer and W. Kuttler. *Witterung und Klima*. B. G. Teubner, Stuttgart, Leipzig, 1998.

IPCC. *Climate Change 2001: The Scientific Basis*. Cambridge University Press, Cambridge, 2001.

I. T. Jolliffe. *Principal Component Analysis*. Springer series in statistics, New York, Berlin, Heidelberg, 2nd edition, 2002.

H. F. Kaiser. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, (23):187–200, 1958.

E. Kalnay. The NCEP/NCAR 50 – Year Reanalysis, 1999. Submitted to the Bulletin of the American Meteorological Society.

S. M. Kanbur and H. Mariani. Principal-component analysis of RR Lyrae light curves. *Monthly Notices of the Royal Astronomical Society*, 355:1361–1368, dec 2004.

A. Kaplunovsky. Why using factor analysis? (dedicated to the centenary of factor analysis). Paper contribution to the conference on "Factor Analysis at 100"; Chapel Hill; NC, 2004.

G. Kirchengast. *Occultation for Probing Atmosphere and Climate: Setting the Scene*, pages 1–8. Occultation for Probing Atmosphere and Climate. Springer, 2004.

G. Kirchengast, J. Fritzer, and J. Ramsauer. End-to-end GNSS Occultation Performance Simulator Version 4 (EGOPS4) Software User Manual (Overview and Reference Manual). Technical report, Inst. for Geophysics, Astrophysics and Meteorology, University of Graz, Austria, 2002. IGAM/UP Technical Report for ESA/ESTEC No. 3/2002.

G. Kirchengast, J. Hafner, and W. Poetzi. The CIRA86aQ_UoG model: An extension of the CIRA-86 monthly tables including humidity tables and a Fortran95 global moist air climatology model. Technical report, Inst. for Meteorol. and Geophys., Univ. of Graz, Austria, 1999. Techn. Rep. for ESA/ESTEC No. 8/1999.

E. R. Kursinski, G. A. Haji, S. S. Leroy, and B. Herman. *The GPS Radio Occultation Technique*, pages 53–114. Springer, Hong Kong, 2001.

E. R. Kursinski, G. A. Hajj, J. T. Schofield, R. P. Linfield, and K. R. Hardy. Observing Earth's atmosphere with radio occultation measurements using the Global Positioning System. *Journal of Geophysical Research*, 102:23 429–23 465, 1997.

K. Labitzke and B. Naujokat. The lower arctic stratosphere in winter since 1952 Update. Available from: `strat-www.met.fu-berlin.de/products/update-np-temp.html` (April 2005), 2004. Stratospheric Research Group FU Berlin.

C. B. Lang and N. Pucker. *Mathematische Methoden in der Physik*. Spektrum Akademischer Verlag, Heidelberg, Berlin, 1998.

D. Lawley and A. Maxwell. *Factor Analysis as a Statistical Method*. Butterworth & Co, London, 2nd edition, 1971.

G. H. Liljequist and K. Cehak. *Allgemeine Meteorologie*. Friedr. Vieweg & Sohn, Braunschweig/Wiesbaden, 3rd edition, 1984.

R. S. Lindzen. *Dynamics in Atmospheric Physics. Lecture Notes for an Introductory Graduate-Level Course*. Cambridge University Press, Cambridge, 1990.

M. Loiselet, N. Stricker, Y. Menard, and J.-P. Luntama. GRAS – Metop's GPS-Based Atmospheric Sounder. *ESA Bulletin*, 102:38–44, May 2000.

G. L. Manney, K. Krüger, J. L. Sabutis, S. A. Sena, and S. Pawson. The remarkable 2003–2004 winter and other recent warm winters in the Arctic stratosphere since the late 1990s. *Journal of Geophysical Research*, 110:D04107, 2005.

K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. Academic Press, Harcourt Brace & Co., London, San Diego, New York, Boston, Sydney, Tokyo, 1979.

T. K. Meehan and G. A. Hajj. Preliminary Results From BlackJack GPS Receiver On SAC-C. *AGU Spring Meeting Abstracts*, pages 31–+, May 2001.

