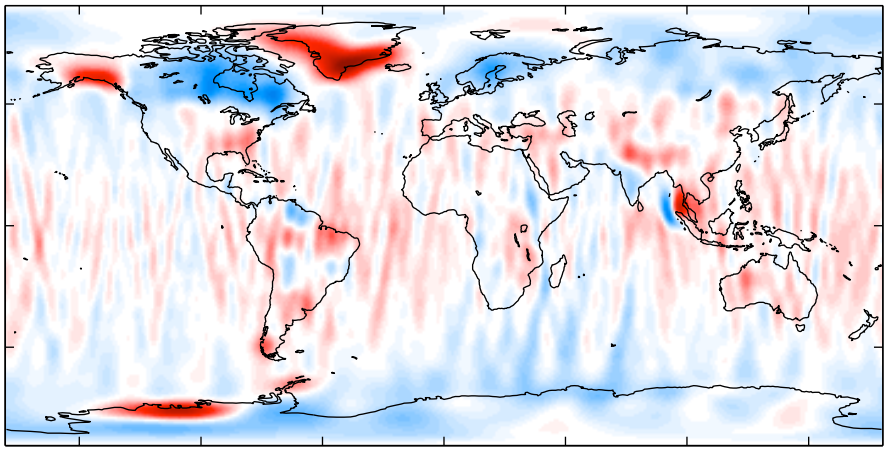
Universität Stuttgart

Geodätisches Institut

Empirical Orthogonal Function Analysis of GRACE Gravity Data



Diplomarbeit im Studiengang

Geodäsie und Geoinformatik

an der Universität Stuttgart

Katrin Bentel

Juni 2009

Betreuer: Prof. Dr.-Ing. Nico Sneeuw

Universität Stuttgart

Erklärung

Hiermit erkläre ich, dass ich diese von mir eingereichte Diplomarbeit zum Thema

Empirical Orthogonal Function Analysis
of GRACE Gravity Data

selbstständig verfasst, keine anderen als die angegebenen Hilfsmittel benutzt, sowie wörtliche und sinngemäße Zitate als solche gekennzeichnet habe.

Stuttgart, den 07. September 2009

Katrin Bentel

Abstract

The Gravity Recovery and Climate Experiment (GRACE) twin-satellite mission has been providing measurements of the time-varying gravity field of the Earth for almost seven years now. Gravity changes on Earth are due to mass changes and play an important role in Earth sciences. Monthly maps of mass changes are derived from the satellite measurements and need to be interpreted. The major difficulty in analyzing GRACE data are North-South stripes in the estimated gravity fields, caused by the fact that the GRACE satellites are flying in a near-polar orbit, one following the other. A microwave ranging instrument is measuring the distance between the two spacecraft, which is about 220 km. Due to these longitudinal stripes, major errors, analyzing the GRACE gravity fields is demanding.

The technique of empirical orthogonal function (EOF) analysis is investigated in this thesis, and it is demonstrated the performance of EOF analysis for separating signal from noise and errors, and for identifying different sources of gravity changes in a real GRACE data set. EOF analysis is explained from a theoretical point of view and is applied to the GRACE data. Basically, the EOF method gives a transformation of the data into a new coordinate frame in the data space, where the axis are chosen according to the data variances. The core of the method is a singular value decomposition of the data matrix. The components obtained from this decomposition need to be interpreted, and signal has to be separated from noise. Additionally, EOF analysis can be used as a filtering tool.

In the detailed data analysis, benefits and shortcomings of the EOF method are studied and described with respect to GRACE data. Global maps of mass changes as well as different smaller regions are analyzed, and global and regional results are compared.

Keywords: empirical orthogonal function, principal component time series, singular value decomposition, GRACE gravity changes, separation of sources

Kurzfassung

Die "Gravity Recovery and Climate Experiment" (GRACE) Mission liefert seit fast sieben Jahren Messungen der zeitlichen Änderungen des Erdschwerefeldes. Diese werden durch Massenverschiebungen auf der Erde hervorgerufen und spielen daher in den Umweltwissenschaften eine bedeutende Rolle. Bei der Auswertung monatlicher Schwerefeldänderungen aus GRACE Messungen bereiten Nord-Süd-Streifen in den Feldern die größten Schwierigkeiten. Diese Streifen entstehen durch den fast polaren Orbit, auf dem die beiden GRACE Satelliten, mit einem Abstand von ungefähr 220 km, einander folgen. Ein Mikrowellenmesssystem misst hochgenau die Abstandsänderungen zwischen den beiden Satelliten. Auf Grund des starken Rauschens, welches die Streifen verursachen, ist die Auswertung der monatlichen Schwerefelder schwierig.

In dieser Diplomarbeit wird die Methode der "Empirical Orthogonal Function (EOF) Analysis" hinsichtlich ihrer Eignung für eine Auswertung echter GRACE Felder untersucht. Anhand dieses Datensatzes wird das Potential der EOF Methode, um Signale von Rauschen zu trennen und einzelne Quellen der Massenänderungen zu identifizieren, demonstriert. Zunächst wird die EOF Analyse von einem theoretischen Standpunkt aus erklärt und dann auf den Datensatz angewendet. Grundsätzlich wird dabei in ein neues Koordinatensystem transformiert, dessen Achsen an den Varianzen der Daten ausgerichtet sind. Der Kern dieser Methode ist eine Singulärwertzerlegung der Datenmatrix. Die Komponenten, die diese Zerlegung leifert, müssen dann hinsichtlich Signal und Rauschen identifiziert werden. Die EOF Analyse mit anschlie§ender Synthese der Felder kann auch als Filter benutzt werden. In einer ausführlichen Untersuchung der GRACE Daten werden Vor- und Nachteile der EOF Zerlegung aufgezeigt und erklärt. Dabei werden sowohl globale Schwerefelder aus GRACE, als auch verscheidene regional Beispiele untersucht. Ergebnisse aus globalen und regionalen Untersuchen werden dabei verglichen.

Schlüsselwörter: Empirische orthogonale Funktion, Hauptkomponentenzeitreihe, Singulärwertzerlegung, GRACE Massenänderungen, Separieren der Quellen

Contents

Chapter 1

Introduction

Models of the Earth's gravity field are widely used in science and engineering. For a long time, there have been terrestrial gravity measurement techniques, which are all very time-consuming and regionally limited. Even the spatial and temporal resolution of airborne gravity measurements is still limited. But with the rise of the satellite-age, gravity measurements covering the whole globe in regular time steps became possible.

There are already three dedicated gravity satellite missions, namely CHAMP, GRACE and GOCE. Of those, GRACE is specifically designed to detect temporal variations of the gravity field. Since its start in 2002, the mission delivers monthly global maps of gravity changes. These changes are related to changes in mass, so the measurements reveal information about mass transport in the Earth system. Many are environmental phenomena directly linked to changes in mass. Thus, identifying different sources of mass changes and observing them over the whole globe is of high interest and relevance for current research.

However, the satellite instruments only measure the integrated gravity effect of all mass changes. And since the satellites fly and measure at an altitude of about 500 km above the Earth's surface, in the case of the GRACE mission, the gravity signal that can be detected by the satellites is attenuated and contaminated with noise. In addition, the measurements are contaminated by instrumental errors and, most of all, errors due to the measurement configuration. The mass changes are determined from highly precise distance measurements between the two GRACE satellites that are aligned collinear along the same orbit. A change in the distance between the two satellites is caused by a mass change on Earth that forces one of the two satellites to speed up. However, the alignment of the two satellites is not optimal for determining global grids of gravity changes. Due to the orbit constellation of

the two satellites all measurements are aligned along-track. The near-circular GRACE orbit with 89.5° inclination has North-South oriented ground track patterns. So sampling in the measurements is very good in North-South direction, but limited in East-West direction. When calculating the spherical harmonic coefficients representing global gravity changes, the differences in the sampling along latitudes and longitudes leads to North-South stripes as the major effect in the maps of gravity changes. These stripes are also called correlated errors, and, unfortunately, bury most of the actual gravity signal.

A major interest when analyzing the maps of gravity changes lies in the contribution of different geophysical sources, separating the individual sources and, most important, separating them from correlated errors. The objective of this thesis is to test the method of empirical orthogonal function (EOF) analysis for its performance in analyzing GRACE gravity data. It is investigated the question if EOF analysis is capable of separating physical effects in the data from noise, and separating the individual effects clearly enough so that they can be interpreted. Therefore, global and regional GRACE maps of gravity changes are used. Concerning the objective to find different sources of mass change, rationally, one has to say that a separation is not possible since only the sum is detected. How shall someone find the components that were contributing to a sum if only the summmation value is given? The GRACE data sets consist of global monthly gravity solutions for almost seven years now. And these time series are a starting point for techniques trying to find different sources of mass change in the gravity signal. The basic idea behind the solution of the problem is to find regular patterns in space and time by using the mathematical tool of EOF analysis. It is investigated into this technique and its applicability to GRACE gravity data. The EOF technique is a powerful tool for analyzing large geophysical fields that are correlated in space and time, different signal parts are separated according to their different variance structure. Thus, EOF analysis is a purely mathematical technique, and the results are primarily just a mathematical decomposition. The decomposition gives patterns in space and time that can be connected to corresponding physical sources.

(Wouters and Schrama, 2007) use EOF analysis on the coefficients of a representation of the gravity field in spherical harmonics and apply EOF analysis as a filter for filtering out correlated errors and (Rangelova et al., 2007) applied the EOF technique on gravity fields

only for North America, and they also proposed EOF-related techniques, such as multiple channel singular spectrum analysis, in (Rangelova et al., 2009). Several authors have applied EOF analysis to large data sets in oceanography and meteorology, see (Kutzbach, 1967) for an early example, or (Wilks, 1995).

In this thesis, first, the mathematical background of EOF analysis is explained. An introductory two-dimensional example visualizes the procedure of EOF analysis and it is followed by the mathematical derivation of the EOF method. Theoretical backgrounds are given and explained. Then, selection methods for distinguishing between signal and noise modes are presented from a theoretical point of view. The mathematical decomposition gives so-called modes, which contain information separated in temporal and spatial dimensions. In the second part of this thesis, a real GRACE data set is analyzed by the EOF method. First, the full method, consisting of analysis, selection of modes, and synthesis of the filtered data, is presented and explained in detail for a global data set. In the following two chapters EOF analysis for different regions is presented. These regions are continents and oceans separately, and an analysis for a smaller region, where Greenland is chosen as an example. In the discussions of the analysis and its results, strengths and limitations of the EOF method are explained, and the different regional examples are compared. Finally, in the outlook, three EOF analysis-related techniques are presented. These are a pre-whitening transformation, multiple channel singular spectrum analysis and canonical correlation analysis.

Chapter 2

Singular Value Decomposition

The primary idea of empirical orthogonal function (EOF) analysis according to (Jolliffe, 2002) is to reduce the dimensionality of a large data set while retaining most of the data sets' variance. By some authors the name principal component analysis (PCA) is used for basically the same method as EOF analysis. For a clarification of the conventions of PCA and EOF analysis see section 2.2.

Reducing the dimensionality and retaining most of the data set's variance is achieved by a linear coordinate transformation to a new set of basis vectors, which are chosen such that the first basis vectors are oriented along the directions of main variances of the data set. Opposite to the variables in the original data set which are correlated, the new basis vectors are chosen to be uncorrelated. In the first subsection 2.1, an illustrative example is given as an introduction to clarify the idea behind the technique. In the following sections, ideas are generalized and the mathematical background is explained in detail.

2.1 An introductory two-dimensional example

Following the examples of (Jolliffe, 2002), (Preisendorfer, 1988), and (Krzanowski, 2007), a two-dimensional example is used to demonstrate the method of Empirical Orthogonal Function Analysis for a simple example.
Suppose we have a data set with two spatial dimensions and $n = 25$ measurements at different times with equal time steps between each measurement. That means two values (x_n, y_n) for each time n. No units for the x and y dimensions are chosen, to keep the example abstract and applicable to any possible measurements. If we regard the bivariate data set as deviations

from a mean, the plot of the values could be the following, see the x, y-coordinate system in figure 2.1.

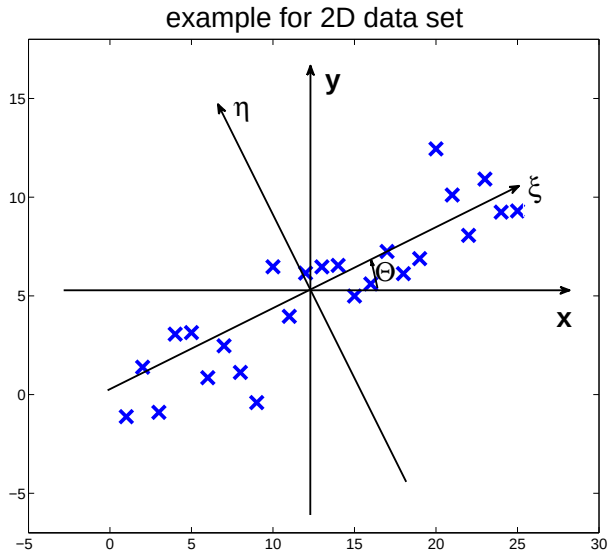


Figure 2.1: Two-dimensional data example

In the method of EOF Analysis, the aim is to find a new basis where the axes are chosen according to the variances in the data. The first axis should be oriented along the main variance and the second axis orthogonal to the first one. To represent the example data set in such a coordinate system, just a rotation about an angle Θ is needed, see figure 2.1. If we denote the points in the example data set as $P = (x(t), y(t))$, a point has the coordinates $P' = (\xi(\Theta, t), \eta(\Theta, t))$ in the rotated frame, where

$$\xi(\Theta, t) = x(t)\cos\Theta + y(t)\sin\Theta \tag{2.1}$$

$$\eta(\Theta, t) = -x(t)\sin\Theta + y(t)\cos\Theta. \tag{2.2}$$

The rotated ξ, η-frame is also shown in figure 2.1.

Now, consider the variances in the rotated frame. If $\mathbf{e}_1(\Theta)$ and $\mathbf{e}_2(\Theta)$ are the new basis vectors in the ξ, η-system, then the variance along $\mathbf{e}_1(\Theta)$ is

$$s^2(\Theta) = (n-1)^{-1}\sum_{t=1}^{n}\xi^2(\Theta, t) \tag{2.3}$$

$$= (n-1)^{-1}\sum_{t=1}^{n}[x(t)\cos\Theta + y(t)\sin\Theta]^2 \tag{2.4}$$

$$s^2(\Theta) = s_{xx}\cos^2\Theta + 2s_{xy}\sin\Theta\cos\Theta + s_{yy}\sin^2\Theta \tag{2.5}$$

with

$$s_{xx} = (n-1)^{-1} \sum_{t=1}^{n} x^2(t) \tag{2.6}$$

$$s_{yy} = (n-1)^{-1} \sum_{t=1}^{n} y^2(t) \tag{2.7}$$

$$s_{xy} = (n-1)^{-1} \sum_{t=1}^{n} x(t)y(t). \tag{2.8}$$

s_{xx}, s_{yy}, and s_{xy} are the variances and covariance in the original data set. Later on, s_{11}, s_{22}, and s_{12} shall be used for variances and covariances in the rotated frame. The data is centered, since the mean has been removed in the very first step. This leads to the factor $(n-1)^{-1}$ in the calculation of the empirical variance.

The angle Θ_m along which $s^2(\Theta)$ is maximum is the first principal angle. For bivariate data sets, as in the example data, the second principal angle is $\Theta_m + \frac{\pi}{2}$. Furthermore, the new basis vectors are chosen in a way that the data in the new frame is uncorrelated. That means the vectors are orthogonal to each other. In the two-dimensional example this is achieved by choosing the second vector in a direction of $\frac{\pi}{2}$ apart from the first one. In the following the angle Θ_m is calculated and it is shown that the data in the new coordinate frame is uncorrelated.

Consider the angle Θ_m, along which $s^2(\Theta)$ is maximum. Θ_m is found as a solution of

$$\frac{ds^2(\Theta)}{d\Theta} = -2s_{xx}\cos\Theta\sin\Theta + 2s_{xy}(-\sin^2\Theta + \cos^2\Theta) + 2s_{yy}\sin\Theta\cos\Theta = 0 \tag{2.9}$$

with following test for signs to distinguish between maxima and minima.

$$\text{Using} \qquad \sin\alpha\cos\alpha = \frac{1}{2}\sin 2\alpha \tag{2.10}$$

$$\text{and} \quad a\cos^2\alpha - a\sin^2\alpha = a\cos 2\alpha, \quad \text{which follows from} \tag{2.11}$$

$$\cos^2\alpha = \frac{1}{2}(1+\cos 2\alpha) \quad \text{and} \tag{2.12}$$

$$\sin^2\alpha = \frac{1}{2}(1-\cos 2\alpha), \tag{2.13}$$

leads to

$$\frac{ds^2(\Theta)}{d\Theta} = (s_{yy} - s_{xx})\sin 2\Theta + 2s_{xy}\cos 2\Theta = 0. \tag{2.14}$$

Solving for Θ yields

$$\Theta_m = \frac{1}{2} \arctan \frac{2s_{xy}}{s_{xx} - s_{yy}}. \tag{2.15}$$

Since $\frac{ds^2(\Theta)}{d\Theta} > 0$ for $\Theta < \Theta_m$, and $\frac{ds^2(\Theta)}{d\Theta} < 0$ for $\Theta > \Theta_m$, the variance as given in equation (2.5) is at maximum for Θ_m.

For any Θ the covariance $s_{12}(\Theta)$ of the given data set in the Θ-rotated frame is

$$s_{12}(\Theta) = (n-1)^{-1} \sum_{t=1}^{n} \xi(\Theta, t)\eta(\Theta, t). \tag{2.16}$$

By using the coordinate transformation equations (2.1), (2.2), and the variances and covariances in the original data set (2.6) - (2.8) the equation reduces to:

$$s_{12}(\Theta) = -s_{xx} \cos\Theta \sin\Theta + s_{xy} \cos^2\Theta - s_{xy} \sin^2\Theta + s_{yy} \sin\Theta \cos\Theta \tag{2.17}$$

and with the sine-cosine relations form equations (2.10) and (2.11) to

$$s_{12}(\Theta) = \frac{1}{2}(s_{yy} - s_{xx}) \sin 2\Theta + s_{xy} \cos 2\Theta. \tag{2.18}$$

For $\Theta = \Theta_m$ we find

$$s_{12}(\Theta_m) = 0 \tag{2.19}$$

that means for a rotation of the coordinate frame about angle Θ_m the data in the new frame will be uncorrelated.

The new basis vectors, which shall be called **Empirical Orthogonal Functions (EOFs)**, for the two-dimensional data example are

$$\mathbf{e}_1 = \begin{bmatrix} \cos\Theta_m \\ \sin\Theta_m \end{bmatrix} \qquad \mathbf{e}_2 = \begin{bmatrix} -\sin\Theta_m \\ \cos\Theta_m \end{bmatrix} \tag{2.20}$$

which are of unit length and orthogonal to each other. For the convention that $-\frac{\pi}{2} < \Theta_m \leq \frac{\pi}{2}$ the two basis vectors are also unique.

The two vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ form a basis in the two-dimensional data space. If we write each data point $(x(t), y(t))$ of the original data set as a vector $\mathbf{d}(t) = \begin{bmatrix} x(t) & y(t) \end{bmatrix}$, the components $u_1(t)$ and $u_2(t)$ along $\mathbf{e}_1$ and $\mathbf{e}_2$, which shall be called **Principal Components (PCs)**, are given by

$$u_1(t) = \mathbf{d}(t)\mathbf{e}_1 \tag{2.21}$$

$$u_2(t) = \mathbf{d}(t)\mathbf{e}_2 \qquad , t = 1, ..., n;\ n = 25 \tag{2.22}$$

Comparing equations (2.21) and (2.22) with (2.1) and (2.2) and using (2.20) we find that

$$u_1(t) = \xi(t) \qquad\qquad u_2(t) = \eta(t) \tag{2.23}$$

Each vector $\mathbf{d}(t)$ may be represented as a linear combination of $\mathbf{e}_1$ and $\mathbf{e}_2$:

$$\mathbf{d}(t) = u_1(t)\mathbf{e}_1^T + u_2(t)\mathbf{e}_2^T \qquad , t = 1, ..., n;\ n = 25 \tag{2.24}$$

This is the desired data representation showing that $\mathbf{e}_1$ and $\mathbf{e}_2$ form a new basis of the data space. $u_j(t)$ can be considered as amplitudes of the data set, and $\mathbf{e}_1$ and $\mathbf{e}_2$ as the modes of the variation of $\mathbf{d}$. Furthermore it can be concluded that the principal components $u_j(t)$ are the projections of the data vectors $\mathbf{d}(t)$ on the principal directions $\mathbf{e}_1$ and $\mathbf{e}_2$. The original data values are exactly reproducible by using the principal directions and principal components.

Matrix Representation

For further general algebraic discussions, considering a matrix form of the complete data set is very useful. The equation (2.24) can be written in matrix form as follows:

$$\begin{bmatrix} \mathbf{d}(1) \\ \vdots \\ \mathbf{d}(n) \end{bmatrix} = \begin{bmatrix} u_1(1) \\ \vdots \\ u_1(n) \end{bmatrix} \mathbf{e}_1^T + \begin{bmatrix} u_2(1) \\ \vdots \\ u_2(n) \end{bmatrix} \mathbf{e}_2^T. \tag{2.25}$$

Using

$$\mathbf{D} = \begin{bmatrix} \mathbf{d}(1) \\ \vdots \\ \mathbf{d}(n) \end{bmatrix} \ (n \times 2) \qquad \mathbf{U} = \begin{bmatrix} u_1(1) & u_2(1) \\ \vdots & \vdots \\ u_1(n) & u_2(n) \end{bmatrix} \ (n \times 2) \qquad \mathbf{E} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix} \ (2 \times 2) \tag{2.26}$$

the matrix representation can be obtained:

$$\mathbf{D} = \mathbf{U}\mathbf{E}^T. \tag{2.27}$$

In this two-dimensional example the matrix $\mathbf{E}$ is fully orthogonal and it holds

$$\mathbf{E}^T\mathbf{E} = \mathbf{E}\mathbf{E}^T = \mathbf{I_2}, \tag{2.28}$$

hereby $\mathbf{I_2}$ is the (2×2) identity matrix.

Note the main property of the principal components $\mathbf{U}$, the so-called PCA property, which is:

$$\mathbf{U}^T\mathbf{U} = \begin{bmatrix} \sum_{t=1}^{n} u_1^2(t) & \sum_{t=1}^{n} u_1(t)u_2(t) \\ \sum_{t=1}^{n} u_1(t)u_2(t) & \sum_{t=1}^{n} u_2^2(t) \end{bmatrix} = (n-1) \begin{bmatrix} s_{11} & 0 \\ 0 & s_{22} \end{bmatrix} \tag{2.29}$$

Equation (2.29) is obtained by using (2.16), (2.19), and (2.23). This means the principal component sets $u_1(t)$ and $u_2(t)$ are uncorrelated and their variances are $(n-1)\, s_{11}$ and $(n-1)\, s_{22}$, respectively.

Finally, we will show that the EOFs $\mathbf{e}_1$ and $\mathbf{e}_2$ are the eigenvectors of the spatial covariance matrix and that the variances in the rotated frame s_{11} and s_{22} are the eigevalues of the covariance matrix of the given data set.

The spatial covariance matrix of the original data set is

$$\mathbf{C} = \begin{bmatrix} s_{xx} & s_{xy} \\ s_{xy} & s_{yy} \end{bmatrix} \tag{2.30}$$

To prove that $\mathbf{e}_j$ and s_{jj} are the eigenvectors and eigenvalues respectively of $\mathbf{C}$, we must show that the equations

$$\mathbf{C}\mathbf{e}_j = s_{jj}\mathbf{e}_j \qquad j = 1, 2 \tag{2.31}$$

hold.

By looking at each line in equation (2.24) individually, squaring each side of these equations, and summing up over t we obtain

$$s_{xx} = s_{11}\cos^2\Theta_m + s_{22}\sin^2\Theta_m \tag{2.32}$$

$$s_{yy} = s_{11}\sin^2\Theta_m + s_{22}\cos^2\Theta_m \tag{2.33}$$

by subtraction equations (2.32) and (2.33) and using

$$a\cos^2\alpha - a\sin^2\alpha = a\cos 2\alpha \tag{2.34}$$

we obtain

$$s_{xx} - s_{yy} = (s_{11} - s_{22})\cos 2\Theta_m \tag{2.35}$$

and by using (2.15)

$$s_{xy} = \frac{1}{2}(s_{11} - s_{22})\sin 2\Theta_m. \tag{2.36}$$

Now, going back to the eigen-equation (2.31), if $j = 1$, then

$$\mathbf{C}\ \mathbf{e}_1 = \begin{bmatrix} s_{xx} & s_{xy} \\ s_{xy} & s_{yy} \end{bmatrix} \begin{bmatrix} \cos\Theta_m \\ \sin\Theta_m \end{bmatrix} = \begin{bmatrix} s_{xx}\cos\Theta_m + s_{xy}\sin\Theta_m \\ s_{xy}\cos\Theta_m + s_{yy}\sin\Theta_m \end{bmatrix} \tag{2.37}$$

$$= s_{11}\begin{bmatrix} \cos\Theta_m \\ \sin\Theta_m \end{bmatrix} = s_{11}\mathbf{e}_1 \tag{2.38}$$

with

$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta. \tag{2.39}$$

The proof for the case $j = 2$ is established similarly.

Define a scatter matrix $\mathbf{R}$ that just differs from the variance-covariance matrix by the factor $(n-1)$ as follows

$$\mathbf{R} \equiv \begin{bmatrix} \sum_{t=1}^{n} x^2(t) & \sum_{t=1}^{n} x(t)y(t) \\ \sum_{t=1}^{n} y^2(t) & \sum_{t=1}^{n} x(t)y(t) \end{bmatrix} \tag{2.40}$$

and define values l_j as multiples of the variances $l_j \equiv (n-1)s_{jj}$, $j = 1, 2$ and arrange them in a diagonal matrix $\mathbf{\Lambda}$

$$\mathbf{\Lambda} = \begin{bmatrix} l_1 & 0 \\ 0 & l_2 \end{bmatrix}. \tag{2.41}$$

Now, the eigen-equation (2.31) can be rewritten as a matrix equation

$$\mathbf{R}\,\mathbf{E} = \mathbf{E}\,\mathbf{\Lambda}, \tag{2.42}$$

matrix $\mathbf{E}$ still contains the same eigenvectors $\mathbf{e}_1$ and $\mathbf{e}_2$. Equation (2.42) bridges the gap to the general theory of Principal Component Analysis, which will follow in the next chapters.

2.2 Empirical Orthogonal Function and Principal Component Analysis

Before deriving the formulas that will be used later for calculations and explaining the mathematical background in detail, some clarification of notation is needed. Literature can be confusing when it comes to the methods of Empirical Orthogonal Function (EOF) Analysis, Principal Component Analysis (PCA) and Singular Value Decomposition (SVD). There are many diverse ideas in literature about how to call the different types of eigenvectors that are needed in the method. There are names as patterns, EOFs, EOF patterns, time series, PCs, PC time series, PC loading patterns, EOF time series and so on. In this subsection, names and conventions are established which are used in this thesis.

While *Empirical Orthogonal Function (EOF) Analysis* emphasizes the fact that the new basis vectors are found empirically and that these vectors are orthogonal to each other, the name *Principal Component Analysis (PCA)* emphasizes that the new basis is found according to the variances in the original data set. While, for example, (Jolliffe, 2002), (Preisendorfer, 1988), and (Krzanowski, 2007) use the term Principal Component Analysis, some authors, such as (von Storch and Zwiers, 2002), (Wilks, 1995), (Björnsson and Venegas, 2007), and (Rangelova, 2007), explicitly state that the two names are used for the same method. Additionally, there is a third group: (Blais, 2008), (Lorenz, 1956), (Schrama et al., 2007), (Hannachi, 2004), and (Eshel, 2003) use the name Empirical Orthogonal Function Analysis. However, all the authors describe the same method. In conclusion, the two names can be used interchangeably, depending on the application context and field of research. We will see later in this chapter that the decomposition yields EOFs and PCs, one gives spatial information and the other one temporal information. The two different names stress one or the other fact, but in the decomposition always both are derived.

In contrast, *Singular Values Decomposition (SVD)*, describes the core of the method. While the eigen-decomposition is only defined for square matrices, singular value decomposition is a generalization of the eigen-decomposition for rectangular matrices. Suppose $\mathbf{A}$ is a square matrix, it can be decomposed into

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1} \tag{2.43}$$

with $\mathbf{Q}$ as a matrix containing the eigenvectors as columns, and $\mathbf{\Lambda}$ a diagonal matrix with the eigenvalues on the main diagonal. When generalizing this method for any rectangular matrix $\mathbf{B}$, we need the singular value decomposition of the matrix. Thereby, the matrix is decomposed into three matrices, two orthogonal matrices and a diagonal matrix as follows:

$$\mathbf{B} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T. \tag{2.44}$$

The columns of $\mathbf{U}$ are called the left singular vectors of $\mathbf{B}$ and they are the eigenvectors of the matrix $\mathbf{B}\mathbf{B}^T$, the columns of $\mathbf{V}$ are called the right singular vectors of $\mathbf{B}$ and are the eigenvectors of $\mathbf{B}^T\mathbf{B}$, and the diagonal matrix $\mathbf{S}$ has the singular values on the main diagonal. The singular values are the square roots of the eigenvectors of $\mathbf{A}\mathbf{A}^T$ which are the same as those of $\mathbf{A}^T\mathbf{A}$. A short summary of the technique of singular values decomposition is given in (Abdi, 2007), the detailed mathematical background in (Lanczos, 1961), and computational aspects in (Press et al., 2007).

Concerning conventions for the matrices $\mathbf{U}$ and $\mathbf{V}$ several different names are used in literature and it can sometimes be confusing. For this thesis, the following conventions will be used: Let $\mathbf{D}$ be the data matrix that is decomposed into $\mathbf{D} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, the column vectors of $\mathbf{U}$ will be called *Principal Components (PCs)* and the column vectors of $\mathbf{V}$ will be called *Empirical Orthogonal Functions (EOFs)*. Matrix $\mathbf{\Sigma}$ still contains the singular values. These conventions are also used for example in (Preisendorfer, 1988). In my application on GRACE data, focus is on the global patterns of gravity changes, denoted as EOFs. For the transformation of the data on the new basis functions in order to obtain the PCs, the name *Empirical Orthogonal Function (EOF) Analysis* is used.

2.3 Derivation of the Method of Empirical Orthogonal Function Analysis

Suppose we have a data matrix $\mathbf{D}$ with the size $(n \times p)$, where each column of the matrix represents the time series for one point of the field, while the grid points of the field of measurements are arranged along the rows of the matrix. That means matrix $\mathbf{D}$ contains points $d(t, x)$, with $t = 1, ...n$ and $x = 1, ...p$ with t representing time and x representing

location. And it has the following form:

$$\mathbf{D} = \begin{array}{c} \\ t \downarrow \end{array} \begin{array}{c} \begin{array}{cccc} & & \rightarrow & x \end{array} \\ \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1p} \\ d_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ d_{n1} & \cdots & \cdots & d_{np} \end{bmatrix} \end{array} \quad (n \times p)$$

The following derivation of the method of EOF Analysis combines the major ideas of (Jolliffe, 2002) and (Preisendorfer, 1988) from their first chapters about the derivation of the method.

Let us define a scatter matrix $\mathbf{\Psi}(\mathbf{v})$ which describes the scatter of the elements of the data $\mathbf{D}$ along a new basis vector $\mathbf{v}$:

$$\mathbf{\Psi}(\mathbf{v}) = \sum_{t=1}^{n} [\mathbf{d}(t)\mathbf{v}]^2 = \sum_{t=1}^{n} [\mathbf{v}^T \mathbf{d}^T(t)][\mathbf{d}(t)\mathbf{v}] \tag{2.45}$$

$$= \mathbf{v}^T \left[\sum_{t=1}^{n} \mathbf{d}^T(t)\mathbf{d}(t) \right] \mathbf{v}, \tag{2.46}$$

where the basis vectors $\mathbf{v}$ are column vectors, and the vectors $\mathbf{d}(t)$ are row vectors of the data matrix, $\mathbf{d}(t) = [d_{t1}\, d_{t2}\, \ldots\, d_{tp}]$. The new basis vectors have been denoted $\mathbf{e}$ in the two-dimensional example to help the reader consider them as eigenvectors. From now on the eigenvectors shall be called $\mathbf{v}$.

Matrix $\mathbf{\Psi}$ can be considered as scatter or variance probe. It samples the scatter or variance of the data $\mathbf{D}$ along each vector $\mathbf{v}$.

The spatial scatter matrix $\mathbf{R_s}$, that differs from the empirical spatial variance-covariance matrix of the data set $\mathbf{D}$ only by the factor $\frac{1}{n-1}$, is defined as

$$\sum_{t=1}^{n} \mathbf{d}^T(t)\mathbf{d}(t) = \mathbf{D}^T\mathbf{D} \equiv \mathbf{R_s}. \tag{2.47}$$

Empirical variances always require deviations from a mean to be calculated. The calculation of the variance-covariance matrix, respectively scatter matrix, can be done that straightforward since we always assume to have the data centered, that means, the values in the data set are always deviations from a mean. This basic assumption is underlying the whole mathematical derivation. From now on, for the sake of simplicity in calculations, we will work with scatter

matrices instead of covariance matrices. Later on we will also need the temporal scatter matrix of the data, that is

$$\mathbf{R_t} \equiv \mathbf{D}\mathbf{D}^T. \tag{2.48}$$

In this case the same applies that was already said for the spatial scatter matrix. It differs from the temporal variance-covariance matrix only by the factor $\frac{1}{p-1}$.

Since the first EOF shall be chosen in a way that the variance along this vector is maximal, to derive the first EOF the term

$$\mathbf{v}_1^T \mathbf{R_s} \mathbf{v}_1 \tag{2.49}$$

is maximized. But by just considering (2.49), the maximum is for an infinite $\mathbf{v}$. A constraint needs to be imposed on the vectors $\mathbf{v}$. This constraint is the normalization of the new basis vectors, that is for the first one

$$\mathbf{v}_1^T \mathbf{v}_1 = 1. \tag{2.50}$$

Other constraints could be useful depending on the application of EOF Analysis, these could be according to (Jolliffe, 2002) $\max_j |v(j) = 1|$ or $\mathbf{v}^T\mathbf{v} =$ constant. For this thesis the normalization constraint mentioned in (2.50) is used.

The standard approach is to maximize a term subject to a condition by using the technique of Lagrange multipliers. This leads to the equation

$$\nabla(\mathbf{v}_1^T \mathbf{R_s} \mathbf{v}_1) - \lambda \nabla(\mathbf{v}_1^T \mathbf{v}_1 - 1) = 0. \tag{2.51}$$

Hereby, λ is the Lagrange multiplier and ∇ the gradient operator. This method is commonly used in mathematical optimization problems, named after Joseph Louis Lagrange. It is a method for finding the extrema of a function of several variables subject to one or more constraints and is the basic tool in nonlinear constrained optimization.
Differentiation of (2.51) with respect to $\mathbf{v}_1$ leads to

$$\mathbf{R_s}\mathbf{v}_1 - \lambda \mathbf{v}_1 = 0 \tag{2.52}$$

$$\text{or} \qquad \mathbf{R_s}\mathbf{v}_1 = \lambda \mathbf{v}_1. \tag{2.53}$$

Thus, λ is an eigenvalue of $\mathbf{R_s}$ and $\mathbf{v}_1$ the corresponding eigenvector.

To find the eigenvector for which $\mathbf{D}\mathbf{v}_1$ has maximum variance, note that the quantity to be maximized is $\mathbf{v}_1^T\mathbf{R}_\mathbf{s}\mathbf{v}_1$, (2.49), which we can manipulate by using (2.52) and (2.50):

$$\mathbf{v}_1^T\mathbf{R}_\mathbf{s}\mathbf{v}_1 = \mathbf{v}_1^T\lambda\mathbf{v}_1 = \lambda\mathbf{v}_1^T\mathbf{v}_1 = \lambda. \tag{2.54}$$

To maximize the variance along the new basis vector, λ must be as large as possible. That means the first eigenvector $\mathbf{e}_1$ is the eigenvector corresponding to the largest eigenvalue λ_1.

For finding $\mathbf{v}_2$, the second one of the new basis vectors, $\mathbf{v}_2^T\mathbf{R}_\mathbf{s}\mathbf{v}_2$ needs to be maximized. Since one of the preliminaries is that the new basis vectors and the PCs are uncorrelated, an additional condition for the optimization problem is one of the following equations, that all express the uncorrelatedness between the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$:

$$\mathbf{v}_1^T\mathbf{R}_\mathbf{s}\mathbf{v}_2 = 0 \tag{2.55}$$

$$\mathbf{v}_1^T\mathbf{v}_2 = 0 \tag{2.56}$$

Since two constraints are needed for finding the second basis vector, that is normalization as in (2.50), but for vector $\mathbf{v}_2$, and orthogonality of the new basis vectors as given in (2.56), we end up with the following Lagrange equation:

$$\nabla(\mathbf{v}_2^T\mathbf{R}_\mathbf{s}\mathbf{v}_2) - \lambda\nabla(\mathbf{v}_2^T\mathbf{v}_2 - 1) - \phi\nabla(\mathbf{v}_2^T\mathbf{v}_1) = 0. \tag{2.57}$$

Hereby, λ and ϕ are both Lagrange multipliers. Differentiating equation (2.57) with respect to $\mathbf{v}_2$ leads to

$$\mathbf{R}_\mathbf{s}\mathbf{v}_2 - \lambda\mathbf{v}_2 - \phi\mathbf{v}_1 = 0. \tag{2.58}$$

And after multiplying (2.58) with $\mathbf{v}_1^T$ on the left we have

$$\mathbf{v}_1^T\mathbf{R}\mathbf{v}_2 - \lambda\mathbf{v}_1^T\mathbf{v}_2 - \phi\mathbf{v}_1^T\mathbf{v}_1 = 0 \tag{2.59}$$

The first two terms are zero according to (2.56) and $\mathbf{v}_1^T\mathbf{v}_1 = 1$ according to (2.50). So (2.59) gives $\phi = 0$. Using $\phi = 0$, equation (2.58) reduces to

$$\mathbf{R}_\mathbf{s}\mathbf{v}_2 - \lambda\mathbf{v}_2 = 0 \tag{2.60}$$

and we have exactly the same form as in (2.52) for the second eigenvector $\mathbf{v}_2$, but with λ being the second largest eigenvalue λ_2.

As long as there are no exactly equal eigenvalues, which is almost never the case in data

sets from geophysical measurements, the eigenvectors are uniquely determined if one additional convention is used: the convention that the first component of each eigenvector is non-negative.

In the same manner all eigenvectors and eigenvalues of $\mathbf{R_s}$ can be derived. Finally, the matrix representation including all the eigenvectors and eigenvalues is:

$$\mathbf{R_s V} = \mathbf{V\Lambda}, \tag{2.61}$$

where the eigenvectors are the columns of matrix $\mathbf{V}$ and $\mathbf{\Lambda}$ is a diagonal matrix with the eigenvalues on the main diagonal. That means the EOFs, that are the new basis vectors, are exactly the eigenvectors of the spatial scatter matrix $\mathbf{R_s}$.

Note the diagonalizing property of the EOFs

$$\mathbf{V}^T\mathbf{RV} = \mathbf{\Lambda}, \tag{2.62}$$

which is proven straightforward from the eigenvalue-eigenvector equation (2.61) when replacing $\mathbf{RE}$ with $\mathbf{E\Lambda}$. The scatter matrix is transformed into a diagonal matrix. That means the data in the new frame is uncorrelated, as all possible covariances are equal to zero.

After having found the EOFs, that are the new basis vectors, the data has to be transformed into the new frame, that means the PCs have to be calculated. Since the PCs can be interpreted as the coordinates of the data in the new frame, the data matrix has to be projected on the new basis vectors in order to find the PCs:

$$\bar{\mathbf{U}} = \mathbf{DV} \tag{2.63}$$

$\bar{\mathbf{U}}$ is used as notation for the unnormalized version of the matrix $\mathbf{U}$, which contains the PCs. $\bar{\mathbf{U}}$ contains the unnormalized PCs. Let us denote the normalization factors for each vector in $\mathbf{U}$ as the elements on the main diagonal of the diagonal matrix $\mathbf{\Sigma}$. So the equation to calculate $\mathbf{U}$ becomes

$$\mathbf{U\Sigma} = \mathbf{DV} \tag{2.64}$$

$$\mathbf{U} = \mathbf{DV\Sigma}^{-1} \tag{2.65}$$

Note an interesting property of the principal components, as given for example in (von Storch and Zwiers, 2002), and demonstrated in the following set of equations. Starting from the

eigen-equation that is used to derive the EOFs in (2.61), we end up with an eigen-equation for the PCs:

$$\begin{aligned}
\mathbf{R_s V} &= \mathbf{V\Lambda} && (2.66)\\
\mathbf{D}^T\mathbf{DV} &= \mathbf{V\Lambda} && (2.67)\\
\mathbf{DD}^T\mathbf{DV} &= \mathbf{DV\Lambda} && (2.68)\\
\mathbf{DD}^T\mathbf{U\Sigma} &= \mathbf{U\Sigma\Lambda} && (2.69)\\
\mathbf{DD}^T\mathbf{U} &= \mathbf{U\Lambda} && (2.70)\\
\mathbf{R_t U} &= \mathbf{U\Lambda} && (2.71)
\end{aligned}$$

In the last line, equation 2.71, $\mathbf{R_t}$ shall be the time scatter matrix of the data as already given in 2.48. So the PCs are exactly the eigenvectors of the time scatter matrix, and can either be derived by means of eigenvectors or by means of a projection of the data onto the EOFs. Furthermore, note that the time and space scatter matrices have the same eigenvalues that are contained in the matrix $\mathbf{\Lambda}$. Of course this holds also for the other way around: Also the EOFs can be derived from a projection if the PCs as eigenvectors of the temporal scatter matrix have been derived first.

