

# GRAS SAF



## **GRAS Satellite Application Facility GRAS SAF Visiting Scientist Project No. 10**

### **Report**

## **Fusion of Radio Occultation Data with ERA40 derived Fields to conduct Climate Studies**

**Version 2.0**

**23 July 2008**


Danish Meteorological Institute (DMI)  
European Centre for Medium-Range Weather Forecasts (ECMWF)  
Institut d'Estudis Espacials de Catalunya (IEEC)  
Met Office (MetO)

Ref: SAF/GRAS/ DMI/REP/VS10/001  
Issue: Version 2.0  
Date: 24 July 2008  
Document: cover\_grassaf\_vs10-  
report\_v2.doc

GRAS SAF  
Visiting Scientist Project No.  
10 Report

*EUMETSAT*  
DMI  
ECMWF  
IEEC  
Met Office



Ref: SAF/GRAS/ DMI/REP/VS10/001 Issue: Version 2.0 Date: 24 July 2008 Document: cover_grassaf_vs10-report_v2.doc	GRAS SAF Visiting Scientist Project No. 10 Report	EUMETSAT DMI ECMWF IEEC Met Office	 www.grassaf.org
---	---	--	--

---




---

## DOCUMENT SIGNATURE TABLE

---



---

	<b>Author(s)</b>	<b>Function</b>	<b>Date</b>	<b>Signature</b>
Prepared by:	Armin Löscher	GRAS SAF Visiting Scientist	08/05/08	
Approved by:	Kent B. Lauritsen	GRAS SAF Project Manager	08/05/08	

---



---

## DOCUMENTATION CHANGE RECORD

---



---

<b>Issue / Revision</b>	<b>Date</b>	<b>By</b>	<b>Description</b>
0.1	01/12/07	ARL	Draft document
1.0	01/08/08	ARL	General Updates and incorporation of comments by S. B. Healy & J. R Eyre
1.1	10/04/08	ARL	General Updates and incorporation of comments by S. Syndergaard
2.0	06/05/08	ARL	Small changes, final editing and formatting

---

# GRAS-SAF Visiting Scientist Report No. 10

Dr. Armin Löscher

July 2008



# Contents

<b>Contents</b>	<b>2</b>
<b>List of Figures</b>	<b>4</b>
<b>List of Tables</b>	<b>5</b>
<b>Colophon</b>	<b>6</b>
<b>Abstract</b>	<b>7</b>
0.1 Section 2 . . . . .	8
0.2 Section 3 . . . . .	8
<b>1 Background Data</b>	<b>9</b>
1.1 ECMWF ERA40 Fields . . . . .	9
<b>2 Radio Occultation Data</b>	<b>9</b>
2.1 The RO Technique . . . . .	11
2.1.1 Unique Advantages for Monitoring of Atmospheric Key Parameters . . . . .	11
2.1.2 Characteristic Horizontal and Vertical Resolution of RO Measurements . . . . .	11
2.2 Retrieval . . . . .	12
2.2.1 General RO Retrieval . . . . .	12
2.3 Data Products . . . . .	15
2.3.1 Refractivity Profiles . . . . .	15
2.3.2 Temperature Profiles . . . . .	15
2.3.3 Humidity Profiles . . . . .	16
2.3.4 Geopotential . . . . .	17
2.3.5 Pressure . . . . .	17
2.3.6 Total Electron Content . . . . .	17
2.4 The CHAMP Satellite . . . . .	17
<b>3 Assimilation System Setup</b>	<b>17</b>
3.1 Coordinate System . . . . .	18
3.2 Temperature, Specific Humidity and Surface Pressure Analysis . . . . .	18
3.3 Variational Assimilation 3D-VAR . . . . .	19
3.3.1 3D-VAR and Incremental 3D-VAR . . . . .	19
3.4 Implementation Technique . . . . .	19
3.4.1 Control Variables . . . . .	19
3.4.2 Minimization . . . . .	20
3.4.3 Preconditioning . . . . .	20
3.4.4 The Adjoint Technique . . . . .	20
3.5 Observation Operators . . . . .	20
3.5.1 Interpolation Operator . . . . .	20
3.5.2 Refractivity Operator . . . . .	21
3.5.3 Vertical Coordinate Operator . . . . .	22
3.6 Observation Error Covariance Matrix . . . . .	24
3.6.1 Formulation of the Observation Error Covariance Matrix . . . . .	24
3.7 Incest Problem . . . . .	26
3.8 Background Error Covariance Matrix . . . . .	26
3.8.1 Variances . . . . .	26
3.8.2 Calculation of Vertical Correlation Matrices . . . . .	26

3.8.3	Calculation of Horizontal Point to Point Correlations . . . . .	27
3.8.4	Vertical and Horizontal ECMWF Temperature Error Correlations Northern High Latitudes $0^\circ - 30^\circ$ (NH) . . . . .	29
3.8.5	Vertical and Horizontal ECMWF Temperature Error Correlations Northern Mid Latitudes $30^\circ - 60^\circ$ (NM) . . . . .	30
3.8.6	Vertical and Horizontal ECMWF Temperature Error Correlations Low Latitudes $60^\circ - 120^\circ$ (LO) . . . . .	31
3.8.7	Vertical and Horizontal ECMWF Temperature Error Correlations Southern Mid Latitudes $120^\circ - 150^\circ$ (SM) . . . . .	32
3.8.8	Vertical and Horizontal ECMWF Temperature Error Correlations Southern High Latitudes $150^\circ - 180^\circ$ (SH) . . . . .	33
3.8.9	Vertical and Horizontal ECMWF Specific Humidity Error Correlations Northern High Latitudes $0^\circ - 30^\circ$ (NH) . . . . .	34
3.8.10	Vertical and Horizontal ECMWF Specific Humidity Error Correlations Northern Mid Latitudes $30^\circ - 60^\circ$ (NM) . . . . .	35
3.8.11	Vertical and Horizontal ECMWF Specific Humidity Error Correlations Low Latitudes $60^\circ - 120^\circ$ (LO) . . . . .	36
3.8.12	Vertical and Horizontal ECMWF Specific Humidity Error Correlations Southern Mid Latitudes $120^\circ - 150^\circ$ (SM) . . . . .	37
3.8.13	Vertical and Horizontal ECMWF Specific Humidity Error Correlations Southern High Latitudes $150^\circ - 180^\circ$ (SH) . . . . .	38
3.8.14	Horizontal ECMWF Temperature and Specific Humidity Error Correlations Global Mean . . . . .	39
3.8.15	Zonal Mean Monthly Temperature and Specific Humidity Variance . . . . .	40
3.8.16	Global Mean Monthly Surface Pressure Variance and Correlation . . . . .	41
3.8.17	Horizontal ECMWF Surface Pressure Error Correlations Northern High ( $0^\circ - 30^\circ$ ) and Mid ( $30^\circ - 60^\circ$ ) Latitudes . . . . .	42
3.8.18	Horizontal ECMWF Surface Pressure Error Correlations Northern Low ( $60^\circ - 120^\circ$ ) and Southern Mid ( $120^\circ - 150^\circ$ ) Latitudes . . . . .	43
3.8.19	Horizontal ECMWF Surface Pressure Error Correlations Southern High ( $150^\circ - 180^\circ$ ) Latitudes . . . . .	44
3.8.20	Horizontal ECMWF Temperature and Specific Humidity Error Correlations Global Mean Derived along Bands of Constant Longitude from North to South Pole . . . . .	45
3.8.21	Horizontal ECMWF Surface Pressure Error Correlation Global Mean Derived along Bands of Constant Longitude from North to South Pole . . . . .	46
3.9	Control Space Transformations . . . . .	46
3.9.1	Vertical Control Variable Transform . . . . .	47
3.9.2	Horizontal Control Variable Transform . . . . .	48
3.9.3	Recursive Filters . . . . .	48
3.10	Horizontal Background Error Covariances . . . . .	51
3.10.1	Horizontal Global Mean ECMWF Temperature Correlations and its Approximation by a Recursive Filter . . . . .	52
3.10.2	Horizontal Global Mean ECMWF Specific Humidity Correlations and its Approximation by a Recursive Filter . . . . .	53
3.10.3	Horizontal Global Mean ECMWF Temperature and Specific Humidity Correlations Approximated by a Recursive Filter Times Two . . . . .	54

3.10.4	Horizontal Global Mean ECMWF Surface Pressure Correlations Approximated by a Recursive Filter and the Respective Approximation Times Two . . . . .	55
3.10.5	Horizontal Global Mean ECMWF Surface Pressure Correlation . . . . .	56
3.11	The Minimization Algorithm . . . . .	56
<b>4</b>	<b>Data Preprocessing</b>	<b>57</b>
<b>5</b>	<b>Quality Control</b>	<b>57</b>
<b>6</b>	<b>Validation</b>	<b>58</b>
6.1	Validation of Adjoint Code . . . . .	58
<b>7</b>	<b>System Test Runs</b>	<b>58</b>
7.1	Convergence Behavior . . . . .	61
	<b>Conclusions</b>	<b>62</b>
<b>A</b>	<b>Notation</b>	<b>64</b>
<b>B</b>	<b>Constants</b>	<b>65</b>
<b>C</b>	<b>List of Acronyms</b>	<b>66</b>
	<b>References</b>	<b>68</b>

## List of Figures

1	Model orography for the T42 resolution. . . . .	10
2	Geometry of a radio occultation measurement including LEO satellite (CHAMP), GPS satellites and fiducial network. . . . .	12
3	Artist view of the CHAMP satellite in orbit (courtesy GFZ Potsdam, NASA picture archive, 2004) . . . . .	18
4	Mean global refractivity profile calculated from T42L60 ECMWF analysis fields. . . . .	22
5	Difference in global mean refractivity calculated with Smith-Weintraub and Thayer Formula. . . . .	23
6	Relative refractivity standard deviation based on the presented formulas and parameters. . . . .	25
7	Derived vertical and horizontal error correlations of ERA40 temperature fields northern high latitudes. . . . .	29
8	Derived vertical and horizontal error correlations of ERA40 temperature fields northern mid latitudes. . . . .	30
9	Derived vertical and horizontal error correlations of ERA40 temperature fields low latitudes. . . . .	31
10	Derived vertical and horizontal error correlations of ERA40 temperature fields southern mid latitudes. . . . .	32
11	Derived vertical and horizontal error correlations of ERA40 temperature fields southern high latitudes. . . . .	33
12	Derived vertical and horizontal error correlations of ERA40 specific humidity fields northern high latitudes. . . . .	34
13	Derived vertical and horizontal error correlations of ERA40 specific humidity fields northern mid latitudes. . . . .	35
14	Derived vertical and horizontal error correlations of ERA40 specific humidity fields low latitudes. . . . .	36

15	Derived vertical and horizontal error correlations of ERA40 specific humidity fields southern mid latitudes. . . . .	37
16	Derived vertical and horizontal error correlations of ERA40 specific humidity fields southern high latitudes. . . . .	38
17	Derived vertical and horizontal error correlations of ERA40 specific humidity and temperature fields global mean. . . . .	39
18	Zonal mean of temperature and specific humidity ERA40 monthly mean variance fields. . . . .	40
19	Global mean ERA40 surface pressure variance and respective derived horizontal correlation. . . . .	41
20	Derived horizontal error correlations of ERA40 surface pressure fields northern high and mid latitudes. . . . .	42
21	Derived horizontal error correlations of ERA40 surface pressure fields low and southern mid latitudes. . . . .	43
22	Derived horizontal error correlations of ERA40 surface pressure field southern high latitudes. . . . .	44
23	Derived horizontal error correlations of ERA40 specific humidity and temperature fields from pole to pole global mean. . . . .	45
24	Derived horizontal error correlation of ERA40 surface prssure fields from pole to pole global mean. . . . .	46
25	Segments along one lattitude or longitude band in original order. . . . .	49
26	Shifted arrangement of segments along one lattitude or longitude band. . . . .	49
27	Global mean horizontal error correlations of ECMWF temperature fields and corresponding filter approximation. . . . .	52
28	Global mean horizontal error correlations of ECMWF specific humidity fields and corresponding filter approximation. . . . .	53
29	Global mean horizontal error correlations of ECMWF temperature and specific humidity fields and corresponding filter approximation times two. . . . .	54
30	Global mean horizontal error correlations of ECMWF surface pressure fields approximated by a recursive filter and the respective approximation times two. . . . .	55
31	Global mean horizontal error correlations of ECMWF surface pressure fields approximated by a recursive filter and the respective approximation times two. . . . .	56
32	Interpolated and thinned measurement distribution. . . . .	57
33	Zonal mean temperature increment of the 00 time layer of the JJA 2004 season. . .	60
34	Zonal mean specific humidity increment of the 00 time layer of the JJA 2004 season. .	60
35	Zonal mean surface pressure increment of the 00 time layer of the JJA 2004 season. .	61

## List of Tables

1	For assimilation study used observations from JJA 2004 time 00. . . . .	10
2	Overview of IGAM CHAMP-RO retrieval schemes (EGOPS/CCR Version 2, March 2004.) . . . . .	15
3	Parameter and Variables used in Thayer & Smith-Weintraub formula. . . . .	21
4	Parameter and Variables used in formulation of observation error covariance matrix formula. . . . .	25
5	CHAMP RO observation distribution summer season (JJA) 2004. . . . .	59

## Colophon

**Serial Title:**

GRAS-SAF Visiting Scientist Project No. 10

**Title:**

Fusion of Radio Occultation Data with ERA40 derived Fields to conduct Climate Studies

**Subtitle:**

The Assimilation of Radio Occultation Data into ERA40 derived Monthly Mean First Guess Fields of Temperature, Specific Humidity and Surface Pressure using the Respective derived Error Patterns

**Author(s):**

Armin Löscher

**Other contributors:**

Christian Retscher (ESA's Grid on-Demand implementation)  
WegCenter (CHAMP Data Provision)  
ECMWF (ERA40 Data)

**Responsible institution:**

DMI

**Language:**

English

**Keywords:**

Radio Occultation, Assimilation, ERA40, Climatology, Temperature, Specific Humidity, Surface Pressure, First Guess, Error Correlation & Covariance Matrices, Monthly Mean

**Doc id:**

SAF/GRAS/DMI/REP/VS10/001

**Version:**

2.0

**Website:**

[www.dmi.dk](http://www.dmi.dk)  
[www.grassaf.org](http://www.grassaf.org)

**Copyright:**

DMI

## Abstract

This document contains a technical description and first results of the assimilation system developed within the framework of the Visiting Scientist project No. 10 at the GRAS SAF (Global Navigation Satellite System Receiver for Atmospheric Sounding Satellite Application Facility) to conduct climate studies using RO (Radio-Occultation) data and ECMWF (European Center for Medium Range Forecast) fields. It is a mature and improved version of the system which is presented detailed in [Loescher(2004)]. The method of choice is 3D-Var (Three Dimensional Variational Assimilation) realized as a system which allows to merge information from different sources (in our case RO data and from ECMWF ERA40 (European 40 Years Reanalysis) data derived first guess fields) under the assumption of known error structures, in a statistically optimal way. In short the whole system is based on a 3D-VAR implementation using climatological background fields and error covariances representing the departure of the atmospheric state from the climatological mean.

The main idea behind this project is to probe the application of a methodology used in numerical weather prediction (NWP) for climate and atmospheric variability studies. To tackle the computational challenge inherent to large scale optimization ESA's Computing Grid on-Demand is used for the trial runs. Since computing power increases constantly such optimization technologies will play a more prominent role in the future not only in climate applications but also in cross validation activities and bias removal procedures. Balancing the shortcoming of different remote sensing techniques by combining them, results in a product of significant added value reflecting, in an optimal case, only the advantages of the different observations used. The use of a time discrete scheme like 3D-Var to merge different kinds of observations (or in this case one kind of observation and a background) to generate analyses of the atmospheric state at different time scales is apparent.

A challenge inherent to assimilation systems is the correct definition of the error patterns. This work is assessing climate change and atmospheric variability studies, thus monthly means are envisaged as baseline. To that end suitable background fields and the respective error characteristics have to be derived since no suitable data is readily available.

Beside the technical aspects the document covers the derivation of the first guess fields and the respective variations which are used as proxies for errors structures from the ERA40 data based on monthly means. The derived first guess fields comprising the respective variances and the vertical correlation matrices (for temperature, specific humidity and surface pressure) had been used within the assimilation test-bed summer season June, July and August (JJA) 2004. A possible method to calculate the horizontal correlations for the respective variables is presented and the results of some example calculations are shown. Within the assimilation test-bed modified global mean horizontal correlations used within ECMWF's IFS (Integrated Forecast System) in 2003 have been used. It is recommended to refer to [Loescher et al.(2008)Loescher, Retscher, Fusco, Goncalves, Brito, and Kirchengast] which complements this report and contains a summary, first results and an exhaustive list of references.

This document describes the methodology used to realize a 3D-Var assimilation system for climate and atmospheric variability studies containing first results and the characteristics of all relevant data is organized as follows:

## **Section 1**

The kind of data which is used for the assimilation experiments is introduced, some background information on ERA40 fields monthly mean fields and ECMWF data is given here. The rationale to select the time frame to derive the first guess data from the reanalysis is explained.

### **0.1 Section 2**

Radio occultation data is specified here in some detail (focus on CHAMP), with a brief outlook on future and present radio occultation missions (e.g. MetOp GRAS & COSMIC).

### **0.2 Section 3**

Here the specific implementation of the temperature, specific humidity and surface pressure (TQPsurf) 3D-Var system is described in some detail. The realization of control space transformations and the preconditioning is explained, as well as the use of recursive filters within this framework. The observation and background covariance matrices are defined, the observation operators are explained in detail and the minimization routine is introduced.

An important part of this section addresses the used background error patterns namely variances and correlations from the ERA40 monthly mean fields. The way which the vertical part of the background error covariance matrix has been derived is explained and possible methods to derive horizontal correlation matrices are presented.

## **Section 4**

Some details on the break down of the radio occultation profile resolution to a background compatible data density (data thinning) and preprocessing procedures are given here.

## **Section 5**

Possible quality control procedures are presented and the check for observations below the orography is introduced here.

## **Section 6**

Brief explanation of adjoint code validation techniques.

## **Section 7**

The results of the first assimilation experiments with quasi operational CHAMP data from the summer season June, July, August (JJA) 2004 on ESA's high performance grid on demand computing cluster are presented and the convergence behavior is discussed.

## **Conclusions**

Some conclusions from the results presented in section 7 are drawn and specific issues are addressed. A short outlook is given and future perspectives are discussed.

## **Appendices**

Additional information concerning the used notation, constants and acronyms.

## 1 Background Data

The use of background data is in general flexible, here background fields derived from 21 years of ECMWF's ERA40 data are used. This approach makes the system model independent in the sense that first guess and error characteristics are static in contrast to the use of any kind of operational NWP output, where the fields have to be adapted to any change in the NWP system itself or the assimilated observation systems. Monthly mean fields for the six hour time layers around 00, 06, 12 and 18 UTC have been averaged to get a first guess of temperature, specific humidity and surface pressure. The same procedure has been applied to derive the fields of variances for these parameters, correlations are derived in a similar way (cf. 3.8.2).

### 1.1 ECMWF ERA40 Fields

As background for the assimilation procedure ECMWF N48L60 ERA40 derived fields, corresponding to a Gaussian grid composed of  $192 \times 96$  geographic areas, and 60 standard model levels up to a height of  $\sim 64$  km are used. The resolution of the vertical grid is highest in the planetary boundary layer and lowest in the stratosphere and lower mesosphere. These vertical levels are realized as hybrid levels, which means they are composed as  $\sigma$  - levels which follow the Earth's surface in the lower and mid troposphere, but are surfaces of constant pressure (pressure levels) in the upper stratosphere and mesosphere with a smooth transition between these types of levels.

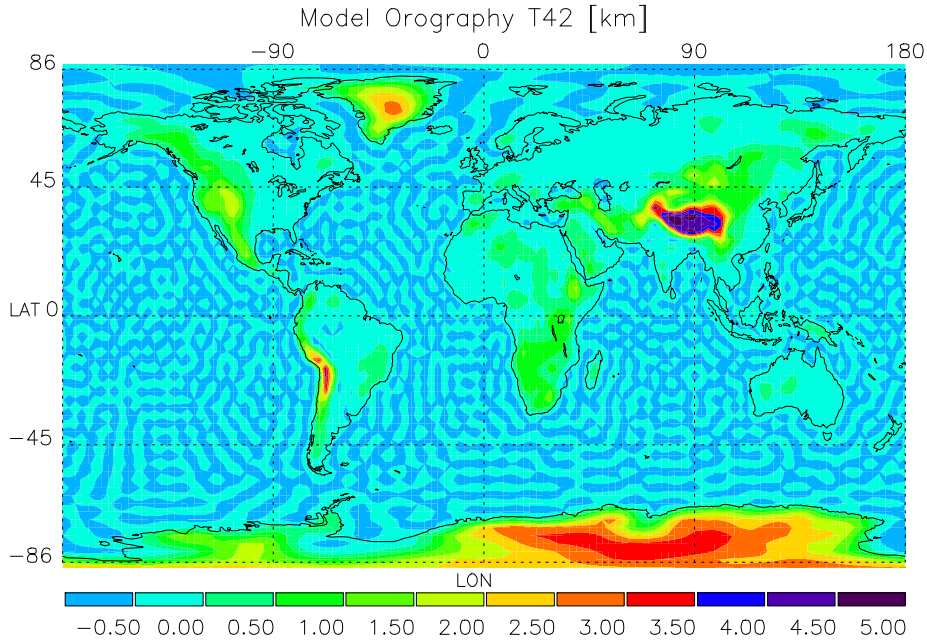
Since a reanalysis has to be done within a reasonable timeframe the computational cost has to be cut down. Using an older version of the system on nowadays hardware is a unsatisfactory solution, omitting advanced and better formulations within the state of the art implementation. Thus a version which represents a compromise had to be developed. In the ERA40 case a modified form of the three dimensional variational analysis scheme used operationally as the ECMWF medium range prediction system between January 1996 and November 1997 was used. It comprises a triangular truncation at wavenumber 159 (T159) for the *spectral* representation of atmospheric fields and 60 vertical levels. In addition there is a grid point representation used for computing dynamic tendencies and the diabatic physical parametrization. This so called Gaussian grid, is regular in longitude but not regular in latitude. Due to the convergence of the longitudes towards the poles, the east-west distance between the grid points decreases polewards. To avoid some numerical problems around the poles and most importantly, to save computing time, a reduced Gaussian grid was introduced by reducing the number of grid points along the shorter latitude lines near the poles, so as to keep the east-west separation between points on different latitudes almost constant ( $\sim 125$ km) [Simmons(2004)], [ECMWF(2003)]. The representation of the *orography* uses the mean orography and is significantly smoother than reality (Fig. 1, from a 2003 operational analysis).

ECMWF's ERA40 covers a 45 year period from September 1957 to August 2002 [Kellberg et al.(2004)Kellberg, Simmons, Uppala, and Fuentes]. To derive the monthly mean fields and variances a period of 21 years has been chosen starting from 1980, covering a period marked by the increasing and lately extensive use of satellite observations within GCM's, improving the analysis quality of remote and data sparse areas (especially over the southern hemisphere) significantly. For our application global temperature, specific humidity, and surface pressure fields are used. As discussed below the spatial characteristics of Radio Occultation data (moderate horizontal, high vertical resolution) fits quite well to this background grid spacing [Gobiet and Kirchengast(2004)]. ECMWF offers a broad spectrum of analysis and forecast products [Person(2003)].

## 2 Radio Occultation Data

With the successful launch of the CHAMP satellite in summer 2000 and the start of its GPS Radio Occultation experiment in February 2001, the number of available RO-based atmospheric profiles





**Figure 1:** Model orography for the T42 resolution.

increased in a way that long term climatological studies become feasible. A good collection of RO related articles can be found in [Kirchengast et al.(2004)Kirchengast, Foelsche, and Steiner] and [Kirchengast et al.(2006)Kirchengast, Foelsche, and Steiner]. In addition to CHAMP, also the RO experiments on the Argentinean SAC-C satellite contributes data (SAC-C data is fragmentary and limited to certain periods, at the moment further data from SAC-C is questionable). The first successful processed GRACE RO profiles were published by JPL on 29 of July 2004, GRACE data is available on a CHAMP like base. Furthermore, a RO receiver (GRAS) [EUMETSAT(2003)] is part of the payload of the METOP series of polar-orbiting, operational meteorological satellites operated by EUMETSAT and ESA, which was successfully launched 19. October 2006, 16:28 UTC on a Soyuz launcher from Baikonur. The COSMIC mission (launched 6:40 p.m. PDT from the Vandenberg Air Force Base, CA, on Friday, April 14. 2006) already delivers a continuous stream of RO data. The global coverage, all-weather capability, high vertical resolution, accuracy and long term stability of RO data makes them an ideal candidate to build global climatologies of fundamental variables such as temperature, geopotential height and water vapor [Kirchengast et al.(2004)Kirchengast, Foelsche, and Steiner], [Gobiet and Kirchengast(2004)].

At the moment a comprehensive CHAMP data set starting 2002 is available. For this study data from the summer season 2004 (JJA 2004) around the 00 analysis time layer has been used (cf. Tab1).

Time	Number of Profiles	Number of Pre-Processed Observations
	JJA 2004	JJA 2004
00	2905	157757

**Table 1:** For assimilation study used observations from JJA 2004 time 00.

## 2.1 The RO Technique

Radio Occultation (RO) is a novel active limb sounding technique to derive atmospheric key parameters. The measurement setup comprises a receiver mounted on a low Earth orbit (LEO) satellite, which tracks the signal of a global navigation system (GPS) satellite positioned in a medium Earth orbit (MEO), in an occultation geometry. The challenges from a technical point of view imply the necessity of an extremely high frequency stability in the signal and the positions and velocities of transmitter and receiver must be known to very high degree of accuracy. The concept was successfully proven on-board the Micro Lab 1 satellite (GPS/MET experiment) [Kursinski *et al.*(1996)] and is now quasi operationally implemented as part of the CHAMP and COSMIC missions and fully operational as part of MetOp. Fig. 2 illustrates the concept of the RO technique, which is the interaction of electromagnetic waves (GPS signals) and the (in our application) terrestrial atmosphere. An electromagnetic ray passing through the atmosphere is bent and retarded due to the ionosphere and the Earth's refractivity field. In our application the signal must be corrected for the influence of the ionosphere which is accomplished by a differential approach. For other applications this information of the signal is used to derive maps of the ionosphere and the total electron content (TEC). The effect of the atmosphere onto the electromagnetic waves can be characterized by a total bending angle ( $\alpha$ ) as a function of the impact parameter ( $a$ ). The impact parameter is defined, assuming spherical symmetry, as the perpendicular distance between the center of local curvature at the perigee of the occultation ray and the ray asymptote at the GPS or LEO satellite.

### 2.1.1 Unique Advantages for Monitoring of Atmospheric Key Parameters

Due to the measurement principle, Radio Occultation features some preferable characteristics which makes it an ideal technique for a long term monitoring of atmospheric key parameters [Anthes *et al.*(2000)Anthes, Rocken, and Kuo]. Its long term stability and self-calibrating concept makes it an ideal candidate for climate studies.

- The atmospheric profiles are not derived from absolute intensities or phase delays.
- The profiles are derived from transmissions (normalized intensities) and the Doppler shift (phase change) profiles (intrinsic self-calibration).
- Only short-term stability is necessary during the occultation event.

