BONNER METEOROLOGISCHE ABHANDLUNGEN

Heft 69 (2015) (ISSN 0006-7156)
Herausgeber: Andreas Hense

Christoph Bollmeyer

A HIGH-RESOLUTION REGIONAL REANALYSIS FOR EUROPE AND GERMANY - CREATION AND VERIFICATION WITH A SPECIAL FOCUS ON THE MOISTURE BUDGET

BONNER METEOROLOGISCHE ABHANDLUNGEN

Heft 69 (2015) (ISSN 0006-7156)
Herausgeber: Andreas Hense

Christoph Bollmeyer

A HIGH-RESOLUTION REGIONAL REANALYSIS FOR EUROPE AND GERMANY - CREATION AND VERIFICATION WITH A SPECIAL FOCUS ON THE MOISTURE BUDGET

A HIGH-RESOLUTION REGIONAL REANALYSIS FOR EUROPE AND GERMANY

-

CREATION AND VERIFICATION WITH A SPECIAL FOCUS ON THE MOISTURE BUDGET

DISSERTATION
ZUR
ERLANGUNG DES DOKTORGRADES (DR. RER. NAT.)
DER
MATHEMATISCH-NATURWISSENSCHAFTLICHEN FAKULTÄT
DER
RHEINISCHEN FRIEDRICH-WILHELMS-UNIVERSITÄT BONN

vorgelegt von
Christoph Bollmeyer
aus
Klosterbrück

Bonn, April 2015

Diese Arbeit ist die ungekürzte Fassung einer der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn im Jahr 2015 vorgelegten Dissertation von Christoph Bollmeyer aus Klosterbrück.

This paper is the unabridged version of a dissertation thesis submitted by Christoph Bollmeyer born in Klosterbrück to the Faculty of Mathematical and Natural Sciences of the Rheinische Friedrich-Wilhelms-Universität Bonn in 2015.

Anschrift des Verfassers: Address of the author:

Christoph Bollmeyer
Meteorologisches Institut der
Universität Bonn
Auf dem Hügel 20
D-53121 Bonn

1. Gutachter: Prof. Dr. Andreas Hense, Rheinische Friedrich-Wilhelms-Universität Bonn
2. Gutachter: Prof. Dr. Leopold Haimberger, Universität Wien

Tag der Promotion: 04. September 2015
Erscheinungsjahr: 2015

Contents

List of Figures and Tables

List of Figures

List of Tables

Abstract

Atmospheric reanalyses represent a state-of-the-art description of the Earth's atmospheric state over the past years or decades. They are comprised of a numerical model for the solution of the equations of motion describing the atmosphere and of a data assimilation system for the use of observational data within the system in order to keep the reanalysis as close to the observed atmospheric state as possible. Several large reanalysis data sets exist, created by the largest meteorological centres and research institutes. Most of them, however, are global reanalyses spanning several decades or even the whole 20th century and are thus of a relatively coarse horizontal resolution of 40 to 120 km and temporal resolution of 3 to 6 hours. Those reanalyses are well suited for studying the global climate conditions and the climate change but are ineligible for regional studies on much smaller domains since they are unable to resolve small scale features in the model domains. When studying the impact of climate change on small domains, e.g. only for Germany, the coupling of atmosphere and surface or sub-surface models or local atmospheric and hydrological features, data sets with a high resolution are needed.

Therefore, the main focus of this work is on developing and operating two high-resolution regional reanalyses for two domains: the first covering Europe at a horizontal resolution of 6 km and the second covering Germany and surrounding states at a horizontal resolution of 2 km, both with a temporal resolution of one hour and less. The setup of the complete system driving the reanalysis is described along with the models behind it. The two models are evaluated against independent observations and the superiority of the regional reanalyses against a global reanalysis and a dynamical downscaling is shown.

A special focus in the verification of the reanalysis for Europe is on the moisture budget, which comprises the divergence of the horizontal moisture transports and the vertically integrated moisture flux divergence. On average time scales of a year, the moisture transport should balance the moisture flux divergence in the reanalysis, which is not the case. An approach for the modification of the moisture transports and flux divergence in a consistent way to fulfill this fundamental balance using finite elements is proposed. The method and consecutive results are presented and discussed.

Zusammenfassung

Atmosphärische Reanalysen repräsentieren den neuesten Stand der Technik in der Beschreibung des atmosphärischen Zustands der Erde über die vergangenen Jahre oder Jahrzente. Sie setzen sich zusammen aus einem numerischen Modell für die Lösung der Zustandsgleichungen der Atmosphäre sowie einem Datenassimilationssystem für die Nutzung von Beobachtungsdaten um die Reanalyse so nah wie möglich am beobachteten Zustand der Atmosphäre zu halten. Es existieren mehrere große Reanalysedatensätze, die von den größten meteorologischen Wetterdiensten und Forschungsinstituten erzeugt wurden. Die meisten davon sind jedoch globale Reanalysen, die mehrere Jahrzente oder sogar das komplette 20. Jahrhundert umfassen, und haben daher eine relativ grobe horizontale Auflösung von 40 bis 120 km und eine zeitliche Auflösung von 3 bis 6 Stunden. Diese Reanalysen eignen sich sehr gut für Untersuchungen des globalen Klimas und des globalen Klimawandels aber sind ungeeignet für regionale Untersuchungen auf viel kleineren Skalen, da sie kleinskalige Strukturen im Modellgebiet nicht auflösen können. Für Untersuchungen des regionalen Klimawandels auf kleinen Gebieten, zum Beispiel für Deutschland, sowie für das Koppeln von Atmosphärenmodellen mit Landoberflächen- oder Bodenmodellen oder die Analyse von lokalen atmosphärischen und hydrologischen Strukturen, werden Datensätze mit höherer Auflösung benötigt.
Daher liegt der Hauptfokus dieser Arbeit auf der Entwicklung und dem operationellen Betrieb zweier hochaufgelöster regionaler Reanalysen für zwei verschiedene Gebiete: das erste Gebiet umfasst Europa mit einer horizontalen Auflösung von 6 km und das zweite umfasst Deutschland und umliegende Staaten mit einer horizontalen Auflösung von 2 km. Beide Reanalysen haben eine zeitliche Auflösung von einer Stunde und weniger. Der Aufbau des kompletten Systems zur Erstellung der Reanalyse wird beschrieben zusammen mit den dahinterstehenden Modellen. Beide Modelle werden gegen unabhängige Beobachtungen evaluiert und die Überlegenheit der regionalen Reanalyse gegen eine globale Reanalyse und ein dynamisches Downscaling wird gezeigt.
Ein besonderer Fokus in der Verifikation der Reanalyse für Europa wurde auf das Feuchtebudget der Atmosphäre gelegt. Dieses setzt sich zusammen aus der Divergenz der horizontalen Feuchtetransporte und der vertikal integrierten Feuchteflußdivergenz. In einer gemittelten Zeitskala von einem Jahr sollte die Feuchtetransportdivergenz die Feuchteflußdivergenz exakt ausgleichen, was in der Reanalyse nicht der Fall ist. Deshalb wird ein Ansatz mit finiten Elementen vorgeschlagen, der sowohl die Feuchtetransporte als auch die Feuchteflußdivergenz in einer konsistenten Art und Weise modifiziert, sodass diese fundamentale Bilanz erfüllt ist. Die Methode und die daraus resultierenden Ergebnisse werden präsentiert und diskutiert.

1. Introduction

An atmospheric reanalysis is a description of the state of the atmosphere in a consistent, four-dimensional way by the use of a numerical weather prediction model and a corresponding data assimilation scheme in order to take as many observations of the atmospheric state into account as possible. Observations are not evenly distributed in space and time but the model, due to its physical formulation, is able to fill the space between the individual observations in a physically consistent way (Bengtsson and Shukla, 1988). A reanalysis is always produced for a past time span and represents the best estimation of the four-dimensional atmospheric state in predefined spatio-temporal boundaries.
Reanalysis data sets are therefore suited for climatological and meteorological, e.g. atmospheric or hydrological, studies on nearly any scale. They have grown to become a key instrument in the monitoring of climate and its attributes (Trenberth *et al.*, 2008). Analyses of the state of the atmosphere are produced at every national weather centre several times a day, e.g. 8 times a day with the COSMO-DE model at the German Meteorological Service (Deutscher Wetterdienst, DWD), since they serve as the basis and initial state for the weather forecasts. But using these operational analyses for climate studies would lead to inconsistencies since the operational model is always subject to improvements and therefore frequently changed (see Bengtsson and Shukla, 1988). Furthermore, during operational production, observations can only be assimilated up to a cut-off time and delayed data cannot be used. In the production of the reanalysis the state-of-the-art model (or model version) is kept fixed during the complete production and is thus used to *re*produce *analyses* for a given past time span on a predefined domain with all available observations assimilated in the model.
Most reanalyses are available for a global domain such as ERA40 (Uppala *et al.*, 2005) and ERA-Interim (Dee *et al.*, 2011) by the European Centre for Medium-Range Weather Forecasts (ECMWF) or the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis (Kalnay *et al.*, 1996) but also the NCEP Climate Forecasting System Reanalysis (CFSR, Saha *et al.*, 2010), the Modern-Era Retrospective Analysis for Research and Applications (MERRA) by the National Aeronautics and Space Administration (NASA) (Rienecker *et al.*, 2011) and the Japanese 25-Year and 55-Year Reanalysis Project (JRA-25 (Onogi *et al.*, 2007) and JRA-55 (Ebita *et al.*, 2011)) by the Japan Meteorological Agency (JMA). All these reanalyses have in common that they use a large observational data set, which is comprised of conventional observations as well as satellite observations, and a global circulation model (GCM) together with a corresponding data assimilation scheme. CFSR even uses a coupled atmosphere-ocean circulation system. Global reanalyses have horizontal resolutions of approximately 125 km to 40 km and a temporal resolution of 6 hours, sometimes of 3 hours when intermediate model forecasts are provided (as e.g. in ERA-Interim). The only exceptions are the MERRA and CFSR reanalyses which provide some of the output fields every hour. An overview of the temporal and spatial scales of the mentioned reanalyses is given in Figure 1.1. Recent reanalysis efforts also cover air quality and global atmospheric composition information, like the Monitoring Atmospheric Composition and Climate (MACC) reanalysis by Inness *et al.*, 2013. All those reanalyses are very useful in the study of atmospheric patterns and phenomena or climate change but due to their rather

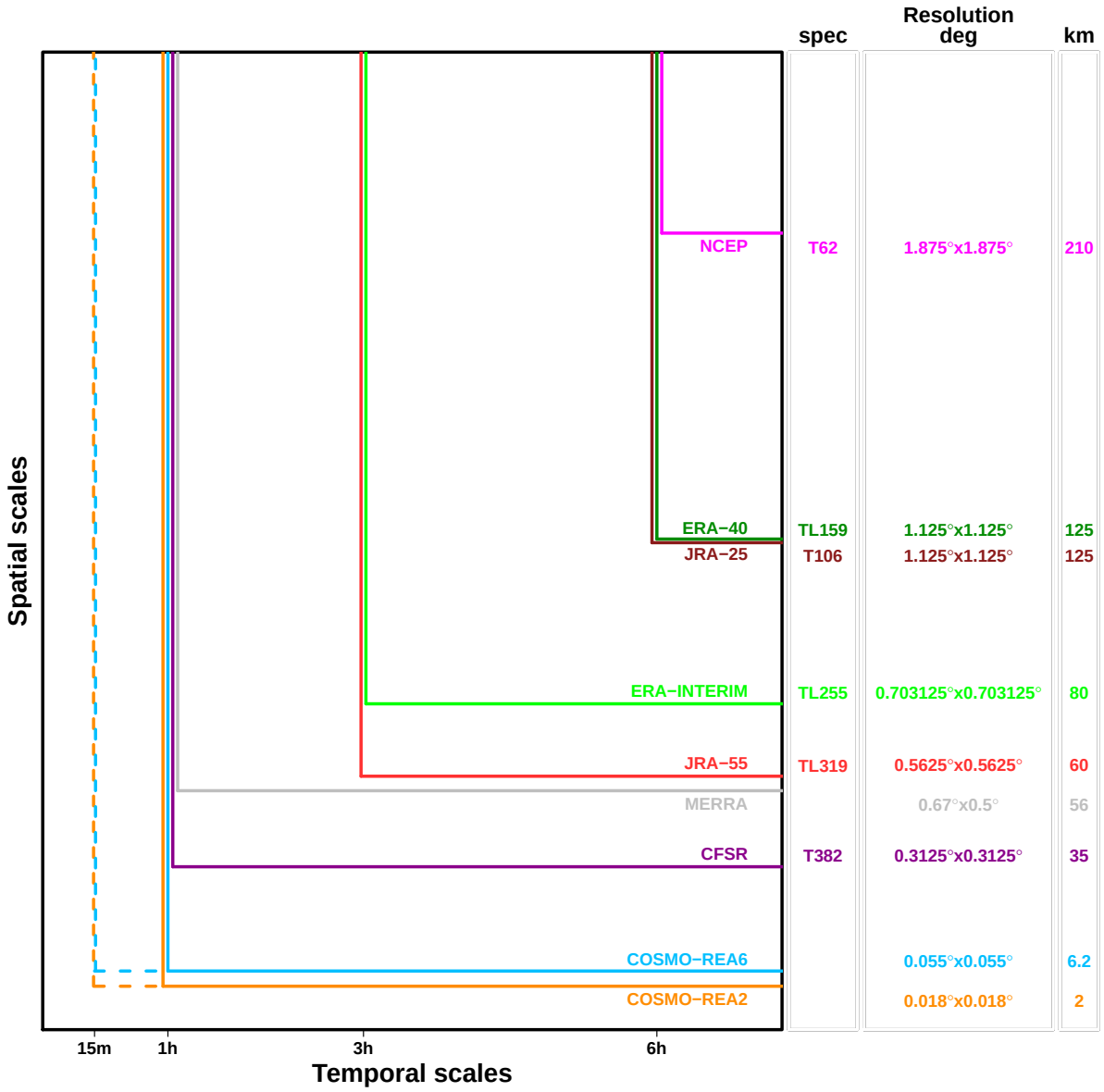


Figure 1.1.: Temporal and spatial scales of different global and regional reanalyses. The used resolution is shown in spectral resolution (if the reanalysis was run in spectral mode), the degrees in longitude times latitude and the approximate horizontal resolution in km. T denotes the spectral truncation for a quadratic Gaussian grid and TL for a linear Gaussian grid.

coarse resolution they are not suited for applications on smaller scales. The scientific community is getting more and more interested in reanalyses on much higher spatial and temporal scales, i.e. below 10 km spatial and below 3 hour temporal resolution for different reasons. For instance, hydrologists require convection-resolving precipitation data sets as boundary forcing for their even finer models to account for local extreme events.
Furthermore, applications of high-resolution reanalysis data can be found in the risk assessment of severe weather events or in the renewable energy sector. The development of forecast systems for wind and solar energy e.g. depends on observational data sets that are consistent

in space, time and between parameters. The covariance structure of e.g. wind speed and cloud cover on small spatio-temporal scales serves to determine optimal locations for power production.
Finally, higher resolution is needed in climate monitoring, especially on the local scale. Firstly, higher resolutions help to improve the estimation of the impact of climate change on those scales and secondly, the smaller resolved scales can support the understanding of mechanisms responsible for local climate features and feedbacks. Furthermore, coupled hydrological reanalyses for the full interaction and description of the water and energy exchange between the atmosphere and the surface rely on high resolution models as well as chemical reanalysis which aim at the local emission and immission scale.
Due to these applications the regional enhancement of the available global reanalysis data has become an important task. One way of addressing this problem, which is often exploited in meteorology, is the use of downscaling techniques, either statistical or dynamical, to obtain data in the desired resolution. In dynamical downscaling a fine-scale numerical atmospheric limited-area climate model is used with boundary conditions coming from a coarser global circulation model, which is an established technique in regional climate models (RCMs), whereas in statistical downscaling a statistical relationship is applied to output data from a GCM to achieve detailed regional atmospheric data (Castro *et al.*, 2005; Wilby and Wigley, 1997). In statistical downscaling, additional information can be introduced into the statistical model by using a priori information such as orography. Dynamical downscaling is often used to generate spatially enhanced data sets from global reanalyses. However, this approach depends on the model to infer fine-scale detail from low-resolution initial and boundary conditions which is always subject to errors. Especially on large domains, nested regional models tend to develop internal variability causing significant differences in the actual spatio-temporal state of the system. This was shown by Simon *et al.*, 2013 using coherence spectra between the boundary model and the regional model. To reduce the errors in high-resolution simulations and avoid the underlying assumption of a perfect model, observations can be used in a data assimilation framework, thus enhancing the quality of the simulations. This approach, i.e. the simulation of regional climate using a high-resolution regional model with the use of observations via a data assimilation approach, is called regional reanalysis.
The first successful implemented long-term regional reanalysis was the North American Regional Reanalysis (NARR). Mesinger *et al.*, 2006 show that NARR outperforms its driving global reanalysis (GR2) in the analysis of 2 m temperature, 10 m wind as well as upper-air temperature and wind. In addition, the moisture budget is closer to closure than in GR2. However, NARR is still using a rather coarse resolution of 32 km. Therefore, current reanalysis efforts of the community aim at higher resolutions of 10 km and beyond, e.g. the Arctic System Reanalysis (Bromwich *et al.*, 2010) as well as efforts in the European Reanalysis and Observations for Monitoring (EURO4M) project (http://www.euro4m.eu).
This work was carried out within the “Hans-Ertel Centre for Weather Research - Climate Monitoring Branch” which is funded by the Federal Ministry of Transport and Digital Infrastructure (BMVI) of Germany. The focus of the research project was a self-consistent assessment and analysis of regional climate in Germany and Central Europe. In order to achieve this goal, two high-resolution regional reanalyses for Europe and Germany at horizontal resolutions of 6 km and 2 km have been developed in this work, providing homogenized data sets for the study of the regional climate and climate change. The documentation of the model and the description of the implementation and setup of the system producing the reanalysis will be presented in the first part of this work. The second part of this work is concerned with a detailed analysis

of the moisture budget. The hydrological water cycle in the atmosphere is of high importance since water is evaporated over the oceans, condensates again into clouds, falls out as rain and snow over land and again reaches the oceans via rivers and groundwater runoff, thus representing one of the main drivers of the global climate. Every atmospheric model should be consistent in the storage of the different moisture components, i.e. total water mass should be conserved over time in the model. This storage is described by the moisture budget. In global models, the model is not dependent on boundary data and inconsistencies in the global budgets are a problem of the model itself. Several variational approaches exist for the correction of the erroneous mass budget in global reanalyses to correct the energy budget and could thus be also used for the closure of the moisture budget (Ehrendorfer *et al.*, 1994; Hacker, 1981; Hantel and Haase, 1983). Regional models however are clearly dependent on the flow into and out of the model domain and therefore add another source of errors in addition to the internal inconsistencies of the model. The moisture budget in the regional reanalysis at hand is not in balance which is why a variational approach has been applied to modify the moisture transports and the vertically integrated moisture flux divergence in a consistent way to fulfill the balance in the reanalysis output. The approach is presented in the second part of this work.
The third part presents results on the performance of the reanalysis and on the verification against independent observations and other reanalysis and data products as well as results on the variational approach. The work is closed with a summary and conclusions.

2. A high-resolution regional reanalysis for Europe and Germany

Every daily numerical weather prediction consists of the following three steps (see Bjerknes, 2009). Firstly, the state of the atmosphere for a given date and time as seen from observations has to be determined. Secondly, based on the observations, the initial physical and dynamical state of the complete atmosphere has to be derived by solving the thermo- and hydrodynamical equations which describe the processes in the atmosphere. Finally, those equations have to be integrated in time and solved for future dates and times. The last two steps, i.e. analysing and forecasting the state of the atmosphere, is done with a numerical model. At the moment there are two different models producing the operational weather forecasts at DWD. These are the global model GME (Majewski and Ritter, 2002) (which will be replaced in the near future by the new Icosahedral non-hydrostatic general circulation model ICON[1]) and the regional model COSMO, which is nested into GME for the European domain (COSMO-EU) and into itself for Germany (COSMO-DE). To start a weather prediction, an estimate of the initial state of the atmosphere is needed. Observations of the various meteorological variables, although subject to observational errors, are the best way to determine this state which is why they are used in a data assimilation scheme. In the data assimilation the observations are passed to the model to produce an analysis which is then the best possible approximation to the initial state of the atmosphere. The data assimilation schemes are usually designed to account for the errors in both the model and the observations and the model is able to fill the space and/or time between observations in a physically consistent way. The analysis serves as the initial state of the weather forecast. In operational production an analysis is usually produced at several times a day, e.g. every three hours. Since analyses provide the best state of the atmosphere they can be used for meteorological and climatological studies. However, a data set comprised of several years of analyses generated during operational production will always be inconsistent since the model used for operational production is being improved permanently and new model versions are implemented frequently. This changes the internal representation of the model atmosphere and can give rise to jumps in such a record (Bengtsson and Shukla, 1988). A reanalysis describes the reproduction of the analysis of historical times with a constant model version, resulting in a consistent data set in both space and time.
In the next sections, the COSMO model which was used for the production of the reanalysis is presented (section 2.1) and afterwards in section 2.2 the setup and technical implementation of the reanalysis with the different production steps is described.

2.1. The COSMO-Model

The COSMO-Model is a non-hydrostatic limited-area atmospheric prediction model developed at DWD. The model has been designed for the operational numerical weather prediction (NWP) and for different scientific applications on the meso-β and meso-γ scale and has been

[1] http://www.mpimet.mpg.de/en/science/models/icon.html

run operationally at DWD since 1999. The COSMO-Model is based on the primitive equations of the atmosphere describing compressible flow in a moist atmosphere. The model equations are formulated in rotated geographical coordinates and a generalized terrain following height coordinate. Various physical processes are taken into account by physical parametrization schemes (Doms *et al.*, 2011; Schättler *et al.*, 2011). The main details of the model formulation are presented in the following.

2.1.1. The model equations

The atmosphere in the COSMO model is described by the Navier-Stokes equations for atmospheric flow. In the following, the basic equations and their transformations for numerical reasons are described. A complete description and derivation of the dynamic equations and their numeric implementation can be found in Doms, 2011.
The atmosphere is considered to consist of dry air, water vapour, liquid water and water in different solid states. Each of this constituents is represented by a prognostic equation, where the liquid and solid forms of water are further subdivided into cloud droplets and raindrops as well as cloud ice and snow, respectively. Considering the conservation laws of momentum, mass and heat, the basic budget equations are as follows:

$$\rho \frac{d\vec{v}}{dt} = -\nabla p + \rho \vec{g} - 2\vec{\Omega} \times (\rho \vec{v}) - \nabla \cdot \underline{\mathbf{t}} \tag{2.1}$$

$$\frac{d\rho}{dt} = -\rho \nabla \cdot \vec{v} \tag{2.2}$$

$$\rho \frac{dq^x}{dt} = -\nabla \cdot \mathbf{J}^x + I^x \tag{2.3}$$

$$\rho \frac{de}{dt} = -p \nabla \cdot \vec{v} - \nabla \cdot (\mathbf{J}_e + \mathbf{R}) + \varepsilon \,. \tag{2.4}$$

The index x represents one of the constituents of the air, namely

$$x = \begin{cases} d & \text{for dry air} \\ v & \text{for water vapour} \\ l & \text{for liquid water} \\ f & \text{for water in frozen form} \,. \end{cases} \tag{2.5}$$

The following symbols and definitions are used here:

$\rho = \sum_x \rho^x$	total density of the air mixture
ρ^x	partial density of mixture constituent x
$\vec{v}$	wind velocity (relative to the rotating earth)
t	time
∇	Nabla operator
p	pressure
$\vec{g}$	earth´s acceleration
$\vec{\Omega}$	earth´s rotation velocity
$\underline{\mathbf{t}}$	stress tensor due to viscosity
$q^x = \rho^x/\rho$	mass fraction of constituent x
$\mathbf{J}^x$	diffusion flux of constituent x
I^x	sources and/or sinks of constituent x
e	specific internal energy
$\mathbf{J}_e$	diffusion flux of internal energy (heat flux)
$\mathbf{R}$	flux density of solar and thermal radiation
$\varepsilon = -\underline{\mathbf{t}} \cdot\cdot \nabla\vec{v}$	kinetic energy dissipation due to viscosity

The budget form of the basic equations (2.1)-(2.4) presented above can easily be transformed into flux form with the help of the following equation

$$\rho \frac{d\psi}{dt} = \frac{\partial(\rho\psi)}{\partial t} + \nabla\cdot(\rho\vec{v}\psi) \tag{2.6}$$

which describes the rate of change of any mass specific quantity ψ. The sources and sinks of the constituents x, which are consolidated in I^x, are generally processes where water undergoes phase changes or where water is generated and lost in chemical reactions with the components of dry air. But since chemical changes can be neglected in mesoscale applications, I^d is set to zero in the budget equation of dry air.
It is now assumed that dry air and water vapour behave like ideal gases and that liquid water and ice are incompressible substances. Under this assumption and the further finding that q^l and q^f are much smaller than 1, the equation of state for a moist atmosphere reads

$$\begin{aligned} p &= \rho(R_d q^d + R_v q^v)T \\ &= \rho R_d\left(1 + (R_v/R_d - 1)q^v - q^l - q^f\right)T \\ &= \rho R_d T_v \end{aligned} \tag{2.7}$$

where R_d and R_v are the gas constants for dry air and water vapour and T is the generalized virtual temperature

$$T_v = \left(1 + (R_v/R_d - 1)q^v - q^l - q^f\right)T \tag{2.8}$$

$$= (1+\alpha)T \tag{2.9}$$

and α abbreviates the moisture term $\alpha = (R_v/R_d - 1)q^v - q^l - q^f$. In the basic set of equations (2.1)-(2.4) and the equation of state (2.7), the temperature is a diagnostic variable which needs to be determined from the internal energy e or from the enthalpy h

$$\rho \frac{dh}{dt} = \frac{dp}{dt} - \nabla \cdot (\mathbf{J}_e + \mathbf{R}) + \varepsilon \,. \tag{2.10}$$

In a numerical treatment, however, it is advantageous to have a prognostic equation for temperature, the so-called heat equation. The heat equation can be obtained from an expansion of the enthalpy $h(T, p, q^x) = \sum_x h_x q^x$ in the way

$$\frac{dh}{dt} = \left(\frac{\partial h}{\partial T}\right)_{p,q^x} \frac{dT}{dt} + \left(\frac{\partial h}{\partial p}\right)_{T,q^x} \frac{dp}{dt} + \sum_x \left(\frac{\partial h}{\partial q^x}\right)_{T,p} \frac{dq^x}{dt} \,. \tag{2.11}$$

The partial specific enthalpies for the different moisture constituents are given by

$$h_x = h_x^0 + c_{px}(T - T_0) \tag{2.12}$$

with the reference temperature $T_0 = 273.15K$, h_x^0 the specific enthalpy of constituent x at reference temperature t_0 and $c_p x$ the specific heat of constituent x at constant pressure. In Eq. (2.12), the variations of h_l and h_f with pressure are assumed to be small and are thus neglected. Using (2.12) in (2.11) results in

$$\begin{aligned}
\left(\frac{\partial h}{\partial T}\right)_{p,q^x} &= c_p = \sum_x c_{px} q^x \\
\left(\frac{\partial h}{\partial p}\right)_{T,q^x} &= 0 \\
\left(\frac{\partial h}{\partial p}\right)_{T,q^x} &= h_x = h_x^0 + c_{px}(T - T_0) + \,.
\end{aligned}$$

Inserting these partial derivatives into the enthalpy equation (2.10) yields then the heat equation

$$\rho c_p \frac{dT}{dt} = \frac{dp}{dt} + l_V I^l + l_S I^f - \nabla \cdot (\mathbf{J}_s + \mathbf{R}) - \sum_x c_{px} \mathbf{J}^x \cdot \nabla T + \varepsilon \,. \tag{2.13}$$

Here, $\mathbf{J}_s$ is the sensible heat flux and l_V and l_S are the latent heat of vaporization and sublimation, respectively, and c_p is the specific heat of moist air at constant pressure. To calculate the temperature from the heat equation (2.13) the total derivative of pressure is needed. The corresponding pressure tendency equation can be obtained by derivation of the equation of state (2.7)

$$\frac{dp}{dt} = \frac{p}{\rho}\frac{d\rho}{dt} + \rho R_d T \frac{d\alpha}{dt} + \rho R_d (1 + \alpha) \frac{dT}{dt} \,. \tag{2.14}$$

Inserting the continuity equation (2.2), the budget equations for the moisture constituents (2.3) and the heat equation (2.13) in (2.14) yields

$$\left\{1 - (1 + \alpha)\frac{R_d}{c_p}\right\} \frac{dp}{dt} = -p \nabla \cdot \vec{v} + (1 + \alpha)\frac{R_d}{c_p} Q_h + Q_m \,. \tag{2.15}$$

Q_h is the diabatic heat production per unit volume of air and Q_m represents the impact of concentration changes of the humidity constituents on the pressure tendency with

$$Q_h = l_V I^l + l_S I^f - \nabla \cdot (\mathbf{J}_s + \mathbf{R}) - \sum_x c_{px} \mathbf{J}^x \cdot \nabla T + \varepsilon \tag{2.16}$$

and

$$Q_m = \rho R_d T \frac{d\alpha}{dt} = -R_v T (I^l + I^f) - R_v T \nabla \cdot \mathbf{J}^v - R_d T \nabla \cdot \mathbf{J}^d \,. \tag{2.17}$$

The term $(1+\alpha)R_d$ can be reformulated

$$(1+\alpha)R_d = R_d q^d + R_v q^v = c_p - c_v \,, \tag{2.18}$$

with c_v the specific heat at constant volume. The liquid and solid forms of water are not included here, since the specific heat at constant pressure and constant volume for these substances are the same due to the assumption of incompressibility above. Inserting (2.18) into the pressure tendency equation (2.15) gives

$$\frac{dp}{dt} = -(c_p/c_v) p \nabla \cdot \vec{v} + (c_p/c_v - 1) Q_h + (c_p/c_v) Q_m \,. \tag{2.19}$$

Now, the continuity equation (2.2) has to be replaced by (2.19) to calculate the pressure tendency. In consequence the total density becomes a diagnostic variable which can be calculated from the equation of state. The state of the atmosphere can then be calculated by the following set of equations:

$$\begin{aligned}
\rho \frac{d\vec{v}}{dt} &= -\nabla p + \rho \vec{g} - 2\vec{\Omega} \times (\rho \vec{v} - \nabla \cdot \underline{\mathbf{t}} \\
\frac{dp}{dt} &= -(c_p/c_v) p \nabla \cdot \vec{v} + (c_p/c_v - 1) Q_h + (c_p/c_v) Q_m \\
\rho c_p \frac{dT}{dt} &= \frac{dp}{dt} + Q_h \\
\rho \frac{dq^x}{dt} &= -\nabla \cdot \mathbf{J}^x + I^x \\
\rho &= p \big(R_d (1+\alpha) T \big)^{-1} \,.
\end{aligned} \tag{2.20}$$

This set of equations has two drawbacks: The first being that the conservation of total mass is not guaranteed but depends on the accuracy of the numerical algorithm. The second being the appearance of the diabatic heating rate Q_h and the moisture source term Q_m in the pressure tendency equation. Both Q_h and Q_m are important for the thermodynamical feedbacks due to diabatic heating as well as the representation of thermal compression waves, but they cause numerical problems. To avoid this, these terms are usually neglected and this neglection results in the introduction of artificial sources and sinks in the continuity equation. This error is considered to be small and the decision to stick with this set of equation is based on numerical efficient schemes for the treatment of sound waves in the model. Sound waves travel at high velocities and therefore require small time steps in the integration to guarantee a stable evolution of the model. These schemes can easily be applied in the set of equations (2.20).