2.4 Mathematical and Statistical Properties of EOFs and PCs

PCA Property

First of all, the so-called **PCA property** shall be introduced. It is an orthogonality property which shows the uncorrelatedness of the PCs:

$$\begin{aligned}
\mathbf{U}^T\mathbf{U} &= \mathbf{\Sigma}^{-1\,T}\mathbf{V}^T\mathbf{D}^T\mathbf{DV\Sigma}^{-1} && (2.72)\\
&= \mathbf{\Sigma}^{-1}\mathbf{V}^T\mathbf{RV\Sigma}^{-1} && (2.73)\\
&= \mathbf{\Sigma}^{-1}\mathbf{\Lambda\Sigma}^{-1} && (2.74)
\end{aligned}$$

For $\mathbf{\Lambda} = \mathbf{\Sigma}^2 \;\rightarrow\; \mathbf{U}^T\mathbf{U} = \mathbf{I}$

That means if the normalization $\mathbf{\Sigma}$ of the principal components $\mathbf{U}$ is chosen so that the normalization factors are the square roots of the eigenvalues of the scatter matrix, contained in $\mathbf{\Lambda}$, then the PCs will be orthonormal to each other. In any case, the product of $\mathbf{U}^T\mathbf{U}$ leads to a diagonal matrix. Again, this shows that the data in the new frame, that is provided by the EOFs, is uncorrelated due to the fact that already the EOFs are chosen orthonormal.

Orthogonality and Completeness Property

Consider once again the EOFs and PCs. At the end of the last chapter we have found the property that the EOFs $\mathbf{V}$ are the eigenvectors of the spatial scatter matrix $\mathbf{R_s}$ and the PCs $\mathbf{U}$ are the eigenvectors of the temporal scatter matrix $\mathbf{R_t}$. This clearly emphasizes the fact that the temporal information from the data is contained in the PCs in $\mathbf{U}$, while the spatial information is contained in the EOFs in $\mathbf{V}$, which is one of the main properties of EOF analysis.

Both scatter matrices have the same eigenvalues λ_i and they both have the same rank. However, as a general data matrix is rectangular and has an arbitrary size of $(n \times p)$, the two scatter matrices will not have the same sizes. $\mathbf{R_s}$ is of the size $(p \times p)$ and $\mathbf{R_t}$ is of the size $(n \times n)$. Depending on whether n or p is smaller, one of the scatter matrices does not have full rank. For a data set of real geophysical measurements we can assume that the rank of the scatter matrices is $\min(n, p)$. So the matrix of eigenvectors of the scatter matrix that does have full rank will be square, while the eigenvector-matrix of the scatter matrix that does not have full rank will be rectangular.
Since eigenvectors are chosen to be orthogonal to each other, and since we normalize EOFs as well as PCs, both fulfill the **orthonormality property**:

$$\mathbf{v}_j^T\mathbf{v}_k = \delta_{jk} \qquad \text{respectively} \qquad \mathbf{u}_j^T\mathbf{u}_k = \delta_{jk} \tag{2.75}$$

where δ_{jk} is Kronecker's delta. The matrix version of equation (2.75) is

$$\mathbf{V}^T\mathbf{V} = \mathbf{I} \qquad \text{respectively} \qquad \mathbf{U}^T\mathbf{U} = \mathbf{I}, \tag{2.76}$$

where $\mathbf{I}$ is the identity matrix.
Only for the square one of the eigenvector matrices the **completeness property** holds:

$$\mathbf{E}\mathbf{E}^T = \mathbf{I}, \tag{2.77}$$

here $\mathbf{I}$ is again the identity matrix, and $\mathbf{E}$ is to be replaced by $\mathbf{U}$ or $\mathbf{V}$, depending on which one of the two scatter matrices has full rank and therefore which one of the eigenvector-matrices is square. See for example (Lanczos, 1961).

In the following, the analysis and synthesis steps are described in detail. An interesting observation is made: It is the orthonormality property (2.76) that allows for the analysis step to be calculated, but it is the completeness property (2.77) that permits the existence of the synthesis formula (Preisendorfer, 1988). In other words, completeness allows for every data set, more generally for every function, to be decomposed according to the new basis vectors and to be exactly reconstructed using all new basis vectors and the corresponding amplitudes.

Analysis

After having found the new basis vectors, that are the EOFs, the data can be analyzed. To analyze the data means to project the data onto the new basis vectors:

$$\mathbf{U\Sigma} = \mathbf{DV} \tag{2.78}$$

$$\mathbf{U} = \mathbf{DV\Sigma}^{-1} \tag{2.79}$$

Synthesis

After the PCs are found in the analysis step, the original data can be reconstructed using the EOFs, PCs and the normalization factors that are the square roots of the eigenvectors λ_i which have been shown to be a measure of the amount of variance contained in each new basis direction in equation (2.54). The equation for the reconstruction of the data is

$$\mathbf{D} = \mathbf{U\Sigma V}^T \tag{2.80}$$

and shall be proven in the following.
First, we consider $\mathbf{D}^T\mathbf{D}$ and compare the new expression for $\mathbf{D}$ in equation (2.80) with the previous definition for the scatter matrices, that are $\mathbf{R_s} = \mathbf{D}^T\mathbf{D}$ and $\mathbf{R_t} = \mathbf{DD}^T$. As it has been mentioned before, the EOFs are the column vectors of $\mathbf{V}$, the PCs are column vectors

of $\mathbf{U}$, and $\mathbf{\Sigma}$ is a diagonal matrix with $\mathbf{\Lambda} = \mathbf{\Sigma}^2$.

$$\begin{aligned} \mathbf{D}^T\mathbf{D} &= \mathbf{V}\mathbf{\Sigma}\mathbf{U}^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T && (2.81)\\ &= \mathbf{V}\mathbf{\Sigma}\mathbf{\Sigma}\mathbf{V}^T && (2.82)\\ &= \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T && (2.83)\\ &= \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T && (2.84) \end{aligned}$$

To proof that equation (2.80), which was used for $\mathbf{D}$, is correct, it has to be shown that the above statement is correct, what can be done by showing that equation (2.47), $\mathbf{R_s} = \mathbf{D}^T\mathbf{D}$, still holds.
From the eigenequation (2.61), $\mathbf{R_s}\mathbf{V} = \mathbf{V}\mathbf{\Lambda}$, follows under the use of the orthonormality property (2.76)

$$\begin{aligned} \mathbf{R_s}\mathbf{V} &= \mathbf{V}\mathbf{\Lambda} && (2.85)\\ \mathbf{R_s}\mathbf{V} &= \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T\mathbf{V} && (2.86)\\ \mathbf{R_s} &= \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T && (2.87) \end{aligned}$$

Thus, from the two equations (2.84) and (2.87) follows $\mathbf{D}^T\mathbf{D} = \mathbf{R_s}$, which proves that $\mathbf{D} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ holds, see (Lanczos, 1961). Of course, this proof can be done in a very similar manner for using the temporal scatter matrix $\mathbf{R_t}$.

Another approach for deriving the synthesis formula is to assume that for the matrix of the EOFs, $\mathbf{V}$, the completeness property $\mathbf{V}\mathbf{V}^T = \mathbf{I}$, equation (2.77), holds.
From

$$\mathbf{D} = \mathbf{D}\mathbf{V}\mathbf{V}^T \qquad (2.88)$$

follows with the analysis formula $\mathbf{U}\mathbf{\Sigma} = \mathbf{D}\mathbf{V}$

$$\mathbf{D} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T. \qquad (2.89)$$

Slightly more complicated is the proof for assuming that the completeness property (2.77) holds for $\mathbf{U}$, the matrix of the PCs. The eigen-equation for the spatial scatter matrix (2.61)

gives $\mathbf{D}^T\mathbf{D} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$, see also equation (2.87). Additionally, the analysis formula (2.79) is needed and leads to:

$$\begin{aligned}
\mathbf{D}^T\mathbf{D} &= \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T && (2.90)\\
\mathbf{D}^T\mathbf{D}\mathbf{V} &= \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T\mathbf{V} && (2.91)\\
\mathbf{D}^T\mathbf{D}\mathbf{V} &= \mathbf{V}\mathbf{\Lambda} && (2.92)\\
\mathbf{D}^T\mathbf{U}\mathbf{\Sigma} &= \mathbf{V}\mathbf{\Lambda} && (2.93)\\
\mathbf{D}^T\mathbf{U} &= \mathbf{V}\mathbf{\Lambda}\mathbf{\Sigma}^{-1} && (2.94)\\
\mathbf{D}^T\mathbf{U} &= \mathbf{V}\mathbf{\Sigma} && (2.95)\\
\mathbf{U}^T\mathbf{D} &= \mathbf{\Sigma}\mathbf{V}^T && (2.96)\\
\mathbf{D} &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T. && (2.97)
\end{aligned}$$

The synthesis formula (2.80), which has now been proven from different approaches, shall be dealt with next. The synthesis formula is exactly the same as the singular value decomposition formula for a rectangular matrix in (2.44). The general **singular value decomposition** including matrix sizes is:

$$\mathbf{D}_{\ (n\times p)} = \mathbf{U}_{\ (n\times\min(n,p))}\ \mathbf{\Sigma}_{\ (\min(n,p)\times\min(n,p))}\ \mathbf{V}^T_{\ (p\times\min(n,p))} \tag{2.98}$$

The values in the diagonal matrix $\mathbf{\Sigma}$ are called the singular values. Each triplet consisting of one singular value, corresponding EOF, and corresponding PC, forms one mode. It was mentioned before, that the singular value is a measure for the variance explained by each mode, while the spatial information is contained in the EOFs, and the temporal information in the PCs. Each mode's EOF and PC are orthogonal to the other EOFs and PCs respectively. In the data analysis, we have to look at the single modes to learn more about the data from the decomposition.

Since the EOFs are chosen according to variances in the data, the goal in EOF Analysis is that the signal parts are mapped into different modes, according to their variance. The question on how to distinguish modes that contain signal from those that contain only noise shall be discussed later in chapter 3. According to equation (2.98) the data can be exactly reconstructed when all modes are used. But when only those modes, that contain signal,

are used for the data synthesis, EOF analysis acts as a filter on the data. Additionally, the dimension of the data can be significantly reduced in the synthesis step by rejecting noisy modes for the synthesis. Often just a few modes are enough to reconstruct the part of interest in the data. In the ideal case, only modes containing signal would be used for reconstruction and all modes containing noise would be omitted. Then an optimal reconstruction of the data would be achieved, with all the noise filtered out. Of course, in reality an exact separation of signal and noise is never possible. The following formula (2.99) shows the reconstruction of the data using only specific modes that are selected by χ_i:

$$\mathbf{D}' = \sum_i \chi_i \, \sigma_i \mathbf{u}_i \mathbf{v}_i^T \qquad \text{with} \qquad \chi_i = \begin{cases} 0 & \text{if mode is rejected} \\ 1 & \text{if mode is accepted.} \end{cases} \tag{2.99}$$

In the following chapter, different methods for selecting the modes that are used for reconstruction of the data are presented.

Chapter 3

Selection Methods

In the different modes obtained from EOF Analysis, we hope to identify different sources of gravity changes. The goal of EOF Analysis is to project each source of mass change into a separate mode. However, EOF Analysis can also be used as a filter, in a way where only signal-containing modes are used for synthesis of the data. Thus, the question that arises is how to choose which modes contain signal and which ones are only representing noise. Furthermore, if a function is established that automatically detects the modes that are containing signal, EOF Analysis can be considered as a filtering function, with the data set as input and the filtered data set as output.

In the first subsection 3.1, an overview of techniques for selecting the signal-representing modes is given. There is a wide range of techniques available, and of course it strongly depends on the application which techniques will perform well, and which one will be less suitable. In the second part, 3.2, a rule based only on the singular values is discussed in detail, and in the third subchapter 3.3, a hypothesis test for selecting the appropriate modes is introduced.

3.1 Overview of Different Techniques

For the synthesis equation $\mathbf{D}' = \sum_i \chi_i \, \sigma_i \mathbf{u}_i \mathbf{v}_i^T$, as given in (2.99), one has to decide which modes to keep for the reconstruction and which modes to consider as noise and therefore not to use in the reconstruction. Of course, the major difficulty is getting rid of as much of the noise as possible, but not loosing signal. Separation of signal and noise into different modes is a preliminary, but in reality it is not possible to perfectly separate signal and noise when

the data is projected onto the EOFs. The quality of the individual modes also depends on the quality of the data.

In literature, different types of procedures for choosing the modes to keep are proposed. Straight-forward ones and very complicated methods. In (Preisendorfer, 1988), three broad procedures are distinguished:

- **Dominant variance rules**

  These methods are associated with the singular values contained in the matrix $\mathbf{\Sigma}$ and are based on the premise that the larger variance terms are associated with signal. Since the variance explained in each mode is represented by the magnitude of the singular values, the magnitude curve of the singular values can be used. One strategy is to look for a drop in magnitude that could mark the border between modes which represent signal and modes which represent noise. Modes corresponding to singular values larger than that one at the drop are considered as signal and those with smaller singular values as noise. In other words, it is searched for a truncation parameter to split the sum in the reconstruction into two sums where the first part is associated with signal and the second part is considered to be noise. Another strategy is to look at the last singular values and declare those close to zero and those very similar to others as representing noise. Only modes with distinct singular values would be used in the synthesis. For more information on the methods based on the singular values please refer to chapter 6 in (Jolliffe, 2002) and to chapter 5 in (Preisendorfer, 1988).

- **Time series rules**

  Methods associated with PCs, the column vectors in the matrix $\mathbf{U}$, are also called time series rules since these methods examine the temporal behavior. The time series for each mode are tested for non-noisy temporal behavior. After performing a hypothesis test for the time series of each mode, these modes with the time series that pass the test are retained for reconstruction. While in the dominant variance rules the major idea is to find an optimal truncation parameter and consider all modes with singular values larger than the truncation values as signal and all smaller than that value as noise, in the time series rules each mode is tested individually. That means the modes

that are kept, those that pass the test, do not necessarily have to be subsequent. For further detailed explanations see (Preisendorfer, 1988).

- **Space-map rules**

  These selection rules use the EOFs, the column vectors of the matrix $\mathbf{V}$, as indicators for the noisiness of a mode. The basic method of most of the space-map techniques is to compare the spatial map of each mode to a pre-defined spatial map or to a pre-defined spatial pattern which can also be smaller than the mode's spatial map. The idea is, if there are known patterns that are expected to appear in the maps, these patterns are compared to all the structures that appear in the EOF maps. It is tested for correlations to find the signal containing modes. For more details and examples for the space-map rules refer to (Preisendorfer, 1988).

In (Jolliffe, 2002), selection rules are not distinguished according to the part of each mode that is used for the evaluation, but the methods are distinguished by the type of technique in use. There are the following two broad groups:

- **Ad hoc rules-of-thumb**

  Some examples of techniques in the group of these more simple methods are

  - The cumulative percentage of total variation: The first few modes are chosen so that the sum of their variances in percent exceeds a certain percentage of the total variance, for example 80% or 90%. Since the singular values are a measure of the amount of variance expressed by each mode, the percentage of total variance can be calculated. The trace of the singular values matrix $\mathbf{\Sigma}$, that is the sum of the singular values for all modes, is a measure for the total variance in the data.
  - Size of variances, for which the size of singular values are a measure, since the variance contained in each mode is a $(n-1)$ multiple of the singular value: A cut-off size for the singular values is chosen, which gives a minimum variance to be explained by each signal mode. All modes expressing less variance than that value are not retained. In the case to perform EOF analysis by using a correlation matrix

instead of a covariance matrix, that cut-off value would be chosen just below one. If all elements of the original data set were independent and uncorrelated, the new basis vectors would be the same as the original ones and the correlation matrix would be an identity matrix. Thus any mode with variance less than one contains less information than one of the original variables and is not worth retaining.

- The eigenvalue diagram: Described for example in (Krzanowski, 2007) or (Jolliffe, 2002). The singular values are plotted in a diagram, and sometimes a logarithmic plot can be helpful, to find a cut-off value from a change in the shape of the curve. This cut-off value marks the highest mode that is used for synthesis.

Compared to the categorization previously mentioned, and given by (Preisendorfer, 1988), these techniques would all be in the group of rules based on the singular values.

- **Rules based on formal tests of hypothesis**

  Some examples for rules based on hypothesis testing as described in (Jolliffe, 2002) are

  - Testing the singular values associated to each mode for inequality. A Bartlett's test is performed on the singular values: The null hypothesis is that two singular values following each other, that is σ_i and σ_{i+1}, are not significantly different. Starting with the first two modes, if the null hypothesis is rejected, the next eigenvalue pair is tested. The procedure is repeated going on to higher modes until two singular values which are no more significantly different are found. These mark the cut-off mode. The idea behind this testing method is the assumption that the higher modes, that are most likely explaining noise, have very similar singular values, if they express the same amount of variance, that it the variance of white noise. Or, stated in an equation, $\sigma_1 > \sigma_2 > \sigma_3 > \sigma_i > \sigma_{i+1} = \ldots = \sigma_m$.
  - Correlations between pre-defined patterns and patterns in the EOFs. Patterns which are expected to be found in the data are used to search for correlations with patterns in the EOFs. The pre-defined patterns do not need to have the same size as the EOFs. Additionally, partial correlations are used, that means the patterns do not have to be fully correlated to identify a mode as being different from white noise. Note that there is a major difference between this testing method

and the other described methods: For this selection method prior information is needed about the signal, respectively the patterns, to be detected.

– Testing time series characteristics: Any white noise test could be used to distinguish between time series that contain signal and time series that contain only white noise. A suitable test is a Kolmogorov-Smirnov test to compare each time series' power spectral density to the power spectral density of white noise. This test shall also be investigated in detail in the following subchapter.

In chapter 3.2, methods based on the singular values and a method based on a hypothesis test will be described in detail. The hypothesis test method will be from the group of time-history rules according to (Preisendorfer, 1988).

3.2 Rules of Thumb Based on Singular Values

The first methods that will be discussed are very simple rules based on the singular values. The selection method which performs best, is highly dependent on the data set that is to be analyzed. Thus, sometimes even those simple rules can lead to good results. Since these rules do not require a lot of calculation effort, it can be useful to check the performance of those simple rules before spending a lot of time on the implementation of a more complicated hypothesis test. The basic idea behind these techniques is that the singular values are a measure for the variance expressed by each mode. It is assumed, that a signal part, that is to be detected, has a higher variance than noise and furthermore that different signal parts have different sizes of variance. It is expected that the first modes, that represent signal, have significantly higher singular values than the noise-representing modes. Furthermore, singular values of noise-representing modes are very similar to each other, since white noise has no characteristic variance structure and they are projected into lower modes since the first EOFs are chosen according to distinguishable structures in the variance.

By dividing each singular value by the sum of all singular values, the percent value of total variance expressed by each mode is obtained. Since the modes are ordered according to the sizes of the singular values, the first modes express most percentage of variance. By summing

up the first few singular values, a very high percentage of total variation can be achieved. Considering data dimension reduction, this approach is very helpful, when with a very few modes most of the variance, about 80 to 95%, can be retained, while the dimensionality of data representation can be significantly reduced. Reconstructing the filtered data set with only those modes retained according to percentage of variance considerations, most of the original information will still be contained, while a lot of useless modes that only contribute noise are omitted. See for example (Preisendorfer, 1988).

Another approach using singular values makes use of the curve of the plotted singular values. It is expected, as already mentioned, that the singular values of modes representing noise are very similar to each other while those representing signal are expected to have individual characteristics. A plot of the singular values for all modes is helpful to examine the sequence of singular values. See figure 3.1 for an example. In this graph we need to look for a characteristic change, most probably a drop in magnitude, that could mark the border between signal-containing modes and modes that are just representing noise. Sometimes a logarithmic plot of the singular values can be helpful, depending on the data set to be analyzed. See also (Jolliffe, 2002). The cut-off value m' that we are looking for divides the sum used in the reconstruction of the data into two sums:

$$\mathbf{D}' = \sum_{i=1}^{m=\min(n,p)} \sigma_i \mathbf{u}_i \mathbf{v}_i^T = \sum_{i=1}^{m'} \sigma_i \mathbf{u}_i \mathbf{v}_i^T + \sum_{i=m'+1}^{m} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \tag{3.1}$$

The second sum is considered to contain a non-significant amount of variance while first sum is thought of as the signal of interest (Preisendorfer, 1988).

Having a closer look at the example singular values graph in figure 3.1 we can see that the singular values of the high modes, except the last four, are very close together. The last four modes whose singular values are extremely small, do not contribute a lot to the reconstruction, especially the last mode, whose singular value is almost zero, can be considered as not contributing any information. The number of non-zero singular values equals the rank of the data matrix. The last singular value is zero since the data matrix does not have full rank, the time mean was removed. So the last singular value can be omitted right away. Due to these last four singular values which are very small compared to the other ones, a logarithmic plot in this special case would not be helpful, as it emphasizes the differences between the

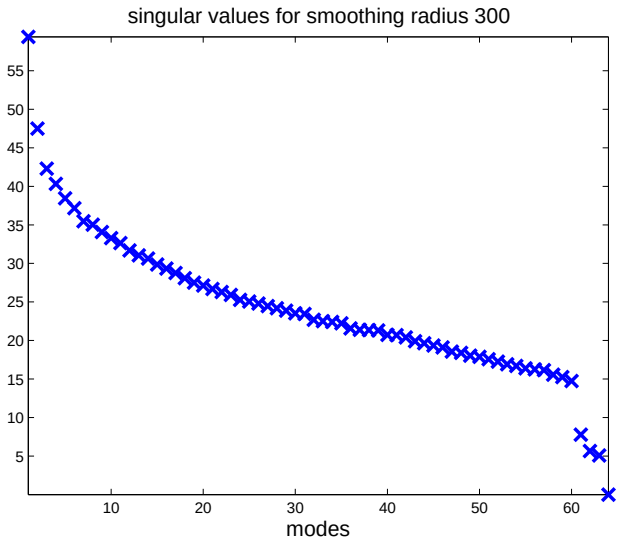


Figure 3.1: Singular values for all modes

majority of the values and those last four, but what would be helpful is an emphasis of the first few largest values, as they are of much more interest than the last ones.
In the middle of the graph, the singular values are very dense and close together in magnitude. The modes in this area are those considered to represent noise. In the first few modes, up to the sixth mode, the values drop significantly in magnitude and especially the first two are definitely separated from the other values. So we would conclude from this diagram in figure 3.1 that the first two modes are for sure representing signal as they are very distinct. Modes three to six are critical to decide on, because there is a difference in the appearance of these four singular values and all the following modes but they are not as significantly different from the others as the first two. All the other higher modes can be considered to be only noise. So the cut-off value drawn from the graph would be $m' = 2$ or $m' = 6$.

While in the first method the cut-off value is found by looking at the cumulative sum of the percentages of the variance expressed, in the second method proposed the cut-off value is found from graph characteristics. Note that for both methods just a cut-off value is searched. That means those modes that are identified as signal are sequential modes. No individual test for each mode is performed, but is looked at the singular values in comparison to each other. In contrast, in the second method that will be described in the following subchapter, 3.3, each mode is tested individually and the outcome of the test does not depend on the other modes. Of course, for this selection method, the modes that are considered to be signal

are not necessarily subsequent modes.

3.3 Kolmogorov-Smirnov Test

This hypothesis test selection rule is based on a Kolmogorov-Smirnov (KS) hypothesis test for spectral whiteness of the PC time series for each mode. For each PC time series the Fourier transform is calculated, which gives the power spectrum density, that is tested for spectral whiteness. The test for spectral whiteness is performed by comparing the cumulative distribution function of the power spectrum density to the cumulative power spectrum density of white noise. The cumulative power spectrum density of white noise is a linear function ranging from zero to one, since power spectral densities are normalized before comparison. Ideal white noise is evenly distributed over all frequencies and its cumulative power spectrum is a line. The time series do not have have endless length, they have the same size as the time domain in the data set. Thus, for performing a useful and reliable hypothesis test in comparing the power spectral density to white noise, the cumulative sum has to be used for the test instead of the spectral density.

Two examples illustrating KS test are given in figure 3.2. The figures show for two different modes, mode 1 and mode 5, the PC time series, the corresponding power spectral density and the cumulative power spectral density that is compared to the cumulative power spectrum density of white noise. In the plots of the cumulative power spectra, white noise is shown in red, and each time series' cumulative power spectrum is compared to this white noise line. The maximum difference gives the test statistic. Mathematical details are given below.

Comparing the two modes in the example, we can clearly see the different between modes that contain noise and modes that contain signal. Taking the example of the annual signal in the first mode we can clearly see the peak in the power spectral density at one cycle per year. Therefore, the cumulative power spectrum is significantly different from that one of white noise. However, in figure 3.2 on the right, there is no pattern that could be identified in the time series and also no signature in the power spectral density. So this cumulative spectrum is not significantly different from the white noise curve, as it can be seen in the

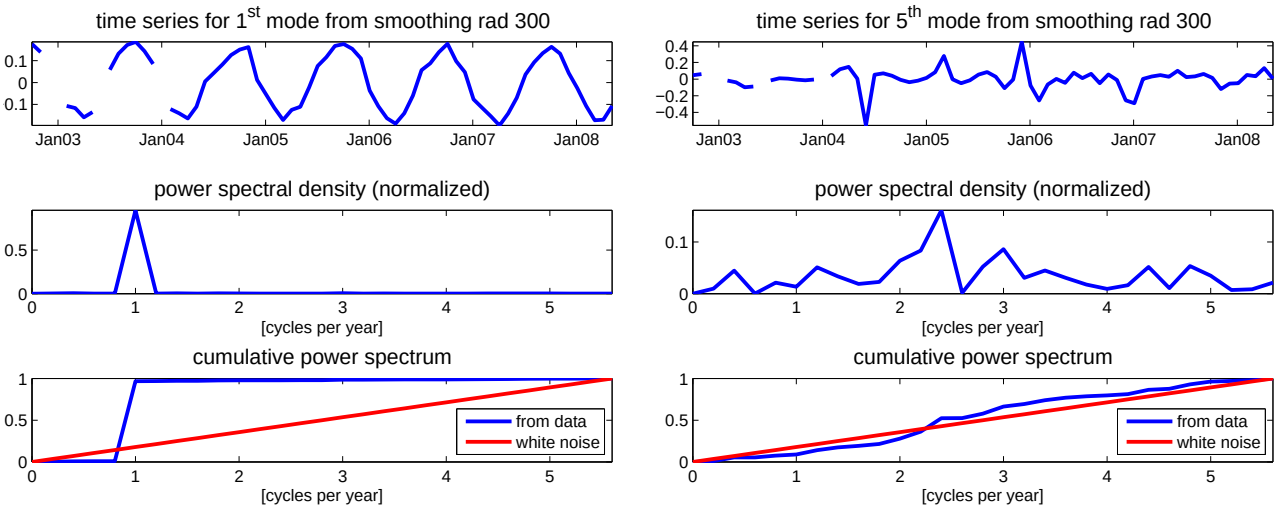


Figure 3.2: Examples for KS test: signal versus noise

plot, and the mode would be considered as representing only noise.

For performing the KS test, as a first step, the spectrum of the time series is estimated. In case of equal spacing between the values and time series without gaps, a simple fast Fourier transform can be used. However, if the time series have gaps, the spectrum has to be estimated using a least squares method. We will see later, that the time series from GRACE have gaps, as well as a lot of other geophysical measurements do. Thus, the technique for handling time series with gaps is introduced here.
Of course, gaps could just be filled with interpolated values for being able to use a Fast Fourier Transform algorithm. But by interpolating no information is added to the time series.

The formula of the discrete Fourier transformation is the following:

$$p(t_i) = a_0 + \sum_{n=1}^{N} a_n \cos \omega_n t_i + b_n \sin \omega_n t_i, \tag{3.2}$$

$$\text{with } \omega_n = n \, \frac{2\pi}{T} = n \, \omega_0,$$

where $p(t_i)$ are the time series values at particular points of time t_i and ω_n are particular frequencies at which the magnitude of the signal is to be determined. Any distinct time vector and any distinct frequency vector can be handled with the least squares Fourier transform

approach.

The equation to estimate the Fourier coefficients is the following, and it is solved for the coefficients a_n and b_n by a standard least squares approach.

$$\begin{bmatrix} p(t_1) \\ \vdots \\ p(t_m) \end{bmatrix} = \begin{bmatrix} \cos\omega_1 t_1 & \sin\omega_1 t_1 & \cos\omega_2 t_1 & \dots & \cos\omega_n t_1 & \sin\omega_n t_1 \\ \vdots & & & & & \vdots \\ \cos\omega_1 t_m & \sin\omega_1 t_m & \dots & \dots & \cos\omega_n t_m & \sin\omega_n t_m \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \\ a_2 \\ \vdots \\ a_n \\ b_n \end{bmatrix} \tag{3.3}$$

Since one of the requirements for EOF Analysis was time centered data, it follows that a_0 is zero and not included in the estimation. In case there is a linear trend in the time series, it would be projected into the lowest frequency. A better solution is to search for a trend and subtract it before the spectrum is estimated. After having estimated a_1 to a_n and b_1 to b_n, by squaring the respective coefficients for each frequency and summing the particular pair, the power spectral density is obtained according to the following formula 3.4:

$$c_n^2 = a_n^2 + b_n^2. \tag{3.4}$$

In general, in the KS test the cumulative distribution function from a sample is compared to a known distribution function. Let $S_N(x)$ denote an estimator of the cumulative distribution function of the probability distribution from which the sample was drawn, and let $P(x)$ denote a known cumulative distribution function. According to (Press et al., 2007) the KS test statistic is

$$D = \max_{-\infty<x<\infty} |S_N(x) - P(x)| \tag{3.5}$$

The null hypothesis is that the sample is drawn from the known distribution. In the application for testing time series for spectral whiteness, the known distribution is from white noise, and the null hypothesis can be formulated as follows:

$$H_0 : \{p(t), t = 1, ..., n\} \quad \text{is a random sample from a white noise process} \tag{3.6}$$

If the null hypothesis is accepted, the times series should be considered as white noise. The test statistic D is compared to a value d that depends on the significance level of the test and

the sample size. For further details of the KS test and the Kolmogorov-Smirnov distribution please refer to (Press et al., 2007). If the value D is larger than d, the time series is considered to be significantly different from white noise. In that way each mode's time series is tested and those modes for which the null hypothesis is accepted are considered as noise while those for which the null hypothesis is rejected, are classified as signal.

For more detailed explanations about the usage of KS test to test PC time series please refer to (Preisendorfer, 1988), and for a practical example of the use of KS test in EOF Analysis on the spherical harmonic coefficients of GRACE data see (Wouters and Schrama, 2007).

Chapter 4

The Data Set

In the second part of this thesis applications of the method of EOF analysis of GRACE gravity data are investigated. The GRACE (Gravity Recovery and Climate Experiment) mission consists of two satellites, which are making detailed measurements of the Earth's gravity field. The two satellites measure changes in their distance very accurately, and the position of both satellites is determined by GPS receivers. From these measurements, a regular grid of gravity changes covering the globe is calculated. These gravity changes are contaminated by measurement and processing errors, and it is an important issue to separate signal from noise.

In the first subchapter details about the satellite mission are given and it is shown how the data is acquired. Then the data in use is described in detail and the construction of the required data matrix for EOF analysis is explained.

4.1 GRACE Gravity Measurements

In this subsection the GRACE mission and the data collection with the GRACE satellites shall be introduced. First, information about the mission and instruments is given, while in the second subsection the data post-processing is explained, which is needed to obtain usable data sets from the satellite measurements.

The Gravity Recovery and Climate Experiment (GRACE) Mission

The objective of the jointly NASA and DLR (German Aerospace Center) two-satellite mission GRACE is to map the Earth's global gravity field with a spatial resolution of about

400 km every thirty days. GRACE was launched on March 17, 2002 and consists of two identical satellites in near-circular orbits at about 500 km altitude and 89.5° inclination. The two satellites are separated from each other along-track by about 220 km. A K-band microwave ranging system is used to measure their distance very accurately. Additionally to the inter-satellite ranging system, each satellite is equipped with Global Positioning System (GPS) receivers, attitude sensors, and accelerometers. Since the two satellites do not fly drag-free, the satellite altitude decays due to atmospheric drag forces. Thus, the ground track does not have a fixed repeat pattern. But the satellites are held in a three-axis-stabilized, nearly Earth-pointed orientation, so that the two K-band antennas are always precisely pointed at each other. The microwave K-band measurements are transmitted to the ground, where they are processed to receive an accurate range measurement. Two different frequencies are used in the measurements for being able to eliminate the atmospheric effects in the data processing. The effects of non-gravitational forces acting on the satellites are removed by using the precise accelerometer measurements of the surface force acceleration. The measurements from the GPS receivers are used for precise time-tagging of the inter-satellite range measurements and to provide the positions of the satellites above the Earth, see (Tapley et al., 2004). For discussion about the errors in each of these systems and about how they contribute to the gravity field estimates please refer to (Kim and Tapley, 2002).

Data Processing

The precise distance between the two GRACE-satellites as they fly in formation over the Earth is constantly measured by the microwave K-band ranging instrument. As the gravitational field changes beneath the satellites, due to changes in mass of the surface beneath, the orbital motion of each satellite is changed. This change in orbital motion causes the distance between the satellites to expand or contract and can be measured using the K-band instrument. Based on these range measurements, the fluctuations in the Earth's gravitational field can be determined in post-processing.

For the raw satellite measurements, which are distances between the two satellites and GPS positions, fairly complicated processing is necessary to obtain spherical harmonic coefficients for the Earth's gravity field and its temporal changes. The GRACE data products are developed, processed and archived in a shared Science Data System (SDS) between the Jet

Propulsion Laboratory (JPL) in Pasadena, the University of Texas Center for Space Research (UTCSR) and the Geoforschungszentrum (GFZ) in Potsdam, Germany. Before the data can be sent to one of these centers for post-processing, the raw data has to be sent to a receiving station on Earth, two stations are in Weilheim and Neustrelitz in Germany. GRACE data products are available at three processing levels. These are:

- **Level 0:** The level 0 data are the result of the data reception, collection and decommutation by the raw data center located in Neustrelitz in Germany. The SDS retrieves these files and extracts and reformats the corresponding instrument and ancillary housekeeping data like GPS navigation solutions, space segment temperatures or thruster firing events. Level 0 products shall be available 24 hours after data reception.

- **Level 1:** Level 1 data are the preprocessed and time-tagged instrument data. These are the K-band ranging, accelerometer, and GPS data of both satellites. Additionally, the preliminary orbits of both GRACE satellites are generated. Software for level 1 data processing has been developed both at JPL and at GFZ. Since level 1 data can be processed at two sites, it can be guaranteed that level 1 data products are available with a delay of 5 days.

- **Level 2:** Level 2 data include the short term (30 days) and mean gravity field derived from calibrated and validated GRACE level 1 data products. This level also includes ancillary data sets (temperature and pressure fields, ocean bottom pressure, hydrological data) which are necessary to eliminate time variabilities in gravity field solutions. Additionally, the precise orbits of both GRACE satellites are generated. All level 2 products shall be available 60 days after data taking. The level-2 processing software was developed independently by all three processing centers JPL, UTCSR, and GFZ. Routine processing is done at UTCSR and GFZ, while JPL will only generate level 2 products at times for verification purposes.

Information about the GRACE data processing levels is derived from (Flechtner, 2001). For further information about the post-processing of GRACE data at the GFZ please refer to (Flechtner, 2007).

For this thesis, level 2 data that was processed at the GFZ in Potsdam is used. The data set contains the spherical harmonic coefficients for a global gravity field solution for every month.

4.2 Details of the Data in Use

The given gravity data from GRACE observations are the coefficients of spherical harmonic functions up to degree and order 120 from the GFZ in Potsdam, from October 2002 to May 2008. The observations are time series of geoid changes with respect to the mean of all months, where the mean is removed from each coefficient for each month.
Then, the monthly maps of mass changes are synthesized from the coefficients and a grid of values in a $1° \times 1°$ size is produced. Before creating the gridded maps in equivalent water height, Gaussian smoothing is applied to the coefficients. This is necessary to obtain a reasonable signal-to-noise ratio. For more details about the smoothing please refer to the upcoming chapter 5.1.

However, some months are completely missing in the data set, these missing months are

- December 2002
- January 2003
- June 2003
- January 2004

For the technique of EOF analysis these gaps in the data do no harm. However, when working with the time series, the PCs from the EOF analysis, we have to take into account the gaps due to the missing months. The reconstruction part of EOF analysis can even be used for an estimation of the missing months, according to (Blais, 2008).

Repeat Orbits of the GRACE satellites

In the year 2004, the ground tracks of the GRACE orbits in September, but also in the preceding and following months, show a pattern that is different from the ordinary ground tracks.

Figure (4.1) shows a plot of the ground tracks for January 2004 compared to a plot of the ground tracks for September 2004. The zoomed figures shall help see the details of the respective ground tracks. January is considered to have an ordinary ground track while in

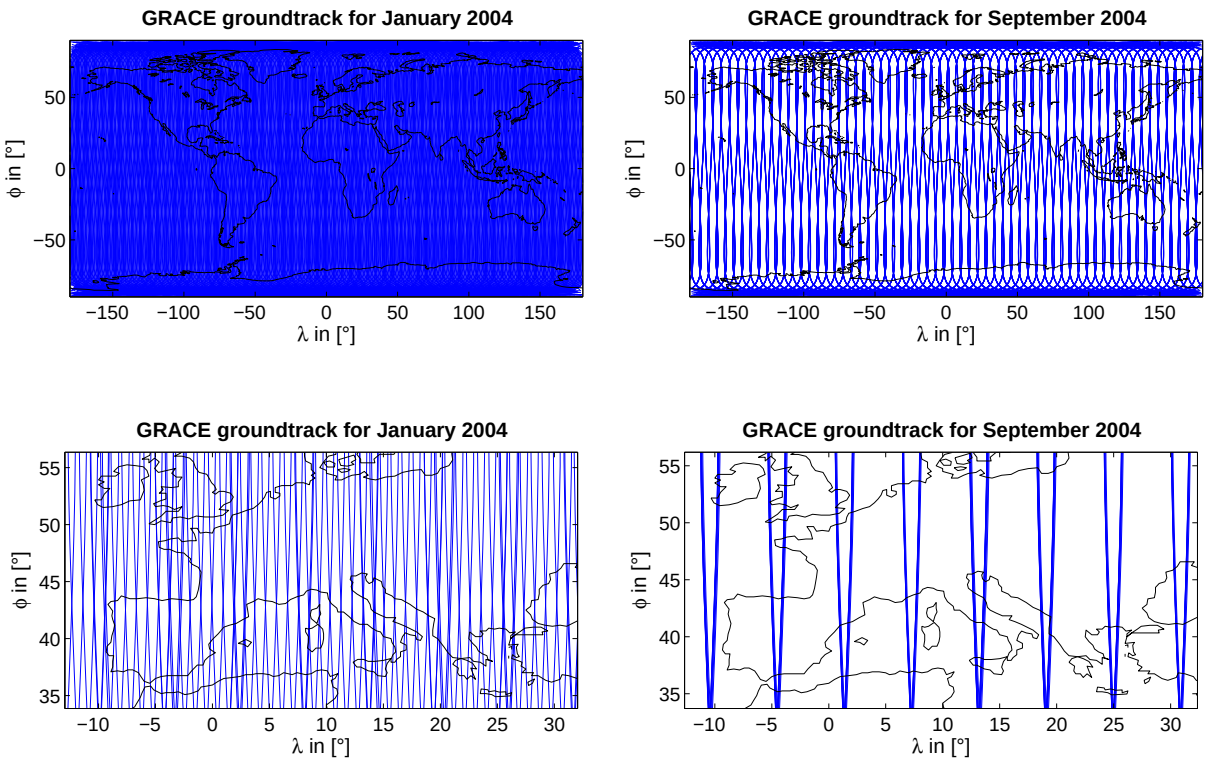


Figure 4.1: GRACE ground track examples, global view and zoomed in to a region

September a very different pattern can be seen, where the ground tracks are no more evenly distributed over the globe. In the global plot for January it is hard to see the single lines of the ground tracks, they are very close to each other and cover the whole globe. In September there are still white spaces in the map where there are no ground tracks. Coverage is much worse.

Generally, for a repeat orbit, that produces a recurring ground track pattern, the ratio between number of revolutions β that the satellite fulfills in a number of days α has to be an integer ratio. In the case of GRACE for September 2004 this ratio is

$$\frac{\beta}{\alpha} = \frac{61}{4}. \tag{4.1}$$

The reason for the different pattern in September 2004 is the repeat orbit of the GRACE satellites in this month. This means the two satellites would only repeat their ground tracks, the measurements in these months are not distributed over the whole globe, but restricted to these tracks. The altitude of the two satellites decays in time, and the orbital parameters change accordingly. In September 2004 GRACE got in the situation of a repeat orbit due to that decay. Even the gravity solutions in the two preceding and following months suffer

from bad spectral resolution. And this has severe impact on the measurements. The configuration of GRACE allows only along track measurements, a fact that leads to unphysical North-South stripes in the monthly maps, already for months that have a normal ground track pattern. This situation is even worsened in the repeat orbit situation. In the data post-processing, when global maps are estimated from the measurements, large areas have to be estimated from measurements concentrated to small tracks. Of course, the errors in the areas where there are no measurements would increase significantly.

Regularized Solutions

Instead of the bad monthly solutions caused by the repeat orbit problem, for the months of July 2004 to October 2004 so-called regularized solutions are provided.
In general, a solution, as for example the monthly gravity field solutions from GRACE measurements, can be unstable due to numerical singularities or numerical irregularities. In the case of the GRACE repeat orbit months these numerical singularities are due to the measurements being concentrated along the repeated ground tracks while numerical solutions are needed and calculated in a regular grid spread all over the globe. To stabilize such solutions, that means in the case of GRACE, make the values in between the repeated tracks more reliable, regularization methods are used. Different regularization methods are available, see for example (Menke, 1989) and (Rauhut, 1992). A problem that can only be solved by a regularization method is called an improperly posed problem. And in general, regularization means that additional information is needed to get a stable numerical solution of the posed problem.