A. G. Mendez. Monitoring the quality of observations. Available from: `http://www.ecmwf.int/newsevents/training/meteorological_presentations/pdf/OP/Quality.pdf` (January 2005), 2004. Met-OP Training Course, Use and Interpretation of ECMWF Products.

V. Natraj, X. Jiang, R. l. Shia, X. Huang, J. S. Margolis, and Y. L. Yung. Application of principal component analysis to high spectral resolution radiative transfer: A case study of the O_2 A band. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 92:539–556, nov 2005.

F. Nielsen. Variational Approach to Factor Analysis and Related Models. Intelligent Signal Processing group at the Institute of Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, 2004. Master Thesis.

K. Pearson. n lines and planes of closest fit to systems of points in space. *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, (2):559–572, 1901.

A. Persson. User Guide to ECMWF Forecast Products. Available from: `http://www.ecmwf.int/products/forecasts/user_guide.pdf` (December 2004), 2001. Designed, edited and printed by ECMWF.

J. M. Picone, D. P. Drob, R. R. Meier, and A. E. Hedin. NRLMSISE-00: A New Empirical Model of the Atmosphere. Available from: `www.nrl.navy.mil/content.php?P=03REVIEW105` (November 2004), 2004.

J. M. Picone, A. E. Hedin, D. P. Drob, and A. C. Aikin. NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues. *Journal of Geophysical Research*, 107(A12):1–16, December 2002.

J. Pohlmann. Factor Analysis Glossary. Available from: `http://www.siu.edu/~epse1/pohlmann/factglossary.htm`, 2005.

W. Randel, M. Chanin, and C. Michaut. SPARC Intercomparison of Middle Atmosphere Climatologies. Technical Report SPARC Report No. 3, SPARC Reference Climatology Group, 2002.

C. Reigber. GFZ Potsdam, Department 1: The CHAMP Mission. Available from: `http://www.gfz-potsdam.de/pb1/op/champ/` (March 2005), 2005.

C. Reigber, H. Lühr, and P. Schwintzer. Announcement of Opportunity for CHAMP. Technical report, GFZ Potsdam, 2001. Available from: `http://op.gfz-potsdam.de/champ` (December 2004).

R. Rew, G. Davis, S. Emmerson, and H. Davies. NetCDF User's Guide. Data Model, Programming Interfaces, and Format for Self-Describing, Portable Data, NetCDF Version 3.6.0. Available from: `www.uni-data.ucar.edu/packages/netcdf/`, 2004. University Corporation for Atmospheric Research, Boulder, Colorado.

R. A. Reyment and K. G. Jöreskog. *Applied Factor Analysis in the Natural Science*. Cambridge University Press, Cambridge, 2nd edition, 1993.

E. Rigdon. Important Continuous Statistical Distributions. Available from: `http://www.gsu.edu/~mkteer/continua.html` (13.06.2005), 1996.

E. Rigdon. Not Positive Definite Matrices–Causes and Cures. Available from: `http://www.gsu.edu/~mkteer/npdmatri.html` (13.06.2005), 1997.

C. Rocken, R. Anthes, M. Exner, D. Hunt, R. Sokolovskiy, S. Ware, , M. Gorbunov, W. Schreiner, D. Feng, B. Herman, Y.-H. Kuo, and X. Zou. Analysis and validation of GPS/MET data in the neutral atmosphere. *J. Geophys. Res.*, 102:29849–29866, 1997.

M. L. Salby. *Fundamentals of Atmospheric Physics*. Academic Press, San Diego, California, 1st edition, 1996.

K. P. Scherer and M. Avellaneda. All For One... One For All? A Principal Component Analysis of the Latin American Brady Bond Debt from 1994 to 2000. EFMA 2001 Lugano, 2001. NYU Courant Inst of Mathmatical Sciences Working Paper.

T. Schmidt, S. Heise, J. Wickert, G. Beyerle, and C. Reigber. GPS radio occultation with CHAMP and SAC-C: global monitoring of the thermal tropopause parameters. *Atmospheric Chemistry and Physics*, 5:1 473–1 488, 2005.

K. Schoellhammer, C. Marquardt, and K. Labitzke. *Comparison of Three Different Meteorological Datasets (ECMWF, Met Office and NCEP)*, pages 528–535. First CHAMP Mission Results for Gravity, Magnetic and Atmospheric Studies. Springer, 2003.