2.1.1.1. Averaging and simplifications

Numerical models cannot solve differential equations exactly, as would be required from mathematics, but only in a discrete formulation with finite grid spacings and finite time steps. In mesoscale applications, as they are used in the COSMO model in this work, the grid spacing is in the order of some kilometers while the time step is in the range of 20 to 50 seconds. A grid spacing of the model in the order of millimeters determined from the Kolmogorov length

scale, as would be required theoretically, is not possible and probably never will be. Therefore, the basic equations have to be averaged over the grid boxes and over the chosen time step.
It is assumed, that every variable ψ can be decomposed into its average and a deviation from this average

$$\psi = \overline{\psi} + \psi' \tag{2.21}$$

where $\overline{\psi}$ is the average over the finite time interval dt and the finite grid spacings in all three dimensions Δx, Δy and Δz

$$\overline{\psi} = \frac{1}{\Delta x \Delta y \Delta z \Delta t} \int\int\int\int \psi dx dy dz dt \,. \tag{2.22}$$

The average of the deviations ψ' is defined to be zero, following Reynolds, 1895:

$$\overline{\psi'} = 0 \,. \tag{2.23}$$

Additionally, a decomposition into a mass-weighted average $\widehat{\psi}$ and its deviation ψ'' is defined

$$\psi = \widehat{\psi} + \psi'' \tag{2.24}$$

where

$$\widehat{\psi} = \overline{\rho\psi}/\overline{\rho} \tag{2.25}$$
$$\widehat{\psi''} = 0 \,. \tag{2.26}$$

The mass-weighted average should be used for velocity and for mass specific variables as e.g. the concentrations q^x of the different constituents (Mieghem, 1973).
One problem common in all models is their finite grid resolution. Some processes can be described by the model, others which are small-scale processes like e.g. turbulence cannot be resolved by the model. The averaging of the model equations helps to separate these processes from each other. The large-scale, slowly varying processes can be resolved by the model and are identified by the mean values $\overline{\psi}$ and $\widehat{\psi}$, whereas the small-scale, fast varying processes are expressed by the deviations ψ' and ψ'', which are, as already mentioned, not resolvable. These not resolvable processes are nevertheless important for the correct evolution of the model and thus need to be parametrized. The most important of the used parametrizations of the COSMO model are described in section 2.1.6.
The previously derived averaging operator can now be applied to the basic equations for the conservation of momentum, mass and water (2.1), (2.2) and (2.3) as well as on the enthalpy equation (2.10). This leads to the following set of averaged equations:

$$\overline{\rho}\frac{\widehat{d\vec{v}}}{dt} = -\nabla\overline{p} + \overline{\rho}\vec{g} - 2\vec{\Omega}\times(\overline{\rho}\widehat{\vec{v}}) - \nabla\cdot(\underline{\bar{\mathbf{t}}} + \underline{\mathbf{T}}) \tag{2.27}$$
$$\frac{\widehat{d\overline{\rho}}}{dt} = -\overline{\rho}\nabla\cdot\widehat{\vec{v}} \tag{2.28}$$
$$\overline{\rho}\frac{\widehat{dq^x}}{dt} = -\nabla\cdot(\overline{\mathbf{J}^x} + \mathbf{F}^x) + \overline{I^x} \tag{2.29}$$
$$\overline{\rho}\frac{\widehat{dh}}{dt} = \frac{\widehat{d\overline{p}}}{dt} + B_h - \nabla\cdot(\overline{\mathbf{J}_e} + \mathbf{F}_h + \overline{\mathbf{R}}) + \overline{\varepsilon} \tag{2.30}$$
$$\overline{p} = \overline{\rho}R_d\left(1 + (R_v/R_d - 1)\widehat{q^v} - \widehat{q^l} - \widehat{q^f}\right)\widehat{T} \tag{2.31}$$
$$= \overline{\rho}R_d\widehat{T}_v \,.$$

As with the basic equations, the above averaged equations can be transformed into flux form with

$$\overline{\rho}\frac{\widehat{d\widehat{\psi}}}{dt} = \frac{\partial(\overline{\rho}\widehat{\psi}}{\partial t} + \nabla\cdot(\overline{\rho}\widehat{\vec{v}}\widehat{\psi}) \,. \tag{2.32}$$

Due to the deviations from the mean the following correlation products that are related to subgrid scale transports appear:

$\underline{\mathbf{T}} = \overline{\rho\vec{v}''\vec{v}''}$	turbulent flux of momentum
$\mathbf{F}^x = \overline{\rho\vec{v}''q^x}$	turbulent flux of constituent x
$\mathbf{F}_h = \overline{\rho\vec{v}''h}$	turbulent flux of enthalpy
$B_h = \overline{\vec{v}''\cdot\nabla p}$	source term of enthalpy due to buoyant heat and moisture fluxes

From Eq. (2.30) the heat equation can be derived, yielding

$$\overline{\rho}\widehat{c_p}\frac{\widehat{d\widehat{T}}}{dt} = \frac{\widehat{d\overline{p}}}{dt} + \overline{Q_h} \,. \tag{2.33}$$

Q_h is the mean diabatic heating and given by

$$\overline{Q_h} = \widehat{l_V}\overline{I^l} + \widehat{l_S}\overline{I^f} + B_h - \nabla\cdot(\overline{\mathbf{J}_s} + \mathbf{H} + \overline{\mathbf{R}}) - \sum_x c_{px}(\overline{\mathbf{J}^x} + \mathbf{F}^x)\cdot\nabla\widehat{T} + \overline{\varepsilon} \tag{2.34}$$

$\widehat{l_V}$ and $\widehat{l_S}$ are again the latent heat of vapourization and sublimation as defined above, but for the mean value of temperature, $\widehat{T}$, and B_h is the buoyancy term which can be written as

$$B_h = (\widehat{c_p} - \widehat{c_v})/\widehat{c_p}\mathbf{H}\cdot\nabla ln\overline{p} + R_d\widehat{T}\left((r_v/R_d - 1)\mathbf{F}^v - \mathbf{F}^l - \mathbf{F}^f\right)\cdot\nabla ln\overline{p} \tag{2.35}$$

where

$$\mathbf{H} = \mathbf{F}_h - \sum_x \widehat{h_x}\mathbf{F}^x \tag{2.36}$$

is the turbulent sensible heat flux.

2.1.1.2. Simplifications of the heat equation

For a fast but nevertheless accurate computation of thermodynamic processes, some simplifications are needed. These are the following

- *Molecular fluxes*
  The atmospheric flow on the used scales is always a turbulent one where the turbulent fluxes of momentum, heat and moisture are in general larger than the corresponding molecular fluxes. In consequence, all molecular fluxes are neglected. An exception to this rule are the diffusion fluxes of liquid and solid forms of water, which are important for microphysical growth processes of water drops and ice crystals. When water drops and ice crystals reach a certain size their fall velocity becomes large enough to create precipitation. Since this is a very important process in the atmosphere those molecular fluxes must be kept. Therefore, the viscous stress tensor and the molecular fluxes of sensible heat and of water vapour are set to zero and the water and ice fluxes are replaced by the sedimentation fluxes

$$\begin{aligned} \underline{\mathbf{t}} &= 0, \mathbf{J}_s = \mathbf{J}^v = 0 \,, \\ \mathbf{J}^l &\simeq \mathbf{P}^l = \rho q^l \vec{v}_T^{\,l} \,, \\ \mathbf{J}^f &\simeq \mathbf{P}^f = \rho q^f \vec{v}_T^{\,f} \,, \end{aligned} \tag{2.37}$$

where $\mathbf{P}^l$ and $\mathbf{P}^f$ are the precipitation fluxes of liquid water and ice and $\vec{v}_T^{\,l}$ and $\vec{v}_T^{\,f}$ their corresponding terminal velocities.

- *Approximations to the heat equation*
  The different water constituents in the atmosphere form only a very small fraction of total mass in the atmosphere. Therefore, all moist air in the atmosphere is treated as though it were dry air and the specific heat of moist air is approximated by the specific heat of dry air. Thus the impacts of the remaining diffusion fluxes of the water constituents on the temperature are neglected and the latent heat of vapourization and sublimation are replaced by their constant values at the reference temperature T_0:

$$\begin{aligned} \widehat{c_p} &= \sum_x c_{px}\widehat{q^x} \simeq c_{pd}\,, \\ \widehat{l_V}(\widehat{T}) &\simeq L_V\,, \\ \widehat{l_S}(\widehat{T}) &\simeq L_S\,, \\ \mathbf{H} &\simeq c_{pd}\overline{\rho\vec{v}''T}\,, \\ \sum_x c_{px}(\overline{\mathbf{J}_x}+\mathbf{F}_x)\cdot\nabla\widehat{T} &\simeq 0\,, \end{aligned} \tag{2.38}$$

- *Approximations to the pressure tendency equation*
  In order to guarantee a conservation of total mass as close as possible, the above mentioned approximations to the heat equation have to be applied to the pressure tendency equation as well, which results in the following form:

$$\frac{\widehat{d\overline{p}}}{dt} = -(c_{pd}/c_{vd})\overline{p}\nabla\cdot\widehat{\vec{v}} + (c_{pd}/c_{vd}-1)\overline{Q_h}\,. \tag{2.39}$$

  According to Doms, 2011, this approximation introduces a small source/sink term in the continuity equation which is believed to have no significant impact on mesoscale numerical modeling.

- *Buoyant heat and moisture fluxes*
  In addition to the simplifications previously described, the buoyant heat and moisture fluxes and the mean dissipation rate due to viscous stresses are neglected completely;

$$B_h = 0\,, \overline{\varepsilon} = 0\,. \tag{2.40}$$

  This simplification is justified by the fact that the forcing function $\nabla ln\overline{p}$ varies only very slowly with height and therefore every temperature change induced by the vertical divergence of $\mathbf{H}$ will be much larger than those caused by the buoyancy term.

Introducing these simplifications into the equations (2.20) leads to the following set of equations for the nonhydrostatic compressible mean flow of the atmosphere

$$\begin{aligned}
\rho\frac{d\vec{v}}{dt} &= -\nabla p + \rho\vec{g} - 2\Omega\times(\rho\vec{v}) - \nabla\cdot\underline{\mathbf{T}} \\
\frac{dp}{dt} &= -(c_{pd}/c_{vd})p\nabla\cdot\vec{v} + (c_{pd}/c_{vd}-1)Q_h \\
\rho c_{pd}\frac{dT}{dt} &= \frac{dp}{dt} + Q_h \\
\rho\frac{dq^v}{dt} &= -\nabla\cdot\mathbf{F}^v - (I^l + I^f) \\
\rho\frac{dq^{l,f}}{dt} &= -\nabla\cdot(\mathbf{P}^{\mathbf{l,f}} + \mathbf{F}^{\mathbf{l,f}}) + I^{l,f} \\
\rho &= p\left[R_d(1+(R_v/R_d-1)q^v - q^l - q^f)T\right]^{-1}
\end{aligned} \tag{2.41}$$

where Q_h is the rate of diabatic heating/cooling

$$Q_h = L_V I^l + L_S I^f - \nabla\cdot(\mathbf{H}+\mathbf{R}) \tag{2.42}$$

and overbars and -hats have been omitted for convenience for now and all the following. The set of equations (2.42) is used as basic equations in the COSMO model.

2.1.2. Rotated spherical coordinates

The equations (2.42) are derived with regard to the rotating earth. In many limited area models, rotated spherical coordinates are applied where the pole of the coordinate system is tilted such that the equator runs approximately through the middle of the coordinate system. In this way, numerical instabilities resulting from the convergence of the meridians and the pole singularities do not occur. Furthermore, when only a small domain is considered where the impact of the curvature of the Earth's surface is negligible, the equations become identical to those for a tangential Cartesian coordinate system. Numerical solutions for the problem of the converging meridians can be found in Haltiner and Williams, 1980.
In the COSMO model, rotated spherical coordinates are used. The effects of the rotation are shown in Fig. 2.1, where the model domain has been plotted in rotated spherical coordinates. In addition, the unrotated lines of longitude and latitude are shown in red.
The basic set of equations (2.42) must now be transcribed to the new rotated coordinate system. For further details on the coordinate transformation see Dutton, 1986 or Zdunkowski and Bott, 2003. First of all, two assumptions are made. Firstly, the earth´s acceleration is assumed to be constant and perpendicular to surfaces of constant radius, i.e.

$$\vec{g} \simeq -g(\vec{r}/r)\,, \tag{2.43}$$

where g is the constant mean value of absolute gravity acceleration.
Secondly, the height z above the ground is much smaller than the radius of the earth a, thus the height can be approximated as

$$r = a + z \simeq a\,, \tag{2.44}$$

In consequence, any reference to r in the dynamic equations can be replaced by the radius of the earth a and any differential variation of r can be replaced by a variation in z, $\partial r = \partial z$,

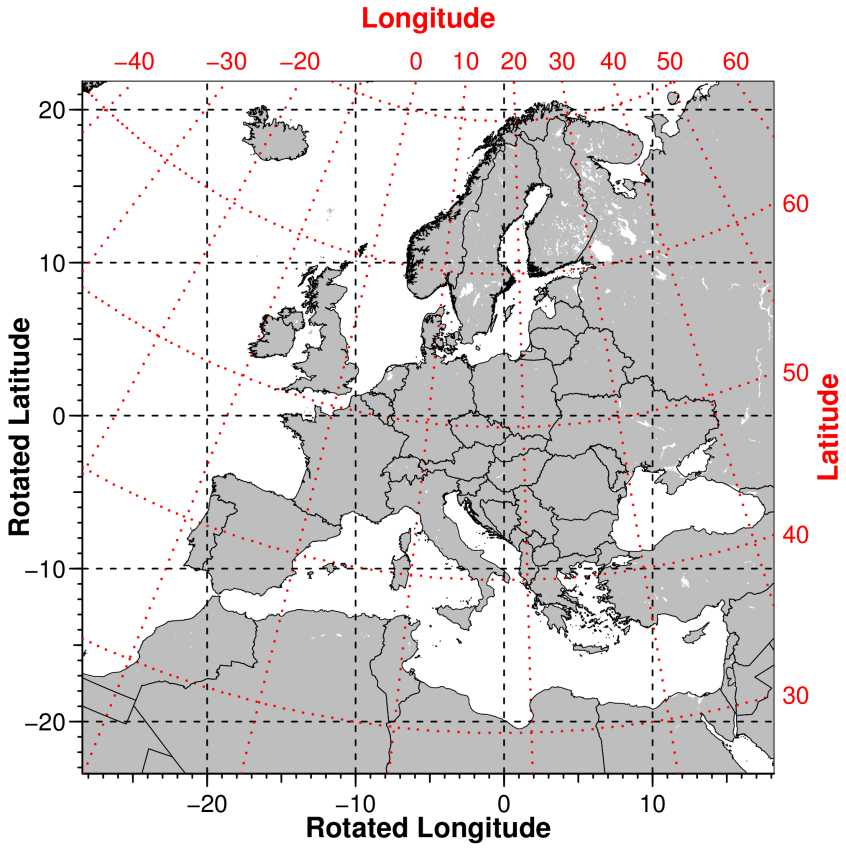


Figure 2.1.: The model domain of COSMO-REA6 in rotated coordinates. Shown in red are the unrotated spherical coordinates.

thereby making z the independent vertical coordinate instead of r. The approximation (2.44) implies that all spherical surfaces of constant vertical coordinate z have the same curvature. This results in two important simplifications:

- A number of metrical accelerations appearing in the equations of motion have to be neglected
- Due to vertical motion and thus its effect on the vertical motion itself the Coriolis effect must be neglected, resulting in a simpler form of the Coriolis acceleration

These simplifications are know as the first metric and the coordinate simplification and further details can be found in Zdunkowski and Bott, 2003. Applying these simplifications, the orthogonal base vectors $\vec{q}_i$ and the Jacobian of the transformation $\sqrt{G^s}$ of the rotated (λ, φ, z) coordinate system become

$$\begin{aligned} \vec{q}_1 &= a\cos\varphi\vec{e}_\lambda \\ \vec{q}_2 &= a\vec{e}_\varphi \qquad (2.45) \\ \vec{q}_3 &= \vec{e}_z = \vec{r}/r \\ \sqrt{G^s} &= a^2\cos\varphi\,, \qquad (2.46) \end{aligned}$$

with $\vec{e}_\lambda$, $\vec{e}_\varphi$ and $\vec{e}_z$ the normalized unit vectors in λ, φ and z direction. $\sqrt{G^s}$ is related to the metric tensor $\mathcal{G}^s$ of the spherical coordinate system via

$$\sqrt{G^s} = \sqrt{\mathcal{G}^s}\,. \tag{2.47}$$

The elements $\mathcal{G}^s_{ij}$ are given by the scalar product of the base vectors, $\mathcal{G}^s_{ij} = \vec{q}_i \cdot \vec{q}_j$. Thus, $\mathcal{G}^s$ is given by

$$\mathcal{G}^s = \begin{pmatrix} a^2\cos^2\varphi & 0 & 0 \\ 0 & a^2 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.48}$$

With the help of the elements $\mathcal{G}_{ii}$, any vector $\vec{A}$ with elements A^i can be written as

$$\vec{A} = \sum_{i=1}^{3} A^N \vec{q}_n = \sqrt{\mathcal{G}^s_{11}} A^1 \vec{e}_\lambda + \sqrt{\mathcal{G}^s_{22}} A^2 \vec{e}_\varphi + \sqrt{\mathcal{G}^s_{33}} A^3\,. \tag{2.49}$$

The three components of the wind vector $\vec{v}$, u for the zonal wind velocity, v for the meridional wind velocity and w for the vertical wind velocity then read

$$u = a\cos\varphi\dot{\lambda},\; v = a\dot{\varphi},\; w = \dot{z} = \dot{r}\,. \tag{2.50}$$

The nabla operator with these simplifications reads

$$\vec{\nabla} = \frac{\vec{e}_\lambda}{a\cos\varphi}\frac{\partial}{\partial\lambda} + \frac{\vec{e}_\varphi}{a}\frac{\partial}{\partial\varphi} + \vec{e}_z\frac{\partial}{\partial z} \tag{2.51}$$

and the basic equations (2.42) become

$$\begin{aligned}
\frac{\partial u}{\partial t} + \vec{v}\cdot\vec{\nabla}u - \frac{uv}{a}\tan\varphi - fv &= -\frac{1}{\rho a\cos\varphi}\frac{\partial p}{\partial\lambda} + M_u \\
\frac{\partial v}{\partial t} + \vec{v}\cdot\vec{\nabla}v - \frac{u^2}{a}\tan\varphi + fu &= -\frac{1}{\rho a}\frac{\partial p}{\partial\varphi} + M_v \\
\frac{\partial w}{\partial t} + \vec{v}\cdot\vec{\nabla}w &= -\frac{1}{\rho}\frac{\partial p}{\partial z} - g + M_w \\
\frac{\partial p}{\partial t} + \vec{v}\cdot\vec{\nabla}p &= -(c_{pd}/c_{vd})pD + (c_{pd}/c_{vd} - 1)\rho c_{pd} Q_T \\
\frac{\partial T}{\partial t} + \vec{v}\cdot\vec{\nabla}T &= \frac{1}{\rho c_{pd}}\left(\frac{\partial p}{\partial t} + \vec{v}\cdot\vec{\nabla}p\right) + Q_T \\
\frac{\partial q^v}{\partial t} + \vec{v}\cdot\vec{\nabla}q^v &= -(S^l + S^f) + M_{q^v} \\
\frac{\partial q^{l,f}}{\partial t} + \vec{v}\cdot\vec{\nabla}q^{l,f} - \frac{1}{\rho}\frac{\partial P_{l,f}}{\partial z} &= S^{l,f} + M_{q^{l,f}} \\
\rho &= p\left[R_d(1 + (R_v/R_d - 1)q^v - q^l - q^f)T\right]^{-1}.
\end{aligned} \tag{2.52}$$

Here the advection operator reads

$$\vec{v}\cdot\vec{\nabla} = \frac{1}{a\cos\varphi}\left(u\frac{\partial}{\partial\lambda} + v\cos\varphi\frac{\partial}{\partial\varphi}\right) + w\frac{\partial}{\partial z} \tag{2.53}$$

and D is defined as the three-dimensional wind divergence

$$D \equiv \vec{\nabla} \cdot \vec{v} = \frac{1}{a\cos\varphi}\left(\frac{\partial u}{\partial \lambda} + \frac{\partial}{\partial \varphi}(v\cos\varphi)\right) + \frac{\partial w}{\partial z}. \tag{2.54}$$

The following abbreviations have been introduced. The Coriolis parameter f now depends on the rotated (λ, φ)-coordinates and on the geographical latitude φ_g^N of the rotated north pole

$$f = 2\Omega\left(\cos\varphi_g^N \cos\varphi\cos\lambda + \sin\varphi\sin\varphi_g^N\right). \tag{2.55}$$

The different M-terms denote the source terms due to turbulent mixing, S^l and S^f represent the cloud microphysical sources and sinks per unit mass of moist air, P^l and P^f are again the precipitation fluxes and Q_T is the diabatic heating term in the heat equation

$$M_u = -\frac{1}{\rho}(\vec{\nabla}\cdot\underline{\mathbf{T}})\cdot\vec{e}_\lambda \tag{2.56}$$

$$M_v = -\frac{1}{\rho}(\vec{\nabla}\cdot\underline{\mathbf{T}})\cdot\vec{e}_\varphi \tag{2.57}$$

$$M_w = -\frac{1}{\rho}(\vec{\nabla}\cdot\underline{\mathbf{T}})\cdot\vec{e}_z \tag{2.58}$$

$$M_{q^x} = -\frac{1}{\rho}\vec{\nabla}\cdot\mathbf{F}^x \tag{2.59}$$

$$M_u = -\frac{1}{\rho c_{pd}}(\vec{\nabla}\cdot\mathbf{H} \tag{2.60}$$

$$S^{l,f} = \frac{1}{\rho} I^{l,f} \tag{2.61}$$

$$\vec{P}^{l,f} = -P_{l,f}\vec{e}_z = -\rho q^{l,f}|\vec{v}_T^{l,f}|\vec{e}_z \tag{2.62}$$

$$Q_T = \frac{1}{\rho c_{pd}}Q_h = \frac{L_V}{c_{pd}}S^l + \frac{L_S}{c_{pd}}S^f + M_T + Q_r \tag{2.63}$$

where Q_r describes the temperature change due to convergence/divergence of the solar and thermal electromagnetic radiation flux

$$Q_r = \frac{1}{\rho c_{pd}}\vec{\nabla}\cdot\mathbf{R}. \tag{2.64}$$

2.1.3. Model reference state

In the COSMO model, any thermodynamic variable is defined as the sum of a reference state and deviation from this reference state

$$\psi(\lambda,\varphi,z,t) = \psi_0(z) + \psi'(\lambda,\varphi,z,t). \tag{2.65}$$

The model reference state in the COSMO model is assumed to be horizontally homogeneous, time invariant, dry and at rest. From this it follows that

$$u_0(z) = v_0(z) = w_0(z) = 0, q_0^v(z) = q_0^l(z) = q_0^f(z) = 0 \tag{2.66}$$

and the model variables are then written as

$$\begin{aligned} u(\lambda,\varphi,z,t) &= u'(\lambda,\varphi,z,t) \\ v(\lambda,\varphi,z,t) &= v'(\lambda,\varphi,z,t) \\ w(\lambda,\varphi,z,t) &= w'(\lambda,\varphi,z,t) \\ T(\lambda,\varphi,z,t) &= T_0(z)+T'(\lambda,\varphi,z,t) \\ p(\lambda,\varphi,z,t) &= p_o(z)+p'(\lambda,\varphi,z,t) \\ \rho(\lambda,\varphi,z,t) &= \rho_0(z)+\rho'(\lambda,\varphi,z,t) \\ q^x(\lambda,\varphi,z,t) &= q^{x\prime}(\lambda,\varphi,z,t)\,. \end{aligned} \tag{2.67}$$

Additionally, the reference state is defined to be hydrostatically balanced, leading to

$$\begin{aligned} p_0 &= \rho_0 R_d T_0 \\ \frac{\partial p_0}{\partial z} &= -g\rho_0 = -\frac{g p_0}{R_d T0}\,. \end{aligned} \tag{2.68}$$

In the used version of the model a reference temperature profile is prescribed

$$T_0(z) = (T_{MSL}-\Delta t)+\Delta t \exp(\frac{-z}{h_{scal}}) \tag{2.69}$$

with T_{MSL} the reference temperature at mean sea level, Δt the temperature difference between sea level and the stratosphere and h_{scal} a scaling height. These are fixed to

$$T_{MSL} = 288.15K,\ \Delta T = 75K,\ h_{scal} = 10000m\,. \tag{2.70}$$

Because the reference pressure is predefined, the COSMO model uses the pressure perturbation p' as dependent prognostic model variable. Since the reference state of pressure is assumed to be horizontally homogeneous, the pressure gradient components become

$$\frac{\partial p}{\partial \lambda} = \frac{\partial p'}{\partial \lambda},\ \frac{\partial p}{\partial \varphi} = \frac{\partial p'}{\partial \varphi} \tag{2.71}$$

and the advection of pressure reduces to

$$\vec{v}\cdot\vec{\nabla}p = \vec{v}\cdot\vec{\nabla}p' - g\rho_0 w\,. \tag{2.72}$$

Lastly, the vertical acceleration due to the pressure gradient and the gravity becomes

$$-\frac{1}{\rho}\frac{\partial p}{\partial z} - g = -\frac{1}{\rho}\frac{\partial p'}{\partial z} + B \tag{2.73}$$

with B being the buoyant vertical acceleration, defined as

$$B = g\frac{\rho_0}{\rho}\left(\frac{T'}{T} - \frac{T_0 p'}{T p_0} + (\frac{R_v}{R_d}-1)q^v - q^l - q^f\right). \tag{2.74}$$

With these assumptions, the set of equations (2.53) becomes

$$\begin{aligned}
\frac{\partial u}{\partial t}+\vec{v}\cdot\vec{\nabla}u-\frac{uv}{a}\tan\varphi-fv &= -\frac{1}{\rho a\cos\varphi}\frac{\partial p}{\partial\lambda}+M_u\\
\frac{\partial v}{\partial t}+\vec{v}\cdot\vec{\nabla}v-\frac{u^2}{a}\tan\varphi+fu &= -\frac{1}{\rho a}\frac{\partial p}{\partial\varphi}+M_v\\
\frac{\partial w}{\partial t}+\vec{v}\cdot\vec{\nabla}w &= -\frac{1}{\rho}\frac{\partial p'}{\partial z}+B+M_w\\
\frac{\partial p'}{\partial t}+\vec{v}\cdot\vec{\nabla}p' &= -(c_{pd}/c_{vd})pD+(c_{pd}/c_{vd}-1)\rho c_{pd}Q_T\\
\frac{\partial T}{\partial t}+\vec{v}\cdot\vec{\nabla}T &= \frac{1}{\rho c_{pd}}\left(\frac{\partial p'}{\partial t}+\vec{v}\cdot\vec{\nabla}p'-g\rho_0 w\right)+Q_T\\
\frac{\partial q^v}{\partial t}+\vec{v}\cdot\vec{\nabla}q^v &= -(S^l+S^f)+M_{q^v}\\
\frac{\partial q^{l,f}}{\partial t}+\vec{v}\cdot\vec{\nabla}q^{l,f}-\frac{1}{\rho}\frac{\partial P_{l,f}}{\partial z} &= S^{l,f}+M_{q^{l,f}}\\
\rho &= p\left[R_d(1+(R_v/R_d-1)q^v-q^l-q^f)T\right]^{-1}.
\end{aligned} \tag{2.75}$$

2.1.4. Terrain-following coordinates

The model equations derived so far are valid on a rotated lat/lon-grid with the height z above mean sea level. Keeping this vertical coordinate will cause numerical problems since the model layers would cut through the orography in the lower levels and will not be defined there. To avoid such problems terrain-following coordinates are introduced in which the lowest model layer is identical to the orography and the higher model layers keep the silhouette of the orography but are damped up to a certain level. The calculus of this transformation is not reproduced here but it is referred to Doms, 2011 for extensive details on the transformation. The COSMO model uses a height-based hybrid terrain-following coordinate as vertical coordinate, which is very similar to the Gal-Chen vertical coordinate proposed by Gal-Chen and Somerville, 1975. The vertical coordinate is depicted as μ. Introducing the vertical coordinate to the model equations (2.76) and following the derivation of Doms, 2011, the final set of model equations reads as follows.