For the following months regularized solutions are used:

- July 2004
- August 2004
- September 2004
- October 2004
- December 2006

The improperly-posedness for the solutions of the months July to October 2004 was already explained in detail in this chapter with the repeat orbit problem. But for December 2006 the necessity of regularization is due to a different cause. For this month only very bad GPS position solutions are available. Towards the end of the month, the GPS phase measurements were corrupted and only less precise GPS range measurements could be used during that time, what worsened the GPS solutions significantly. For calculating reasonable gravity field solutions for this month regularization was used. For details please refer to (Flechtner et al., 2007).

4.3 Reorganization of the Data Matrix

In Chapter 2.3 it was mentioned that the required input for EOF analysis is a two-dimensional data matrix with the spatial dimension along the columns and the time dimension along the rows. It should have to following form, see also (2.3)

$$\mathbf{D} = \begin{array}{c} \\ \downarrow \\ t \end{array} \begin{array}{c} \begin{array}{cccc} & \rightarrow & & x \end{array} \\ \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1p} \\ d_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ d_{n1} & \cdots & \cdots & d_{np} \end{bmatrix} \end{array} \quad (n \times p)$$

Where n is the number of observations and p is the number of locations. Or, in other words, each row is a map for time t and each column is a time series for location x, with $t = 1, ..., n$ and $x = 1, ..., p$.

However, the data set with the monthly synthesized gravity changes is organized in a three-dimensional matrix. This matrix has a two-dimensional matrix of size (180×360) representing a global grid of points for every point of time, with time running along the third dimension. So these map grid have to be rearranged to vectors, to obtain one vector of locations for each point of time. The question arises how to order the grid points in the vector. Of course, after computation of the EOFs, PCs and singular values, the vectors

that are in the EOFs represent the locations which were originally contained in the grids. For visualizing the results of EOF analysis these vectors have to be rearranged again back to the two-dimensional grids. The question if the map grids should be put into the vector column-wise or row-wise is up to the user, the results of EOF analysis are invariant for the column- or row-wise organization of the map grids.
For the final computations, each grid was reorganized column-wise into a row vector of the new data matrix. Since data are available in (180×360) grids for 64 months, the data matrix has the following size: $(64 \times (180 \times 360)) = (64 \times 64800)$. Or $t = 1, ..., 64$ and $x = 1, ..., 64800$.

In chapter 2 it has been pointed out that a centered data set is one of the preliminaries for EOF analysis. The monthly maps from GRACE are synthesized as monthly maps of gravity changes. Thus, this important preliminary is already fulfilled by the GRACE data set.

Chapter 5

Empirical Orthogonal Function Analysis of GRACE Data

In this chapter the analysis of the GRACE data set is performed. The EOF decomposition and the selection of significant modes for synthesis, that has been derived theoretically in the first two chapters, are applied now to the GRACE data set which was described in chapter 4. The results of EOF analysis on the global data set are described and discussed, as well as the performance of different selection methods for the modes. In the first subchapter, the necessary Gaussian pre-smoothing is explained, and in the following chapter the results of EOF analysis are presented. In the third subchapter modes are selected according to different methods for the synthesis presented in the last subchapter.

The MATLAB functions that have been implemented for the different steps in EOF analysis are described in detail with respect to the presented results in *italic* text font in the chapters. The reader who is not interested in the MATLAB implementation shall skip these short paragraphs.

5.1 Gaussian Smoothing

The instruments on the GRACE satellite observe time varying gravity which is the sum of all mass variations within the Earth system. This makes the separation of contributions of different sources difficult. Additionally, the GRACE gravity field solutions are subject to errors that vary from month to month because of the combined action of many factors including satellite ground track and temporal coverage, uncertainties of dealiasing models of short-term

mass variations and tides, space craft maneuvers, and data processing strategies (Rangelova et al., 2007). Thus, the direct interpretation of the observed gravity changes is complicated and therefore requires pre-processing, such as filtering, in order to smooth the random errors that are present in the short-wavelength spherical harmonic coefficients (Rangelova, 2007) and (Rangelova et al., 2007). Without any pre-smoothing, the technique of EOF analysis would not be able to deliver reasonable solutions. If the data are still too noisy, no signal can be detected using EOF analysis, since a signal can only be detected if it has a characteristic variance structure that can be identified in the overall variance. Gaussian smoothing with different radii is performed before analyzing the data with EOF analysis. The performance of EOF analysis after Gaussian smoothing with different radii shall be investigated. Since the GRACE mission only delivers along-track measurements, a typical dominant feature in the resulting monthly maps of gravity changes are artifacts in form of unphysical longitudinal stripes, sometimes called correlated errors. These stripes are non-isotropic, while the Gaussian smoothing function is an isotropic filter. To reduce non-isotropic errors by an isotropic smoothing function, large radii would be necessary.

Of course, there are other filtering techniques that have been developed especially to filter out the stripes, for example (Svenson and Wahr, 2006) as one of the most famous. But the aim in this chapter is to investigate the potential of EOF analysis to filter out the stripes. Unfortunately, the method of EOF analysis is not capable of filtering signal from noise without any previous smoothing. One of the main objectives is to find a smoothing radius where EOF analysis leads to reasonable results, and simultaneously the smoothing radius is as small as possible, as smoothing always implies a loss of signal, too. Different Gaussian smoothing radii are tested.

Before the data matrix as input for the EOF analysis can be set up from the monthly maps of gravity changes, the gravity fields have to be synthesized from the given spherical harmonic coefficients. Gaussian smoothing is a pre-processing step, it is performed on the coefficients, before synthesizing the gravity fields in equivalent water height (EWH). It has been shown, that the errors of unphysical longitudinal striping are due to an increasing correlated error spectrum at higher degrees in the Stokes coefficients C_{lm} and S_{lm}. The striping is suppressed by weighting the Stokes coefficients by means of a Gaussian smoothing function,

which decreases in value with an increasing degree and thus attenuates the higher degree coefficients that are responsible for the striping. See for further information (Wahr et al., 1998) and (Jekeli, 1981).

5.2 Empirical Orthogonal Function Analysis

The rearranged data matrix $\mathbf{D}$ is of the size (64×64800), see chapter 4.3, and is decomposed in the analysis-step into

$$\mathbf{D} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \tag{5.1}$$

where $\mathbf{U}$ is of size (64×64) and contains the principal components (PCs) as column vectors, $\mathbf{\Sigma}$ is a diagonal matrix of size (64×64) and contains the singular values on the main diagonal, and $\mathbf{V}$ has the size (64800×64) with the empirical orthogonal functions (EOFs) as column vectors. In the data set GRACE gravity changes are used. The data is already centered, which is one of the requirements for EOF analysis. However, the temporal mean was removed column-wise in the data matrix before Gaussian smoothing. To allow EOF analysis for the best possible results, immediately before performing EOF analysis the mean is again removed in MATLAB to take care of the changes imposed by the smoothing function. Even if only very small values are removed, this step is important since EOF analysis relies on the variances. In the data matrix from the GRACE data, the time dimension is much smaller than the spatial dimension (time dimension is 64 and spatial dimension is 64800). By removing the temporal mean, the data matrix get one rank defect. The number of non-zero singular values equals the rank of the data matrix. Thus, the last singular values is zero due to the removal of the temporal mean. See the singular values plot in figure 5.8.

To decompose the two-dimensional data matrix according to the singular value decomposition in equation (5.1), MATLAB provides a tool called svd. *Using a call to the function in the form of* $[\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}] = \text{svd}(\mathbf{D})$ *gives exactly the same components in equation (5.1), these are the EOFs, the PCs, and the singular values. Since calculation all of the modes can be a bit time consuming, and often only the first few modes are needed, in the case of GRACE data only the temporal scatter matrix is calculated in the first step. This matrix is the smaller one of the two possible scatter matrices for the GRACE data set. Due to the size of the data*

matrix, $\mathbf{D}$ (64×64800)*, the size of the temporal scatter matrix is* $\mathbf{R_t}$ (64×64) *while the size of the spatial scatter matrix is* $\mathbf{R_s}$ (64800×64800)*. Both scatter matrices have the same rank and the same eigenvalues, as proven in chapter 2.3. The singular values are the square roots of the eigenvectors of the scatter matrix. In chapter 3.2 ad-hoc rules of thumb based on singular values for the selection of modes are presented. A plot of the singular values can help to decide on the number of modes to be calculated in the EOF analysis. Of course, a more specific selection of modes can be performed afterwards, but it speeds up calculation times significantly if a certain number of modes can be excluded before even calculating them. For only calculation a certain number of singular values, MATLAB provides a function called* svds*, a singular value decomposition, but only for a certain number of modes.*

The singular values are a measure of the amount of variance that is explained by the corresponding mode. The PCs are plotted as time series while the EOF vectors are rearranged into two-dimensional grids of patterns and plotted as maps, see explanations in chapter 4.3. Since EOF and PC vectors are both orthonormal, the scaling that leads back to values of Equivalent Water Height (EWH), in which the data sets are given, is contained in the singular values. In the following figures 5.1 to 5.7, the first few modes from EOF analysis that was performed on data pre-filtered with a Gaussian smoothing filter of different radii, are shown. Each row of images contains one mode's EOF, its singular value, and its PC, each triplets is considered to be one mode.

Unfortunately, in the first modes for smoothing radius 200 km no signal can be found in the EOF patterns and the PC time series. All there is to be seen in the plots are the characteristic longitudinal stripes from GRACE. The same happens for smoothing radius 250 km. It is impossible to identify anything else than longitudinal stripes, resulting from GRACE correlated errors, in the EOF patterns, and noise in the PC time series. The signal magnitude of noise and errors is still too large compared to the signal magnitude of the signal we want to identify, the variances of noise and errors still exceed the variance structure of the signal. Thus, EOF analysis is not capable of identifying the signal in such a bad signal-to-noise ratio. Higher smoothing radii smooth out more of the correlated errors and noise, and yield better results.

In the plots of the first few modes from radius 300 km, figure 5.3, interpretable patterns

EOFs and PCs for smoothing radius 200 km

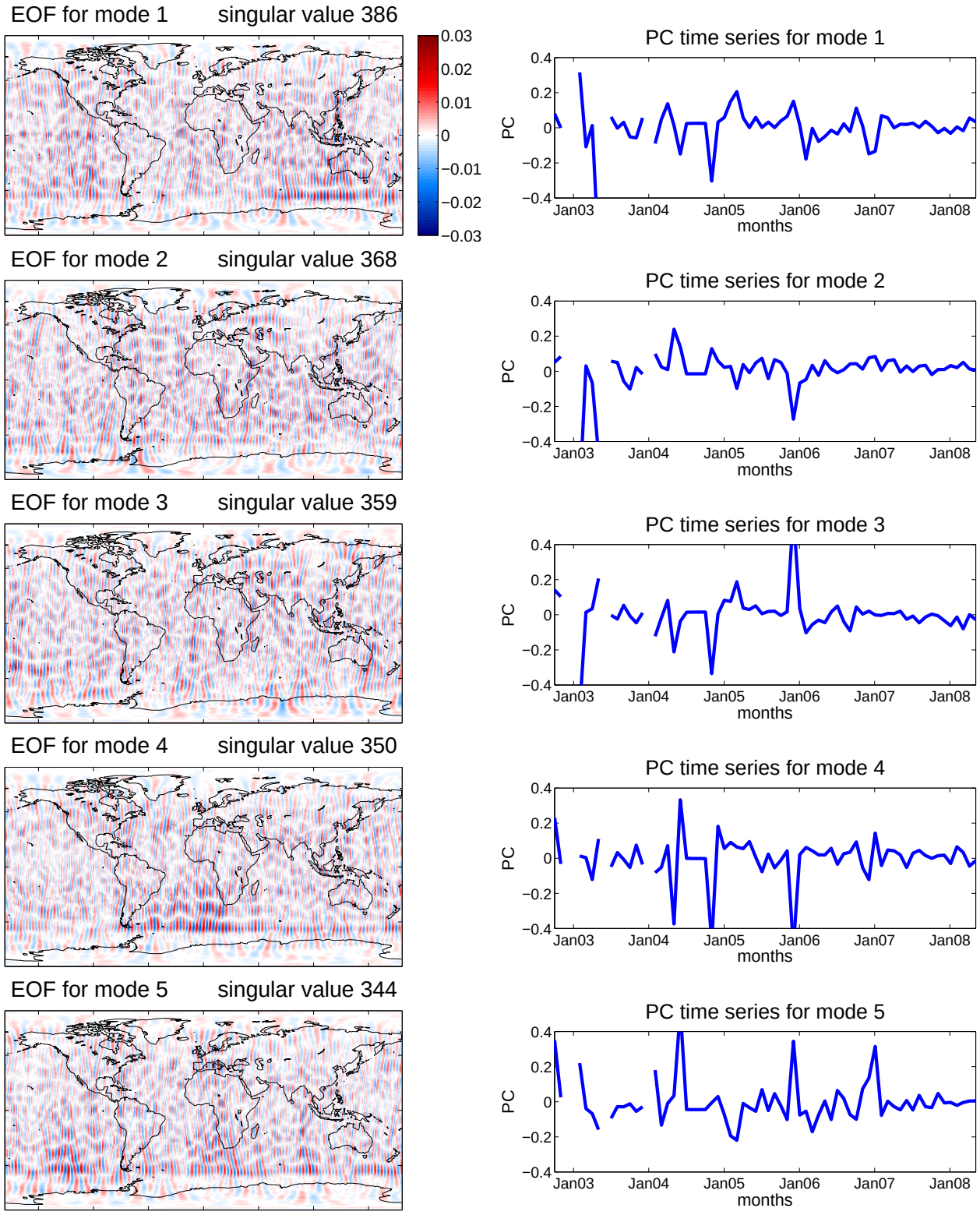


Figure 5.1: EOF patterns and PC time series for the first modes with a Gaussian smoothing radius of 200 km

EOFs and PCs for smoothing radius 250 km

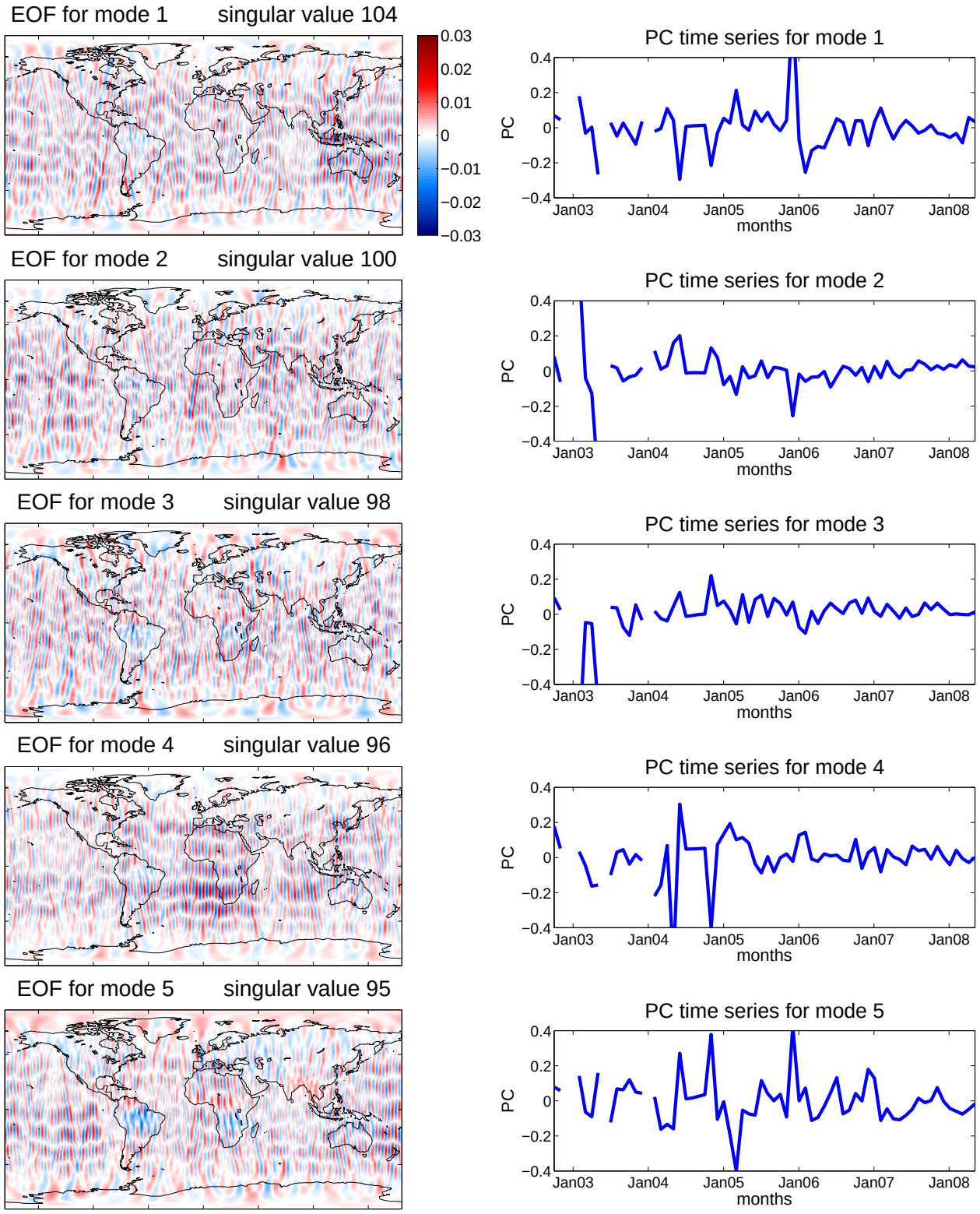


Figure 5.2: EOF patterns and PC time series for the first modes with a Gaussian smoothing radius of 250 km

EOFs and PCs for smoothing radius 300 km

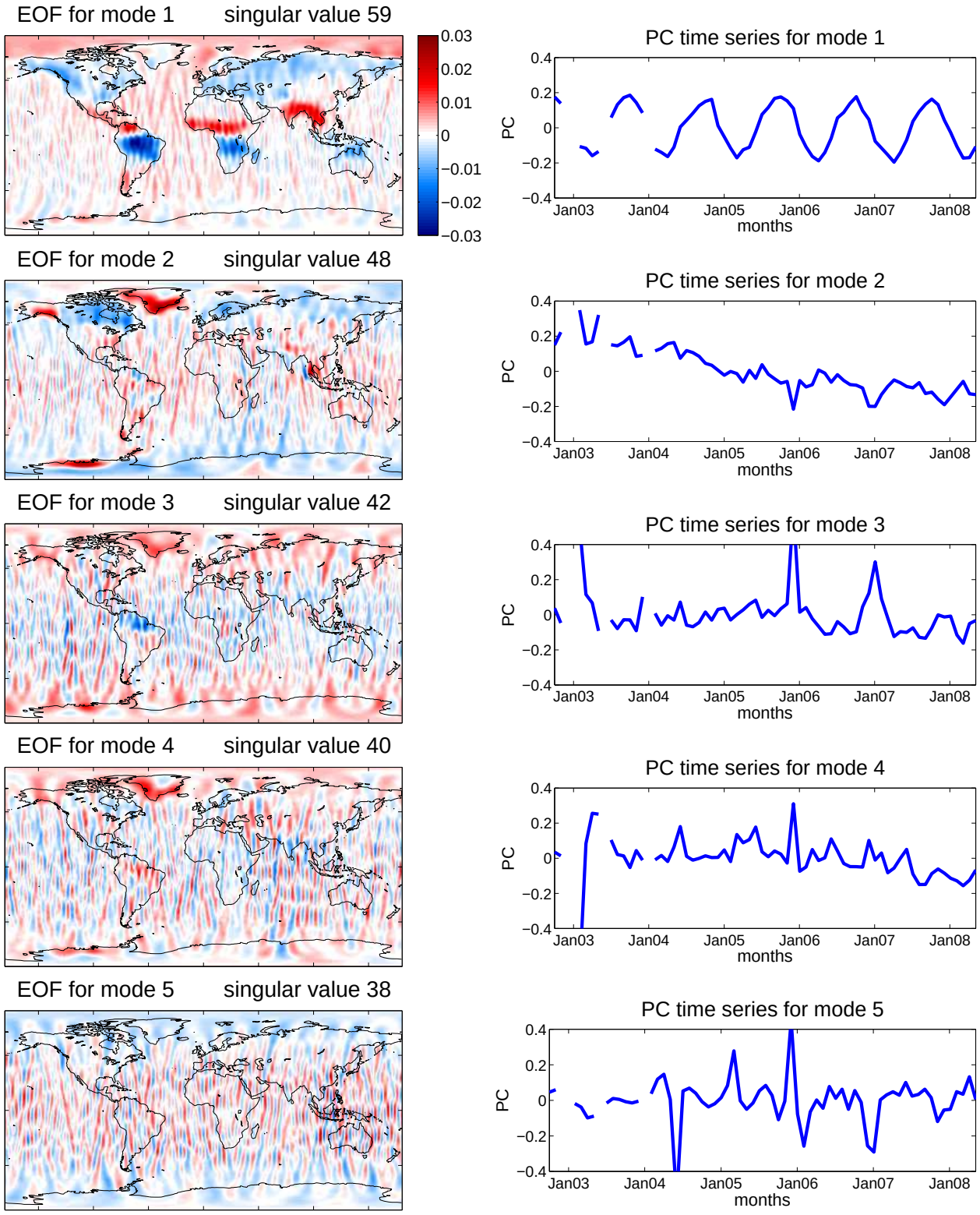


Figure 5.3: EOF patterns and PC time series for the first modes with a Gaussian smoothing radius of 300 km

EOFs and PCs for smoothing radius 350 km

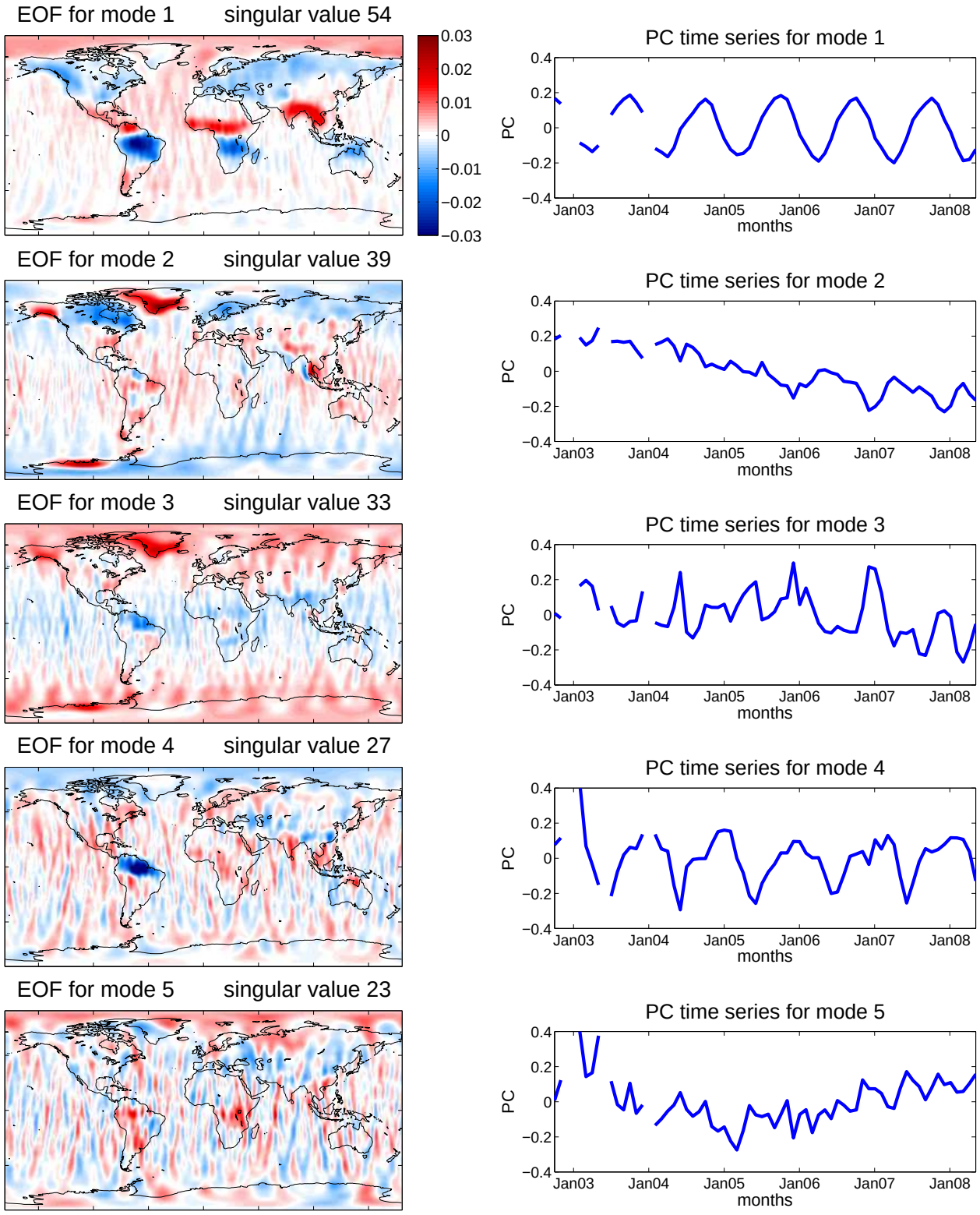


Figure 5.4: EOF patterns and PC time series for the first modes with a Gaussian smoothing radius of 350 km

EOFs and PCs for smoothing radius 400 km

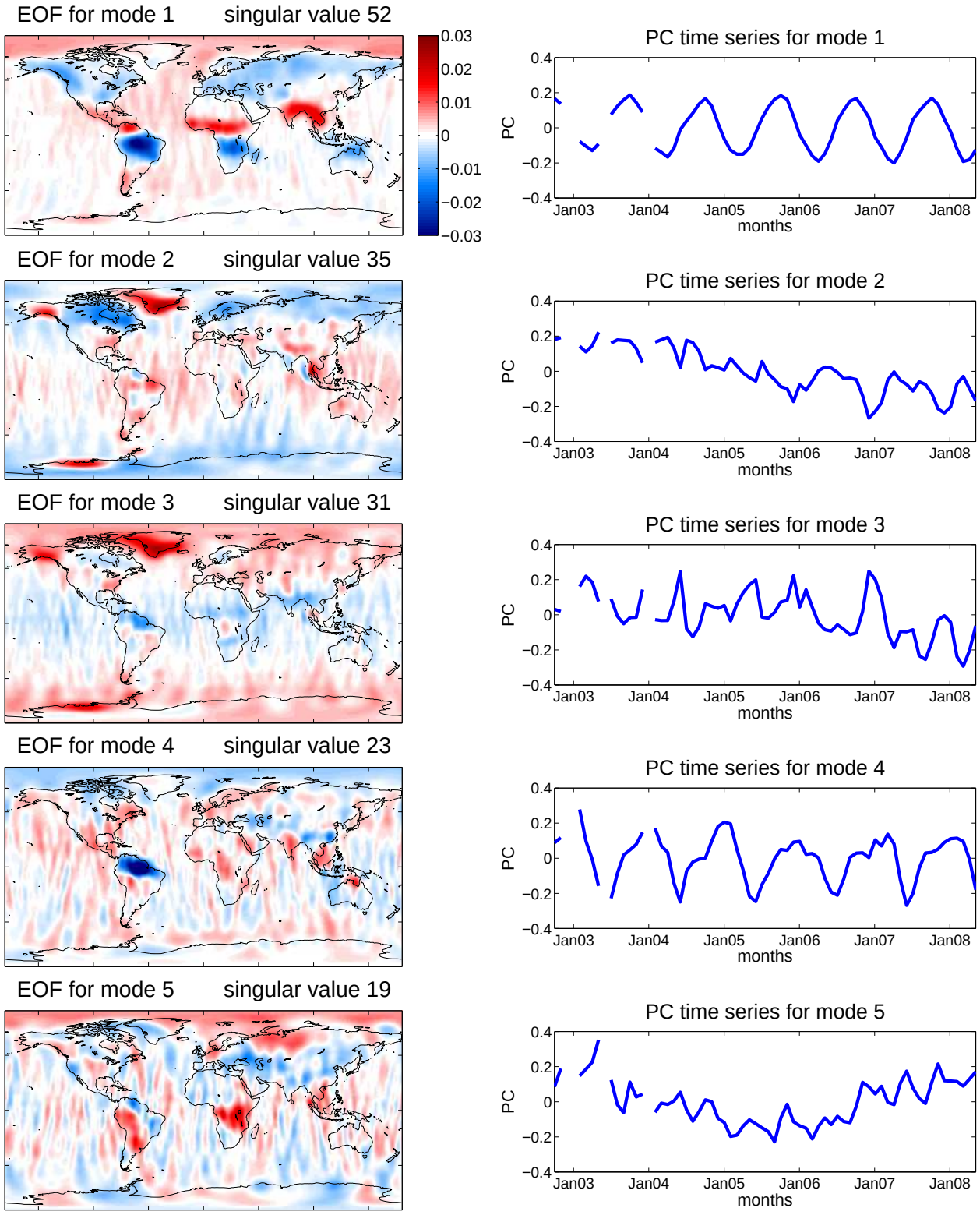


Figure 5.5: EOF patterns and PC time series for the first modes with a Gaussian smoothing radius of 400 km

EOFs and PCs for smoothing radius 450 km

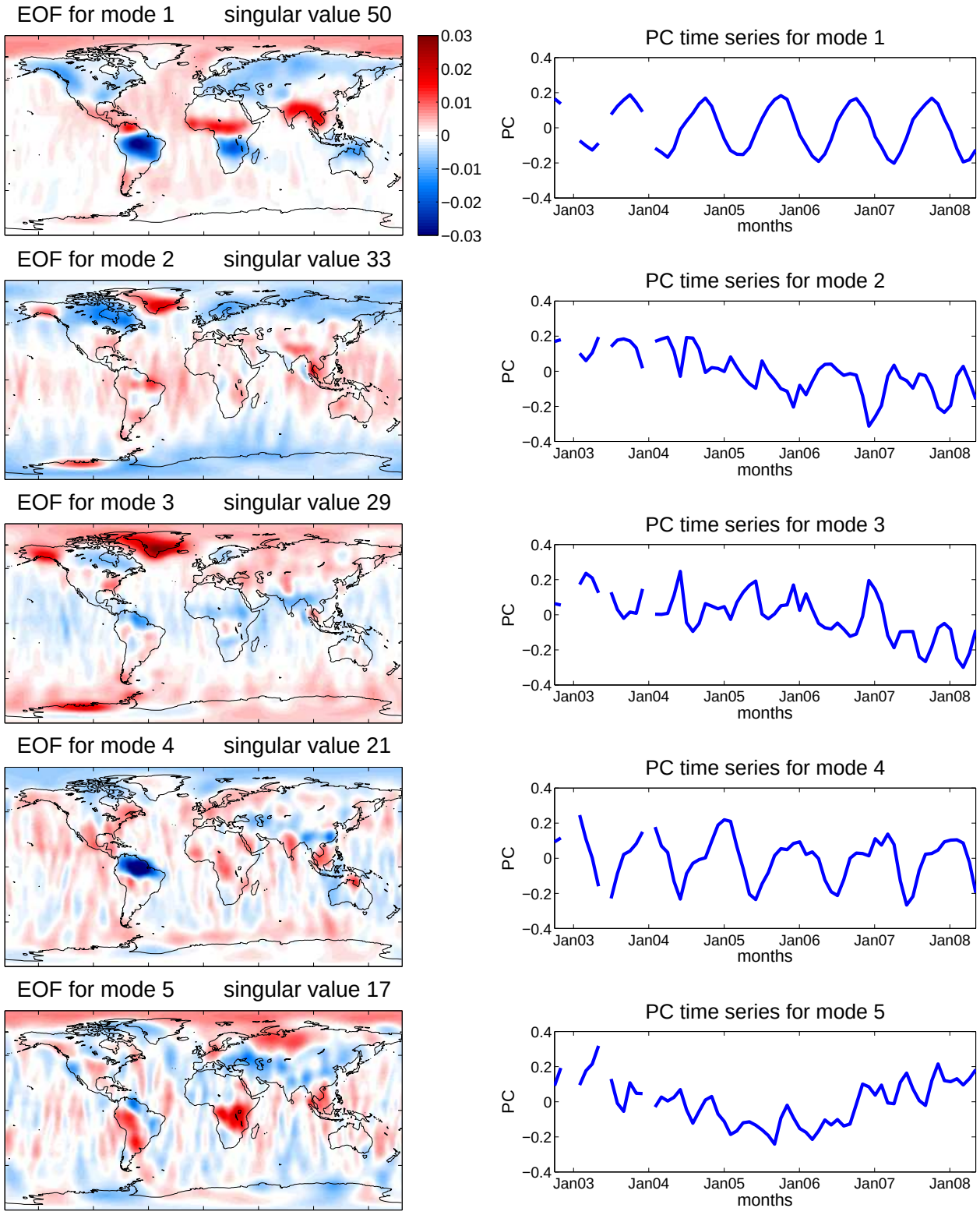


Figure 5.6: EOF patterns and PC time series for the first modes with a Gaussian smoothing radius of 450 km

EOFs and PCs for smoothing radius 500 km

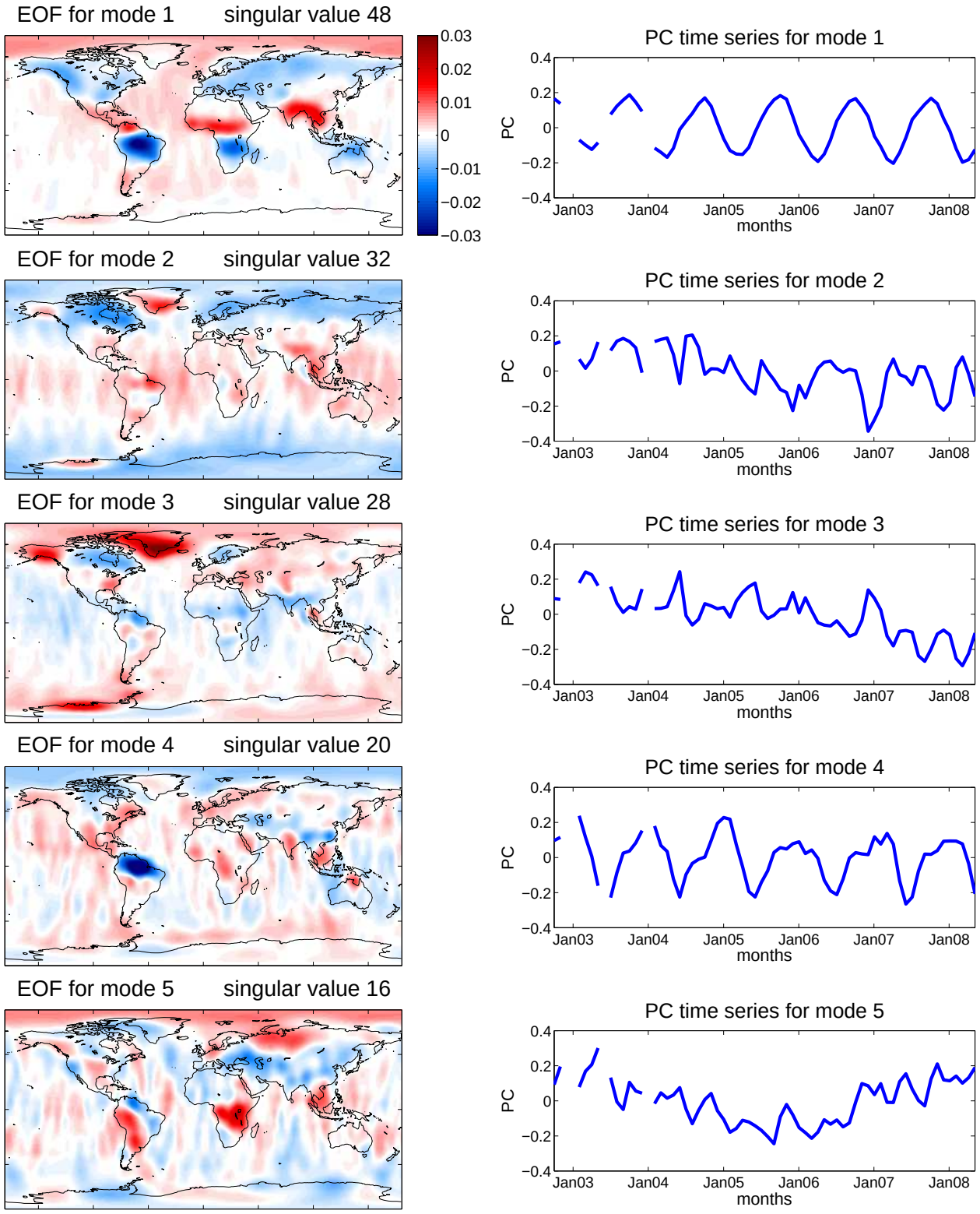


Figure 5.7: EOF patterns and PC time series for the first modes with a Gaussian smoothing radius of 500 km

can be found. There is an annual signal in the first mode: In the PC time series is an oscillation with an annual period and in the EOF pattern the corresponding regions are the tropics. The border between the positive and negative areas on the continents in the EOF pattern is situated along the equator, and since the positive areas in the pattern are North of the equator, and the negative patterns are South of it, this means that there is a mass change going on just North of the equator with the given periodicity and there is a mass change going on just South of the equator with an oscillation anti-cyclic to that one given by the time series. This signal of mass change is due to the annual change of rainfall in the tropics. It can also be interpreted as the annual shift of the area with the major rain falls due to the changing seasons and shift of the inner tropic weather zone. Since it has the largest singular value, this signal is the one expressing the most characteristic structure in variance in the data set, and therefore the strongest signal in the GRACE data set.

The second mode's PC time series is a bit noisy at the first sight, but a trend can be identified. The EOF pattern shows us where this trend is going on: The major areas are Greenland, Alaska, and West Antarctica. Since the trend in the time series is a downward trend and the patterns in these named regions have a positive sign, the signal we have detected here is loss of mass in these regions. And of course this is a signal that was expected in the GRACE measurements. The loss of mass is due to glacier and ice sheet melting in Greenland, Alaska and West Antarctica. However, there are also regions in this mode that appear negative in the EOF patterns. The magnitude is not as strong as for the named regions where ice sheet melting is going on, as it can be seen from the coloring in the EOF pattern. Since we have a downward trend and a negative sign in these areas, that are Scandinavia and in Canada a large area around Hudson Bay, a positive mass change is detected in these areas. Of course slighter than the ice sheet melting signal, but still with the same ongoing trend. These mass changes are due to post-glacial rebound in Scandinavia and North America. For a correct interpretation of the EOFs and the PCs it is important to keep in mind that both the EOFs and PCs come from normalized vectors and they lead back to the gravity signal by multiplying together the EOFs, PCs and the singular values. Therefore, the sign of the EOFs and PCs can be chosen arbitrary as long as the sign of the other part of the decomposition, that is the PC or the EOF, is chosen alike so that it leads to the same result after multiplication. The singular values are always positive and therefore do not need to be included

in discussions about how to choose the sign. For the plots the convention was chosen that the first element of each PC is positive. Finally, it is remarkable that in this second mode of smoothing radius 300 km in figure 5.3 shows the signature of the Sumatra earthquake from December 2004 in the EOF. The red and blue patterns can be seen at the west coast of Sumatra. The mass changes that caused the earthquake have been that strong that they can be detected by GRACE. The positive and negative patterns can even be interpreted tectonically.

In the third mode, there is in the EOF pattern signal visible again in Greenland and in Northern Brazil, but these patterns are very slight compared to the surrounding noise patterns and in the PC time series there is no signal traceable. Thus, this mode should be considered as noise. There is no signal part to identify, as there was in the first few modes. The same holds for the fourth mode, even if there is again a pattern visible in Greenland. Modes three and four are two of the modes that are hard to decide on, whether to consider them as signal, even is there is no real signal, or consider them as noise, even if there is a slight pattern visible. Finally, mode five is definitely considered as noise, as neither in the EOF patterns nor in the PC time series there is any signal traceable.

Consider now the modes from smoothing radius 350 km in figure 5.4. The signal of the first modes resembles that one that was already found in the first mode for smoothing radius 300 km. For an explanation of this signal please refer to mode one for 300 km smoothing radius (figure 5.3). As well for mode two, the signal is the same as the one already described for mode two and 300 km smoothing radius. However, mode three becomes more distinguished: Even if the time series are very noisy, and there is only a very slight trend to be interpreted, the patterns show a signal in Greenland, Antarctica, Alaska and the Arctic. In comparison to the preceding mode, these patterns cover Greenland, Alaska and West Antarctica as well, but they do not show the post-glacial rebound at all. And there seems to be also a slight signal in the polar regions. The time series of mode three are too noisy to see the actual difference to the trend in mode two, it is hard to find a physical meaning that could have caused the signal in this mode. It is a signal of a mass change trend that comes from the polar regions. It could be due to changes in the ice masses caused by temperature changes. By definition of the EOF analysis, it has to be independent from and orthogonal to the other modes. So we do not have the same long term trend here as in mode two. There must be

some other temporal structure underlying this signal.

Moving on to mode four, another interesting pattern is found: The time series show an annual pattern with minima in May and maxima in December and January. In the EOF pattern there is one distinguished area: The Amazon basin. So the time series give the changes in water level of the Amazon river with the changes of the season. Please note that these annual patterns are very different from those in mode one. First, the time series have a very different characteristic, with distinct peaks for the minima instead of being a sine-like function. And second, the pattern shown in the EOF is a bit more North than the patterns in South America in mode one. Additionally, the pattern in the Amazon basin is negative. Multiplied together with the negative peaks in the time series, these two negative components lead to a positive signal, that marks the peaks of water drainage of the Amazon. It might be confusing for interpretation that both interesting features have a negative sign. This is just due to the convention of the first element of the PC to be chosen positive. Signs for the EOFs and PCs can be, as already mentioned, chosen arbitrarily, besides the chosen convention. It only matters that they give the correct sign after multiplication. Finally, a completely different signal in mode four from mode one is a preliminary of EOF analysis, as already said, the EOFs of the different modes as well as the PCs are orthogonal to each other.