### 2.1.2 Characteristic Horizontal and Vertical Resolution of RO Measurements

One of the characteristics of a RO measurement is its high vertical ( $\Delta z$ ) and moderate to low horizontal ( $\Delta L$ ) spatial resolution. The following relation is valid

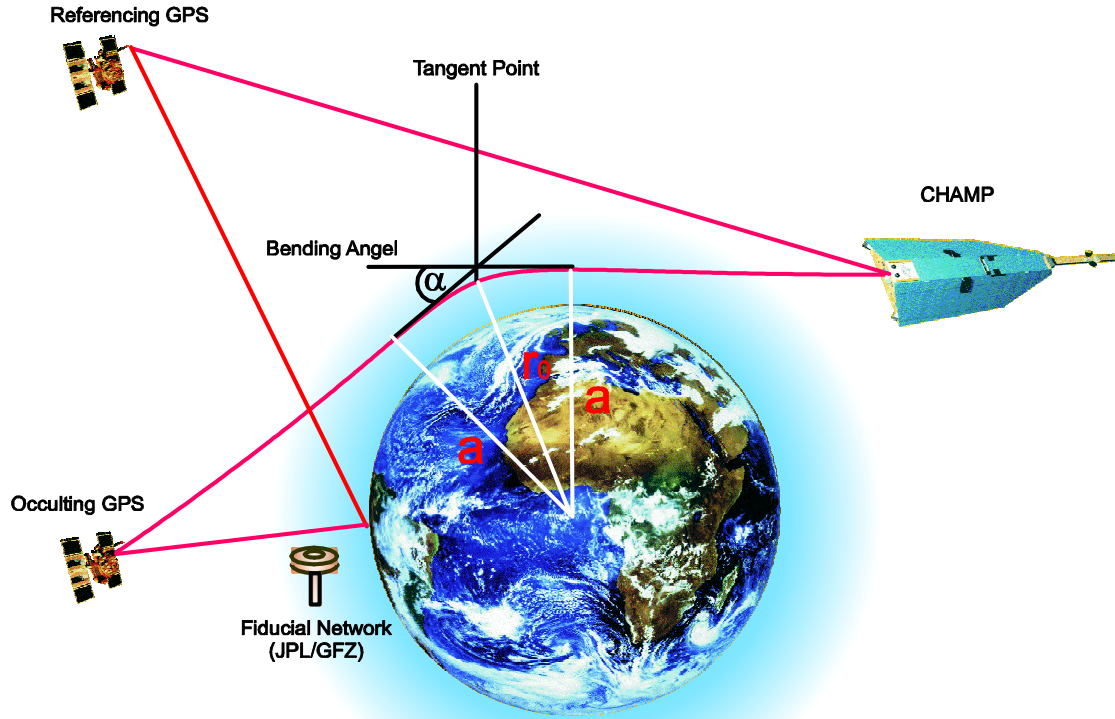
$$\Delta L = 2 \cdot \sqrt{2 \cdot R \cdot \Delta z} , \quad (1)$$

where  $\Delta L$  denotes the chord, which is defined by the tangent of the inner of two concentric circles with radii which differ about  $\Delta z$ .  $R$  is the radius of the inner circle, which is in fact the atmospheric radius at the tangent point of the ray path. If geometric optics is applied, the vertical resolution is limited by the diameter of the first Fresnel Zone  $d_F$ . For occultation geometry, this can be calculated by negligible atmosphere (stratosphere) as follows

$$d_F = 2 \cdot \sqrt{\lambda \cdot D} ,$$

where  $\lambda$  denotes the wavelength of the GPS signal and  $D$  the distance between the GPS receiver on-board CHAMP and the tangent point. With  $\lambda = 19$  cm and  $D = 2.600$  km (orbit height of 500 km) for the diameter of the first Fresnel Zone a value of 1.4 km follows. Using Eq. 1 a horizontal resolution  $\Delta L$  of  $\sim 270$  km can be calculated. Caused by the exponential increase of the refractivity towards the Earth's surface (troposphere)  $d_F$  becomes smaller and reaches close

## Geometry of GPS limb sounding with CHAMP



**Figure 2:** Geometry of an radio occultation measurement including LEO satellite (CHAMP), GPS satellites and fiducial network.

to the Earth's surface a value of 0.5 km [Kursinski et al.(1997)Kursinski, Hajj, Schofield, K., and Hardy] which translates into a horizontal resolution of about  $\sim 80$  km. It is possible to enhance the vertical resolution by using methods which take diffraction effects into account [Gorbunov and Gurvich(1998)], [Wickert(2002)].

## 2.2 Retrieval

### 2.2.1 General RO Retrieval

As a basic idea the radio signals emitted by the GNSS satellites can be treated as rays, which means a geometric optics assumption. This is a valid simplification from the mid-troposphere upwards. However below some 7 km wave optics methods, which can cope with complex signal structures in the presence of strong refractivity gradients enhance the retrieval performance significantly [Gorbunov(2002)], [Hocke et al.(1999)Hocke, Pavelyev, Yakovlev, Barthes, and Jakowski], [Sokolovskiy(2003)]. At the moment the IGAM retrieval for CHAMP data uses only the geometric optics approach. The use of a retrieval procedure which blends geometric optics assumption from the mid-troposphere upwards with wave optics methods derived data in the lower regions is

foreseen to be used in the near future.

$$\begin{aligned} \text{GPS Frequencies } L_1 &= 1575.42 \text{ MHz} \\ L_2 &= 1227.60 \text{ MHz} \end{aligned}$$

The primary observables are the phase delays of GNSS signals, resulting from the deceleration of the electromagnetic wave's phase velocity by the atmosphere. The Doppler shifts and total bending angles  $\alpha$  as function of the ray's impact parameter  $a$  can be deduced from the phase delays [Kursinski et al.(1997)Kursinski, Hajj, Schofield, K., and Hardy]. The basis to derive  $\alpha$  is the Doppler-Shift equation [Gorbunov et al.(1996)Gorbunov, Sokolovskiy, and Bengtsson]:

$$f_d = f_c \left( \frac{c - \vec{v}_2 \vec{m}_2 n_2}{c - \vec{v}_1 \vec{m}_1 n_1} - 1 \right), \quad (2)$$

where  $\vec{v}_1$  and  $\vec{v}_2$  denote the velocity vectors of the occulting GPS and CHAMP satellites,  $\vec{m}_1$  and  $\vec{m}_2$  are the unit vectors of the wave vector of the transmitted and received signals,  $n_1$  and  $n_2$  are the refractivities at the corresponding satellite positions. The Doppler shift  $f_d$  corresponds to the measured phase as

$$f_d = -\frac{f_c}{c} \frac{dL}{dt}, \quad (3)$$

with the carrier frequency  $f_c$  and the vacuum speed of light  $c$ . It is possible to decompose  $L$  as

$$L = L_0 + dA_{L_0}, \quad (4)$$

so it is possible to decompose the Doppler shift

$$f_d = f_{d_0} + f_{d_A}. \quad (5)$$

The first term describes the frequency shift without atmospheric influence, caused by the relative motion of the satellites and can be calculated from precise orbit data. The second term is the time derivative of calibrated atmospheric induced signal delay of the occultation link which is composed of an ionospheric part and a part of the neutral atmosphere. The angle of refraction  $\alpha$  can be derived as follows

$$\alpha = \Phi_1 + \Phi_2 + \Theta - \pi. \quad (6)$$

The angles  $\Phi_1$  and  $\Phi_2$  are the only unknowns in Eq. 6, after solving the scalar product and the introduction of the measured Doppler shift in Eq. 2 [Kursinski et al.(1997)Kursinski, Hajj, Schofield, K., and Hardy] and can be derived under the assumption of local spherical symmetry of the refractivity  $n = n(r)$  using Snells law

$$r_1 n(r_1) \sin \Phi_1 = r_2 n(r_2) \sin \Phi_2 = a. \quad (7)$$

Eq. 2 and Eq. 7 are a non linear system which can not be solved analytically, but with a simple iterative procedure [Gorbunov et al.(1996)Gorbunov, Sokolovskiy, and Bengtsson]. Eq. 7 also provides the impact parameter  $a$ . Furthermore an ellipsoid and an ionospheric correction have to be applied. As a next step the refractivity index  $n$  can be derived via an inverse Abel transform [Fjeldbo et al.(1971)Fjeldbo, Eshleman, and Kliore]

$$n(a) = \exp \left[ \frac{1}{\pi} \cdot \int_a^\infty \frac{\alpha(a')}{\sqrt{a'^2 - a^2}} da' \right]. \quad (8)$$

The refractivity as a function of height  $N(a)$  is obtained via Eq. 8

$$\begin{aligned} N(a) &= 10^6 \cdot (n(a) - 1) \\ z(a) &= \frac{a}{n(a)} - R_c \end{aligned}$$

where  $R_c$  denotes the local radius of curvature of the Earth's ellipsoid at the occultation location. Bending angles above  $\sim 45$  km are dominated by ionospheric effects [Hocke(1997)]. Since the ionosphere is a dispersive medium and thus causes different  $L_1$  and  $L_2$  phase delays, it is possible to remove these effects to first order by linear combination of these two signals. The method of linear correction of bending angles Eq. 9 has been applied most successfully [Vorobev and Krasnilnikova(1994)] by

$$\alpha_{LC}(a) = \frac{f_1^2 \alpha_1(a) - f_2^2 \alpha_2(a)}{f_1^2 - f_2^2}, \quad (9)$$

where  $\alpha_{LC}$  denotes the ionosphere corrected bending angle,  $\alpha_1$  and  $\alpha_2$  the uncorrected bending angles of the  $L_1$  and  $L_2$  signals. Still, retrieval results above 20 – 30 km are sensitive to residual ionospheric noise (resulting from higher order terms, which are not corrected by Eq. 9) and other errors like receiver noise, residual clock errors, local multipath and orbit uncertainties. Since the upper integration limit of the inverse Abel transform in Eq. 8 ranges to infinity it needs in practice some kind of high altitude initialization. To avoid downward propagation of errors via the Abel transform itself and subsequently via the hydrostatic integration Eq. 24 one has to be careful. To minimize these errors the concept of statistical optimization is applied [Sokolovskiy and Hunt(1996)]. The best linear unbiased estimator (BLUE Eq. 10; [Loescher(2004)])

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K} (\mathbf{y} - H(\mathbf{x}_b)) \quad (10)$$

$\alpha_{opt}$  is derived from an observed  $\alpha_O$  and a background  $\alpha_B$  bending angle profile under the assumption of unbiased Gaussian errors. The  $\mathbf{O}$  and  $\mathbf{B}$  are the observation and background error covariance matrices, respectively. The  $\alpha_{opt}$  is derived by

$$\alpha_{opt} = \alpha_B + (\mathbf{B}^{-1} + \mathbf{O}^{-1})^{-1} \cdot \mathbf{B}^{-1} \cdot (\alpha_O - \alpha_B) . \quad (11)$$

Where  $\alpha_{opt}$  is a fused bending angle profile dominated by the background in the upper part and by the observation in the lower part. The IGAM retrieval scheme integrates background information only at one point of the retrieval (at bending angle level), so that the results have well defined error characteristics. Since Eq. 11 assumes that the errors of  $\alpha_O$  and  $\alpha_B$  are uncorrelated nowadays ECMWF fields already containing RO information might pose a problem using this approach. One has to be careful if background information is used in a retrieval procedure, if the retrieved data is used in a consecutive assimilation framework. If the assimilation framework uses the same background data as the retrieval we end up with a so-called *incest problem*<sup>1</sup>. The analysis in the assimilation procedure is artificially drawn closer to the background than justified. At IGAM, statistical optimization is implemented in two ways, both relying on Eq. 11, but using different sources of background information and different ways of preprocessing of this information. IGAM/MSIS uses bending angle profiles extracted from the MSIS-90 climatology [Hedin(1991)] and applies best fit profile library search and bias correction procedures [Gobiet et al.(2004b)Gobiet, Steiner, Retscher, Foelsche, and Kirchengast] in order to diminish known biases in the climatology [Randel et al.(2002)Randel, Chanic, and Michaut]. IGAM/ECMWF uses bending angle profiles derived from ECMWF operational analyses. For the assimilation experiment data derived with the IGAM/ECMWF framework has been used, since the first guess fields are derived from ERA40 data no incest problem occurs.

---

<sup>1</sup>cf. 3.7.

	<b>IGAM/MSIS</b>	<b>IGAM/ECMWF</b>
<b>Outlier Rejection and Smoothing</b>	$3\sigma$ outlier rejection on phase delays and smoothing using regularisation	Like IGAM/MSIS
<b>Ionospheric Correction</b>	Linear combination of bending angles. Correction is applied to low-pass filtered bending angles (1 km sliding average), $L_1$ height-pass contribution is added after correction. $L_2$ bending angles < 15 km derived via $L_1 - L_2$ extrapolation.	Like IGAM/MSIS
<b>Bending Angle Initialisation</b>	Statistical optimisation of bending angles 30-120 km. Vertical correlated background (corr. length $L=6$ km) and observation ( $L=1$ km) errors. Obs. error estimated from obs. profile > 60 km. Background error:15%. Background information: MSIS-90 best fit-profile, bias corrected.	Like IGAM/MSIS, but co-located bending angle profile derived from ECMWF operational analysis (above ~60 km: MSISE-90) as background information. No further processing.
<b>Hydrostat. Integral Init.</b>	At 120 km: pressure= $p(\text{MSISE-90})$ .	Like IGAM/MSIS
<b>Quality Control</b>	Refractivity 5-35 km $\frac{\Delta N}{N} < 10\%$ ; Temperature 8-25 km: $\frac{\Delta T}{T} < 25K$ ; Reference: ECMWF analysis.	Like IGAM/MSIS

**Table 2:** Overview of IGAM CHAMP-RO retrieval schemes (EGOPS/CCR Version 2, March 2004.)

## 2.3 Data Products

From the phase delay measurements a variety of atmospheric parameters can be derived. In theory, some parameters (e.g. dry temperature) could be derived without any background information, but in practice, as mentioned above, the retrieval procedure has to be initialized. For a detailed description of the refractivity formulas (Smith-Weintraub and Thayer formula) cf. Section 3.5.2.

### 2.3.1 Refractivity Profiles

Refractivity profiles are derived as described above from the statistical optimized bending angle  $\alpha$ . This is the retrieval product which is used within the assimilation framework. For data quality reasons only data processed with the IGAM/ECMWF retrieval scheme below 35 km and above 5 km is used. At that altitude (35 km) the influence of the initialization on the retrieved refractivity is expected to be small enough, to pose no major problem.

### 2.3.2 Temperature Profiles

The dry temperature  $T_{\text{Dry}}$  can then be derived from Eq. 37 or Eq. 38 by neglecting the effect of water vapor (ignoring the wet terms,  $k_1$  empirical constant cf. Section 3.5.2)

$$N = k_1 \cdot \frac{pA}{T_{\text{Dry}}} , \quad (12)$$

which is valid in the mid- to upper troposphere and stratosphere. If the temperature is below 250 K, the temperature error caused by a 50% error of the water vapor climatology is less than 1 K [Kursinski et al.(1996)]. So the assumption of a *dry atmosphere* can be expanded down to the ground at high latitudes beginning from the subpolar regions on. If this assumption does not hold, *a priori* information about the humidity below  $\sim 7$  km is necessary to solve the ambiguity. Assuming a *dry atmosphere*, using Eq. 12 and introducing the ideal gas law

$$p_A = \frac{\rho_A T_{\text{Dry}} R_{\text{Dry}}}{m_A}, \quad (13)$$

where  $\rho_A$  denotes the dry air density,  $T_{\text{Dry}}$  the *dry air temperature*,  $R_{\text{Dry}}$  the universal gas constant for dry air,  $p_A$  the *dry air pressure* and  $m_A$  the mean molar mass of dry air, it follows with the use of Eq. 12

$$\rho_A = \frac{m_A}{k_1 R_{\text{Dry}}} \cdot N, \quad (14)$$

what means that the density of air is directly proportional to the refractivity and thus can be derived directly. If the vertical air density  $\rho(z)$  is known the vertical pressure can be derived using the equation of hydrostatic equilibrium

$$dp_A(z) = -g(z)\rho_A(z)dz, \quad (15)$$

and integration over  $z$

$$p_A(z) = \int_z^\infty g(z')\rho_A(z')dz'.$$

A second application of Eq. 13 allows to derive the vertical profile of the dry temperature  $T_{\text{Dry}}$

$$T_{\text{Dry}} = k_1 \frac{p_A(z)}{N(z)} \quad (16)$$

### 2.3.3 Humidity Profiles

To derive humidity profiles *a priori* information about the humidity is necessary to resolve the ambiguity. An iterative procedure to calculate specific humidity profiles works as follows:

1. Assumption of dry atmosphere

$$q(z) = 0. \quad (17)$$

2. Calculation of the virtual temperature profile

$$T_v(z) = T(z) \cdot (1 + 0.608 \cdot q(z)). \quad (18)$$

3. Calculation of the pressure profile as in Eq. 24

$$p(z) = \frac{T(z)^2}{c_2} \cdot \left( n(z) - 1 - c_1 \cdot \frac{p(z)}{T(z)} \right), \quad (19)$$

4. Calculation of the specific humidity profile

$$q(z) = \frac{0.622 \cdot p_w(z)}{(p(z) - 0.378 \cdot p(z))}. \quad (20)$$

With the calculated  $q(z)$  the iteration starts again at step 2; the procedure converges fast [Gorbunov and Sokolovskiy(1993)]. An other approach would be a 1D-Var procedure to determine the most likely state of the atmosphere taking background information into account [Healy and Eyre(2000)].

### 2.3.4 Geopotential

The geopotential height profile can be computed corresponding to a given geometric height profile. The (geodetic) latitude dependence of gravitation weighted by  $\frac{g_{\text{Equ}}}{g_0}$  is needed and calculated as factor

$$g_{\text{fact}} = \frac{g_{\text{Equ}}}{g_{\text{Mean}}} + \frac{0.00531}{g_{\text{Mean}}} \cdot \sin(\varphi_j)^2 . \quad (21)$$

To calculate the geopotential height the relation

$$dZ = \frac{g}{g_0} dz \quad (22)$$

is used. The geopotential height is calculated from the geometric height by integration

$$Z(z) = \int_{z_1}^{z_n} \left( \frac{r_{\text{Mean}}}{r_{\text{Mean}} + \frac{1}{2} \cdot h(z')} \right)^2 \cdot g_{\text{fact}} \cdot h(z') dz' . \quad (23)$$

### 2.3.5 Pressure

Eq. 24 describes the calculation of dry pressure  $p_A(z)$  which is equal to the atmospheric pressure if humidity  $p_W(z)$  can be neglected, i.e., everywhere above the lower to middle troposphere

$$p_A(z) = \frac{M_d}{k_1 R} \cdot \int_z^{\infty} g(z') \cdot N(z') dz' . \quad (24)$$

### 2.3.6 Total Electron Content

For the ionosphere, phase changes measured with a dual band GPS receiver can be used to calculate electron density profiles. This specific data product is especially valuable for Space Weather applications [Jakowski et al.(2004)Jakowski, Heise, Wehrenpfennig, and Tsybulya].

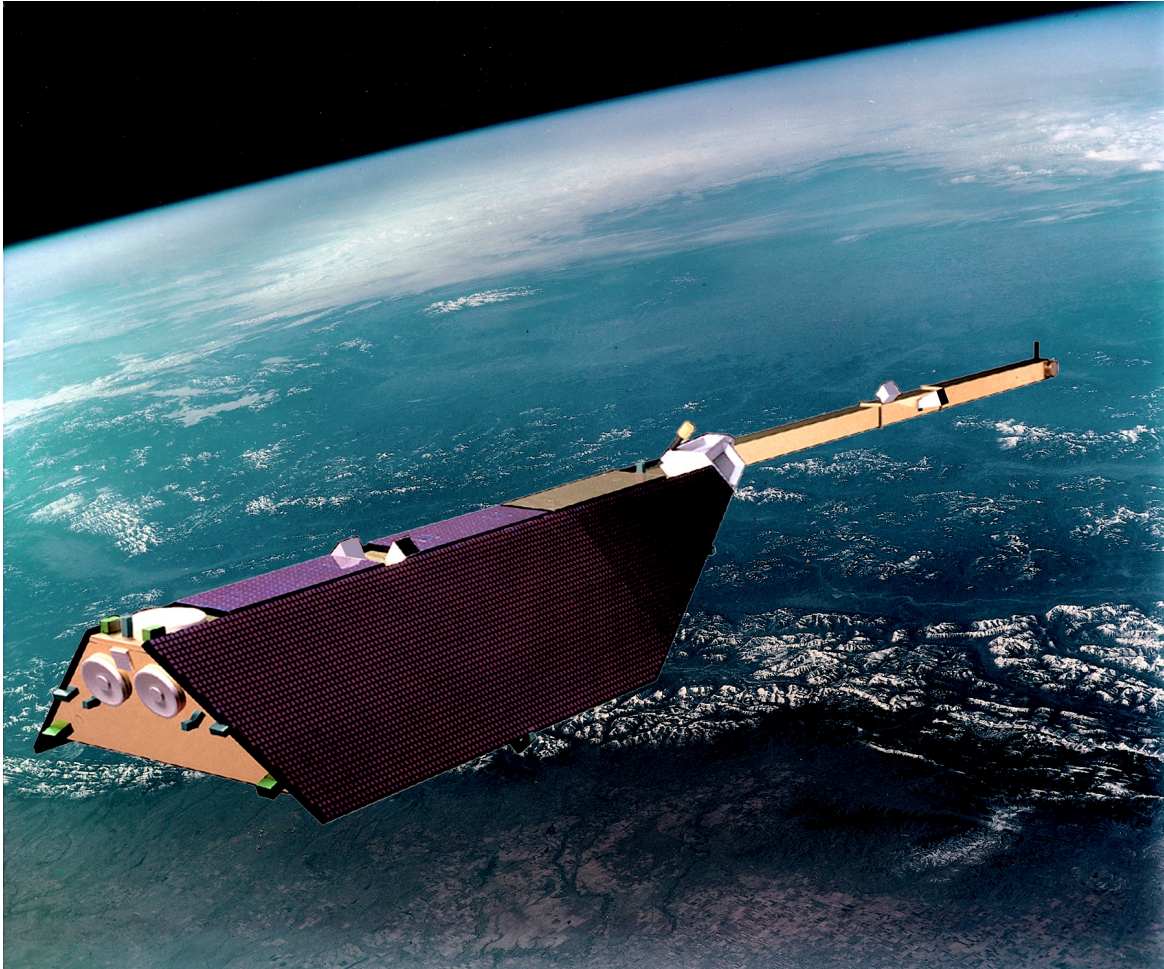
## 2.4 The CHAMP Satellite

The CHAMP satellite was launched from the COSMODROM at Plesetzsk,  $\sim 800$  km North of Moscow, at the 15 July 2000, 12:00 UTC on-board a COSMOS-3B launcher. The initial orbit was nearly circular ( $\varepsilon = 0.004$ ) at a height of 454 km and an inclination of  $87.3^\circ$ . Due to the atmospheric drag the orbit height is decreasing approximately 50 to 200 m/day, depending on the solar activity. The TRSR-2 (*Black Jack*) receiver used for the RO experiment is a key component of the science payload and serves several purposes onboard CHAMP [Reigber et al.(1995)Reigber, Schwinzer, and Kohlhasse], [Wickert et al.(2002)Wickert, Schmidt, Marquardt, Reigber, Neumayer, Beyerle, Galas, and Grunwald], [Wickert et al.(2001)], [Reigber et al.(2003)Reigber, Lühr, and Scheintzer].

## 3 Assimilation System Setup

The system is implemented as a 3D-Var scheme using control space transformations (projection from model levels onto the weighted eigenvectors of the vertical component of the background error covariance matrix) and recursive filters (to represent the horizontal component of the background error covariance matrix). The dimensions of the background are flexible, but due to the specific application a GCM compliant N48 Gaussian grid corresponding to, 96 latitudes  $\times$  192 longitudes comprising 60 model levels is used.





**Figure 3:** Artist view of the CHAMP satellite in orbit (courtesy GFZ Potsdam, NASA picture archive, 2004)

### 3.1 Coordinate System

As mentioned above a GCM compliant N48 Gaussian grid i.e.  $96 \text{ latitude} \times 192 \text{ longitude}$ , comprising 60 model levels is used. The vertical coordinate system is derived from the surface pressure and the A and B vectors provided by ECMWF. This vertical grid comprises 60 hybrid levels. From this basic vertical coordinate system grids of geopotential height and geometric height (over reference ellipsoid) can be derived. The assimilation scheme can be used either with geometric height or geopotential height. The necessary operators are discussed later within this Chapter.

### 3.2 Temperature, Specific Humidity and Surface Pressure Analysis

The assimilation scheme directly updates the temperature, specific humidity, and surface pressure input fields, meaning that all fields are interpolated separately. At each iteration the new vertical fields (like pressure, geopotential and height grid) have to be derived from the updated surface pressure field. That means that the whole vertical coordinate system is shifted up and down during the optimization process. That implies the necessity to calculate the interpolation coefficients for the background fields at observation location at every cycle (*simulation*) before the refractivity

can be derived. After comparison of background refractivity (also denoted as *model observation*) and observation the gradients of the input fields and observations are calculated and a suitable correction is applied.

### 3.3 Variational Assimilation 3D-VAR

The solution of the minimization problem can be performed either in terms of full-fields  $J(\mathbf{x}_a)$  or the analysis of increments  $J(\delta\mathbf{x}_a = \mathbf{x}_a - \mathbf{x}_b)$  [F. Bouttier(1999)]. The latter solution method provides *optimal* analysis increments, which are added to the unmodified background field. This procedure has a number of advantages like the use of linearized control variable transforms which allow the straightforward use of adjoints to calculate the gradient of the cost function [Barker et al.(2004)Barker, Huang, Guo, Bourgeois, and Xiao]. Another advantage is that any imbalance introduced through the analysis procedure is limited to the small increments which are added to the balanced first guess. Since the system doesn't *know* that negative humidities are physically not existing they occur during the minimization cycles. It is taken care of these values by setting them to  $10^{-12}$  and the corresponding gradients to 0.

#### 3.3.1 3D-VAR and Incremental 3D-VAR

3D-VAR uses observations and background information in a statistical optimal way to derive combined analysis fields. The principal formulation is as follows:

$$\mathbf{x}_a = \text{Arg min } J(\mathbf{x}) \quad (25)$$

$$J(\mathbf{x}) = J_b(\mathbf{x}) + J_o(\mathbf{x}) \quad (26)$$

$$J(\mathbf{x}) = \frac{1}{2} \left\{ (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{y} - H(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - H(\mathbf{x})) \right\} \quad (27)$$

$$\nabla J_a = \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_b) + \mathbf{H}^T \mathbf{R}^{-1} [H(\mathbf{x}_0) - \mathbf{y}] = 0. \quad (28)$$

The Solution of the minimization problem requires the calculation of the gradient  $\nabla J_a$  and can be performed either in terms of full fields  $\mathbf{x}$  or in terms of an analysis of the increments  $\delta\mathbf{x}$  :

$$J(\delta\mathbf{x}) = J(\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_b) \quad (29)$$

$$J(\mathbf{x}) = \frac{1}{2} \left\{ \delta\mathbf{x}^T \mathbf{B}^{-1} \delta\mathbf{x} + (\mathbf{H}(\delta\mathbf{x}) - \mathbf{d})^T \mathbf{R}^{-1} (\mathbf{H}(\delta\mathbf{x}) - \mathbf{d}) \right\} \quad (30)$$

$$\mathbf{d} = \mathbf{y} - H(\mathbf{x}_b) \quad (31)$$

$$\nabla J = \mathbf{B}^{-1} \delta\mathbf{x} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}(\delta\mathbf{x}) - \mathbf{H}^T \mathbf{R}^{-1} \mathbf{d} \quad (32)$$

where the analysis is found by adding the final increment to the first guess:

$$\mathbf{x}_a = \mathbf{x}_b + \delta\mathbf{x}_a \quad (33)$$

$\mathbf{x}$  is the atmospheric state vector containing the control variables,  $\mathbf{x}_a$  the updated state vector at convergence,  $J_b(\mathbf{x})$  and  $J_o(\mathbf{x})$  are the background cost function respectively observation cost function,  $\nabla J_a$  the gradient at convergence,  $H$  the potentially non-linear observation operator,  $\mathbf{H}$  the linear approximation (tangent linear operator) of  $H$ ,  $\mathbf{H}^T$  the adjoint operator,  $\mathbf{y}$  the observation vector,  $\mathbf{R}$  and  $\mathbf{B}$  are the observation respectively background covariance matrices.

## 3.4 Implementation Technique

### 3.4.1 Control Variables

The control variables used in the analysis are temperature, specific humidity and surface pressure. To avoid negative specific humidities in the analysis the field is checked for negative values at every

iteration, which are corrected to a value of  $10^{-12}$ , the corresponding gradients are set to 0. The cross correlations between the control variables are assumed to be small enough to be neglected. This assumption serves to effectively block-diagonalize the background error covariance matrix. For each control variable there still remains both, horizontal and vertical correlations. These are assumed to be separable, which is a widely used simplification.

### 3.4.2 Minimization

The cost function is minimized by using an iterative descent algorithm, which is in our case the L-BFGS-B routine. The cost of the analysis is proportional to the number of cost function and its gradient evaluations, denoted as *simulations*. If a new state  $\mathbf{x}$  is found, an *iteration* is performed, which means that to find a new state  $\mathbf{x}$ , several simulations may be required (cf. Section 3.11).

### 3.4.3 Preconditioning

A preconditioning of the problem is performed as part of the control space transformations (cf. Section 3.9), a good paper on preconditioning is [Zupanski(1993)].

### 3.4.4 The Adjoint Technique

The adjoint technique allows to calculate the gradients, needed for the minimization procedure, at reasonable numerical cost [INRIA(2002)].

## 3.5 Observation Operators

### 3.5.1 Interpolation Operator

To calculate the background values at the spatial location of the measurements two bilinear horizontal and one logarithmic vertical interpolation are performed for each observation.

**Horizontal Interpolation** The bilinear interpolation consists of a weighted average of the four surrounding grid points to determine their interpolated value. Two linear interpolations on opposite sites are performed followed by a consecutive interpolation of these intermediate results. This horizontal interpolation is performed for the atmospheric layer above and below the observation

$$f(x, y) = (1 - u)(1 - v)f_{i,j} + u(1 - v)f_{i+1,j} + (1 - u)v f_{i,j} + uv f_{i+1,j+1} , \quad (34a)$$

$$u = \frac{(x - x_i)}{(x_{i+1} - x_i)} , \quad (34b)$$

$$v = \frac{(y - y_j)}{(y_{j+1} - y_j)} , \quad (34c)$$

where  $x_i < x < x_{i+1}$  and  $y_j < y < y_{j+1}$ .

**Vertical Interpolation** Due to the fact of a globally non uniform vertical grid, the heights of the horizontal interpolated values are also calculated by bilinear interpolation from the vertical background grid.