- *Zonal wind velocity*

$$\frac{\partial u}{\partial t}=-\left(\frac{1}{a\cos\varphi}\frac{\partial E_h}{\partial\lambda}-vV_a\right)-\dot{\mu}\frac{\partial u}{\partial\mu}-\frac{1}{\rho a\cos\varphi}\left(\frac{\partial p'}{\partial\lambda}-\frac{1}{\sqrt{\gamma}}\frac{\partial p_0}{\partial\lambda}\frac{\partial p'}{\partial\mu}\right) \tag{2.76}$$

- *Meridional wind velocity*

$$\frac{\partial v}{\partial t}=-\left(\frac{1}{a}\frac{\partial E_h}{\partial\varphi}+uV_a\right)-\dot{\mu}\frac{\partial v}{\partial\mu}-\frac{1}{\rho a}\left(\frac{\partial p'}{\partial\varphi}-\frac{1}{\sqrt{\gamma}}\frac{\partial p_0}{\partial\varphi}\frac{\partial p'}{\partial\mu}\right) \tag{2.77}$$

- *Vertical wind velocity*

$$\begin{aligned}
\frac{\partial w}{\partial t} &= -\left[\frac{1}{a\cos\varphi}\left(u\frac{\partial w}{\partial\lambda}+v\cos\varphi\frac{\partial w}{\partial\varphi}\right)\right]-\dot{\mu}\frac{\partial w}{\partial\mu}+\frac{g}{\sqrt{\gamma}}\frac{\rho_0}{\rho}\frac{\partial p'}{\partial\mu}+M_w\\
&\quad +g\frac{\rho_0}{\rho}\left[\frac{(T-T_0)}{T}-\frac{T_0p'}{Tp_0}+\left(\frac{R_d}{R_v}-1\right)q^v-q^l-q^f\right]
\end{aligned} \tag{2.78}$$

- *Pressure perturbation*

$$\frac{\partial p'}{\partial t} = -\left[\frac{1}{a\cos\varphi}\left(u\frac{\partial p'}{\partial \lambda} + v\cos\varphi\frac{\partial p'}{\partial \varphi}\right)\right] - \dot{\mu}\frac{\partial p'}{\partial \mu} + g\rho_0 w - \frac{c_{pd}}{c_{vd}}pD \tag{2.79}$$

- *Temperature*

$$\frac{\partial T}{\partial t} = -\left[\frac{1}{a\cos\varphi}\left(u\frac{\partial T}{\partial \lambda} + v\cos\varphi\frac{\partial T}{\partial \varphi}\right)\right] - \dot{\mu}\frac{\partial T}{\partial \mu} - \frac{1}{\rho c_{vd}}pD + Q_T \tag{2.80}$$

- *Water vapour*

$$\frac{\partial q^v}{\partial t} = -\left[\frac{1}{a\cos\varphi}\left(u\frac{\partial q^v}{\partial \lambda} + v\cos\varphi\frac{\partial q^v}{\partial \varphi}\right)\right] - \dot{\mu}\frac{\partial q^v}{\partial \mu} - \left(S^l + S^f\right) + M_{q^v} \tag{2.81}$$

- *Liquid and solid forms of water*

$$\begin{aligned}\frac{\partial q^{l,f}}{\partial t} &= -\left[\frac{1}{a\cos\varphi}\left(u\frac{\partial q^{l,f}}{\partial \lambda} + v\cos\varphi\frac{\partial q^{l,f}}{\partial \varphi}\right)\right] - \dot{\mu}\frac{\partial q^{l,f}}{\partial \mu} \\ &\quad -\frac{g}{\sqrt{\gamma}}\frac{\rho_0}{\rho}\frac{\partial P_{l,f}}{\partial \mu} + S^{l,f} + M_{q^{l,f}}\end{aligned} \tag{2.82}$$

- *Total density of air*

$$\rho = p\left[R_d\left(1 + \left(R_v/R_d - 1\right)q^v - q^l - q^f\right)T\right]^{-1} \tag{2.83}$$

Some additional terms are added here. E_h and V_a describe the kinetic energy of horizontal motion and the vertical component of absolute vorticity, respectively

$$E_h = \frac{1}{2}\left(u^2 + v^2\right) \tag{2.84}$$

$$V_a = \frac{1}{a\cos\varphi}\left[\frac{\partial v}{\partial \lambda} - \frac{\partial}{\partial \varphi}(u\cos\varphi)\right] + f \tag{2.85}$$

The abbreviation $\sqrt{\gamma}$ stands for the variation of reference pressure with the vertical coordinate μ

$$\sqrt{\gamma} = \frac{\partial p_0}{\partial \mu}. \tag{2.86}$$

The contravariant vertical velocity $\dot{\mu}$ and the divergence of the wind field D are defined as

$$\dot{\mu} = -\frac{1}{\sqrt{\gamma}}\left(\frac{u}{a\cos\varphi}\frac{\partial p_0}{\partial \lambda} + \frac{v}{a}\frac{\partial p_0}{\partial \varphi} + g\rho_0 w\right) \tag{2.87}$$

$$D = \frac{1}{a\cos\varphi}\left[\frac{\partial u}{\partial \lambda} - \frac{1}{\sqrt{\gamma}}\frac{\partial p_0}{\partial \lambda}\frac{\partial u}{\partial \mu} + \frac{\partial}{\partial \varphi}(v\cos\varphi) - \frac{\cos\varphi}{\sqrt{\gamma}}\frac{\partial p_0}{\partial \varphi}\frac{\partial v}{\partial \mu}\right] - \frac{g\rho_0}{\sqrt{\gamma}}\frac{\partial w}{\partial \mu}. \tag{2.88}$$

The equations (2.76) - (2.83) form a complete set of prognostic equations to predict the model variables u, v, w, T, p', ρ, q^v, q^l and q^f. In order to solve them, the various mixing terms M_ψ, the cloud microphysical source and sink terms S^l and S^f, the associated precipitation fluxes P^l and P^f as well as the radiative heating term Q_r, which is a part of the total diabatic heating Q_T in (2.63), need to be known. Since these terms describe mainly subgrid-scale processes, they need to be parametrized. The main points of these physical parametrization schemes are described in section 2.1.6

2.1.5. Model grid structure

The equations of motion are numerically solved using finite differences. In this way the model domain in (λ, φ, μ)-coordinates is approximated by a finite number of grid points in (i, j, k) where i, j and k correspond to the λ-, φ- and μ-directions, respectively. The model grid points are then defined by

$$\begin{aligned} \lambda_i &= \lambda_0 + (i-1)\Delta\lambda, \; i = 1, \ldots, N_\lambda \\ \varphi_j &= \varphi_0 + (j-1)\Delta\varphi, \; j = 1, \ldots, N_\varphi \\ \mu_k &= k, \; k = 1, \ldots, N_\mu \, . \end{aligned} \tag{2.89}$$

$\Delta\lambda$ and $\Delta\varphi$ are the grid spacings in the corresponding directions. The grid spacing in μ-direction is set to 1. N_λ, N_φ and N_μ are the number of grid points in the three dimensions. The points λ_0 and φ_0 define the southwest corner of the model domain in the rotated coordinate system.
Due to the use of finite differences the model domain is subdivided into a finite number of grid boxes with a volume $\Delta V = \Delta\lambda\Delta\varphi\Delta\mu$. Every grid point (i, j, k) then defines the centre of this grid box with the faces of this box located in the middle between the grid-points, i.e at $\lambda_{\pm 1/2}$, $\varphi_{\pm 1/2}$ and $\mu_{\pm 1/2}$. In vertical direction the k-levels are referred to as main levels with N_{ml} levels and the grid box faces as half levels with $N_{hl} = N_{ml} + 1$ levels. The model variables are staggered on an Arakawa-C-grid (Arakawa and Lamb, 1981; Arakawa and Lamb, 1977) where all scalar variables Ψ are defined in the grid centre at (i, j, k) whereas the components of the wind vector are defined at the box faces. Therefore the vertical velocity is always defined at the half levels. The zonal and meridional velocities are shifted by half a grid point to the east for the zonal component and to the north for the meridional component. Divergences can easily be computed at the grid box centres in this way. This basic concept of the grid structure is depicted in Fig. 2.2.

2.1.6. Physical parametrizations

The physical parametrizations accounting for the subgrid-scale processes are as follows. For the grid-scale precipitation a bulk-water continuity model is applied which computes the effects of the precipitation formation on temperature and on water vapour, cloud water and cloud ice. Precipitation is treated prognostically.
The radiative transfer scheme used in the COSMO model is described in Ritter and Geleyn, 1992. The scheme is based on the δ-two-stream solution of the radiative transfer equation for plane-parallel horizontally homogeneous atmospheres. The radiative transfer equation is solved in five spectral intervals in the thermal and for three spectral intervals in the solar part of the spectrum. As input to the radiation parametrization grid- and subgrid scale water clouds are considered. The radiation scheme is applied once every 15 minutes.
For the parametrization of subgrid-scale convection a Tiedtke mass flux scheme is used (Tiedtke, 1989). Within the scheme, the feedback of subgrid-scale vertical fluxes of heat, mass, moisture and momentum in up- and downdrafts is calculated by use of a bulk cloud model.
Another important parametrization is that of vertical turbulent transport. The parametrization scheme used in the COSMO model is based on prognostic turbulent kinetic energy and the second order moments of the basic equations. The scheme is formulated in terms of liquid water potential temperature and total water content and includes subgrid thermal inhomogeneities.

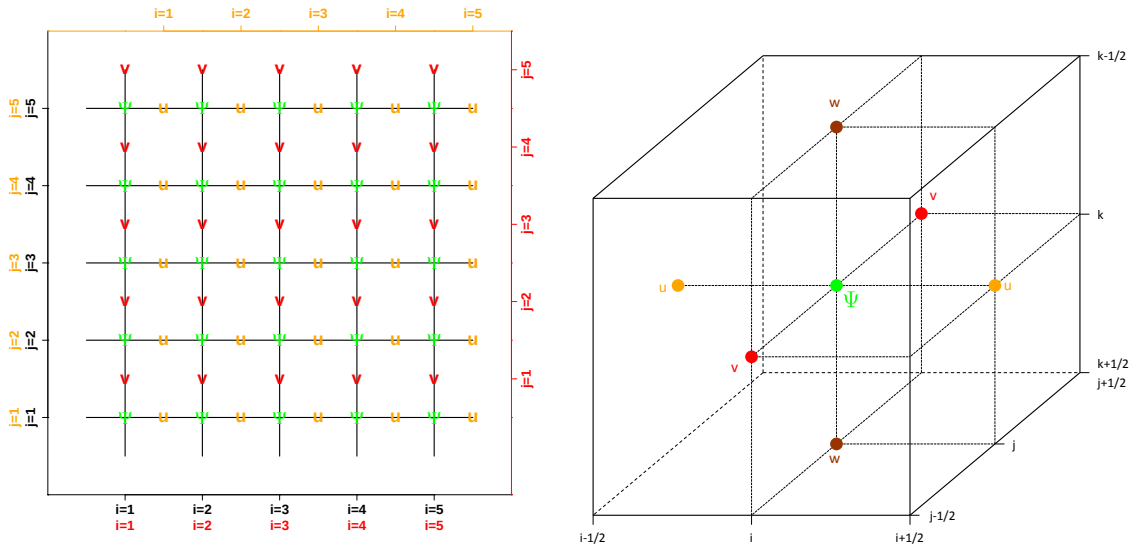


Figure 2.2.: Example of the staggered Arakawa-C-grid used in the COSMO-model. Left: The horizontal grid structure with the positions of the U-Wind component, the V-Wind component and the scalar variables Ψ. Right: The 3D-grid structure with the positions of the zonal wind component u, the meridional wind component v, the vertical wind component w and the scalar variables Ψ.

Extensive details on the used and optional parametrization schemes can be found in Doms *et al.*, 2011.

2.1.7. Data assimilation and surface analysis modules

The data assimilation in the reanalysis consists of an online and an offline part. The online part is formed by the data assimilation within the COSMO model, which is performed during the forward integration of the model. This part is explained in the following section 2.1.7.1. The offline part of the data assimilation scheme is comprised of the surface analyses modules, which are

- An analysis of the snow depth (see section 2.1.7.2).
- A sea surface temperature (SST) analysis (see section 2.1.7.3).
- A variational soil moisture analysis that uses 2-m temperature observations for the derivation of optimized soil moisture fields (see section 2.1.7.4).

2.1.7.1. General data assimilation method

The data assimilation within the COSMO model is based on the nudging technique or Newtonian relaxation which consists of relaxing the model´s prognostic variables towards prescribed, i.e. observed, values within a predetermined time window. Detailed descriptions are provided in Davies and Turner, 1977 and Stauffer and Seaman, 1990. A complete description on the implementation of the nudging in the COSMO model can be found in Schraff and Hess, 2003. The basic concept is as follows.

Let $\psi(x,t)$ be any prognostic variable. The nudging introduces a relaxation term, the so-called nudging term, into the tendency equation for the prognostic variable which then reads

$$\frac{\partial}{\partial t}\psi(\vec{x},t) = F(\psi,\vec{x},t) + G_\psi \cdot \sum_{k_{(obs)}} W_k(\vec{x},t)\cdot[\psi_k^{obs} - \psi(\vec{x}_k,t)]\,. \tag{2.90}$$

Here F denotes the model dynamics and physical parametrizations, G_ψ is the constant nudging-coefficient and W_k an observation dependent weight which always takes values between 0 and 1. ψ_k^{obs} is the value of the k^{th} observation influencing the grid point $\vec{x}$ at time t and x_k are the observation locations. The difference $[\psi_k^{obs} - \psi(x_k,t)]$ between observed and model value is called the observation increment and the complete last term of (2.90) is called the nudging term which determines the analysis increments. The analysis increments are the values added to the model fields by the nudging within one timestep.
If the physics and dynamics are neglected and a single observation with a weight W_k equal to 1 is assumed than the model value at the observation location relaxes exponentially towards the observed value with an e-folding decay rate of $\frac{1}{G_\psi}$. In this way the nudging equation (2.90) describes a continuous adaptation of the model towards the observed values during the forward integration of the model. Usually the nudging term remains smaller than the largest term of the dynamics. This is guaranteed by the nudging coefficient, which is in the range $10^{-4}s^{-1}$. When this weight is chosen too large, the nudging term could become overly dominant or an almost replacement of the model variable by the observed value could occur. The latter would destroy the internal balance of the model (Anthes, 1974). By controlling the nudging term in this way, the model fields are relaxed towards the observed values without significantly disturbing the dynamic balance of the model.
Every observation at a specific grid point is assigned a specific weight w_k and the factors W_k determine the relative weights to the observations at this grid point. For a single observation, this weight w_k computes as

$$w_k = w_t \cdot w_{xy} \cdot w_z \cdot \varepsilon_k \tag{2.91}$$

where w_t, w_{xy} and w_z are the weights depending on the temporal, horizontal and vertical difference between the observation and the target grid point, respectively, and ε_k holds information on the quality of the observation. The temporal weight for single observations is a hat function, linearly decreasing for -3 h and $+1$ h for radiosonde data and -1.5 h and $+0.5$ h for all other data. The relative weight W_k takes the effect of multiple observations into account and thus prevents the nudging term from becoming dominant over the dynamics.

$$W_k = \frac{w_k}{\sum_j w_j}\cdot w_k \tag{2.92}$$

There are of course disadvantages to the nudging in comparison to other widely used data assimilation schemes. For example, in contrast to optimal interpolation (OI) and 3- or 4-dimensional variational techniques (3D-VAR, 4D-VAR) there is no mathematical formalism to determine a theoretically optimal solution to the analysis problem. This results in several free parameters which cannot be determined in an optimal way using theoretical considerations but rather only rough estimates. Bergemann and Reich, 2012 have shown just recently that a mathematical formalism for continuous data assimilation in an ensemble context exists and could be used for ensemble runs but not in a deterministic run.

Furthermore the observation increments have to be expressed in model space rather than in observation space which is the biggest disadvantage compared to 3D-VAR or 4D-VAR. This means that for any observational information used by the nudging the observation increments have to be derived always in terms of the prognostic model variable. For example if radar reflectivities or satellite radiances are assimilated using 3D- or 4D-VAR, corresponding reflectivity or radiance values can be computed from model fields. These are then used to compute observation increments which can be used directly for the assimilation. In a nudging scheme, observation increments for the prognostic variables such as temperature, wind or humidity need to be derived from the reflectivities or radiances. This limits the choice of observational information to some extent which does not apply for variational methods.
Lastly, cross-correlations of observation and model errors such as error correlations between the wind and mass fields can not be taken into account unless additional balancing steps are added. In fact, there are three types of balancing steps applied to the analysis increment fields before they are added to the model fields. These are

- a hydrostatic upper-air temperature correction which balances the pressure analysis increments at the lowest model layer.
- a geostrophic wind correction which balances the wind field with respect to the mass field increments
- an upper-air pressure correction which balances the total analysis increments of the mass field hydrostatically

The whole nudging scheme consists of the following steps.
In the beginning, the observation processing is used to assign the observations temporally and spatially to the model space, exploit the quality flags, apply the bias corrections and finally to check for gross errors and redundancies. Afterwards the observation increments are computed and a quality control of the observations is applied. This step is followed by the computation of the weights and the spreading of increments provided with the weights to the target grid points for each observation. These weighted increments are then subject to the above mentioned balancing steps. The nudging scheme is finished by summing up the final weighted increments to form the analysis increments, i.e. the second term on the right hand side of Eq. (2.90), which are then added to the model equations.

2.1.7.2. Snow analysis

Since main surface parameters such as surface albedo, the turbulent surface fluxes of energy, moisture and momentum and the temperature at the surface depend on the presence of snow it is important to know the distribution of the snow cover over the model domain. To obtain the best possible distribution a snow depth analysis is performed in a separate module and afterwards transformed into the model variable snow water content for the use in COSMO. For extensive details on the snow analysis it is referred to Schraff and Hess, 2003. This module will be referred to as Snow Analysis.
SYNOP reports form the basic source of information for the snow analysis. Only in data-poor regions where the weight of the SYNOP observations falls beneath a given threshold COSMO model values of snow depth are needed as a background field. From SYNOP reports, the total

snow depth is extracted. If this information is missing, 6-hourly precipitation sum is extracted. In this case, the 2-m temperature T_{2m} is also needed and extracted from the SYNOP report or, if missing, is taken from the model. If T_{2m} is below 0°C and the present and past weather observations indicate snowfall, then the precipitation sum is converted into snow depth.
The snow depth determined in this way is then subject to a plausibility check. The snow depth is rejected if it exceeds an acceptance limit $d_{al} = 1.5[m] \cdot (1 + z_{sh}/800[m])$ which depends on the station height z_{sh}. After this, a quality control check is performed. The snow depth observation is rejected if it deviates from a first guess by more than a threshold value d_{thr}

$$d_{thr} = 0.8[m] \cdot \left(1 + \frac{z_{sh}}{2000[m]}\right) \cdot \max\left(0, \min\left(1, \frac{287.16[K] - T_{2m}}{10[K]}\right)\right). \tag{2.93}$$

As first guess, the previous snow depth analysis is used.
The analysis method is based upon a simple weighted averaging of observed values. A background field is only used in data-poor regions. The individual weight w_k of an observation k at a target grid point depends on the horizontal and vertical distance Δh and Δz between the observed location and the target grid point in the following way

$$w_k = \max\left(\frac{r_h^2 - \Delta h^2}{r_h^2 + \Delta h^2}, 0\right) \cdot \max\left(\frac{r_z^2 - \Delta z^2}{r_z^2 + \Delta z^2}, 0\right) \tag{2.94}$$

In this way, the weight becomes zero if the distance between the observation location and the target grid point is larger than the radii of influence r_h and r_z. The horizontal radius of influence r_h is set to 120 (200) km for data-dense (data-poor) areas whereas the vertical radius of influence r_z depends on the height z of the target grid point $r_z = 0.4 \cdot z + 180$ m. The weighted averages of snow depth and of snow depth increments along with the total weights then form the basis for the analysed snow depth. A final check is performed to ensure that the analysis increment at any grid point does not exceed a height and temperature dependent limit.

2.1.7.3. Sea surface temperature analysis

The sensible and latent energy fluxes as well as the development of cyclones over the ocean strongly depend on the temperature of the sea surface. It is therefore important to specify the sea surface temperature correctly. This specification is done in a separate module by analysing both the sea surface temperature and the location of the sea ice boundary two-dimensionally. This module will be referred to as SST-Analysis. The analysis method is outlined below and is described in Schraff and Hess, 2003.
At first the sea ice cover in the Baltic Sea is analysed. For this purpose an external weekly analysis from the Federal Maritime and Hydrographic Agency of Germany (BSH) with a resolution of $0.1\overline{6}$ degrees in longitudinal and 0.1 degrees in latitudinal direction is directly interpolated onto the model grid.
The sea surface temperature is then analysed by means of a correction scheme. In the vicinity of the observations, weighted observation increments are added to a first guess field. For the latter, the interpolated SST analysis of ERA-Interim is used. Additionally, observational data comprising all the ship and buoy data from the previous five days are used. These data are checked against the first guess and against other stations in the near vicinity. The first guess value is then corrected by a weighted mean of all the observation increments at each grid point, forming the analysis. The individual weights depend on the temporal distance between

analysis and observation time, on the observation type, and on the spatial distance of the observation location and the target grid point as in the Snow Analysis.

2.1.7.4. Soil Moisture Analysis

The soil water content has a large influence on the near surface values of temperature and relative humidity, especially on clear-sky days. Additionally, the soil water content serves as a memory in the reanalysis in the way that precipitation from days ago is kept in the soil and influences, e.g. the sensible and latent energy fluxes. Due to a lack of direct measurements of the soil moisture contents the analysis of soil moisture is done by a variational method using 2-m temperature observations which are not assimilated by the nudging. The optimal soil moisture contents minimize a cost functional that expresses the differences between modeled and observed surface temperature. The method was described by Hess, 2001 and Bouyssel *et al.*, 1999 and applications can be found in Callies *et al.*, 1998, Rhodin *et al.*, 1999 and Mahfouf, 1991. The following remarks on the method, as applied in COSMO as well as further details, can be found in Schraff and Hess, 2003.
2-m temperature observations close to noon are assimilated by the variational soil moisture analysis (SMA) since the soil-atmosphere coupling is strongest with high radiative impact. But this coupling is not always strong enough to derive sufficient information on the soil moisture contents to compute them in a reliable way. Therefore, a Kalman filter cycled analysis is applied that incorporates a background state along with background error estimates. The variational Kalman filter analysis scheme requires one additional forecast run for each of the three analysed soil moisture layers which makes it computationally expensive. Since no adequate alternative could be found at DWD this scheme is used in the reanalysis cycle as well.
The variational analysis scheme is based on the minimization of a cost functional. This minimization problem is generally a high-dimensional problem since the moisture contents of every grid column for every soil layer have to be retrieved. To reduce the high-dimensional minimization problem to a large series of low-dimensional minimizations, a horizontal decoupling of surface temperatures and the soil moisture contents is assumed. This assumption holds because surface temperatures and humidities are mainly vertically coupled to the soil moisture of the same grid column, at least for horizontal grid sizes of 7 km, as was tested by DWD (Schraff and Hess, 2003). The cost functional $\mathcal{J}$ to be minimized then reads

$$\mathcal{J}(\eta) = \mathcal{J}^o(\eta) + \mathcal{J}^b(\eta) \tag{2.95}$$

with the observation term

$$\mathcal{J}^o(\eta) = \frac{1}{2}\left(T^o - T(\eta)\right)^T \boldsymbol{R}^{-1}\left(T^o - T(\eta)\right) \tag{2.96}$$

and the background term

$$\mathcal{J}^b(\eta) = \frac{1}{2}(\eta - \eta^b)^T \boldsymbol{B}^{-1}(\eta - \eta^b) \tag{2.97}$$

Here, η and η^b denote the vectors of dimension n_{soil} containing the moisture contents of the analysed and background state soil layers, respectively. T^o and $T(\eta)$ are the vectors of dimension n_{obs} containing analysed and model forecasted values of 2-m temperature for specified observation times. The matrices $\boldsymbol{R}$ of dimension $(n_{obs} \times n_{obs})$ and $\boldsymbol{B}$ of dimension $(n_{soil} \times n_{soil})$ denote the observation error and background error covariance matrices, respectively, which

are both symmetric and positive definite.
To obtain the 2-m temperature values, a linearization around the background state is assumed. The dependency between the 2-m temperature and the soil moisture is nonlinear and rather complex in general, but a linearization provides good approximations, at least as long as the retrieved values do not differ too much from the background state. The linearization reads

$$T(\eta) = T(\eta^b) + \mathbf{\Gamma}(\eta - \eta^b) \tag{2.98}$$

with $\mathbf{\Gamma}$ the Jacobian of dimension $(n_{obs} \times n_{soil})$ which is approximated by

$$\Gamma_{i,j} = \min\left(\frac{T_i(\eta^j) - T_i(\eta^b)}{n_j^j - n_j^b}, 0\right) \tag{2.99}$$

with $i = 1,\ldots,n_{obs}$ and $j = 1,\ldots,n_{soil}$. To obtain the approximation (2.99) when COSMO is run in forecast mode, as it is done operational at DWD, the routine forecast based on the background moisture contents η^b is needed as well as n_{soil} additional forecast runs with varied soil moisture contents η^j. There is no routine forecast in reanalysis mode, so in order to perform the SMA the routine forecast has to be computed additionally. This is the reason why the SMA is very expensive in computational resources and very time consuming.
The vector components of η^j are set to

$$\eta_k^j = \begin{cases} n_j^j \text{ for } k = j \\ n_j^b \text{ for } k \neq j \end{cases}, \; k = 1,\ldots,n_{soil} \tag{2.100}$$

where the varied soil moisture content η_j^j is altered depending on air dryness point (ADP) and field capacity (FC) of the soil model. This alteration is done in the actual call of the SMA program in the SMA module and is done to reduce the influence of the soil type of the actual horizontal grid point. Using the linearization (2.98) in the cost function (2.95) and computing the gradient results in

$$\nabla\mathcal{J}(\eta^b) = -\mathbf{\Gamma}^T\boldsymbol{R}^{-1}\left(T^o - T(\eta^b) - \mathbf{\Gamma}(\eta - \eta^b)\right) + \boldsymbol{B}^{-1}(\eta - \eta^b). \tag{2.101}$$

The minimization of $\mathcal{J}$ results in the analysed soil moistures η^a

$$\mathcal{J}(\eta^a) \leq \mathcal{J}(\eta) \; \forall \; \eta \neq \eta^a \tag{2.102}$$

which can be obtained by

$$\nabla\mathcal{J}(\eta^a) \stackrel{!}{=} 0. \tag{2.103}$$

Some calculation results in

$$\eta^a = \eta^b + \left(\mathbf{\Gamma}^T\boldsymbol{R}^{-1}\mathbf{\Gamma} + \boldsymbol{B}^{-1}\right)^{-1}\mathbf{\Gamma}^T\boldsymbol{R}^{-1}\left(T^o - T(\eta^b)\right). \tag{2.104}$$

This equation (2.104) is actually computed in the SMA program to obtain η^a.
For the start of the cycled soil moisture analysis scheme the background error covariance matrix $\boldsymbol{B}$ is initialized with estimated error variances and covariances of first guess moisture fields that are used as initial background η^b. The background state $(\eta^b)^{t+1}$ and the background error covariance matrix $(\boldsymbol{B})^{t+1}$ for the following day are provided in a Kalman filter cycled

analysis in the following way.
The background state $(\eta^b)^{t+1}$ for the following day is computed as

$$(\eta^b)^{t+1} = (\eta^a)^t + \left(M_t^{t+1}((\eta^b)^t) - (\eta^b)^t\right) \tag{2.105}$$

where $M_t^{t+1}((\eta^b)^t)$ are the soil moisture values of at 00 UTC of the following day, obtained from the 24 h routine forecast started at 00 UTC.
The analysis error covariance matrix $\boldsymbol{A}$ of dimension $(n_{soil} \times n_{soil})$ represents the confidence in the retrieved values $(\eta^a)^t$

$$(\boldsymbol{A})^t = \left(\nabla\mathcal{J}\right)^{-1} = \left(\boldsymbol{\Gamma}^T\boldsymbol{R}^{-1}\boldsymbol{\Gamma} + ((\boldsymbol{B})^t)^{-1}\right)^{-1} \tag{2.106}$$

and is given by the Hessian of $\mathcal{J}$ (Tarantola, 2005). The new background error covariance matrix $(\boldsymbol{B})^{t+1}$ can then be computed by

$$(\boldsymbol{B})^{t+1} = \boldsymbol{\mathcal{M}}(\boldsymbol{A})^t\boldsymbol{\mathcal{M}}^T + \boldsymbol{Q} \tag{2.107}$$

with $\boldsymbol{\mathcal{M}}$ an estimation of the tangent linear of the forecast operator M_t^{t+1} and $\boldsymbol{Q}$ the assumed error matrix of $\boldsymbol{\mathcal{M}}$.
The actual execution of the SMA is schematically shown in Fig. 2.3. At first, the 24 hour "routine" forecast, here depicted as first forecast, is performed. The first forecast starts at 00 UTC of the previous day with the reanalysis fields as input. From this run the soil moisture fields [W_SO] at 00 UTC from the previous and starting day are needed as well as the 2-m temperature [T2M] at 12 and 15 UTC of the previous day. The analysed soil moisture from 00 UTC and 2-m temperature from 12 and 15 UTC of the previous day together with the 2-m temperature from 12 and 15 UTC of the previous day from the first forecast run are then passed to the SMA program where a new, perturbed background state for the soil moisture is calculated. This is done twice with two different perturbations, resulting in two different soil moisture contents, marked with status 21 and 23. These are passed to the second respectively third forecast run which run for 15 hours and give new 2-m temperatures again from 12 and 15 UTC of the previous day. The final execution of the SMA program (status 27), where the new soil moisture contents for 00 UTC of the actual day are computed needs the following parameters:

- The analysed soil moisture content at 00 UTC from the previous day
- The analysed soil temperatures at 00 UTC from the previous and actual day
- The analysed 2-m temperature at 12 and 15 UTC from the previous day
- The 2-m temperature at 12 and 15 UTC from the previous day from the first forecast
- The 2-m temperature at 12 and 15 UTC from the previous day from the second forecast
- The 2-m temperature at 12 and 15 UTC from the previous day from the third forecast
- The soil moisture contents at 00 UTC from the previous and actual day from the first forecast
- The perturbed soil moisture content from the first call to the SMA program (status 21)
- The perturbed soil moisture content from the second call to the SMA program (status 23)

The resulting new soil moisture contents are then used to start the new COSMO run.