Mode five is hard to interpret. In the EOF pattern there is a slight signal in Africa in the region of Kilimanjaro and Lake Victoria. The time series are also very noisy, but show a downward trend for the first two and a half years, until March 2005, followed by an upward trend until May 2008. Finding a physical explanation of that signal, in case it is really a signal, is very hard. Also, since this signal is not very distinguished.

In figure 5.5 for smoothing radius 400 km, the signals that can be found in the modes one to five are very similar to those that we have found in the modes from smoothing radius 350 km in figure 5.4. The EOF patterns as well as the PC time series are just a bit less noisy than for smoothing radius 350 km. For explanations of each mode please refer to the explanations given for smoothing radius 350 km. With a larger smoothing radius more of the correlated errors, the longitudinal stripes, are smoothed out and thus there are less stripes in the data set which makes it easier for EOF analysis to identify clear signals since there is

less contamination with errors and noise.

For smoothing radius 450 km in figure 5.5, the signal that there is to interpret is still the same and has already been interpreted in the explanations for smoothing radius 350 km. Still, the only difference is that more smoothing leads to less stripy patterns. While there is not much change in the time series, the larger smoothing radius can be clearly seen in its impact in the EOF patterns. The patterns are much clearer and a lot of striping has been eliminated compared to the shorter smoothing radii.

Finally, in figure 5.7 for smoothing radius 500 km, the first mode is still the same as expected, showing an annual signal as explained before for the first mode. But in the modes two and three a change from the preceding smoothing radii is found. While the time series of mode two is getting noisier, the trend that has been found before in mode two seems to be divided into individual parts. The downwards trend in mode three is more distinct than for the previous smoothing radii. Concerning the EOF patterns, there is almost no signal any more in West Antarctica and Alaska in mode two. But there is still a signal in Greenland, however, not as strong as it has been for the previous smoothing radii. The signal in North America and in Scandinavia from the post-glacial rebound is still present. In mode three the signal in Greenland is very strong, as well as the signal in West Antarctica and Alaska. It seems like the larger smoothing radius lead to these two signals being clearer distinguished into two different modes. Mode four shows again the expected patterns, there is the annual signal in the time series with the pattern in the Amazon as it was explained before. Mode five is also very similar to the fifth mode in smaller smoothing radii, there is a signal in Africa that is hard to interpret.

In conclusion, in the different smoothing radii, figures 5.1 to 5.7, it can be seen, that the most interesting modes are those for smoothing radii 300 km and 350 km. For smoothing radius 300 km the first detectable signal appears, and for smoothing radius 350 km more smoothing reveals more different signals. In the results of 400 km smoothing the different signals in each mode are a bit clearer, since they are no more as much contaminated by the correlated errors as for smoothing radius 350 km, where the EOF patterns are still very stripy. Larger smoothing radii do not reveal additional signals.

5.3 Selection of Modes

In the preceding subchapter 5.2, the modes resulting from EOF analysis after Gaussian pre-smoothing with different radii have been shown. Naturally, the question arises which modes are containing signal and which ones only noise, and it has been discussed already in the interpretation for the plots. But for a synthesis step a selection method different from interpretation and visual judgement is desirable. Out of the numerous methods possible for selecting an optimal subset of modes for a reconstruction of the data with the aim not to loose signal, and to filter out as much of the noise as possible, two different techniques are applied to the results of EOF Analysis on the GRACE data set. The theoretical background of the methods that are applied has already been given in chapters 3.2 and 3.3. First, the ad-hoc rules of thumb based on the singular values shall be applied on the results for different smoothing radii, and second, a KS hypothesis test is performed for the EOF analysis results.

Singular Values

The first method that shall be applied is based on the singular values. The goal is to find a cut-off value in the singular values of each smoothing radius so that all modes before the cut-off mode are considered to be signal and all higher modes are considered to be noise. Depending on the singular values different techniques to find the cut-off value can be useful. To show some of the techniques and discuss their performance for the GRACE data set, the singular values are given in different representations: Each first eight singular values for smoothing radii from 200 km to 500 km are given in the following table 5.1, and in figure 5.8 all singular values for these smoothing radii are given for graphical interpretation. One of the two plots represents the singular values while the second one gives the normalized singular values. For some applications a logarithmic plot of the singular values can be useful, however, in the case of GRACE, the last four singular values are very small compared to the other ones, so a logarithmic plot mainly emphasizes the differences in magnitude between the last few values and the majority of their preceding values. But the more interesting part, the differences between the first few values, about up to mode fifteen or twenty, is not emphasized in a logarithmic plot. It would be even worse than in the diagram in figure 5.8.

Singular values for different Gaussian smoothing radii

smoothing	singular values							
radii	mode 1	mode 2	mode 3	mode 4	mode 5	mode 6	mode 7	mode 8
200 km	386	368	359	350	344	322	316	315
250 km	104	100	98	96	95	92	88	85
300 km	59	48	42	40	38	37	35	35
350 km	54	39	33	27	23	21	21	20
400 km	52	35	31	23	19	16	15	14
450 km	50	33	29	21	17	14	12	11
500 km	48	32	28	20	16	12	11	10

Table 5.1: First eight singular values for different smoothing radii

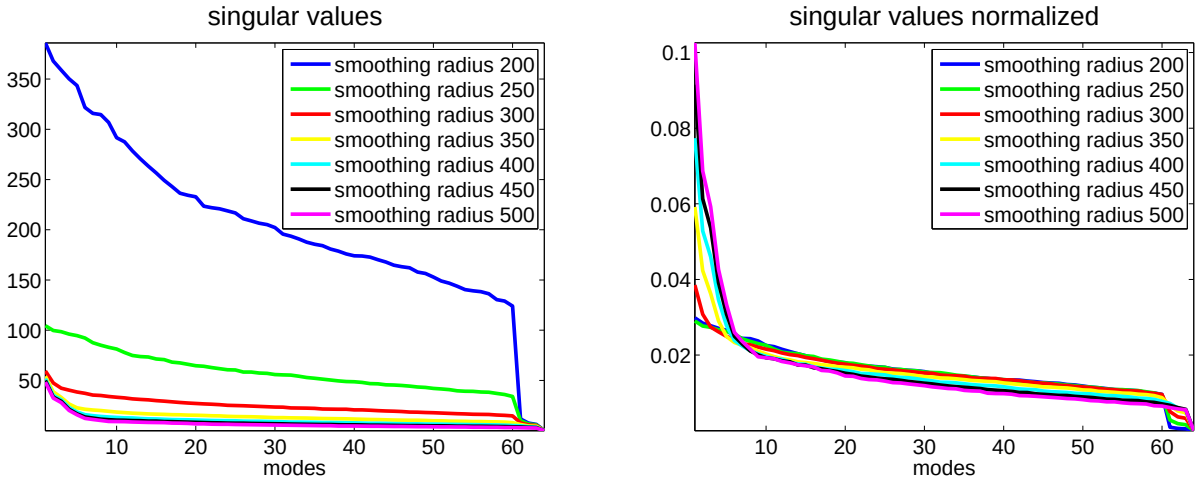


Figure 5.8: Singular values for different smoothing radii

First of all, in the singular values in table 5.1, it is remarkable that the higher the smoothing radius the smaller the singular values. It has already been said that since the EOFs and PCs are normalized vectors, the only information about magnitude is contained in the singular values. And it has also been explained that the singular values are a measure of the amount of variance expressed by each mode. With larger smoothing radii more of the noise

and correlated errors are smoothed out and the total variance is diminished, by the amount of the smoothed out part of the noise and errors. So if the total variance is reduced by the portion of the smoothed out noise and errors, the share of the signal variance part of the total variance is increased. Since in the procedure of EOF analysis the new basis, the EOFs, are chosen according to the main variances, the variances of the signal to be detected need to have a certain amount of total variance to be found by the EOF analysis method. This is shown in the plot of the first few modes for each smoothing radius in the preceding subchapter 5.2. For smoothing radii 200 km and 250 km the major part of variance comes from noise and errors, which are plotted in these first modes. The signal variance is too small compared to the noise variance, and thus covered by noise variance. Only when enough of the noise variance is smoothed out, the relative share of the signal increases and the signal can be detected.

In table 5.1 the relative differences between the singular values of each smoothing radius reveal information. For 200 km smoothing no significant drop in magnitude can be found, the difference between the first two singular values is 18, the following three differences are only 9 and 6, while the difference between the fifth and sixth singular value is again as high as 22. For smoothing radius 250 km the differences are all in the same range of magnitude, between 1 and 4. In contrast, for smoothing radius 300 km, the differences between the first three singular values are 9 and 6, respectively, while all the following differences between singular values following each other are no larger than 2. For 350 km smoothing, a similar change can be detected, starting from mode six the difference is no larger than 1. The singular values for smoothing radius 400 km are very similar to those of smoothing radius 350 km, starting from mode six, the difference is no larger than 1, too. Smoothing radii 450 km and 500 km are again very similar to each other, however, the singular values being similar to each other start at higher modes. It is hard to tell from the values given in table 5.1, but from the values given it could be about mode seven, that all the higher modes can be considered as noise.

The singular value plots in figure 5.8 should reveal more help. In the non-normalized plot we can see again that the singular values in general decrease with increasing smoothing radii. There are extreme decreases in the magnitudes from smoothing radius 200 km to smoothing radius 350 km. The reason is the same as stated already above for the interpretation of the values in table 5.1. While the noise variance greatly exceeds the signal variance, smooth-

ing significantly reduces the major share of the variance, that is the noise. For radii where signal is already the main pattern in the variance, more Gaussian smoothing, as explained in chapter 5.1, attenuates the contribution of higher degree coefficients and therefore only the noise parts. Of course, there is always some signal reduced, too. But since the main variance characteristics are now due to signal that is more concentrated to the lower degree coefficients, the major variance is no more changed that much.
In the diagram of the normalized singular values, the plot on the right in figure 5.8, the answer to the question about the behavior of the singular values of the higher smoothing radii can be found. The larger the smoothing radius, the more distinguished the drop in magnitude between the first and about eighth singular value. This drop in magnitude can be interpreted to mark the border between signal and noise-containing modes, according to (Jolliffe, 2002), since the noise-containing modes are assumed not to differ significantly in their variance, which is represented by the singular value. It is assumed that the noisy modes all have similar singular values. The mode associated to the drop is chosen as cut-off mode. All higher modes are considered as noise and not used for synthesis.

From the plot it is concluded, that for smoothing radius 500 km, the singular values are very similar to each other, starting only from mode nine. There is a distinguished sharp bend in the curve. Smoothing radius 450 km shows almost the same behavior. The smaller the smoothing radius, the less distinguished is the bend in the curve. And, the smaller the smoothing radius, the earlier the bend, that means the earlier the singular values start being very similar to each other. Down to smoothing radius 300 km this sharp bend is detectable. But for smoothing radii 250 km and 200 km, there is no more any change or drop detectable in this normalized plot of singular values. Both curves are very similar to each other and both are very smooth, there are no noticeable changes. This confirms what has been already seen in the previous chapter in the plot of the modes and in the table of singular values: For these smoothing radii EOF analysis is not capable of detecting any signal. Since the signal expressed is all noise and correlated errors, there are no significant differences between the modes.

Furthermore, it is noticeable that the last few singular values are very small compared to the others. For small smoothing radii, 200 km and 250 km, a significant drop in magnitude

appears for the modes 61 to 64. It was already stated, that removing the time mean in the data matrix prior to all computations, led to a rank deficiency in the data matrix. And the number of non-zero singular values equals the rank of the data matrix. The rank of the data matrix is 63, so there are only 63 non-zero singular values. The last singular value is zero for all smoothing radii.

Kolmogorov-Smirnov Test

The performance of a KS hypothesis test on the modes from EOF analysis of the GRACE data set shall be investigated here. The mathematical background of the KS test applied to the power spectral density of PC time series has been given in chapter 3.3. Here, the application on the results of EOF analysis for different smoothing radii is presented.

The following figure 5.9 shows the results of the KS test for two different significance levels. Comparing the two figures it can be seen that the results of the test are relatively stable under changes of significance level. The main characteristics are the same for both significance levels. The dots mark those modes, that are significantly different from white noise due to the test.

It is remarkable that no mode for smoothing radii 200 km and 250 km is significantly different from white noise in the test. This corresponds to what we have found from the singular values and in the plots of the first few modes' EOFs and PCs. Also, the hypothesis test proves our finding that EOF analysis is not capable of detecting signal in the GRACE fields that are only smoothed with a Gaussian filter of radius up to 250 km. Regarding smoothing radius 300 km, the first two modes are signal-containing modes due to the test, which matches our findings in the plots of EOFs and PCs in figure 5.4. However, mode 55 is also considered to be signal in the test and will be investigated later in detail. For smoothing radius 350 km the first six modes for significance level $\alpha = 5\%$ and the first five modes for significance level $\alpha = 2.5\%$ are considered to be signal. These findings match the results from the plot of the normalized singular values and the absolute singular values of the first modes. From table 5.1 we have concluded that the first five modes could represent signal and from the normalized plot of the singular values in figure 5.8 we have also found the cut-off value

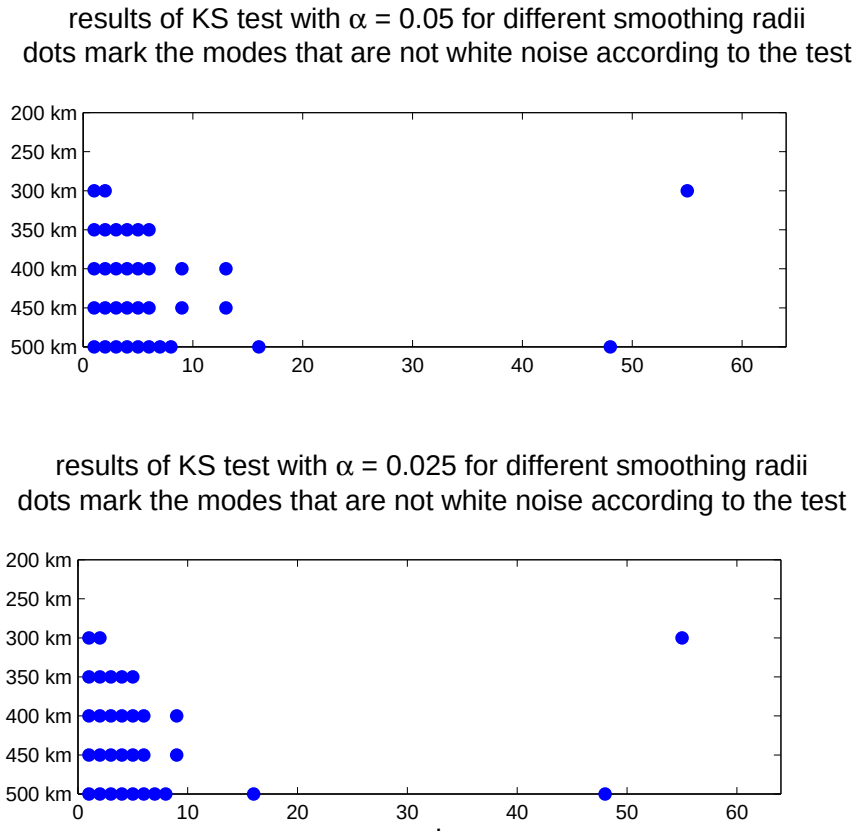


Figure 5.9: KS test results for $\alpha = 5\%$ (top) and $\alpha = 2.5\%$ (bottom)

in mode five or six. However, in the plots of the EOF and PC for each mode in chapter 5.2, it was hard to find any physical interpretation for mode five and smoothing radius 350 km, figure 5.4. Since in the selection methods it is only cared about finding structures that differ from white noise, this selection methods are helpful for finding non-noisy modes. But finding the physical impact that caused the signal that is contained in each mode, can be a very hard task.

Consider smoothing radii 400 km and 450 km in figure 5.9. The results of the KS test are the same for both smoothing radii in both significance levels. This similarity between the two smoothing radii has already been noticeable in the plots of the first EOFs and PCs in chapter 5.2, but also in the singular values. It might be interesting to have a closer look at the EOF patterns and PC time series of modes nine and thirteen, since these two modes are considered to be signal for both, smoothing radius 400 km and 450 km, while all other modes higher than mode six are considered to be noise. The first five modes' EOFs and PCs have already been plotted in chapter 5.2 please refer to figures 5.5 and 5.6, and in figures 5.10 and 5.11 modes six to fifteen for smoothing radius 400 km are plotted. The higher modes for

radius 450 km are not plotted here, since they are very similar to the modes found for radius 400 km.

Only the modes for smoothing radius 400 km are discussed as one representative example of the two very similar results from smoothing radius 400 km and 450 km. Additionally, not all the modes six to fifteen will be discussed in detail. Only those modes which are significantly different from white noise in the KS hypothesis test are investigated further. It was already hard to find a physical meaning for the signal represented by mode five, please refer to chapter 5.2. In mode six in figure 5.10, the time series shows the reason why this mode is identified to be signal in the KS test: There is a downward trend in the first two years, followed by a steeper upward trend roughly from end of 2004 to the end of 2005, which is again followed by a downward trend similar to the first one until end of 2007. Finally, the start of another steep upward trend is shown. This could be interpreted as an inter annual cycle. However, it is extremely hard to find a physical impact that could have caused this signal since there is no corresponding pattern to identify. A speculation could be that this time series pattern is caused by aliasing effects. An explanation for the significance of mode nine is hard to find in the plots of EOFs and PCs in figure 5.10. The details of the KS test need to be studied in detail for further investigation. Please refer to figure 5.15 and the discussions below.
In figure 5.11 no distinguishable pattern can be found in the EOF for mode thirteen, neither any distinct signal in the time series. Since mode thirteen was not accepted anymore as signal for the lower significance level of $\alpha = 2.5\%$, the hypothesis test was passed for $\alpha = 5\%$ most likely because of the five peak-like features in the time series. But since this mode is considered not to be signal in the hypothesis test for $\alpha = 2.5\%$ it is for sure one of the critical modes that are hard to decide about. Furthermore, according to the singular values graph, mode thirteen should not represent signal anymore, while mode nine is very close to the cut-off value and is therefore worth further investigation. In combination of all the insights we have from the two hypothesis tests, the singular values and the plots of the EOFs and PCs, mode thirteen can be considered as noise and therefore no further investigation is done for this mode.

For smoothing radius 500 km, the results from the hypothesis test with different significance levels are both the same. Modes one to eight are identified as being signal, additionally

EOFs and PCs for smoothing radius 400 km

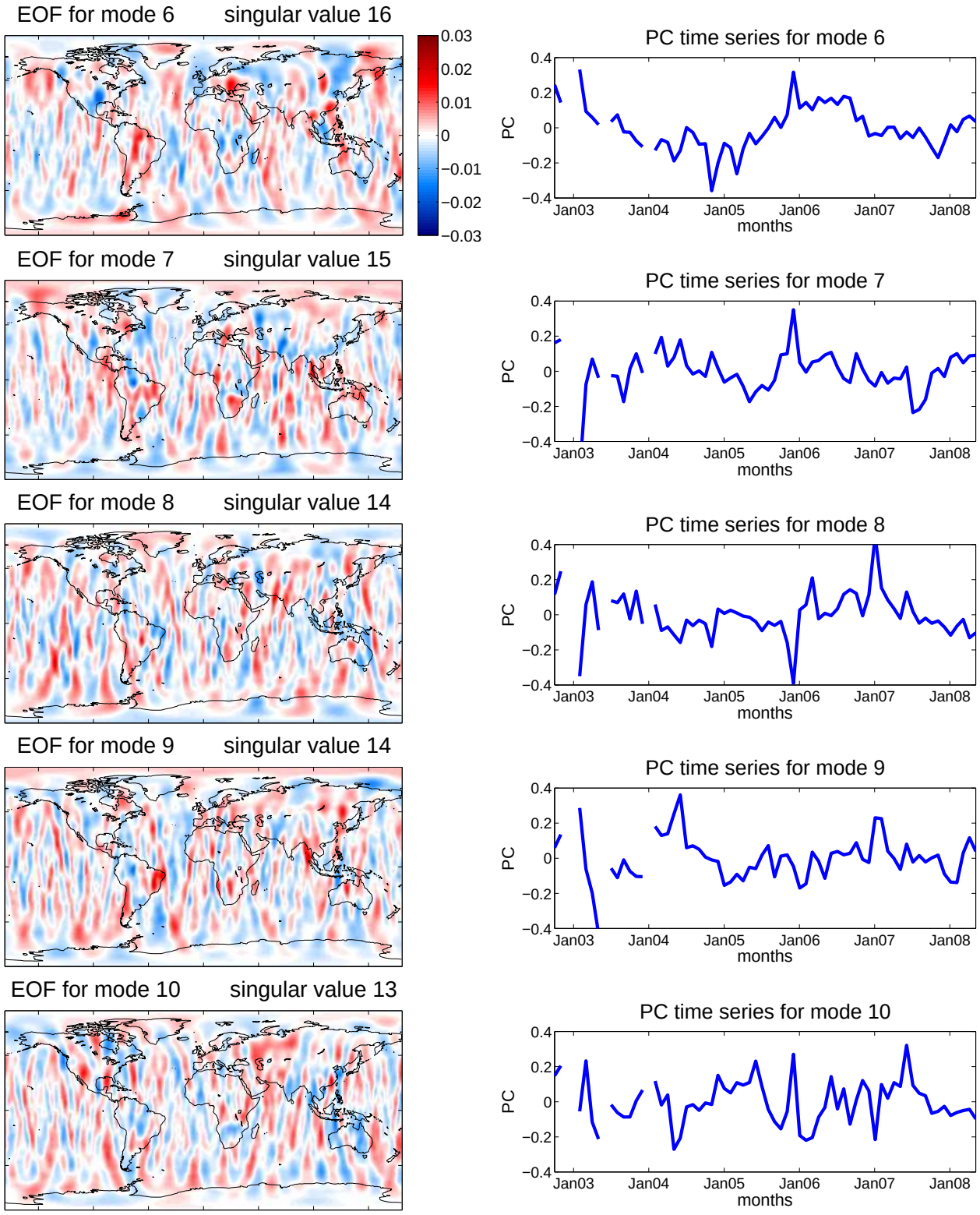


Figure 5.10: EOF patterns and PC time series for modes six to ten with a Gaussian smoothing radius of 400 km

EOFs and PCs for smoothing radius 400 km

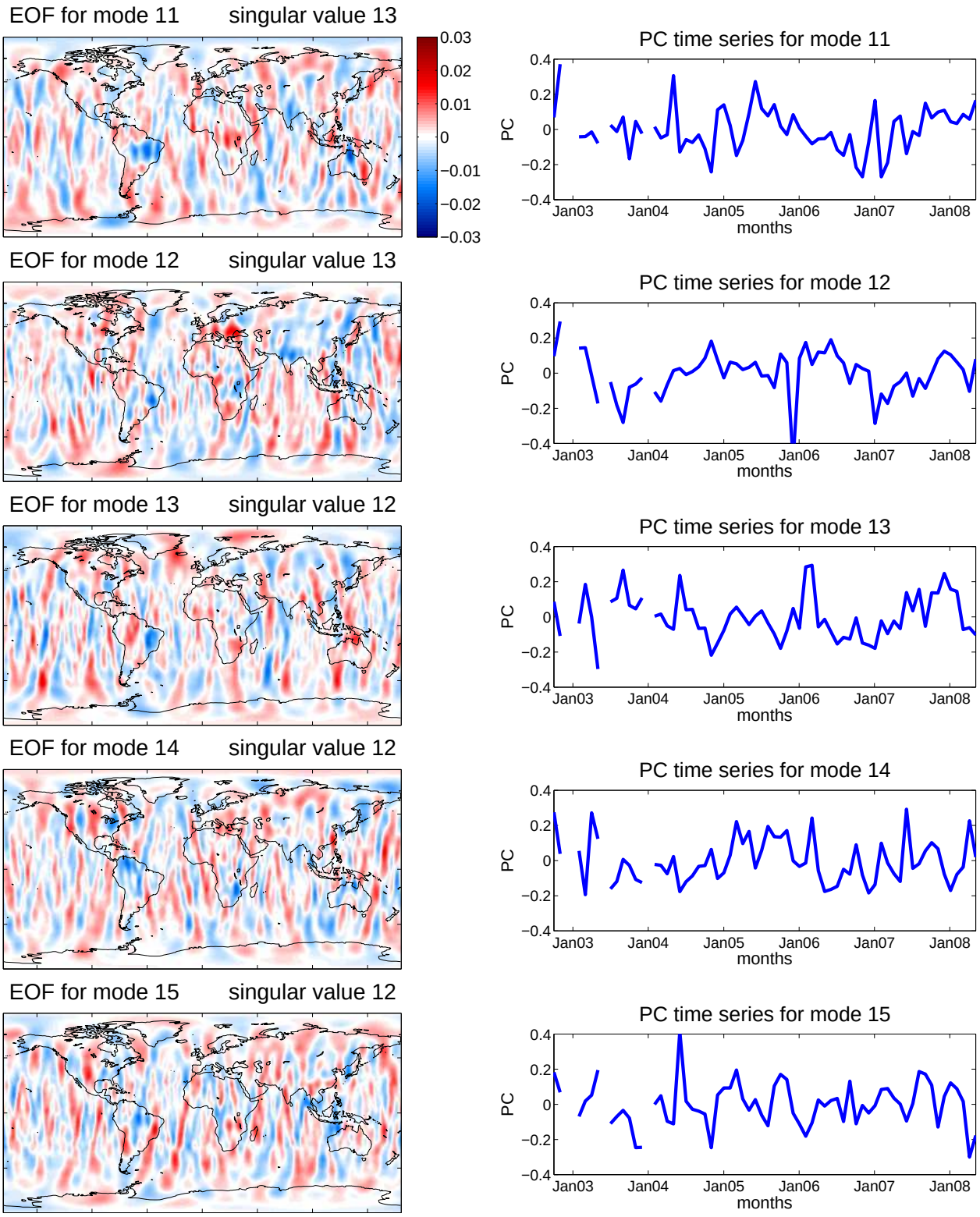


Figure 5.11: EOF patterns and PC time series for modes eleven to fifteen with a Gaussian smoothing radius of 400 km

modes sixteen and 48. For modes one to five please refer to figure 5.3, EOFs and PCs for modes six to ten are plotted in figure 5.12. Modes one up to five have already been discussed in chapter 5.2, and in modes six to eight, there is again a distinguishable behavior in the time series that resembles to that one described for mode six in smoothing radius 400 km. Again, it is extremely hard to find the corresponding physical impact since there are no distinguishable patterns in the EOFs. But from the time series it can be clearly seen why modes six to eight passed the hypothesis test and why modes nine and ten were identified as white noise. The normalized singular values plot in figure 5.8 gave a cut-off value between mode eight and nine. This finding resembles the first significant modes in the KS test. The other modes that are considered to be significant according to the test, modes sixteen and 48, are investigated later in detail.

Details in the outcomes of the KS test have already been explained. Furthermore, there are some outlier modes that have been considered as signal in the test. The most unexpected are modes 55 in smoothing radius 300 km and mode 48 in smoothing radius 500 km. Additionally, there is mode sixteen in smoothing radius 500 km and modes nine for both smoothing radius 400 km and 450 km. It seems really strange that these outlier modes are distributed over all different smoothing radii. This could be a hint for limitations of KS test, since it is very unlikely that any real signal appears in one single high mode. These outliers are investigated in detail now. Therefore it is necessary to look into the procedure of a KS hypothesis test for the power spectral density of each mode's time series.

First, let us have a closer look at mode 55 in smoothing radius 300 km. For being able to perform a KS hypothesis test, the Fourier coefficients of the time series have to be estimated, then the spectrum is normalized and its cumulative sum is calculated. The cumulative sum of the spectrum is compared to the cumulative sum of an optimal white noise spectrum, that is a linear function ranging from zero to one over the whole frequency range of the power spectrum that is to be compared. First, EOFs and PCs for modes 54, 55, and 56 and smoothing radius 300 km are plotted in figure 5.13. Then the time series, power spectrum, and cumulative power spectrum for modes 55 and 56 are given in figure 5.14. The details of the power spectral density, which is used for the hypothesis test, is plotted for modes 55 and 56. Mode 55 is to be investigated, and mode 56 is plotted to have an ordinary white noise

EOFs and PCs for smoothing radius 500 km

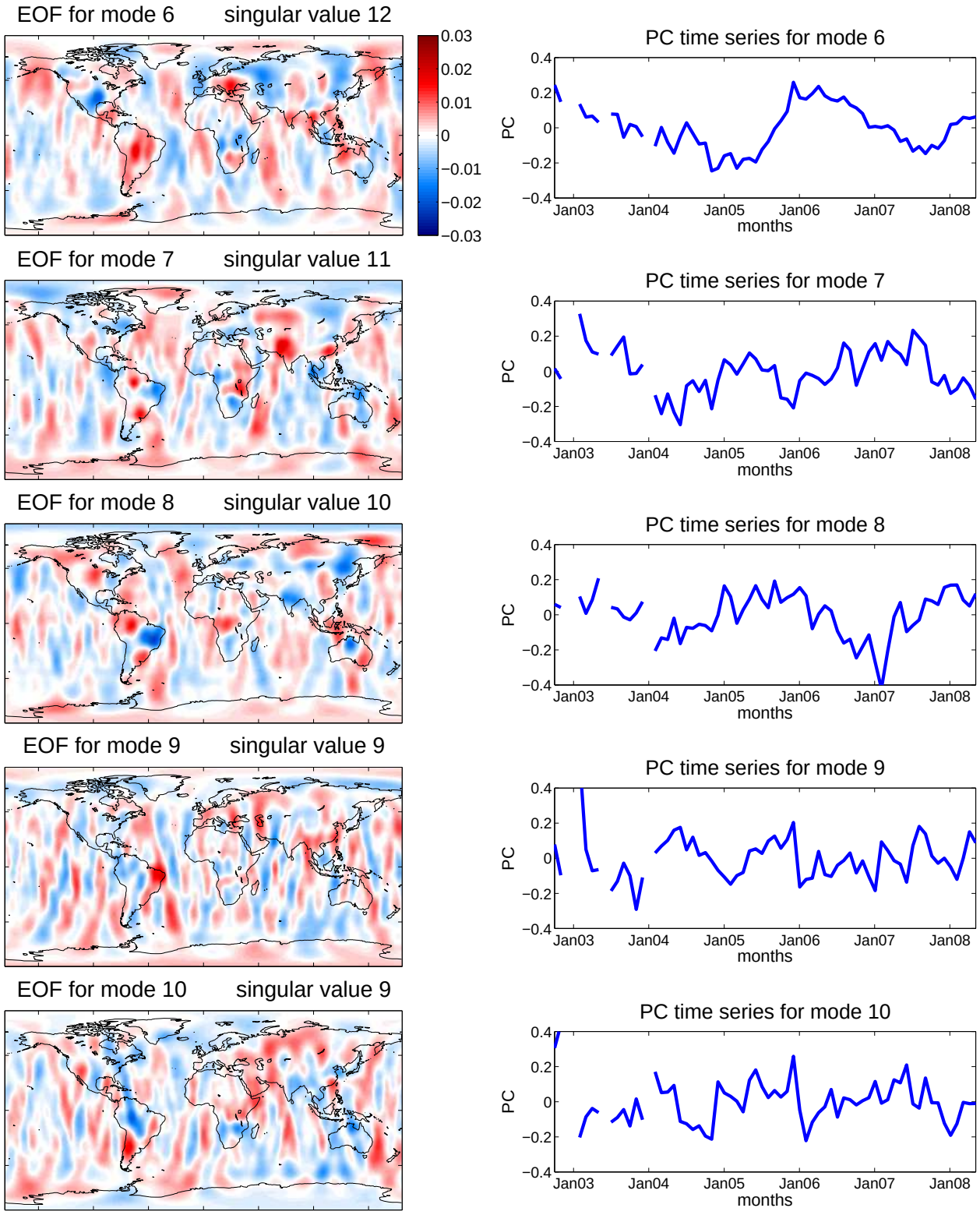


Figure 5.12: EOF patterns and PC time series for modes six to ten with a Gaussian smoothing radius of 500 km

mode to which mode 55 can be compared.

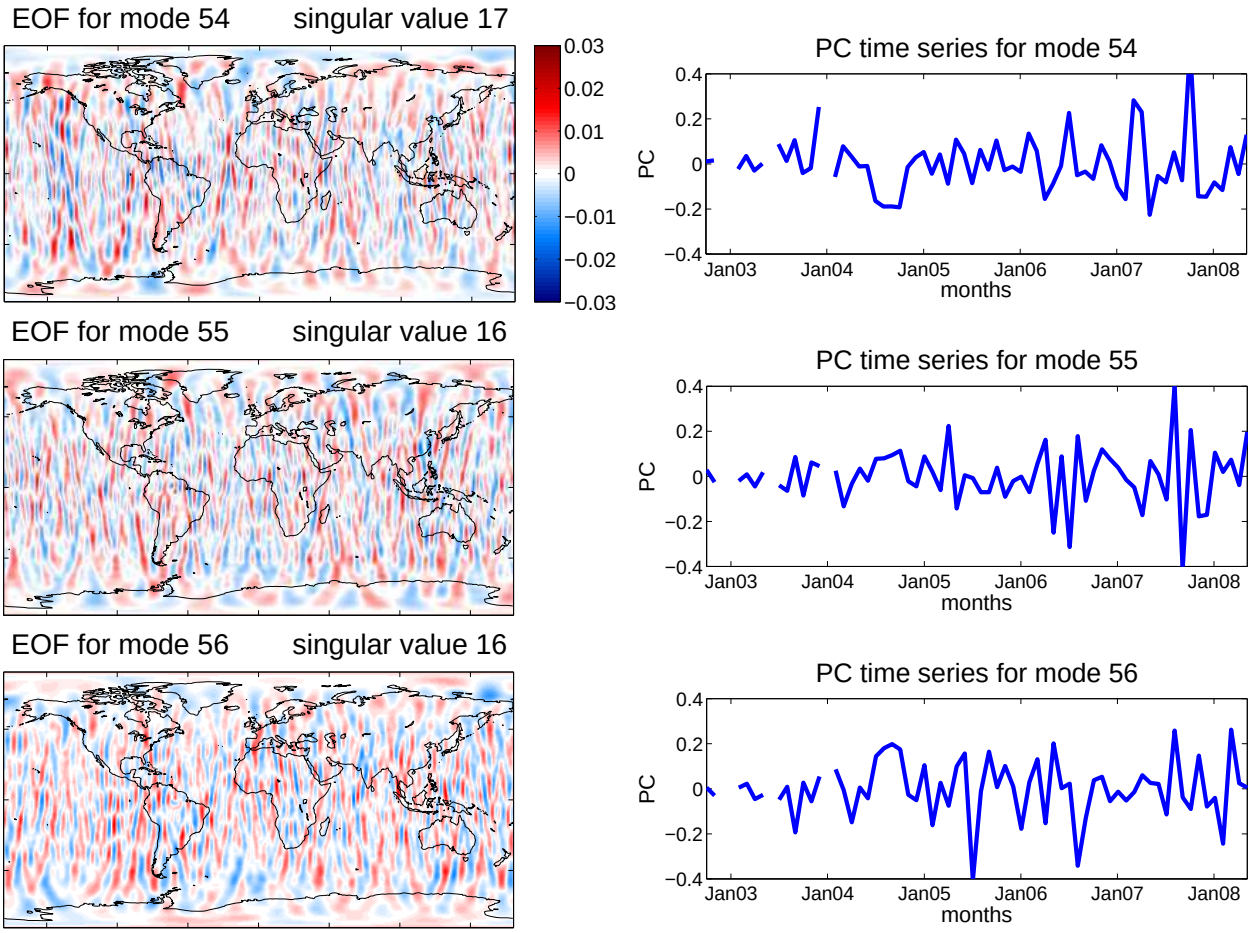


Figure 5.13: Modes 54, 55, and 56 for smoothing radius 300 km

In the plots of EOF s and PCs in figure 5.13 no significant differences can be seen between the three modes 54, 55, and 56. All the three time series do not show anything but noise, and also the EOF patterns do not display anything but the longitudinal stripes due to the correlated errors from GRACE. But in the plots of the power spectral density in figure 5.14, the difference in mode 55 is revealed: This mode contains mainly high frequency information. Going on to the cumulative power spectra, which are compared to white noise, we can see for mode 55 a cumulative power spectrum that differs significantly from white noise. Due to the high frequency, the curve from mode 55 is significantly below the curve of white noise. And of course, therefore considered as not being equal to white noise by the hypothesis test. The test statistic just compares the maximum difference between the calculated cumulative power spectrum and white noise. Since the spectrum for mode 56 is distributed over all frequencies

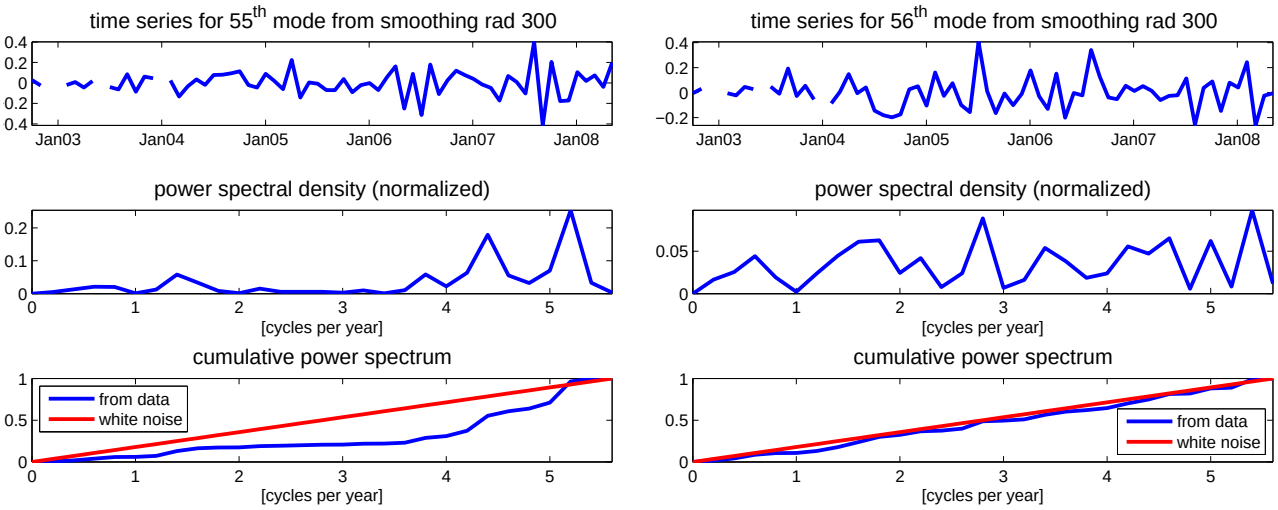


Figure 5.14: Details of Kolmogorov-Smirnov test for modes 55 and 56 from smoothing radius 300 km

its cumulative spectrum is much closer to that one of white noise and therefore considered to be white noise by the test. This is the correct decision, of course. For a correct decision for mode 55 it would help to change the KS test statistic to considering cumulative power spectra below the white noise curve as not significant, too. The high frequencies which differ from white noise are indeed worse than white noise.

The next irregularity in the results of the KS hypothesis test that shall be studied in detail is mode nine from both smoothing radii 400 km and 450 km. The EOFs and PCs for smoothing radius 400 km have been already plotted in figure 5.10. The plots of the cumulative power spectrum densities for both smoothing radii are given in figure 5.15.

The PC time series of these two modes are fairly similar to each other, so they also have similar cumulative power spectral densities. They both have peaks at low frequencies of less than one cycle per year in their power spectral densities. Even if it is hard to find a physical impact that could have caused that signal, it could be a measured signal, on the one hand since it is still a fairly low mode and on the other hand since it is contained in two different smoothing radii. But of course, the low frequencies could also be due to artificial effects from smoothing, since 400 km and 450 km are already large smoothing radii, or due to any other artificial effect. These two modes are critical to decide on.

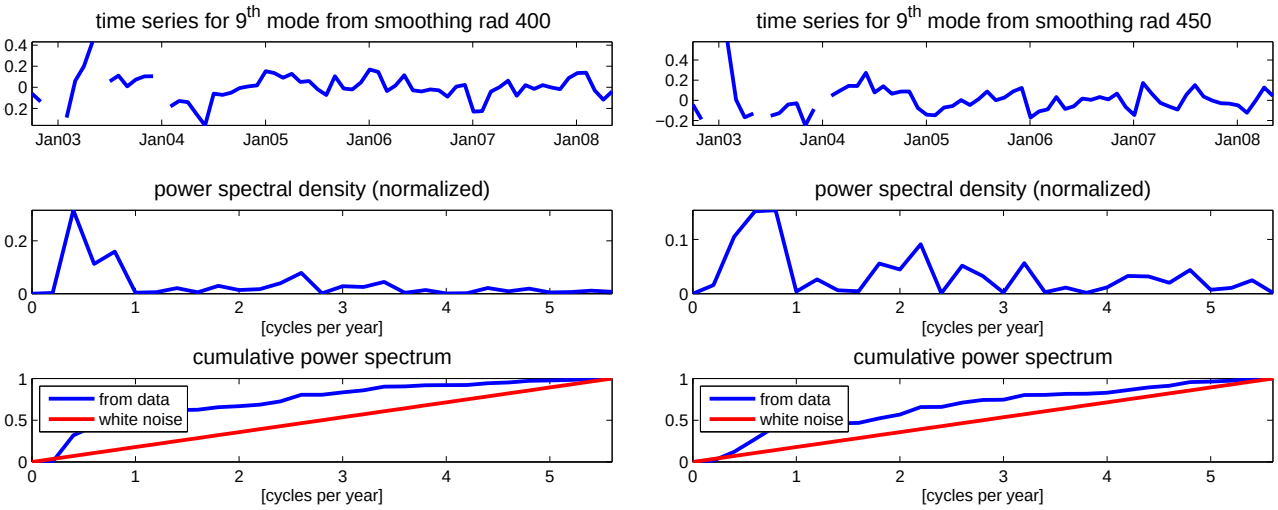


Figure 5.15: Details of Kolmogorov-Smirnov test for mode 9 from smoothing radii 400 km and 450 km

Finally, modes 16 and 48 for smoothing radius 500 km need further investigation. For these two modes, the EOFs and PCs as well as the PC's power spectral density are shown in figures 5.16 and 5.17.