**Linear and Logarithmic Interpolation** Given the background value above and below the spatial location of the observation either a logarithmic or a linear interpolation is performed to get the final value of the background at the location of the observation.

$$\text{Linear Interpolated Value} = Z_1 * \text{Weight}_1 + Z_2 * \text{Weight}_2 \quad (35)$$

$$\text{Logarithmic Interpolated Value} = e^{(\log(Z_1)*\text{Weight}_1)+(\log(Z_2)*\text{Weight}_2)} \quad (36)$$

where  $Z_1$  and  $Z_2$  denote the horizontal interpolated values of the layers above and below the observation which are weighted with  $\text{Weight}_1$  and  $\text{Weight}_2$ , calculated from the vertical distance between observation and  $Z_1$ ,  $Z_2$ . Temperature is linear interpolated, for specific humidity and pressure a logarithmic interpolation is used.

### 3.5.2 Refractivity Operator

To calculate the refractivity at a given point, it is necessary to know the atmospheric variables absolute temperature, specific humidity and pressure. There are two standard formulas used, which are the Thayer and the Smith-Weintraub formula. The Thayer Formula is the more accurate one, the Smith-Weintraub formula is basically the same, but assumes an ideal gas.

#### Thayer Formula

$$N = k_1 \cdot \frac{p_A}{T} \cdot \frac{1}{z_A} + k_2 \cdot \frac{e}{T} \cdot \frac{1}{z_W} + k_3 \cdot \frac{e}{T^2} \cdot \frac{1}{z_W} \quad (37)$$

#### Smith-Weintraub Formula

$$N = k_1 \cdot \frac{p_A}{T} + k_2 \cdot \frac{e}{T} + k_3 \cdot \frac{e}{T^2} \quad (38)$$

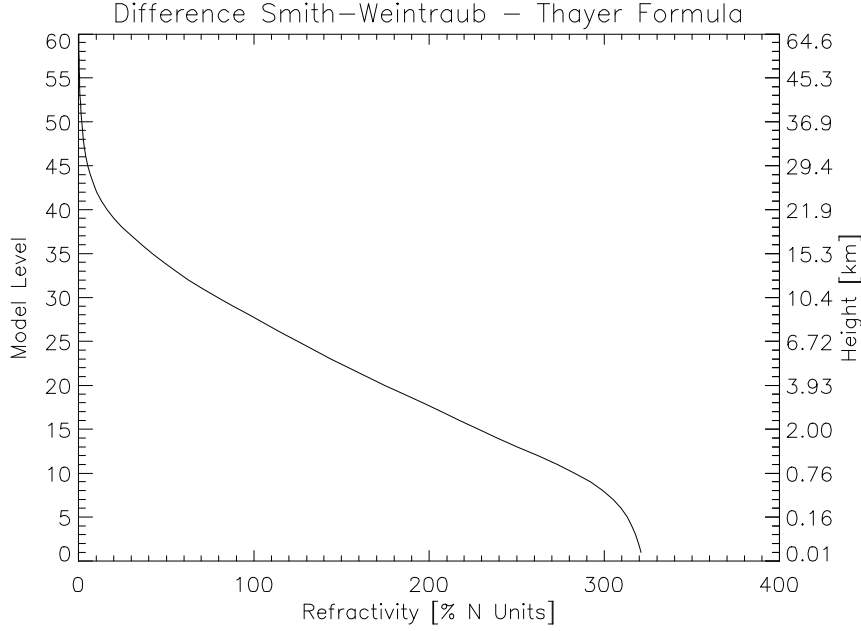
$N$	=	Refractivity	[ ]
$T$	=	Absolute Temperature	[K]
$e$	=	Partial pressure of water vapor	[hPa]
$p_A$	=	Pressure of "dry Air"	[hPa]
$z_D$	=	Compressibility factor of "dry Air"	[ ]
$z_W$	=	Compressibility factor of water vapor	[ ]
$k_1$	=	Empirical constant Thayer Formula	[K/hPa]
$k_2$	=	Empirical constant Thayer Formula	[K/hPa]
$k_3$	=	Empirical constant Thayer Formula	[K <sup>2</sup> /hPa]

**Table 3:** Parameter and Variables used in Thayer & Smith-Weintraub formula.

Further explanations of the parameters and their respective values cf. Appendix B.

**Calculation of Refractivity Fields from ECMWF Fields** Given the fields of temperature, surface pressure and specific humidity (in our case derived fields from ECMWF's ERA40 data), we can calculate the field of refractivity using either the Thayer or Smith Weintraub- formula. A comparison of the two formulas shows no significant differences. A simplified version of the Smith Weintraub - formula was chosen as forward operator, to calculate refractivity from temperature, humidity, and pressure analysis fields, the latter derived from the surface pressure.

Fig. 5 shows the negligible difference concerning our application of the mean global refractivity profile calculated with the Thayer and the Smith Weintraub formula using an ECMWF T42L60 analysis field (Date: 2003.01.03; 12 UTC).



**Figure 4:** Mean global refractivity profile calculated from T42L60 ECMWF analysis fields.

The use of a simple formulation to calculate refractivities is justified within this framework so the refractivity forward operator can be written as:

$$N = k_1 \cdot \frac{p_A}{T} + k_3 \cdot \frac{e}{T^2}. \quad (39)$$

This simple formulation is generally accepted and introduces no significant errors. Further we need the partial pressure of water vapor  $e$  which is given by

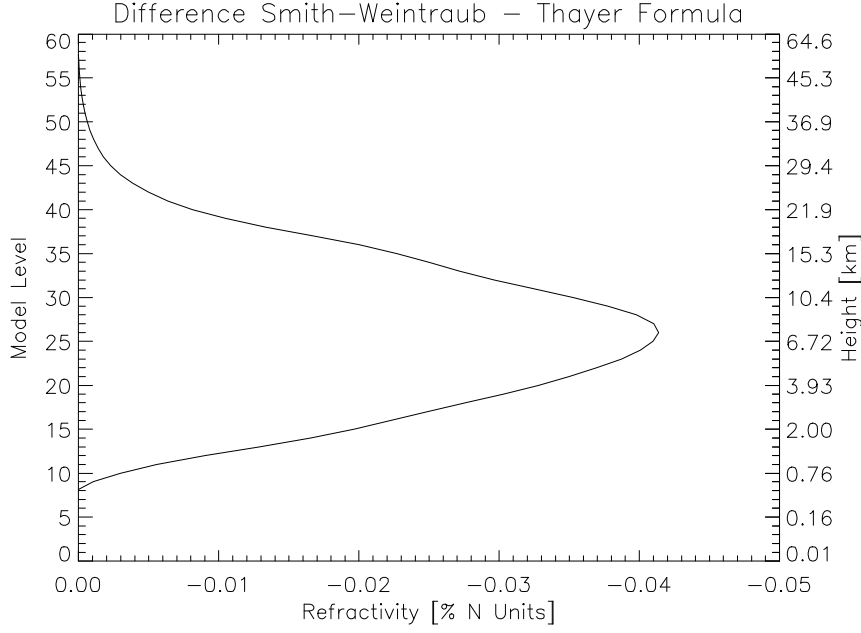
$$p_W = \frac{M_W}{M_D}, \quad (40)$$

$$e = \frac{q \cdot p}{p_W + (1.0 - p_W) \cdot q}, \quad (41)$$

where  $e$  denotes the water vapor pressure,  $M_W$ , respectively  $M_D$  the molecular mass of water vapor and dry air.

### 3.5.3 Vertical Coordinate Operator

As can be seen in Eq. 37 and Eq. 38 the pressure at the location of the observation is needed to calculate the refractivity. ECMWF provides temperature, specific humidity, and surface pressure fields. The background pressure field is derived by a series of operators. These operators are also used to set up the vertical coordinate system of the assimilation scheme either as a vertical grid of geopotential heights or a vertical grid of geometric heights. Since the pressure field is also derived, pressure coordinates are possible but not implemented at the moment. The values of temperature  $T_{i,j,z}$  and specific humidity  $q_{i,j,z}$ , are given for the Gaussian grid of the latitudes  $\varphi_j$  and the homogenous grid of the longitudes  $\lambda_i$ , and an irregular spaced height grid  $z_{i,j,z}$ . For the N48 Gaussian grid the index ranges are  $i = 1 \dots 96$ ,  $j = 1 \dots 192$ , and  $z = 1 \dots 60$  for full level quantities and  $z = 0 \dots 60$  for half level quantities. Here and in other parts of this report  $z$  simply



**Figure 5:** Difference in global mean refractivity calculated with Smith-Weintraub and Thayer Formula.

denotes the height coordinate whether it is geometrical height or geopotential height, but further on in this Section  $z$  denotes geometrical height and  $\phi$  geopotential height. The vertical index of all quantities is always denoted as  $z$ . Coordinates are geocentric except denoted otherwise. The pressure corresponding to the  $z$ th half and full levels are calculated by the means of formulas, e.g., [Roeckner et al.(2003)]

$$P_{i,j,z+\frac{1}{2}} = A_{z+\frac{1}{2}} + B_{z+\frac{1}{2}} \cdot P_{\text{Surf}_{i,j}} , \quad (42)$$

$$P_{i,j,z} = \frac{1}{2} \left( P_{i,j,z+\frac{1}{2}} + P_{i,j,z-\frac{1}{2}} \right) , \quad (43)$$

where  $P_{\text{Surf}_{i,j}}$  denotes the surface pressure at the  $i$ th longitude and the  $j$ th latitude. The  $A_{z+\frac{1}{2}}$  and  $B_{z+\frac{1}{2}}$  are the vertical coordinate parameters provided by ECMWF. The calculation of the geopotential heights is based on the hydrostatic equation and on an interpolation between the half and the full levels [Gorbunov and Kornbluh(2003)]

$$\phi_{i,j,z+\frac{1}{2}} - \phi_{i,j,z-\frac{1}{2}} = R_{\text{Dry}} T_{v,i,j,z} \cdot \ln \left( \frac{P_{i,j,z+\frac{1}{2}}}{P_{i,j,z-\frac{1}{2}}} \right) , \quad (44)$$

$$\phi_{i,j,z_{\text{max}}+\frac{1}{2}} = \phi_{i,j,z_{\text{Surf}}} , \quad (45)$$

$$\phi_{i,j,z} = \phi_{i,j,z+\frac{1}{2}} + \alpha_{i,j,z} \cdot R_{\text{Dry}} T_{v,i,j,z} , \quad (46)$$

$$\alpha_{i,j,z} = \text{Ln}(2) \text{ for } z = 1 , \quad (47)$$

$$\alpha_{i,j,z} = 1 - \frac{P_{i,j,z-\frac{1}{2}}}{P_{i,j,z+\frac{1}{2}} - P_{i,j,z-\frac{1}{2}}} \cdot \text{Ln} \left( \frac{P_{i,j,z+\frac{1}{2}}}{P_{i,j,z-\frac{1}{2}}} \right) \text{ for } z > 1 , \quad (48)$$

where  $T_v$  denotes the virtual temperature as defined in Eq. 18 and  $\phi_{\text{Surf}}$  is the surface geopotential which is equal the orography. The geometrical heights over reference ellipsoid is calculated from

the geopotential heights by the approximate formula from the *US Standard Atmosphere*:

$$R_0 = \frac{2 \cdot 10^{-3} \cdot g_{\text{Surfi},j}}{3.085462 \cdot 10^{-6} + 2.27 \cdot 10^{-9} \cdot \cos(2\varphi_j) - 2 \cdot 10^{-12} \cdot \cos(4\varphi_j)}, \quad (49)$$

$$z_{i,j,z} = R_0 \cdot \frac{\phi_{i,j,z} \cdot (g_{\text{Mean}} - \phi_{i,j,z})}{g_{\text{Surfi},j} \cdot R_0}, \quad (50)$$

where  $R_0$  denotes the effective Earth's radius,  $g_{\text{Mean}}$  the mean gravity acceleration and  $g_{\text{Surfi},j}$  the local gravity acceleration on the surface.  $g_{\text{Surfi},j}$  itself is calculated by the *International gravity formula plus standard z dependence*:

$$g_{\text{Equator}} = 9.7803 \text{ m/s}^2 \quad (51)$$

$$g_{\text{Surfi},j} = g_{\text{Equator}} \cdot \left( 1 + 0.00531 \cdot \sin(\varphi_{\text{Geodetic},j})^2 \right), \quad (52)$$

finally the standard height dependency of  $g_{\text{Surfi},j}$  is calculated

$$R_{\text{Mean}} = 6371.0 \text{ [km]}, \quad (53)$$

$$f_{gz} = \left( \frac{R_{\text{Mean}}}{R_{\text{Mean}} + z} \right)^2, \quad (54)$$

$$g_{\text{Surfi},j,z} = g_{\text{Surfi},j} \cdot f_{gz}, \quad (55)$$

$g_{\text{Equator}}$  is the gravity acceleration at the equator and  $R_{\text{Mean}}$  the mean Earth's radius.

### 3.6 Observation Error Covariance Matrix

The observation error covariance matrix takes only vertical correlations within a profile into account. Due to the separation in space and time between the different RO events this assumption is justified.

#### 3.6.1 Formulation of the Observation Error Covariance Matrix

A simple error covariance matrix formulation was deduced from empirically estimated matrices [Steiner(2004)], [Kirchengast et al.(2006)Kirchengast, Foelsche, and Steiner] (not latitude dependent). A least square method was used to fit analytical functions to the relative standard deviation which shows a different behavior below and above the tropopause height. The empirical relative standard deviation can be approximated with an exponential increase above the upper troposphere/lower stratosphere region between about 14-20 km, where it is closely constant, and with a decrease from near 14 km downwards proportional an inverse law. To be able to scale the error magnitude, which is receiver dependent, the standard deviation in the upper troposphere/lower stratosphere domain ( $s_{\text{utls}}$ ) can be tuned. Eq. 56 gives the analytical functions for the relative standard deviation  $s_z$  over all altitude domains, where  $z$  denotes the height,  $z_{\text{Tropotop}}$  the top level of the "troposphere domain",  $z_{\text{Stratobot}}$  the bottom level of the "stratosphere domain" and  $H_{\text{Strato}}$  which is the scale height of the error increase over the stratosphere [Steiner and Kirchengast(2004)].

$$s_z = \begin{cases} s_{\text{utls}} + s_0 \cdot \left[ \frac{1}{z^p} - \frac{1}{z_{\text{Tropotop}}} \right], & \text{for } 2 \text{ km} < z \leq z_{\text{Tropotop}} \\ s_{\text{utls}}, & \text{for } z_{\text{Tropotop}} < z \leq z_{\text{Stratobot}} \\ s_{\text{utls}} \cdot \exp \left[ \frac{z - z_{\text{Stratobot}}}{H_{\text{Strato}}} \right], & \text{for } z_{\text{Stratobot}} < z \leq 50 \text{ km} \end{cases}. \quad (56)$$

To be able to derive the error covariance matrix the correlation length  $L(z)$  has to be determined. The best values for  $L(z)$  are 2 km within the troposphere (up to ~15km) and a linear decrease of

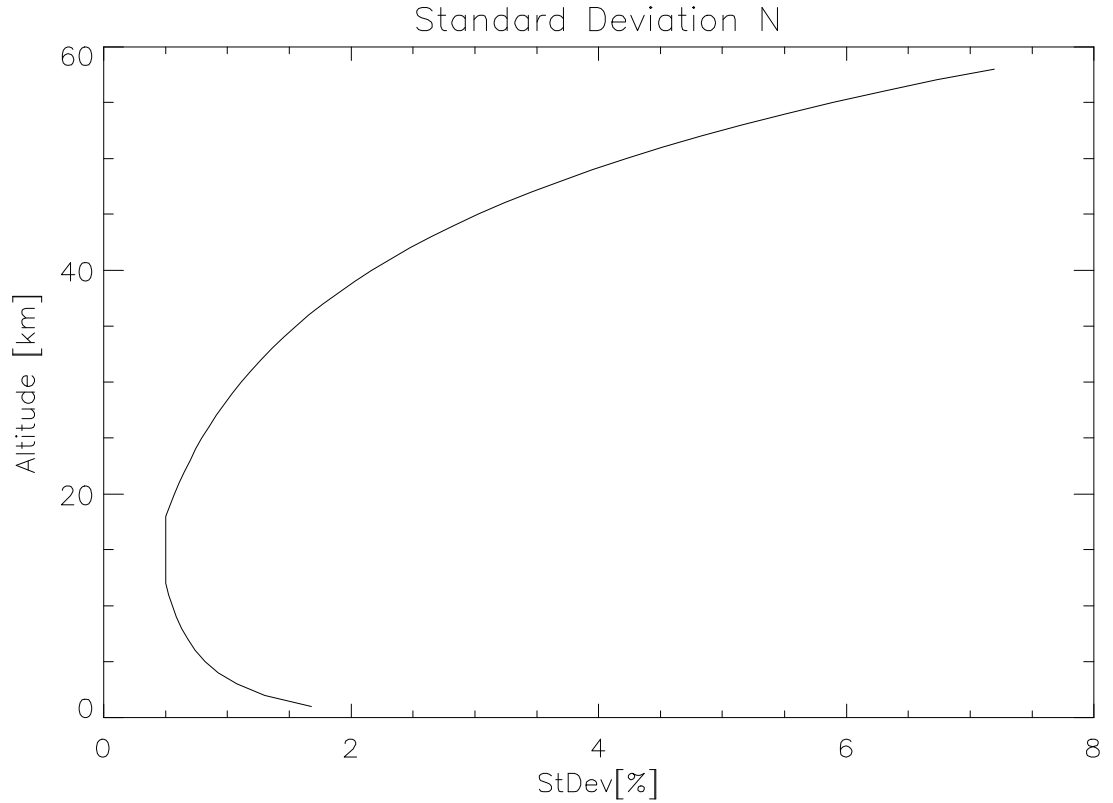
$L(z)$  above the troposphere to 1 km at 50 km altitude. The observation error covariance matrix  $\mathbf{S}$  can now be described as:

$$\mathbf{S} = S_{ij} = s_i \cdot s_j \cdot \exp\left(-\frac{|z_i - z_j|}{L(z)}\right) \quad (57)$$

This formulation of the observation error covariance also accounts for the error of representativeness, so there is no additional specification within the assimilation framework necessary. The values which are currently used for CHAMP data are:

$s_{utls}$ :	0.5%
$s_0$ :	4.5%
$z_{Stratobot}$ :	20 km
$z_{Tropotop}$ :	14 km
$H_{Strato}$ :	15 km
$p$ :	1.0
$L(z)$ , $15 \text{ km} \leq z$ :	Linear decrease to 1 km at 50 km
$L(z)$ , $2 \text{ km} \leq z \leq 15 \text{ km}$ :	2 km

**Table 4:** Parameter and Variables used in formulation of observation error covariance matrix formula.



**Figure 6:** Relative refractivity standard deviation based on the presented formulas and parameters.



### 3.7 Incest Problem

The assumption that there is no correlation between background and observation errors is usually justified, since the causes of the errors are supposed to be completely independent. However, one must be careful about observation preprocessing, like retrieval procedures, which use background information. These procedures can cause a bias of the observations towards the background. If observations, containing background information are used in an assimilation procedure, we are confronted with the so-called *incest problem*. The analysis is drawn closer to the background than justified, which is caused by observations already containing background information, reducing the apparent background departures. If auxiliary information is necessary in the observation preprocessing procedure, one should carefully decide, which additional information will not influence the analysis result. The analysis is only optimal if the assumption of bias-free observations holds. In practice background and observations are often significantly biased. If the biases are known, they can be subtracted from the background and observations, which is in practice a delicate problem. Bias monitoring and removal are subjects of ongoing improvement and research.

### 3.8 Background Error Covariance Matrix

As a suitable time frame to generate first guess fields from ERA40 monthly mean data the years 1980 to 2000 (21 years) have been chosen. This time covers a period of increasingly and lately massive use of satellite observations, improving the analysis over remote and data sparse areas like the southern hemisphere significantly. These monthly mean fields are averaged separately for the four 6 hour time windows around 00,06, 12 and 18 UTC. For this data set the monthly mean variances of the used atmospheric parameters (temperature, specific humidity and surface pressure) are available too.

The approach to use these variances is somewhat sound, problems occur concerning the correlation matrices, since they are not readily available, they had to be derived. The background error covariance matrices are composed of global mean horizontal and vertical correlation structures and the three dimensional respectively two dimensional (for surface pressure) variance fields.

The vertical part of the control space transformation is based on positive definite vertical background covariance matrices (in general covariance matrices have to be positive definite by definition).

#### 3.8.1 Variances

For this ERA40 data set the monthly mean variances of the used atmospheric parameters (temperature, specific humidity and surface pressure) are available directly from the MARS system and calculated as follows:

$$a_{var} = \frac{1}{k-1} \sum_{n=1}^{n=k} (a_n - \hat{a})^2 \quad (58)$$

where  $k$  denotes the number of days,  $a_{var}$  the monthly mean variance of a parameter,  $a$  the daily analysis value and  $\hat{a}$  the corresponding monthly mean. Since monthly mean error patterns are not available this daily variations compared to the monthly mean had been used to derive a proxy which is referred to as *error* from now on.

#### 3.8.2 Calculation of Vertical Correlation Matrices

The vertical correlation matrices are calculated from a subset of the used 21 years period, namely from the years 1980, 1985, 1990, 1995 and 2000 under the assumption of "true" daily analysis fields which are compared with the monthly mean analysis<sup>2</sup>. First the vertical error covariance matrix

<sup>2</sup>Personal communication Mike Fisher 2006

$\mathbf{B}$  is computed from the differences  $\Delta\mathbf{x}_n$  between the monthly mean and the *true values* as with the daily analyses serve, for the  $k$  vertical levels.

$$\mathbf{B} = \left[ \frac{1}{k-1} \right] \sum_{n=1}^{n=k} (\Delta\mathbf{x}_n)(\Delta\mathbf{x}_n)^T \quad (59)$$

The diagonal elements of  $\mathbf{B}$  represent the variances ( $B_{ii}$ ) at vertical level  $i$ ; with its non-diagonal elements representing the covariances ( $B_{ij}$ ) between the vertical levels  $i$  and  $j$ . From  $\mathbf{B}$  the error correlation matrix  $\mathbf{C}$  with its elements  $C_{ij}$  denoting the error correlations between  $\Delta x_i$  at level  $i$  and  $\Delta x_j$  at level  $j$  can be derived.  $\mathbf{C}$  is finally calculated by dividing the covariance  $B_{ij}$  by the square root of the product of variances  $B_{ii}$  and  $B_{jj}$ .

$$\mathbf{C} \text{ with } C_{ij} = \frac{B_{ij}}{\sqrt{B_{ii}B_{jj}}} \quad (60)$$

During calculation of the  $\Delta\mathbf{x}$  using the "true" daily analysis fields which are compared with the monthly mean analysis it turned out that some values of the daily analysis fields had been identical with the corresponding values of the monthly mean analysis ( $\Delta\mathbf{x} = \mathbf{0}$ ). That has been true for temperature, specific humidity and surface pressure, a fact that seems to be odd. During coding of the program to derive the horizontal correlation (cf. 3.8.3) a bug was found concerning the mentioned problem. If  $\Delta\mathbf{x} = \mathbf{0}$ , numerical problems occur, in that case  $\Delta\mathbf{x}$  has to be set to a small value, but due to an implementation bug not only the respective value but the whole column has been set to a small value. Tests within the horizontal correlations framework showed no significant impact with and without the bug due to the averaging procedures. Since there are significantly more values averaged in the calculation of the vertical correlation matrices than in the horizontal test case the occurrence of the bug can be seen as negligible without any real impact on the result.

The same averaging procedure as used for the first guess fields (separate averaging of the four time layers of monthly means over 21 years) had been applied to derive the variances of the control variables. To calculate the vertical correlation matrices the averaging procedure included all four time layers otherwise it had been applied the same way as in the variance case, just restricted to the years 1980, 1985, 1990, 1995 and 2000, separated for 5 latitude regions. Individual correlation matrices for northern high latitudes NH ( $0^\circ - 30^\circ$ ), northern mid latitudes NM ( $30^\circ - 60^\circ$ ) low latitudes LO ( $60^\circ - 120^\circ$ ), and the respective southern counterparts (SM, SH) had been derived and are used within the assimilation scheme.

### 3.8.3 Calculation of Horizontal Point to Point Correlations

To calculate horizontal point to point correlation the following equations could be used

$$\begin{aligned} \mathbf{c}_1 &= c_{11} = 1 \\ \mathbf{c}_i &= \frac{1}{i} (\mathbf{C}_{12} + \mathbf{C}_{23} + \dots + \mathbf{C}_{ii+1}) \\ \mathbf{c}_i &= \frac{1}{i} \sum_{i=1}^{Dim(\mathbf{C})-1} \mathbf{C}_{ii+1} \\ i &= 1..Dim(\mathbf{C}) \end{aligned} \quad (61)$$

which in principal is an averaging of the diagonal and off diagonal elements of a correlation matrix (the averaging of the diagonal elements always equals 1). This procedure gives a rough estimate of the horizontal correlation patterns. Bear in mind that the number of averaged correlated grid points is reduced by distance (meaning that the averaged number is reduced by one for stepping one point ahead, since in fact the diagonal and off diagonal elements of the correlation

matrix are used). The correlation matrices themselves are derived by taking the grid points in east-west direction for every single latitude by means of the equations used to calculate the vertical correlation matrices (cf. 3.8.2) and subsequent averaging over the respective latitudes and all time layers. Since we have a closed surface the correlations are expected to drop but increase again after a while since the last grid point (as expressed in the matrix) is in fact next to the first one.

Another fact to be aware of is that the following plots showing these correlation patterns depict the correlation as function of grid points not as function of distance! That implies that 192 grid points represent a different distance for the high mid and low latitude plots (indicated by H, M, L in the plots where N denotes the northern and S the southern hemisphere). One shouldn't be surprised by the strong correlation at high latitudes (temperature Figs. 7, 10 lower panels, specific humidity Figs. 12, 16 lower panels and surface pressure Figs. 20 upper panel and Fig. 22) since the distance is much smaller than in the low latitude cases (temperature Fig. 9 lower panel, specific humidity Fig. 14 lower panel and surface pressure Fig. 20 lower panel).

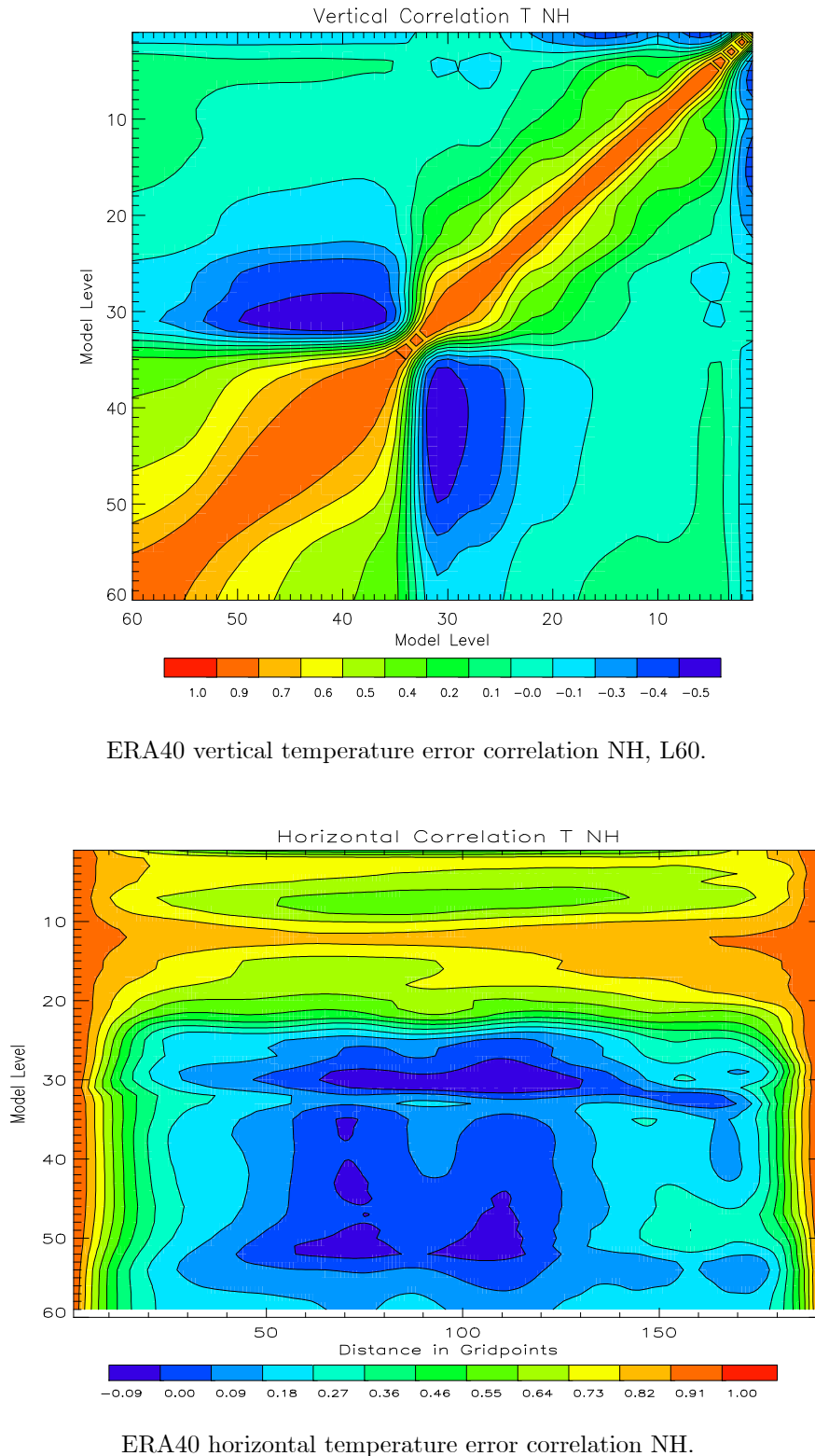
Alternatively the same procedure comprising the same data set has been applied in the north-south direction meaning that the equations had been applied to every single longitude with subsequent averaging as explained before. In this case we don't have a closed surface (bands are running from north to south pole) resulting in matrices of only half the dimension of the east-west case. Since in this case the grid point distance is roughly equal and there are naturally no latitude bands only global means are shown for temperature in Fig. 23 upper panel, for specific humidity in Fig. 23 lower panel and for surface pressure in Fig. 24.