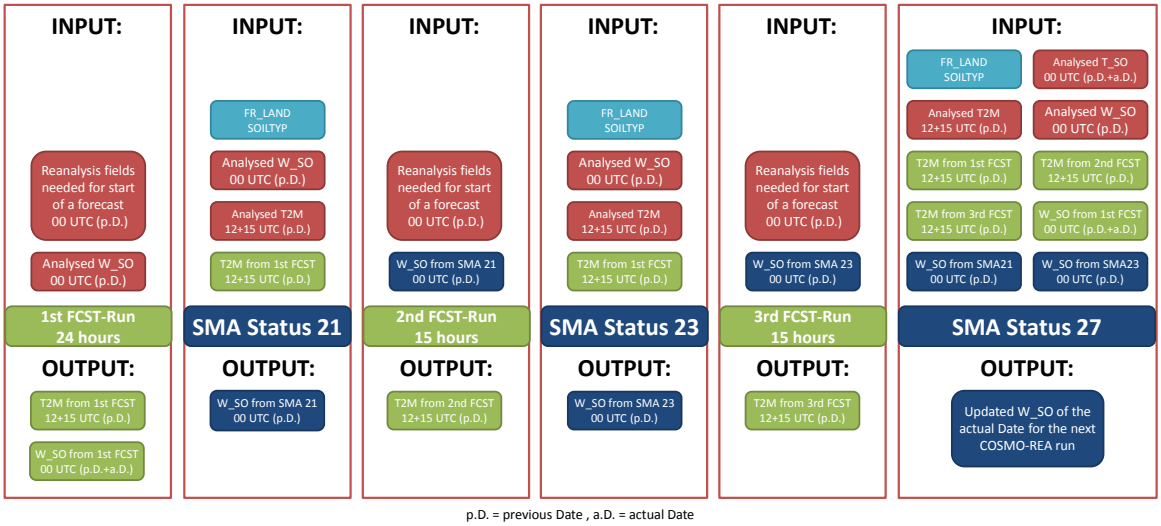


Figure 2.3.: A schematic overview of the complete SMA module. For details see text.

2.1.8. Assimilation of precipitation data

Precipitation is one of the most important quantities in the atmosphere. It is therefore of importance to have a realistic analysis of precipitation within the model, since the right amount of precipitation influences the soil moisture among many other parameters, which then for example influences the triggering of convection.
Precipitation is measured in many different ways, e.g. with rain gauges or with radar and satellites. This information can be passed to the model via the data assimilation in different ways. The different data assimilation techniques have been introduced extensively in section 2.1.7. Since the COSMO model uses the nudging the focus will be on the assimilation of precipitation via the nudging technique which can be assimilated in two different ways: via specific/relative humidity or via temperature increments. Davolio and Buzzi, 2004 modified the specific humidity profiles of their model such that they improve the accumulated rainfall over a time period. Another approach is the introduction of temperature increments into the model. The idea behind this approach is that due to condensation of water vapour in the clouds the vertically integrated latent heating rate should be approximately proportional to the net precipitation rate (Holton, 1979; Jones and Macpherson, 1997). This method is hence called latent heat nudging (LHN). The temperature increments for the LHN are usually obtained from radar or satellite measurements which have a high temporal resolution. Jones and Macpherson, 1997 applied the LHN with radar observations over the United Kingdom to the NWP model of the UK Meteorological Office (UK Met. Office). They have shown that the LHN improved the skill of the forecast as well as the analysis of precipitation.
The latent heat nudging is used in the COSMO-model as well. The method will be shortly outlined here, an extensive description can be found in Stephan *et al.*, 2008 and Schraff and Hess, 2003.
The LHN scheme introduces temperature increments into the thermodynamic equation in cases where the observed rain rate, taken from radar measurements, differs from the precipitation produced by the model. The radar network coverage, which is used by DWD, is shown with the COSMO-REA2 domain in Fig. 2.4. A scan of each radar is available every five minutes with

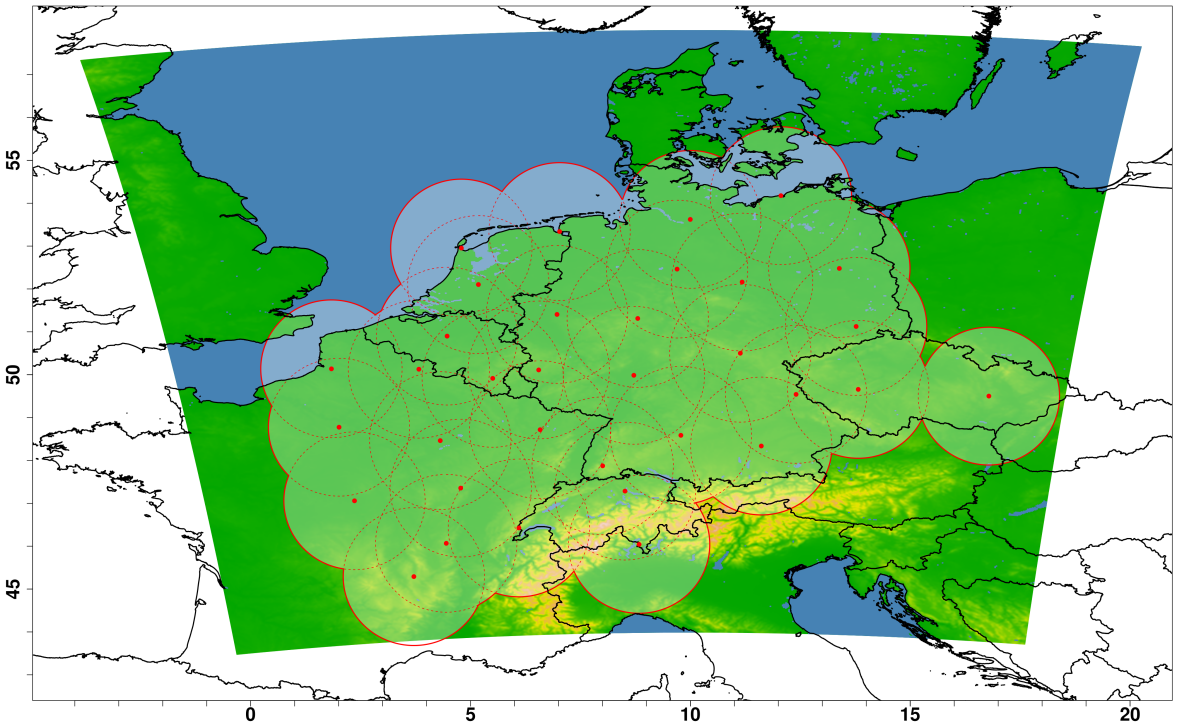


Figure 2.4.: The radar network of DWD and the surrounding national meteorological and hydrological services (MeteoFrance, MeteoSwiss, KNMI The Netherlands, KMI Belgium, Czech Hydrometeorological Institute), displayed on the COSMO-REA2 domain.

a spatial resolution of 1 km in range and 1 degree in azimuth. After applying a quality control, the reflectivity values are converted into precipitation rates by use of empirical reflectivity-rainfall, so called Z-R relationships which depend on the meteorological situation. As already mentioned above, the precipitation rate R can be assumed to be proportional to the latent heat release δLH within the cloud

$$R(l_0) \propto \int_{l_F}^{l_0} \Delta(LH(l))dl. \tag{2.108}$$

Here, l_0 is the ground, l_F is the height at which the precipitating particle is formed and l is the path between these two points. Usually, the path l is associated with the time the particle needs to reach the ground. During this time it can be transported within the cloud or advected during its fall into other regions, depending on the local wind field. However, the LHN scheme ignores this realistic but nevertheless complicated feature and assumes that the path l is run through within one model column and one time step, so that the precipitation rate is proportional to the latent heat release within the model column. The temperature increment due to the release of latent heat is added to the thermodynamic model equation and reads

$$\begin{aligned} \Delta T_{LHN}(l) &= (\alpha - 1) \cdot \frac{1}{c_p} \Delta(L_H(l)) \\ \alpha &= (\frac{R_{obs}}{R_{mod}}). \end{aligned} \tag{2.109}$$

This is the formulation of the conventional LHN scheme described by Jones and Macpherson, 1997 which they used in a model with a horizontal resolution of 17 km and a diagnostic computation of precipitation. Here, the scaling factor has upper and lower limits of 3 and $1/3$, respectively. However, the COSMO-DE configuration of DWD runs with a horizontal resolution of 2.8 km and COSMO-REA2 runs with an even higher horizontal resolution of 2 km as was described in section 2.2.1.3. Both models treat the precipitation as a prognostic variable. Therefore the LHN scheme implemented in the COSMO model works with some modifications compared to the conventional LHN scheme (see Stephan *et al.*, 2008):

- Due to the prognostic treatment the precipitation needs some time to reach the ground because in the prognostic formulation the parcel can be advected and transported after generation. The conventional scheme notices the precipitation not directly after generation but with some delay and will thus continue to add or take away energy for some time although it may not be required any longer. This problem is circumvented by introducing a reference precipitation which has the immediate information on the already initialized precipitation rate. This reference precipitation is the vertically averaged precipitation flux

$$R_{ref} = \frac{1}{z_{top} - z_0} \int_{z_0}^{z_{top}} \left[\sum_i (\rho(z) q_i(z) v_i) \right] dz, \tag{2.110}$$

  with q_i the mass fraction of the different forms of precipitation (rain, snow or graupel) and v_i their corresponding sedimentation velocity. The top layer z_{top} is defined as the first layer from above in which the sum of the fluxes is higher than a pre-specified threshold of 0.1 mm h^{-1}. This reference pressure replaces R_{mod} in equation (2.109).

- Only the vertical model layers with a positive model latent heat release are used to compute and insert the LHN increments. This is done to avoid negative LHN temperature increments and therefore cooling where the precipitation rate should be increased. For example, this is the case in regions with strong downdrafts where high precipitation rates with weak or negative latent heat release can occur.

- Although the precipitation rates R_{mod} are low updraught regions with very high values of latent heat release $\Delta(L_H)$ often occur at the leading edge of convective cells. This results in high values of the scaling factor α leading to very high temperature increments ΔT_{LHN}. To avoid these high values, the limits for the scaling factor are changed from 3 to 2 for the upper and from $1/3$ to 0.5 for the lower limit. In addition, the scaling factor in equation (2.109) is replaced by the logarithmic formulation $\{lhn(\alpha) + 1\}$, which yields effective upper and lower scaling limits of 1.7 and 0.3, respectively.

The implemented LHN scheme was always evaluated with respect to the improvement of the following forecast. The assimilation of radar images with the scheme presented above showed a positive impact on the prediction of precipitation in the first hours of the forecast (Stephan *et al.*, 2008). In COSMO a so-called LHN-coefficient was introduced which controls how much temperature increment is added to the prognostic equation. In operational settings this coefficient is set to 1. For the production of the reanalysis, however, it is not of interest to improve the forecasts, but the analysis itself. Therefore experiments were carried out to test different LHN-coefficients for the improvement of the analysis (section 2.2.1.5).

2.2. The reanalysis framework

In the following, the overall setup of the reanalysis together with its different steps in the cycle will be explained.

2.2.1. Setup of the system

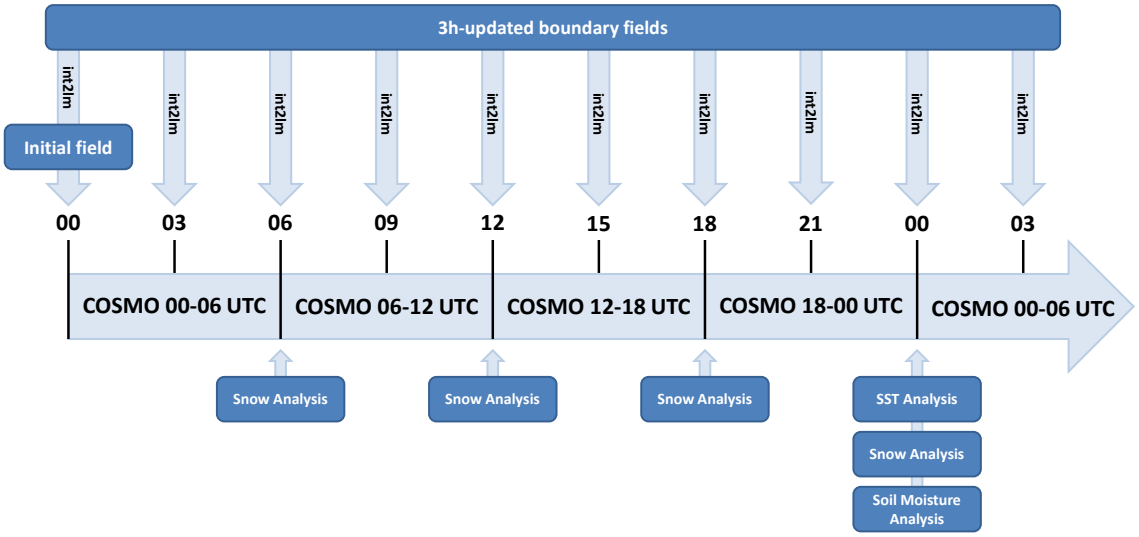


Figure 2.5.: The basic setup of the reanalysis system for the production of COSMO-REA6.

The reanalysis system which produces the reanalysis itself has been set up as a script-based system. The script-based system is written as a bash script to make it as flexible as possible. The bash-shell is usually provided on all major super-computing centres, making it suitable for a portable system. For convenience the script-based system will from now on be referred to as Reanalysis Production System, short RPS. The RPS is started on one of the login nodes of the super computing centres and from there submits the various jobs needed for the production of the reanalysis. Since the RPS is designed for the production of a reanalysis with the COSMO model it can only produce a reanalysis on a limited area. Therefore, the first step in the RPS is the interpolation of a coarser resolved model which is available on a larger grid than that of the reanalysis itself. The interpolation is done by the program *int2lm*, which is used by DWD to interpolate global or regional model outputs onto the COSMO grid. *int2lm* provides boundary fields in an update frequency which equals the availability of the coarse model data and an initial field for the start of a COSMO model run. From this initial field which is provided at 00 UTC of the start date the first six hour run of the COSMO model is started with the full data assimilation. Since DWD works with a Snow-Analysis every six hours, a six hour run has been chosen accordingly. When this model run is finished, the first Snow Analysis for 06 UTC is conducted and the updated fields are passed to the 06 UTC field for the next start of a six hour COSMO run. This continues till 00 UTC of the following day. At this point, a SST analysis is performed, followed by another Snow Analysis. The SMA is performed last, in the way as explained in section 2.1.7.4. The updated fields from these three separate procedures then

overwrite some of the output fields for 00 UTC. From this, the next day is started and so forth till the last production date. In order to speed up the production cycle, *int2lm* is carried out only every 24 hours, calculating the boundary fields for one day. Those fields are then linked for every 6 hour COSMO-run accordingly.
As mentioned above, the RPS submits the various jobs, extracts and post-processes the data and goes to sleep in between waiting for the jobs to finish. If the job ran successfully to the end, the next program is started, unless an error occurs. In such a case the program sends a e-mail indicating where the error occurred to notify the user where he/she needs to check and restart the system. The system has been tested and run successfully on the IBM Power-7 machine of the ECMWF and its successor, the CRAY XC-30, as well as on the IBM Power-6 machine called *blizzard* of the Deutsches Klimarechenzentrum (DKRZ). An overview of this basic setup for the case of the COSMO-REA6 reanalysis is shown in Figure 2.5.
Unfortunately, the flexibility of the bash-script based RPS is also its drawback. Since the system goes to sleep between the different actions of RPS (submitting jobs, post-processing, etc.) it is necessary that the RPS is started with a no-hangup call on the login node. In this way, the RPS runs as long as it takes the system to produce the complete reanalysis, i.e. several weeks to months. This became a problem, when different streams were started (see section 2.2.1.1) and operated in parallel. The login nodes at the major supercomputing centres are not made for operating several programs on such timescales, since this would cause too much traffic on the login nodes, in the worst case making it impossible for users to login. Therefore, when the new CRAY XC-30 machine was introduced at ECMWF in summer 2014, the RPS was migrated to the work flow package *ecflow*[2] provided by ECMWF which enables users to run a large number of programs in a large environment and is thus perfectly suited for the production of the reanalysis. The setup in *ecflow* follows in general the above depicted structure of the RPS, the technical details of the implementation are nevertheless outlined in Appendix C.

2.2.1.1. COSMO-REA6

The overall vision of the project was to create a high-resolution reanalysis for Europe and an even higher resolved one for Germany, following the setup of the Numerical Weather Prediction (NWP) system of DWD with their two domains for Europe and Germany, COSMO-EU and COSMO-DE, which are operated with the COSMO model. COSMO-EU is run with a horizontal grid resolution of 0.0625° which corresponds to roughly 7 km whereas COSMO-DE runs with a resolution of 0.025° (2.8 km). COSMO-EU is driven with boundary data from the global model GME of DWD and COSMO-DE is subsequently driven by COSMO-EU.
In recent years, regional climate modeling efforts were centralized within the CORDEX framework, which stands for COordinated Regional Downscaling EXperiment (Giorgi *et al.*, 2009). CORDEX provides global coordination of Regional Climate Downscaling for improved regional climate change adaptation and impact assessment. The framework defined several domains across the globe on which the modeling community should focus in their research. The European reanalysis created during this work was therefore chosen to match the specifications of the high-resolution CORDEX-EURO-11 domain. The CORDEX domains work on a rotated grid, just as the COSMO model does. The rotated north pole of the model domain lies at $\lambda_N = -162°$ and $\varphi_N = 39.25°$. The south-west corner in rotated coordinates lies at $\lambda_{SW} = -28.375°$ and $\varphi_{SW} = -23.375°$. The number of grid points in λ/φ-direction is 424/412 with a grid spacing of

[2] https://software.ecmwf.int/wiki/display/ECFLOW/What+is+ecFlow

$\Delta\lambda = \Delta\varphi = 0.11°$. These specifications were the basis for the reanalysis. The horizontal resolution of the produced reanalysis, however, was supposed to be more similar to the COSMO-EU resolution of 7 km rather than 12 km as in CORDEX-EURO-11. Therefore, the horizontal grid spacing was reduced by a factor of 2 to 0.055° to be closer to the COSMO-EU resolution. A resolution of 0.055° corresponds to a grid spacing of approximately 6.2 km. Due to this grid spacing, the created reanalysis for the European domain was named and is referred to in the following as COSMO-REA6.
This change in the model resolution leads to a slight modification in the model grid depicted in Figure 2.6. In both plots, the black dots and black dashed lines indicate the CORDEX-EURO-11

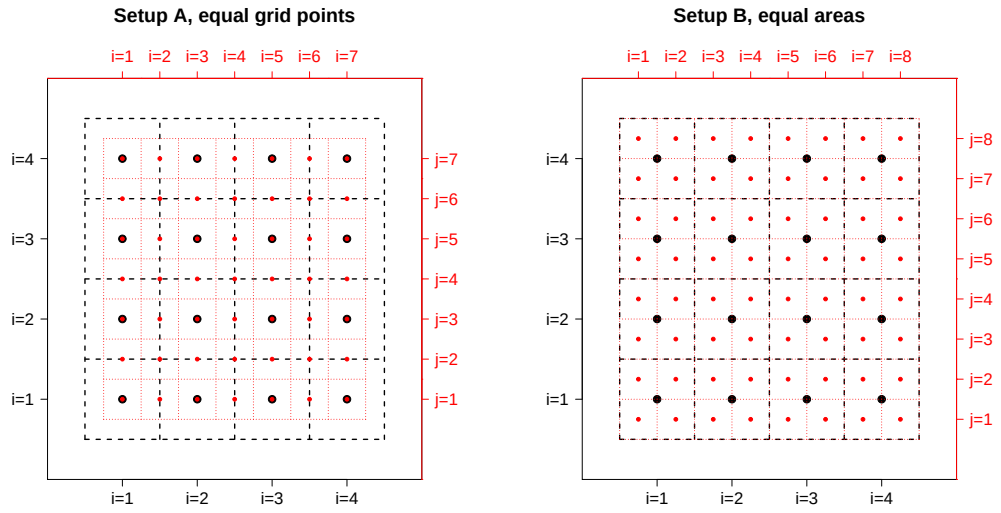


Figure 2.6.: Comparison between the grid matching of equal grid points and the matching of equal volumes.

grid points and grid box faces and the red dots and red dotted lines depict those of COSMO-REA6. If the reanalysis matched the CORDEX specifications exactly, this would result in the left plot, having the COSMO-REA6 grid points at the same locations as in CORDEX-EURO-11 and in the middle between two CORDEX points. But as a consequence of the smaller grid spacing, the grid boxes in COSMO-REA6 are representative for a smaller area than the grid boxes in CORDEX-EURO-11. This could lead to problems in comparing different model outputs. Therefore, the reanalysis model setup was chosen as shown in the right plot of Figure 2.6. By shifting the grid points of COSMO-REA6 half a grid point to the left and bottom and adding one grid box in each direction, four grid boxes in COSMO-REA6 can be combined to match one grid box in CORDEX-EURO-11.
Another problem to be adressed is the so-called “sponge” of a model. This is the region at the boundary of the model domain where the influence of the boundary fields is still noticeable. The sponge in COSMO-REA6 is comprised of approximately the first 10 to 15 grid points. These grid points are affected by spurious errors, resulting e.g. in very large precipitation fields being transported into the model domain. Therefore, the model domain was enhanced by 16 grid points in each direction in order to skip those grid points afterwards. In this way, the actual domain matching CORDEX-EURO-11 is not affected by the boundaries.
This two modifications result in a few differences of the setup. The south-west corner of COSMO-REA6 is now $\lambda_{SW} = -29.2825°$, $\varphi_{SW} = -24.2825°$ and the number of grid points is

Table 2.1.: Main parameters of the model domain in CORDEX-EURO-11 and COSMO-REA6.

	CORDEX-EURO-11	COSMO-REA6 computational model domain	COSMO-REA6 evaluated model domain
Coordinates rotated north pole $(\lambda,\varphi)_N$	$(-162.0, 39.25)$	$(-162.0, 39.25)$	$(-162.0, 39.25)$
Horizontal resolution	0.11°	0.055°	0.055°
Coordinates south-west corner $(\lambda,\varphi)_{SW}$	$(-28.375, -23.375)$	$(-29.2825, -24.2825)$	$(-28.4025, -23.4025)$
N_λ	424	880	848
N_φ	412	856	824
N_μ	n.a.	40	40

$N_\lambda = 880$ and $N_\varphi = 856$ for the computational domain. Subtracting the 16 grid points on each side of the computational domain results in the actual evaluated model domain with $N_\lambda = 848$, $N_\varphi = 824$ and the south-west corner at $(\lambda,\varphi)_{SW} = (-28.4025, -23.4025)$. These fixed parameters are summarized in table 2.1 and the model domain is shown in Figure 2.7. The model domain covers the whole of Europe with a large coverage of the northern Atlantic and an easterly coverage into western Russia. The northernmost parts of Africa are covered as well.
COSMO-REA6 is run with 40 layers in the vertical, resulting in 41 half levels which is the same vertical resolution as in COSMO-EU. The time step of the model runs is 50 seconds. The model version of COSMO for the production of COSMO-REA6 is 4.25 (28 September 2012).
The reanalysis was produced in different streams. The first stream consisted of the six years 2007-2012. In order to account for some spin-up of the model the production of this stream was started at the 1st November 2006. This stream was produced at the IBM Power-7 machine of ECMWF. In 2014 it was decided to extend the six years to the past and present. Therefore, 3 additional streams were started. The year 2013 was produced by simply extending the first stream. To the past, it was decided to extend the reanalysis back to 1997. In order to be finished with the production in a reasonable time, two parallel streams were set up, one producing the reanalysis for 1997-2001 and the other from 2002-2006. These streams were again started 2 months in advance, i.e. at the 1st November 1996 and 2001, respectively. Unfortunately, in 2014 ECMWF prepared the change from the IBM Power-7 machine to a new super computing cluster, the CRAY XC-30. The IBM Power-7 was switched off on the 30th September 2014. Therefore, the RPS was migrated to the CRAY and was implemented into ecflow in mid 2014 and the production of the three additional streams was carried out on the CRAY and was finished in October and November 2014 for the two historic streams. The stream for 2013 was extended into 2014 as soon as ERA-Interim boundary fields were available. 2014 was finished in late January 2015, extending the COSMO-REA6 data set to eighteen years (1997-2014).

2.2.1.2. COSMO-DS

In order to assess the added value of a full reanalysis an additional dynamical downscaling was performed for the year 2011. The dynamical downscaling was initialised at 00 UTC on the 1st January 2011 from the corresponding state of the full reanalysis and ran freely with the same setup and boundary conditions but without any data assimilation. For a correct specification of the lower boundary conditions the same 24h cycle of the SST analysis as in the full reanalysis was implemented in the downscaling. The other analysis modules were not

used in the downscaling. The dynamical downscaling will be called COSMO-DS hereafter and serves as a reference.

2.2.1.3. COSMO-REA2

For Germany an even higher resolved reanalysis was performed, called hereafter COSMO-REA2. The basic setup is that of a 2 km-Version of COSMO-DE, i.e. with a horizontal resolution of 0.018°. The rotated north pole is slightly different than for COSMO-REA6 with $\lambda_N = -170^\circ$ and $\varphi_N = 40^\circ$. The number of grid points in longitude and latitude direction is $N_\lambda = 724$, $N_\varphi = 780$ with the start longitude and latitude being $\lambda_{SW} = -7.5^\circ$, $\varphi_{SW} = -6.0^\circ$. In COSMO-REA2, there are 50 vertical levels and the time step is 18 seconds. The model domain covers Germany, Austria, Switzerland, Belgium, the Netherlands and Denmark as well as parts of their surrounding states such as Poland, Italy, France and Great Britain. The model domain of COSMO-REA2 is shown in Figure 2.7 as well, indicated by the yellow dotted quadrilateral. COSMO-REA2 was produced with the Snow and SST-Analysis, but without the SMA. However, in contrast to COSMO-REA6, the assimilation of precipitation information from radar via latent heat nudging was used (see Section 2.1.8). COSMO-REA2 was produced with the COSMO model version 5.00.2 (21st February 2014) for the years 2007-2013 on the CRAY, using *ecflow*.

2.2.1.4. Performance of the model

When running such a system, performance is an important issue. The first 6 years of the reanalysis to be computed were the years 2007-2012. These ran on the IBM-Power 7 machine of the ECMWF. Every node there consisted of 32 CPUs which could be run in hyperthread mode resulting in 64 logical CPUs. For the COSMO model runs, 10 nodes in hyperthread mode were used and the model domain was separated into 20 boxes in longitude and 32 in latitude direction. With this configuration one 6 hour run could be finished in ≈ 10 minutes. Due to the three additional forecast runs for the SMA this resulted in ≈ 90 minutes for one complete day of the reanalysis which additionally depended on the queueing times. Due to occasional errors and aborts and longer queueing times this first stream took about half a year to be finished.
Since ECMWF switched to the new CRAY XC-30 machine in early 2014, the production of the two other streams, 1997-2002 and 2002-2006, had to be switched to the new machine as well. On the CRAY one node consists of 24 CPUs, i.e. 48 logical CPUs in hyperthread mode. For the COSMO model runs 25 nodes in hyperthread mode on the larger and faster CRAY machine were used resulting in ≈ 75 minutes of production time for one day. COSMO-REA2 was completely produced on the CRAY with 25 nodes for each run. Due to the much smaller time step one run takes a while longer than one for COSMO-REA6 resulting in approximately 100 minutes for one day. One year of reanalysis took about a month to be produced.