D. J. Seidel, R. J. Ross, J. A. Angell, and G. C. Reid. Climatological characteristics of the tropical tropopause as revealed by radiosondes. *Journal of Geophysical Research*, 106(D8):7 857–7 878, 2001.

D. J. Shea, S. J. Worley, U. R. Stern, and T. J. Hoar. An Introduction to Atmospheric and Oceanographic Data. Technical Report NCAR/TN-404+IA, National Center for Atmospheric Research, Boulder, Colorado, Climate and Global Dynamics Division, 1994.

J. Stackpole. Guide to WMO Binary Code Form GRIB 1. Technical report.

A. K. Steiner, A. Gobiet, U. Foelsche, and G. Kirchengast. Radio Occultation Data Processing Advancements for Optimizing Climate Utility. Technical report, IGAM/UniGraz, 2004. Technical Report for ASA No. 3/2004.

A. K. Steiner, G. Kirchengast, U. Foelsche, L. Kornblueh, E. Manzini, and L. Bengtsson. GNSS Occultation Sounding for Climate Monitoring. *Phys. Chem. Earth (a)*, 26:113–124, 2001.

A. K. Steiner, G. Kirchengast, and H. P. Ladreiter. Inversion, error analysis, and validation of GPS/MET occultation data. *Annales Geophysicae*, 17:122–138, 1999.

A. K. Steiner, A. Löscher, and G. Kirchengast. Error Characteristics of Refractivity Profiles Retrieved from CHAMP Radio Occultation Data.

G. Stoehr. Freunde alter Wetterinstrumente: Themoetrie – Geschichte. Available from: `http://www.freunde-alter-wetterinstrumente.de/22theges.htm` (August 2005), 2004.

L. L. Thurstone. Vectors in Mind. *The University of Chicago Press*, 1935.

P. Thy and K. H. Esbensen. Seafloor spreading and the ophiolitic sequences of the Troodos complex: A principal component analysis of lava and dike compositions. *Journal of Geophysical Research*, 98(B7):11799–11805, may 1993.

W. Torge. *Geodäsie*. Walter de Gruyter, Berlin, 2nd edition, 2003.

L. Tucker and R. MacCallum. Exploratory Factor Analysis. Available from: `http://www.unc.edu/~rcm/book/factor.pdf`, 1997.

A. Untch and A. Simmons. Increased stratospheric resolution in the ECMWF forecasting system. *ECMWF Newsletter*, (82):2–8, 1998/99.

G. van der Grijn. Satellite Observations. Available from: `http://www.ecmwf.int/newsevents/training/meteorological_presentations/pdf/OP/SatObs.pdf` (January 2005), 2004. Met-OP Training Course, Use and Interpretation of ECMWF Products.

Vintersol. The Northern Hemisphere Stratosphere in the 2002/03 Winter. Preliminary Results from the first phase of VINTERSOL (Validation of International Satellites and Study of Ozone Loss). University of Cambridge, 2003. European Ozone Research Coordinating Unit.

H. von Storch and F. W. Zwiers. *Statistical Analysis in Climate Research*. Cambridge University Press, Cambridge, UK, 2003.

V. V. Vorob'ev and T. G. Krasil'nikova. Estimation of the Accuracy of the Atmospheric Refractive Index Recovery from Doppler Shift Measurements at Frequencies Used in the NAVSTAR System. *Physics of the Atmosphere and Ocean*, 29:602–609, 1994.

R. Ware, C. Rocken, F. Solheim, M. Exner, R. Schreiner, W. Anthes, D. Feng, B. Herman, M. Gorbunov, S. Sokolovskiy, K. Hardy, Y.-H. Kuo, X. Zou, K. Trenberth, T. Meehan, W. Melbourne, and S. Businger. GPS soundings of the atmosphere from low earth orbit: Preliminary results. *Bull. Amer. Meteor. Soc.*, 77:19–40, 1996.