For this setup the grid point separation represents  $\sim 208$  km, taking 20 to 30 grid points into account (translating into a distance of 4160 to 6240 km) might be a good solution to describe the horizontal correlation properties. Although this seems to result in quite broad correlations it might be close to the truth of monthly mean fields of this spatial resolution. Some corrections around the high latitudes might be necessary to avoid numerical problems. These horizontal correlation patterns have still to be tested in practice.

The quantities in the following figures are averaged over latitude bands where  $0^\circ$  stands for the North Pole  $180^\circ$  for the South Pole or global means. In detail:

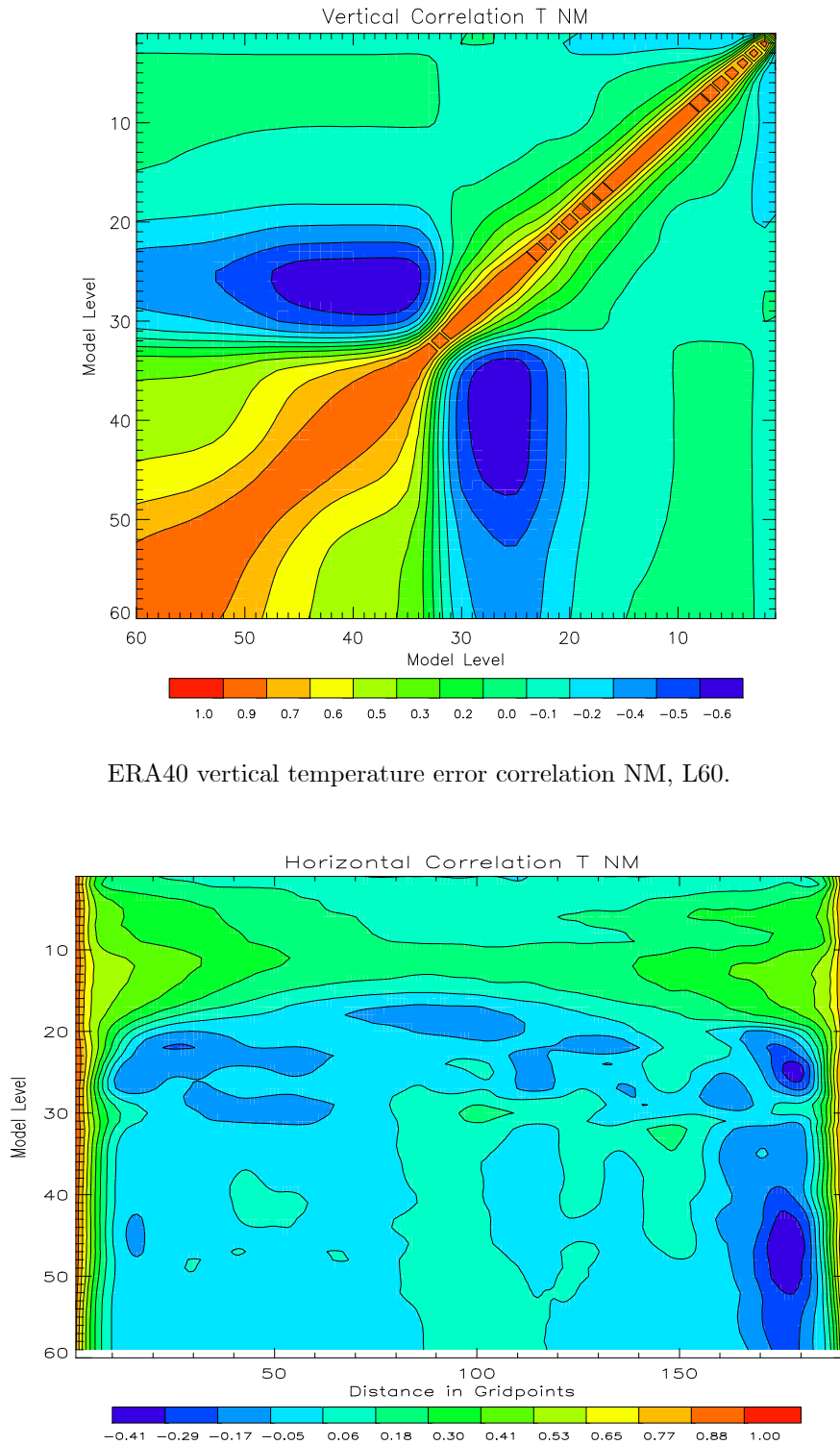
- $0^\circ - 30^\circ$  NH: Northern High Latitudes
- $30^\circ - 60^\circ$  NM: Northern Mid Latitudes
- $60^\circ - 120^\circ$  LO: Low Latitudes
- $120^\circ - 150^\circ$  SM: Southern Mid Latitudes
- $150^\circ - 180^\circ$  SH: Southern High Latitudes

### 3.8.4 Vertical and Horizontal ECMWF Temperature Error Correlations Northern High Latitudes $0^\circ - 30^\circ$ (NH)



**Figure 7:** Derived vertical and horizontal error correlations of ERA40 temperature fields northern high latitudes.

### 3.8.5 Vertical and Horizontal ECMWF Temperature Error Correlations Northern Mid Latitudes 30° - 60° (NM)

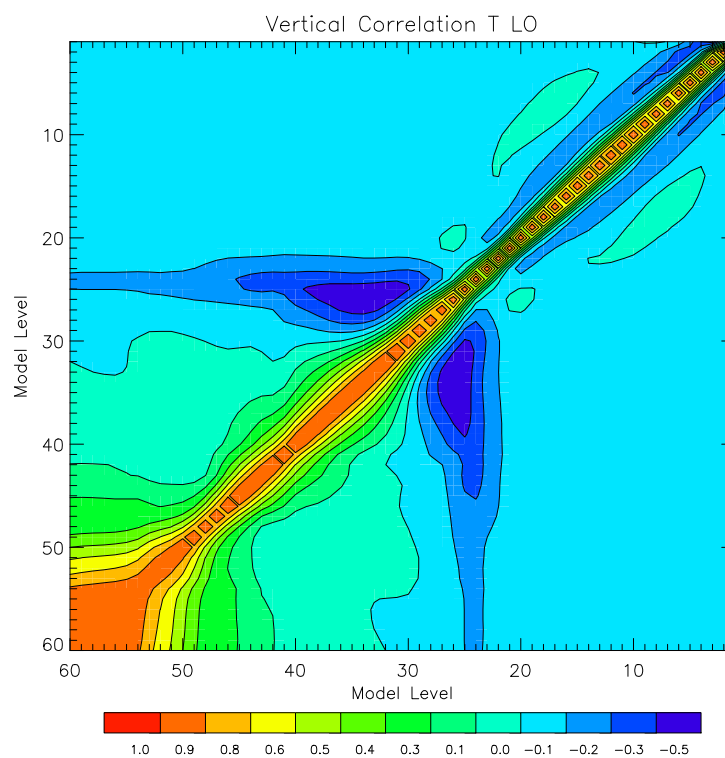


ERA40 vertical temperature error correlation NM, L60.

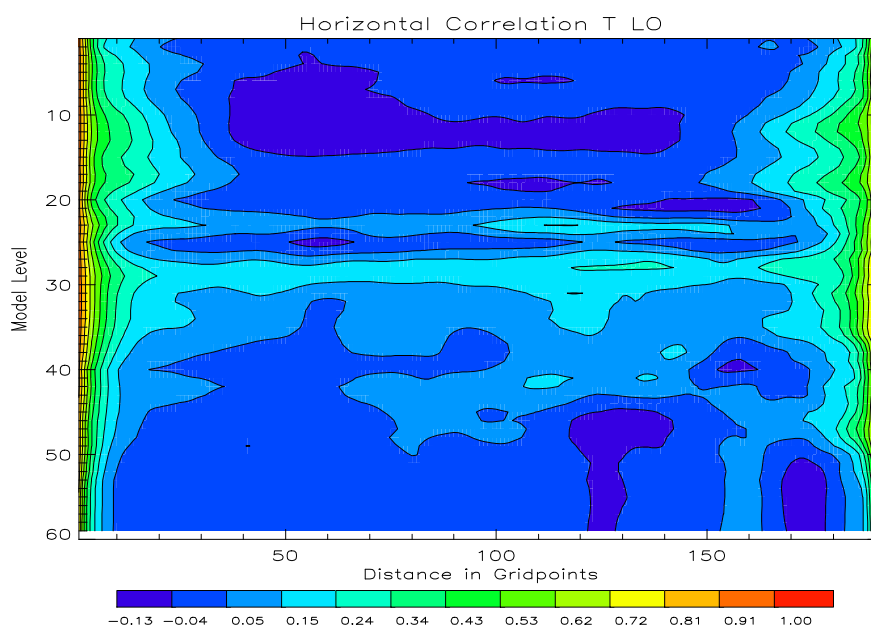
ERA40 horizontal temperature error correlation NM.

**Figure 8:** Derived vertical and horizontal error correlations of ERA40 temperature fields northern mid latitudes.

### 3.8.6 Vertical and Horizontal ECMWF Temperature Error Correlations Low Latitudes 60° - 120° (LO)



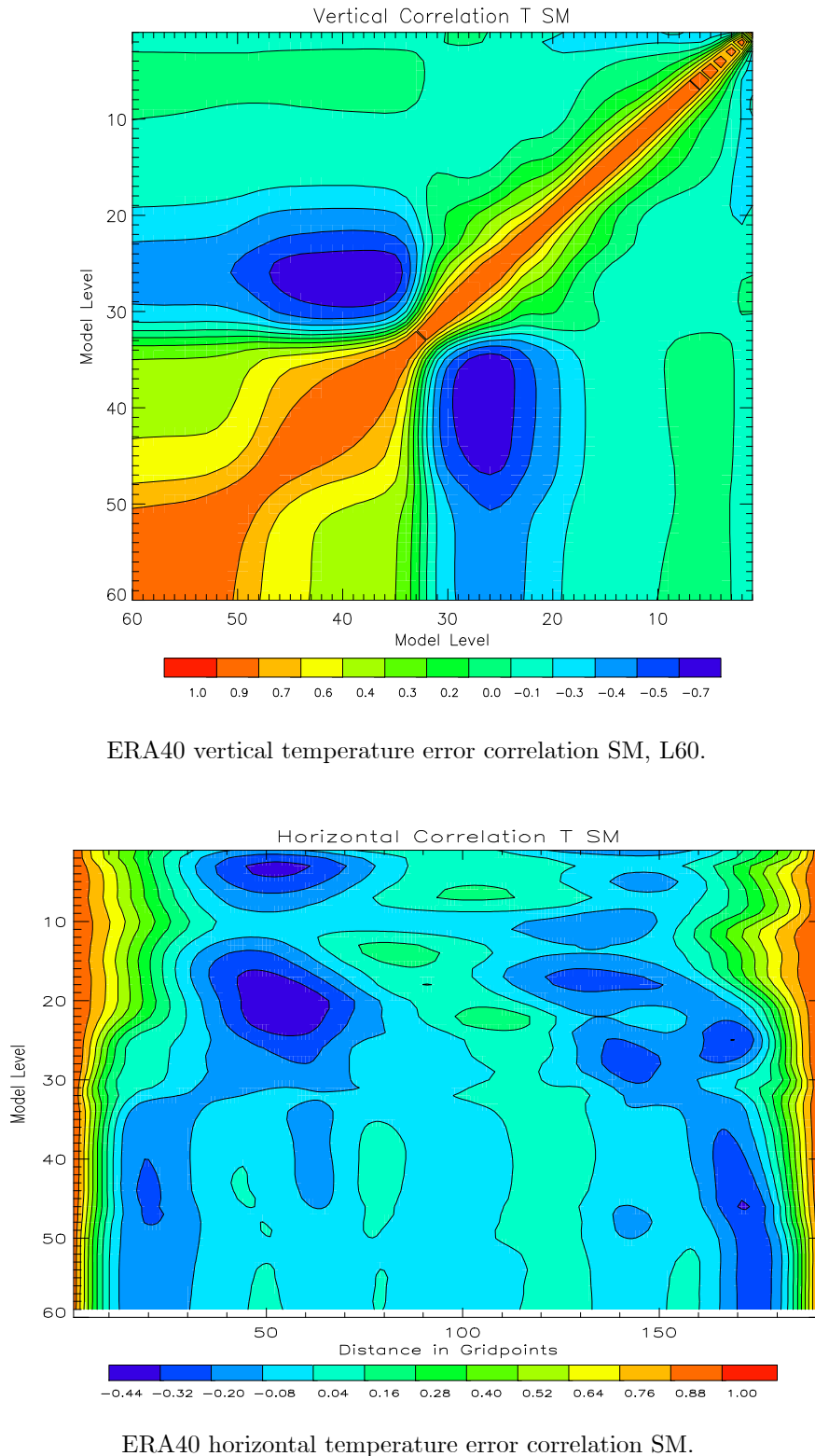
ERA40 vertical temperature error correlation LO, L60.



ERA40 horizontal temperature error correlation LO.

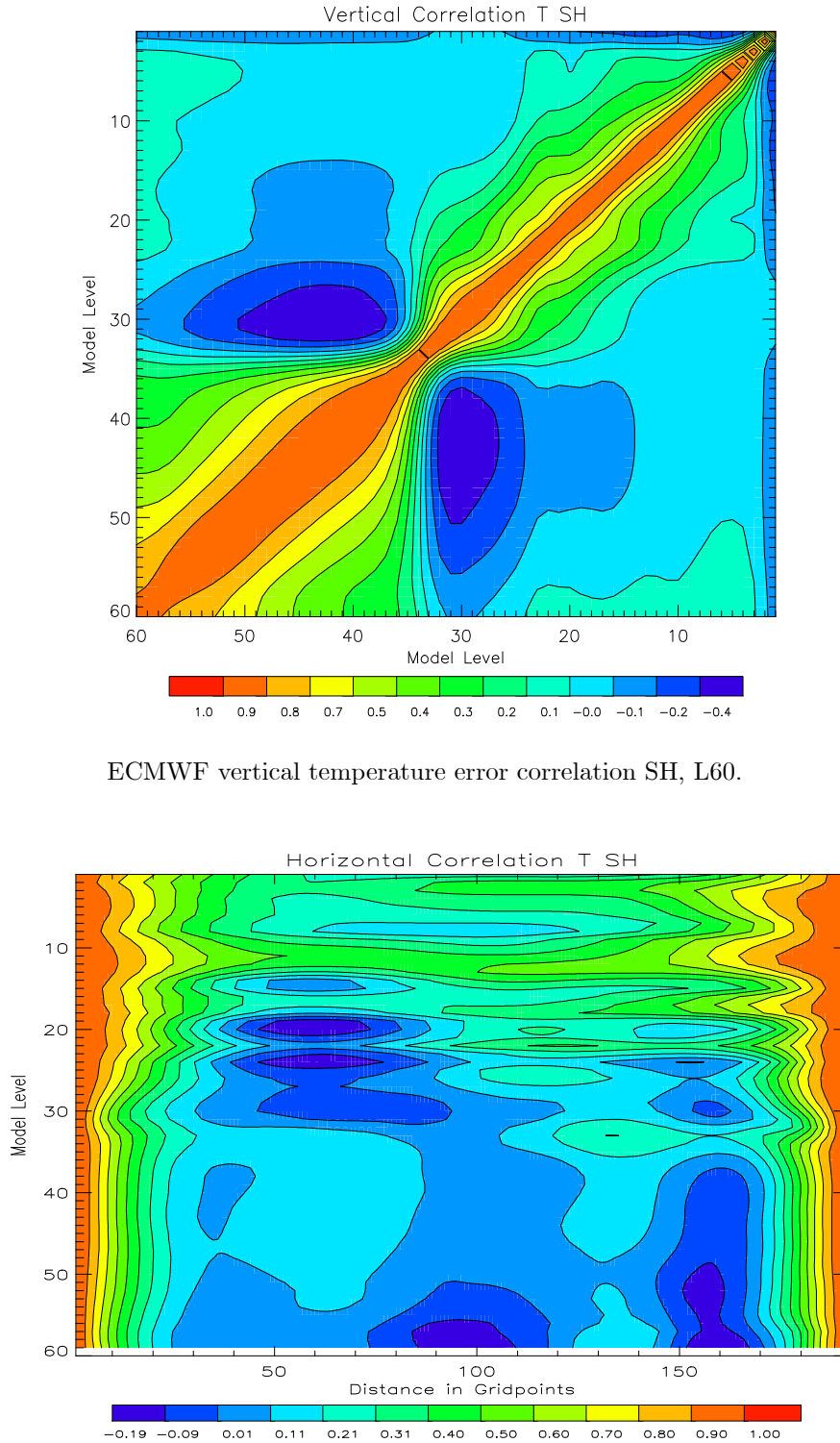
**Figure 9:** Derived vertical and horizontal error correlations of ERA40 temperature fields low latitudes.

### 3.8.7 Vertical and Horizontal ECMWF Temperature Error Correlations Southern Mid Latitudes 120° - 150° (SM)



**Figure 10:** Derived vertical and horizontal error correlations of ERA40 temperature fields southern mid latitudes.

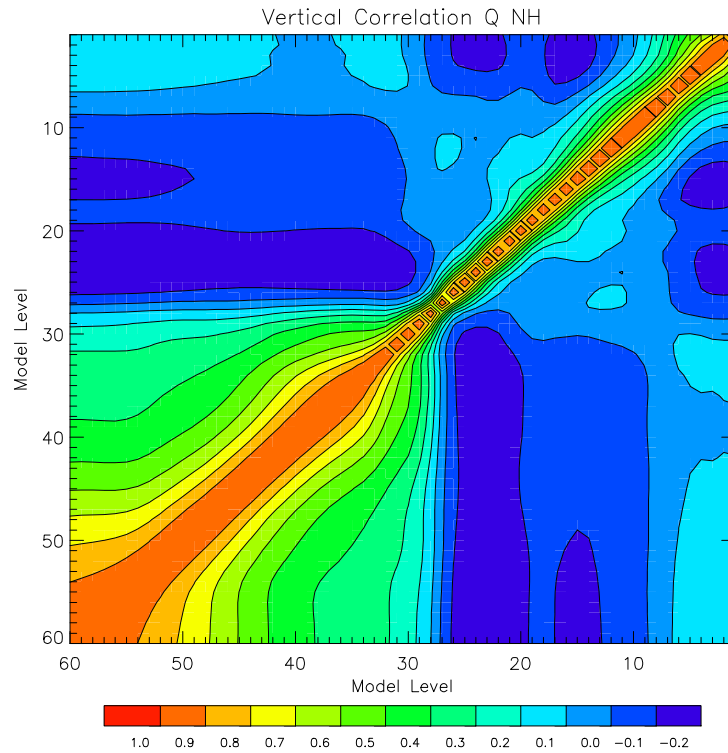
### 3.8.8 Vertical and Horizontal ECMWF Temperature Error Correlations Southern High Latitudes 150° - 180° (SH)



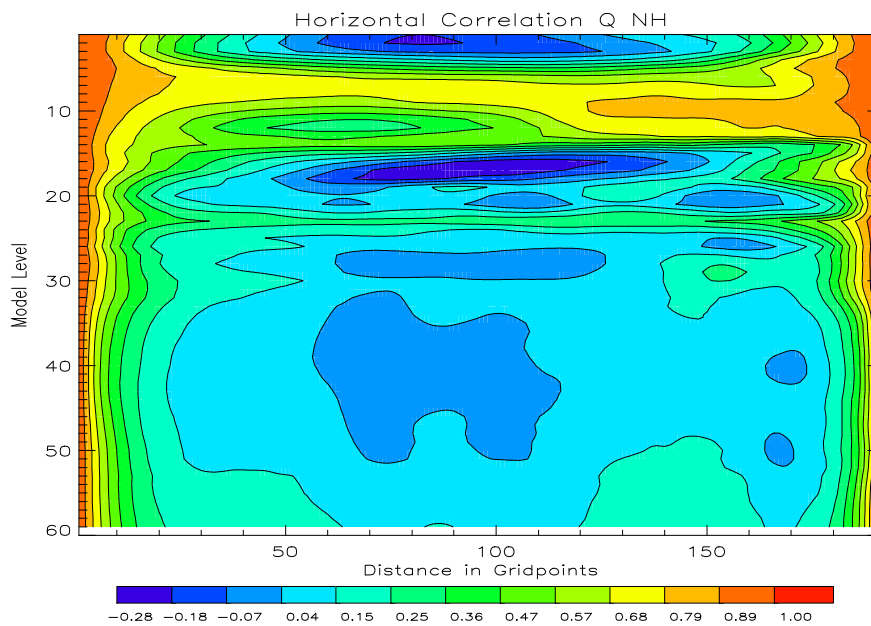
**Figure 11:** Derived vertical and horizontal error correlations of ERA40 temperature fields southern high latitudes.



### 3.8.9 Vertical and Horizontal ECMWF Specific Humidity Error Correlations Northern High Latitudes $0^\circ - 30^\circ$ (NH)



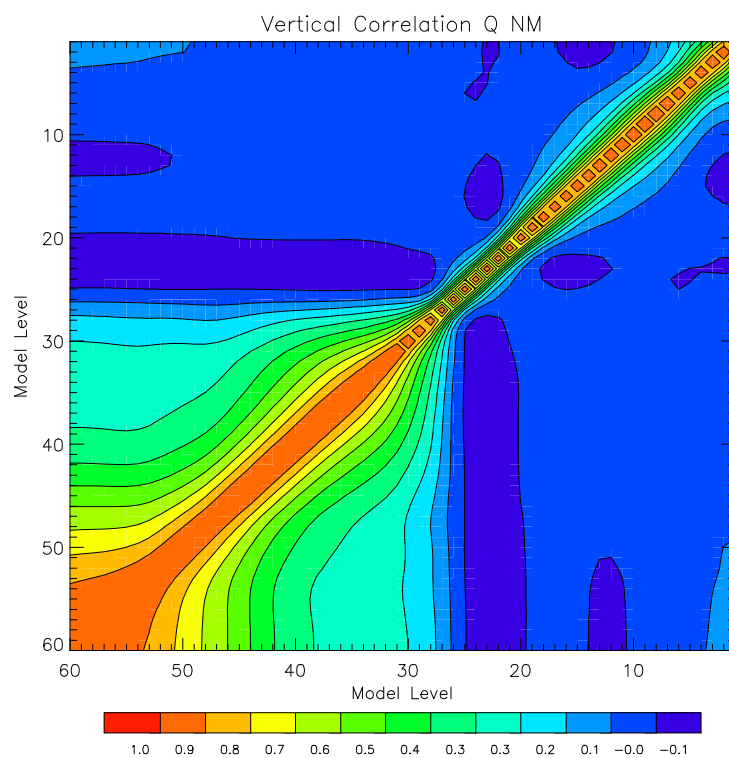
ERA40 vertical specific humidity error correlation NH, L60.



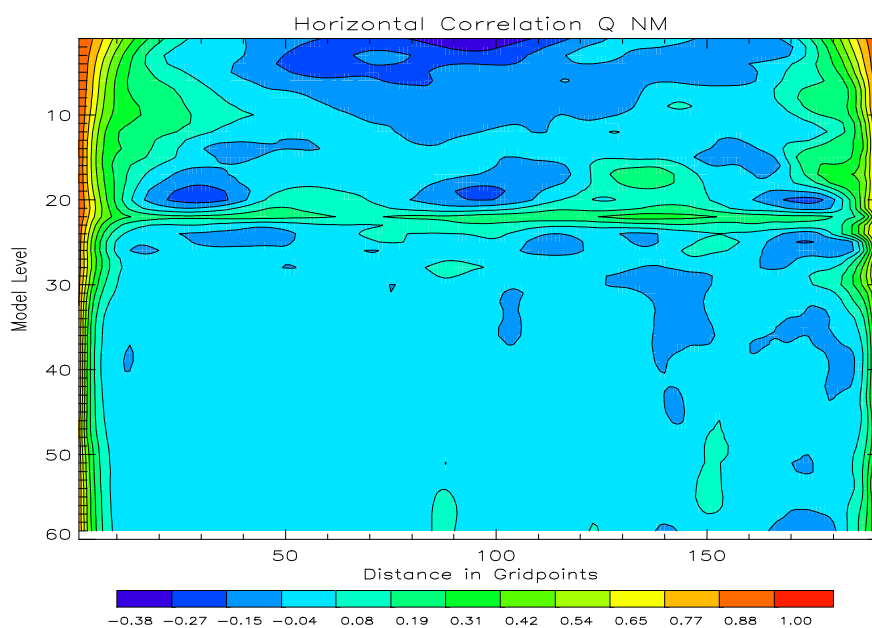
ERA40 horizontal specific humidity error correlation NH.

**Figure 12:** Derived vertical and horizontal error correlations of ERA40 specific humidity fields northern high latitudes.

### 3.8.10 Vertical and Horizontal ECMWF Specific Humidity Error Correlations Northern Mid Latitudes 30° - 60° (NM)



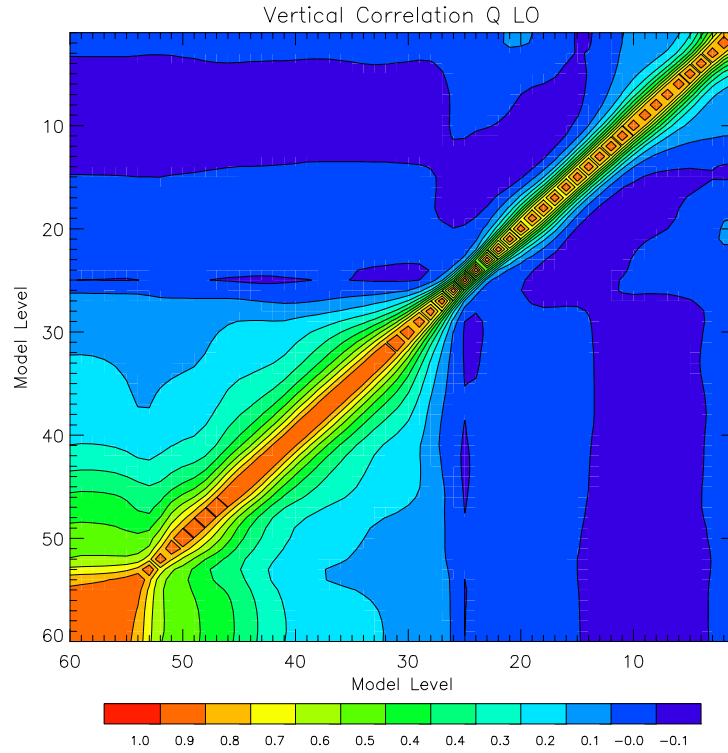
ERA40 vertical specific humidity error correlation NM, L60.



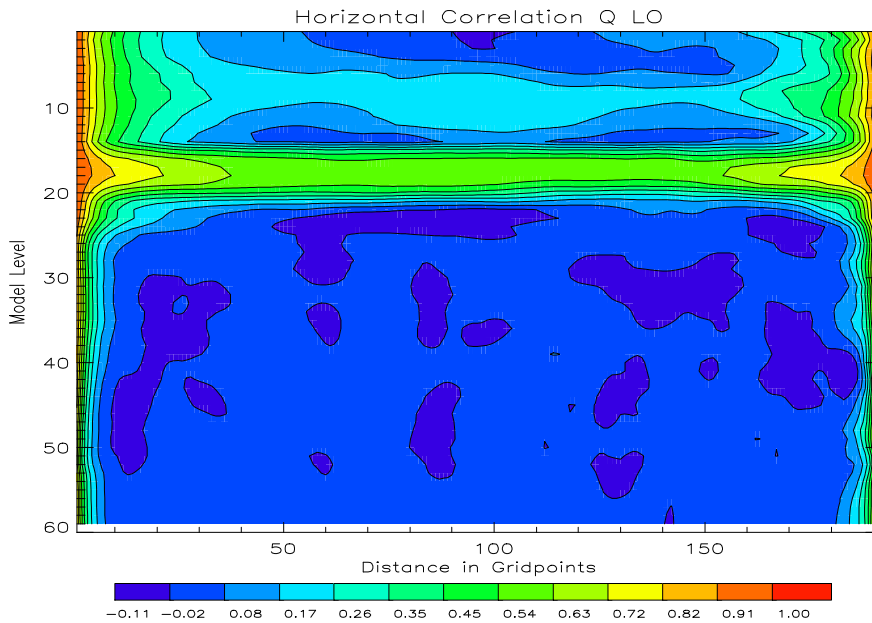
ERA40 horizontal specific humidity error correlation NM.