2.2.1.5. Experiments with LHN

The assimilation of precipitation information from radars via the latent heat nudging is operationally used in COSMO-DE only with a fixed LHN-coefficient of 1. Some preliminary studies showed that a coefficient of 1.0 disturbed the stability of the model too much in the analysis, resulting in spurious, unrealistic cloud structures. Therefore, it was decided to reduce the coefficient to 0.25 which lead to good results without the drawbacks of a coefficient of 1 (not

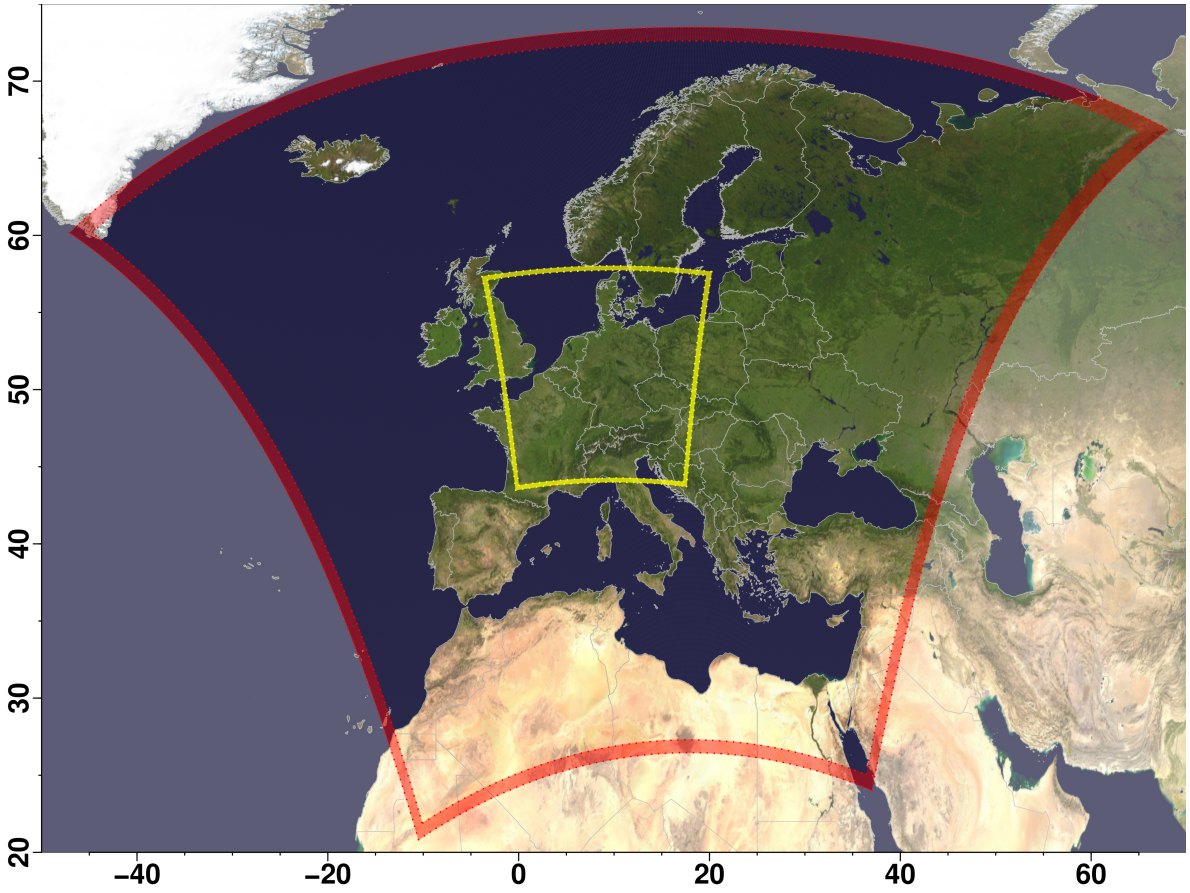


Figure 2.7.: Model domains of COSMO-REA6 and COSMO-REA2. The black dotted quadrilaterals indicate the computational and the actual model domain of COSMO-REA6, separated by the sponge layer which is indicated in red. The yellow quadrilaterals show the computational and actual model domain of COSMO-REA2 with the sponge layer indicated in yellow.

shown). Since the LHN improves the representation of precipitation on the 2 km scale it was decided to test the LHN also on the 6 km scale. Therefore, different experiments for the LHN were set up for summer 2011 (JJA). These were

- COSMO-REA6 without LHN (setup of the complete reanalysis)
- COSMO-REA6 with LHN enabled and a coefficient of 0.25
- COSMO-REA2 without LHN
- COSMO-REA2 with LHN enabled and a coefficient of 0.25 (setup of the complete reanalysis).

The results of the comparisons of the different experiments are shown in sections 4.2.1.1 and 4.2.2.1.

2.3. Data

In the following, an overview of all used and produced data of this work is given.

2.3.1. Observations for COSMO-REA6 and COSMO-REA2

In order to produce the reanalysis, observations for COSMO as well as for the additional analysis modules are needed. Those observation data were extracted from DWDs operational database where all observations needed for conducting COSMO are stored. The observation types together with the corresponding assimilated variables are summarized in Table 2.2. The soundings comprise radiosonde ascents (TEMP) and pilot balloon ascents (PILOT). Aircraft observations are divided in aircraft reports (AIREP), automatic reports of the type AMDAR (Aircraft Meteorological Data Relay) and reports from ACARS (Aircraft Communication Addressing and Reporting System). Surface level observations compose of the SYNOP stations, both manual and automatic reports as well as METAR (Meteorological Aerodrome Report), SHIP reports, both manual and automatic, and drifting buoys (DRIBU). No satellite observations were actively assimilated in COSMO-REA6.
As mentioned in Section 2.1.7.1, the observations for one model run started at time XX UTC are at least needed one hour before until three hours after the starting time, i.e. from XX-1 UTC till XX+3 UTC. Due to these time windows, there would have been many overlaps in data if the observations would have been obtained for every model run separately. Therefore, the observations for one complete day were obtained from DWDs database with the time window as mentioned above. E.g. for the 1st June 2011 the observations for the time window from 23 UTC at the 31st May until 03 UTC at the 2nd June were retrieved. These observations were then passed to the 4 COSMO runs at that day and COSMO then took only the observations relevant in the corresponding time. In this way, the overlap in the observations and therefore also storage space was reduced. In addition, the observation files remained still small enough and did not disturb the performance of the model, which would have been the case when observations for e.g. a whole month would have been stored into one file. The observations for the Snow and SST-Analysis were also retrieved from the database but for every start time separately.

2.3.2. ERA-Interim

Since every regional model covers only a limited area, frequently updated boundary data are needed in order to ensure the interaction with the surrounding environment. Lateral and lower boundary fields for COSMO-REA2 are provided by COSMO-REA6 whereas the boundary fields for COSMO-REA6 were chosen to be provided by ERA-Interim. ERA-Interim is a global reanalysis produced at ECMWF (Dee *et al.*, 2011) which currently covers the years 1979 to 2014 and is frequently extended to the present. ERA-Interim produces analysis fields every 6 hours and additionally performs forecast runs at 00 and 12 UTC with 3-hourly output. In this way output fields from ERA-Interim are available every 3 hours.
The boundary data for COSMO-REA6 are updated every three hours and linearly interpolated in between. Since analysis steps from ERA-Interim are only available every 6 hours the analysis fields from 00 and 12 UTC and the 3, 6 and 9 hour forecast integrations from these analyses

Table 2.2.: Observation types and corresponding assimilated variables used in the nudging scheme of COSMO-REA6.

Observations	Observation type	Assimilated variables	Mean number of assimilated reports per 6 hours
Radiosondes	PILOT	upper-air wind	187
	TEMP	temperature	134
		upper-air humidity	
		upper-air-wind	
		surface-level wind	
		surface-level temperature	
		surface-level humidity	
		geopotential	
Aircraft	AIREP	wind	868
		temperature	
	AMDAR	wind	6343
		temperature	
	ACARS	wind	7310
		temperature	
Wind profiler		upper-air wind	186
Surface level	SYNOP	station pressure	18580
		10-m wind	
		2-m humidity	
	SHIP	station pressure	166
		10-m wind	
		2-m humidity	
	DRIBU	station pressure	310
		10-m wind	
		2-m humidity	

are taken as boundary fields for 03, 06 and 09 UTC respectively 15, 18 and 21 UTC for consistency reasons. The resolution of the used boundary files is 0.5°. As boundary input from ERA-Interim the following variables are used: the geopotential, skin temperature, the soil temperature as well as the volumetric soil water, snow depth and temperature of the snow layer, skin reservoir content, sea-ice cover and ice temperature in the first layer, surface pressure as ground variables and specific humidity, temperature, wind and cloud liquid and ice water content in all heights. Additionally, the SST analysis module used the sea surface temperature of ERA-Interim as lower boundary condition. This results in a passive assimilation of satellite data for the SST information in COSMO-REA6, since ERA-Interim assimilates satellite data.
ERA-Interim further serves as reference reanalysis for the verification of COSMO-REA6. For this purpose the necessary variables with 3-hourly temporal and 0.5° horizontal resolution are used.

2.3.3. GRIB-Output from COSMO-REA6 and COSMO-REA2

The output fields from both reanalyses are written in GRIB1-Format and are divided into three-dimensional and two-dimensional fields. Three-dimensional fields are stored every hour whereas two dimensional fields are stored every 15 minutes. With that both reanalyses have a so far unmatched highly resolved output rate among reanalyses and can serve for many purposes. A list of output variables from both COSMO-REA6 and COSMO-REA2 is provided in Appendix B. In addition to this output synthetic satellite images for eight infrared channels onboard the geostationary satellites MeteoSat Second Generation (MSG) are produced every 15 minutes for COSMO-REA6, using the Radiative Transfer model for TIROS Operational Vertical Sounder 9 (RTTOV-9) (Saunders *et al.*, 2007) which is implemented in the COSMO model. The RTTOV model produces clear-sky and cloudy radiances and brightness temperatures. Due to a technical issue, the compilation of the COSMO model with RTTOV-9 enabled was not successful which is why the synthetic satellite images for COSMO-REA6 are only available for the first stream, i.e. for the years 2007-2012. In model version 5.00 of the COSMO model, the RTTOV-10 operator was introduced (Saunders *et al.*, 2012). The production of the synthetic satellite images in COSMO-REA2 was switched to RTTOV-10 for the complete data set.

2.3.4. GPCC

The Global Precipitation Climatology Centre (GPCC) provides global precipitation analyses for the monitoring and research of the earth´s climate and is operated by DWD. There are several different data sets like first guess products, climatologies and reanalysis products. In this work, GPCC´s full data reanalysis version 6 (Schneider *et al.*, 2014) is used for comparison with the COSMO-REA output. This GPCC reanalysis, based on quality controlled in-situ data, is available for the years 1901-2010 and the highest available resolution is 0.5°.

2.3.5. E-OBS

The European Climate Assessment & Dataset have created a European land-only daily gridded data set for precipitation and minimum, maximum and mean temperature, called E-OBS (Haylock *et al.*, 2008). The data is available from 1950 to 2013 on a regular grid of 0.25° and the data of version 10.0 are used here.

2.3.6. Rain gauges

DWD is operating a large number of rain gauges across Germany with a high temporal observation frequency all across Germany. A quality controlled data set of 1034 stations measuring the precipitation in hourly intervals is available from April 2003 until 2014. Since all analyses will be done on complete years the data set is used for the years 2004-2014 for the results shown in chapter 4.

2.3.7. CERES-EBAF

For the evaluation of radiation fluxes data from the Clouds and the Earth´s Radiant Energy System - Energy Balanced and Filled (CERES-EBAF). The flux data at the top of the atmosphere (Loeb *et al.*, 2009) are available for March 2000 to October 2014 and the flux data at the surface (Kato *et al.*, 2013) are available from March 2000 to September 2014. The evaluations

will be restricted to complete years, which is why only data for the years 2001-2013 are used. The data are available on a regular $1^\circ \times 1^\circ$ grid.

2.4. Climate classification

In the late 19th and early 20th century Wladimir Köppen (Köppen, 1884; Köppen, 1918) developed a classification of climate regimes. Since he lacked the full bandwidth of observations available nowadays, he focused on two easily measurable variables while classifying different climates: temperature and precipitation. Geiger updated Köppens´ work in the middle of the century and the newest maps of world climates are updated by the University of Vienna (Kottek *et al.*, 2006).
The climate types are depicted with two or three letters. The first letter characterizes the main climates with the subsequent letters characterizing the precipitation and temperature conditions of the region. Five main climates of the earth are distinguished, namely equatorial, arid, warm temperate, snow and polar climates, indicated by the large letters A to E. The precipitation conditions for the arid climates are distinguished by desert and steppe conditions with large letters W and S. The other main climates are characterized by small letters as fully humid (f), summer dry (s), winter dry (w) and monsoonal (m). The equatorial climates are independent of any temperature conditions and only characterized by precipitation conditions whereas the polar climates are independent of precipitation conditions and only characterized by the temperature conditions "frost" (F) and "tundra" (T). The arid climates are further subdivided by the temperature conditions hot arid (h) and cold arid (k). The warm temperate and snow climates receive the additional characteristics of hot (a), warm (b) or cool summer (c) and extremely continental (d).Table 2.3 reproduces the criteria for the main climates and the precipitation conditions and Table 2.4 for the temperature conditions of these climates from Kottek *et al.*, 2006.
The following variables are needed for the classifications: The annual mean near-surface, i.e. 2-m, temperature (T_{ann}) as well as the monthly mean temperatures of the coldest (T_{min}) and warmest (T_{max}) months. To obtain the coldest and warmest month the mean of every month over the years (T_{mon}) is calculated and then the coldest and warmest of these twelve months is determined. For precipitation the accumulated annual precipitation (P_{ann}) as well as the precipitation of the driest month P_{min} is needed. The driest month is determined similar to the coldest month, i.e. calculate the mean monthly precipitation for the twelve months and then find the driest month from that. In addition the lowest and highest monthly precipitation values for the summer and winter half years (P_{smin}, P_{smax}, P_{wmin} and P_{wmax}) on the hemisphere under consideration are needed. On the northern hemisphere the winter half year is defined as the months October to March and the summer half year as April to September. For the arid climates (B) a dryness threshold P_{th} in mm is introduced which is calculated as follows

$$P_{th} = \begin{cases} 2|T_{ann}| & \text{if at least 2/3 of the annual precipitation occurs in winter} \\ 2|T_{ann}| + 28 & \text{if at least 2/3 of the annual precipitation occurs in summer} \\ 2|T_{ann}| + 14 & \text{otherwise} \end{cases} \quad (2.111)$$

where $|T_{ann}|$ is the absolute value of the annual mean temperature in °C. The world map of Köppen-Geiger classifying climates from Kottek *et al.*, 2006 is reproduced in Fig. 2.8. This map is based on temperature data from the Climatic Research Unit (CRU, Mitchell and Jones,

Table 2.3.: Type, description and criterions for the main climates and the subsequent precipitation conditions for the Köppen-Geiger climate classifications.

Type	Description	Criterion
A	**Equatorial climates**	$T_{min} \geq +18°C$
Af	Equatorial rainforest, fully humid	$P_{min} \geq 60mm$
Am	Equatorial monsoon	$P_{ann} \geq 25(100 - P_{min})$
As	Equatorial savannah with dry summer	$P_{min} < 60mm$ in summer
As	Equatorial savannah with dry summer	$P_{min} < 60mm$ in winter
B	**Arid climates**	$P_{ann} < 10P_{th}$
BS	Steppe climate	$P_{ann} > 5P_{th}$
BW	Desert climate	$P_{ann} \leq 5P_{th}$
C	**Warm temperate climates**	$-3°C < T_{min} < +18°C$
Cs	Warm temperate climate with dry summer	$P_{smin} < P_{wmin}, P_{wmax} > 3P_{smin}$ and $P_{smin} < 40$ mm
Cw	Warm temperate climate with dry winter	$P_{wmin} < P_{smin}$ and $P_{smax} > 10P_{wmin}$
Cf	Warm temperate climate, fully humid	neither Cs nor Cw
D	**Snow climates**	$T_{min} \leq -3°C$
Ds	Snow climate with dry summer	$P_{smin} < P_{wmin}, P_{wmax} > 3P_{smin}$ and $P_{smin} < 40$ mm
Dw	Snow climate with dry winter	$P_{wmin} < P_{smin}$ and $P_{smax} > 10P_{wmin}$
Df	Snow climate, fully humid	neither Ds nor Dw
E	**Polar climates**	$T_{max} < +10°C$
ET	Tundra climate	$0°C \leq T_{max} < +10°C$
EF	Frost climate	$T_{max} < 0°C$

Table 2.4.: Type, description and criterions for the temperature conditions for the arid climates (B) and for the warm temperate (C) and snow climates (D) for the Köppen-Geiger climate classifications.

Type	Description	Criterion
h	Hot steppe / desert	$T_{ann} \geq +18°C$
k	Cold steppe / desert	$T_{ann} < +18°C$
a	Hot summer	$T_{max} \geq +22°C$
b	Warm summer	not (a) and at least 4 $T_{mon} \geq +10°C$
c	Cool summer and cold winter	not (b) and $T_{min} > -38°C$
d	extremely continental	not (c) and $T_{min} \leq -38°C$

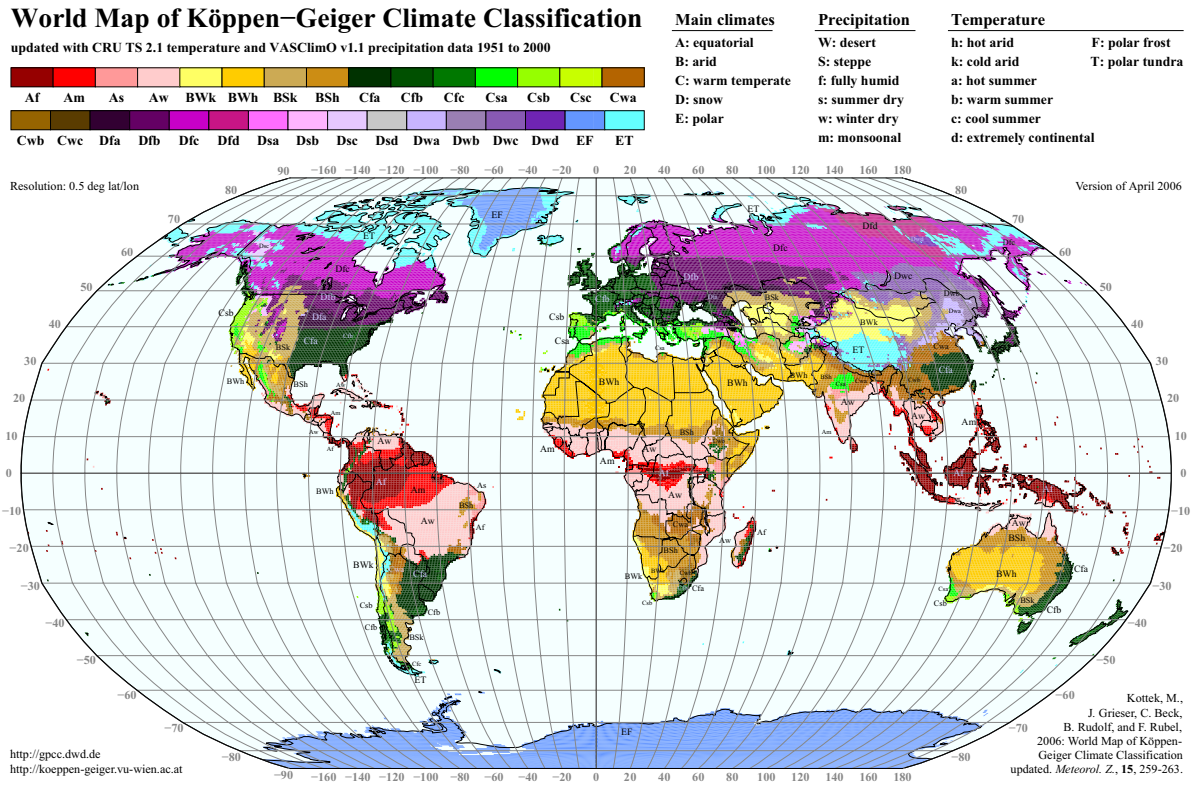


Figure 2.8.: World map of Köppen-Geiger climate classifications from Kottek *et al.*, 2006, based on 50 years of CRU and GPCC data.

2005) and on precipitation data from GPCC for the years 1951-2000 (Beck *et al.*, 2004). The results for different reanalyses and data sets can be found in section 4.4.

3. Variational approach for the moisture budget

The moisture budget is defined as (see e.g. Trenberth *et al.*, 2011)

$$\frac{\partial}{\partial t}\left(\frac{1}{g}\int q dp\right) + \vec{\nabla}\cdot\left(\frac{1}{g}\int q\vec{v}dp\right) = E - P \tag{3.112}$$

and comprises the storage term, the horizontal transport divergences of vertically integrated moisture as well as evapotranspiration E and precipitation P. The storage term is defined as the local change of the vertically integrated moisture over time. In a long-term average, the local tendencies are very small and can therefore be ignored. This then leads to the fundamental balance

$$\begin{aligned} \vec{\nabla}\cdot\left(\frac{1}{g}\int q\vec{v}dp\right) &= E - P \\ \vec{\nabla}\cdot\vec{M} &= E - P \\ \vec{\nabla}\cdot\vec{M} - (E-P) &= 0 \\ \vec{\nabla}\cdot\vec{M} - S &= 0 \end{aligned} \tag{3.113}$$

where $\vec{M} = \frac{1}{g}\int q\vec{v}dp$ is the vertically integrated moisture transport and $S = E - P$ is the vertically integrated vertical moisture flux divergence, abbreviated as VMD in the following. Eq. (3.113) is not fulfilled in the data, therefore the transports and the VMD need to be corrected in a consistent way. Bollmeyer and Hense, 2014 proposed a way of modifying the energy transports together with the vertically integrated vertical energy flux divergence in a self-consistent way to fulfill the energy balance in a combination of ERA40 and ISCCP data which will be adopted here for the moisture budget.
For this purpose, a cost function is introduced to obtain the modified transports and vertical fluxes:

$$\begin{aligned} J &= \frac{1}{2}\int \frac{1}{\sigma_M^2}(\vec{M}-\vec{M}_0)^T(\vec{M}-\vec{M}_0)dF + \frac{1}{2}\int \frac{1}{\sigma_S^2}(S-S_0)^T(S-S_0)dF \\ &\quad + \int \gamma^T(\vec{\nabla}\cdot\vec{M} - S)dF + \oint_F \mu^T(\vec{M}-\vec{M}_B)\vec{n}d\Omega \stackrel{!}{=} \min . \end{aligned} \tag{3.114}$$

Here, $\vec{M}_0$ and S_0 are the original and $\vec{M}$ and S the modified moisture transports and VMD. σ_M and σ_S are the errors of the moisture transports and the VMD. These are assumed to be independent for each grid point and have been set to the values $\sigma_S = 5\cdot 10^{-6}$ kg m^{-2} s^{-1} and $\sigma_M = a\cdot\sigma_S\cdot\kappa$ with a the earth's radius and κ a scaling parameter. For the remainder of this study, κ is set to 0.1 and its influence on the outcomes is discussed in section 4.5. In Eq. (3.114), two Lagrangian multiplicators have been introduced. The first one, γ, defines the side constraint that the divergence of the moisture transports equals the VMD. The second side constraint, introduced via the Lagrangian multiplicator μ, states that the modified moisture transport into or out of the model should equal the moisture transports coming from the

boundary fields, $\vec{M}_B$. This is a direct implementation of boundary conditions into the variational approach.
In the following, the moisture transport is divided into its zonal and meridional part

$$\begin{aligned} \vec{M} &= M_\lambda \vec{e}_\lambda + M_\varphi \vec{e}_\varphi \\ \vec{M}_0 &= M_{\lambda,0}\vec{e}_\lambda + M_{\varphi,0}\vec{e}_\varphi \,. \end{aligned} \tag{3.115}$$

This leads to

$$\begin{aligned} J = \; & \frac{1}{2}\int \frac{1}{\sigma_M^2}\Big[(M_\lambda - M_{\lambda,0})^T(M_\lambda - M_{\lambda,0}) + (M_\lambda - M_{\lambda,0})^T(M_\varphi - M_{\varphi,0}) \\ & + (M_\varphi - M_{\varphi,0})^T(M_\lambda - M_{\lambda,0}) + (M_\varphi - M_{\varphi,0})^T(M_\varphi - M_{\varphi,0})\Big] dF \\ & + \frac{1}{2}\int \frac{1}{\sigma_S^2}(S - S_0)^T(S - S_0) dF \\ & + \int \gamma^T(\vec{\nabla}\cdot(M_\lambda \vec{e}_\lambda + M_\varphi \vec{e}_\varphi) - S) dF \\ & + \oint_F \mu^T\big((M_\lambda \vec{e}_\lambda + M_\varphi \vec{e}_\varphi) - (M_{\lambda,B}\vec{e}_\lambda + M_{\varphi,B}\vec{e}_\varphi)\big)\vec{n} d\Omega \,. \end{aligned} \tag{3.116}$$

At this point, a discretization in finite elements is applied. This is explained in more detail in appendix A.1. The discretization is done by means of the following formula

$$\psi(\vec{x}) = \sum_{i=1}^{N} \psi_i h_i(\vec{x}) \tag{3.117}$$

with ψ_i constant coefficients and h_i a set of basis functions. Applying this discretization to both the wanted and observed moisture transports and vertical fluxes, the boundary moisture transport and the Lagrangian multiplicators γ and μ

$$M_\lambda = \sum_{i=1}^{N} m_\lambda^i h_i(\lambda,\varphi) \tag{3.118}$$

$$M_{\lambda,0} = \sum_{i=1}^{N} m_{\lambda,0}^i h_i(\lambda,\varphi) \tag{3.119}$$

$$M_\varphi = \sum_{i=1}^{N} m_\varphi^i h_i(\lambda,\varphi) \tag{3.120}$$

$$M_{\varphi,0} = \sum_{i=1}^{N} m_{\varphi,0}^i h_i(\lambda,\varphi) \tag{3.121}$$

$$S = \sum_{i=1}^{N} s^i h_i(\lambda,\varphi) \tag{3.122}$$

$$S_0 = \sum_{i=1}^{N} s_0^i h_i(\lambda,\varphi) \tag{3.123}$$

$$\gamma = \sum_{i=1}^{N} \gamma^i h_i(\lambda,\varphi) \tag{3.124}$$

$$M_{B,\lambda} = \sum_{i=1}^{N} m^i_{B,\lambda} h_i(\lambda,\varphi) \tag{3.125}$$

$$M_{B,\varphi} = \sum_{i=1}^{N} m^i_{B,\varphi} h_i(\lambda,\varphi) \tag{3.126}$$

and inserting into the cost function leads to

$$\begin{aligned}
J = & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(m^i_\lambda - m^i_{\lambda,0})^T(m^j_\lambda - m^j_{\lambda,0})\int \frac{1}{\sigma_M^2} h_i(\lambda,\varphi)^T h_j(\lambda,\varphi)dF \\
& + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(m^i_\lambda - m^i_{\lambda,0})^T(m^j_\varphi - m^j_{\varphi,0})\int \frac{1}{\sigma_M^2} h_i(\lambda,\varphi)^T h_j(\lambda,\varphi)dF \\
& + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(m^i_\varphi - m^i_{\varphi,0})^T(m^j_\lambda - m^j_{\lambda,0})\int \frac{1}{\sigma_M^2} h_i(\lambda,\varphi)^T h_j(\lambda,\varphi)dF \\
& + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(m^i_\varphi - m^i_{\varphi,0})^T(m^j_\varphi - m^j_{\varphi,0})\int \frac{1}{\sigma_M^2} h_i(\lambda,\varphi)^T h_j(\lambda,\varphi)dF \\
& + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(s^i - s^i_0)^T(s^j - s^j_0)\int \frac{1}{\sigma_S^2} h_i(\lambda,\varphi)^T h_j(\lambda,\varphi)dF \\
& + \sum_{i=1}^{N}\sum_{j=1}^{N}\gamma^{iT} m^j_\lambda \int h_i(\lambda,\varphi)\frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda} h_j(\lambda,\varphi)dF \\
& + \sum_{i=1}^{N}\sum_{j=1}^{N}\gamma^{iT} m^j_\varphi \int h_i(\lambda,\varphi)\frac{1}{a}\frac{\partial}{\partial\varphi} h_j(\lambda,\varphi)dF \\
& - \sum_{i=1}^{N}\sum_{j=1}^{N}\gamma^{iT} s^j \int h_i(\lambda,\varphi) h_j(\lambda,\varphi)dF \\
& + \sum_{i=1}^{N}\sum_{j=1}^{N}\mu^{iT}(m^j_\lambda - m^j_{\lambda,B})\oint \overline{\left(h_i(\lambda,\varphi)h_j(\lambda,\varphi)\right)}\vec{n}d\Omega \\
& + \sum_{i=1}^{N}\sum_{j=1}^{N}\mu^{iT}(m^j_\varphi - m^j_{\varphi,B})\oint \overline{\left(h_i(\lambda,\varphi)h_j(\lambda,\varphi)\right)}\vec{n}d\Omega\,.
\end{aligned} \tag{3.127}$$