E. Weber. *Einführung in die Faktorenanalyse*. VEB Gustav Fischer, Jena, 1st edition, 1974.

J. Wickert. *Das CHAMP-Radiookkultationsexperiment: Algorithmen, Prozessierungssystem und erste Ergebnisse*. PhD thesis, Institut für Geophysik, Astrophysik und Meteorologie, Karl-Franzens-Universität Graz, Graz, 2002.

J. Wickert. GNSS Radio occultation: Introduction and selected results. Available from: `http://143.50.39.126/OPAC2/pdf_presentation/opac2_tutorial_wickert_presentation.pdf` (March 2005), 2004. OPAC-2, September 13-17, 2004, Graz, Tutorial talk.

J. Wickert, C. Reigber, G. Beyerle, R. König, T. Marquardt, C. Schmidt, , L. Grunwaldt, R. Galas, T. K. Meehan, W. G. Melbourne, and K. Hocke. Atmosphere sounding by GPS radio occultation: First results from CHAMP. *Geophysical Research Letters*, 28: 3263–3266, 2001a.

J. Wickert, T. Schmidt, C. Marquardt, C. Reigber, K.-H. Neumayer, G. Beyerle, R. Galas, and L. Grunwaldt. GPS radio occultation with CHAMP: First results and status of the experiment. In *IAG 2001 Scientific Assembly*, pages 1–6, 2001b.

Abbreviations

4D-Var	Four Dimensional VARiational Analysis
BC	Barnett and Corney
BLUE	Best Linear Unbiased Estimator
CDC	U.S. Climate Diagnostic Center
CDF	Common Data Format
CHAMP	CHAllenging Minisatellite Payload
CIRA	COSPAR International Reference Atmosphere
CIRA86aQ_UoG	CIRA 1986 – and humidity (University of Graz CIRA model)
COSPAR	COMmittee on Space Research
ECMWF	European Centre for Medium-Range Weather Forecasts
EOF	Empirical Orthogonal Function
EV	Eigenvalue
FA	Factor Analysis
GFZ	GeoForschungsZentrum (Potsdam)
GLONASS	GLObal NAvigation Satellite System (Russia)
GNSS	Global Navigation Satellite System
GPS	Global Positioning System (US)
GSFC	Goddard Space Flight Center
HDF	Hierarchical Data Format
IDL	Interactive Data Language
IGAM	Institutsbereich Geophysik, Astrophysik und Meteorologie (Uni-Graz)
LEO	Low Earth Orbiting (satellite)
ML-FA	Maximum Likelihood Factor Analysis
MSIS	Mass Spectrometer and Incoherent Scatter Radar (model)
NASA	U.S. National Space and Aeronautics Administration
NCAR	U.S. National Centers for Atmospheric Research (former NMC)
NCEP	U.S. National Centers for Environmental Prediction
netCDF	Network Common Data Format
NMC	U.S. National Meteorological Center (now NCEP)
NOAA	U.S. National Oceanic and Atmospheric Administration
NRL	Naval Research Laboratory

PCA	Principal Component Analysis
PFA	Principal Factor Analysis
RO	Radio Occultation
SNR	Signal to Noise Ratio
UT	Universal Time
VAR	VARiational assimilation (usually 1-, 2-, 3- or 4-D)
WMO	World Meteorological Organization

List of Tables

List of Figures

A Linear Algebra

A.1 Definitions

A.1.1 Data Matrix

The data matrix $\boldsymbol{X}$ contains any information given from measurements. $\boldsymbol{X}_{(n\times p)}$ consists of n rows and p columns, whereas each row represents the object and each column lists one variable.

A.1.2 Deviation Scores and Standard Scores

Deviation scores are obtained by subtracting the mean of one variable $\bar{x}_j$ from the raw scores of the variable,

$$y_{ij} = x_{ij} - \bar{x}_j \qquad i = 1, \ldots, n$$

with

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}, \qquad j = 1, \ldots, p.$$

The mean of deviation scores is always zero, the standard deviation is still the same as from the raw scores.

Standard scores (z-scores) are obtained by dividing deviation scores by the standard deviation s_j

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \qquad i = 1, \ldots, n$$

with

$$s_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}, \qquad j = 1, \ldots, p. \tag{A.1}$$

The mean of standard scores is always zero and the standard deviation (and variance) is always one.