**Figure 13:** Derived vertical and horizontal error correlations of ERA40 specific humidity fields northern mid latitudes.

### 3.8.11 Vertical and Horizontal ECMWF Specific Humidity Error Correlations Low Latitudes 60° - 120° (LO)



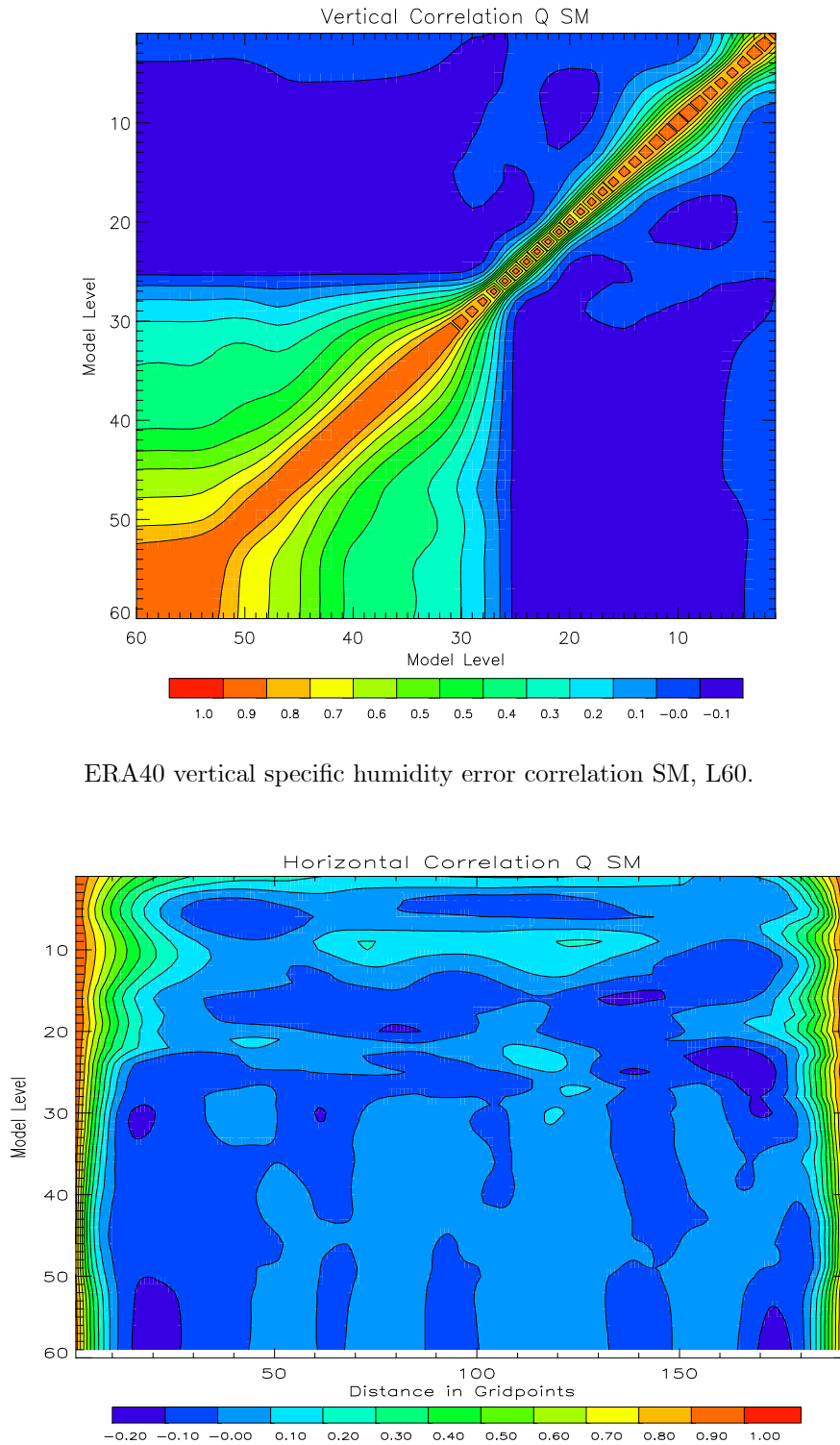
ERA40 vertical specific humidity error correlation LO, L60.



ERA40 horizontal specific humidity error correlation LO.

**Figure 14:** Derived vertical and horizontal error correlations of ERA40 specific humidity fields low latitudes.

### 3.8.12 Vertical and Horizontal ECMWF Specific Humidity Error Correlations Southern Mid Latitudes 120° - 150° (SM)

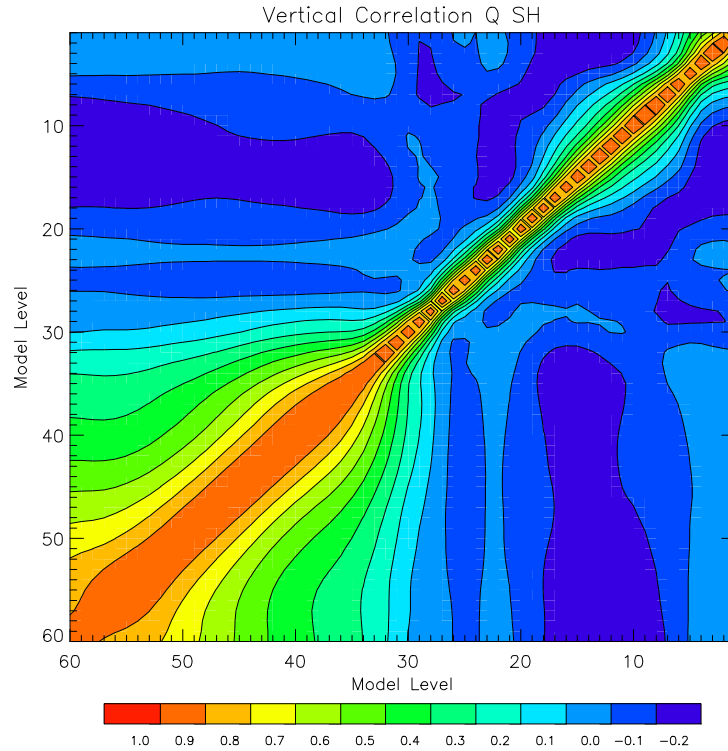


ERA40 vertical specific humidity error correlation SM, L60.

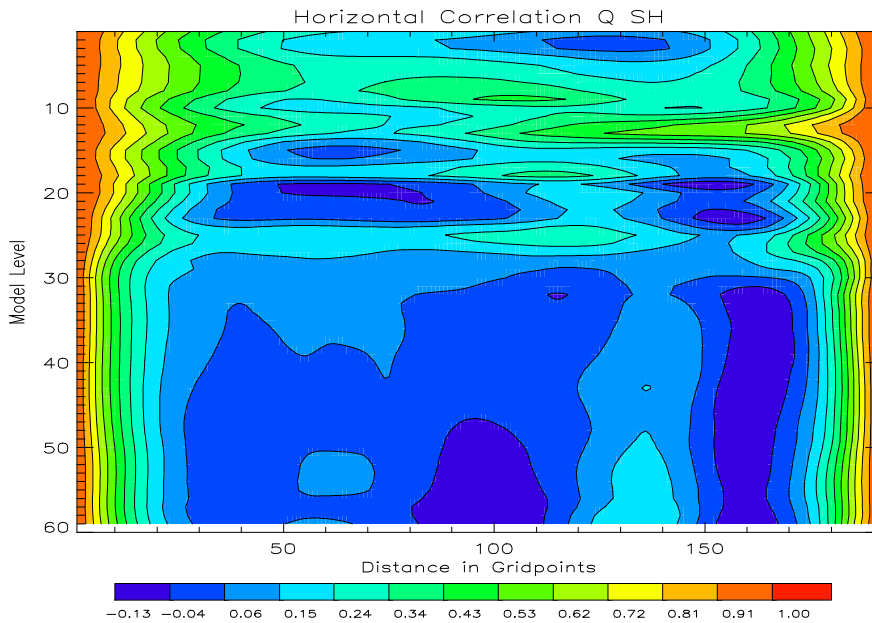
ERA40 horizontal specific humidity error correlation SM.

**Figure 15:** Derived vertical and horizontal error correlations of ERA40 specific humidity fields southern mid latitudes.

### 3.8.13 Vertical and Horizontal ECMWF Specific Humidity Error Correlations Southern High Latitudes 150° - 180° (SH)



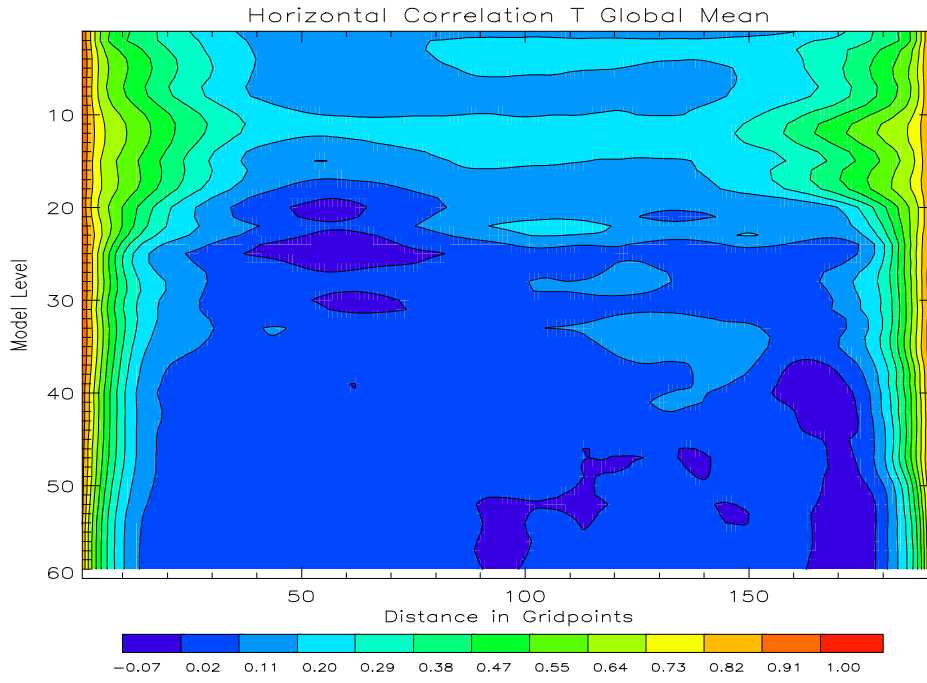
ERA40 vertical specific humidity error correlation SH, L60.



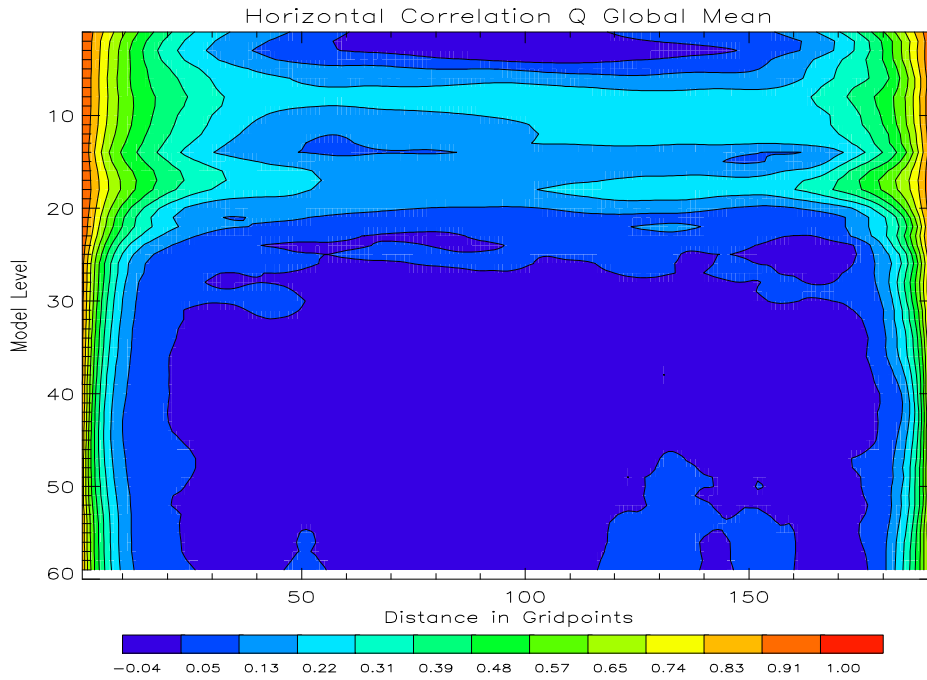
ERA40 horizontal specific humidity error correlation SH.

**Figure 16:** Derived vertical and horizontal error correlations of ERA40 specific humidity fields southern high latitudes.

### 3.8.14 Horizontal ECMWF Temperature and Specific Humidity Error Correlations Global Mean



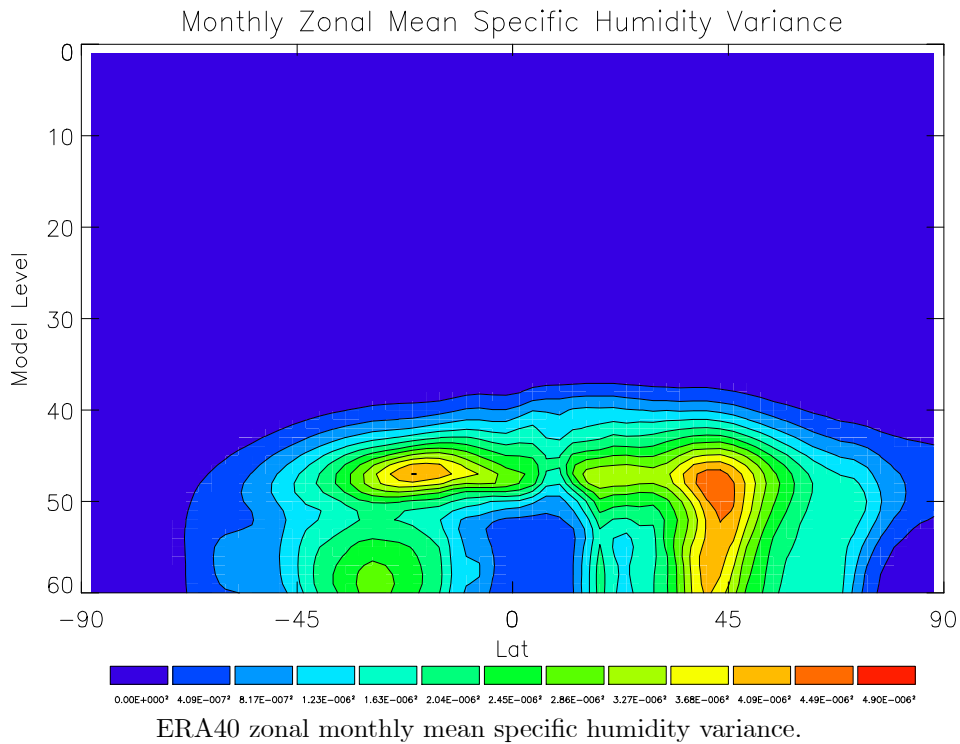
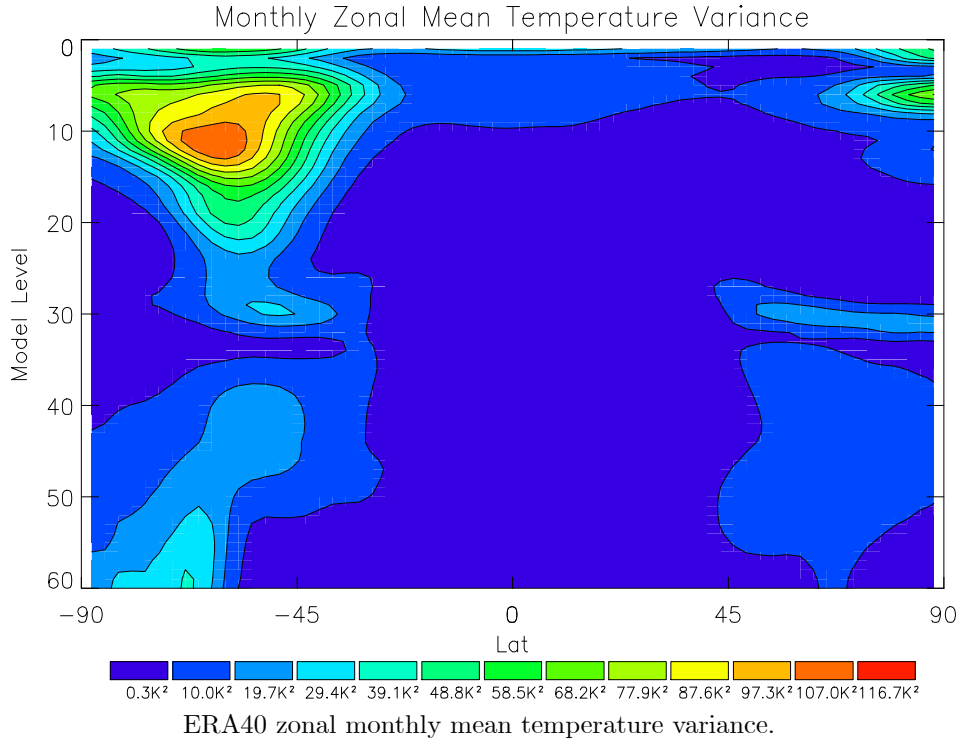
ERA40 vertical temperature error correlation global mean, L60.



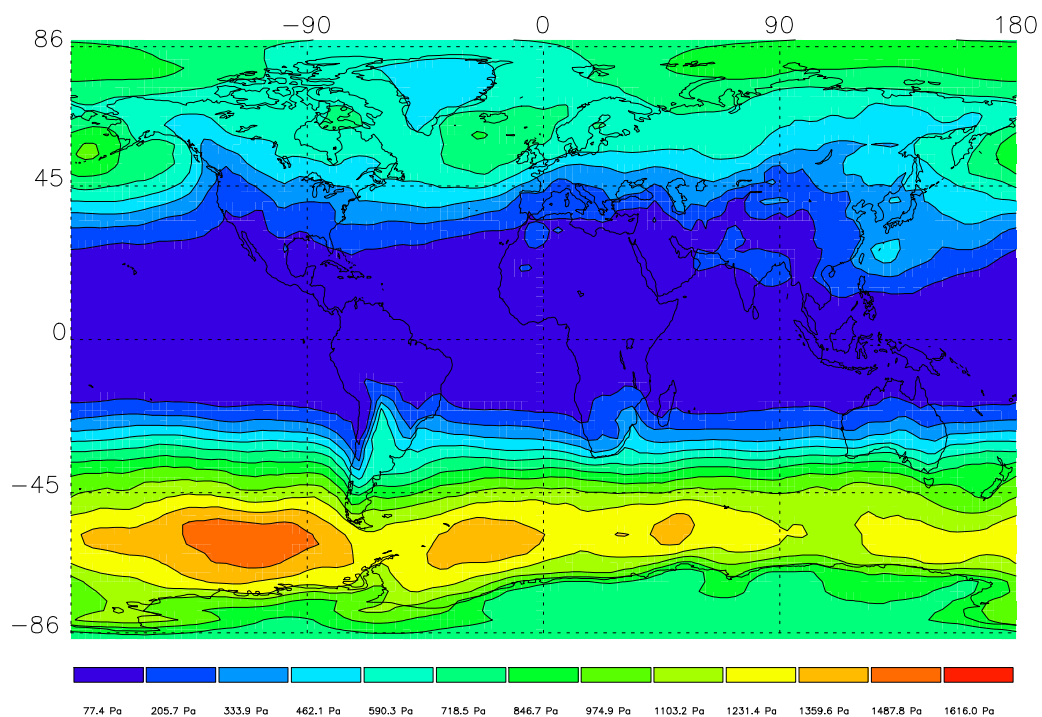
ERA40 horizontal specific humidity error correlation global mean.

**Figure 17:** Derived vertical and horizontal error correlations of ERA40 specific humidity and temperature fields global mean.

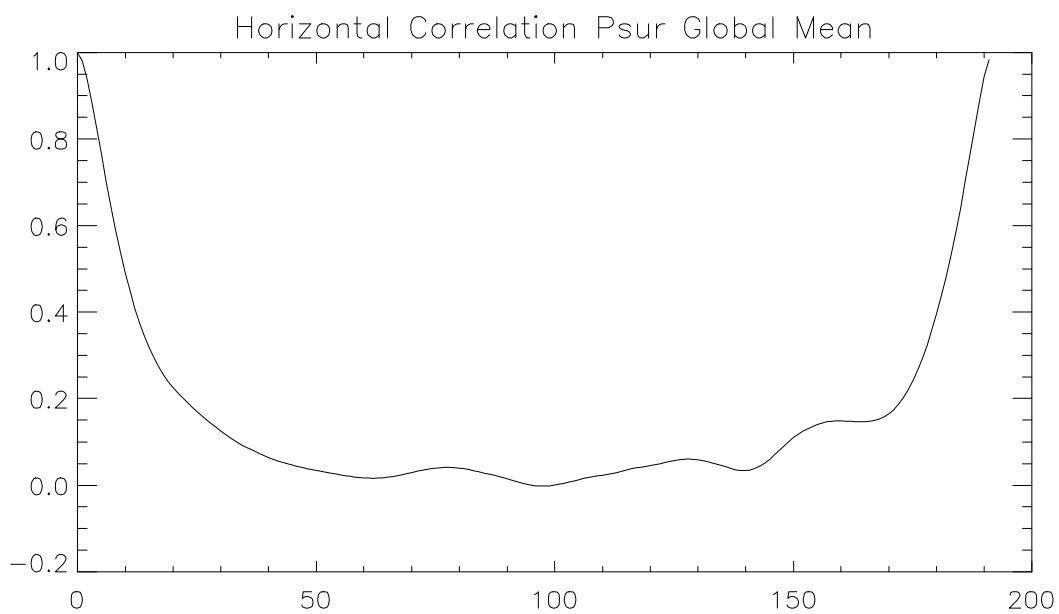
### 3.8.15 Zonal Mean Monthly Temperature and Specific Humidity Variance



**Figure 18:** Zonal mean of temperature and specific humidity ERA40 monthly mean variance fields.

**3.8.16 Global Mean Monthly Surface Pressure Variance and Correlation**

ERA40 global variance of surface pressure.

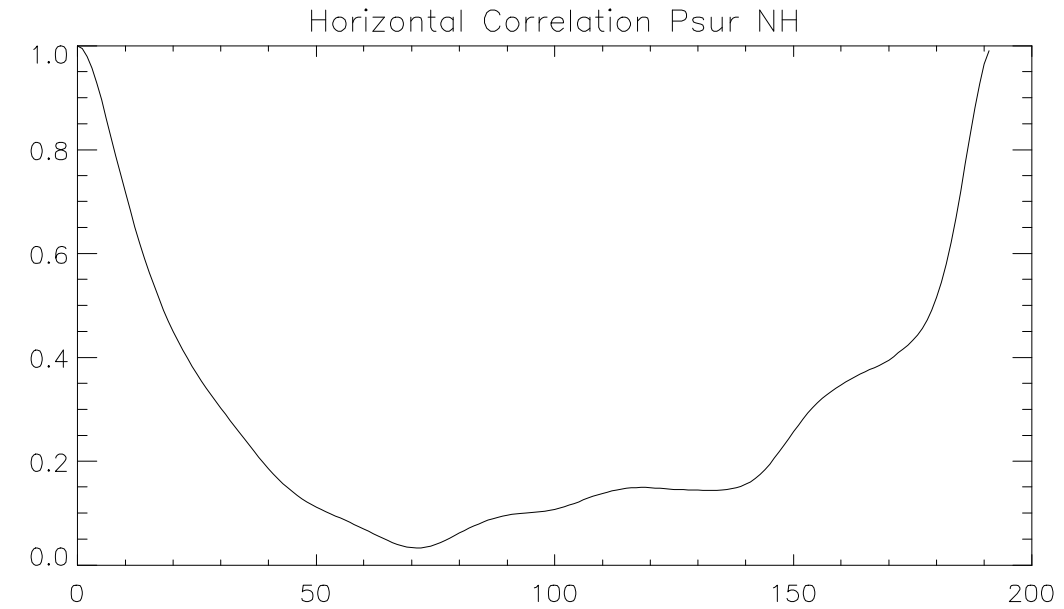


ERA40 error correlation of surface pressure global mean.

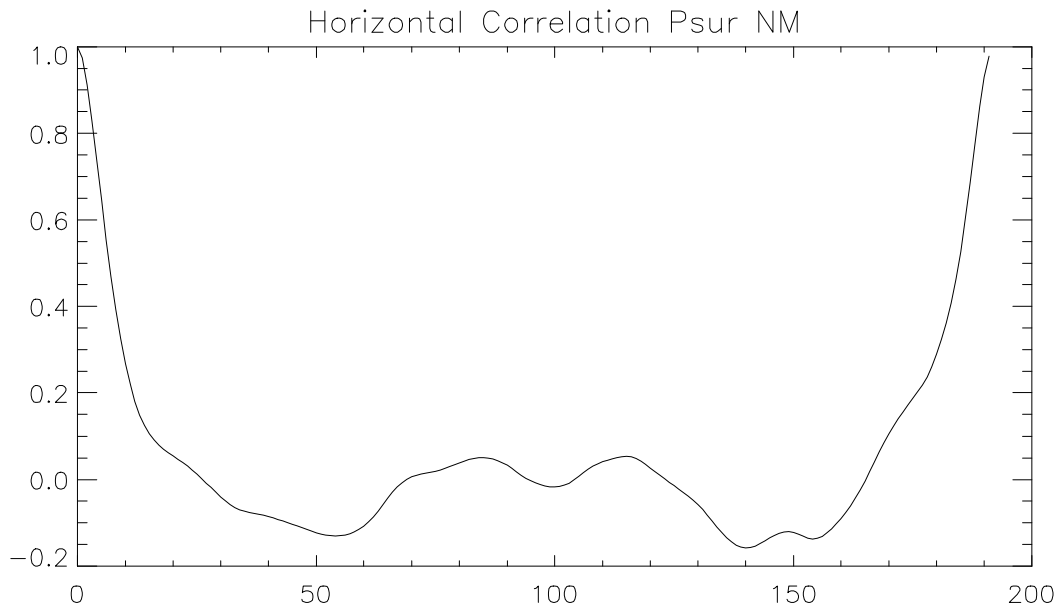
**Figure 19:** Global mean ERA40 surface pressure variance and respective derived horizontal correlation.



### 3.8.17 Horizontal ECMWF Surface Pressure Error Correlations Northern High ( $0^\circ - 30^\circ$ ) and Mid ( $30^\circ - 60^\circ$ ) Latitudes

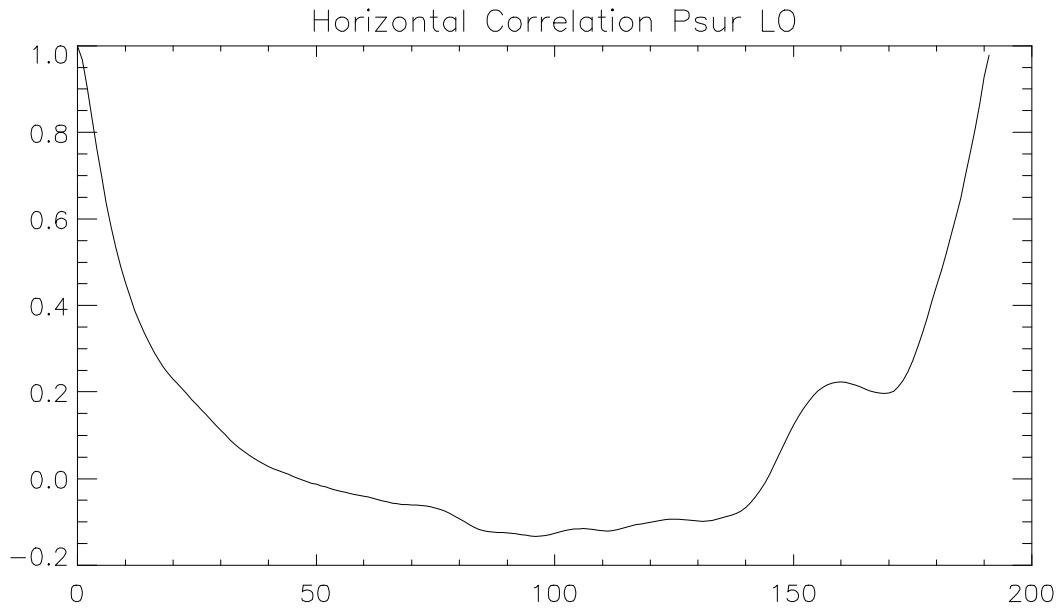


ERA40 correlation of surface pressure NH.