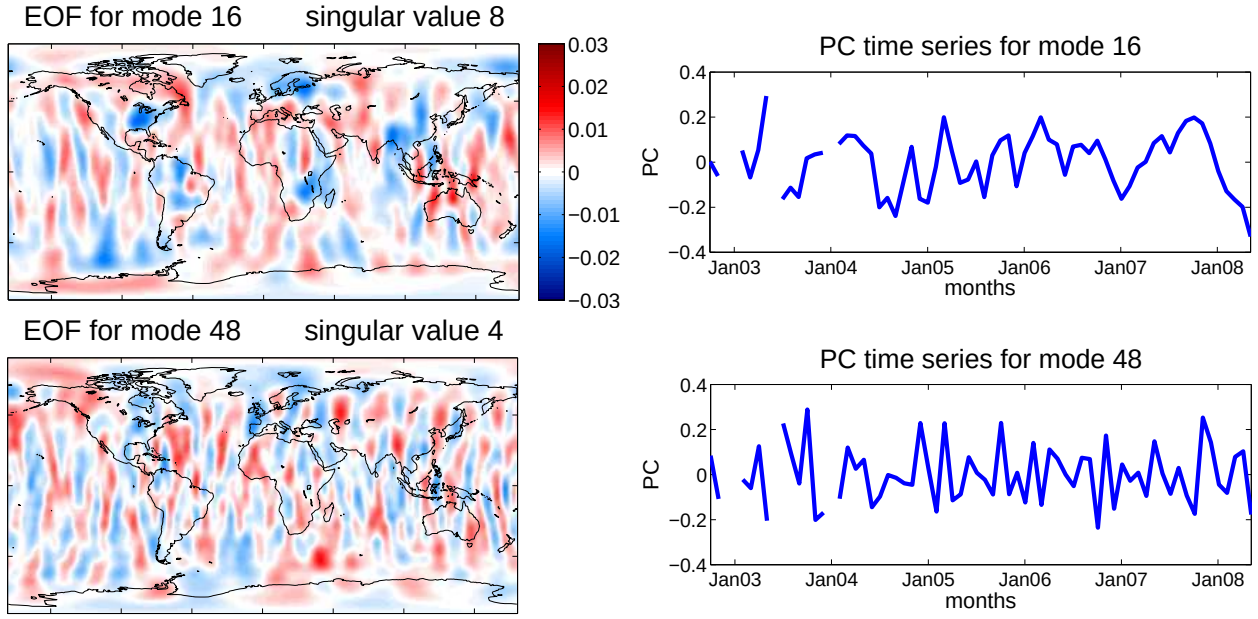


Figure 5.16: Modes 16 and 48 for smoothing radius 500 km

In the EOF patterns of these two modes no major differences can be found, both are dominated by North-South stripes. The PC of mode sixteen shows lower frequencies than the PC

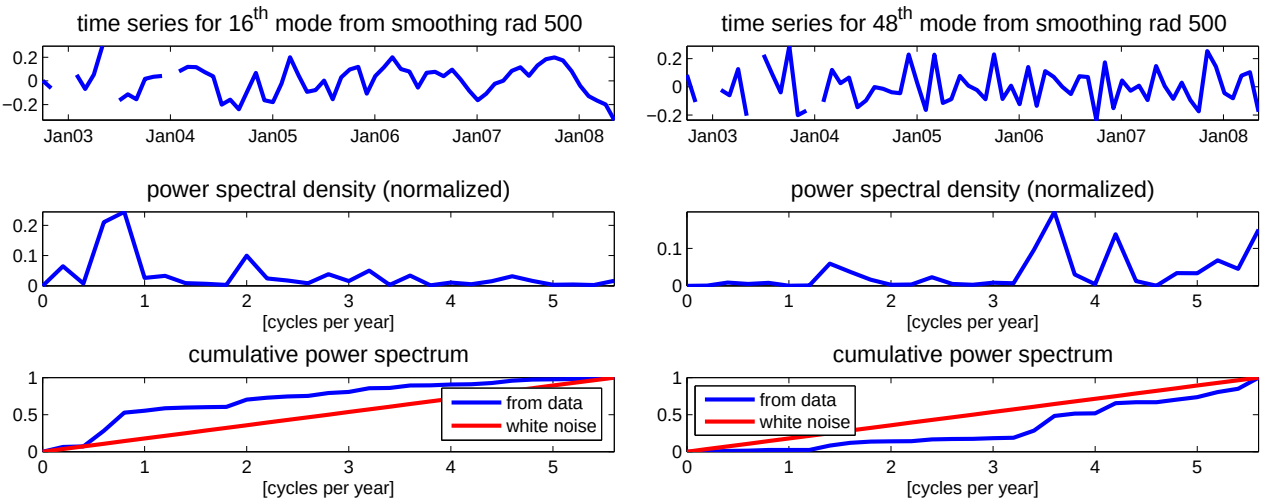


Figure 5.17: Details of Kolmogorov-Smirnov test for modes 16 and 48 from smoothing radius 500 km

of mode 48. But in the time series plot no specific pattern can be detected. However, going to figure 5.17, in the power spectral density it can be seen that in the time series of mode sixteen there is a wide peak at a frequency between half a cycle per year and one cycle per year. Since this is exactly what shall be detected by the hypothesis test, the test performed adequately. However, it is extremely hard to find a physical impact that could have caused this signal. It needs more investigation into GRACE signal characteristics, detectable physical mass change effects, and possible aliasing effects in the GRACE measurements to decide about this mode. Mode 48 shows the same problem as mode 55 for smoothing radius 300 km. The time series contains high frequencies that lead to a cumulative power spectrum density below white noise. Again, a correct result from the hypothesis test could be obtained by modifying the hypothesis test in a way that a cumulative power spectrum curve below the white noise curve would be considered as error.

For performing a KS hypothesis test, MATLAB provides a function called kstest2*. Input for the function are each modes power spectral density, the white noise line and the significance level. The first step is calculating the power spectral density for each mode's PC time series. The MATLAB function* kstest2 *only performs the actual KS test by comparing the calculated cumulative power spectrum density to that one of white noise, which is just a linear function ranging from zero to one over all frequencies. The result of the MATLAB* kstest2

function is a true or false value depending on the test result, indicating if the null hypothesis was accepted or rejected. The procedure of estimating the Fourier coefficients and performing the hypothesis test has to be done for each mode individually.

Improved Kolmogorov-Smirnov Test

In the previous examples, we can clearly see the limitations of the KS hypothesis test. Modes with time series that contain high frequencies were in some cases considered to be signal, however the high frequency time series in the high modes are most likely due to noise and errors. Of course, the physical meaning of the individual modes has to be investigated more to be sure not to dismiss any mode that contains signal. Since the GRACE satellites measure all global mass change signal that there is, there is the possibility that some modes contain at least traces of a signal that would not be considered as real signal at first sight. And of course, we have to be careful for not taking aliasing effects as signal. But let us stick to the assumption stated above that these high frequencies in the high modes are due to noise. One way to improve the results of the KS hypothesis test is not taking the absolute difference between the calculated cumulative power spectra and white noise as test statistic, but taking the real value of the maximum difference and considering the cumulative power spectra that are below the white noise line as noise, too. That would mean the null hypothesis is changed from an absolute value to a real value. This new test was implemented in MATLAB, and it resulted in the expected effect. The high modes are no more considered as signal. See figure 5.18. But note in the two plots for the results of the new hypothesis test, that for a significance level of $\alpha = 5\%$ there are more of the lower modes passing the test than in the original test for $\alpha = 5\%$. Test new hypothesis test with $\alpha = 2.5\%$ leads to exactly the same results in the low modes as the original test for $\alpha = 5\%$. So note that the significance level has to be adjusted if only real difference values are tested.

However, in general, a signal can also be a high frequency signal. So using the new null hypothesis that considers all high frequency signal as noise would restrict our ability to find signal. An alternative idea would be to use the original form of the hypothesis test, with the absolute values as test statistic, which also permits high frequency signal. Then, combine the KS test results with the singular values method. To combine both methods as proposed

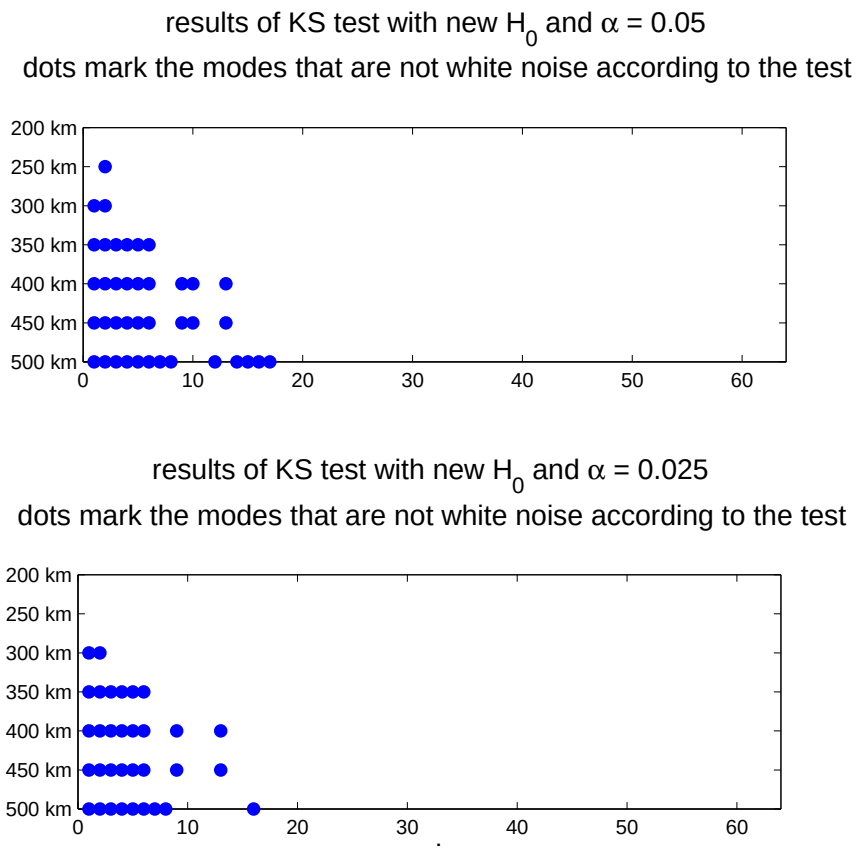


Figure 5.18: KS test results for $\alpha = 5\%$ (top) and $\alpha = 2.5\%$ (bottom) for the new null hypothesis and different smoothing radii

would permit finding high frequency signal as long as its contribution to the total variance is high enough. A hypothesis test could be performed first, and could then be followed by truncation with a cut-off value. The cut-off mode can be chosen high enough to be very sure not to truncate signal, that means a cut-off value higher than where it would have been for the singular values method only can be chosen, since a KS test is performed afterwards. For the GRACE data example, a cut-off value could be chosen from figure 5.8 as high as mode 25. If the results from the original KS test, figure 5.9, are taken and then truncated by a cut-off mode 25, we would just loose mode 55 for smoothing radius 300 km and mode 48 for smoothing radius 500 km. This method would be a very viable alternative for a reasonable selection of modes.

Alternatively, the results from the KS test could be weighted by the normalized singular values to decide which modes from the outcomes of the hypothesis test should be considered as signal.

5.4 Synthesis

After the selection of signal-containing modes, the synthesis step can follow, using only the signal modes. The reconstruction completes the procedure when EOF analysis is used as a filter and the reconstructed data is the filtered result. Of course, the whole GRACE data set covering 68 months is reconstructed, but only examples can be plotted. See figures 5.19 to 5.21. The first example in figure 5.19 shows the month May 2008 from a reconstruction with the first four modes from smoothing radius 300 km on the left, and in the second plot on the right the input data for EOF analysis, only filtered with a Gaussian smoothing function of radius 300 km. The first four modes were used for synthesis based on the result from the singular values plot. However, the KS hypothesis test gave a different result, only the first two modes are significantly signal due to the KS test. For comparison, a second reconstruction of the same month, May 2008, is plotted with only two modes in the synthesis step. See figure 5.20. Again, the synthesized data is plotted on the left, and the input data, which is exactly the same as in figure 5.19, on the right. In comparison of the reconstruction in figure 5.20 with the reconstruction using the first four modes in figure 5.19, only slight differences can be seen. This shows that the major part of signal variance is contained in the first two modes, and the other modes do not contribute much to the signal part. The hypothesis test gave better results than the singular values method in this case.

The third example in figure 5.21 again shows the same month, May 2008, but for smoothing radius 400 km. In the first plot, the reconstruction with those modes that have been considered to be signal in the hypothesis test, that are modes one to six, and nine, are shown on the left. On the right in figure 5.21 the input data for EOF analysis is displayed, the data set only filtered by a Gaussian smoothing of radius 400 km.

For a reconstruction, the synthesis step, the matrices $\mathbf{U}$*,* $\mathbf{\Sigma}$*, and* $\mathbf{V}$ *have to be multiplied according to the synthesis equation (2.99).* χ_i *has to be set to one or zero according to the results of the KS test, or the decision whether a mode should be retained or not.*
Once the data matrix has been calculated, it has to be reorganized into its prior shape and dimensions. The reorganization step of the matrix that has been done as he very first step before all the analysis has to be reversed. Then, each month's reconstructed map can be plot-

ted.

When plotting the EOF patterns, we also have to keep the previous organization of the data matrix in mind. Each EOF is contained in a vector in matrix $\mathbf{V}$*. The vector has to be rearranged into a map before plotting.*

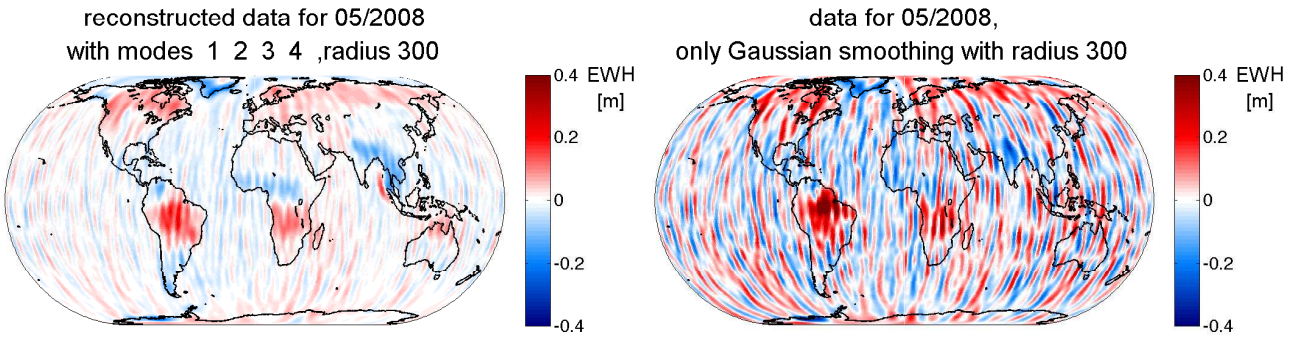


Figure 5.19: Reconstruction using the first four months and original data for May 2008 with smoothing radius 300 km, in Eckert IV Projection

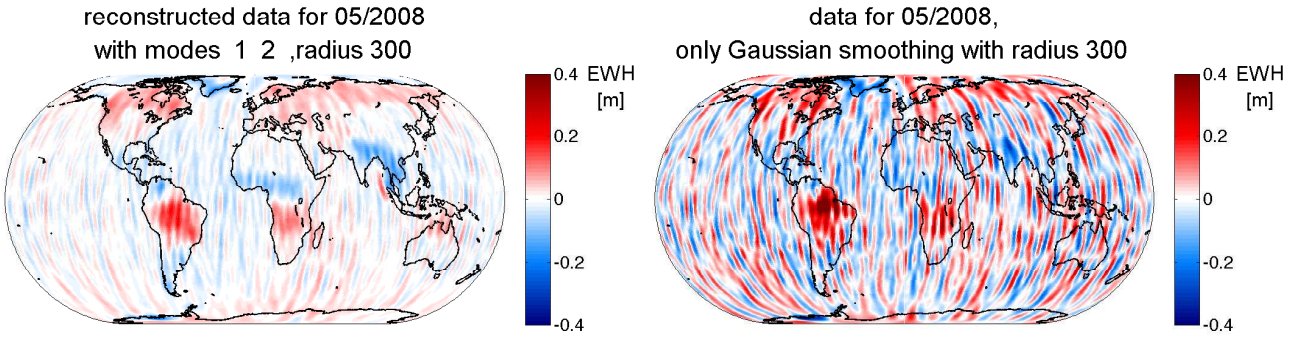


Figure 5.20: Reconstruction using the first two months and original data for May 2008 with smoothing radius 300 km, in Eckert IV Projection

Concluding from the discussions in this chapter and figures 5.19 to 5.21, we have to point out that EOF analysis serves as a good filtering tool to reduce the GRACE characteristic North-South stripes. In figure 5.20 the stripes have been reduced significantly by performing EOF analysis. In the data set before EOF analysis the characteristic stripes have approx-

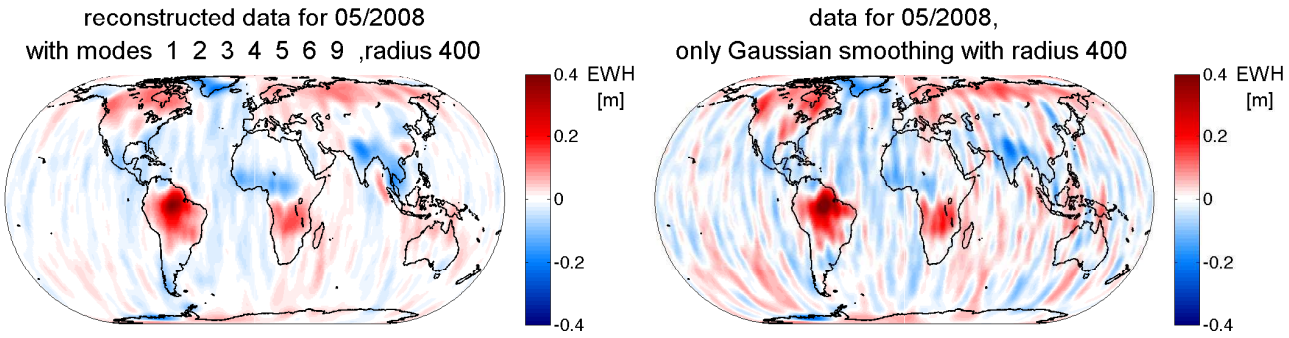


Figure 5.21: Reconstructed and original data for May 2008 with smoothing radius 400 km, in Eckert IV Projection

imately the same magnitude as the signal. It is hard to identify signal parts, they seem to be covered by the stripes. The only signal that is detectable visually before filtering by EOF analysis is in the Amazon basin and in Greenland. In the filtered result, there are still North-South stripes present. But their magnitude is much lower compared to the map before EOF analysis. And, even more important, in the map of the EOF-filtered result the magnitude of the stripes is much lower than the magnitude of the signal. In this map, the signal North and South of the equator can be clearly seen, as well as a signal in the Sumatra earthquake region, and a signal in Greenland and in North America. The stripes are only a minor effect in the map after EOF analysis.

In the results of EOF analysis after more pre-smoothing, a Gaussian smoothing radius of 400 km, in figure 5.21, a significant reduction of the stripes can be seen, too. Even if the North-South stripes are no more that dominant in the map before performing EOF analysis, there is still a certain amount of striping that has been removed by EOF analysis. It is remarkable, that in the EOF-filtered result the signal magnitude is still about the same as in the map before EOF analysis. EOF analysis has only removed the major part of stripes.

Chapter 6

Continents and Oceans

One of the major advantages of EOF analysis, compared to various other methods available for analyzing gravity data, is the method's independence of the kind of data to be analyzed. The method is just a mathematical matrix decomposition and physical background or size of the region to be analyzed does not matter. The ability of the EOF method to detect signal components just depends on the significance of the signal's variance structure and its separability from the overall variance. A second advantage is that EOF analysis can be applied to any distinct region or subset of the data.

In this chapter, continents and oceans are studied separately. In the first two subchapters analysis and synthesis are described for EOF analysis preformed on the continental part of the same data set which was already used in chapter 5. In the subchapters 6.3 and 6.4 another subset of the original data set is analyzed: only data from the oceans.

6.1 EOF Analysis for Continents Only

For EOF analysis only on the continents, the data matrix has to be changed in an appropriate way to have only gravity data from the continents in the data matrix. For this new data matrix EOF analysis is performed. When plotting the results, in the EOFs as well as in a reconstructed data set, the values have to be rearranged properly. Of course it is very important that this rearranging just reverses the procedure where the data matrix was built. For building the new data matrix no gaps were used for the oceans, but the data matrix is filled tight with the values to be used. The spatial dimension of the data matrix decreases according to the smaller number of values in the global grid for continents only.

In MATLAB there are different ways to filter the continental data from the data over the oceans. The set topo *provides coastlines which can be used to build a matrix, in the same size as the data grids, containing a one for each continental point and a zero for each ocean point. With the help of this matrix, the continental points can be selected and rearranged in a data matrix as input for EOF analysis. When reorganizing the patterns after EOF analysis this filter matrix helps as well, since it gives the continental structure that is filled with the data points.*

Figures 6.1 - 6.3 show the first few modes for different smoothing radii from EOF analysis on data from only the continents.
It is remarkable in the continents-only modes, that there is signal to be detected already for smoothing radius 250 km, while in the global analysis, the first signals were detected for smoothing radius 300 km. The first mode in figure 6.1 is still very noisy, it is hard to identify a signal. But even if the time series in the second mode are still noisy, too, the underlying annual signal can be recognized. And the corresponding EOF pattern shows a positive and negative band just north and south of the equator. This mode represents the annual changes in tropical rainfall, as it was seen in the first mode for smoothing radius 300 km for the global analysis in figure 5.3. In the third mode there is still a noisy annual pattern in the time series to be seen, but the pattern is less distinguishable. There is again something to see around the equator, but less distinct than in mode two. Modes four and five are in the time series as well as in the patterns very noisy, no physical signal can be identified. The most interesting finding in the modes from smoothing radius 250 km is that there is signal at all, while for the global solution in figure 5.2 no signal could be seen. This is explained as follows: Since the most dominant mass changes are due to continental effects, and these are the major signal measured by GRACE, the measurements over the oceans contribute much more noise than signal. This large amount of additional noise covers the signals in the global analysis. When only continental data is used, a certain amount of noise is no more contained in the data right from the start. Since there is less noise variance contained in the continental data, the annual signal is easier to detect an can already be found for less smoothing.

For smoothing radius 300 km, in figure 6.2, the expected annual signal in the tropics appears very clearly in mode one, much smoother and clearer than for the global solutions.

EOFs and PCs for smoothing radius 250 km

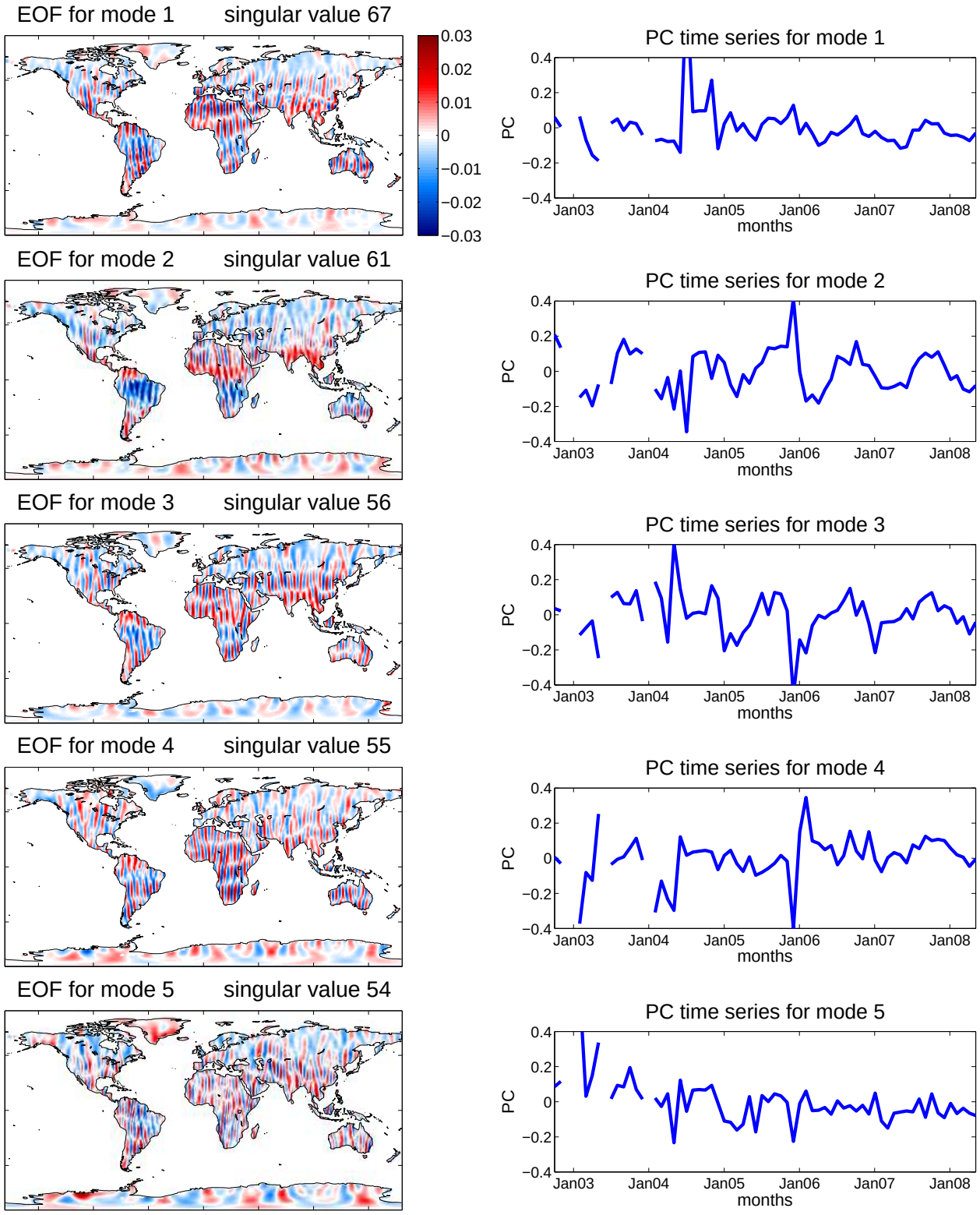


Figure 6.1: EOF patterns and PC time series for the first modes with a Gaussian smoothing radius of 250 km

EOFs and PCs for smoothing radius 300 km

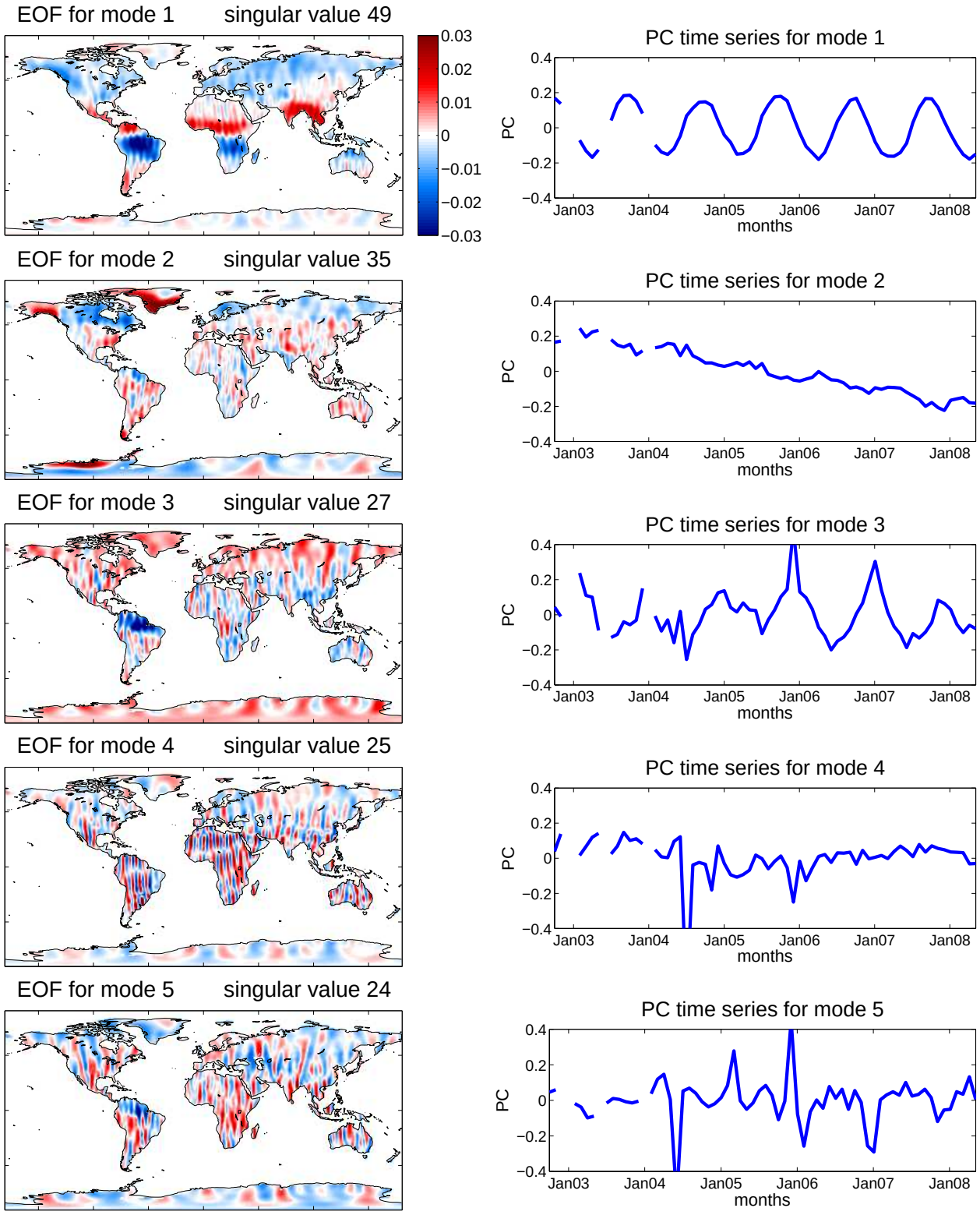


Figure 6.2: EOF patterns and PC time series for the first modes with a Gaussian smoothing radius of 300 km

EOFs and PCs for smoothing radius 350 km

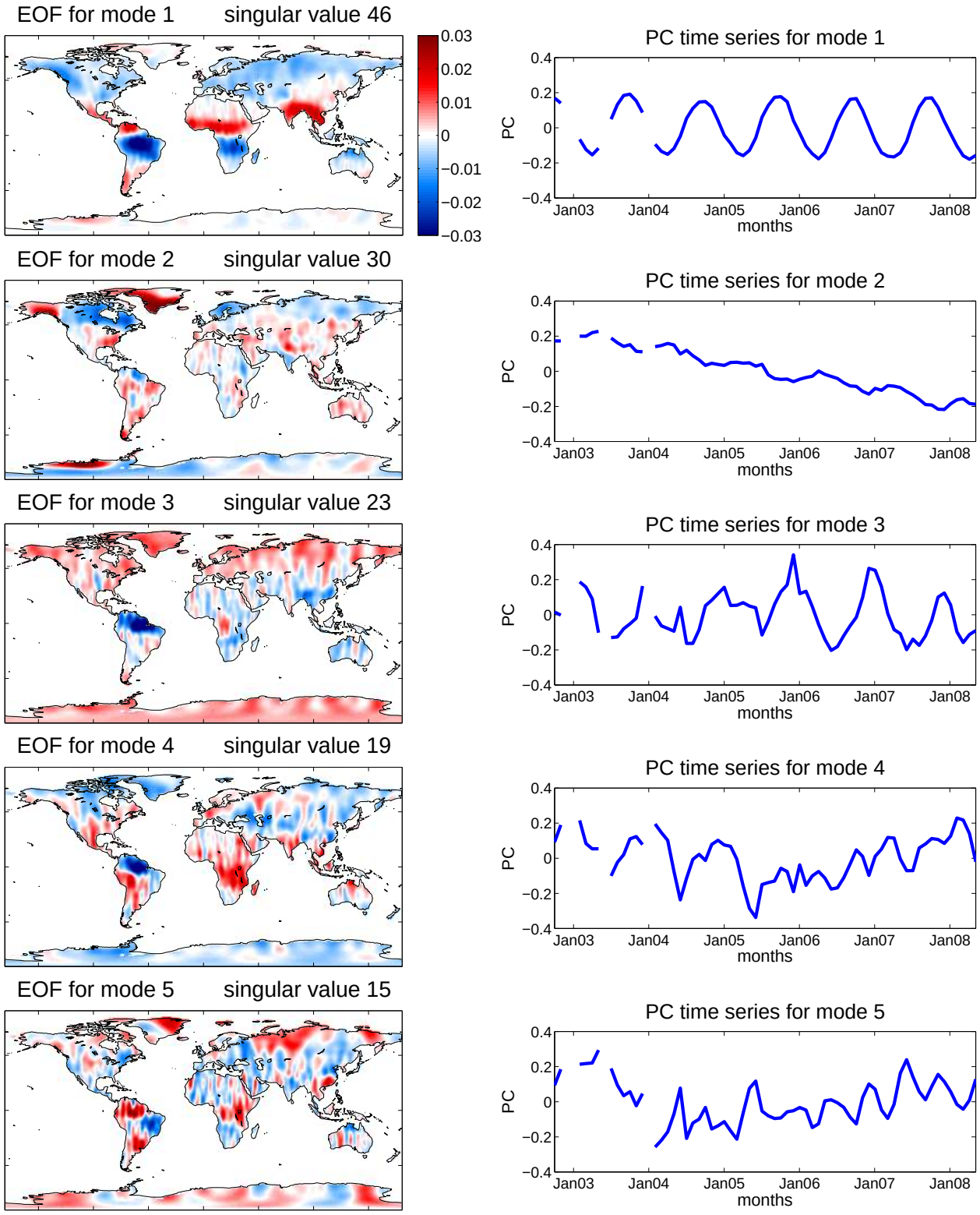


Figure 6.3: EOF patterns and PC time series for the first modes with a Gaussian smoothing radius of 350 km

Remarkable is mode two. While in the global solutions, even for very large smoothing radii, the trend signal in the time series was always very noisy, it is in figure 6.2 very clearly a trend signal and much smoother than in the global solutions. As expected, the areas where the downward trend takes place are again in Greenland, West Antarctica and Alaska. Again, the post-glacial rebound can be seen in North America and Scandinavia. It is surprising, that especially the trend signal gets so significantly smoother when using continents only. Maybe ice masses reaching into the ocean in Greenland and Antarctica cause the noise in the signal of the global data. These are cut off in the continents-only data. Furthermore, the third mode shows the annual signal in the Amazon basin, with the characteristic peaks in the time series. Remember that this signal first appeared for 350 km smoothing radius in the global solutions in figure 5.4. Again, a signal could be detected at 50 km less smoothing compared to the global analysis. Most likely due to the noise imposed by the oceans in the global fields. Modes four and five are hard to interpret, both EOFs and PCs are very noisy.

Finally, for 350 km smoothing radius in figure 6.3, the expected signals are encountered again, and again surprisingly clear and distinguished. The first three modes show nicely the expected signals of the annual signal in the tropics, the trend mainly in Greenland, and the annual peaks in the Amazon drainage basin. In contrast, in the global solutions there were two different trend signal, in modes two and three. See for example figure 5.4. Here, the signal from the Amazon basin appears already in mode three. The second trend signal in the global solutions must originate from an oceanic effect, since it does not appear in the continents-only decomposition. Modes four and five are very different from white noise, but they are hard to interpret. Mode four shows a PC time series signal that has annual character for the first two and a half years, but then it changes its characteristic to a kind of upward trend with noisy periodic effects with two peaks a year imposed. The corresponding regional patterns are in Africa, in the region of Lake Victoria and Kilimanjaro and in South America, slightly north of the Amazon basin and a smaller effect in southern Peru and Bolivia. Mode five shows a noisy behavior in the time series that is hard to interpret. Also the EOF patterns are hard to interpret, but they do not mainly show the GRACE-characteristic longitudinal stripes.

6.2 Selection of Modes and Reconstruction for Continents Only

Selection methods can be used for the modes from continental data in the same way as they are used for global data. All the background information has been given already in chapters 3 and 5. In this subchapter, the singular value plots are shown and the already well-known selection method of a KS hypothesis test is applied to the time series. The reconstruction shall be also demonstrated in this subchapter, again for an example month.

First of all, a plot of the singular values and a normalized version of this plot are given in figure 6.4. Second, the results of the KS hypothesis test for the power spectral densities of the time series are shown in figure 6.5.

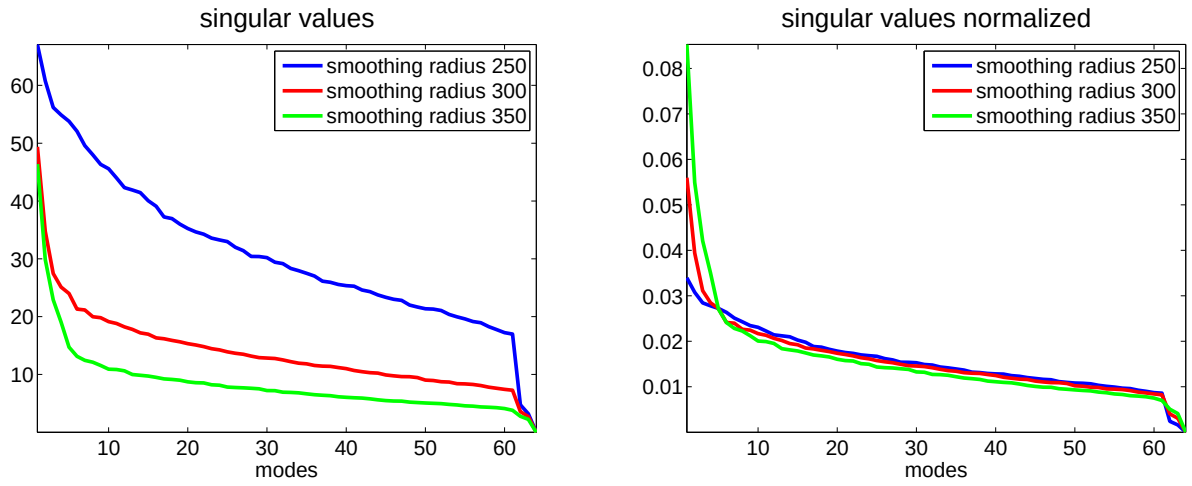


Figure 6.4: Singular values for different smoothing radii

The first significant mode in the hypothesis test for smoothing radius 250 km resembles the mode that would have been chosen from the visual impression of the first five EOFs and PCs in figure 5.2. Out of the first few modes, mode two is the only one that seems to be different from white noise in the plots in figure 6.1. For getting an impression of mode nine, that is significant according to the test, mode nine is plotted in figure 6.6 and the details of KS test for mode nine in figure 6.7. For a comparison, the KS test details of mode ten are also plotted in this figure.

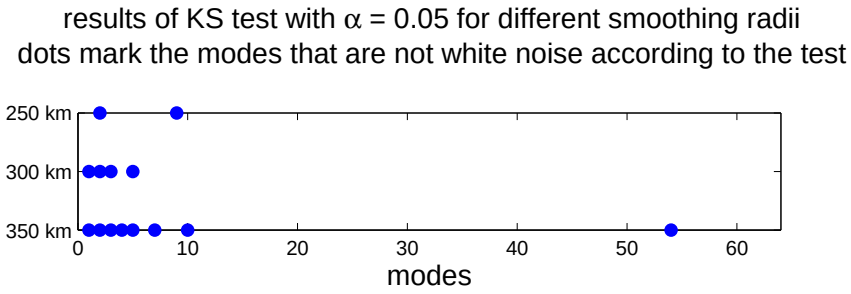


Figure 6.5: KS test results for $\alpha = 5\%$

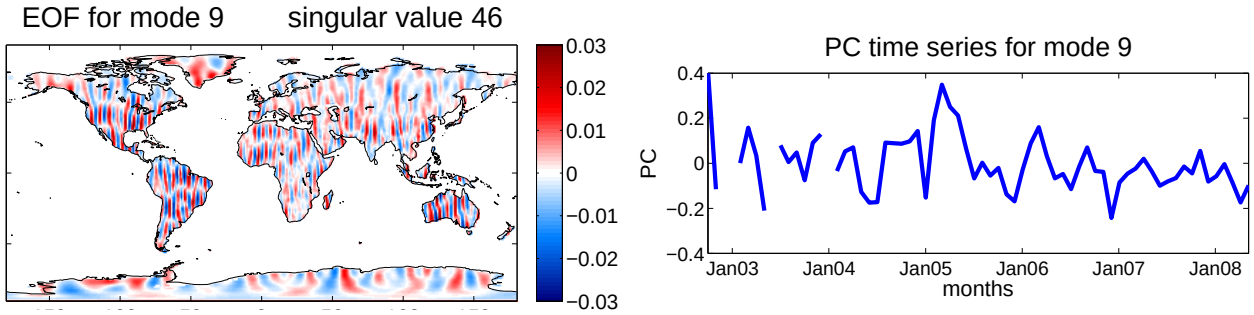


Figure 6.6: EOF pattern and PC time series for the ninth mode with a Gaussian smoothing radius of 250 km

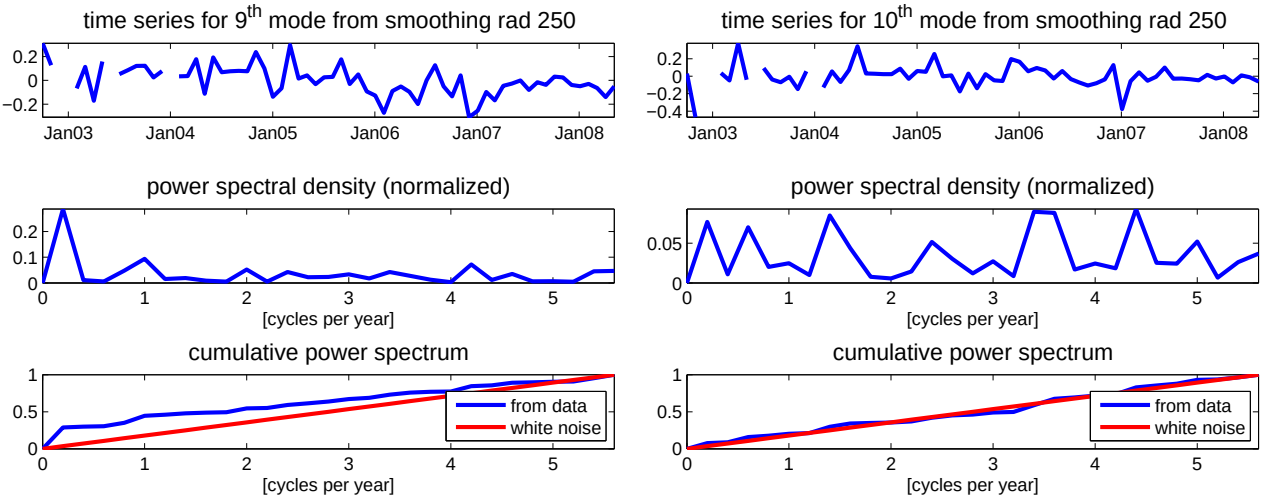


Figure 6.7: Details of Kolmogorov-Smirnov test for modes 9 and 10 from smoothing radius 250 km

The time series of mode nine has a pattern where a low frequency signal of less than half a cycle per year can be found. This is considered to be signal in the test, it is significantly different from white noise. However, it is very hard to recognize physical signal in this mode. Since all the other modes, besides mode two, have been considered to be noise, most likely this mode is noise, too. The singular values plot in figure 6.4 also does not give a hint for a significance in mode nine.