This rather lengthy equation can be rearranged with vectors and matrices, leading to the following system of equations

$$\begin{aligned}
J = & \frac{1}{2\sigma_M^2}(\vec{m}_\lambda - \vec{m}_{\lambda,0})^T\boldsymbol{\mathcal{T}}(\vec{m}_\lambda - \vec{m}_{\lambda,0}) + \frac{1}{2\sigma_M^2}(\vec{m}_\lambda - \vec{m}_{\lambda,0})^T\boldsymbol{\mathcal{T}}(\vec{m}_\varphi - \vec{m}_{\varphi,0}) \\
+ & \frac{1}{2\sigma_M^2}(\vec{m}_\varphi - \vec{m}_{\varphi,0})^T\boldsymbol{\mathcal{T}}(\vec{m}_\lambda - \vec{m}_{\lambda,0}) + \frac{1}{2\sigma_M^2}(\vec{m}_\varphi - \vec{m}_{\varphi,0})^T\boldsymbol{\mathcal{T}}(\vec{m}_\varphi - \vec{m}_{\varphi,0}) \\
+ & \frac{1}{2\sigma_S^2}(\vec{s} - \vec{s}_0)^T\boldsymbol{\mathcal{T}}(\vec{s} - \vec{s}_0) \\
+ & \vec{\gamma}^T\left(\boldsymbol{\mathcal{D}}_\lambda\vec{m}_\lambda + \boldsymbol{\mathcal{D}}_\varphi\vec{m}_\varphi - \boldsymbol{\mathcal{T}}\vec{s}\right) + \mu^T\Big(\sum_{\lambda_B}(\vec{m}_\lambda - \vec{m}_{\lambda,B})|_{\lambda_B} + \sum_{\varphi_B}(\vec{m}_\varphi - \vec{m}_{\varphi,B})|_{\varphi_B}\Big)\,.
\end{aligned} \tag{3.128}$$

Here, the operators $|_{\lambda_B}$ and $|_{\varphi_B}$ indicate that the term is evaluated only at the boundary points in λ and φ. The various vectors and matrices are formed as follows

$$\vec{m}_\lambda = \begin{pmatrix} m_\lambda^1 \\ m_\lambda^2 \\ \vdots \\ m_\lambda^{N-1} \\ m_\lambda^N \end{pmatrix}, \vec{m}_\varphi = \begin{pmatrix} m_\varphi^1 \\ m_\varphi^2 \\ \vdots \\ m_\varphi^{N-1} \\ m_\varphi^N \end{pmatrix}, \vec{m}_{\lambda,0} = \begin{pmatrix} m_{\lambda,0}^1 \\ m_{\lambda,0}^2 \\ \vdots \\ m_{\lambda,0}^{N-1} \\ m_{\lambda,0}^N \end{pmatrix}, \vec{m}_{\varphi,0} = \begin{pmatrix} m_{\varphi,0}^1 \\ m_{\varphi,0}^2 \\ \vdots \\ m_{\varphi,0}^{N-1} \\ m_{\varphi,0}^N \end{pmatrix} \tag{3.129}$$

$$\vec{s} = \begin{pmatrix} s^1 \\ s^2 \\ \vdots \\ s^{N-1} \\ s^N \end{pmatrix}, \vec{s}_0 = \begin{pmatrix} s_0^1 \\ s_0^2 \\ \vdots \\ s_0^{N-1} \\ s_0^N \end{pmatrix}, \vec{\gamma} = \begin{pmatrix} \gamma^1 \\ \gamma^2 \\ \vdots \\ \gamma^{N-1} \\ \gamma^N \end{pmatrix} \tag{3.130}$$

$$\vec{m}_{B,\lambda} = \begin{pmatrix} m_{B,\lambda}^1 \\ m_{B,\lambda}^2 \\ \vdots \\ m_{B,\lambda}^{N-1} \\ m_{B,\lambda}^N \end{pmatrix}, \vec{m}_{B,\varphi} = \begin{pmatrix} m_{B,\varphi}^1 \\ m_{B,\varphi}^2 \\ \vdots \\ m_{B,\varphi}^{N-1} \\ m_{B,\varphi}^N \end{pmatrix} \tag{3.131}$$

$$\begin{aligned} \mathcal{T} = & \begin{pmatrix} \int h_1(\lambda,\varphi)^T h_1(\lambda,\varphi) dF & \int h_1(\lambda,\varphi)^T h_2(\lambda,\varphi) & \cdots \\ \int h_2(\lambda,\varphi)^T h_1(\lambda,\varphi) dF & \int h_2(\lambda,\varphi)^T h_2(\lambda,\varphi) & \cdots \\ \vdots & & \cdots \\ \int h_{N-1}(\lambda,\varphi)^T h_1(\lambda,\varphi) dF & \int h_{N-1}(\lambda,\varphi)^T h_2(\lambda,\varphi) & \cdots \\ \int h_N(\lambda,\varphi)^T h_1(\lambda,\varphi) dF & \int h_N(\lambda,\varphi)^T h_2(\lambda,\varphi) & \cdots \end{pmatrix} \\ & \begin{pmatrix} \cdots & \int h_1(\lambda,\varphi)^T h_{N-1}(\lambda,\varphi) & \int h_1(\lambda,\varphi)^T h_N(\lambda,\varphi) \\ \cdots & \int h_2(\lambda,\varphi)^T h_{N-1}(\lambda,\varphi) & \int h_2(\lambda,\varphi)^T h_N(\lambda,\varphi) \\ \cdots & & \vdots \\ \cdots & \int h_{N-1}(\lambda,\varphi)^T h_{N-1}(\lambda,\varphi) dF & \int h_{N-1}(\lambda,\varphi)^T h_N(\lambda,\varphi) \\ \cdots & \int h_N(\lambda,\varphi)^T h_{N-1}(\lambda,\varphi) & \int h_N(\lambda,\varphi)^T h_N(\lambda,\varphi) \end{pmatrix} \end{aligned} \tag{3.132}$$

$$
\boldsymbol{\mathcal{D}}_\lambda = \begin{pmatrix}
\int h_1(\lambda,\varphi)^T \frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda} h_1(\lambda,\varphi)dF & \int h_1(\lambda,\varphi)^T \frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda} h_2(\lambda,\varphi) & \cdots \\
\int h_2(\lambda,\varphi)^T \frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda} h_1(\lambda,\varphi)dF & \int h_2(\lambda,\varphi)^T \frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda} h_2(\lambda,\varphi) & \cdots \\
\vdots & & \cdots \\
\int h_{N-1}(\lambda,\varphi)^T \frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda} h_1(\lambda,\varphi)dF & \int h_{N-1}(\lambda,\varphi)^T \frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda} h_2(\lambda,\varphi) & \cdots \\
\int h_N(\lambda,\varphi)^T \frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda} h_1(\lambda,\varphi)dF & \int h_N(\lambda,\varphi)^T \frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda} h_2(\lambda,\varphi) & \cdots
\end{pmatrix}
$$

$$
\begin{pmatrix}
\cdots & \int h_1(\lambda,\varphi)^T \frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda} h_{N-1}(\lambda,\varphi) & \int h_1(\lambda,\varphi)^T \frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda} h_N(\lambda,\varphi) \\
\cdots & \int h_2(\lambda,\varphi)^T \frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda} h_{N-1}(\lambda,\varphi) & \int h_2(\lambda,\varphi)^T \frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda} h_N(\lambda,\varphi) \\
\cdots & & \vdots \\
\cdots & \int h_{N-1}(\lambda,\varphi)^T \frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda} h_{N-1}(\lambda,\varphi)dF & \int h_{N-1}(\lambda,\varphi)^T \frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda} h_N(\lambda,\varphi) \\
\cdots & \int h_N(\lambda,\varphi)^T \frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda} h_{N-1}(\lambda,\varphi) & \int h_N(\lambda,\varphi)^T \frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda} h_N(\lambda,\varphi)
\end{pmatrix} \tag{3.133}
$$

$$
\boldsymbol{\mathcal{D}}_\varphi = \begin{pmatrix}
\int h_1(\lambda,\varphi)^T \frac{1}{a}\frac{\partial}{\partial\varphi} h_1(\lambda,\varphi)dF & \int h_1(\lambda,\varphi)^T \frac{1}{a}\frac{\partial}{\partial\varphi} h_2(\lambda,\varphi) & \cdots \\
\int h_2(\lambda,\varphi)^T \frac{1}{a}\frac{\partial}{\partial\varphi} h_1(\lambda,\varphi)dF & \int h_2(\lambda,\varphi)^T \frac{1}{a}\frac{\partial}{\partial\varphi} h_2(\lambda,\varphi) & \cdots \\
\vdots & & \cdots \\
\int h_{N-1}(\lambda,\varphi)^T \frac{1}{a}\frac{\partial}{\partial\varphi} h_1(\lambda,\varphi)dF & \int h_{N-1}(\lambda,\varphi)^T \frac{1}{a}\frac{\partial}{\partial\varphi} h_2(\lambda,\varphi) & \cdots \\
\int h_N(\lambda,\varphi)^T \frac{1}{a}\frac{\partial}{\partial\varphi} h_1(\lambda,\varphi)dF & \int h_N(\lambda,\varphi)^T \frac{1}{a}\frac{\partial}{\partial\varphi} h_2(\lambda,\varphi) & \cdots
\end{pmatrix}
$$

$$
\begin{pmatrix}
\cdots & \int h_1(\lambda,\varphi)^T \frac{1}{a}\frac{\partial}{\partial\varphi} h_{N-1}(\lambda,\varphi) & \int h_1(\lambda,\varphi)^T \frac{1}{a}\frac{\partial}{\partial\varphi} h_N(\lambda,\varphi) \\
\cdots & \int h_2(\lambda,\varphi)^T \frac{1}{a}\frac{\partial}{\partial\varphi} h_{N-1}(\lambda,\varphi) & \int h_2(\lambda,\varphi)^T \frac{1}{a}\frac{\partial}{\partial\varphi} h_N(\lambda,\varphi) \\
\cdots & & \vdots \\
\cdots & \int h_{N-1}(\lambda,\varphi)^T \frac{1}{a}\frac{\partial}{\partial\varphi} h_{N-1}(\lambda,\varphi)dF & \int h_{N-1}(\lambda,\varphi)^T \frac{1}{a}\frac{\partial}{\partial\varphi} h_N(\lambda,\varphi) \\
\cdots & \int h_N(\lambda,\varphi)^T \frac{1}{a}\frac{\partial}{\partial\varphi} h_{N-1}(\lambda,\varphi) & \int h_N(\lambda,\varphi)^T \frac{1}{a}\frac{\partial}{\partial\varphi} h_N(\lambda,\varphi)
\end{pmatrix} \tag{3.134}
$$

Now to get the minimum, Eq. (3.128) needs to be differentiated with respect to the unknown vectors $\vec{m}_\lambda$, $\vec{m}_\varphi$, $\vec{s}$ and $\vec{\gamma}$ and the scalar Lagrangian multiplicator μ. The differentiation leads to the following five equations

$$
\begin{aligned}
\vec{\nabla}_{\vec{m}_\lambda} J &: \frac{1}{\sigma_M^2}\boldsymbol{\mathcal{T}}\vec{m}_\lambda + \boldsymbol{\mathcal{D}}_\lambda^T\vec{\gamma} = \frac{1}{\sigma_M^2}\boldsymbol{\mathcal{T}}\vec{m}_{\lambda,0} \\
\vec{\nabla}_{\vec{m}_\varphi} J &: \frac{1}{\sigma_M^2}\boldsymbol{\mathcal{T}}\vec{m}_\varphi + \boldsymbol{\mathcal{D}}_\varphi^T\vec{\gamma} = \frac{1}{\sigma_M^2}\boldsymbol{\mathcal{T}}\vec{m}_{\varphi,0} \\
\vec{\nabla}_{\vec{s}} J &: \frac{1}{\sigma_S^2}\boldsymbol{\mathcal{T}}\vec{s} - \boldsymbol{\mathcal{T}}\vec{\gamma} = \frac{1}{\sigma_S^2}\boldsymbol{\mathcal{T}}\vec{s}_0 \\
\vec{\nabla}_{\vec{\gamma}} J &: \boldsymbol{\mathcal{D}}_\lambda\vec{m}_\lambda + \boldsymbol{\mathcal{D}}_\varphi\vec{m}_\varphi - \boldsymbol{\mathcal{T}}\vec{s} = 0 \\
\vec{\nabla}_{\mu} J &: \sum_{\lambda_B}(\vec{m}_\lambda - \vec{m}_{\lambda,B})|_{\lambda_B} + \sum_{\varphi_B}(\vec{m}_\varphi - \vec{m}_{\varphi,B})|_{\varphi_B} = 0
\end{aligned} \tag{3.135}
$$

where the symmetry of $\boldsymbol{\mathcal{T}}$ was used. This complete system of five differential equations can be written in a compact form as

$$\begin{pmatrix} \frac{1}{\sigma_M^2}\boldsymbol{\mathcal{T}} & \mathbf{0} & \mathbf{0} & \boldsymbol{\mathcal{D}}_\lambda^T & \vec{1}|_{\lambda_B} \\ \mathbf{0} & \frac{1}{\sigma_M^2}\boldsymbol{\mathcal{T}} & \mathbf{0} & \boldsymbol{\mathcal{D}}_\varphi^T & \vec{1}|_{\varphi_B} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\sigma_S^2}\boldsymbol{\mathcal{T}} & -\boldsymbol{\mathcal{T}} & \vec{0} \\ \boldsymbol{\mathcal{D}}_\lambda & \boldsymbol{\mathcal{D}}_\varphi & -\boldsymbol{\mathcal{T}} & \mathbf{0} & \vec{0} \\ \vec{1}|_{\lambda_B}^T & \vec{1}|_{\varphi_B}^T & \vec{0}^T & \vec{0}^T & 0 \end{pmatrix} \cdot \begin{pmatrix} \vec{m}_\lambda \\ \vec{m}_\varphi \\ \vec{s} \\ \vec{\gamma} \\ \mu \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma_M^2}\boldsymbol{\mathcal{T}}\vec{m}_{\lambda,0} \\ \frac{1}{\sigma_M^2}\boldsymbol{\mathcal{T}}\vec{m}_{\varphi,0} \\ \frac{1}{\sigma_S^2}\boldsymbol{\mathcal{S}}\vec{s}_0 \\ \vec{0} \\ \sum_{\lambda_B}\vec{m}_{\lambda,B}|_{\lambda_B} + \sum_{\varphi_B}\vec{m}_{\varphi,B}|_{\varphi_B} \end{pmatrix}. \quad (3.136)$$

In the last row and last column on the points of the boundaries there are 1's and 0 everywhere else. Solving Eq. (3.136) results in the corrected moisture transports $\vec{m}_\lambda$ and $\vec{m}_\varphi$ and the corrected VMD $\vec{s}$.

By its very nature, the variational approach only provides the divergent part of the moisture transports. Therefore, in order to compare the outcome of the variational approach the divergent part of the original moisture transports needs to be extracted. For this purpose, the following equation is solved:

$$J = \frac{1}{2}\int \frac{1}{\sigma_M^2}(\vec{M}-\vec{M}_0)^T(\vec{M}-\vec{M}_0)dF + \int \gamma^T(\vec{\nabla}\times\vec{M})dF \stackrel{!}{=} \min . \quad (3.137)$$

Here, the modified transports should fulfill the side constraint that the rotation of the moisture transports is zero. Every vector field $\vec{X}$ can be divided into a purely rotational and a purely divergent part following Helmholtz-Theorem

$$\begin{aligned} \vec{X} &= \vec{X}_{div} + \vec{X}_{rot} \\ &= \vec{\nabla}\cdot\chi + \vec{e}_r \times \vec{\nabla}\psi \end{aligned} \quad (3.138)$$

with

$$\begin{aligned} \vec{\nabla}\times\vec{X}_{div} &= 0 \\ \vec{\nabla}\cdot\vec{X}_{rot} &= 0 \end{aligned} \quad (3.139)$$

So by constraining the rotation of the moisture transports to be zero only the divergent part of the flow will remain. Now the same steps as for the variational approach above are applied. First, the moisture transports are divided into their zonal and meridional part leading to

$$\begin{aligned} J &= \frac{1}{2}\int \frac{1}{\sigma_M^2}\Big[(M_\lambda - M_{\lambda,0})^T(M_\lambda - M_{\lambda,0}) + (M_\lambda - M_{\lambda,0})^T(M_\varphi - M_{\varphi,0}) \\ &\quad + (M_\varphi - M_{\varphi,0})^T(M_\lambda - M_{\lambda,0}) + (M_\varphi - M_{\varphi,0})^T(M_\varphi - M_{\varphi,0})\Big]dF \\ &\quad + \int \gamma^T\left(\vec{\nabla}\times(M_\lambda\vec{e}_\lambda + M_\varphi\vec{e}_\varphi)\right)dF . \end{aligned} \quad (3.140)$$

Then, the discretization in finite elements is applied and the cost function becomes

$$\begin{aligned} J &= \frac{1}{2\sigma_M^2}(\vec{m}_\lambda - \vec{m}_{\lambda,0})^T\boldsymbol{\mathcal{T}}(\vec{m}_\lambda - \vec{m}_{\lambda,0}) + \frac{1}{2\sigma_M^2}(\vec{m}_\lambda - \vec{m}_{\lambda,0})^T\boldsymbol{\mathcal{T}}(\vec{m}_\varphi - \vec{m}_{\varphi,0}) \\ &+ \frac{1}{2\sigma_M^2}(\vec{m}_\varphi - \vec{m}_{\varphi,0})^T\boldsymbol{\mathcal{T}}(\vec{m}_\lambda - \vec{m}_{\lambda,0}) + \frac{1}{2\sigma_M^2}(\vec{m}_\varphi - \vec{m}_{\varphi,0})^T\boldsymbol{\mathcal{T}}(\vec{m}_\varphi - \vec{m}_{\varphi,0}) \\ &+ \vec{\gamma}^T\left(\boldsymbol{\mathcal{D}}_\lambda\vec{m}_\varphi - \boldsymbol{\mathcal{D}}_\varphi\vec{m}_\lambda\right). \end{aligned} \quad (3.141)$$

Here, the notation with vectors and matrices is used directly. The differentiation of the cost function to get the minima leads to the differential set of equations

$$\begin{pmatrix} \frac{1}{\sigma_M^2}\boldsymbol{\mathcal{T}} & \mathbf{0} & -\boldsymbol{\mathcal{D}}_\varphi^T \\ \mathbf{0} & \frac{1}{\sigma_M^2}\boldsymbol{\mathcal{T}} & \boldsymbol{\mathcal{D}}_\lambda^T \\ -\boldsymbol{\mathcal{D}}_\varphi & \boldsymbol{\mathcal{D}}_\lambda & \mathbf{0} \end{pmatrix} \cdot \begin{pmatrix} \vec{m}_\lambda \\ \vec{m}_\varphi \\ \vec{\gamma} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma_M^2}\boldsymbol{\mathcal{T}}\vec{m}_{\lambda,0} \\ \frac{1}{\sigma_M^2}\boldsymbol{\mathcal{T}}\vec{m}_{\varphi,0} \\ \vec{0} \end{pmatrix}. \tag{3.142}$$

The solution of (3.142) then yields the divergent parts of the moisture transports which can be used for comparison with the results from Equation (3.136). The numerical implementation of the finite element method and how to solve Eq. (3.136) (and thus also Equation (3.142)) is explained in Appendix A. The above presented method requires the handling of very large matrices. If this approach would be computed for the whole COSMO-REA6 domain, the matrix $\boldsymbol{\mathcal{T}}$ would be of the size (N_x*N_y $\times$ N_x*N_y) = (848*824 $\times$ 848*824) = (698752 $\times$ 698752), which can not be handled by a standard PC. Therefore, this work can be seen as a study of feasibility and the extension to the whole domain remains work for future studies. For this work, a subdomain was chosen of the size $N_x = N_y = 120$, i.e. $\boldsymbol{\mathcal{T}}$ is of dimension (14400×14400), leading to a matrix of size (57601×57601) in Eq. (3.136), which could be handled in a reasonable amount of time.
The results of the presented approach are shown in section 4.5.

4. Results

4.1. Analysis increments

As mentioned in section 2.1.7.1 analysis increments are the values added to the model fields by the nudging. Analysis increments show the impact of the data assimilation on the model variables, i.e. the deviation of the model prognostic variables from the observations. Persistent positive or negative increments can reveal systematic errors within the forecast model or the observations, for example residual biases or false tendencies. These analysis increments for the model variables are available every hour as integrated quantities over the last hour for every model level.
Figures 4.1 and 4.2 show the analysis increments for temperature and horizontal wind speed in all 40 model levels with daily averages calculated from the hourly model output over the full domain exemplary for the year 2011. For temperature, the daily averages exhibit a rather small amplitude of ± 0.10 K over the year. The model apparently tends to be too cold during the wintertime with positive analysis increments from November to April from the surface up to model level 17, approximately corresponding to 450 hPa. On the other hand, the lower 5 km of the atmosphere appear to be too warm during the summer months from May to September since negative values of the analysis increments are present, most pronounced between model layers 33 and 19 (approximately 900 and 500 hPa, respectively). Over the whole year high positive values occur close to the upper boundary and in a prominent band spanning layers 7–9 (i.e. 100–50 and 200–150 hPa) which is associated with the tropopause and therefore with higher variability as underlined by the standard deviations which show larger values of around 0.2 K. Similarly the highest values of standard deviation (up to 0.3 K) are present in the boundary layer decreasing with height up to level 15.
For the horizontal wind speed four different vertical ranges of analysis increments can be distinguished. The first and most pronounced band of negative increments up to –0.10 m s^{-1} reaches from the surface to layer 31 ($\approx$ 900 hPa). Apparently, the boundary-layer winds in the COSMO model are too strong throughout the year, which could, together with the high standard deviations in the lower layers, point to problems with boundary layer processes such as mixing in the boundary layer. The second range with positive values reaches from model layer 30 to approximately layer 23 with a very sharp change at the intersection of layers 30 and 31. The middle troposphere around 500 hPa is characterized by slightly negative values and low standard deviations whereas the upper troposphere/lower stratosphere is dominated by positive values up to 0.07 m s^{-1} with higher values of standard deviation. This probably shows an underestimation of the speed of the jet which might be related to the problems in timing and positioning of the tropopause as mentioned above.
The overall very small values of analysis increments show that the model is performing reasonable since larger or increasing/decreasing increments with time would have indicated systematic errors. Furthermore the small analysis increment values show that the nudging data assimilation scheme does not disturb the balance of the model too much by drawing the model in one direction or another.

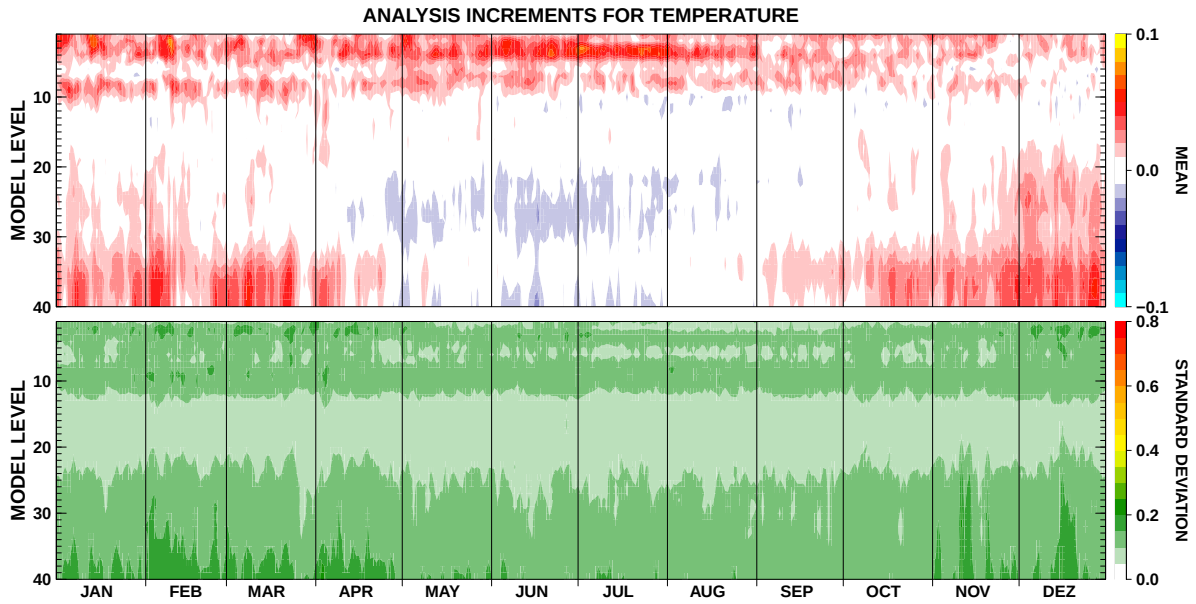


Figure 4.1.: Daily mean averages of hourly aggregated and area-averaged analysis increments for temperature over the complete model domain for all 40 hybrid vertical levels (top) and corresponding standard deviations (bottom) in K.

4.2. Precipitation

Precipitation is a very important part of the global water cycle and therefore a key parameter for the evaluation of the reanalysis. Further, precipitation is not assimilated into COSMO-REA6 at all and only via the LHN in COSMO-REA2, making precipitation fields at the surface suitable as independent measurements for the verification. This is why the spatial distribution of total accumulated yearly precipitation is assessed and compared to the GPCC precipitation data. Bollmeyer *et al.*, 2015 compared the annual precipitation for 2011 of GPCC, COSMO-REA6, COSMO-DS and ERA-Interim (their Figure 5) and found that all reanalyses overestimate the total precipitation amounts over Scandinavia, Russia and Iceland and to smaller extents over Turkey, Scotland and parts of the Alps. Over the remainder of the domain, the total annual precipitation is underestimated compared to GPCC. With eighteen years of reanalysis available, the same comparison is applied to the longer time series. For this, the full data reanalysis of GPCC is used, as explained in section 2.3.4. The overlapping years between GPCC, COSMO-REA6 and ERA-Interim are 1997-2010, so the yearly accumulated precipitation for these years is calculated, averaged and is then compared between the three data sets. The results are shown in Figures 4.3 and 4.4. Figure 4.3 shows the spatial distribution of the mean yearly accumulated precipitation of GPCC, REA6 and ERA-Interim for their respective resolutions of 0.5°, 0.055° and 0.5°. In GPCC the highest precipitation values of up to 3000 mm are located in areas of high latitude, i.e. in Norway, Iceland, the Scottish Highlands, the Dinaric Alps and the Caucasus. In large parts of Europe, precipitation rates of 500-1000 mm are most prevalent. Both REA6 and ERA-Interim show very similar patterns and the prominent precipitation structures can be identified in both reanalyses. For purposes of creating a difference plot COSMO-REA6 was aggregated onto the GPCC grid which is the same grid as for ERA-

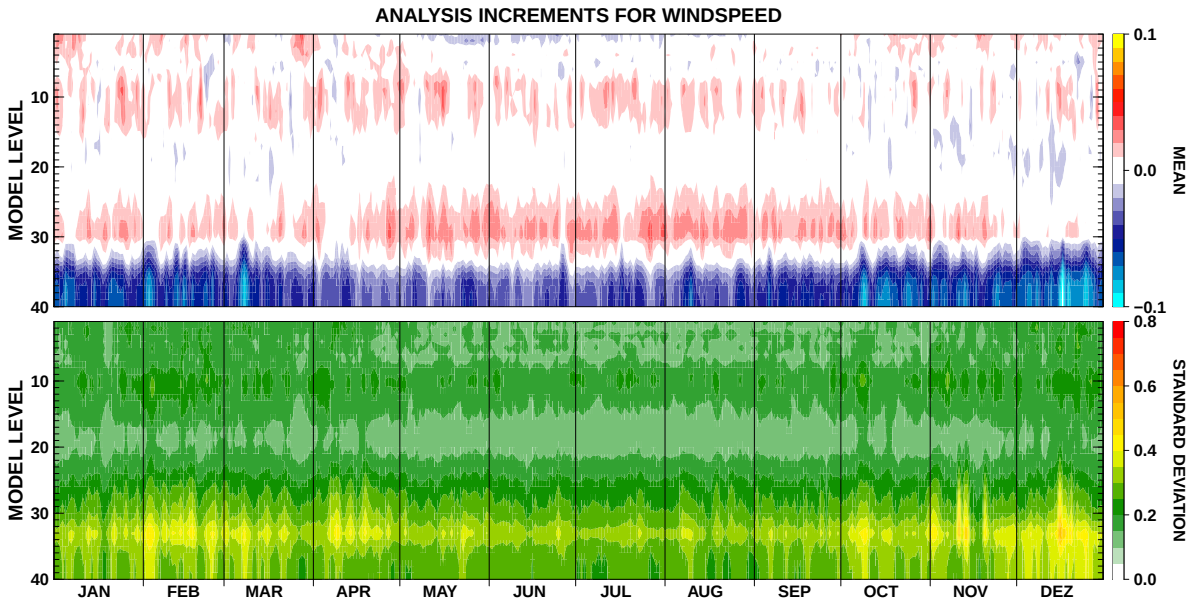


Figure 4.2.: The same as Figure 4.1, but for wind speed in m s^{-1}.

Interim. Figure 4.4 shows the differences from GPCC for both COSMO-REA6 and ERA-Interim and reveals the same features as have been found by Bollmeyer *et al.*, 2015. Precipitation is overestimated over high topography and in parts of Russia and underestimated almost everywhere else. Since this result matches with the findings for the year 2011 this could hint to a problem in either the models or in GPCC. But especially over high terrain, the errors possibly stem from GPCC since no height correction is applied in GPCC, reducing its performance in these areas.

4.2.1. Diurnal cycle

An important aspect of precipitation is the reproduction of the diurnal cycle which is calculated at 1034 rain gauge stations across Germany and calculated separately for winter and summer months. The diurnal cycle is assessed in two ways: once with the precipitation values below 0.1 mm h^{-1} included in the analyses (precipitation sums) and once with these values excluded (precipitation intensities), taking only actual precipitation events into account. The point measurements of the stations are always compared to the corresponding nearest grid point of the different models (COSMO-REA6/REA2/DS and ERA-Interim[1]).