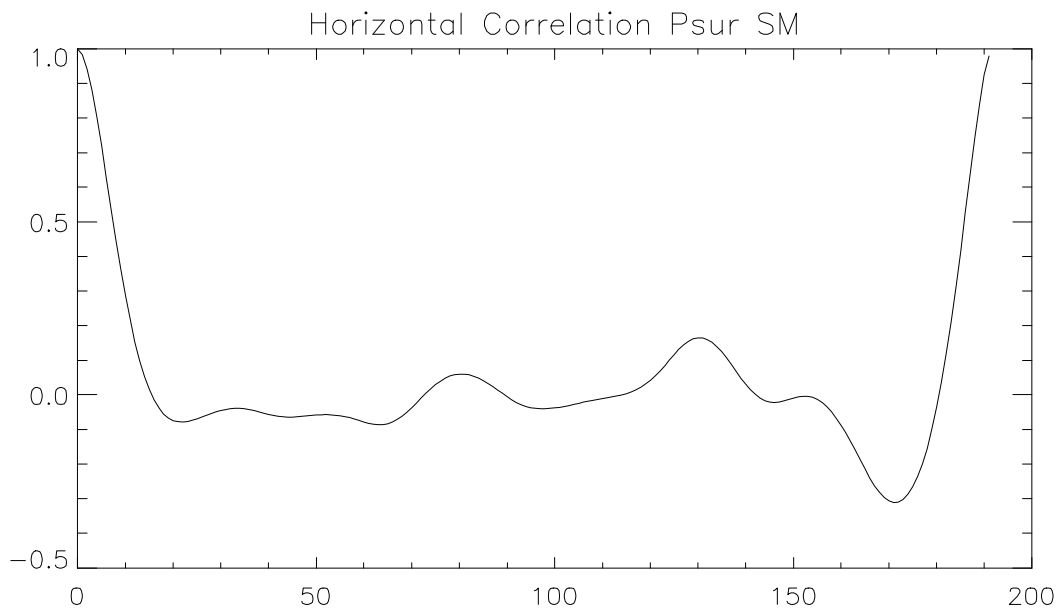


ERA40 correlation of surface pressure NM.

**Figure 20:** Derived horizontal error correlations of ERA40 surface pressure fields northern high and mid latitudes.

**3.8.18 Horizontal ECMWF Surface Pressure Error Correlations Northern Low ( $60^\circ - 120^\circ$ ) and Southern Mid ( $120^\circ - 150^\circ$ ) Latitudes**

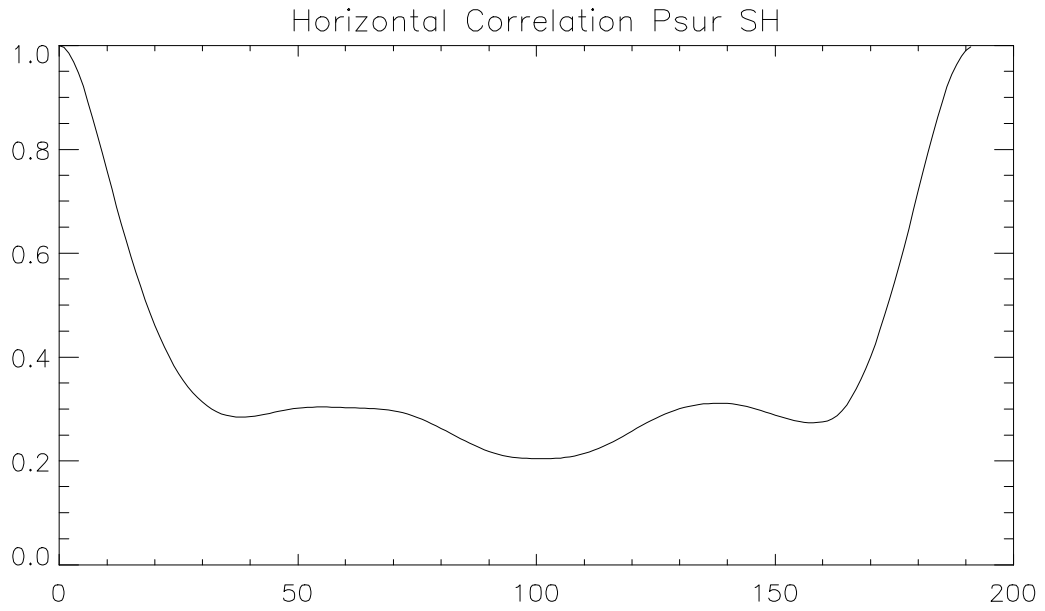
ERA40 correlation of surface pressure LO.



ERA40 correlation of surface pressure SM.

**Figure 21:** Derived horizontal error correlations of ERA40 surface pressure fields low and southern mid latitudes.

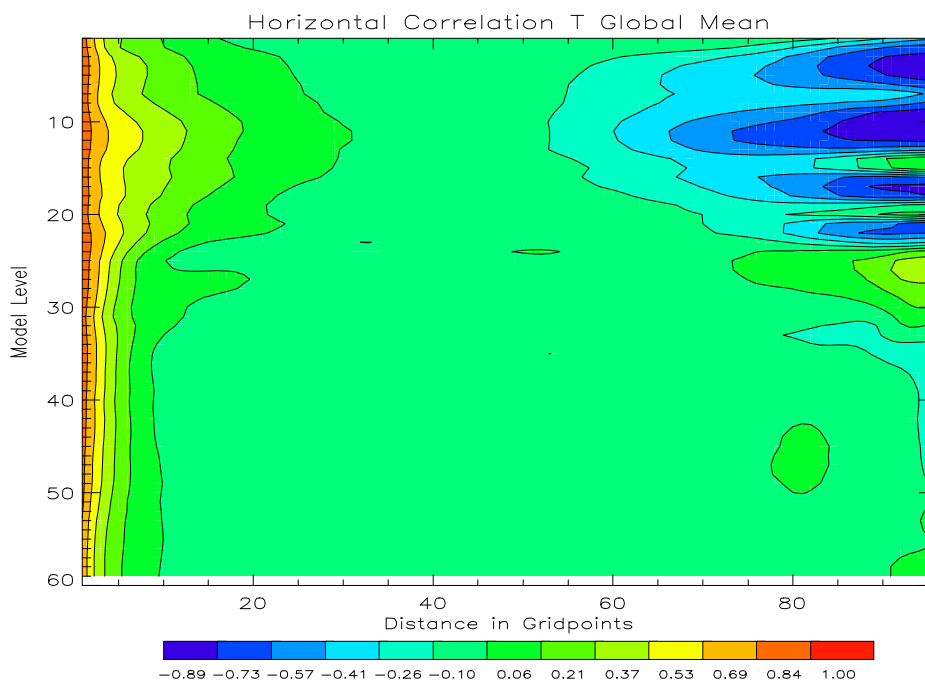
### 3.8.19 Horizontal ECMWF Surface Pressure Error Correlations Southern High (150° - 180°) Latitudes



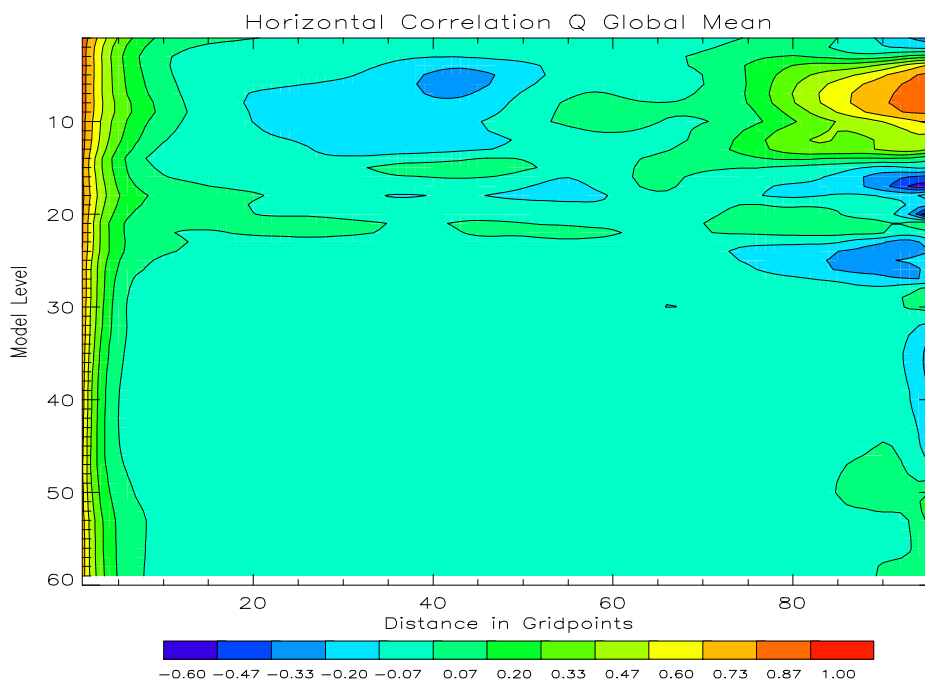
ERA40 correlation of surface pressure SH.

**Figure 22:** Derived horizontal error correlations of ERA40 surface pressure field southern high latitudes.

### 3.8.20 Horizontal ECMWF Temperature and Specific Humidity Error Correlations Global Mean Derived along Bands of Constant Longitude from North to South Pole



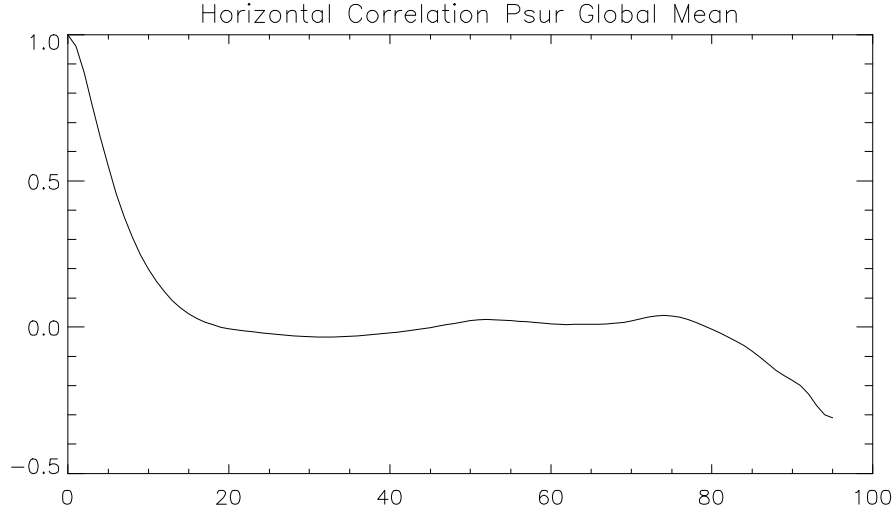
ERA40 horizontal temperature error correlation global mean, L60.



ERA40 horizontal specific humidity error correlation global mean, L60.

**Figure 23:** Derived horizontal error correlations of ERA40 specific humidity and temperature fields from pole to pole global mean.

### 3.8.21 Horizontal ECMWF Surface Pressure Error Correlation Global Mean Derived along Bands of Constant Longitude from North to South Pole



ERA40 error correlation of surface pressure global mean.

**Figure 24:** Derived horizontal error correlation of ERA40 surface pressure fields from pole to pole global mean.

## 3.9 Control Space Transformations

The principles of 3D-Var are briefly described in Sub Section 3.3. For a model state  $\mathbf{x}$  with  $n$  degrees of freedom minimization of the cost function is numerically costly [*F. Bouttier(1999)*] and becomes prohibitively expensive for usual  $n$ 's. One practical solution to this problem is to perform the minimization in a control variable space  $\mathbf{v}$  given by

$$\mathbf{x} = \mathbf{U}\mathbf{v} .$$

The transform  $\mathbf{U}$  has to be chosen in a way that

$$\mathbf{B} = \mathbf{U}\mathbf{U}^T , \tag{62}$$

is approximately satisfied. In the control space  $\mathbf{v}$  the number of required minimization calculations is reduced. Furthermore by using the transform Eq. 62, the background error covariance matrix becomes  $\mathbf{B}_c = \mathbf{I}$ , hence effectively preconditioning the problem.  $\mathbf{I}$  denotes the identity matrix,  $\mathbf{B}_c$  the vertical background covariance matrix. In terms of increments the control variable transform can be written as

$$\delta\mathbf{x} = \mathbf{U}\mathbf{v} .$$

The transformation

$$\mathbf{v} = \mathbf{U}^{-1}\delta\mathbf{x}$$

can be specified in different ways. The definition must provide a way to break down the atmospheric state  $\mathbf{x}$  into uncorrelated but physically realistic error modes which can be *penalized* in  $J_b$  according to their estimated error magnitude.

### 3.9.1 Vertical Control Variable Transform

The vertical transform serves to project control variables from model levels onto the weighted eigenvectors of the vertical component of the background error covariance matrix where the  $\epsilon_v$  represent the background departures from the *true* state of the atmosphere.

$$\mathbf{B}_v = \overline{\epsilon_v \epsilon_v^T} . \quad (63)$$

Eq. 63 considers the vertical transformation of  $v$  at a single horizontal location. For practical reasons approximations must be made like the use of climatological eigenvectors and eigenvalues and the averaging over a geographical domain of these structures. At the moment five different correlation matrices which are latitudinal means are used, and completely spatially dependent standard deviations (cf. 3.8.2).

**General Formulation of  $U_v$**  The vertical component of the covariance matrix  $\mathbf{B}_v$  is given as  $K \times K$  positive-definite symmetric matrix where  $K$  is equal the number of vertical levels. These are properties which allow to perform an eigendecomposition

$$\begin{aligned} \mathbf{B}_v &= \mathbf{P}^{-1} \mathbf{E} \mathbf{\Lambda} \mathbf{E}^T \mathbf{P}^{-1} , \\ \widehat{\mathbf{B}}_v &= \mathbf{E} \widehat{\mathbf{B}}_v \mathbf{E}^T . \end{aligned} \quad (64)$$

The inner product  $\mathbf{P}$  defines a weighted error  $\widehat{\epsilon}_v = \mathbf{P} \epsilon_v$  which might be used to allow for variable model level thickness or introduce synoptic dependencies. In the current version this option is not used. The columns of the matrix  $\mathbf{E}$  are the  $K$  eigenvectors  $\mathbf{e}(m)$  of  $\mathbf{B}_v$  which obey the orthogonality relationship,

$$\mathbf{E} \mathbf{E}^T = \mathbf{I} .$$

$\widehat{\mathbf{B}}_v$  denotes a latitude dependent domain-averaged  $\mathbf{B}_v$ . The diagonal matrix  $\mathbf{\Lambda}$  contains the  $K$  eigenvalues  $\lambda(m)$ . With this standard theory it is possible to define a transform  $\mathbf{U}_v$  between variables  $\delta \mathbf{x}(k)$  on model levels  $k$  and their projection onto vertical modes  $m$  defined by

$$\mathbf{B}_v = \mathbf{U}_v \mathbf{U}_v^T . \quad (65)$$

The comparison of Eq. 64 and Eq. 65 allows to derive

$$\begin{aligned} \delta \mathbf{x} &= \mathbf{U}_v \mathbf{v}_v \\ \delta \mathbf{x} &= \mathbf{P}^{-1} \mathbf{E} \mathbf{\Lambda}^{\frac{1}{2}} \mathbf{v}_v . \end{aligned} \quad (66)$$

If Eq. 66 is inserted into the control variable space form of the background error cost function

$$J_b = \frac{1}{2} \delta \mathbf{x}^T \mathbf{B}_v^{-1} \delta \mathbf{x} ,$$

which gives

$$\begin{aligned} J_b &= \frac{1}{2} \mathbf{v}_v^T \mathbf{v}_v \\ J_b &= \frac{1}{2} \sum_m \mathbf{v}_v(m)^2 , \end{aligned} \quad (67)$$

and for the gradient

$$\nabla J_b = \mathbf{v}_v . \quad (68)$$

As can be seen easily there are several effects of the  $\mathbf{U}_v$  transform:

- The projection onto uncorrelated eigenvectors of  $\mathbf{B}_v$  leads to very significant CPU savings as can be seen via Eq. 67 in the calculation of the background cost function and in its adjoint (gradient) calculations.
- The scaling by the square root of the eigenvalues  $\lambda^{\frac{1}{2}}(m)$  serves as preconditioning.
- The eigenvectors are ordered by the size of their respective eigenvalues what means  $\lambda(1)$  is the dominant structure and  $\lambda(k)$  essentially contains low amplitude noise. This ordering can be used to filter vertical grid scale noise which reduces CPU still further by neglecting small-scale eigenvalue structures, which contribute little to the total error.

**Approximated Eigenstructures** Assuming a single column model, with knowledge of the background covariance matrix and hence the eigenvectors and eigenvalues the  $\mathbf{U}_v$  transform Eq. 66 is an efficient means of reducing CPU without any loss of information. In reality the background covariance matrix is not exactly known, so approximations have to be made. Furthermore our application is 3D-Var where averaging is necessary compared to the 1D-Var case. At the moment five different correlation matrices which are latitudinal means are used (cf. 3.8.2).

### 3.9.2 Horizontal Control Variable Transform

A recursive filter (RF) is used to represent the horizontal component of the background error covariance matrix. The implementation is based on the description of RF's in [Lorenc(1992)].

### 3.9.3 Recursive Filters

**RF Basic Algorithm** The basic algorithm for a recursive filter is quite simple. The RF is presented with an initial function  $A_j$  at grid points  $j$  where  $1 \leq j \leq J$  where the  $\alpha$ 's are the filter coefficients.  $A_j$  is the initial value at gridpoint  $i$   $B_j$  is the value after filtering from  $i = 1$  to  $J$ ,  $C_j$  is the value after one pass of the filter in each direction. A single pass of the RF consists of an initial smoothing from *left to right*

$$B_j = \alpha B_{j-1} + (1 - \alpha)A_j \text{ for } j = 1 \dots J, \quad (69)$$

followed by another pass from *right to left*

$$C_j = \alpha C_{j+1} + (1 - \alpha)B_j \text{ for } j = J \dots 1. \quad (70)$$

The application of the RF in each direction is performed to ensure zero phase change. So a 1-pass filter is defined as a single application of Eq. 69 and Eq. 70. A  $N$ -pass RF is defined by  $N$  sequential applications. In fact the  $A$ 's are the values before and the  $C$ 's the values after application of the filter.

**RF Boundary Conditions** Eq. 69 and Eq. 70 are used to compute recursively the RF response at all points  $j = 2 : J - 1$  interior to the boundary. Explicit boundary conditions are required to specify the response at points  $j = 1$  and  $J$ . If there is a limited area and thus a *real* boundary a method of Hayden & Purser [Hayden and Lorenc(1995)] can be used to specify boundary conditions which assume a given decay-tail outside the domain. This technique assures that the response to observations near the boundary is equivalent to the response within the center of the domain. The boundary conditions for  $B_1$  and  $C_{J+1}$  depend on the particular number of passes  $p$  of the filter in *opposite directions*. Assuming no previous pass of the left moving filter ( $p = 0$ ) we have

$$B_1 = (1 - \alpha)A_1. \quad (71)$$

Following one pass of the filter in the opposite direction the  $p = 1$  boundary condition is

$$(C_J, B_1) = \frac{1 - \alpha}{(1 - \alpha^2)^2} [(B_J, A_1) - \alpha^3(B_{J-1}, A_2)]. \quad (72)$$

Hayden & Purser [Hayden and Lorenc(1995)] suggest to use the  $p = 2$  boundary condition also for  $p > 2$ . In our application there is no *real* boundary but the "boundary conditions" for  $B_1$  and  $C_{J+1}$  still have to be defined

$$B_1 = \alpha A_J + (1 - \alpha)A_1 , \quad (73)$$

and

$$C_J = \alpha B_1 + (1 - \alpha)B_J . \quad (74)$$

**Handling of Boundary Conditions** Since the filter runs from a grid point 1 to a grid point  $n$  and returns it is well suited for areas with defined borders. The boundary conditions can be specified and are a function of the number of filter passes. In our case we operate on a closed surface so information must be transferred from grid point  $n$  to grid point one and vice versa. The solution found doubles the number of filter operations, which can be handled without problems due to the numerical properties of the filter procedure. The approach can be explained with a simple graphical representation. Fig. 25 shows a latitude or longitude band divided into four equal elements.



**Figure 25:** Segments along one latitude or longitude band in original order.

To be able to transfer information from A to D and vice versa the filter procedure is applied a second time to a shifted arrangement of boxes:



**Figure 26:** Shifted arrangement of segments along one latitude or longitude band.

After the second filtering process the original order of the boxes is reconstructed using the two middle segments from both filtering runs, respectively B and C from the original order and D and A from the shifted version. As can be seen, this approach ensure a smooth transition between the boxes A and D.

**Matching of RF Output and Analytical Functions** The smoothing operations performed by the RF algorithm are related to certain analytical functions. In particular, for  $N = 2$ , the RF output approximates a second order autoregressive (SOAR) function

$$\mu_s(r) = \left(1 + \frac{r}{s}\right) e^{-\frac{r}{s}} . \quad (75)$$

In the limit  $N \rightarrow \infty$  it can be shown that the RF output tends to a Gaussian function

$$\mu_g(r) = \exp \left[ -\frac{1}{2} \left( \frac{r}{2s} \right)^2 \right] \quad (76)$$



where  $r$  is the distance at which we want to know the correlation and  $s$  is a characteristic length scale which in fact controls the *width* and is related to the FWHM (Full Width at Halve Maximum). The  $\alpha$  is calculated as

$$\frac{\alpha}{(1 - \alpha)^2} = \frac{1}{2E}, \quad (77)$$

where

$$E = \frac{N(\Delta x)^2}{s^2}. \quad (78)$$

The definition of  $E$  is in this particular case the same for the SOAR and the Gaussian function. This arises from the particular scaling of the Gaussian function given by equation 76.  $\Delta x$  denotes the grid spacing,  $N$  and  $s$  are also known parameters, thus  $E$  can be calculated from Eq. 78.  $\alpha$  can be calculated as follows

$$\alpha = 1 + E - \sqrt{E(E + 2)}. \quad (79)$$

This approach is matching the large-scale response of the RF to that of a SOAR for  $N = 2$  and approaches that of a Gaussian for increasing  $N$ . The matching of the large scale response to analytical SOAR and Gaussian functions serves the definition of  $\alpha$  via Eq. 79. It is also required that the RF conserves the background error variance, for the zero distance case. The calculation of this scaling factor  $S$  is realized as the inverse of the zero distance response of a 1D  $N$  - pass RF to a delta function. A two dimensional  $N$  - pass RF is realized by performing  $N$  applications of multiple 1D RF's in one direction followed by the multiple application of 1D RF's in the orthogonal direction.  $\alpha$  and  $E$  are calculated in the same way as in the 1 dimensional case, however the RF output has to be scaled by  $S^2$  instead of  $S$  which is defined as in the one dimensional case [Lorenz(1992)].

**Transform to Non Dimensional Space** The two dimensional field is transformed to a non dimensional space prior to the filter procedure. This transform is realized as an inner product which is defined as

$$\frac{\text{Increment in Control Space}}{\sqrt{\text{Grid Box Area}}}. \quad (80)$$

The grid box area is calculated by subtracting fractions of ellipsoid areas between the equator and the pole and subsequent division by the number of longitudes. The ellipsoid (WGS84) areas are calculated by

$$\begin{aligned} \text{Ellipsoid Area} = 2\pi (r_{\text{Equator}} + z)^2 \cdot & (0.996647190 \cdot \sin(\text{Lat}) - 0.001116660 \\ & \sin(3.0 \cdot \text{Lat}) + 1.68880838^{-6} \cdot \sin(5.0 \cdot \text{Lat}) - 2.70005436^{-9} \\ & \sin(7.0 \cdot \text{Lat}) + 4.41731436^{-12} \cdot \sin(9.0 \cdot \text{Lat})), \end{aligned} \quad (81)$$

where  $z$  denotes the height over the reference ellipsoid [Laufer(1983)].

**RF Representation of Background Error Covariances** The control variable transform uses the identity

$$\mathbf{B} = \mathbf{U}\mathbf{U}^T, \quad (82)$$

to define a transform  $\delta\mathbf{x} = \mathbf{U}\mathbf{v}$  which relates preconditioned control variables  $\mathbf{v}$  to analysis increments  $\delta\mathbf{x}$  in model space. The horizontal component  $\mathbf{U}_h$  defined by

$$\mathbf{B}_h = \mathbf{U}_h\mathbf{U}_h^T, \quad (83)$$

is realized by scaled recursive filters where  $\mathbf{B}_h$  is the horizontal part of the background covariance matrix  $\mathbf{B}$  (size: (number of latitudes  $\cdot$  number of longitudes)  $\times$  (number of latitudes  $\cdot$  number of longitudes)). The RF has to be applied in a non dimensional space

$$\hat{\mathbf{v}} = \mathbf{F}^{\frac{1}{2}}\mathbf{v},$$

where the scaling factors  $\mathbf{F}$  contains the grid box area as described above. The relation between model and non dimensional space background error covariance matrix  $\widehat{\mathbf{B}}$  is given as

$$\mathbf{B}_h = \mathbf{F}^{-\frac{1}{2}} \widehat{\mathbf{B}}_h \mathbf{F}^{-\frac{1}{2}} . \quad (84)$$

The comparison between Eq. 82 and Eq. 84 indicates that the horizontal component of the control variable transform  $\mathbf{U}_h$  relating model space control variables  $\mathbf{v}$  to model space analysis variables  $\mathbf{x}$  via  $\mathbf{x} = \mathbf{U}_h \mathbf{v}$  can be represented by using a recursive filter  $\widehat{R}$  in non dimensional space as

$$\delta \mathbf{x} = \sigma_b \mathbf{F}^{-\frac{1}{2}} \widehat{R} \mathbf{F}^{\frac{1}{2}} \mathbf{v} . \quad (85)$$

When the two dimensional recursive filter  $\widehat{R}$  is applied, only  $\frac{N}{2}$  passes are performed, as indicated in Eq. 83 the other  $\frac{N}{2}$  passes are performed by the adjoint transform [*Barker(1999)*].

### 3.10 Horizontal Background Error Covariances

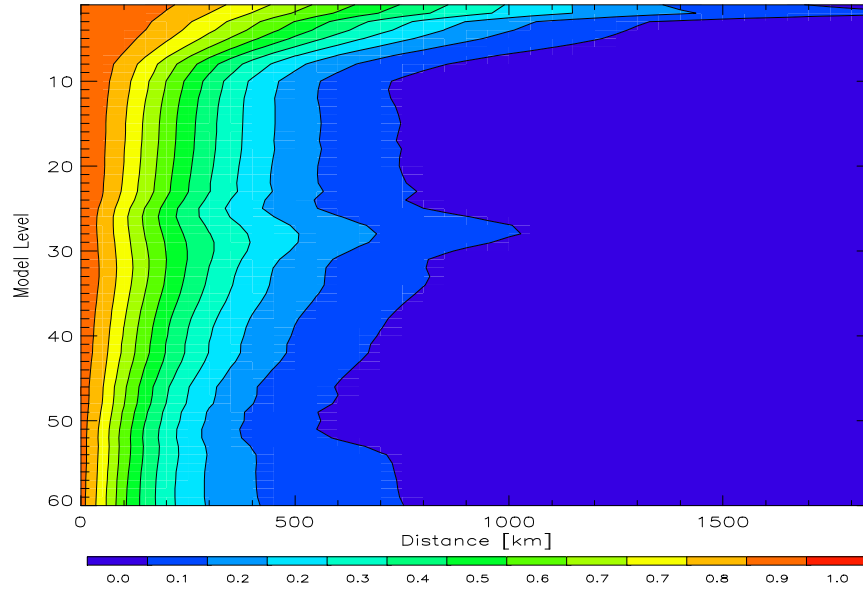
The horizontal correlations are functions of point to point separation which follow roughly the characteristics of the global mean horizontal error correlations in use at ECMWF within the IFC framework in operational use 2003. An alternative which is more suited and possible ways to derive it have been presented in 3.8.3, a description which follows roughly the method which has been used to derive the vertical correlations (cf. 3.8.2).

As recursive filters approximate analytical functions it naturally occurs to be difficult to match them with statistically derived correlation functions. Nevertheless it is possible to archive a quite good agreement with the ECMWF provided horizontal correlations. To take the different grid point distances into account (nearly constant along the latitudes, getting smaller toward the poles along the longitudes), the filter coefficients are calculated separately along the latitude, longitude directions. The following plots show the horizontal correlation characteristics expressed by recursive filters. As can be easily seen the correlation patterns agree qualitatively with the structures presented in 3.8. The fit can be even more perfect, one just has to optimize the characteristic length scale for every single vertical level by hand. The level of accuracy reached here is adequate for these assimilation experiments.