From the other results that were already presented in this thesis, we know that the annual signal in the tropics is the strongest expected signal. For this reason it can be concluded that this strongest signal could be detected in this low smoothing radius due to less noise surroundings than in the global analysis, it is projected into mode two. And since the second strongest signal expected is the trend mainly in Greenland, which has not been found in the other modes, the most likely other signal to be detected besides the annual pattern is this trend. However, mode nine is far from being a trend and therefore not considered as signal. On the other hand, the very low frequency in this mode can be a hint for a hidden trend, but it is not evident here.

For smoothing radius 300 km modes one, two, three, and five are considered to be signal in the KS hypothesis test. The physical meaning for modes one to three has already been explained in the first subchapter 6.1, so of course these modes should be containing signal. When looking again at the EOF and PC plots in figure 5.3 the findings in the hypothesis test for mode four can be confirmed. Since there is nothing in the EOF and PC that is different from white noise by visual judgement, it seems to be correct to denote it as noise. However, mode five is considered to be signal in the test. In the PC time series in figure 6.2, a repeating pattern can be found during the last two years. And there are some sharp peaks for which a regularity can be guessed during 2004 and 2005. But during the first year, no significant behavior can be seen. In the EOF a very slight pattern can be interpreted in South American and Africa. Even if it cannot be clearly associated to a physical effect, this mode can be seen as a signal-containing one.

The results of the KS test (figure 6.5) denote the first five modes of smoothing radius 350 km as signal. Comparing these results to the first five plotted EOFs and PCs we can

EOFs and PCs for smoothing radius 350 km

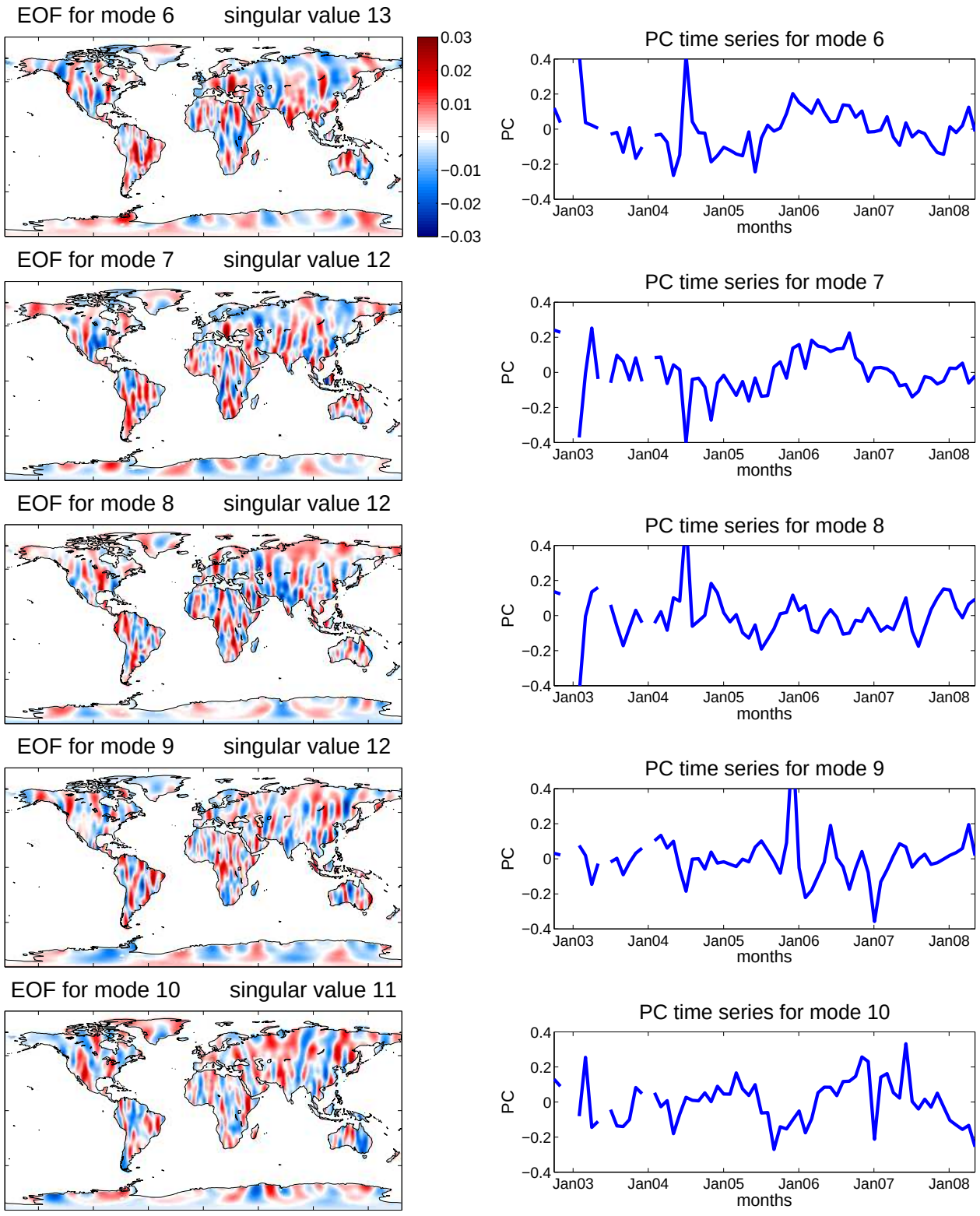


Figure 6.8: EOF patterns and PC time series for modes six to ten with a Gaussian smoothing radius of 350 km, modes seven and ten were identified as signal by the KS test

agree of course for the first three modes, these signals have been named very often by now. The time series of mode four show an interesting pattern, it seems to be divided into two parts. During the first two and a half years there is an annual pattern, imposed by a slight downward trend. The second two and a half years show a very different annual pattern imposed by a slight upward trend. The corresponding regions in the EOF are in Africa the region around Kilimanjaro and Lake Victoria, and in South America a region slightly north of the Amazon basin. These annual patterns could be again due to water levels in drainage basin of a river, but this is only a guess, there is no really strong indication. The time series in mode five show a regular series of smaller peaks roughly about half a year apart. This regularity must have been the indication for the hypothesis test to consider this mode as signal . The EOF patterns are hard to interpret. Modes seven and ten are also considered as signal, and it might be helpful to have a look at the EOFs and PCs of these modes. The EOFs and PCs of modes six to ten for smoothing radius 350 km are shown in figure 6.8.

Modes seven and ten, which are considered to be signal, show some indication of having low frequency signal in their time series. But since for all modes six to ten the time series and also the patterns are hard to interpret, all these modes could also be considered as noise by judgement from the EOF and PC plots. Again, there are limitations of the KS hypothesis test, since by visual examination of the modes plotted in figure 6.8 none of those modes would be considered to be signal. Furthermore, also mode 54 is considered to be signal, the same effect that was already described at the end of chapter 5.3 appears. The mode is again considered as signal due to some high frequency pattern in the time series. Maybe these could be due to aliasing effects. But quite sure, this is not one of the signals we are looking for. From the singular values plot, we can decide on a cut-off values by mode 30, and still being very sure not to loose any signal. With this reason mode 54 is eliminated as not containing signal.

Finally, to give a reconstruction example, reconstruction is done again for May 2008 for smoothing radius 300 km in figure 6.9. Since modes one to three are for sure signal, a reconstruction using only those three modes is plotted, while a reconstruction with the modes that were given from the Kolmogorov-Smirnov test is plotted in figure 6.10. Both plots are for the same smoothing radius. The maps on the left show the reconstructed maps, while

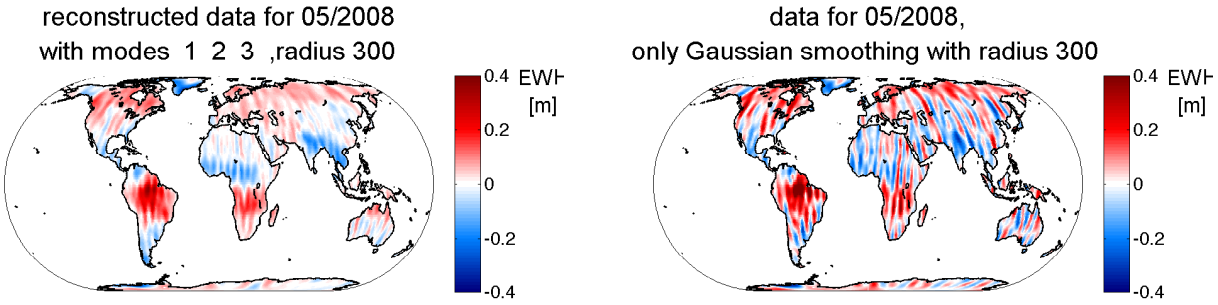


Figure 6.9: Reconstruction using the first three modes and original data for May 2008 with smoothing radius 300 km, in Eckert IV Projection

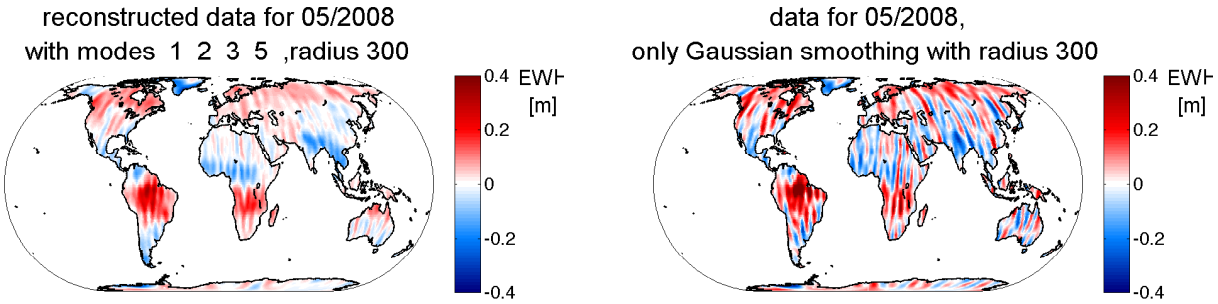


Figure 6.10: Reconstruction using the first three and the fifth modes and original data for May 2008 with smoothing radius 300 km, in Eckert IV Projection

on the right the input data to EOF analysis is plotted for comparison. In the two different reconstructions we can see that mode five adds a bit more signal. The overall signal seems to become a bit stronger in the reconstructed data in figure 6.10. But from visual judgement no additional correlated errors can be seen in the second reconstruction. Mode five must have been a signal-containing mode. Again, as in the reconstruction for the global maps, EOF analysis has filtered out a large amount of the North-South stripes in the continental maps from GRACE. This effect can be clearly seen in comparison of the input data of EOF analysis on the right in figures 6.9 and 6.10 and the filtered results on the left.

6.3 EOF Analysis for Oceans Only

First of all, the first few modes for different smoothing radii for the oceans only are presented in figures 6.11 to 6.13. It has already been mentioned, that most of the mass change signal, that is detectable by the satellites, originates from continental areas. Ocean currents of course also cause mass changes, but these are hardly detectable by a satellite since its signal is not strong and distinguished enough. Thus, more noisy data is expected from the oceans than from the continents. This proves to be the right assumption, the smallest smoothing radius where there is signal detectable in the oceans is 350 km. So the plots start with smoothing radius 350 km.

The first mode's PC time series show an annual cycle, however not as smooth and sine-like in shape as we saw it in the first mode from the global data. The annual cycle's lowest points are sharp peaks each year in March. And the corresponding EOF patterns show the Arctic region. This pattern in the Arctic could be due to inconsistencies in modeling the global gravity field from GRACE measurements in the term for degree one, the static part.

In modes two and three we can find a trend signal. The trend signal is very noisy, as we have already seen it in the global analysis. The signs of the time series as well as the signs of the EOF patterns are the opposite of each other in both modes. In the synthesis step this would lead to a trend in the same direction. Both time series have similar characteristics in the noise that is imposed to the trend. In this imposed signal there could be another noisy annual signal included. In the EOF patterns in mode two, the main region is the Hudson Bay, but also a little slighter pattern around the coast of Greenland. In contrast, in mode three the dominant pattern is around the coast of Greenland with only a slighter pattern in the Hudson Bay. Even if these two modes seem to be very similar, they are uncorrelated, since EOFs and PCs are orthogonal by definition. At least the trend-imposed signal structure seems to be different between the modes. This could be due to leakage effects of land signal from Greenland to the ocean, but also from the North American continent to the Hudson Bay. Finally, note that in the EOF pattern of mode two the Sumatra earthquake for December 2004 can be clearly seen at the coast of Sumatra. Modes four and five do not show any visible signal content, they are considered as noise from the visual judgement.

EOFs and PCs for smoothing radius 350 km

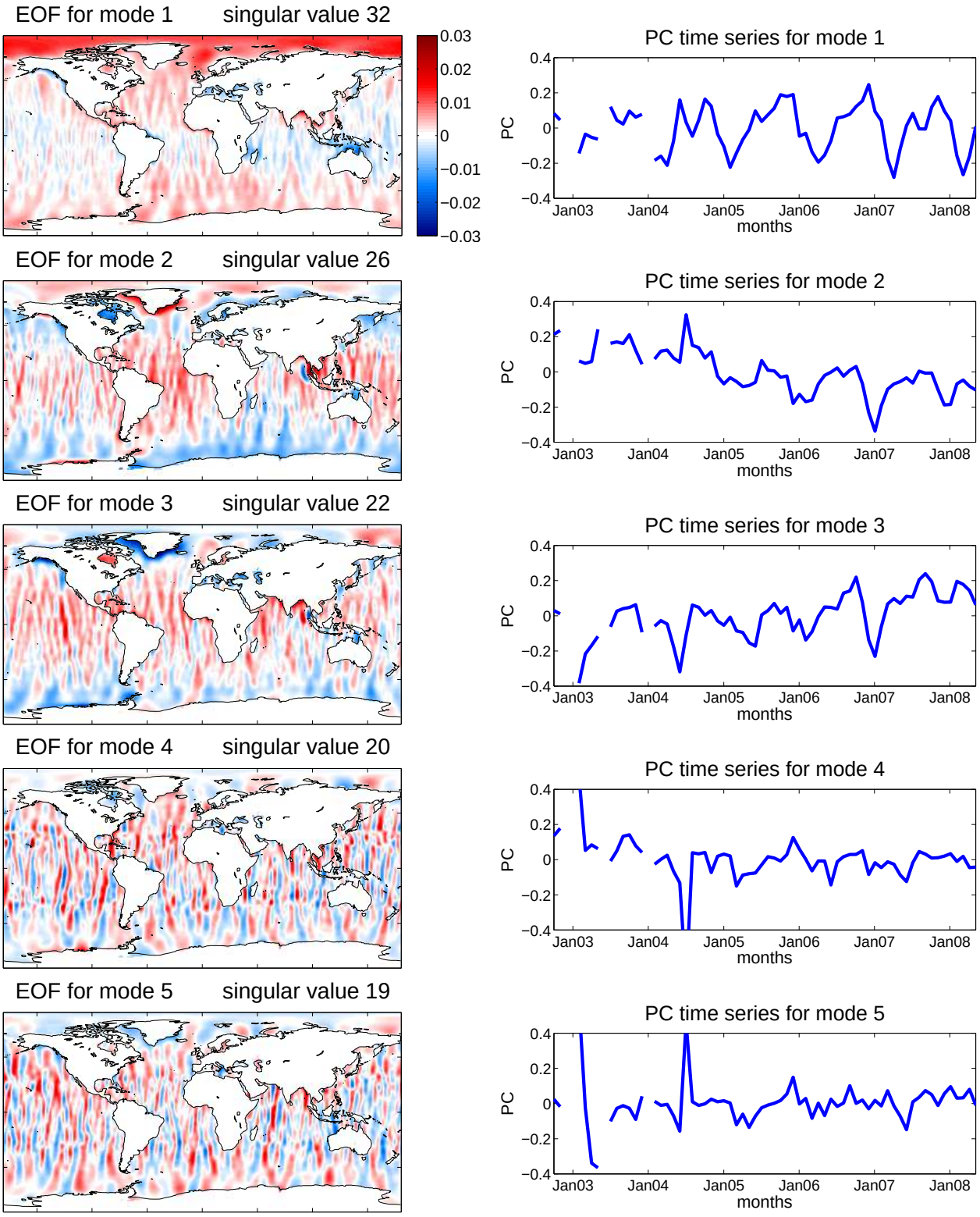


Figure 6.11: EOF patterns and PC time series for the first modes with a Gaussian smoothing radius of 350 km

EOFs and PCs for smoothing radius 400 km

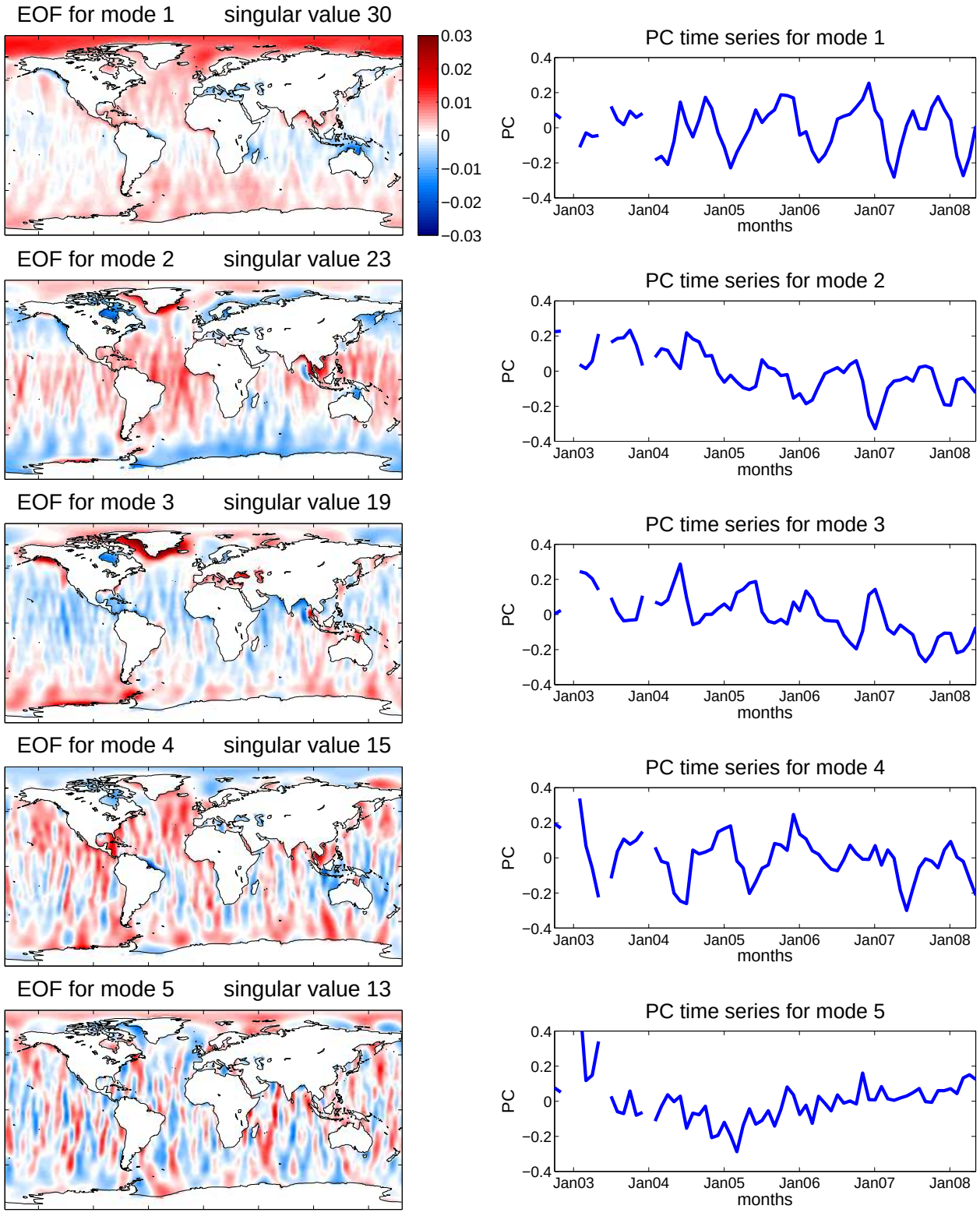


Figure 6.12: EOF patterns and PC time series for the first modes with a Gaussian smoothing radius of 400 km

EOFs and PCs for smoothing radius 500 km

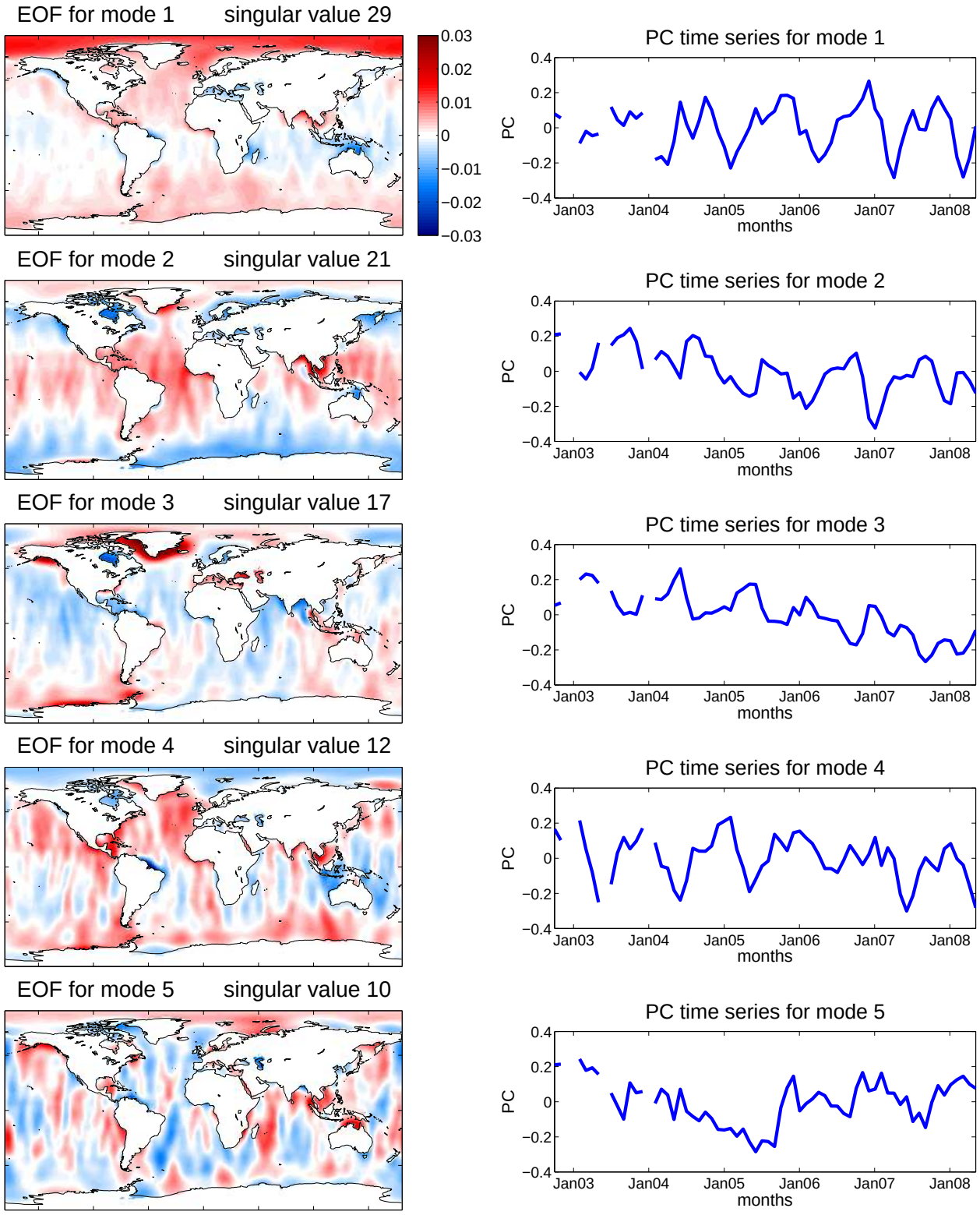


Figure 6.13: EOF patterns and PC time series for the first modes with a Gaussian smoothing radius of 500 km

The first modes for smoothing radius 400 km, see figure 6.12, show the same effects for modes one to three that have been already described for smoothing radius 350 km. Mode one has the annual signal in the Arctic, a bit more distinguished now, since there is more smoothing. In modes two and three, the two trends are still present, a bit less noisy than in the figure before, but still the trend in mode two has a stronger signal in the Hudson Bay, while the trend in mode three is more focused on Greenland's coast. Mode four now reveals an annual signal, too, with a low in each year's late spring. A region is very hard to find in the EOF pattern to which this temporal behavior belongs. There is a small pattern in Central America and another small pattern in the Gulf of Siam and at the coast of Thailand. But maybe this pattern is still due to the effect of the Sumatra earthquake. It could also be the case that this annual effect is caused by a continental drainage basin and the signal spreads itself into the measurements over the oceans. But this is only speculation. Finally, mode five does not show any signal structure in the PC time series and also no signal pattern in the EOF.

The next figure, figure 6.13, shows already smoothing radius 500 km. This large smoothing radius was chosen to show that more smoothing does not reveal any more signal. Of course, the modes, especially the EOF patterns become much smoother, but this does not show any additional signal. For modes one to four exactly the same interpretation holds, which was already given for 400 km smoothing radius. EOF patterns and PC time series for mode five do not have a characteristic white noise shape, but this mode is really hard to interpret, especially when we keep in mind that the larger smoothing radius can lead to low frequency effects which are not due to actual signal but to smoothed errors.

6.4 Selection of Modes and Reconstruction for Oceans Only

In the selection of modes for synthesis only two smoothing radii, 350 km and 500 km, shall be dealt with. Smoothing radius 350 km since it it the smallest smoothing radius that reveals any signal and smoothing radius 500 km as an example for a large smoothing radius since

we expect the signal from the oceans to be relatively noisy. These two examples should be enough to demonstrate the selection of modes. First of all, the singular value plots and the results of the KS hypothesis test are given in figures 6.14 and 6.15.

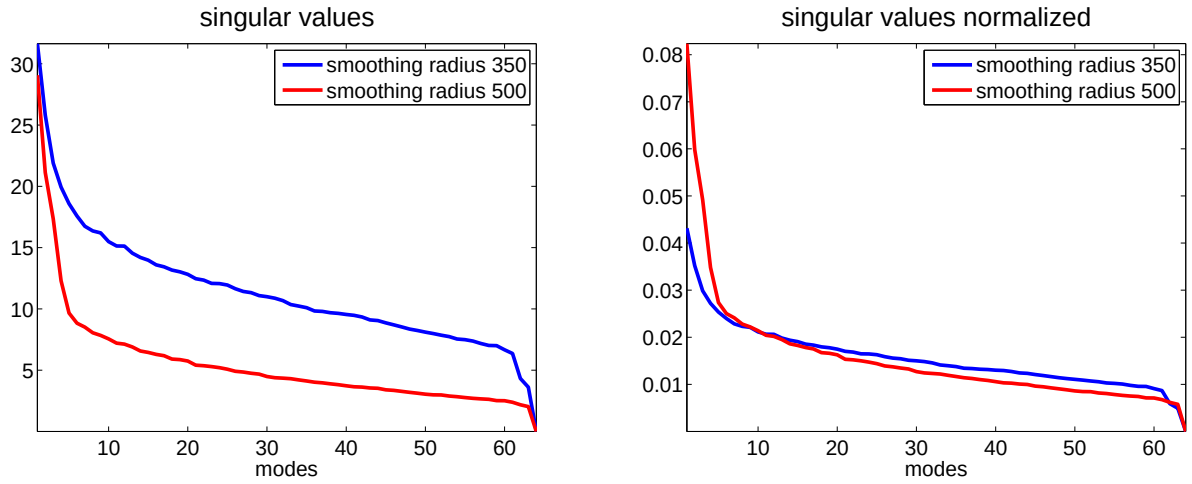


Figure 6.14: Singular values for different smoothing radii

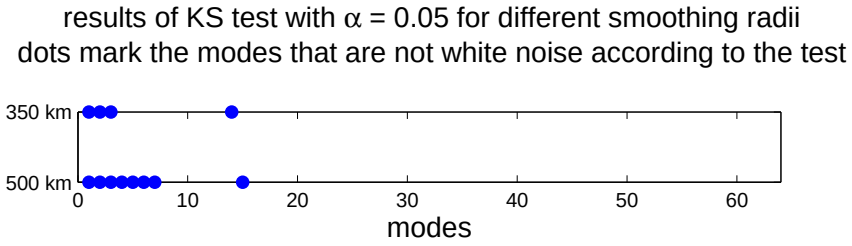


Figure 6.15: KS test results for $\alpha = 5\%$

According to the singular value plots, for smoothing radius 350 km a cut-off mode can be chosen at mode three and for smoothing radius 500 km at mode five. The KS hypothesis test gives modes one to three for radius 350 km to which we can agree as significant modes from visual judgement of the EOFs and PCs in figure 6.11. But the test also identifies mode fourteen as signal. This mode shall be investigated later in detail. For smoothing radius 500 km, modes one to seven and mode fifteen are considered as being significantly signal in the hypothesis test in figure 6.15. When looking at the first five modes' EOFs and PCs in

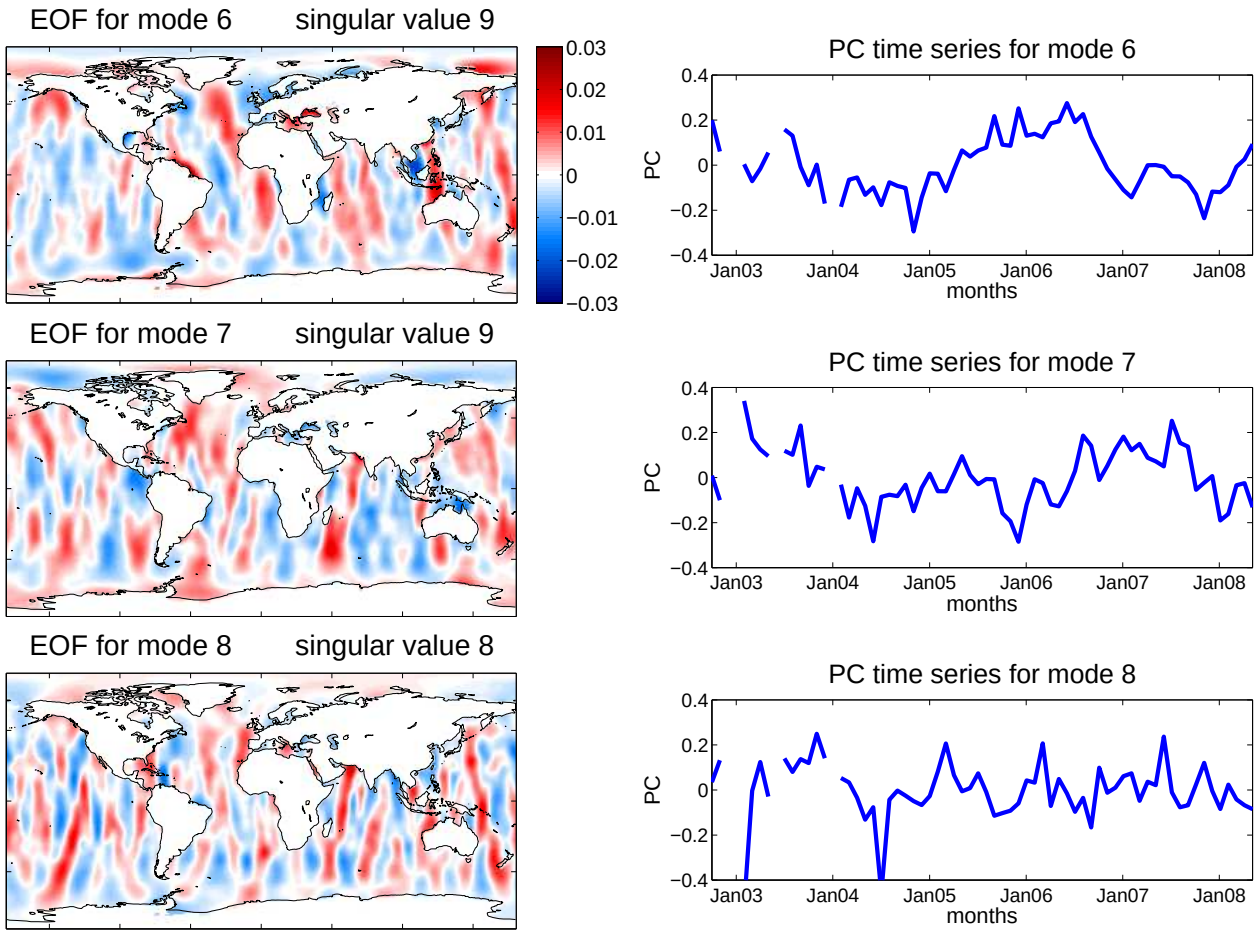


Figure 6.16: EOF patterns and PC time series for higher modes with a Gaussian smoothing radius of 500 km

figure 6.13, it seems understandable why those time series have passed the hypothesis test, even if it is hard to find a physical impact that could have caused that signal. Since modes six and seven passed the test, too, modes six to eight are plotted in figure 6.16.

From the time series of modes six to eight it is understandable why modes six and seven have been considered as being signal by the test. These two mode's time series show a certain structure, however it is again very hard to find any physical effect that could correspond to that modes, since there is no pattern in the EOFs to detect. Finally, we should study the test statistics for modes fourteen and fifteen for the two smoothing radii, respectively. Please refer to figure 6.17 for the KS test details and to figures 6.18 and 6.19 for their EOF patterns

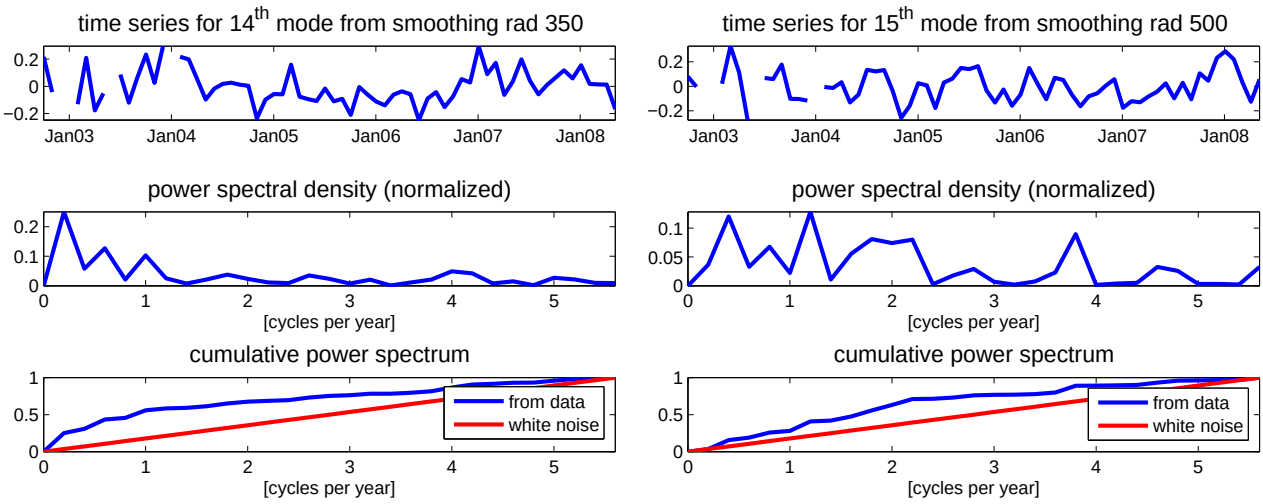


Figure 6.17: Details of Kolmogorov-Smirnov test for mode 14 for smoothing radius 350 km and mode 15 for smoothing radius 500 km

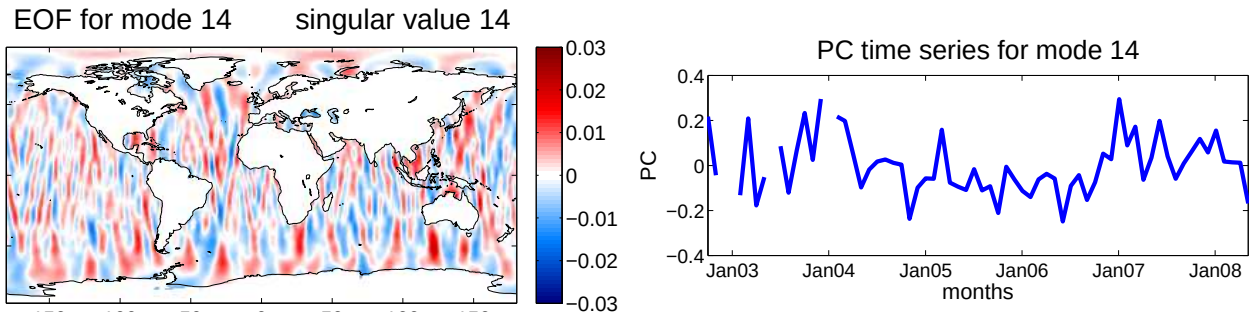


Figure 6.18: EOF patterns and PC time series for mode 14 with a Gaussian smoothing radius of 350 km

and PC time series.

Even if both time series look very noisy, they do have low frequency content that causes the hypothesis test to consider them as signal. Additionally, it is surprising that these two modes that were considered to be signal according to the test are so close together in the range of modes. But since the time series do not have anything in common, we must assume that these two modes do not have any signal in common, and are not due to any weaker signal that manifests itself only in higher modes. This would be another idea how to interpret these higher, and according the the KS hypothesis test signal-containing, modes. But this

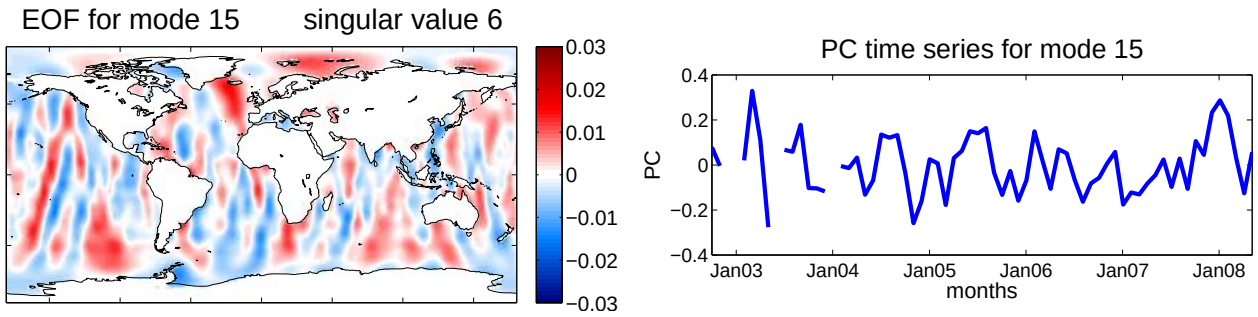


Figure 6.19: EOF patterns and PC time series for mode 15 with a Gaussian smoothing radius of 500 km

interpretation seems not suitable here.

Now, the reconstruction step is performed. For smoothing radius 350 km, modes one, two, and three are used in the synthesis according to the test, and higher modes are left out for reasons from combining the test results with all other considerations and information from the EOF, PC, and singular value plots. These were discussed previously in this subchapter. For smoothing radius 500 km it is not easy to decide on the number of modes. Modes one to seven were chosen due to the hypothesis test results. But this is an arbitrary choice. One could also choose less modes due to visual analyzation of EOFs and PCs and the singular values. Refer to figures 6.20 and 6.21 for the reconstruction plots. Again, the reconstructed maps for May 2008 are plotted on the left side and the example from the input for EOF analysis, also for May 2008, on the right side.

It is very surprising, how little signal is retained in the reconstruction. In figure 6.20 the reconstruction can hardly be seen. The signal before performing EOF analysis is relatively light and it was mentioned already that the signal over the oceans is very noisy. The signal content is much lower than for the continents. For smoothing radius 500 km modes one to seven were used for reconstruction, so the reconstructed signal's intensity is a bit higher. But the signal over the oceans before performing EOF analysis is already very light for such a large smoothing radius. More pre-smoothing reduces the magnitude retained in the data maps.