A.1.3 Major and Minor Product Moments

The major product moment $\boldsymbol{C}_{(n\times n)}$ is defined as the product of the data matrix $\boldsymbol{X}_{(n\times p)}$ postmultiplied by its transpose (Reyment and Jöreskog 1993):

$$\boldsymbol{C} = \boldsymbol{X}\boldsymbol{X}'.$$

The minor product moment $\boldsymbol{E}_{(p\times p)}$ is defined as the premultiplication of the matrix $\boldsymbol{X}_{(n\times p)}$ by its transpose (Reyment and Jöreskog 1993):

$$\boldsymbol{E} = \boldsymbol{X}'\boldsymbol{X}.$$

Both major product moment and minor product moment are square symmetrical matrices.

A.1.4 Orthogonal, Orthonormal Matrices

A square matrix $\boldsymbol{E}_{(p\times p)}$ is said to be orthogonal if its columns, considered as vectors, are mutually perpendicular and

$$\boldsymbol{E}'\boldsymbol{E} = \boldsymbol{D},$$

where $\boldsymbol{D}$ is a diagonal matrix. $\boldsymbol{E}_{(p\times p)}$ is said to be orthonormal if its columns, considered as vectors, are mutually perpendicular and have unit length.

$$\boldsymbol{E} \text{ is orthonormal} \Leftrightarrow \boldsymbol{E}'\boldsymbol{E} = \boldsymbol{E}\boldsymbol{E}' = \boldsymbol{I}_p \Leftrightarrow \boldsymbol{E}^{-1} = \boldsymbol{E}'.$$

$\boldsymbol{I}_p$ is the $(p \times p)$-identity matrix.

A.1.5 Singular/Nonsingular

A square matrix $\boldsymbol{E}_{(p\times p)}$ is said to be nonsingular, if $\boldsymbol{E}\boldsymbol{x} = 0$ implies that $\boldsymbol{x} = 0$. A square matrix $\boldsymbol{E}$ is said to be singular, if there exists an $\boldsymbol{x} \neq 0$ such that $\boldsymbol{E}\boldsymbol{x} = 0$.

Equivalently, a square matrix $\boldsymbol{E}$ is said to be nonsingular if its rank (cf., Section A.1.9) is equal to its number of rows (or columns).

A.1.6 Determinant

Determinants are scalar mathematical objects that are derived from mathematical operations on matrices. Determinants are defined only for square matrices.

If the determinant of a matrix is zero, the matrix is a singular matrix, nonsingular matrices have nonzero determinants. See, e.g., Lang and Pucker (1998) for details.

A.1.7 Minor

Given a square matrix $\boldsymbol{E}_{(p\times p)}$, the minor M_{ij} is defined as the determinant of the matrix formed by deleting the i^{th} row and j^{th} column of $\boldsymbol{E}$. There will therefore be one minor corresponding to each element of $\boldsymbol{E}$.

A.1.8 Positive Definite

A square, real matrix $\boldsymbol{E}_{(p\times p)}$ is said to be positive definite, if

$$\boldsymbol{x}'\boldsymbol{E}\boldsymbol{x} > 0 \qquad \forall \boldsymbol{x} \neq 0.$$

Equivalently, a square symmetric matrix $\boldsymbol{E}$ is said to be positive definite if all minors associated with the elements of the principal diagonal are greater than zero. If $\boldsymbol{E}$ is positive definite, all diagonal elements and all eigenvalues are greater than zero. The determinant of a positive definite matrix is always positive, so it is always nonsingular. $\boldsymbol{E}$ is invertible and its inverse is also positive definite.

If $\boldsymbol{x}'\boldsymbol{E}\boldsymbol{x} \geq 0$, $\forall \boldsymbol{x} \neq 0$, $\boldsymbol{E}$ is said to be positive semi definite. In that case some eigenvalues and the determinant are equal to zero.

A.1.9 Rank

The row rank of a matrix is the maximum number of linearly independent rows considering the rows as vectors. The columns rank of a matrix is the rank of its set of columns, considered as vectors. The row rank and the column rank are equal and are called the rank of the matrix.