4.2.1.1. LHN experiments

First, the results for the LHN experiments explained in section 2.2.1.5 will be discussed. Figures 4.5 and 4.6 show the diurnal cycle of the precipitation intensities and the precipitation sums for the rain gauge observations, COSMO-REA6 with and without LHN, COSMO-REA2 with and without LHN, COSMO-DS and ERA-Interim for June to August 2011. In the observations, a clear diurnal cycle with a minimum of precipitation intensities around 10 UTC

[1] 3 h precipitation values are divided by 3 to obtain hourly values for ERA-Interim.

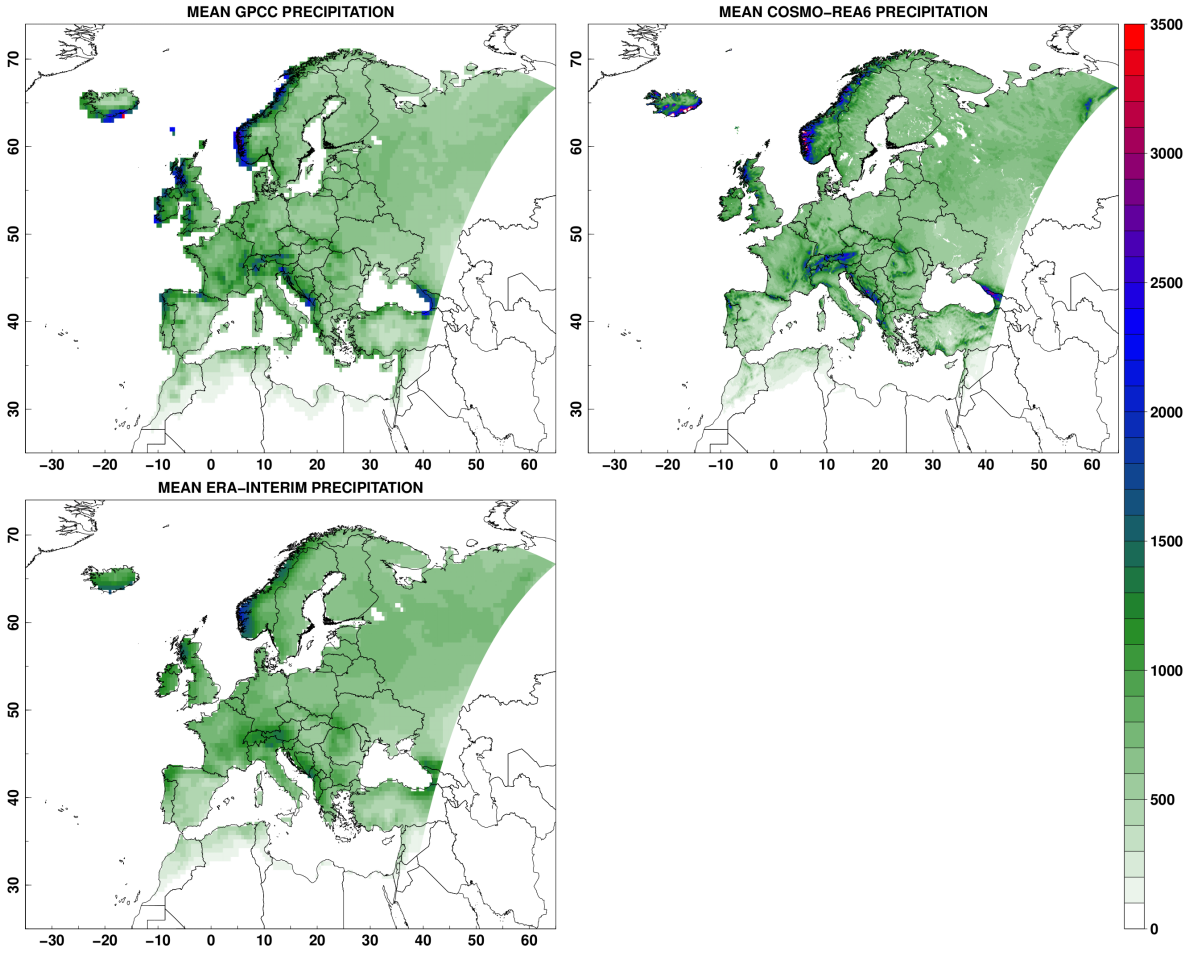


Figure 4.3.: Mean yearly accumulated precipitation for the years 1997-2010 for GPCC, COSMO-REA6 and ERA-Interim in mm.

and a maximum in the early evening between 17 and 18 UTC can be observed which is very similar in the precipitation sums. All COSMO runs are able to reproduce a diurnal cycle, especially in the precipitation intensities, although with different outcomes. In COSMO-REA6 and COSMO-DS, the diurnal cycle of precipitation intensities is similar to the observations in the maximum intensities and a bit too low in the minimum. In the precipitation sums, both COSMO-REA6 and COSMO-DS follow the observations closely in the early day until noon but produce values too low in the afternoon and evening with COSMO-DS showing greater deficiencies than COSMO-REA6. Further, the diurnal cycle in the intensities of both models is deferred in time towards the night, resulting in a minimum around 12/15 UTC and the maximum around 21/22 UTC for COSMO-REA6/DS, respectively. One reason for this is probably that convective rainfall events are initiated too late in the COSMO model, showing deficiencies in the parametrization schemes. The assimilation of radar information via the LHN in COSMO-REA6 improves the representation of the diurnal cycle by shifting the diurnal cycle in the intensities forward in time with the maximum between 19 and 20 UTC and the minimum at 11 UTC, although overestimating the intensities a bit. In the precipitation sums, the diurnal cycle is much better represented with LHN. However, the results are overestimated.

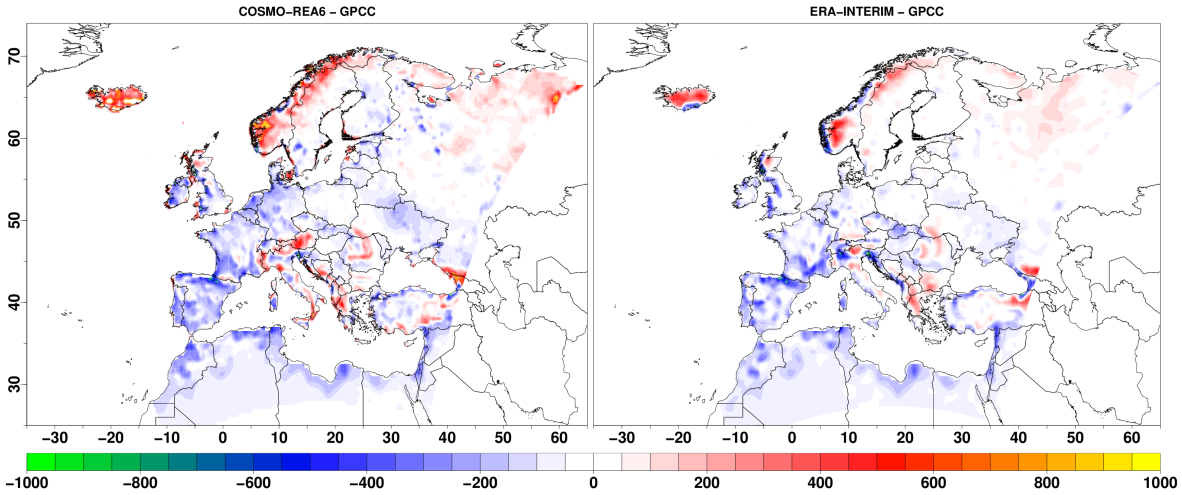


Figure 4.4.: Differences from GPCC in mean yearly accumulated precipitation for the years 1997-2010 for COSMO-REA6 and ERA-Interim in mm.

COSMO-REA2 shows even better results in the diurnal cycle of both precipitation intensities and sums than COSMO-REA6. It is interesting to observe that COSMO-REA2 without the LHN fits the intensity observations best whereas COSMO-REA2 with LHN shows some lower values throughout the day and a slight shift of the maximum towards 20 UTC. In the precipitation sums though, COSMO-REA2 with LHN shows better results between 18 and 23 UTC compared to COSMO-REA2 without LHN. In contrast, ERA-Interim precipitation sums show that the precipitation is produced too early in the morning and ERA-Interim precipitation intensities are significantly lower than the observations which seems to be a natural consequence of the coarser resolution. Current global reanalyses such as ERA-Interim are not designed to simulate distinct precipitation features.
It can be noted that the assimilation of LHN potentially improves the representation of precipitation at 6 km as well as 2 km resolution, although these results are only from one summer. For more solid results further experiments would be needed, especially in the other seasons.

4.2.1.2. Complete reanalysis

Now the results of the complete reanalysis will be presented. As mentioned above, the observed rain gauge measurements are only used for the years 2004-2014 and the COSMO-REA2 results are of course only available for the years 2007-2013. Since the results for the diurnal cycle for COSMO-REA6 are qualitatively the same for the years 2004-2014 as for the years 2007-2013, the analysis is done only on the years 2007-2013 in order to better compare the results with those of COSMO-REA2. The results are shown in Figures 4.7 (intensities) and 4.8 (sums) for both summer (top) and winter months (bottom). The diurnal cycle in intensities during summer is reproduced in both COSMO reanalyses with the diurnal cycle in COSMO-REA6 still being shifted towards late evening and too low minimal values during midday. The diurnal cycle in COSMO-REA2 follows the observations closely although showing too low intensities of 1.2 mm h^{-1} compared to 1.4 mm h^{-1} between 8 and 24 UTC. During winter, both reanalyses follow the observations very closely with no observable diurnal cycle. Apparently, the diurnal cycle stems mostly from convective precipitation events which are not that

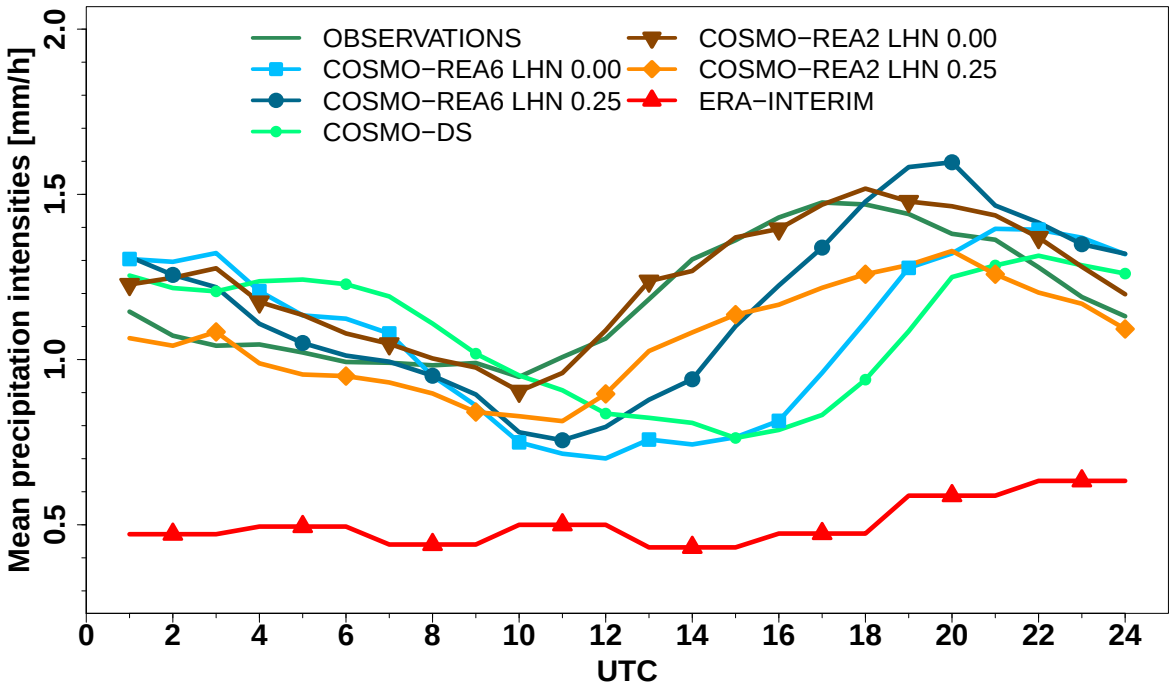


Figure 4.5.: Mean diurnal cycle of precipitation intensities (mm h^{-1}) at 1034 rain gauge stations throughout Germany for JJA 2011. The intensity is calculated only from events above a threshold of 0.1 mm h^{-1}.

distinct during winter. Again, ERA-Interim precipitation intensities and sums are significantly lower and show no real diurnal cycle which is also the case for winter. The observed diurnal cycle in the precipitation sums in summer is represented in COSMO-REA2 very well. COSMO-REA6 shows deficiencies with values too low in the afternoon and evening hours, whereas ERA-Interim shows too high values in the morning and midday and too low values during the afternoon and evening. During winter all reanalyses follow the observations well.

4.2.2. Distribution of precipitation

As a further result, probability distributions of precipitation intensities are presented, where the same rain gauge data as above are used as the observational data set. Histograms for hourly accumulated precipitation rates are calculated for different precipitation intensities and presented in the following.

4.2.2.1. LHN experiments

The histograms for summer 2011 for the LHN experiments are presented in Figure 4.9. The accumulations for the lower intensities (0.1 to 5.0 mm h^{-1} reveal similar relative frequencies for the observations and the different reanalyses experiments. ERA-Interim underestimates the frequency of values below 0.1 mm h^{-1} and strongly overestimates the small values between 0.1 and 1.0 mm h^{-1}. However, it should be noted that this is a comparison of different grid resolutions to point measurements and its significance depends on the application. From a

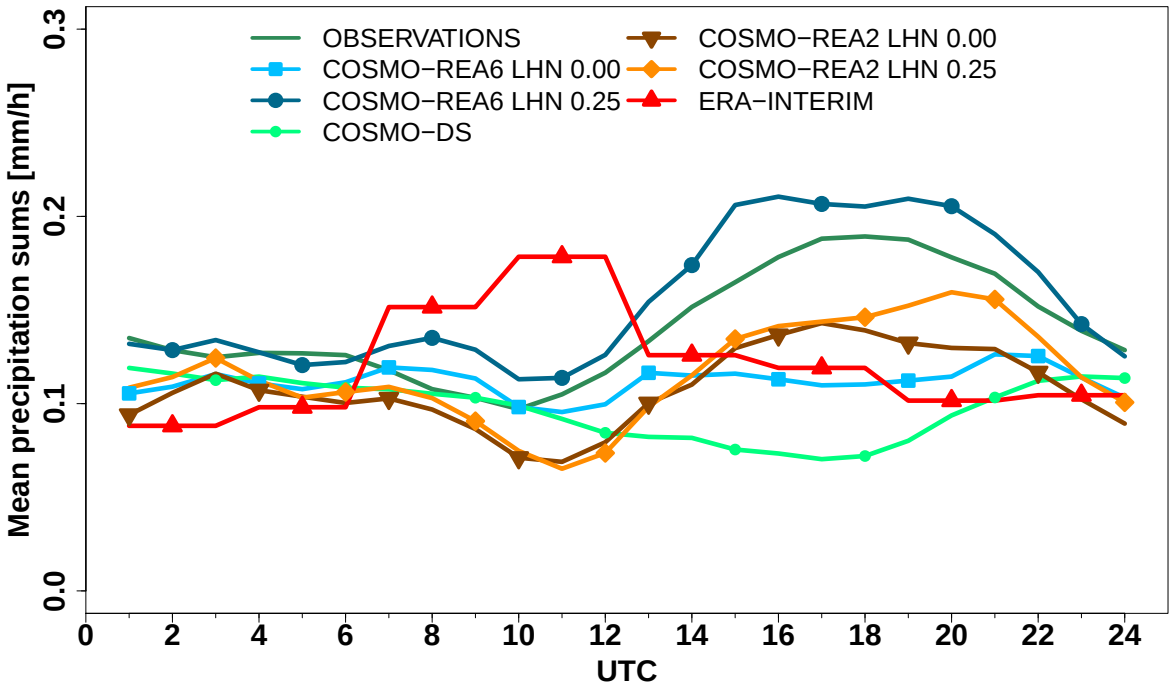


Figure 4.6.: Mean diurnal cycle of precipitation sums (mm h^{-1}) at 1034 rain gauge stations throughout Germany for JJA 2011. The sum is calculated over all hours with events without precipitation included.

modeler's perspective a spatio-temporal aggregation to the coarsest grid would be a prerequisite to assess the quality of the model simulation. Nevertheless, from a user's perspective, the results indicate that a grid-spacing of approximately 6 km respectively 2 km already produces precipitation characteristics that are comparable to point observations. In order to evaluate less frequent heavy precipitation events, the plot on the right of Figure 4.9 focuses the relative frequency histogram on values above 2 mm in one hour. Some differences between the experiments can be observed here. The models without LHN, i.e. COSMO-REA6 and COSMO-REA2 without LHN and COSMO-DS, show too few occurrences of minus 20-25% in all threshold categories. Using the latent heat nudging in COSMO-REA6 leads to enhanced precipitation in all categories, resulting in overestimated amounts between 2.0 and 10.0 mm h^{-1}. Nevertheless, the absolute differences to the observations decrease. The LHN in COSMO-REA2 has the same enhancing effect between 0.1 and 5.0 mm h^{-1} but shows fewer occurrences in the higher categories. The double penalty problem becomes more prominent at 2 km resolution. It occurs at high resolutions when an event, e.g. the occurrence of precipitation, is observed in a grid cell and the model reproduces the precipitation, but in an adjacent grid cell. The model would then be penalized twice, once for not reproducing rain in the observed grid cell and once for producing rain in a grid cell where no precipitation was observed. This double penalty problem could produce some misleading results for COSMO-REA2. ERA-Interim strongly underestimates the higher precipitation values and is not able to produce rain rates larger than 10 mm per hour.

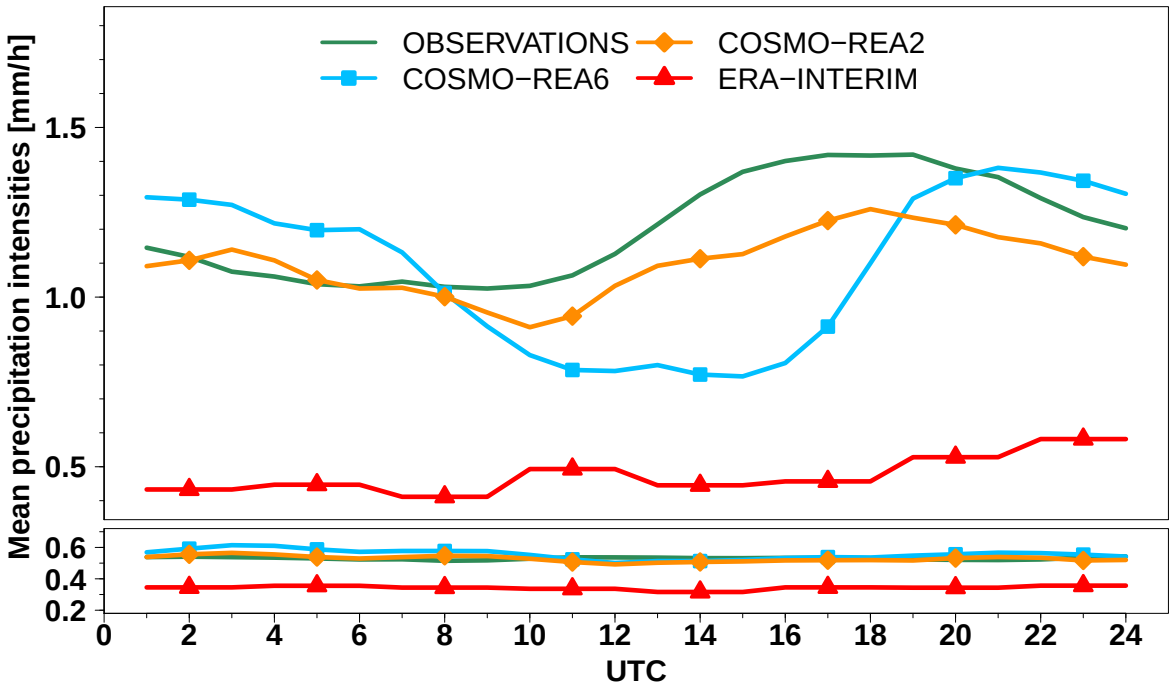


Figure 4.7.: Mean diurnal cycle of precipitation intensities (mm h^{-1}) at 1034 rain gauge stations throughout Germany for summer months (JJA, top) and winter months (DJF, bottom) for the years 2007-2013. The intensity is calculated only from events above a threshold of 0.1 mm h^{-1}.

4.2.2.2. Complete reanalyses

Figure 4.10 shows the histograms for the years 2007-2013 for the observations, COSMO-REA6, COSMO-REA2 and ERA-Interim. Again, ERA-Interim underestimates the events of no precipitation, overestimates smaller rain rates and fails in simulating the higher ones. Both COSMO reanalyses show a slight underestimation of events smaller than 0.1 mm h^{-1} and a slight overestimation of rain rates between 0.1 and 1.0 mm per hour. Principally, COSMO-REA2 shows more occurrences in all categories than COSMO-REA6, leading to a better representation of the high rain rates of 2 mm per hour and above, where COSMO-REA6 is always underestimating the frequencies.

4.2.2.3. Bias and log-odds ratio

In order to further evaluate precipitation in the reanalyses, correspondence between modeled and observed precipitation is assessed using a contingency table. It is build upon the rain gauge observations and the corresponding values of the different models, respectively, classifying precipitation according to its exceedance over a certain threshold as a yes/no event. The contingency table is filled with hits a, false alarms b, misses c and correct rejections d. The contingency tables are assessed using two measures, namely the frequency bias and the log-odds ratio. The bias in such a contingency table is defined as the ratio between the number of

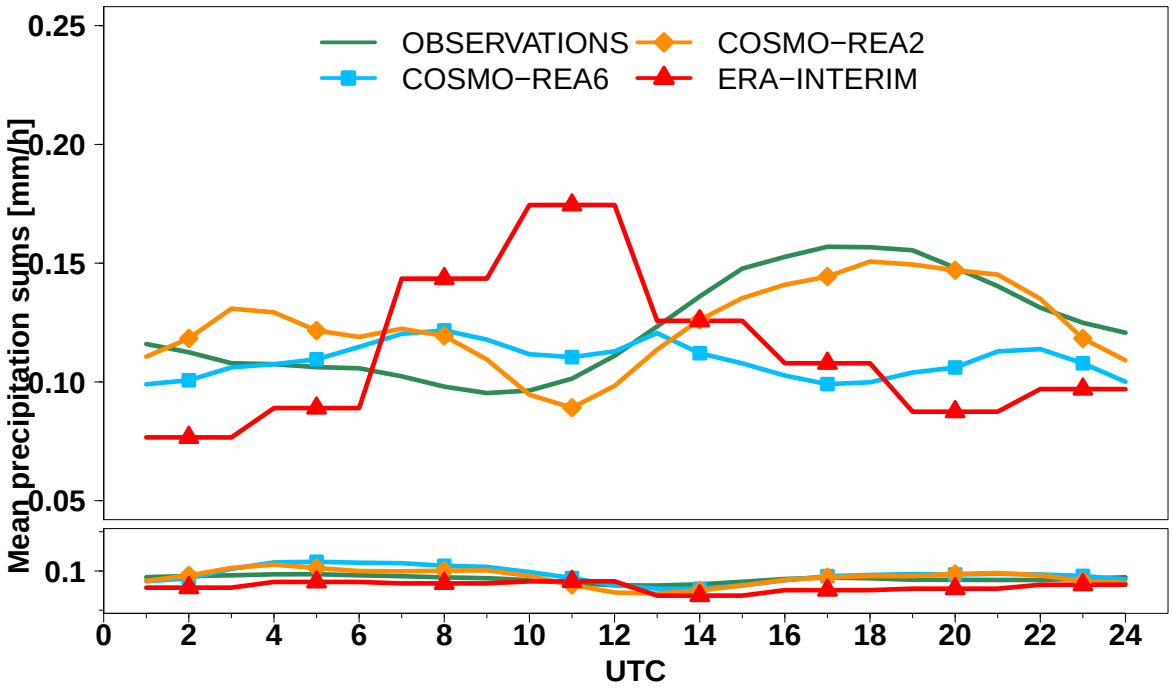


Figure 4.8.: Mean diurnal cycle of precipitation sums (mm h^{-1}) at 1034 rain gauge stations throughout Germany for summer months (JJA, top) and winter months (DJF, bottom) for the years 2007-2013. The sum is calculated over all hours with events without precipitation included.

modeled and observed (SYNOP) events (see e.g. Jolliffe and Stephenson, 2012).

$$BIAS = \frac{a+b}{a+c}. \tag{4.143}$$

A bias of one indicates that every observed event is modeled, i.e. indicates a perfect reproduction of the events, and values above (below) one indicate that more (less) events are modeled than observed. Contrary to the bias which assesses the climatology of precipitation, the log-odds ratio also assesses the skill of the precipitation estimate, i.e. the agreement between timing and location of precipitation events. It is calculated from the contingency table as

$$\log\theta = \log(\frac{a\cdot d}{b\cdot d}) \tag{4.144}$$

or, if written in terms of hit rate $H = \frac{a}{a+c}$ and false alarm rate $F = \frac{b}{b+d}$ as

$$\log\theta = \log(\frac{H}{1-H}\cdot\frac{1-F}{F}). \tag{4.145}$$

The log-odds ratio is greater than zero if the hit rate exceeds the false alarm rate, see e.g. Stephenson, 2000. Values larger than one indicate a significantly skillful reanalysis, whereas values smaller than one show that the reanalysis is no more skillful than a random forecast. So using both bias and log-odds ratio together helps in identifying a skillful reanalysis.
Figures 4.11-4.13 show maps for both the bias and the log-odds ratio across Germany for

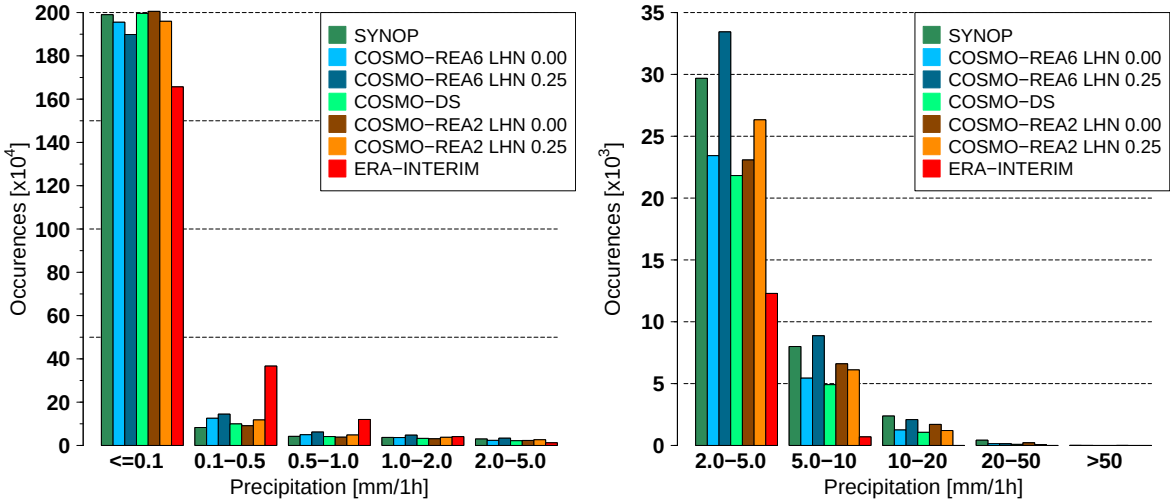


Figure 4.9.: Histograms of precipitation events for different thresholds for SYNOP-Stations, COSMO-REA6 with and without LHN, COSMO-DS, COSMO-REA2 with and without LHN and ERA-Interim, evaluated at 1034 stations across Germany and at the corresponding nearest grid points for JJA 2011.

COSMO-REA6, COSMO-REA2 and ERA-Interim for three different thresholds, 0.10, 0.50 and 1.00 mm h^{-1}. The bias for COSMO-REA6 is about one in the lowest threshold, decreasing slightly with increasing threshold. In COSMO-REA2, the bias starts around one with slightly positive values in central Germany and decreases with increasing thresholds but stays slightly closer to one than COSMO-REA6. This confirms the good representation of precipitation in these models, but also indicates a slight underestimation of rainfall events with higher rain rates and is therefore in good agreement with Figure 4.10. ERA-Interim, however, simulates too many precipitation events above 0.10 mm h^{-1}, resulting in bias values of two and above, and shows a strong underestimation of higher rain rates above 1.00 mm h^{-1} with bias values around 0.5.

Considering the log-odds ratio, all models show a skillful analysis with no values less than 2.0. COSMO-REA6 and ERA-Interim show very similar structures with values between 2.7 and 3.3 for the 0.10 mm h^{-1} threshold and increasing values up to 4.0. COSMO-REA2 is characterized by the highest values in all thresholds starting from values of $3-3.5$ and increasing up to values of 5.0. The increasing values in the log-odds ratio can be explained by the decreasing number of actual events which increases the numerator in Eq. (4.144). The log-odds ratio can also be tested for significance (see Stephenson, 2000) when calculating the standard error of the log-odds ratio which is given by $1/(n_h)^{1/2}$, where n_h is the effective number of degrees of freedom $1/n_h = 1/a+1/b+1/c+1/d$. When the log-odds ratio values are larger than two times the standard error then there is 95% confidence that the skill does not stem from pure chance. The values for the standard error are in the range of $0-0.083$ for the 0.10 mm h^{-1} threshold and in the range of $0-0.247$ for the 1.00 mm h^{-1} threshold with only minor differences between the three models. This shows that all models are significant in the log-odds ratio.