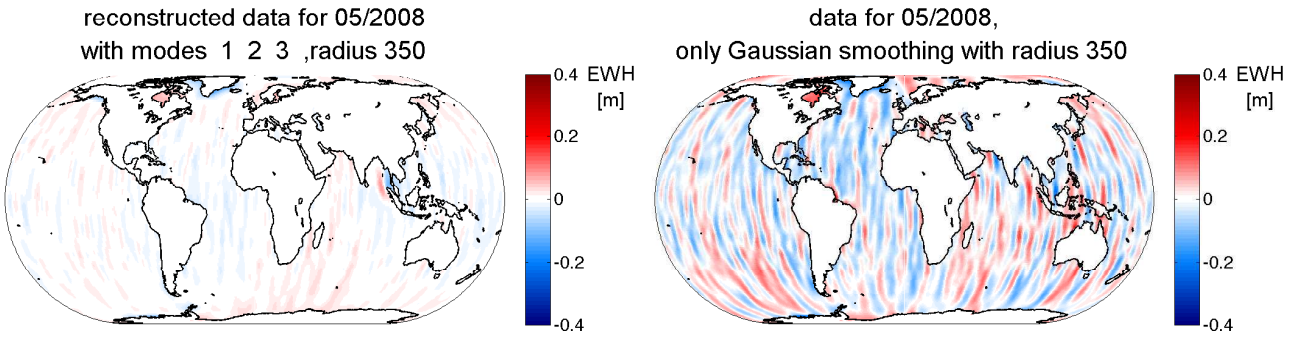


Figure 6.20: Reconstruction using the first three modes and original data for May 2008 with smoothing radius 350 km, in an orthographic Eckert IV Projection

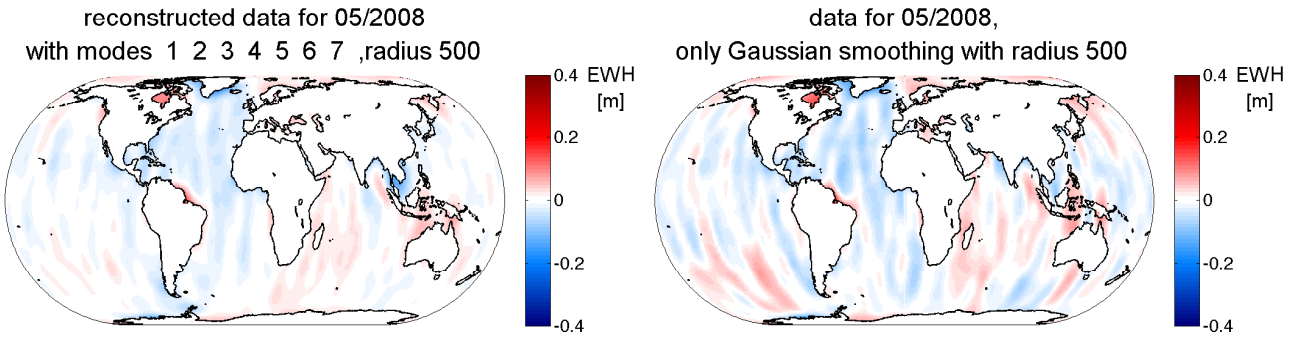


Figure 6.21: Reconstruction using the first seven modes and original data for May 2008 with smoothing radius 500 km, in an orthograhic Eckert IV Projection

All the reconstruction plots for May 2008, that are given throughout this thesis, have the same color scale. The color scale for these plots is not changed for reasons of comparability. These figures shall demonstrate that there is already a lot of magnitude filtered out.

Chapter 7

Regional EOF Analysis

After having used EOF analysis for decomposing global fields as well as continents and oceans separately, in this chapter it is demonstrated that EOF analysis can also be successfully used for regional analysis. The success of the method just depends on the variance structure of the signal in the data set. Any physical meaning is subject to interpretation of the modes that have been found from the EOF decomposition. As a regional example, on which to demonstrate the ability of the EOF method for regional analysis, Greenland is chosen, since the expected gravity signal is well-know and its characteristics have already been investigated.

In this chapter, the analysis and synthesis step are given and analyzed for the regional example of Greenland. Results are compared to the results from EOF analysis in the two previous chapters.

7.1 Regional Analysis of Greenland

For being able to perform EOF analysis only for the region of Greenland, a rectangular region situated between 55° North and 85° North, and 10° West and 75° West, is cut from the global data set, so that a new regional grid of the size (31×66) is obtained for each of the 64 months. This is the new data set that is reorganized in the form of the required data matrix to perform EOF analysis. The first modes from regional EOF analysis in Greenland for different smoothing radii are given in figures 7.1 to 7.4.

For performing a EOF analysis of Greenland in MATLAB, the same functions svd *or* svds*, which has been previously used for the complete data set, can be used again. The area*

of Greenland has to be taken from the global data grids as an input to this EOF analysis procedure. When plotting the EOF maps and results of the EOF analysis, an appropriate map projection should be used, since Greenland is far in the North.

In the first figure, 7.1, the first four modes for smoothing radius 200 km are plotted. Having a look at the patterns as well as at the time series, we can see that all the four modes are only representing noise. Also for regional analysis of Greenland, 200 km radius is not enough smoothing. In figure 7.2, for smoothing radius 250 km, the first signal appears. Again, for a 50 km smaller smoothing radius than in the global analysis, with the same smoothing radius as in the continental analysis. In the first mode we see the expected trend of mass loss in Greenland. The EOF pattern shows clearly the region over Greenland and the time series shows a well-recognizable downward trend. As expected, much smoother than the trend that was revealed in the global analysis. Modes two, three, and four show was well in the patterns as in the time series only noise and the typical longitudinal striping. An interesting observation is made here: The singular value for the first mode, which is 25, is significantly higher than the singular values of the other three modes, which are 17, 16, and 14. However, for smoothing radius 200 km, in figure 7.1, the first four singular values are 66, 62, 58, and 53. The differences between these singular values are all in the same order, exactly what would have been expected for modes that represent only noise, according to the selection methods based on the singular values, see chapter 3.2. In comparison, there is a large difference between the first singular value and the other three for smoothing radius 250 km. Again, exactly what would have been expected. The major variance comes from the trend signal, since noise variances are not that distinguished, which leads to a singular value for the signal-containing mode that differs significantly from the group of singular values of the noisy modes. The question arises, if more smoothing reveals more signals in the data. Therefore, the first modes for smoothing radii 300 km and 400 km are shown in figures 7.3 and 7.4.

For smoothing radius 300 km no other signal than the trend in the first mode can be found from visual interpretation of the EOF patterns and PC time series of the first four modes. The first singular value becomes even more outstanding for these higher smoothing radii. For 300 km smoothing the first singular value is 23, while the other three are 9, 8, and 7. And

EOFs and PCs for smoothing radius 200 km

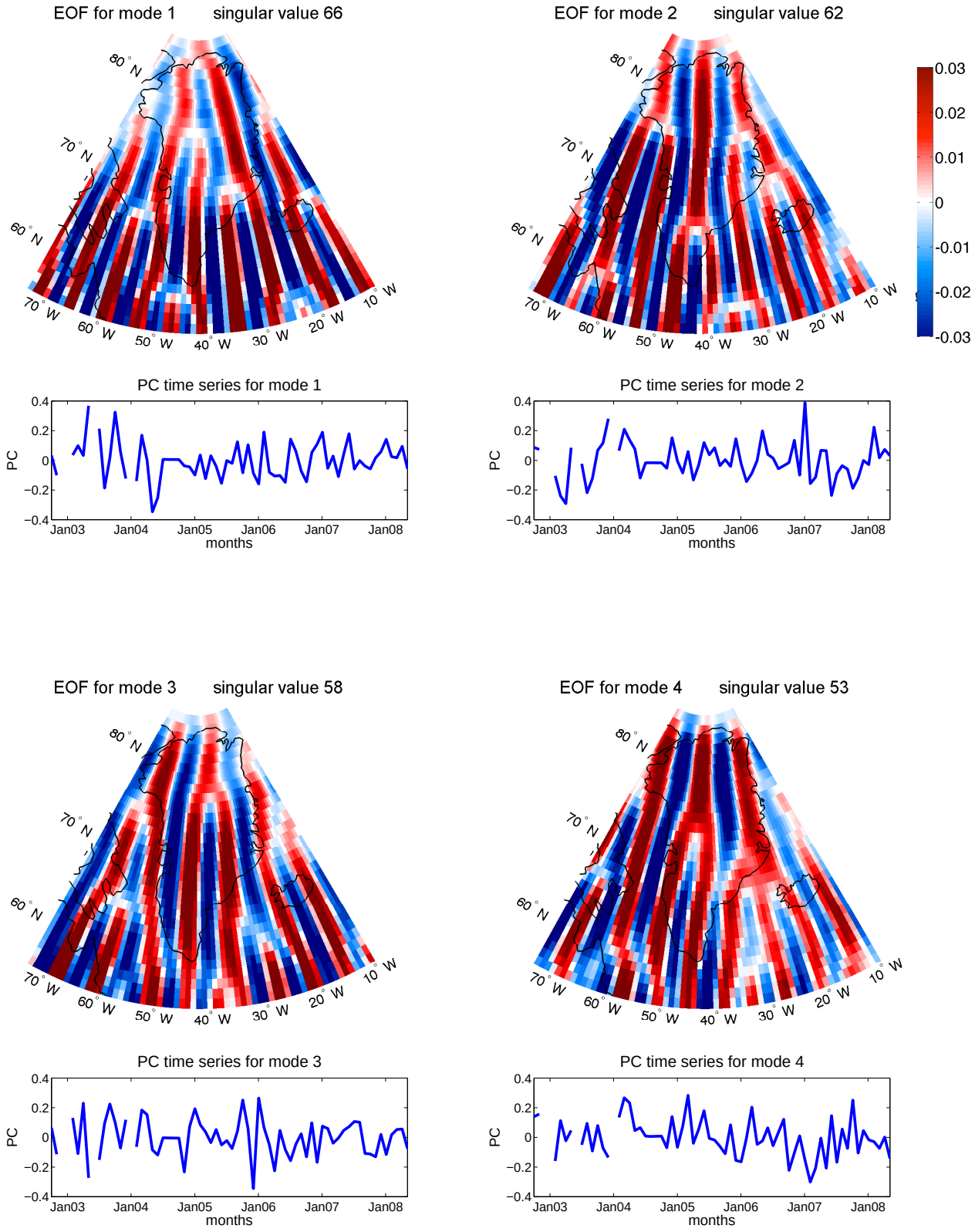


Figure 7.1: EOF patterns and PC time series for the region of Greenland with a Gaussian smoothing radius of 200 km, patterns are shown in an orthographic projection

EOFs and PCs for smoothing radius 250 km

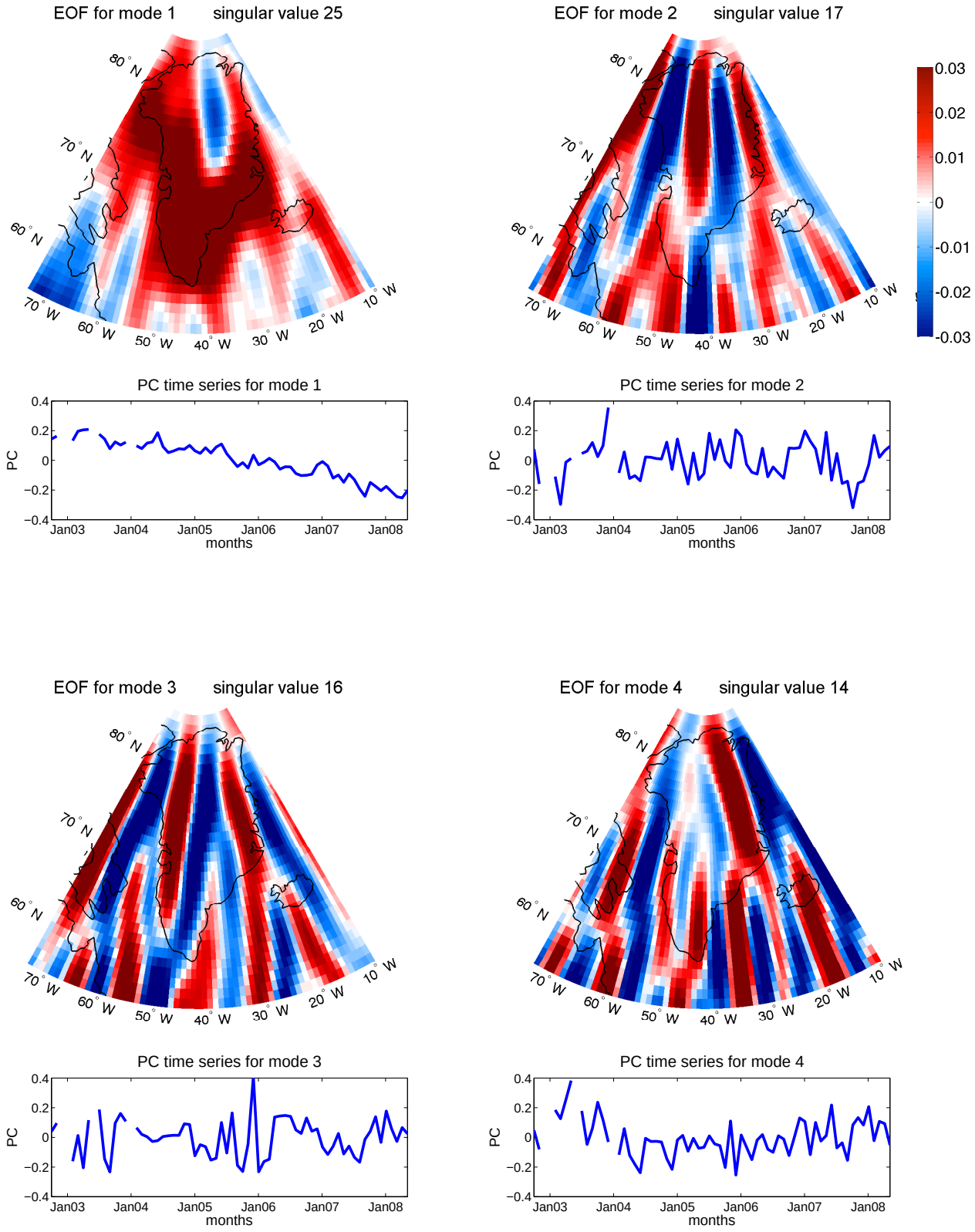


Figure 7.2: EOF patterns and PC time series for the region of Greenland with a Gaussian smoothing radius of 250 km, patterns are shown in an orthographic projection

EOFs and PCs for smoothing radius 300 km

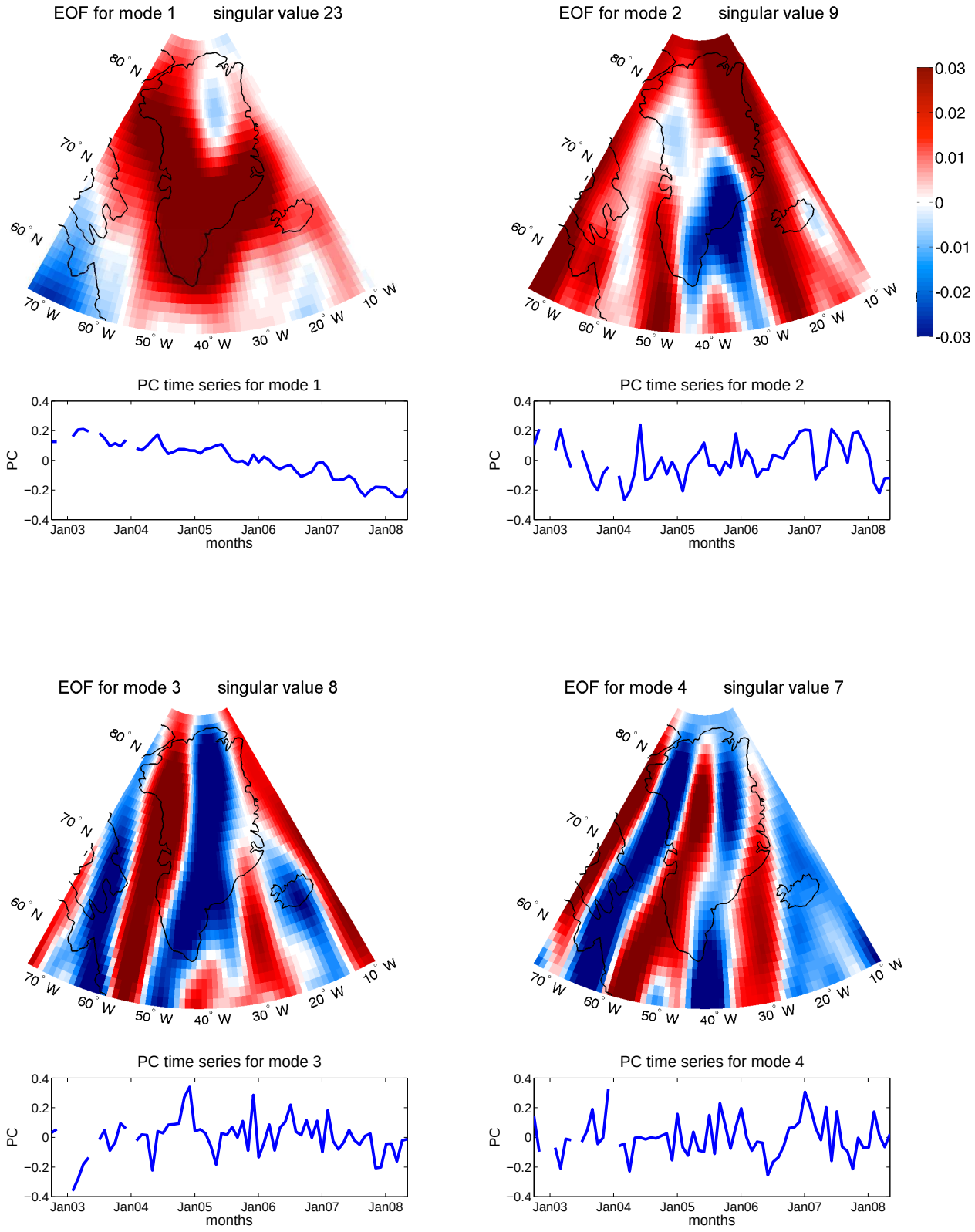


Figure 7.3: EOF patterns and PC time series for the region of Greenland with a Gaussian smoothing radius of 300 km, patterns are shown in an orthographic projection

EOFs and PCs for smoothing radius 400 km

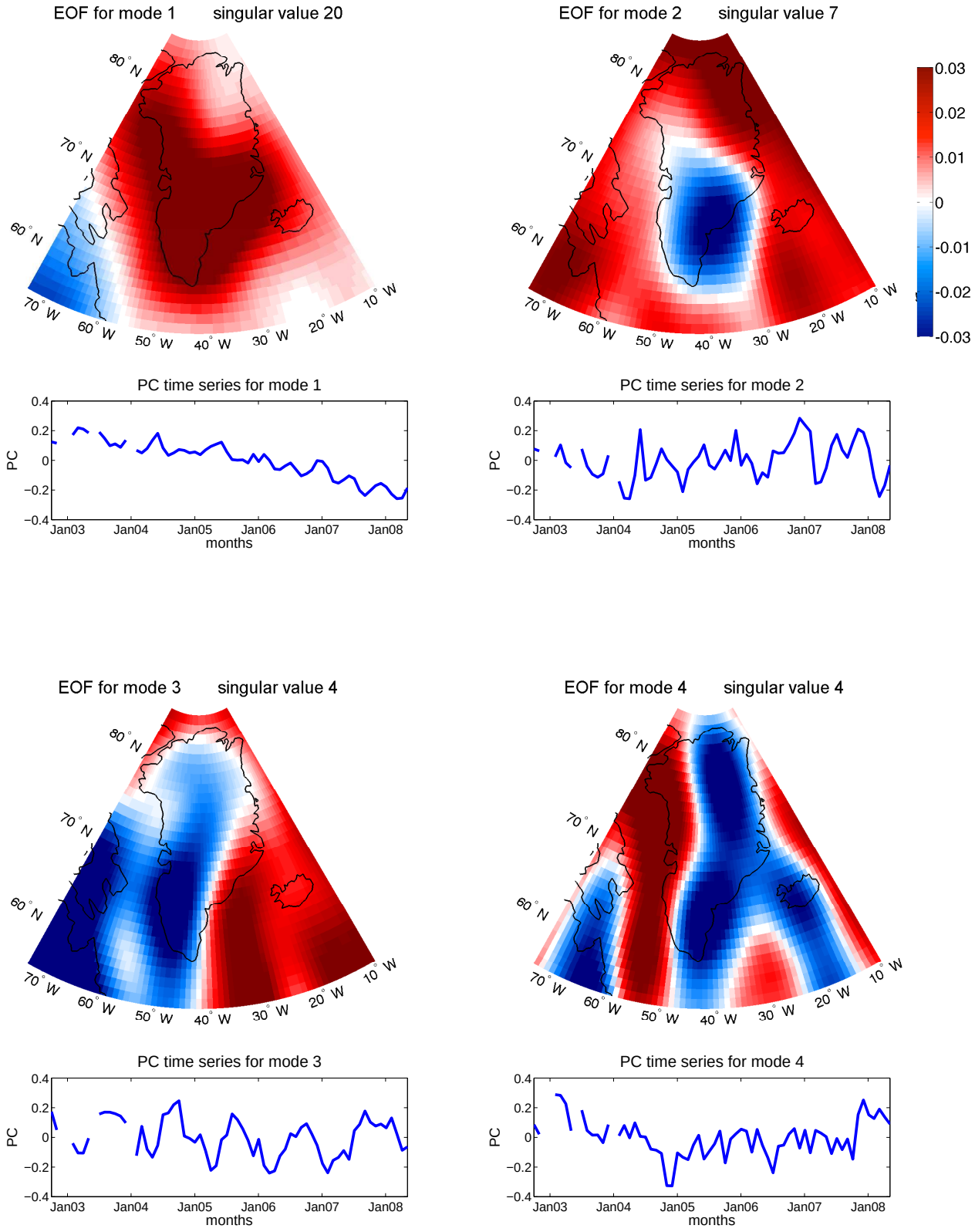


Figure 7.4: EOF patterns and PC time series for the region of Greenland with a Gaussian smoothing radius of 400 km, patterns are shown in an orthographic projection

for 400 km smoothing the first singular value is 20, while the others are 7, 4, and 4. This shows us, that the technique of selecting modes according to significant singular values can seriously work and be reliable in this regional example of Greenland and smoothing radius 300 km. For the global analysis but also for continents and oceans only judgement only from the singular values did not prove that useful. It was already said in chapter 3, when the selection methods were described, and we have found it here again, the dependence of the usefulness of a section method on the data set. While the singular value based methods did not prove that useful and reliable in the global and continental and oceanic analysis, it proves very useful and reliable here. Since it is one of the fastest methods it is always preferable, in case it is successfully applicable.

However, in the EOF patterns and PC time series of smoothing radius 400 km there seems to be an annular signal in mode three. For further investigation, a KS test is performed for the time series from EOF analysis of Greenland only. See figure 7.5 for the results of the KS test for two different significance levels.

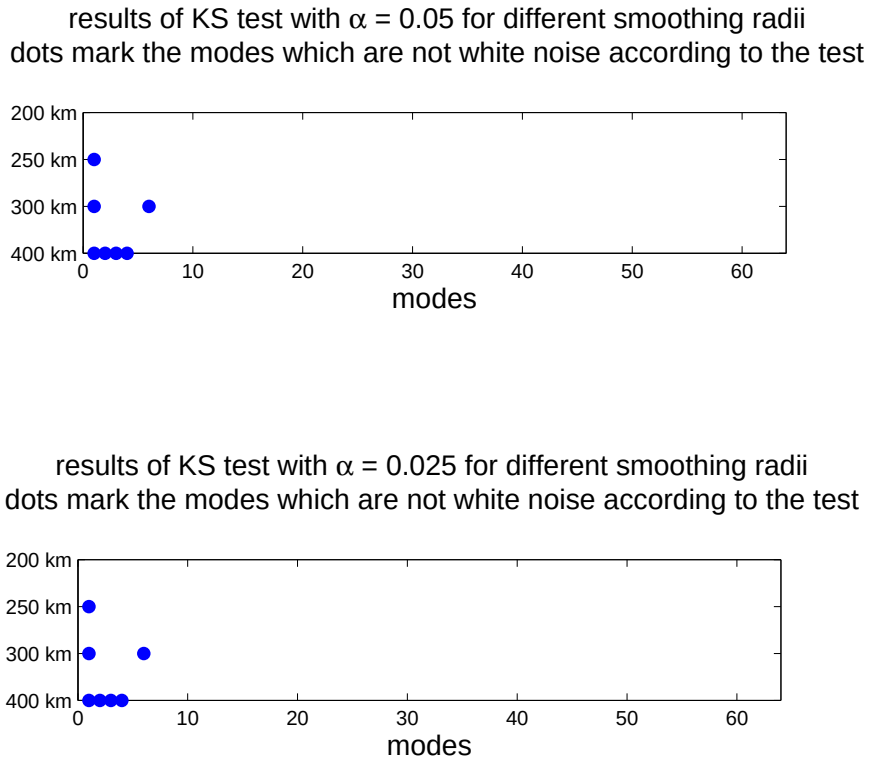


Figure 7.5: KS test results for $\alpha = 5\%$ (top) and $\alpha = 2.5\%$ (bottom)

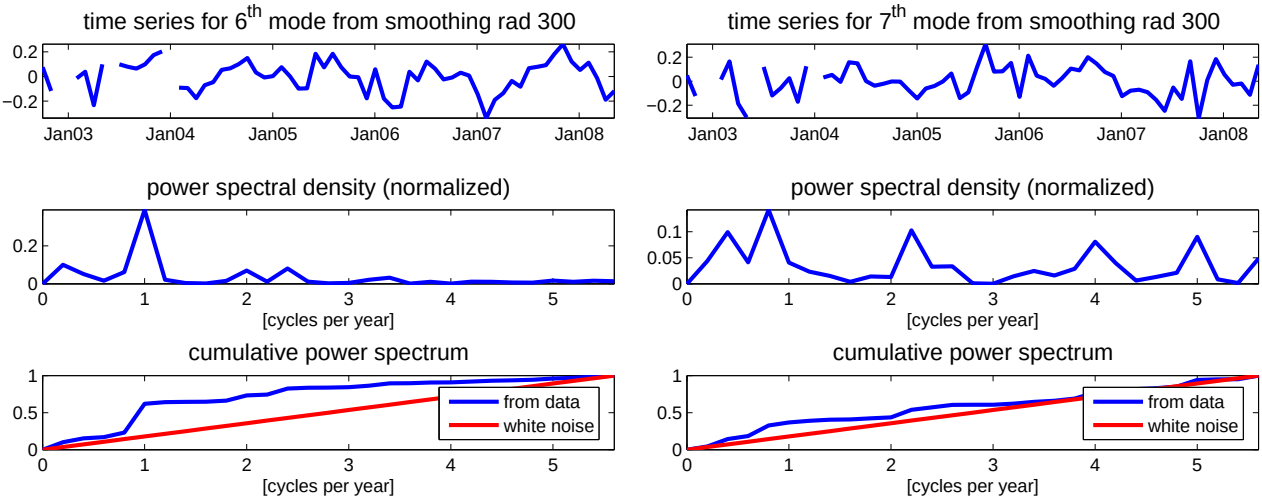


Figure 7.6: Details of Kolmogorov-Smirnov test for modes 6 and 7 for smoothing radius 300 km

KS test leads to exactly the same results for both significance levels. For smoothing radius 200 km the test results confirm what we have already seen in the EOF patterns and time series: no signal can be detected for this smoothing radius. 200 km radius is not enough pre-smoohting. For smoothing radius 250 km our findings are confirmed, too. Only mode one, which contains the trend signal, is different from white noise. From the plots of the first four EOF patterns and PC time series we would have expected the same for smoothing radius 300 km. But the KS test also gives modes six as significant. This mode shall be investigated in detail, see the plots of EOF patterns and PC time series in figure 7.7 for modes five to eight, end the details of KS test for modes six and seven in figure 7.6. Mode six was identified as signal by the KS test , and mode seven is plotted for comparison. In the time series of mode six, a noisy annual signal can be recognized. However, there is nothing interpretable in the EOF pattern of mode six. The details of KS test in figure 7.6 show the annual signal of mode six in the peak in the power spectral density. This is why mode six was interpreted as signal by KS test and mode seven, which has no such interpretable feature, as noise.

The first four modes from smoothing radius 400 km were interpreted to be signal according to the test. The first mode is again the expected trend, and both modes two and three show an annual signal in their time series, but it is very noisy. The PC time series of these two modes can be interpreted as seasonal changes in the ice masses in Greenland. But the

EOFs and PCs for smoothing radius 300 km

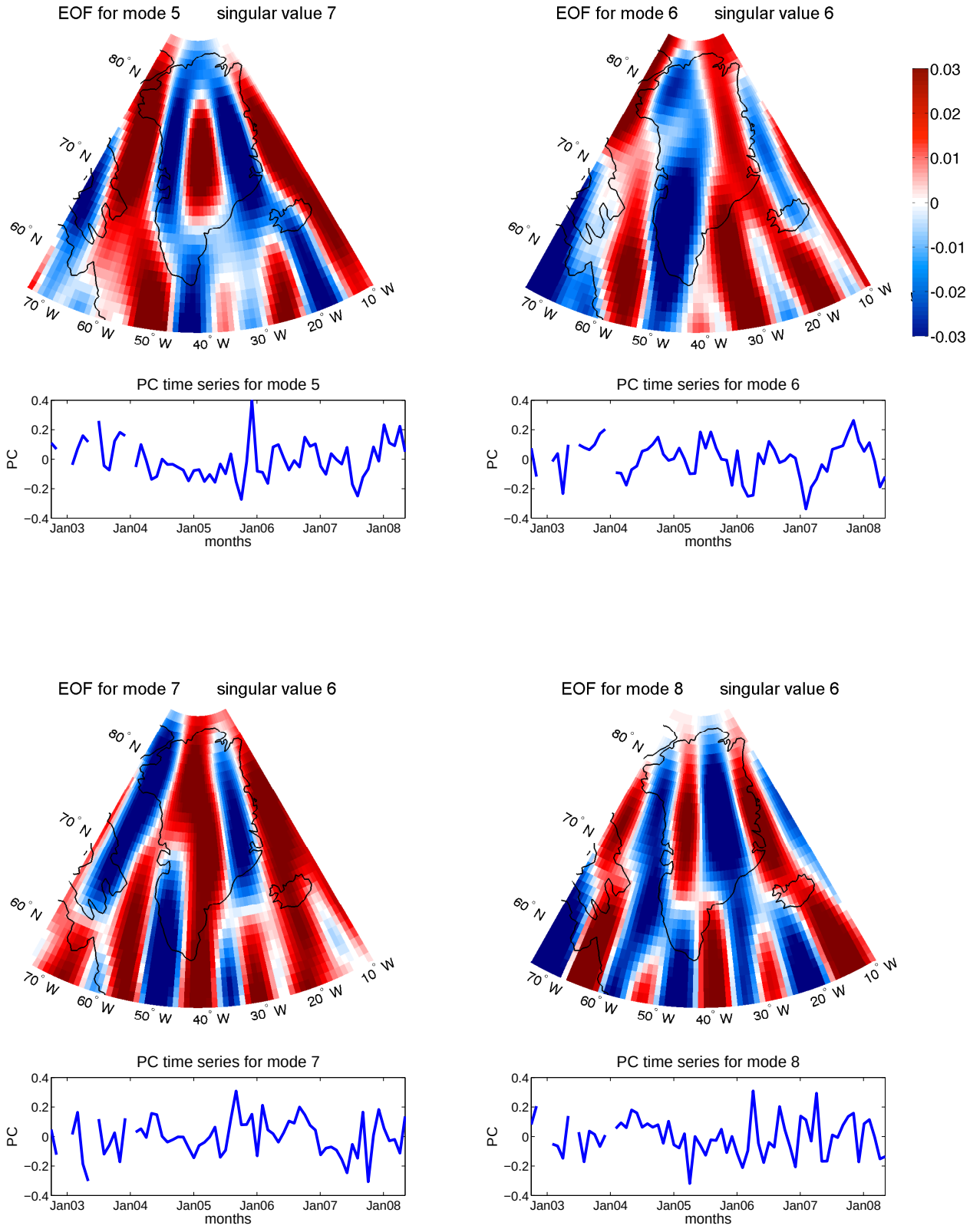


Figure 7.7: EOF patterns and PC time series modes five to eight for the region of Greenland with a Gaussian smoothing radius of 300 km, patterns are shown in an orthographic projection

corresponding patterns are difficult to interpret. Mode four is as well identified as signal by the test, but it is very hard to find a physical interpretation.

7.2 Synthesis of the Regional Signal from Greenland

Before synthesis can be performed, the number of modes to use for reconstruction needs to be determined. In the case of the regional example of Greenland this has already been done in the interpretation of the modes in the analysis step. When examining the EOFs, PCs and corresponding singular values in the previous subchapter, we have seen that the singular value based selection method is very applicable for Greenland for lower smoothing radii, such as 250 km and 300 km. And we have found from the singular values, also proven by visual examination of the EOFs and PCs, that starting from smoothing radius 250 km only the first mode is a signal mode. It might sound strange, but a reconstruction for the Greenland region is done by only using the first mode, since all the signal is contained in this mode. In terms of data dimensionality reduction, this EOF analysis was extremely rewarding, 63 modes of noise are skipped in the reconstruction. An example of the reconstructed and original data for May 2008 and smoothing radius 250 km is given in figure 7.8.

Both graphics have the same scale, so even if the signal in the reconstruction is weaker than in the original data set on the right in the figure, it is considerably strong, keeping in mind that the reconstruction is based on only one mode out of 64. This shows, that by decomposing the data set in EOF analysis, all trend signal is projected into the first mode. Since this trend in Greenland is a very strong signal compared to global gravity changes, it is not surprising the first mode is that dominant. Furthermore, we can assume that there is no other gravity signal in Greenland but this trend, so the difference between the two plots in figure 7.8 is just due to the noise in the data set. In the plot on the right, the stripes, caused by correlated errors, are dominating the image. However, in the plot on the left, almost all of the stripes have been filtered out, there are only miner effects from the stripes remaining. That means the major part of difference between the two plots are the stripes, which have been successfully filtered out by EOF analysis. In comparison to EOF results from the global and continental data sets, filtering of the stripes performed better in this

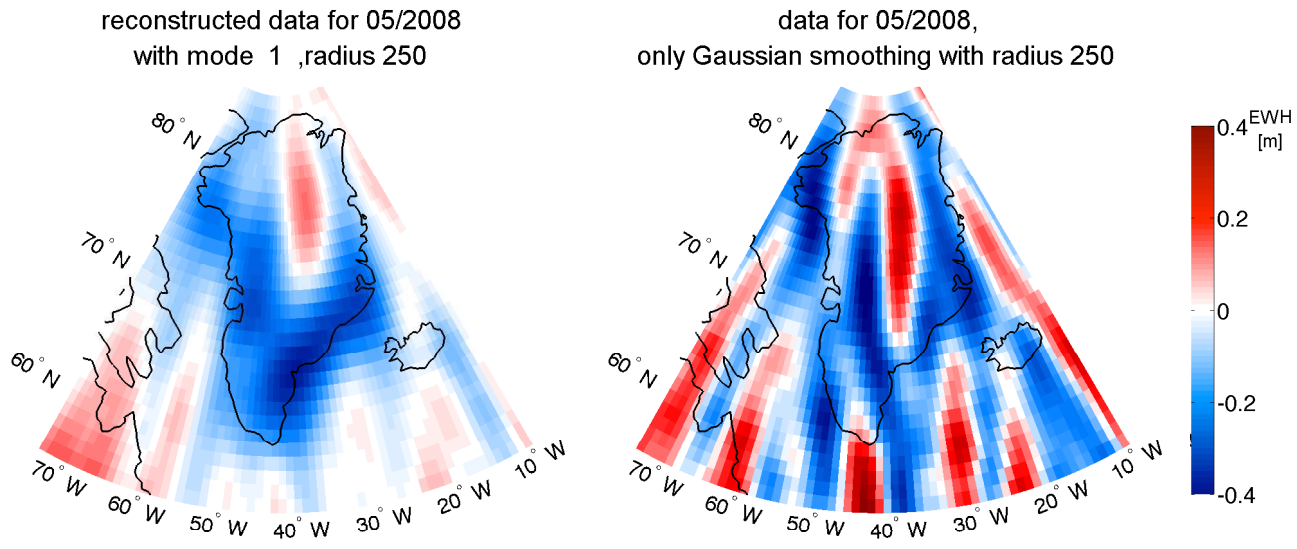


Figure 7.8: Reconstructed data using only the first mode and original data for May 2008 with smoothing radius 250 km, in an orthographic projection

regional example. This is due to the well-determined trend signal that is comparably strong and therefore clearly detectable in the EOF analysis.

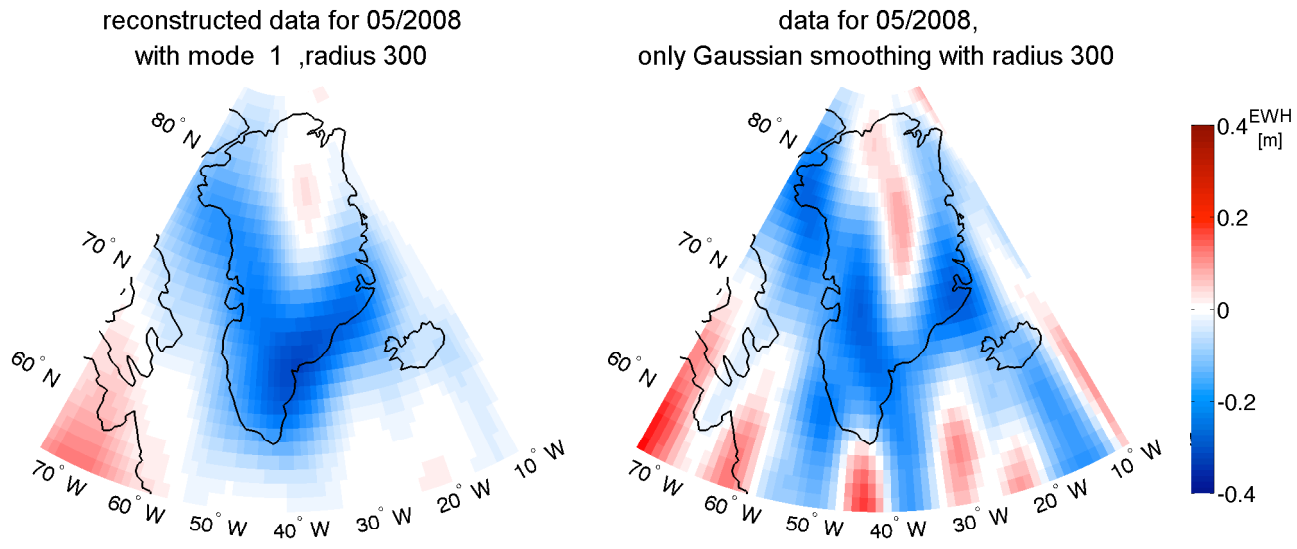


Figure 7.9: Reconstructed data using only the first mode and original data for May 2008 with smoothing radius 300 km, in an orthographic projection

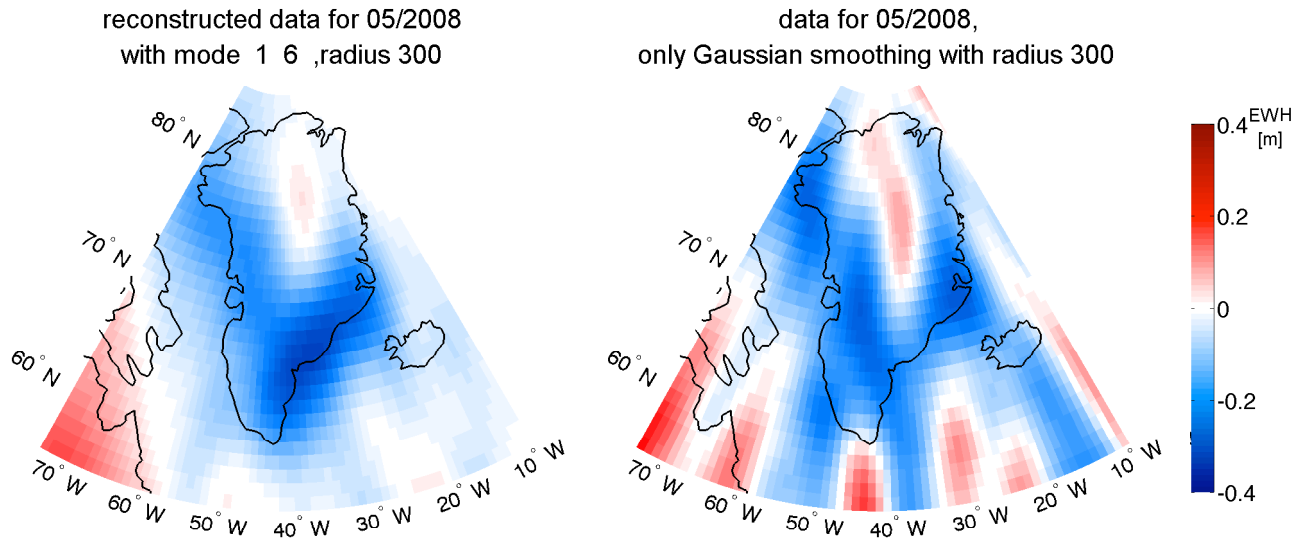


Figure 7.10: Reconstructed data using the first and sixth mode and original data for May 2008 with smoothing radius 300 km, in an orthographic projection

For smoothing radius 300 km mode six showed an annual signal in the time series, while modes two to five were considered as noise. The expected effect in EOF analysis was that the modes containing signal would be projected into the first few modes, followed by higher modes which represent noise. It is unusual that the annual signal only appears in mode six. For comparison, two reconstruction examples are plotted for smoothing radius 300 km. First, only modes one used for synthesis and second, modes one and six. See figures 7.9 and 7.10. Not much difference can be seen between the two reconstructed fields in Greenland. The intensities in the reconstruction with only mode one even seem to be a bit higher than in the reconstruction with modes one and six. Probably the annual effect from mode six which is superimposed on the trend from mode one in the synthesis step attenuates the trend signal. Concluding, from the visual judgement of this reconstruction, only using mode one would be sufficient. Compared to the data before EOF analysis, the North-South striping was significantly reduced in EOF analysis.