A.1.10 Eigenvalues/Eigenvectors

If $\boldsymbol{E}$ is a real, symmetric $(p \times p)$-matrix and $\boldsymbol{I}_p$ is the $(p \times p)$-identity matrix, then the scalars $\lambda_1, \lambda_2, \ldots, \lambda_p$ satisfying the polynomial equation

$$\det(\boldsymbol{E} - \lambda \boldsymbol{I}_p) = 0$$

are the eigenvalues of $\boldsymbol{E}$. $\lambda_1, \lambda_2, \ldots, \lambda_p$ are also referred to characteristic roots; $\det(\boldsymbol{E} - \lambda \boldsymbol{I}_p) = 0$ is known as the characteristic equation.

If $\boldsymbol{x}$ is a non-zero vector such $\boldsymbol{E}\boldsymbol{x} = \lambda \boldsymbol{x}$, then $\boldsymbol{x}$ is said to be an eigenvector (characteristic vector) of the matrix $\boldsymbol{E}$ associated with λ.

Properties of Eigenvalues and Eigenvectors

1. The sum of the eigenvalues of a matrix $\boldsymbol{E}$ is equal to the sum of the elements in the principal diagonal of the matrix $\boldsymbol{E}$.

2. The product of the eigenvalues equals to the determinant of the matrix. If one or more eigenvalues are zero, the determinant of the matrix will be zero and it is singular.

3. The number of nonzero eigenvalues equals to the rank of the matrix.

4. Eigenvectors are always normalized, they are of unit length.

5. All eigenvectors associated with different eigenvalues are orthogonal to each other.

A.1.11 Sample Covariance Matrix

The sample covariance matrix $\boldsymbol{S}$ is defined as the minor product moment of the data matrix expressed in deviate form ($\boldsymbol{Y'Y}$) divided by $n-1$, n is the number of objects

$$\boldsymbol{S} = \frac{\boldsymbol{Y'Y}}{n-1}.$$

The elements of the principal diagonal are the variances of the variables, the other elements are the covariances of the variables.

$$\boldsymbol{S} = \begin{pmatrix} \mathrm{Var}(x_1), & \mathrm{Cov}(x_1, x_2) & \dots & \mathrm{Cov}(x_1, x_p) \\ \mathrm{Cov}(x_2, x_1) & \mathrm{Var}(x_2) & \dots & \mathrm{Cov}(x_2, x_p) \\ \vdots & \vdots & & \vdots \\ \mathrm{Cov}(x_p, x_1) & \mathrm{Cov}(x_p, x_1) & \dots & \mathrm{Var}(x_p) \end{pmatrix} \tag{A.2}$$

The covariance of two data sets can be defined as there tendency to vary together. The covariance value will be larger than zero if two variables tend to increase together, below zero if they tend to decrease together, and zero if they are independent. Due to the symmetry property of covariances, it is necessarily a symmetric matrix. Covariance matrices are positive definite or positive semi definite.

A.1.12 Sample Correlation Matrix

The sample correlation matrix $\boldsymbol{R}$ is defined as the minor product moment of the standardized data matrix ($\boldsymbol{Z'Z}$) divided by $n-1$, n is the number of objects

$$\boldsymbol{R} = \frac{\boldsymbol{Z'Z}}{n-1}.$$

The correlation matrix is a symmetric matrix (the correlation of variable x with variable y is always equal to the correlation of variable y with variable x) and its diagonal elements are one because these are the correlations between each variable and itself.

$$\boldsymbol{R} = \begin{pmatrix} 1, & \mathrm{r}(x_1, x_2) & \dots & \mathrm{r}(x_1, x_p) \\ \mathrm{r}(x_2, x_1) & 1 & \dots & \mathrm{r}(x_2, x_p) \\ \vdots & \vdots & & \vdots \\ \mathrm{r}(x_p, x_1) & \mathrm{r}(x_p, x_1) & \dots & 1 \end{pmatrix} \tag{A.3}$$

Each element (correlation coefficient) in the correlation matrix ranges from -1.0 to $+1.0$, the closer to $+1$ or -1, the more closely the two variables are related. If the correlation is positive, it means that as one variable gets larger the other gets larger and if it is negative it means that as one gets larger, the other gets smaller.