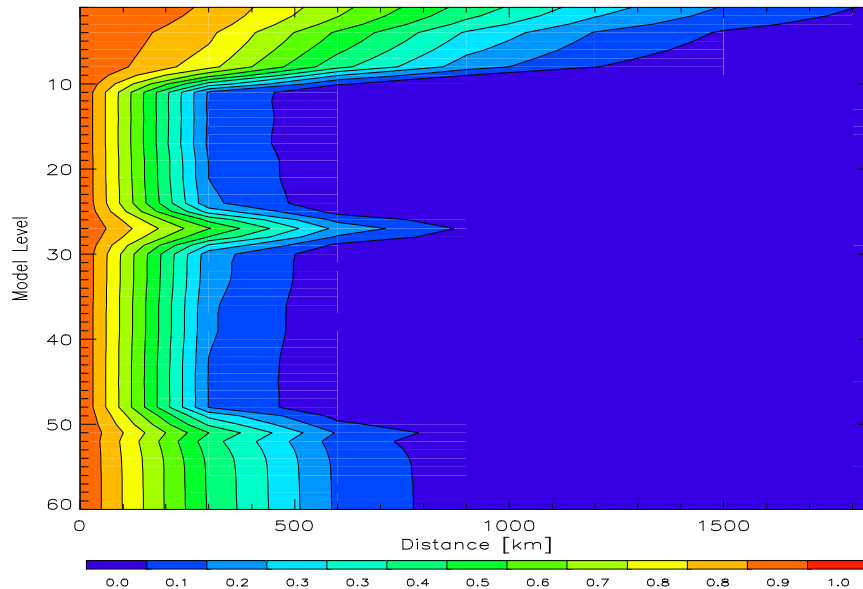
The plots Fig. 27 and Fig. 28 show the ECMWF provided mean correlation patterns for temperature and specific humidity and their by filters approximated counterparts. Fig. 29 depicts in addition the filter approximations using a correlation length of the original pattern times two. Fig. 30 shows the filter approximation of the surface pressure correlation and the corresponding approximation using the original correlation length pattern multiplied by a factor of two. The original ECMWF correlation structure is shown in Fig. 31.

The same procedure could be also applied to horizontal correlation derived as presented in 3.8.3. The filter correlation lengths could be equally tuned to approximate these more appropriate correlation patterns by the use of RF's.

### 3.10.1 Horizontal Global Mean ECMWF Temperature Correlations and its Approximation by a Recursive Filter



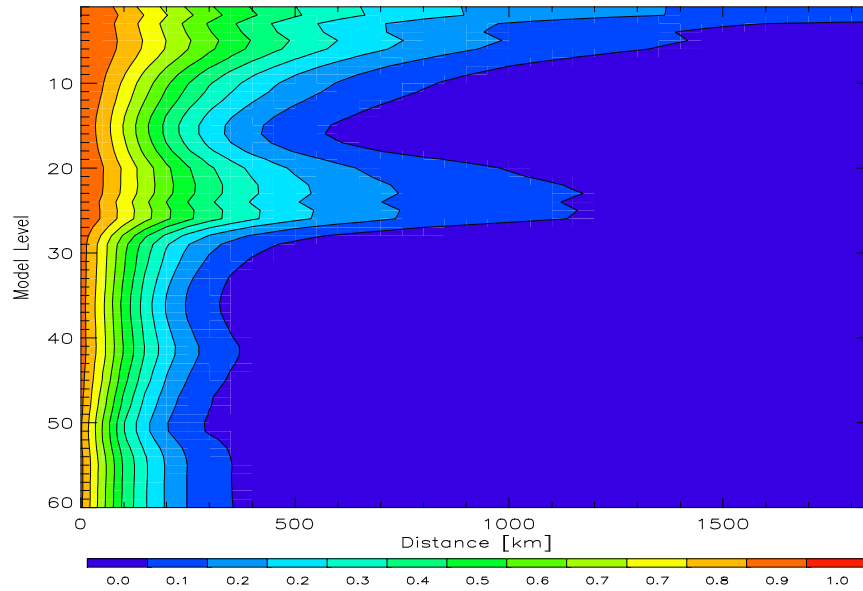
ECMWF horizontal global mean temperature error correlation.



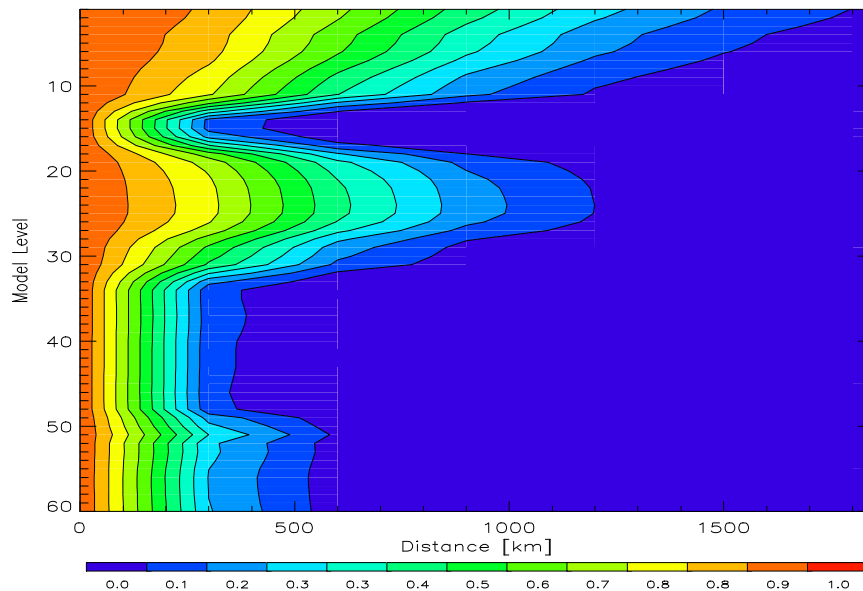
ECMWF horiz. gl. mean temperature corr. approximated by a recursive filter.

**Figure 27:** Global mean horizontal error correlations of ECMWF temperature fields and corresponding filter approximation.

### 3.10.2 Horizontal Global Mean ECMWF Specific Humidity Correlations and its Approximation by a Recursive Filter



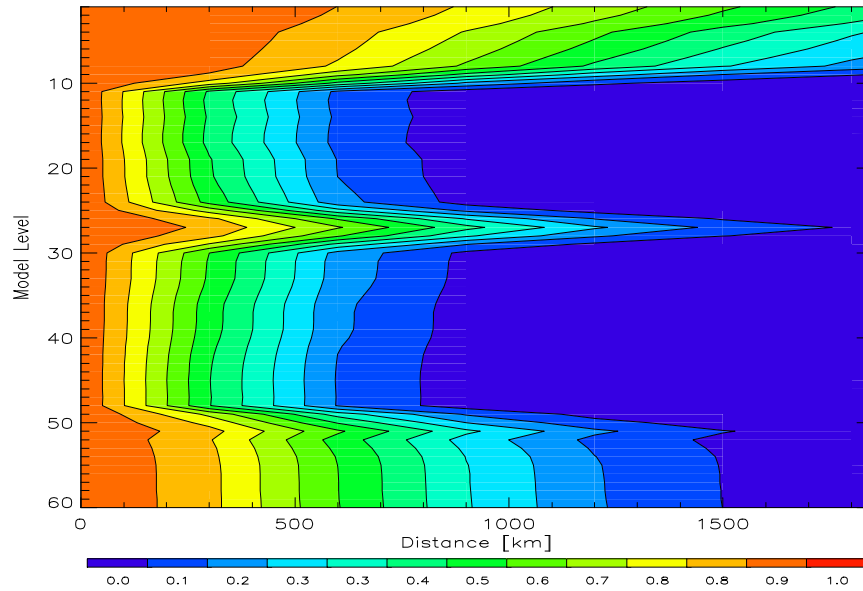
ECMWF horizontal global mean specific humidity error correlation.



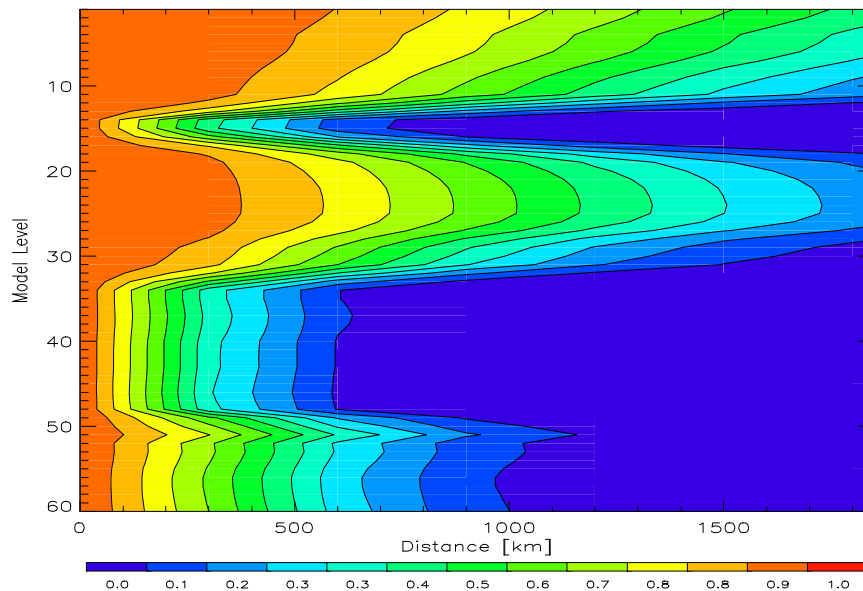
ECMWF horiz. gl. mean spec. humi. corr. approximated by a recursive filter.

**Figure 28:** Global mean horizontal error correlations of ECMWF specific humidity fields and corresponding filter approximation.

### 3.10.3 Horizontal Global Mean ECMWF Temperature and Specific Humidity Correlations Approximated by a Recursive Filter Times Two



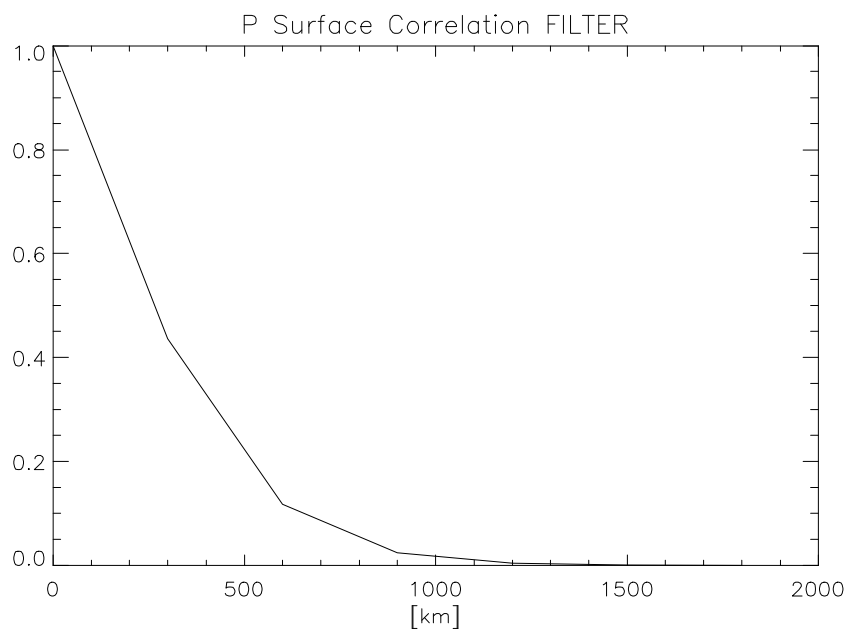
Filter approximation of temperature error correlations 2x.



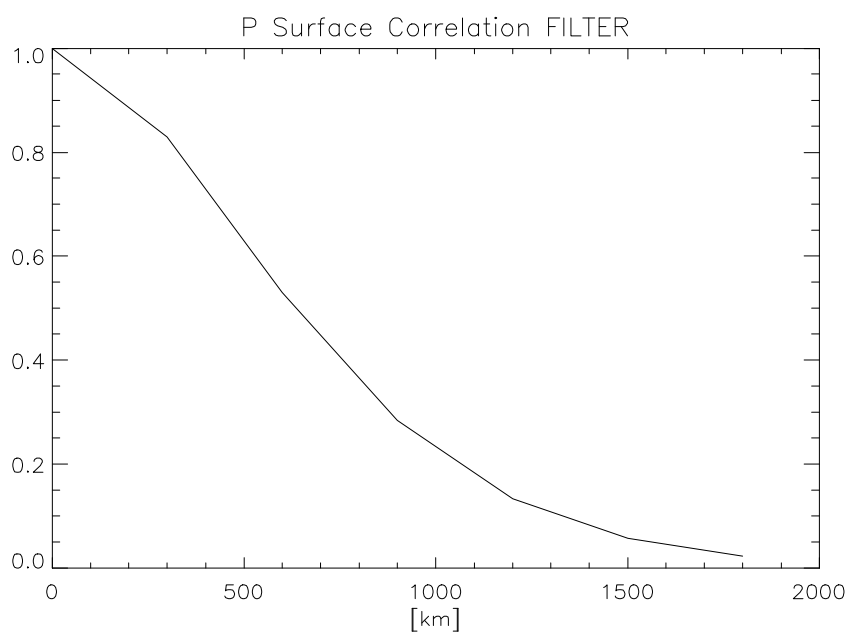
Filter approximation of specific humidity error correlations 2x.

**Figure 29:** Global mean horizontal error correlations of ECMWF temperature and specific humidity fields and corresponding filter approximation times two.

### 3.10.4 Horizontal Global Mean ECMWF Surface Pressure Correlations Approximated by a Recursive Filter and the Respective Approximation Times Two



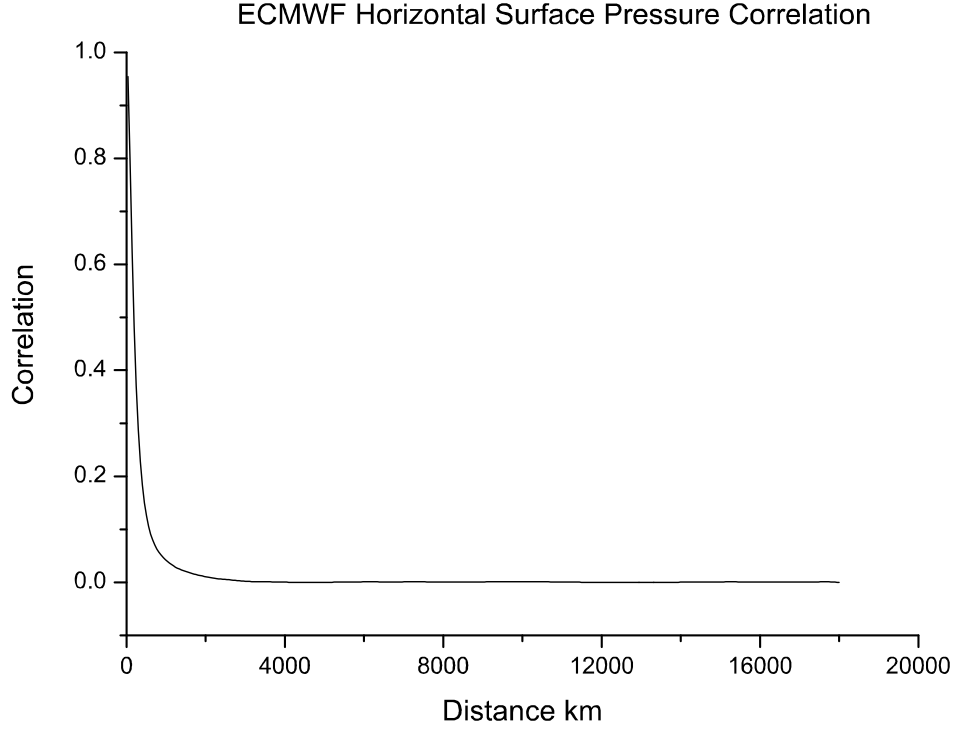
Filter approximation of the global mean surface pressure correlation.



Filter approximation of the global mean surface pressure correlation 2x.

**Figure 30:** Global mean horizontal error correlations of ECMWF surface pressure fields approximated by a recursive filter and the respective approximation times two.

### 3.10.5 Horizontal Global Mean ECMWF Surface Pressure Correlation



Global mean ECMWF horizontal error correlation of surface pressure.

**Figure 31:** Global mean horizontal error correlations of ECMWF surface pressure fields approximated by a recursive filter and the respective approximation times two.

## 3.11 The Minimization Algorithm

The L-BFGS-B algorithm is a limited memory algorithm (L) for solving large nonlinear optimization problems subject to simple bounds on the variables. It is based on the Broyden-Fletcher-Goldfarb-Shanno Method (BFGS), which is from the class of Quasi-Newton methods, the most common. BFGS uses the following basic update for  $A_i$

$$A_{i+1} = A_i + \frac{s_i s_i^T}{s_i^T v_i} + \frac{A_i v_i v_i^T}{v_i^T A_i v_i} + (v_i^T A_i v_i) \cdot u_i u_i^T, \quad (86)$$

with

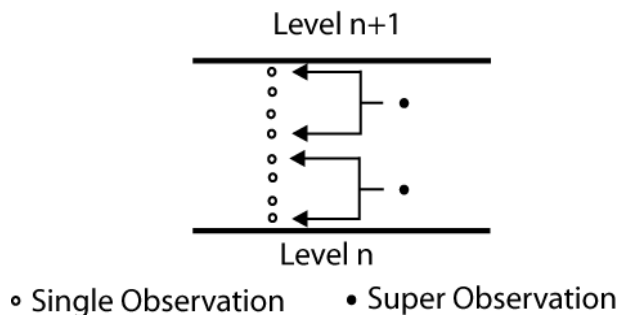
$$u_i = \frac{s_i}{s_i^T} - \frac{A_i v_i}{v_i^T A_i v_i}, \quad (87)$$

where  $s_i = x_{i+1}$  and  $v_i = \nabla f_{i+1} - \nabla f_i$ . For a symmetric positive definite matrix  $A_i$  the matrix  $A_{i+1}$  is also symmetric positive definite, and thus the Quasi-Newton condition is fulfilled.

This version was chosen to be able to apply simple bounds within the assimilation framework (denoted by the B). This option is currently not used. It is intended for problems in which information on the Hessian matrix is difficult to obtain or for large dense problems. L-BFGS-B can also be used for unconstrained problems, as currently in our application, and in this case performs similarly to its predecessor algorithm L-BFGS (Harwell routine VA15). The algorithm is implemented in Fortran 77 [Byrd et al.(1994)Byrd, Peihuang, Nocedal, and Ciyou], [Dong and Nocedal(1989)] the basic input is current function values and the respective gradients.

## 4 Data Preprocessing

Since the RO data products consist of profiles with a vertical resolution (300 - 400 observations within the interesting altitude domain), which exceeds the vertical resolution of the used hybrid level grid (60 vertical levels) by far, a data thinning procedure has to be applied prior to the use of the data. Several studies, also indicate that CHAMP observations have the best quality between 5 and 35 km [Beyerle et al.(2006)Beyerle, Schmidt, Wickert, Heise, Rothacher, and Koenig-Langlo], so only observations within this altitude domain are used. Within the preprocessing step the quality flags of the observations (IGAM processed CHAMP data) are utilized to reject suspicious data.



**Figure 32:** Interpolated and thinned measurement distribution.

During the preprocessing the number of measurements is reduced by averaging, taking the background grid into account. The linear averaging procedure in *LOG* space, takes the spacing of the background levels into account which is derived from the mean global vertical grid [Loeschner(2004)]. Fig. 32 depicts the chosen averaging pattern, giving two super observations between two hybrid levels. The observation error assumptions (see section 3.6) can be seen as very conservative taking the fact of the observation averaging procedure within the pre-preprocessing step into account. This profile thinning procedure leads to a significant reduction in CPU without noticeable loss of information.

## 5 Quality Control

The assimilation system allows for additional quality control steps and observation screening, (e.g. certain criteria to refuse observation too far from the background values, taking the error characteristic from both, background and observation into account), which are not implemented at the moment. As a "quasi" quality control step the observations are screened for cases below the model orography (cf. model orography Fig. 1), which might happen if the profiles are used within their full vertical domain (cf. horizontal interpolation of height grid Eq. 34a).



## 6 Validation

### 6.1 Validation of Adjoint Code

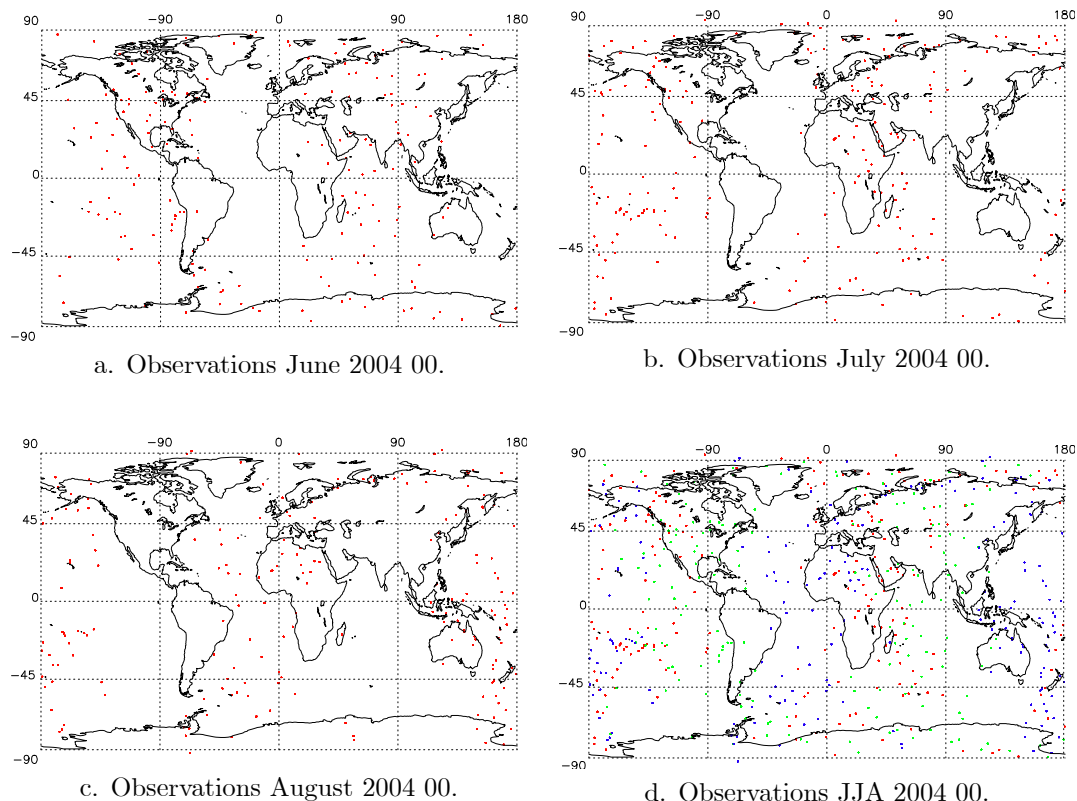
Beside general validation strategies like test-bed setups, adjoint code can be verified with numerical methods,

$$\langle TL(x), TL(x) \rangle = \langle (x), AD(TL(x)) \rangle. \quad (88)$$

The identity expressed in Eq. 88 should hold up to machine accuracy, where  $TL$ , denotes the tangent linear,  $AD$ , the adjoint and  $x$  the original input. These tests have been performed for the observation operator adjoints (interpolation, forward model and filter). This procedure can be applied to single do loops as well as to whole subroutines or even larger sections of the code. It is to mention that this methodology verifies the adjoint code with respect to the forward model. It will not show any bugs in the forward model (the adjoint code can be correct with respect to a erroneous forward model). In our case, single or sets of subroutines as a whole have been tested depending on the setup of the individual operators.

## 7 System Test Runs

The efficiency of an assimilation system can only be verified by real runs, using original data sets. As a system test-bed the summer season 2004 covering the month June, July, August (JJA) was chosen, using the ERA40 derived background fields for the 00 time layer and the respective CHAMP observations. The system has been implemented on ESA's high performance computing Grid on-Demand [*Retscher et al.(2006)Retscher, Goncalves, Brito, and Fusco*], offering a large number of data, computing and storage resources.

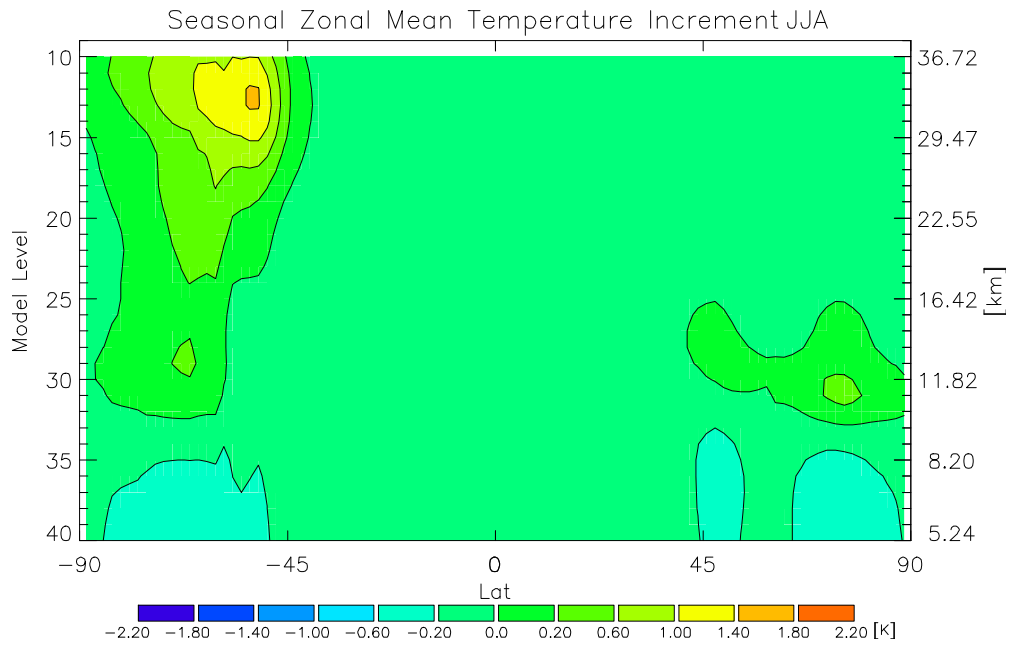


**Table 5:** CHAMP RO observation distribution summer season (JJA) 2004.

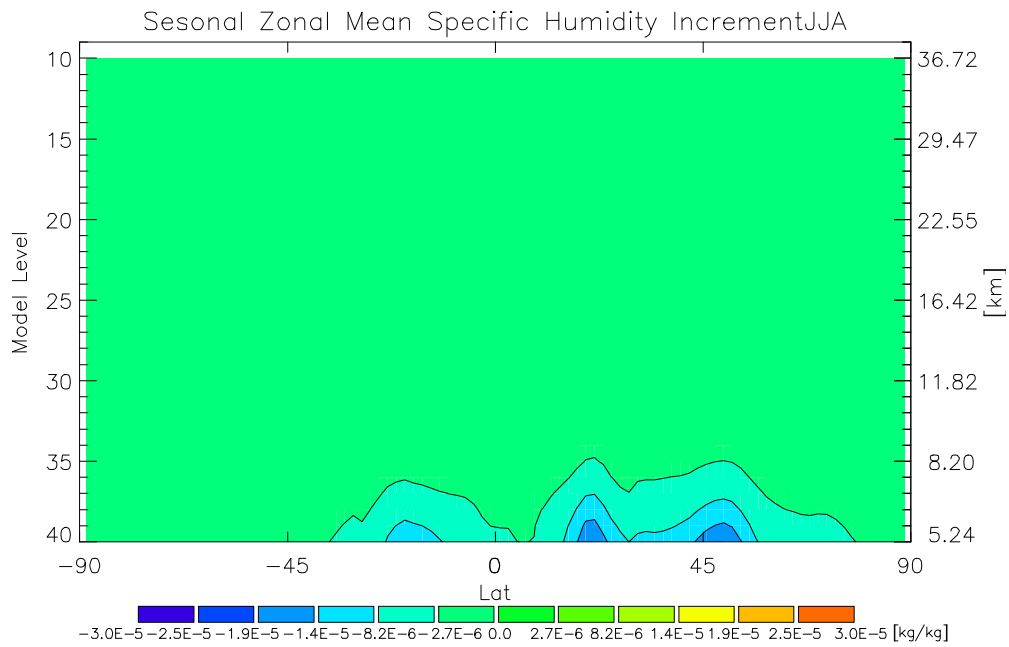
The plots shown in Fig. 5 illustrate the global observation distribution for the 00 time layer in June (a), July (b) and August (c) 2004 with the plot (d) showing the combined observation distribution for the whole JJA 2004 season. The following sample plots are showing the results of test runs comprising 30 simulations within the iterative procedure. The assimilation procedure works on a monthly base; to derive the seasonal results the output of the respective monthly runs is averaged. The presented plots cover the altitude domain predominately influenced directly by the presence of observations. Since the cut-off height of 35 respectively 5 km the increments above and below these limits are caused due to the information spreading within the assimilation system by the background covariance matrices. Model level 20 corresponds approximately to 4 km, model level 50 to 37 km.

The impact of the RO observations on the temperature background is mainly apparent over the southern high latitudes (Fig.33) which is in-line with the error characteristics of the background (Fig.3.8.15). Wave like increment structures appearing predominately over Antarctica and recently over the northern high latitude within the ECMWF temperature analysis fields as in [Gobiet *et al.*(2004a)Gobiet, Foelsche, Steiner, Borsche, Kirchengast, and Wickert] when compared with RO observations are not apparent. This model behavior might be less distinct in the ERA40 version but more likely the use of monthly means and the averaging over 21 years removes these strange patterns from the temperature background fields. Another explanation might be the convergence behavior which is addressed in 7.1.