There is another remarkable effect that was not clearly evident in the plots of the individual modes, but it is the more in the reconstruction. When reconstructing the data for smoothing radius 400 km, and plotting again as an example month May 2008, see figure 7.11, there is not much difference in the intensities of the signal between the reconstructed

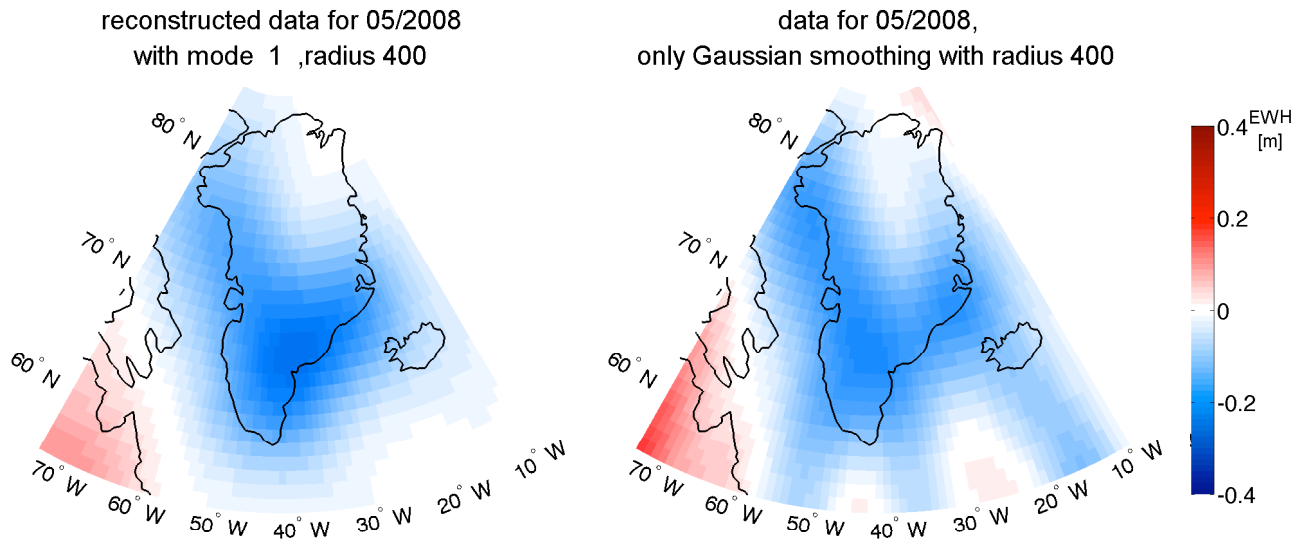


Figure 7.11: Reconstructed data using only the first mode and original data for May 2008 with smoothing radius 400 km, in an orthographic projection

data and the original data. Just the remnants of the North-South stripes are removed by EOF analysis, and the location of the minimum of the signal is changed. Due to the large smoothing radius of 400 km compared to the size of the region of Greenland, that is in the order of a few thousands of square kilometers, a lot of the striping has already been removed by pre-smoothing. It is remarkable, that EOF analysis could even filter out the remnants of the North-South stripes, even if they were no more that dominant.

400 km smoothing radius was probably too much smoothing for this specific area. EOF analysis leads to very good results already for smoothing radii as low as 250 km for detecting the main trend signal. But it is remarkable that in radius 300 km an annual signal appeared in mode six. Again we find that an appropriate radius for Gaussian pre-smoothing is a sensitive and very important parameter to choose.

Chapter 8

Conclusions and Outlook

In this concluding chapter, the results from the work presented in this thesis are briefly summed up and conclusions are drawn. Some options for further research are proposed and in the second subchapter three EOF-related methods are presented. While the first method, the pre-whitening transformation, seems very promising for pre-filtering, the other two methods, multiple channel singular spectrum analysis and canonical correlation analysis, share similar mathematics with EOF analysis. The basic idea behind one of them is an augmentation of the data matrix, and the other one is applied to compare two different data matrices.

8.1 Summary of Results and Conclusions

The advantages and shortcomings of EOF analysis for GRACE gravity data were studied in this thesis. In chapter 5, details of the analysis procedure and the results for analyzing monthly global maps of gravity changes from GRACE are shown. Chapter 6 shows the studies of EOF analysis of continents and oceans separately and chapter 7 gives the results of EOF analysis on a selected region, where Greenland was chosen as an example. In the regional examples, which are Greenland and the continents and oceans only, different regions gave a new truncated data set with different characteristics concerning size and signal quality to be studied.

Preliminary, it has to be pointed out that the performance of EOF analysis strongly depends on the noise level and noise structure in the input data set. Gaussian smoothing as a pre-filter plays an important role. Since EOF analysis depends on the separability of the signal variance structure from noise variance, EOF analysis cannot be successfully applied

on non-filtered GRACE data. Of course, the mathematical decomposition is always possible, but it would not reveal any signal containing modes. An interesting effect of Gaussian pre-smoothing can be seen in the analysis of global maps. A smoothing radius of at least 300 km was necessary to detect signal at all, and a 350 km smoothing radius revealed a signal that could not be detected yet in the results of radius 300 km. Generally, we can conclude that more smoothing led to more distinguished signals in the modes. However, very large smoothing radii also led to modes that were not significantly due to noise but also not identifiable as being due to any physical effect. With increasing smoothing radii more signal effect could be detected, but at a certain smoothing radius the modes following those, that could be identified as being due to actual physical effect, became relatively smooth in time series as well as patterns, because of large pre-smoothing radii. It is very hard, if possible at all, to distinguish between these low frequency modes, which are most possible due to error, and actual signal modes.

The least smoothing radii that is necessary strongly depends on the significance of the signal variance in contrast to the noise variance. For the Greenland region and the continents, smaller smoothing radii were sufficient, since the signal in these regions was more dominating compared to the noise. For the oceans only, 100 km larger smoothing radii were necessary to detect the first signal. Of course, smoothing is applied to all the data, and smoothing always causes a loss of signal, too. Since the signal over the oceans, that is not that strong compared to the continental signals, is already hard to find, the large smoothing radii could also smooth out a lot of it. The question is to find a smoothing radius that is large enough to suppress enough errors to allow the following EOF analysis to find signal, but on the other hand not too large, since signal that has been smoothed out cannot be detected anymore. Depending on the region to be analyzed, the range of smoothing radii where a meaningful EOF analysis can be performed can be very small. Greenland was an example where the signal was strong and distinguished enough for being separated from noise by EOF analysis even for small smoothing radii, but also strong enough for not getting lost when too much smoothing is implied. However, this does not hold for data sets where the signal is not that strong. The example of analyzing the oceans only is one of these cases where signal detectability is very sensitive to the pre-smoothing radius. Too much or too less smoothing does not allow EOF analysis to detect the signal content. From all the examples we can draw the conclusion that

EOF analysis is capable of detecting major gravity signal from GRACE data after Gaussian pre-smoothing of about 300 km. However, the EOF method has difficulties to deal with the longitudinal stripes in the GRACE gravity maps. When the error variances of the stripes are on the same level of intensity and distinctiveness as the signal variances, the signal cannot be detected by EOF analysis. Different regions as data examples have been studied and hereby, the oceans-only analysis is considered as an example for regionally restricted EOF analysis for a data set where no strong signal is expected, compared to Greenland as an example for excellent signal quality. The contribution of this thesis is to give an overview of the EOF analysis method and study the method's ability of analyzing GRACE data globally and regionally. It might be interesting to perform EOF analysis on more different smaller regions to see if it is possible by just taking a part of the data to get a better signal-to-noise ratio and maybe find another signal that has not been detected so far. The distinctiveness, with which the trend signal in Greenland has been detected in the regional analysis in contrast to the global analysis where the trend time series was extremely noisy, leads to the idea that maybe analyzing other particular regions could reveal another signal. If the global map is divided into several smaller regions to be analyzed, maybe EOF analysis can be able to detect additional mass change signals. Of course, the size of these regions and their arrangement would play an important role, since the regions have to be cut in a way that in the new smaller data matrix the signal variance is improved compared to the new data set's noise variance.

Different techniques for selecting the signal-containing modes were proposed and applied to the data. From the different examples the conclusion is drawn that it strongly depends on the area that is analyzed, and sometimes even on the radius of Gaussian pre-smoothing, which method is useful and leads to reliable results. Unfortunately, none of the selection methods proved reliable, there are very many cases where the outcomes of the section methods, from singular value based methods as well as from the Kolmogorov-Smirnov hypothesis test, had to be checked and sometimes were doubted from the visual interpretation of the corresponding EOFs and PCs. In the results from large pre-smoothing radii it becomes more and more difficult to decide whether a mode contains signal or noise. Since the errors get smoother, they are no more distinguishable by resembling to white noise. Too much smoothing leads to low frequencies caused by smoothed errors. And these can no more be identified

as noise by a white noise test for time series of only 68 months. Due to the difficulties in the selection of signal-containing modes, EOF analysis for GRACE data can not be implemented as a filter in the form of a black box with just an input and a filtered output signal. A reasonable choice of modes is difficult and needs to be adapted individually for the data set and modes. The user's expertise and judgement is required. In most cases different methods need to be compared to draw a reliable conclusion from combined results.

Comparing the results in chapters 5, 6, and 7 it can be seen that EOF analysis is successfully applied to regions of different sizes. It has been demonstrated that the size of the region in the input data set to be analyzed does not matter. Only the signal and noise variance content matters. This is a major strength of the method. Even more general, EOF analysis can be applied to any multi-dimensional data set of measurements over a certain period of time. A very obvious other possibility for analyzing GRACE gravity data is performing EOF analysis on the spherical harmonic coefficients instead of on the synthesized fields. Analyzing the spherical harmonic coefficients with EOF analysis was already described by (Wouters and Schrama, 2007) and a first comparison of EOF analysis on the spherical harmonic coefficients and on the synthesized fields is given in (Iran Pour et al., 2009). When analyzing spherical harmonic coefficients by EOF analysis, the filtering results, concerning the correlated errors, are much better than the filtering results from EOF analysis being performed on the synthesized fields as done in this thesis. However, for example the signal from the Sumatra earthquake, which could be found in the analysis, was lost when performing EOF analysis on the coefficients. Again, the typical problem in analyzing GRACE gravity changes arises: The difficulty to distinguish between correlated errors and signal.

The advantage of EOF analysis compared to other methods for analyzing GRACE data is that EOF analysis does not search for any pre-defined signal or pre-defined type of signal, as a lot of other methods do. Anything that is distinguishable from white noise is projected by the method into a different mode. Of course, colored noise, as the correlated errors in GRACE, cause problems. It is important to keep in mind that EOF analysis is basically a transformation of the data set into a new coordinate frame, where the new axes are aligned along the data variances. So in the process of EOF analysis itself, no signal content is lost. By transforming the data into a new coordinate frame we just hope to see more of the con-

tained signal by looking at the data from the new point of view, that is the new coordinate frame. The worst that can happen in the analysis procedure is that a signal's variance is not distinguished enough for being assigned an individual axis in the new frame, that is an EOF, and therefore not being visible. Problems arise when selecting the modes that contain signal and identifying signal that is actually caused by a physical impact. Here is where more investigation could be useful, in the physical interpretation of the modes.

For selecting the signal modes, the method of the KS hypothesis test proved to be a more reliable method to decide on the number of modes than just relying on the singular values. However, one of the difficulties of the test is that some modes were interpreted as signal, since the structure of the time series was different from white noise, even if it was extremely hard to decide on their physical interpretation. There were a couple of cases where very high modes were interpreted as signal by the test. In fact, these time series showed very high frequencies, their cumulative power spectral density curves were below the white noise line. They were even worse than white noise. The KS test should be improved in a way that these modes are no more identified as noise. In future work, a better hypothesis test should be developed, which leads to more reliable results. Maybe such a test could also include information from the corresponding EOF patterns.

Summing up, EOF analysis as a filtering tool leads to better results than just smoothing the data, as it has been shown in different examples. And by decomposing GRACE data with the EOF method, some insight in the data could be gained. Besides all difficulties, it revealed interesting details in the GRACE data, and quite a few signals have been identified to be due to physical effects. EOF analysis is for sure not the perfect method to solve all troubles when analyzing GRACE data, but it is a remarkable method that performs very well in some cases but also has some shortcomings that were described and analyzed. Since EOF analysis is just a mathematical decomposition, it could be combined with any other analysis technique and maybe lead in combination to better results than just performing EOF analysis on pre-smoothed GRACE data.

8.2 Outlook: EOF-related Techniques

Pre-Whitening Transformation

In (Allen and Smith, 1997) a method is proposed to pre-filter the data before performing EOF analysis that helps to identify the signal part in a data set that consists of signal and noise. EOF analysis is capable of separating a signal from white noise due to its variance structure. However, in the presence of any other colored noise, the signal variances have to exceed the noise variances for being able to be detected. This is one of the limitations of EOF analysis that we have seen in the results of analyzing GRACE data. Gaussian smoothing with a radius of at least 250 km had to be applied before performing EOF analysis to suppress enough of the high frequency correlated errors for being able to detect the signal parts in the data. The idea in (Allen and Smith, 1997) is to perform a so-called pre-whitening transformation on the data under the use of a known noise covariance matrix. In case the noise covariance is known, this transformation could prevent genuine signals from being obscured with high-variance noise components.

The EOFs are eigenvectors of the scatter matrix, that is a linear multiple of the data covariance matrix, as it was shown in chapter 2.

$$\mathbf{Ce} = \lambda\mathbf{e} \tag{8.1}$$

$$\text{or} \quad (\mathbf{C} - \lambda\mathbf{I})\mathbf{e} = 0 \tag{8.2}$$

hereby, $\mathbf{C}$ is the data covariance matrix and λ and $\mathbf{e}$ the eigenvalues and eigenvectors, respectively.

The data covariance contains signal and noise, so it can be denoted as $\mathbf{C} = \mathbf{C_S} + \mathbf{C_N}$ where $\mathbf{C_S}$ is the signal covariance part and $\mathbf{C_N}$ the noise covariance part. If the noise is white noise, the noise covariance matrix has the form $\mathbf{C_N} = \sigma^2\mathbf{I}$. And for the noise being white noise, the eigenvectors of the data covariance matrix $\mathbf{C}$ are the same as the eigenvectors of the signal covariance matrix $\mathbf{C_S}$ since

$$(\mathbf{C} - \lambda\mathbf{I})\mathbf{e} = 0 \tag{8.3}$$

$$(\mathbf{C_S} + \mathbf{C_N} - \lambda\mathbf{I})\mathbf{e} = 0 \tag{8.4}$$

$$(\mathbf{C_S} + \sigma^2\mathbf{I} - \lambda\mathbf{I})\mathbf{e} = 0 \tag{8.5}$$

$$(\mathbf{C_S} - (\lambda - \sigma^2)\mathbf{I})\mathbf{e} = 0. \tag{8.6}$$

Only the eigenvalues decrease by σ^2. If the noise is only white noise, EOF analysis can optimally separate signals into individual modes and the reconstructed data set has maximum signal-to-noise ratio. But this fact shall no longer bother, it was just mentioned as interesting side note.

Unfortunately, the property of the eigenvectors, that are the EOFs or PCs in the EOF analysis, as shown in (8.6) only holds if the noise is pure white noise. In general, the eigenbasis of the sum of two matrices does not share eigenvectors with either of the two constituent matrices. Only when the noise is white noise, that means it has equal variance in all directions, the eigenvectors of the data covariance matrix are the same as the eigenvectors of the signal covariance matrix. So the idea is, that if the noise covariance matrix is known, a coordinate transformation can be performed, so that the noise in the new coordinates is only white noise, that is that the noise covariance matrix is the identity matrix. Let us assume that the noise covariance $\mathbf{C_N}$ is known. A coordinate transformation into a new frame where the noise part is uncorrelated, that means where the noise part is reduced to white noise, is performed by choosing the eigenvectors of the noise covariance matrix as new basis. That means the new coordinates are obtained by multiplying the data $\mathbf{D}$ with the matrix containing the eigenvectors of the noise covariance matrix $\mathbf{E_N}$. The transformed data matrix would be

$$\mathbf{D}' = \mathbf{D}\mathbf{E_N} \tag{8.7}$$

and the transformed covariance matrix

$$\mathbf{C}' = \mathbf{E_N}^T \mathbf{C} \mathbf{E_N}. \tag{8.8}$$

EOF analysis would be performed for the transformed data set, and if the covariance matrix of the noise was exactly known, the signal variances could be found since the white noise scatter matrix is the identity matrix. From a theoretical point of view, this technique could be applied successfully to GRACE gravity data, since the noise covariance matrix of the correlated errors can be determined. The main reason why such large radii in Gaussian pre-smoothing were necessary are the correlated errors. The idea is, that if the correlated errors could be reduced by such a pre-whitening transformation, much less pre-smoothing would be enough for meaningful results from EOF analysis. Of course, all depends on the quality of an error covariance matrix. But if the error covariance matrix is known precisely enough, the

pre-whitening transformation might lead to better results from EOF analysis, since smoothing causes a loss of signal. Smoothing treats all components in the data, that means also the signal, while the pre-whitening transformation would only reduce the correlated errors.

Multiple Channel Singular Spectrum Ananlysis

The technique of multiple channel singular spectrum analysis (MSSA) has been proposed by different authors as a tool for analyzing space and time correlated data fields. Background of the technique can be found in (Allen and Robertson, 1996) and (Plaut and Vautard, 1993) but also in (Robertson, 1995). While in EOF analysis dominant variance structures are identified, spatial patterns dominating the variability for the EOF and dominating time variance structures for the PCs. The EOFs and PCs are the eigenvectors of the spatial and, respectively, temporal scatter matrix. Since geophysical and meteorological data also often have space-time correlated patterns, MSSA provides a technique for analyzing space-time patterns. (Rangelova et al., 2009) use the MSSA method to analyze GRACE-derived mass variation in North America. In that paper, the method is investigated and the results from GRACE data are compared to information from continental water storage models. MSSA is also called extended EOF analysis, a name that is for example used in (Allen and Robertson, 1996) but also in (Hannachi, 2004) who gives a nice overview over the EOF analysis and extended EOF analysis methods.

While EOF analysis focuses on the data variances, MSSA tries to put focus on covariances, too. The basic idea behind the method is to augment the data matrix which is the input to EOF analysis. Temporal lags are included in the spatial dimension of the augmented data matrix. That means the derived EOFs will also have temporal information included, and therefore be extended. But it can also be spatial information included in the temporal dimension. It is made use of spatial and temporal correlations by extending the time respectively space dimension by including information from the other dimension. The augmented

data matrices are constructed as follows. Let $\mathbf{D}$ be an example matrix as follows:

$$\mathbf{D} = \begin{bmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} & d_{17} & d_{18} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} & d_{27} & d_{28} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} & d_{37} & d_{38} \\ d_{41} & d_{42} & d_{43} & d_{44} & d_{45} & d_{46} & d_{47} & d_{48} \\ d_{51} & d_{52} & d_{53} & d_{54} & d_{55} & d_{56} & d_{57} & d_{58} \end{bmatrix} \tag{8.9}$$

Using a lag of 3 leads to $\mathbf{D}_\mathbf{t}^*$ for including a spatial lag in the time domain

$$\mathbf{D}_\mathbf{t}^* = \begin{bmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} \\ d_{12} & d_{13} & d_{14} & d_{15} & d_{16} & d_{17} \\ d_{13} & d_{14} & d_{15} & d_{16} & d_{17} & d_{18} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} \\ d_{22} & d_{23} & d_{24} & d_{25} & d_{26} & d_{27} \\ d_{23} & d_{24} & d_{25} & d_{26} & d_{27} & d_{28} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} \\ d_{32} & d_{33} & d_{34} & d_{35} & d_{36} & d_{37} \\ d_{33} & d_{34} & d_{35} & d_{36} & d_{37} & d_{38} \\ d_{41} & d_{42} & d_{43} & d_{44} & d_{45} & d_{46} \\ d_{42} & d_{43} & d_{44} & d_{45} & d_{46} & d_{47} \\ d_{43} & d_{44} & d_{45} & d_{46} & d_{47} & d_{48} \\ d_{51} & d_{52} & d_{53} & d_{54} & d_{55} & d_{56} \\ d_{52} & d_{53} & d_{54} & d_{55} & d_{56} & d_{57} \\ d_{53} & d_{54} & d_{55} & d_{56} & d_{57} & d_{58} \end{bmatrix} \tag{8.10}$$

and to $\mathbf{D}_\mathbf{s}^*$ for including a temporal lag in the spatial domain

$$\mathbf{D}_\mathbf{s}^* = \begin{bmatrix} d_{11} & d_{21} & d_{31} & d_{12} & d_{22} & d_{32} & d_{13} & d_{23} & d_{33} & \dots \\ d_{21} & d_{31} & d_{41} & d_{22} & d_{32} & d_{42} & d_{23} & d_{33} & d_{43} & \dots \\ d_{31} & d_{41} & d_{51} & d_{32} & d_{42} & d_{52} & d_{33} & d_{43} & d_{53} & \dots \end{bmatrix}. \tag{8.11}$$

These augmented data matrices are the input for a regular EOF analysis. For reconstructing the original data sizes after EOF analysis, the points from the augmented data matrix have to be averaged along the diagonals, but those going from top right to bottom

left.

To demonstrate the basics of the MSSA technique, a data matrix has been filtered with MSSA after Gaussian pre-smoothing of radius 300 km. Then, a regular EOF analysis is performed on the MSAA-filtered data matrix, to show the effects of MSSA filtering and to compare the results from MSSA with temporal and spatial lags. The size of the augmented data matrices were, for the GRACE data set and a lag of 30, for spatial lag included in the time series (64771×1920) and for temporal lag included in the spatial domain (35×1944000). Of course, a crucial step is choosing the number of modes in the MSSA process. In this example, 25 modes for the spatial lags and 6 modes for the temporal lags were chosen. The singular values of MSSA are plotted in figure 8.1.

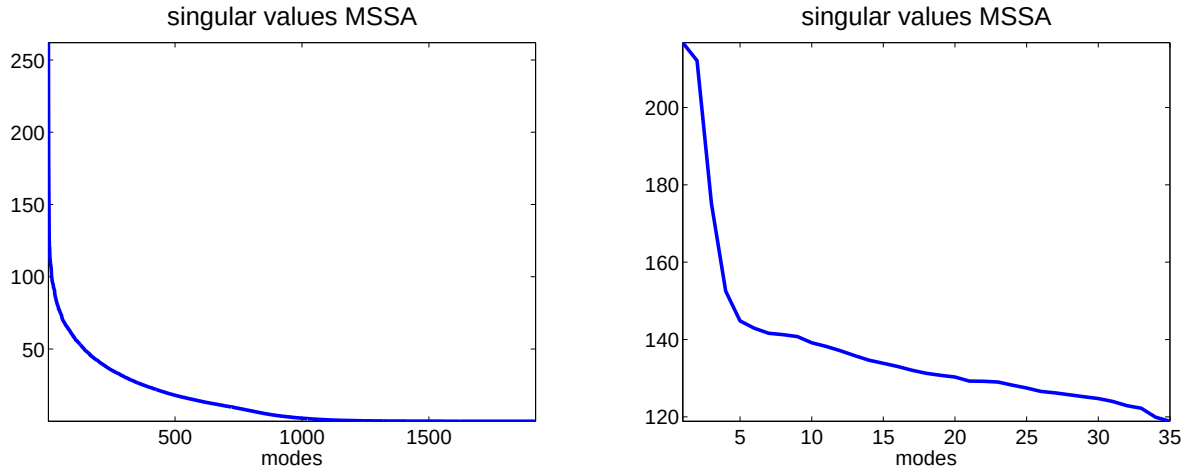


Figure 8.1: Singular values from MSSA for a spatial lag on the left and for a temporal lag on the right

The modes from EOF analysis of the MSSA-filtered data matrix are plotted in figures 8.3 and 8.2 for using a temporal and a spatial lag.

The differences between using a spatial and temporal lag can be clearly seen. The time series for using a temporal lag, where the time dimension was shortened in the MSSA procedure, show a very strange smooth and sine-like behavior in all modes but the first one. This annual signal could be recovered very clearly, but for the temporal lag version it was not possible to recover any other signal. That is why just modes one and two are plotted.

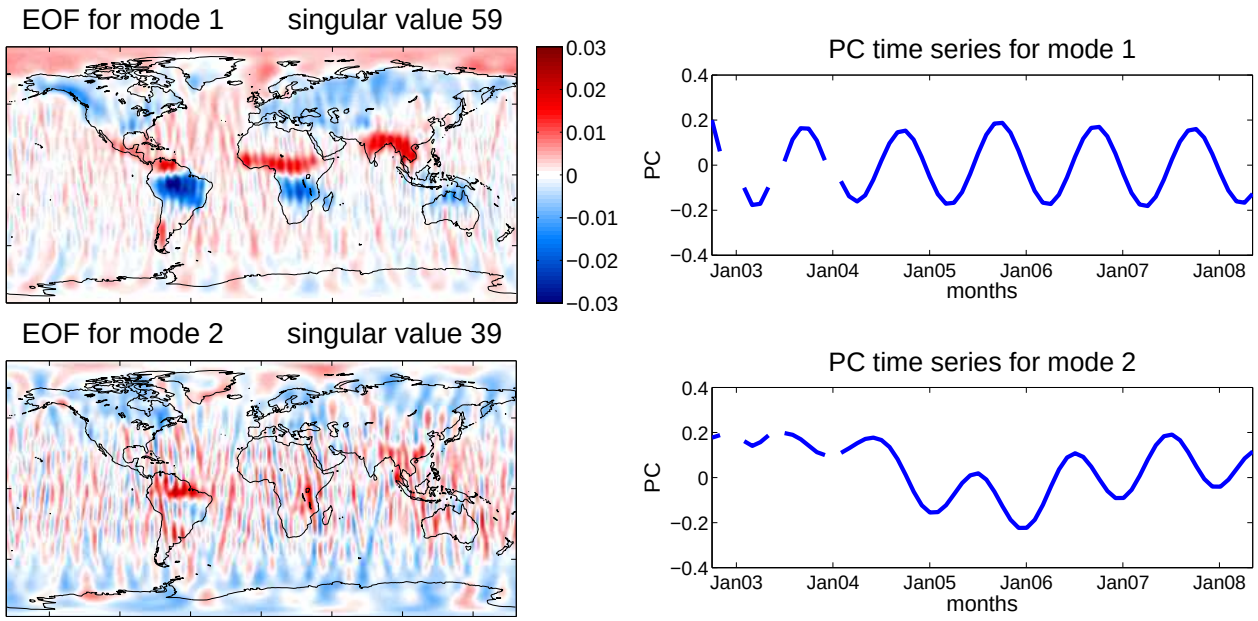


Figure 8.2: EOF patterns and PC time series for the first modes of EOF analysis of a data matrix that was pre-filtered with MSSA and a temporal lag after a Gaussian smoothing with radius 300 km

The higher modes are similar to mode two. The time-lag effect probably leads to the very smooth curve in the first two and a half years, before the amplitude gets larger.

In the results after pre-filtering with MSSA and spatial lag of 30 included in the time dimension, in figure 8.3 the annual signal that we know from previous EOF analysis is found in mode one, and also the well-known trend signal is in mode two. But all higher modes do no more reveal any signal. They do not differ much from mode three that is plotted. Note that the patterns in the EOFs are very clear and more distinguished than they have been in a regular EOF analysis.

The plots of these first few modes were just to give a little insight in the MSSA method and to show the effects of including lags in the data matrix. The method has to be investigated much more for finding good results in the GRACE data set. Here the method shall just be presented. Generally, it can be concluded that a temporal lag helps finding different frequencies in the time series, while a spatial lag helps finding spatial patterns in the maps. So for specific requirements MSSA can be very helpful, but for finding any non-specified

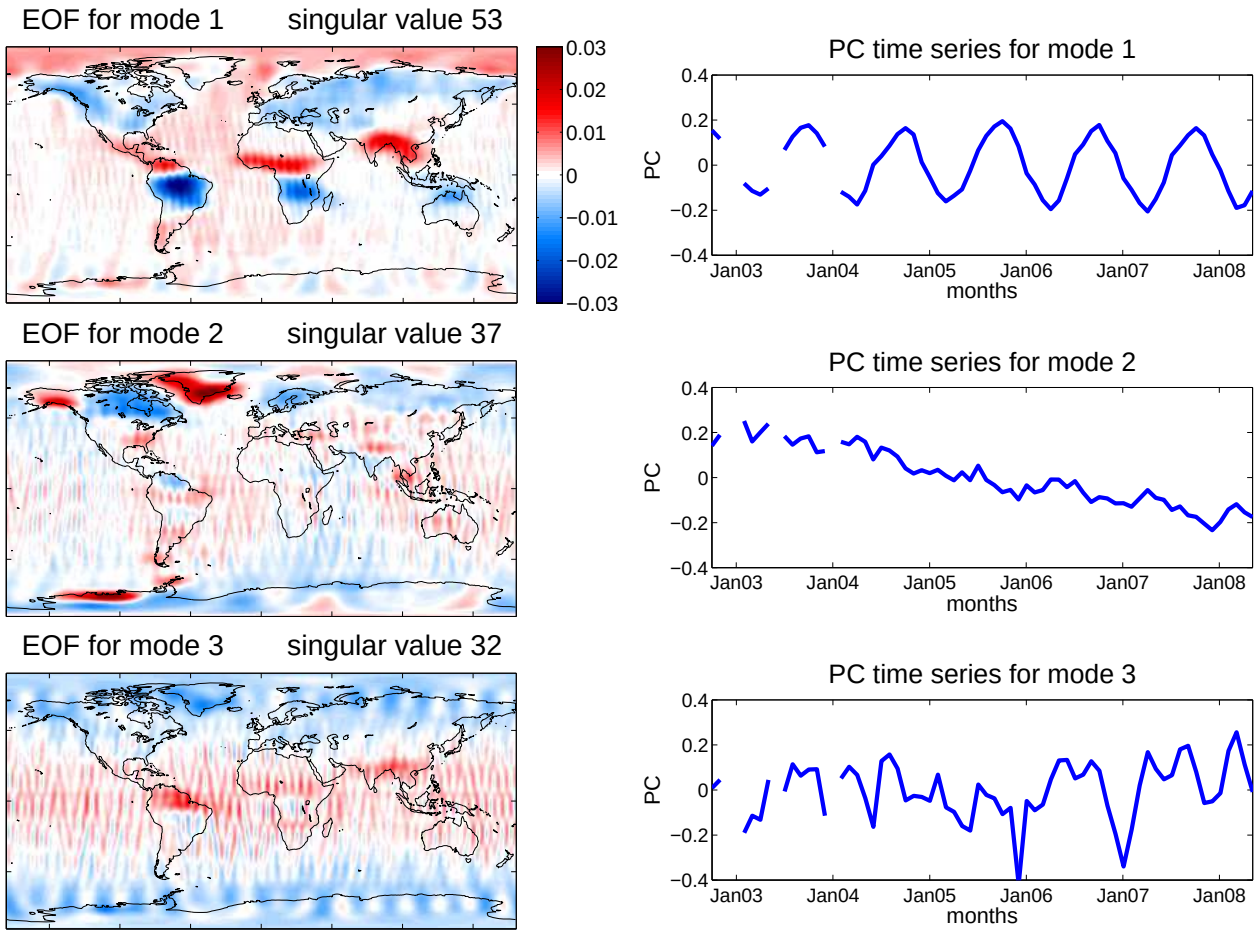


Figure 8.3: EOF patterns and PC time series for the first modes of EOF analysis of a data matrix that was pre-filtered with MSSA and a spatial lag after a Gaussian smoothing with radius 300 km

signal in a data set this method is too specific.

Canonical Correlation Analysis

In Canonical Correlation Analysis (CCA) focus is on the linear relationship between two multidimensional variables, that are two data sets. The method is again an EOF analysis related method. It is searched for new basis vectors for each of the two data sets that maximize the correlation between the two data sets. That means two new bases are found, one for each of the data sets, which are optimal with respect to correlations. That means, the new bases are found in way that the correlation matrix between the data in the new bases is close to diagonal, with correlations on the main diagonal. In other words, the relation-

ship between the two data sets is identified and qualified. In comparison to EOF analysis, which depends on the variances of the data set, CCA depends on the correlations between two data sets. In EOF analysis, the first new basis vector is found according to the major variance, and the second one orthogonal to the first, still variance maximizing. In CCA, a pattern-pair is found that maximizes the correlation between the linear combinations that are the projections of each data set onto the pattern to be found. The second pattern pair is uncorrelated to the first one and is found in the same way as the first one. We have seen that EOF analysis and CCA share similar strategies and mathematics. Background information about CCA can be found in (von Storch and Zwiers, 2002), (Dehon et al.) and (Borga, 2001).

The canonical correlations for two data sets ($\mathbf{X}$) ($\mathbf{Y}$) can be drived as follows: According to (Dehon et al.) and (von Storch and Zwiers, 2002) the canonical correlations can be found by solving the eigenvalue equations

$$\begin{aligned} \mathbf{R}_{xx}^{-1}\mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}\mathbf{e}_x &= \sigma^2\mathbf{e}_x \\ \mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}\mathbf{R}_{xx}^{-1}\mathbf{R}_{xy}\mathbf{e}_y &= \sigma^2\mathbf{e}_y \end{aligned}$$

where $\mathbf{R}_{yx} = \mathbf{R}'_{xy}$, the eigenvalues σ^2 are the squared canonical correlations, and the eigenvectors $\mathbf{e}_x$ and $\mathbf{e}_y$ are the normalized canonical basis vectors. Generally, $\mathbf{R}_{ij}$ are the data scatter matrices.

According to (von Storch and Zwiers, 2002) the canonical correlation patterns $\mathbf{v}$ can be obtained from

$$\begin{aligned} \mathbf{V}_x &= \mathbf{R}_{xx}\mathbf{E}_x \\ \mathbf{V}_y &= \mathbf{R}_{yy}\mathbf{E}_y \end{aligned}$$

and the so-called canonical variates or canonical correlation coordinates $\mathbf{u}$, which can be viewed as the result of coordinate transforms that have been applied to the original data sets $\mathbf{X}$ and $\mathbf{Y}$, can be obtained from

$$\begin{aligned} \beta_x &= \mathbf{e}'_x\mathbf{X} \\ \beta_y &= \mathbf{e}'_y\mathbf{Y} \\ \mathbf{U}_x &= \mathbf{X}'\mathbf{E}_x\sqrt{\mathbf{\Sigma}_x}^{-1} \\ \mathbf{U}_y &= \mathbf{Y}'\mathbf{E}_y\sqrt{\mathbf{\Sigma}_y}^{-1} \end{aligned}$$

Remark: the matrices $\mathbf{R}_{ij}$ are meant to be data scatter matrices, that means they differ from the data covariance matrices by the factor $(n-1)$ and are calculated as follows:

$$\begin{aligned} \mathbf{R}_{xx} &= \mathbf{X}\mathbf{X}' \\ \mathbf{R}_{yy} &= \mathbf{Y}\mathbf{Y}' \\ \mathbf{R}_{xy} &= \mathbf{X}\mathbf{Y}' \end{aligned}$$

CCA can be a very helpful tool for finding correlations, and it is of interest here since it shares ideas and mathematics with the EOF analysis method. CCA could be especially helpful when combining different measurements of one phenomenon since the input data sets just need to be matrices. Like EOF analysis, any physical background does not matter, since it is just a mathematical decomposition. One application of CCA related to this thesis could be for example analyzing space borne gravity measurements and ground based or air borne measurements together.

Bibliography

H. Abdi. Singular value decomposition and generalized singular value decomposition. Technical report, Herve Abdi, Program in Cognition and Neurosciences, The University of Texas at Dallas, Richardson, TX, USA, 2007. Encyclopedia of Measurements and Statistics.

M. R. Allen and L. A. Smith. Optimal filtering in singular spectrum analysis. *Physics Letters A*, (234):419–428, October 13 1997.

M.R. Allen and A.W. Robertson. Distinguishing modulated oscillations from coloured noise in multivariate datasets. *Climate Dynamics*, 12:775–784, 1996.

H. Björnsson and S. A. Venegas. *A Manual for EOF and SVD Analyses of Climatic Data*. McGIll University, Department of Atmospheric and Ocean Sciences and Centre for Climate and Global Change Research McGill University, 2007.

J. A. R. Blais. Lecture notes for ENGO 699: Estimation and spectral analysis. lecture notes, Winter 2008.

M. Borga. *Canonical Correlation - a Tutorial*, January, 12 2001.

C. Dehon, P. Filzemoser, and C. Croux. *Robust Methods for Canonical Correlation Analysis*. Université Libre de Bruxelles, Vienna University of Technology.

G. Eshel. Geosci236: Empirical orthogonal functions, 2003. Department of the Geophysical Sciences, University of Chicago.

F. Flechtner. *Webpage GeoForschungsZentrum (GFZ) Potsdam, Pages on "GRACE"*, April 2001. URL:http://op.gfz-potsdam.de/grace/products/products.html, last accessed: 06/25/09.

F. Flechtner. GRACE GFZ level-2 processing standards document. For Level-2 Product Release 0004 GRACE 327-743, GeoForschungszentrum Potsdam, Department 1: Geodesy and Remote Sensing, February 2007.

F. Flechtner, S. Bettadpur, M. Watkins, and G. Kruizinga. GRACE science data system monthly report January 2007. Technical report, GFZ Potsdam, 2007.

A. Hannachi. *A Primer for EOF Analysis of Climate Data*. Department of Meteorology, University of Reading, Reading U.K., March 24 2004.

S. Iran Pour, K. Bentel, and N. Sneeuw. EOF-based filtering of GRACE gravity field solutions: A comparison between spectral and spatial approaches. EGU General Assembly, Vienna 2009, April 2009.

C. Jekeli. Alternative methods to smooth the earth's gravity field. *Technical Report No. 327*, 1981.

I. T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer, New York, 2nd edition, 2002.

J. Kim and B. Tapley. Error analysis of a low-low satellite-to-satellite tracking mission. *AIAA Journal of Guidance, Control and Dynamics*, 25(6):1100–1106, 2002.

W. J. Krzanowski. *Principles of Multivariate Analysis*. Number 23 in Oxford Statistical Science Series. Oxford Univ. Press., revised edition, 2007.

J.E. Kutzbach. Empirical eigenvectors of sea-level pressure, surface temperature and precipitation complexes over North America. *Journal of Applied Meteoroloy*, 6:791–802, October 1967.

C. Lanczos. *Linear Differential Operators*. Dover Publications, Mineola, New York, 1961.

E. N. Lorenz. Empirical orthogonal functions and statistical weather prediction. *Sci. Rep. 1. Statistical Forecasting Project*, 1956.

W. Menke. *Geophysical Data Analysis: Discrete Inverse Theory*, volume 45. Academic Press, 1989. ISBN: 0 12 490921-3.

G. Plaut and R. Vautard. Spells of low-frequency oscillations and weather regimes in the Northern hemisphere. *Journal of the Atmospheric Sciences*, 51(2):210–236, May 18 1993.

R. W. Preisendorfer. *Principal Component Analysis in Meteorology and Oceanography*. Developments in Atmospheric Sciences 17. Elsevier, 1988.

W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, 3 edition, 2007.

E. Rangelova, W. van der Wal, A. Braun, M. G. Sideris, and P. Wu. Analysis of gravity recovery and climate experiment time-variable mass redistribution signals over North America by means of principal component analysis. *Journal of Geophysical Research*, 112(F03002), 2007. doi:10.1029/2006JF000615.

E. Rangelova, W. van der Wal, M.G. Sideris, and P. Wu. Spatiotemporal analysis of the grace-derived mass variations in North America by means of multi-channel singular spectrum analysis. *Proceedings of the IAG Symposia*, June 2009.

E. V. Rangelova. *A Dynamic Geoid Model for Canada*. Phd thesis, University of Calgary, Department of Geomatics Engineering, 2007.

A. C. Rauhut. *Regularization of the Inverse Stokes Problem*. Phd thesis, University of Calgary, Department of Geomatics Engineering, 1992.

A. W. Robertson. Interdecadal variability over the north pacific in a multi-century climate simulation. *Climate Dynamics*, 12:227–241, 1995.

E. J. O. Schrama, B.Wouters, and D. A. Lavalleé. Signal and noise in gravity recovery and climate experiment (GRACE) observed surface mass variations. *Journal of Geophysical Research*, 12(B08407), 2007. doi:10.1029/2006JB004882.

S. Svenson and J. Wahr. Post-processing removal of correlated errors in GRACE data. *Geophysical Research Letters*, 33(L08402), 2006. doi:10.1029/2005GL025285.

B.D. Tapley, S. Bettadpur, M.Watkins, and Ch. Reigber. The gravity recovery and climate experiment: Mission overview and early results. *Geophysical Research Letters*, 31(L09607), 2004. doi:10.1029/2004GL019920.

H. von Storch and F. W. Zwiers. *Statistical Analysis in Climate Research*. Cambridge, UK, Cambridge University Press, 2002. ISBN:0521450713.

J. Wahr, M. Molenaar, and F. Bryan. Time variability of the earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE. *Journal of Geophysical Research*, 103(B12):30,305–30.229, 1998.

D. S. Wilks. *Statistical Methods in the Atmospheric Sciences*, volume 59 of *International Geophysics Series*. Academic Press, 1995.

B. Wouters and E. J. O. Schrama. Improved accuracy of GRACE gravity solutions through empirical orthogonal function filtering on spherical harmonics. *Geophysical Research Letters*, 34(L23711), 2007. doi:10.1029/2007/GL032098.