If the correlation coefficient is close to zero, the variables do not have any relationship.

A.1.13 Covariance Matrix in Comparison With the Correlation Matrix

The difference between these two matrices is how the data are scaled before the matrix multiplication is executed. In case of the covariance matrix the mean of each variable is subtracted before multiplication and the correlation matrix results from standardized variables (mean subtracted, then divided by standard deviation).

B Differential Calculus

B.1 Lagrange Multiplication

Lagrange multipliers can be used to find the maximum of a multivariate function subject to a constraint.

If $f(x_1, x_2, \ldots, x_p)$ is the function being maximized and $g(x_1, x_2, \ldots, x_p) = c$ being the constraint, there is a λ (Lagrange multiplier) conforming to the equation

$$\frac{\partial f}{\partial x_i} - \lambda \frac{\partial g}{\partial x_i} = 0, \qquad i = 0, \ldots, p \tag{B.1}$$

and in a stationary point a the partial derivatives are zero: $\frac{\partial f}{\partial \boldsymbol{x}}(a) = 0, \boldsymbol{x} = (x_1, \ldots, x_p)$. The function $L(\boldsymbol{x}, \lambda)$ is defined with

$$L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) - \lambda[g(\boldsymbol{x}) - c]. \tag{B.2}$$

Because of the constraint, $g(\boldsymbol{x}) - c = 0$, equation (B.1) can be written as

$$\frac{\partial L}{\partial \boldsymbol{x}} = 0. \tag{B.3}$$

Abstract:
This report deals with CHAMP radio occultation climatologies generated at the Wegener Center, University of Graz. In the first part of the report, the climatologies were compared to analyses and reanalyses (from the European and U.S. meteorological services) as well as to two different climatological models called NRLMSISE-00 and CIRA86aQ_UoG.
One focus was the investigation of the influence of the local time at which radio occultation measurements were taken, revealing that monthly climatologies do not show appreciable problems and that the influence of local time is generally negligible.
The second part of the report deals with the analyses of CHAMP radio occultation data by the use of principal component analysis and factor analysis (the latter being implemented with four different calculation procedures).
First, the methods were examined theoretically, afterwards the techniques were compared to each other, and finally they were applied to atmospheric fields in search of atmospheric patterns.
Depending on the location of the data as well as on the applied mean correction the origins of temperature fluctuations were identified as the seasonal cycle, the QBO (Quasi Biennial Oscillation), SSWs (Sudden Stratospheric Warmings), and the polar vortex.

Zum Inhalt:
Diese Arbeit beschäftigt sich mit CHAMP Radiookkultations-Klimatologien, welche am Wegener Zentrum der Karl-Franzens-Universität Graz erstellt wurden. Im ersten Teil der Arbeit wurden die Daten mit Analysen bzw. Reanalysen (vom europäischen und U.S.-amerikanischen Wetterdienst) sowie mit zwei verschiedenen Klimatologie-Modellen (genannt NRLMSISE-00 und CIRA86aQ_UoG) verglichen.
Ein Schwerpunkt lag auf der Untersuchung des Einflusses der Lokalzeit, zu welcher die Radiookkultationsmessungen stattgefunden haben. Es stellte sich heraus, dass der Einfluss der Lokalzeit im Allgemeinen vernachlässigbar ist.
Im zweiten Teil der Arbeit wurden die CHAMP Radiookkultationsdaten mit Hilfe der Hauptkomponentenanalyse und der Faktorenanalyse (letztere durch vier unterschiedliche Verfahren implementiert) untersucht.
Nach einer theoretischen Betrachtung der Methoden wurden diese miteinander verglichen und dann für die Suche nach Mustern in atmosphärischen Feldern angewandt.
Je nach der geographischen Lage der Datenfelder sowie der angebrachten Mittelwertkorrektur konnten als Ursachen für Temperaturschwankungen der Jahresgang, die QBO (Quasi Biennial Oscillation), SSWs (Sudden Stratospheric Warmings) und der polare Vortex identifiziert werden.

Wegener Center for Climate and Global Change
University of Graz
Leechgasse 25
A-8010 Graz, Austria
www.wegcenter.at
ISBN 3-9502126-4-7