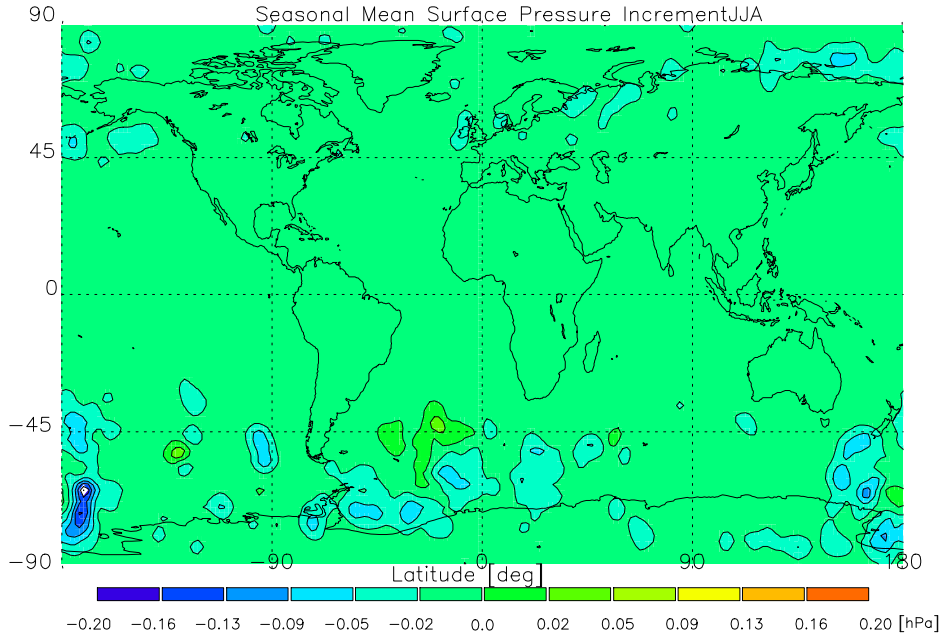
The increment of the specific humidity analysis is limited to the lower most levels as can be seen



**Figure 33:** Zonal mean temperature increment of the 00 time layer of the JJA 2004 season.



**Figure 34:** Zonal mean specific humidity increment of the 00 time layer of the JJA 2004 season.



**Figure 35:** Zonal mean surface pressure increment of the 00 time layer of the JJA 2004 season.

in Fig. 34, a result which is expected, since the atmosphere is relatively dry above the troposphere and the use of RO observations is limited to altitudes above 5 km in this study.

The surface pressure increment plots illustrate the horizontal spread of information due to the filter procedure, distinct increments appear predominately at southern high latitudes. The observation distribution (Fig. a, b, c, and d table 5) shows an increased amount of observation at high latitudes, a fact that can be contributed to the orbital characteristics of the CHAMP satellite. The more distinct impact at the southern high latitudes can be explained by the orography of the antarctic plateau. Since here the distance between the last observation ( $\sim 5$  [km]) and the surface ( $\sim 3$  [km]) is drastically reduced compared to mean sea level, resulting in a more pronounced impact of pressure information propagated down to the surface by the operators.

## 7.1 Convergence Behavior

The test runs had been performed using a hard limit of 30 function and gradient evaluations (simulations) on ESA's high performance Grid on Demand which resulted in 1 to 3 new iterates indicating a bad convergence behavior (a simulation denotes the attempt to find a new minima whereas an iterate denotes a *successful* simulation thus a new minima was found). Other experiments using analysis or forecast fields within six hour assimilation windows with a very similar assimilation system produce  $\sim 20$  new iterates using the same 30 function and gradient evaluations limit [Loescher and Kirchengast(2008)]. The difference is easy to explain since within an observation set covering a six hour time window spatially close observations will naturally have similar values. In contrast observations within a six hour time window covering a whole month might be quite different even if their separation in space is small since their separation in time might be significant (and in reality is). Since the observations are quite accurate which can be seen in the characterization of the observation error covariance matrix (cf. 3.6) the minimization algorithm tries to fit the background to observations which are spatially close but differ significantly in value (fitting one observation increases the costfunction value of the other and vice versa). Due to this inherent problem the minimization procedure becomes numerical ill posed.

## Conclusions

As shown in the last section, the first results are mixed but still promising. Due to the observation distribution I would advise to stick to the chosen approach and analyze monthly and even better seasonal means. The approach to use six hour assimilation time windows around the synoptic hours 00, 06, 12, and 18 is appropriate to take the local time issues into account. To derive total monthly means the resulting fields might be averaged separately. With the COSMIC constellation up and in operation it will be possible to address the questions concerning local time after the satellites reached their final orbits (expected in late 2007), since the distribution of observations in space and time will be most likely sufficient.

The use of variational techniques is not widespread within the climate community, mostly due to the principle of using data as *unaffected* and *uncontaminated* as possible. Fact is that most remote sensing data, since the measurements are in general indirect, not fulfil these requirements anyway. Although RO data is closer to that *ideal observation*, than most other observations, a variational approach is promising, since the *bias free* requirement (paramount to variational methods, and also important for climate applications) might be nearly fulfilled for RO data in the near future. If one is interested in trend monitoring only relative changes are of interest, thus a bias which is static would not pose a major problem. Nevertheless it is difficult to assess if a possible bias is static over time if the reason is not entirely clear. Advanced retrieval techniques like wave optics are able to cope with the problems occurring in the lower troposphere and hence are superior to geometric optics below about 7 km [Gorbunov(2002)], [Hoche et al.(1999)Hoche, Pavelyev, Yakovlev, Barthes, and Jakowski], [Sokolovskiy(2003)], [Jensen et al.(2003)Jensen, Benzou, and Nielsen], [Beyerle et al.(2003)Beyerle, Wickert, Schmidt, and Reigber]. One could also argue to use RO data predominately in the stratosphere, an altitude domain where temperature data derived from RO profiles already exhibit a high quality. Future RO data products will take advantage of the improved retrieval techniques and thus deliver high quality observations down into the lower troposphere at least to the top of the planetary boundary layer. A next step should be the implementation of a bending angle operator, following another principle of data assimilation, using observations as raw and unprocessed as possible [Ringer and Healy(2008)], [von Engeln(2006)]. This approach mitigates introduction of errors due to processing and auxiliary data (*insect problem*). It allows for a simpler observation error characterization, avoiding correlations caused by the data processing steps itself. Furthermore the concept can be expanded to other observations in future to get multi instrument analyses. Using multiple instruments would improve the analysis in a very elegant way not only by sheer numbers of observations. If combined in a clever way one type of observation may balance weaknesses of another type and vice versa (for example a combination of RO data comprising an excellent vertical resolution with radiometer data from a nadir sounder exhibiting high spatial resolution). One might also think about joint retrievals if the number of coincident events is high enough, to get a data product of improved quality.

The derivation of first guess fields from ERA40 data results in a data set being completely independent from model changes. This methodology offers the unique opportunity to derive climatological fields to complement the traditional climate products, which are in general based on interpolation techniques. Assimilation technique can combine different sources of information in a optimal and consistent way, a clear advantage to interpolation techniques which show their weakness in merging different sources of information.

Nevertheless the characterization of the background errors poses a difficult problem, especially for monthly mean fields. This is an exercise which still has to be studied more in depth. The way chosen here to derive the vertical correlations is valid and can be seen as a good approximation of the truth which is in fact unknown. The derivation of horizontal correlations still has to be conducted thoroughly; a possible way comprising examples has been shown here.

Concerning the observation error assumption the used approach has to be seen as a conservative one, since the data is averaged (cf. 4) within the preprocessing step. This data thinning procedure which results in a kind of *super observation* profile should reduce the overall standard deviation

a bit, even if the data is a little bit blurred over the vertical domain. This is taken to a certain extent into account by using the background grid to determine the vertical averaging interval. Since the refractivity shows an exponential decrease with altitude, this is accounted for by the dense background grid spacing within the troposphere which smoothly decreases throughout the tropopause into the stratosphere, thus the averaging intervals are growing with altitude suitable for the characteristics of refractivity observations. Furthermore this preprocessing step ensures an even vertical distribution of the observations within the background grid and a smoothing (the RO profiles capture finer atmospheric structures than the model grid is capable to represent).

The convergence issue has been already addressed and will not be easy to solve. It stems mainly from an inconsistency between profiles which are close in distance but separated in time. In that case the errors are not agreeing and the system tries to fit values which can't be fit. The background errors are significantly larger than the observation errors so the weight of the observations is high and the algorithm can not solve the discrepancy between observations close in space but quite different in value. One way maybe to relax the observation error structures to allow for more variability, which would mean in fact to assume more variance and thus using intentionally suboptimal errors. That approach has not been tested and the way to relax the errors (basically the *shape* of the error structure could stay the same but the magnitude has to be changed, which can be easily done in the used formulation) has to be justified. Maybe a relaxation reassembling the underlying variability might be a possible approach. Using shorter timeframes and thus reducing the separation in time of observations is possible and would work [Loescher and Kirchengast(2008)] but the computing time would increase significantly (extreme case instead of 4 runs per month up to 44 runs per month (4 per day)) although the number of observations within every single assimilation run would decrease resulting in less cpu time.

## A Notation

$\mathbf{x}_t$	: True State of the Atmosphere	Dimension $n$
$\mathbf{x}_b$	: Background Model State	Dimension $n$
$\mathbf{x}_a$	: Analysis Model State	Dimension $n$
$\mathbf{y}$	: Observation Vector	Dimension $p$
$H$	: Observation Operator	Dimension $n \rightarrow p$
$\mathbf{H}$	: Linear Observation Operator	Dimension $n \rightarrow p$
$\mathbf{B}$	: Background Covariance Matrix	Dimension $n \times n$
$\mathbf{R}$	: Observation Covariance Matrix	Dimension $p \times p$
$\mathbf{A}$	: Analysis Covariance Matrix	Dimension $n \times n$
$\mathbf{K}$	: Gain Matrix	Dimension $n \times n$
$\mathbf{I}$	: Identity Matrix	
$J_b$	: Background Cost Function	
$J_o$	: Observation Cost Function	
$J$	: Total Cost Function	
$\mathbf{v}$	: Background State Vector in Control Space	Dimension $n$
$\mathbf{E}$	: Matrix Containing Eigenvectors (Columns)	Dimension $n \times n$
$U$	: Control Space Transformation Operator	
$\lambda$	: Eigenvalues	
$P$	: Inner Product	
$z$	: Geometric Height	
$\phi$	: Geopotential Height	
$\varphi$	: Latitude	
$\lambda$	: Longitude	
$\sigma$	: Standard Deviation	
$\alpha$	: Filter Coefficients	
$\mathbf{F}$	: Scaling Factors	

## B Constants

$k_1$	=	77.60	[K/hPa]	Empirical Constant Thayer Formula & Smith Weintraub
$k_2$	=	70.40	[K/hPa]	Empirical Constant Thayer Formula & Smith Weintraub
$k_3$	=	373900.00	[K <sup>2</sup> /hPa]	Empirical Constant Thayer Formula & Smith Weintraub
$A$	=	$6.02214 \times 10^{23}$	[mol <sup>-1</sup> ]	Avogadro Number
$R$	=	8.3145	[Pa·m <sup>3</sup> /K·mol]	Universal Gas Constant
$R_{Dry}$	=	287.06	[J/K·Kg]	Dry Air Gas Constant
$R_{WatVap}$	=	461.52	[J/K·Kg]	Water Vapor Gas Constant
$m_A$	=	28.964	[kg/kmol]	Molar Mass of Dry Air
$m_W$	=	18.015	[kg/kmol]	Molar Mass of Water Vapor
$g_{Mean}$	=	9.80665	[m·s <sup>-1</sup> ]	Mean Acceleration of Gravity
$r_{Mean}$	=	6371.0	[Km]	Mean Radius of Earth
$r_{Pol}$	=	6356.752314	[Km]	Polar Radius of Earth
$r_{Equ}$	=	6378.137	[Km]	Equatorial Radius of Earth
$M^*$	=	$6.022140 \times 10^{26}$	[Kmol <sup>-1</sup> ]	Kilo Mol
$Earth\_Flattening$	=	$\frac{r_{Equ}}{r_{Pol}} - r_{Equ}$	[km]	Earth Flattening
$J_2$	=	$1.08263 \times 10^{-3}$		GRS-80 zonal coefficient
$g_{Equ}$	=	9.7803	[m·s <sup>-1</sup> ]	Acceleration of Gravity at Equator



## C List of Acronyms

<b>ECMWF</b>	:	European Center for Medium Range Weather Forecast
<b>CHAMP</b>	:	Challenging Mini Satellite Payload
<b>RO</b>	:	Radio Occultation
<b>LEO</b>	:	Low Earth Orbit Satellite
<b>MEO</b>	:	Medium Earth Orbit Satellite
<b>GEO</b>	:	Geo Stationary Orbit Satellite
<b>TEC</b>	:	Total Electron Content
<b>GFZ</b>	:	Geo Forschungszentrum Potsdam
<b>3D-VAR</b>	:	Three Dimensional Variational Data Assimilation
<b>4D-VAR</b>	:	Four Dimensional Variational Data Assimilation
<b>pdf</b>	:	Probability Density Function
<b>BLUE</b>	:	Best Linear Unbiased Estimator
<b>TL</b>	:	Tangent Linear
<b>AD</b>	:	Adjoint
<b>IFS</b>	:	ECMWF Integrated Forecast System
<b>TLE</b>	:	Two Line Element
<b>COSMIC</b>	:	Constellation Observing System for Meteorology, Ionosphere & Climate
<b>CIRA</b>	:	Cospar International Reference Atmosphere
<b>COSPAR</b>	:	Committee on Space Research
<b>MSISE</b>	:	Mass Spectrometry Incoherent Scatter (Extended)
<b>RAOB</b>	:	Radio (Balloon) Observations
<b>MIPAS</b>	:	Michelson Interferometer for Passive Atmospheric Sounding
<b>PCA</b>	:	Principal Component Analysis
<b>ICA</b>	:	Independent Component Analysis
<b>GFZ</b>	:	Geo Forschungs Zentrum Potsdam
<b>DMI</b>	:	Danish Meteorological Institute
<b>ENVISAT</b>	:	Environment Satellite
<b>ERA40</b>	:	ECMWF Re-Analysis 40
<b>GRAS</b>	:	GNSS Receiver for Atmospheric Sounding
<b>GNSS</b>	:	Global Navigation Satellite System
<b>WMO</b>	:	World Meteorological Organization
<b>WCRP</b>	:	World Climate Research Program
<b>ECHAM5</b>	:	European Center Hamburg Model 5
<b>METOP</b>	:	Meteorological Operational Satellite
<b>EPS</b>	:	Eumetsat Polar System
<b>SAF</b>	:	Satellite Application Facility
<b>WRF</b>	:	Weather Research and Forecasting Model
<b>MM5</b>	:	Mesoscale Model 5
<b>GENESIS</b>	:	GPS Environmental & Earth Science Information System
<b>IPCC</b>	:	Intergovernmental Panel on Climate Change
<b>GCM</b>	:	Global Circulation Model
<b>NWP</b>	:	Numerical Weather Prediction
<b>OI</b>	:	Optimal Interpolation
<b>RF</b>	:	Recursive Filter
<b>INRIA</b>	:	Institut National de Recherche en Informatique et en Automatique

<b>r.m.s.</b>	:	Root Mean Square
<b>NASA</b>	:	National Aeronautics and Space Administration
<b>NOAA</b>	:	National Ocean and Atmospheric Administration
<b>UCAR</b>	:	University Cooperation for Atmospheric Research
<b>CDAAC</b>	:	COSMIC Data Analysis and Archive Center
<b>GPS</b>	:	Global Positioning System
<b>SOAR</b>	:	Second Order Autoregressive Function
<b>FGAT</b>	:	First Guess at Appropriate Time
<b>ESA</b>	:	European Space Agency
<b>NCEP</b>	:	National Climate and Environmental Prediction
<b>BFGS</b>	:	Broyden-Fletcher-Goldfarb-Shanno Method
<b>NPOESS</b>	:	National Polar Orbiting Environmental Satellite System
<b>GPSOS</b>	:	GPS Occultation Sensor

## References

- [*Anthes et al.(2000)*Anthes, Rocken, and Kuo] Anthes, R., C. Rocken, and Y. Kuo, Applications of cosmic to meteorology and climate, *TAO*, 1, 115–156, 2000.
- [*Barker(1999)*] Barker, D., A general formulation for 3dvar control variables, 1999.
- [*Barker et al.(2004)*Barker, Huang, Guo, Bourgeois, and Xiao] Barker, D. M., W. Huang, Y.-R. Guo, A. J. Bourgeois, and Q. N. Xiao, A three-dimensional variational data assimilation system for mm5: impementation and initial results, *Monthly Weather Review*, 132, 897–914, 2004.
- [*Beyerle et al.(2003)*Beyerle, Wickert, Schmidt, and Reigber] Beyerle, G., J. Wickert, T. Schmidt, and C. Reigber, Atmospheric sounding by gnss radio occultation: An analysis of negative refractivity bias using champ observations, *J. Geophys. Res.*, 2003.
- [*Beyerle et al.(2006)*Beyerle, Schmidt, Wickert, Heise, Rothacher, and Koenig-Langlo] Beyerle, G., T. Schmidt, J. Wickert, S. Heise, M. Rothacher, and G. Koenig-Langlo, Observations and simulations of receiver-induced refractivity biases in gps radio occultation, *JGR*, 111, 2006.
- [*Byrd et al.(1994)*Byrd, Peihuang, Nocedal, and Ciyou] Byrd, R. H., L. Peihuang, J. Nocedal, and Z. Ciyou, A limited memory algorithm for bound constrained optimization, *Tech. rep.*, North-western University, 1994.
- [*Dong and Nocedal(1989)*] Dong, L. C., and J. Nocedal, On the limited memory bfgs methode for large scale optimisation, *Mathematical Programming*, 5, 503–528, 1989.
- [*ECMWF(2003)*] ECMWF, *MARS User Guide*, 2003.
- [*EUMETSAT(2003)*] EUMETSAT (Ed.), *2nd GRAS SAF User Workshop, Workshop Proceedings*, vol. EUM P 40, EUMETSAT, 2003.
- [*F. Bouttier(1999)*] F. Bouttier, P. C., *Data Assimilation Concepts and Methodes, Meteorological Training Course Lecture Series*, 1999.
- [*Fjeldbo et al.(1971)*Fjeldbo, Eshleman, and Kliore] Fjeldbo, G. F., V. R. Eshleman, and A. J. Kliore, The neutral atmosphere of venus as studied with the mariner v radio occultation experiments, *Astron. J.*, 76, 123–140, 1971.
- [*Gobiet and Kirchengast(2004)*] Gobiet, A., and G. Kirchengast, Advancements of global navigation satellite system radion occultation retrieval in the upper stratosphere for optimal climate monitoring utility, *J. Geophys. Res.*, 109, 2004.
- [*Gobiet et al.(2004a)*Gobiet, Foelsche, Steiner, Borsche, Kirchengast, and Wickert] Gobiet, A., U. Foelsche, A. K. Steiner, M. Borsche, G. Kirchengast, and J. Wickert, Climatological validation of stratospheric temperatures in ecmwf operational analysis with champ radio occultation data, *J. Geophys. Res.*, 32, 2004a.
- [*Gobiet et al.(2004b)*Gobiet, Steiner, Retscher, Foelsche, and Kirchengast] Gobiet, A., A. K. Steiner, C. Retscher, U. Foelsche, and G. Kirchengast, Algorithms validation based on champ/gps data, *Tech. rep.*, Inst. for Geophys., Asrophys., and Meteorol., Univ. of Graz, Austria, 2004b.
- [*Gorbunov(2002)*] Gorbunov, M. E., Canonical transform methode for processing radio occultation data in the lower troposphere, *Radio Sci.*, 37(5), 1057, 2002.
- [*Gorbunov and Gurvich(1998)*] Gorbunov, M. E., and A. S. Gurvich, Microlab-1 experiment: Multipath effects in the lower troposphere, *J. Geophys. Res.*, 103, D12, 13,819–13,826, 1998.

- [*Gorbunov and Kornblueh(2003)*] Gorbunov, M. E., and L. Kornblueh, Principles of variational assimilation of gnss radio occultation data, *Tech. Rep. Report No.350*, Max-Planck-Institut für Meteorologie, Hamburg, 2003.
- [*Gorbunov and Sokolovskiy(1993)*] Gorbunov, M. E., and S. V. Sokolovskiy, Remote sensing of refractivity from space for global observations of atmospheric parameters, *Tech. rep.*, Max Planck Institute for Meteorology, Hamburg, 1993.
- [*Gorbunov et al.(1996)Gorbunov, Sokolovskiy, and Bengtsson*] Gorbunov, M. E., S. Sokolovskiy, and L. Bengtsson, Space refractivity tomography of the atmosphere: Modeling of direct and invers problems, *Tech. rep.*, Max Planck Institute for Meteorology, Hamburg, 1996.
- [*Hayden and Lorenc(1995)*] Hayden, C. M., and A. C. Lorenc, Recursive filter for objective analysis of meteorological fields. applications to nesdis operational processing, *J. Appl. Meteor.*, *34*, 3–15, 1995.
- [*Healy and Eyre(2000)*] Healy, S. B., and J. Eyre, Retrieving temperature, water vapor and surface pressure information from refractive index profiles derived by radio occultations: a simulation study, *Quart. J. Roy. Meteorol. Soc.*, pp. 1661–1683, 2000.
- [*Hedin(1991)*] Hedin, A. E., Extension of the msis thermosphere model into the middle and lower atmosphere, *J. Geophys. Res.*, *96*, 1159–1172, 1991.
- [*Hocke(1997)*] Hocke, K., Inversion of gps meteorology data, *Annales. Geophysicae*, *15*, 443–450, 1997.
- [*Hocke et al.(1999)Hocke, Pavelyev, Yakovlev, Barthes, and Jakowski*] Hocke, K., A. G. Pavelyev, O. I. Yakovlev, L. Barthes, and N. Jakowski, Radio occultation data analysis by the radiographic methode, *J. Atmos. Terr. Phys.*, *61*, 1169–1177, 1999.
- [*INRIA(2002)*] INRIA, Software tapenade inria 2002, version 2.0, *Tech. rep.*, Domaine de Voluceau, Rocquencourt - BP 105, 78153 Le Chesnay Cedex, FRANCE, 2002.
- [*Jakowski et al.(2004)Jakowski, Heise, Wehrenpfennig, and Tsybulya*] Jakowski, N., S. Heise, A. Wehrenpfennig, and K. Tsybulya, *Ionospheric Radio Occultation Measurements and Space Weather*, pp. 383–392, in [*Kirchengast et al.(2004)Kirchengast, Foelsche, and Steiner*], 2004.
- [*Jensen et al.(2003)Jensen, Benzon, and Nielsen*] Jensen, A. S., H.-H. Benzon, and A. Nielsen, Full spectrum inversion of radio occultation signals, *Radio Sci.*, *18(3)*, 1159, 2003.
- [*Kellberg et al.(2004)Kellberg, Simmons, Uppala, and Fuentes*] Kellberg, Simmons, Uppala, and Fuentes, The era40 archive, *Tech. rep.*, Tech. Rep. ECMWF, 2004.
- [*Kirchengast et al.(2004)Kirchengast, Foelsche, and Steiner*] Kirchengast, G., U. Foelsche, and A. Steiner (Eds.), *Occultations for Probing Atmosphere and Climate*, Springer, 2004.
- [*Kirchengast et al.(2006)Kirchengast, Foelsche, and Steiner*] Kirchengast, G., U. Foelsche, and A. Steiner (Eds.), *Atmosphere and Climate, Studies by Occultation Methodes*, Springer, 2006.
- [*Kursinski et al.(1997)Kursinski, Hajj, Schofield, K., and Hardy*] Kursinski, R. E., G. A. Hajj, J. T. Schofield, R. P. L. K., and R. Hardy, Observing eart’s atmosphere with radio occultation measurements using the global positioning system, *J. Geophys. Res.*, *120*, 23,429–23,465, 1997.
- [*Kursinski et al.(1996)*] Kursinski, R. E., et al., Initial results of radio occultation observations of earth’s atmosphere using the global positioning system, *Science*, *271*, 1107–1110, 1996.
- [*Lauf(1983)*] Lauf, G. B., Geodesy and map projections, *TAFE Publications Unit*, 1983.

- [Loescher(2004)] Loescher, A., Assimilation of gnss radio occultation data into gcm fields for global climate analysis), *Tech. rep.*, Inst. for Geophys., Asrophys., and Meteorol., Univ. of Graz, Austria, 2004.
- [Loescher and Kirchengast(2008)] Loescher, A., and G. Kirchengast, Variational data analysis for deriving global climate analyses from gnss radio occultation data, *GPS Solutionst*, 2008.
- [Loescher et al.(2008)Loescher, Retscher, Fusco, Goncalves, Brito, and Kirchengast] Loescher, A., C. Retscher, L. Fusco, P. Goncalves, F. Brito, and G. Kirchengast, Variational optimization for global climate analysis on esa's high performance computing grid, *Remote Sensing of Environment*, 112, 2008.
- [Lorenc(1992)] Lorenc, A. C., Analysis methodes for numerical weather prediction, *Qart. J. Ray. Meteor. Soc.*, 112, 1177–1194, 1992.
- [Person(2003)] Person, A., *User Guide to ECMWF Forecast Products*, 2003.
- [Randel et al.(2002)Randel, Chanic, and Michaut] Randel, W., M. L. Chanic, and C. Michaut, Sparc intercomparison of middel atmosphere climatologies, *Tech. rep.*, WCRP 116, 2002.
- [Reigber et al.(1995)Reigber, Schwinzer, and Kohlhasse] Reigber, C., P. Schwinzer, and A. Kohlhasse, Champ - a challenging micro satellite payload for geophysical research and application, feasibility study for dara, final report, *Tech. rep.*, Geo Forschungs Zentrum Potsdam, Germany, 1995.
- [Reigber et al.(2003)Reigber, Lühr, and Scheintzer] Reigber, C., H. Lühr, and P. Scheintzer (Eds.), *First CHAMP Mission Results for Gravity, Magnetic and Atmospheric Studies*, Springer, 2003.
- [Retscher et al.(2006)Retscher, Goncalves, Brito, and Fusco] Retscher, C., P. Goncalves, F. Brito, and L. Fusco, Grid on-demand enabling earth observation applications: The atmosphere, in *Grid-Enabling Legacy Applications and Supporting End Users (GELA)*, 2006.
- [Ringer and Healy(2008)] Ringer, M. A., and S. B. Healy, Monitoring twenty-first century climate using gps radio occultation bending angles, *Geophys. Res. Lett.*, 35, 2008.
- [Roeckner et al.(2003)] Roeckner, E., et al., The atmospheric general circulation model echam5 parti, *Tech. Rep. Report No.349*, Max-Plank-Institut für Meteorologie, Hamburg, 2003.
- [Simmons(2004)] Simmons, A. J., Development of the era40 data assimilation system, *ECMWF*, p. 30, 2004.
- [Sokolovskiy(2003)] Sokolovskiy, S. V., Effects of superrefraction on inversions of radio occultation signals in the lower troposphere, *Radio Sci.*, 38(3), 1058, 2003.
- [Sokolovskiy and Hunt(1996)] Sokolovskiy, S. V., and D. Hunt, Statistical optimisation approach for gps/met data inversions, paper presented at the ursi gps/met workshop, tucson, arizona, in *URSI GPS/Met Workshop*, 1996.
- [Steiner(2004)] Steiner, A. K., Error analyses of refractivity profiles retrieved from champ radio occultation data, *Tech. Rep. Scientific Report 04-02*, Danish Meteorological Institute, 2004.
- [Steiner and Kirchengast(2004)] Steiner, A. K., and G. Kirchengast, Error analysis for gnss radio occultation dada based on ensembles of profiles from end-to-end simulations, *J. Geophys. Res.*, 110, 2004.
- [von Engeln(2006)] von Engeln, A., A first test of climate monitoring with radio occultation instruments: comparing two processing centers, *Geophys. Res. Lett.*, 33, 2006.

- [*Vorobev and Krasnilnikova(1994)*] Vorobev, V. V., and T. G. Krasnilnikova, Estimation of the accuracy of the atmospheric refractive index recovery from doppler shift measurements at frequencies used in the navstar system, *Phys. Atmos. Ocean*, 29, 602–609, 1994.
- [*Wickert(2002)*] Wickert, J., Das champ-radiookkultationsexperiment: Algorithmen, prozessierungssystem und erste ergebnisse, *Tech. Rep. Scientific Technical Report STR02/07*, Geo-ForschungsZentrum Potsdam, Germany, 2002.
- [*Wickert et al.(2002)*] *Wickert, Schmidt, Marquardt, Reigber, Neumayer, Beyerle, Galas, and Grunwald* Wickert, J., T. Schmidt, C. Marquardt, C. Reigber, K. H. Neumayer, G. Beyerle, R. Galas, and L. Grunwald, Gps radio occultation with champ: First results and status of the experiment, proc. of iag scientific assambly, 2.-7. september, budapest, in *Proc. of IAG Scientific Assambly*, vol. 125, 2002.
- [*Wickert et al.(2001)*] Wickert, J., et al., Atmosphere sounding by gps radio occultation: First results from champ, *J. Geophys. Res.*, 28, 3263–3266, 2001.
- [*Zupanski(1993)*] Zupanski, M., A precondition algorithm for large-scale minimisation problems, *Tellus*, 45A, 478–492, 1993.