Bollmeyer *et al.*, 2015 have found a similar behaviour of COSMO-REA6 and ERA-Interim for the year 2011 and showed that COSMO-DS, while having similar results in the bias compared to COSMO-REA6, has significantly lower values around 2 in the log-odds ratio. So although

the downscaling performs as good as the reanalysis when climatological precipitation fields are considered, a closer look at the spatial–temporal distribution reveals that the use of the data assimilation has a large impact and improves the analysis of those fields. However, only COSMO-REA6 and COSMO-REA2 are able to improve both attributes of performance: measures of accuracy, e.g. log-odds ratio, as well as measures of bias, e.g. the frequency bias with COSMO-REA2 being superior to COSMO-REA6.

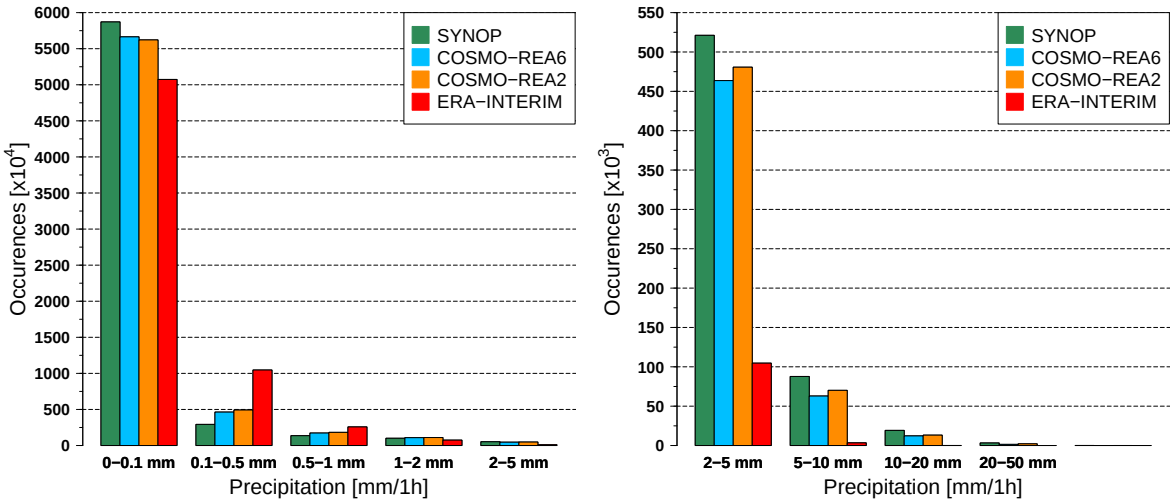


Figure 4.10.: Histograms of precipitation events for different thresholds for SYNOP-Stations, COSMO-REA6, COSMO-REA2 and ERA-Interim, evaluated at 1034 stations across Germany and at the corresponding nearest grid points for the years 2007-2013.

4.3. Radiation

The radiative fluxes are the main drivers of the energy cycle in the atmosphere. They are divided into shortwave and longwave radiation. Shortwave radiation lies in the visible part of the radiative spectrum and is essentially the solar insolation. It is incoming at the top of the atmosphere and is scattered by clouds and the atmosphere and absorbed and partly reflected back into space by the surface, the clouds and the atmosphere. Longwave radiation on the other hand is radiation in the infrared part of the spectrum and comprises the emitted radiation of the Earth‘s surface, the clouds and the atmosphere. The sensible and latent heat fluxes close the global energy budget (Trenberth *et al.*, 2009). A correct specification of the longwave and shortwave radiation at both the surface and the top of atmosphere is therefore of importance for the correct representation of energy in the model. For this reason, those fluxes are compared between COSMO-REA6 and CERES-EBAF data for the years 2001 to 2013. For the difference plots, COSMO-REA6 has been aggregated to the CERES-EBAF grid and COSMO-REA6 minus CERES-EBAF is depicted.
In Figure 4.14 the shortwave, longwave and net radiation at the surface for both CERES-EBAF and COSMO-REA6 as well as their difference is shown. Note that fluxes directed towards the surface are counted positive and fluxes directed away from the surface are counted negative. In CERES-EBAF a clear north-south gradient in shortwave radiation can be observed with the smallest values of 50 W m^{-2} in the northern Atlantic and the highest values of 246 W m^{-2}

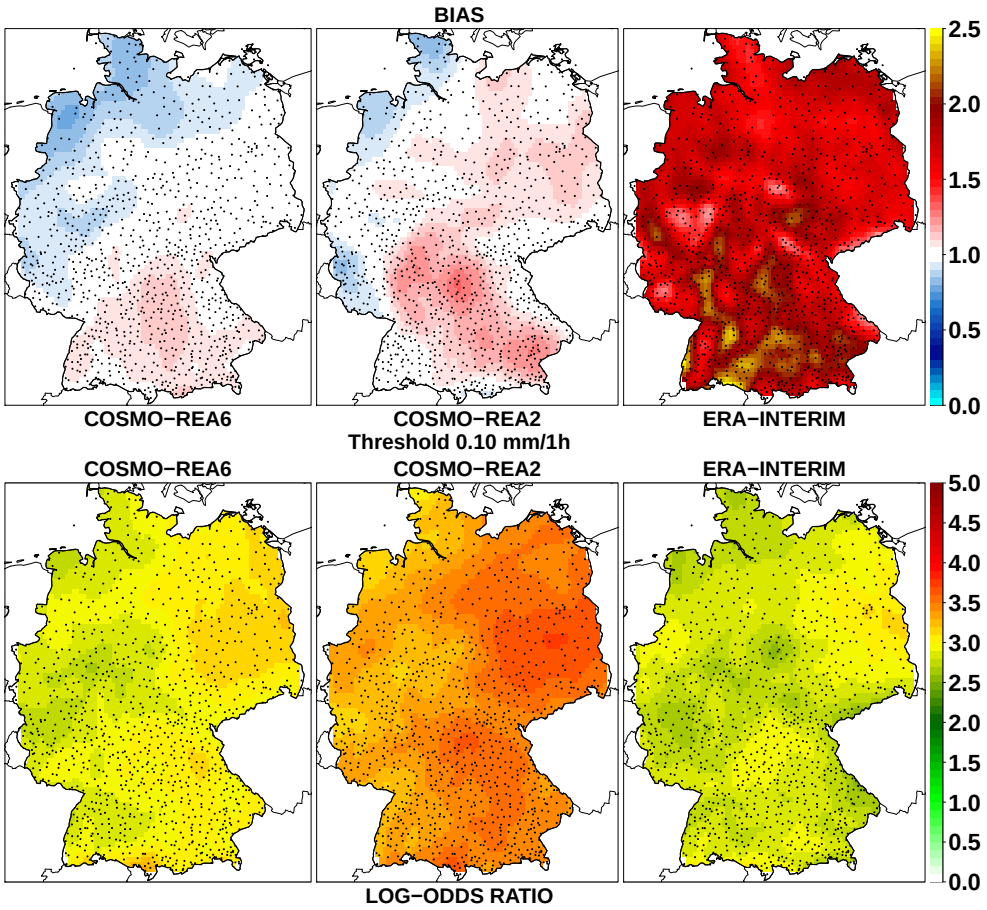


Figure 4.11.: Frequency bias (top) and log-odds ratio (bottom) of precipitation events between SYNOP stations and COSMO-REA6, COSMO-REA2 and ERA-Interim (from left to right) for a threshold value of 0.10 mm h^{-1} for the years 2007-2013.

in the Red Sea. High values between 175 and 220 W m^{-2} are present in the Mediterranean Sea, where the incoming solar radiation is strongly absorbed. In Northern Africa net shortwave radiation is lower with 110 to 190 W m^{-2} due to the comparably high albedo values and the different capacity of heat storage in the ocean. The overall structure of the shortwave radiation at the surface is reproduced by COSMO-REA6 fairly well, especially for latitudes north of approximately 45°. In the Mediterranean Sea, shortwave radiation values are about 20 W m^{-2} lower compared with CERES-EBAF and between $20-40$ W m^{-2} higher over Africa. The differences over the Mediterranean Sea can probably be explained by a wrong sea surface temperature in COSMO-REA6. Over Africa, the differences stem in part from different albedo values specified in COSMO. In addition, some obscure features are visible in COSMO-REA6 in Africa with sharp changes in the radiation values. These features coincide with changes in the soiltype in COSMO, where the lower values are sandy areas and the areas with higher values are clay types. The surface albedo in COSMO is a function of soiltype, which explains the sharp jumps at a change in the soiltype. The longwave radiation at the surface, Figure 4.14 b), shows also a north-south gradient with small outgoing values of around -20 W m^{-2} in northern Europe and the highest values between -90 and -110 W m^{-2} in northern Africa with peaks up to -128 W m^{-2}. What catches the eye is a jump towards lesser values at exactly 60°N in CERES-

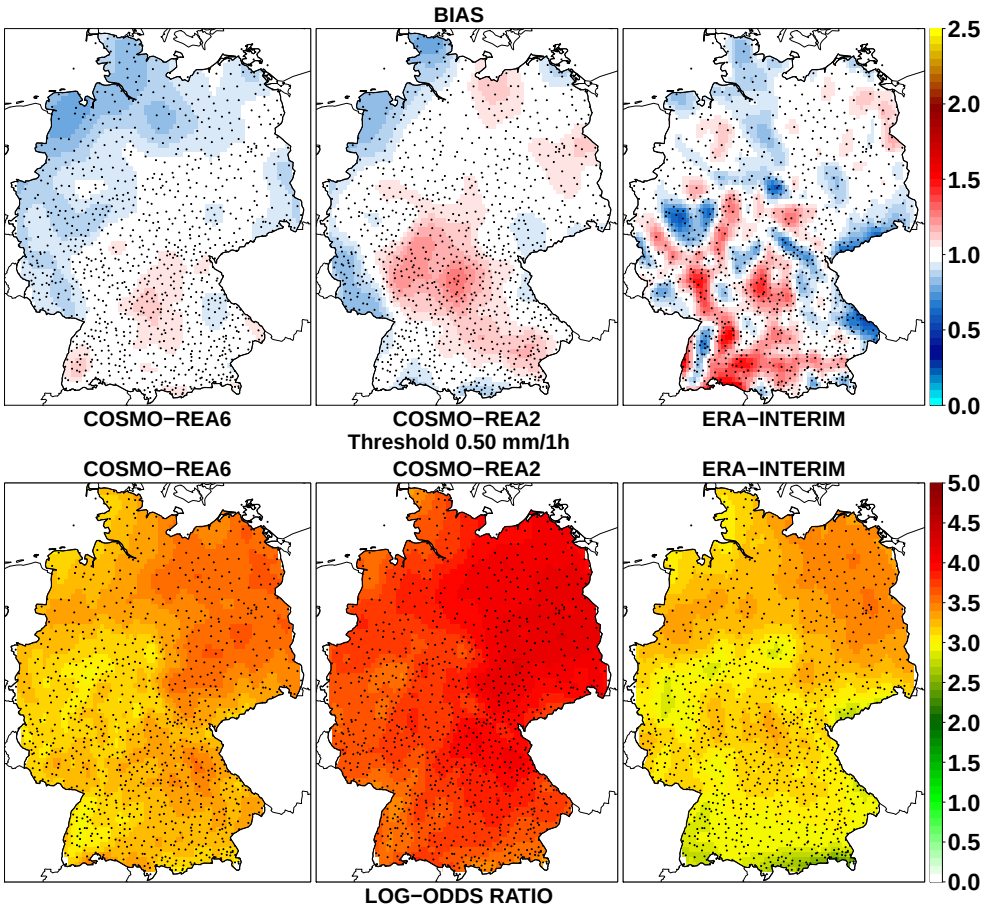


Figure 4.12.: Frequency bias (top) and log-odds ratio (bottom) of precipitation events between SYNOP stations and COSMO-REA6, COSMO-REA2 and ERA-Interim (from left to right) for a threshold value of 0.50 mm $h^{-1}$for the years 2007-2013.

EBAF. This can be explained by a switch from a geostationary to a non-geostationary satellite in the processing of the radiative fluxes in CERES-EBAF (Loeb *et al.*, 2009). COSMO-REA6 again shows very similar structures but produces higher values in outgoing longwave radiation over Africa between -100 and -140 W m^{-2}. The observed features over Africa which are visible in the shortwave cannot be observed in the longwave, which supports the hypothesis of the different albedo for the different soiltypes because longwave radiation is not significantly affected by the albedo. The differences between CERES-EBAF and COSMO-REA6 in the longwave radiation at the surface are in the range of -10 W m^{-2} in most parts of the domain and reach up to -41 W m^{-2} in the Sahara. Conclusively, the net radiation at the surface in Figure 4.14 c) reveals differences between -10 and -25 W m^{-2} over the oceans and north of 60° N with more incoming radiation in CERES-EBAF. In the Mediterranean Sea, the differences increase to about -40 W m^{-2}. Values between ± 5 W m^{-2} are observed in Mideurope whereas over Africa too much radiation in COSMO-REA6 is observed with a difference of 20 W m^{-2} compared to CERES-EBAF.

Except for Africa, the shortwave radiation values show negative differences, i.e. CERES-EBAF has larger values in incoming radiation and smaller values in outgoing radiation than COSMO-REA6. Some possible reasons are the following. First of all, the albedo values at the surface are

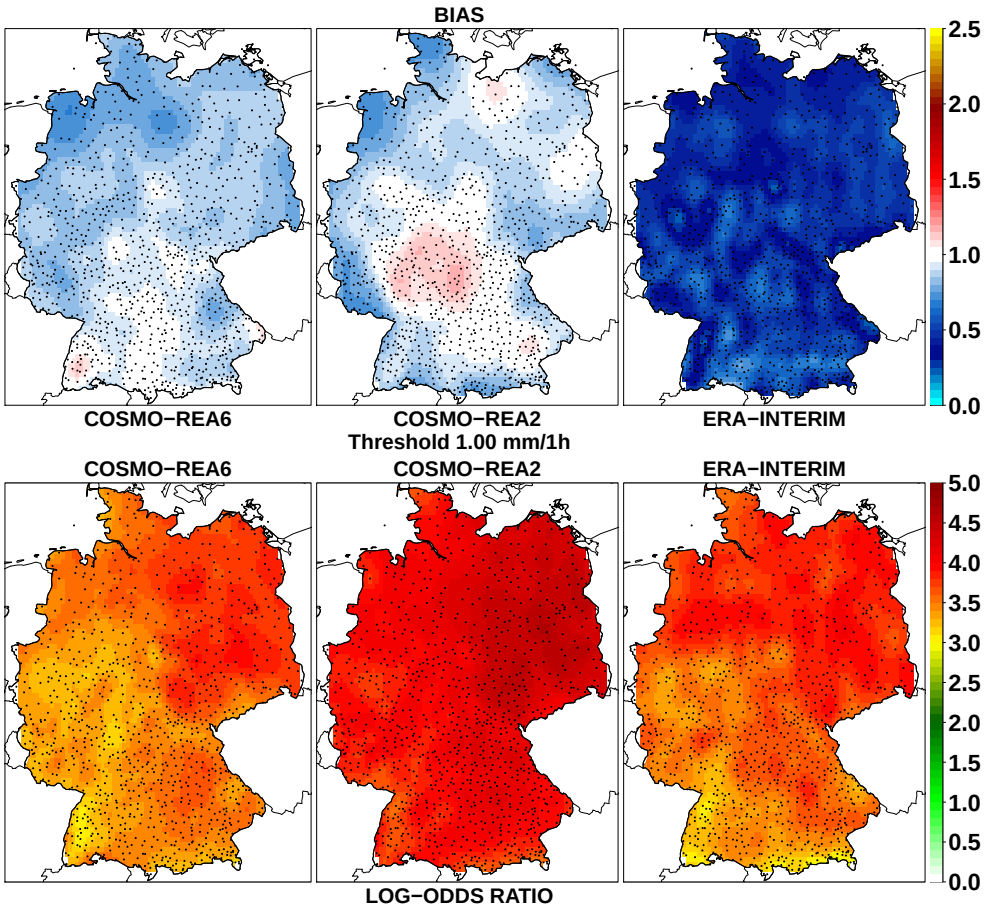
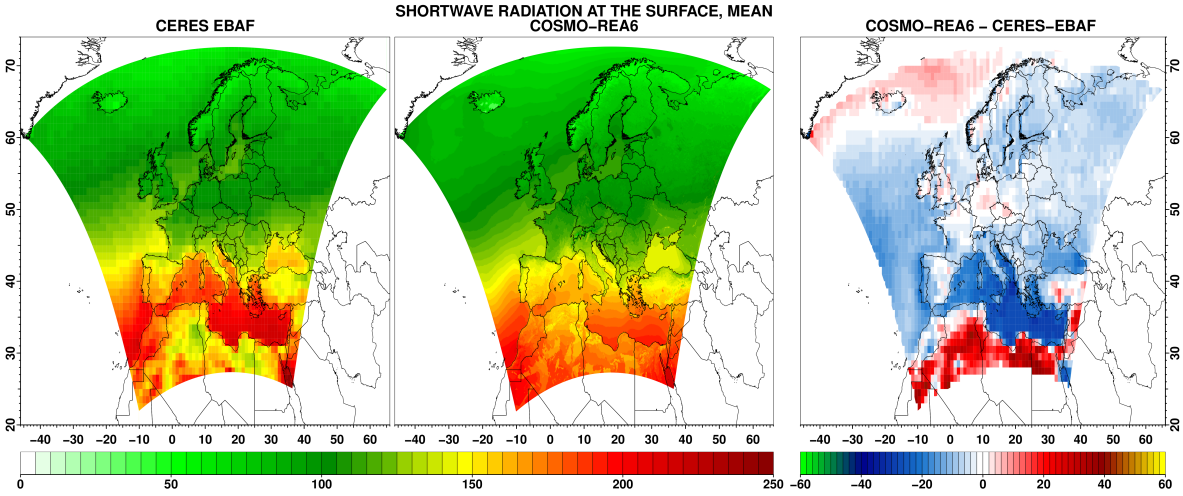
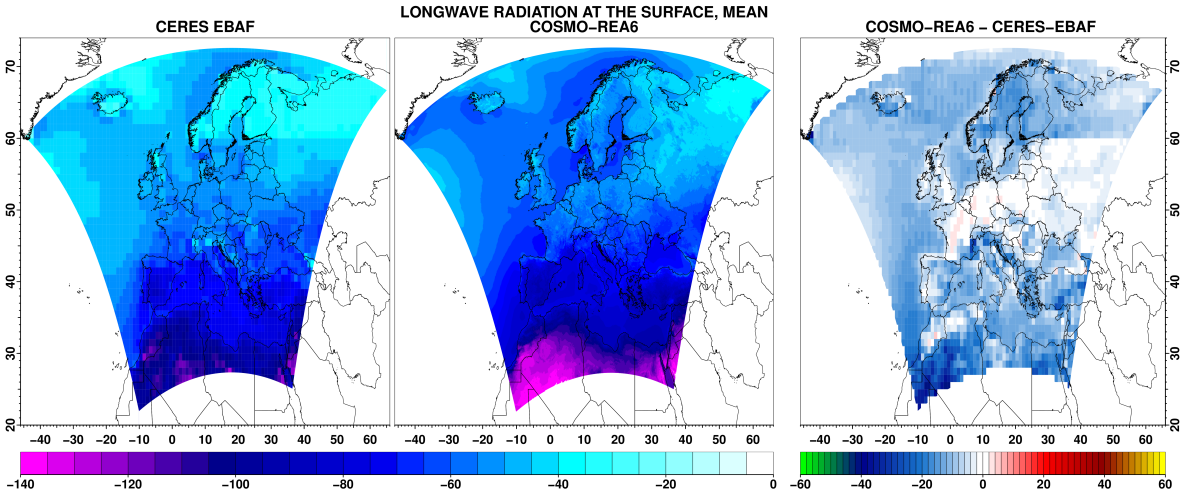


Figure 4.13.: Frequency bias (top) and log-odds ratio (bottom) of precipitation events between SYNOP stations and COSMO-REA6, COSMO-REA2 and ERA-Interim (from left to right) for a threshold value of 1.00 mm h^{-1} for the years 2007-2013.

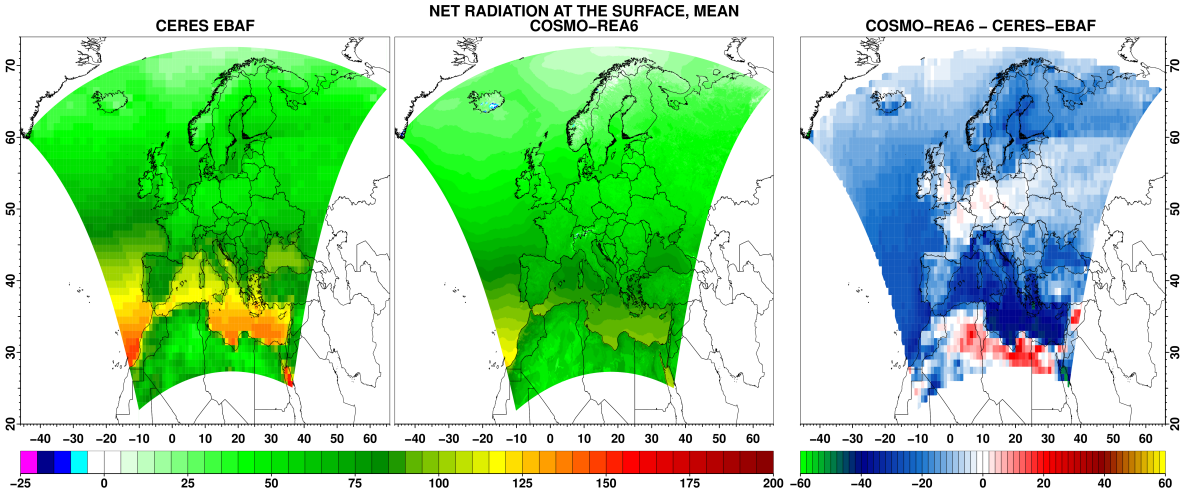
different in the two models. E.g., in COSMO-REA6 surface albedo is a function of soiltype and is only slightly changed by the radiation parametrization over land and not at all over water faces. The surface albedo of water faces is a constant value of 0.07 in the COSMO model. In CERES-EBAF on the other hand a scene-dependent diurnal albedo model for the computation of albedo changes with solar zenith angle is used. This could of course lead to differences especially in the reflected shortwave radiation. Especially over the Mediterranean Sea the differences between COSMO-REA6 and CERES-EBAF are quite large which could be caused by a bias in sea surface temperatures in COSMO-REA6. Since COSMO-REA6 is based on the setup of the operational NWP production, sea surface temperature is not changed during the COSMO runs but is kept constant till the next SST analysis which uses sea surface temperatures from ERA-Interim. A further evaluation of sea surface temperatures would be advantageous in this regard. Another important aspect is the calculation of the surface fluxes from measured radiances at satellite instruments. Since the satellites measure the radiances high over the atmosphere, every surface radiation product relies on radiative transfer calculations. For those calculations, temperature and humidity profiles of the atmosphere need to be prescribed (Kato *et al.*, 2013). In CERES-EBAF the temperature and humidity profiles are taken from the Goddard Earth Observing System (GEOS-4 and GEOS-5) Data Assimilation System reanalysis



(a)



(b)



(c)

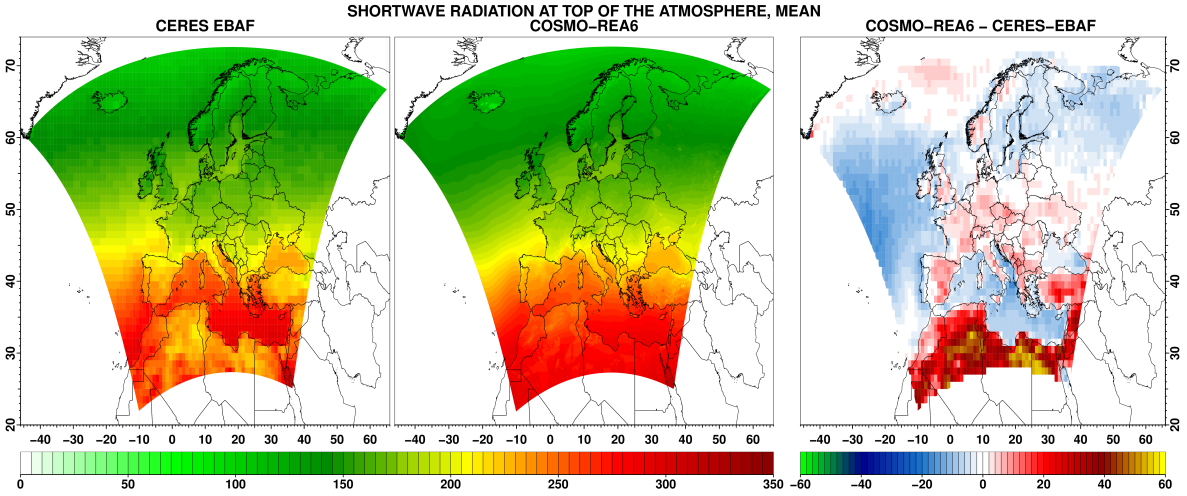
Figure 4.14.: Mean net radiation at the surface for a) shortwave radiation, b) longwave radiation and c) net, i.e. shortwave+longwave, radiation, for the years 2001-2013 for CERES-EBAF and COSMO-REA6 and the difference from CERES-EBAF in $\mathrm{W\ m^{-2}}$.

(Rienecker *et al.*, 2008) which is also used by the MERRA reanalysis (Rienecker *et al.*, 2011). These profiles have a temporal resolution of 6 h and the skin temperatures used from GEOS-4 and GEOS-5 have a 3-hourly resolution. The differences in both the data assimilation system and the temporal resolutions used can explain part of the differences in the surface radiation between CERES-EBAF and COSMO-REA6. Further, Kato *et al.*, 2013 point out that the estimated uncertainty in the radiation fluxes is in the range of about $9-12$ W m^{-2} for shortwave and $12-19$ W m^{-2} for longwave fluxes. So differences in surface radiation between COSMO-REA6 and CERES-EBAF in the range of 10 to 15 W m^{-2} are within the measurement and processing uncertainty of CERES-EBAF. Only the shortwave values over Africa show clear deficiencies in COSMO-REA6 which need to be further investigated as albedo differences can only explain parts of these high differences.

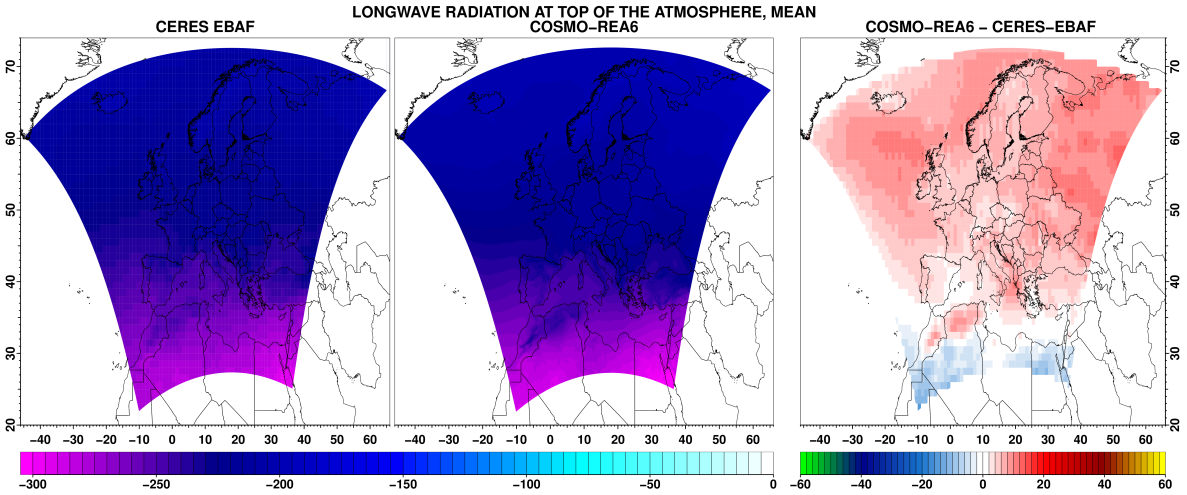
The radiative fluxes at the top of the atmosphere (TOA) are presented in Figure 4.15. The shortwave radiation at the TOA is very similar in both CERES-EBAF and COSMO-REA6 with values ranging from 57 to 333 W m^{-2} with increasing values from north to south. The highest values in CERES-EBAF can be observed over the Mediterranean and Red Sea whereas in COSMO-REA6 comparably high values are present not only over those seas but also over Africa as is the case in the shortwave radiation at the surface as well. The differences over the Mediterranean Sea and the Atlantic are slightly negative with values between -5 and -10 W m^{-2} and slightly positive around 10 W m^{-2} over the continent and the Norwegian Sea. Over Africa however, the differences are unrealistically high between 30 and 60 W m^{-2}. Values up to 20 W m^{-2} can also be found over the Middle East and Turkey.

Longwave radiation at the TOA ranges from -296 (-305) to -202 W m^{-2} (-190 W m^{-2}) in CERES-EBAF (COSMO-REA6). The differences are quite small with positive values between 0 and 10 W m^{-2} north of about 35° N and only slightly negative values over Africa. As a consequence of the longwave and shortwave radiation, the net radiation at the TOA shows a good representation in the overall structure over Europe, but large differences up to 60 W m^{-2} over Africa. Small negative differences can be seen at the western boundary of the domain and over the Mediterranean Sea and positive differences between 0 and 18 W m^{-2} are present over the rest of the domain.

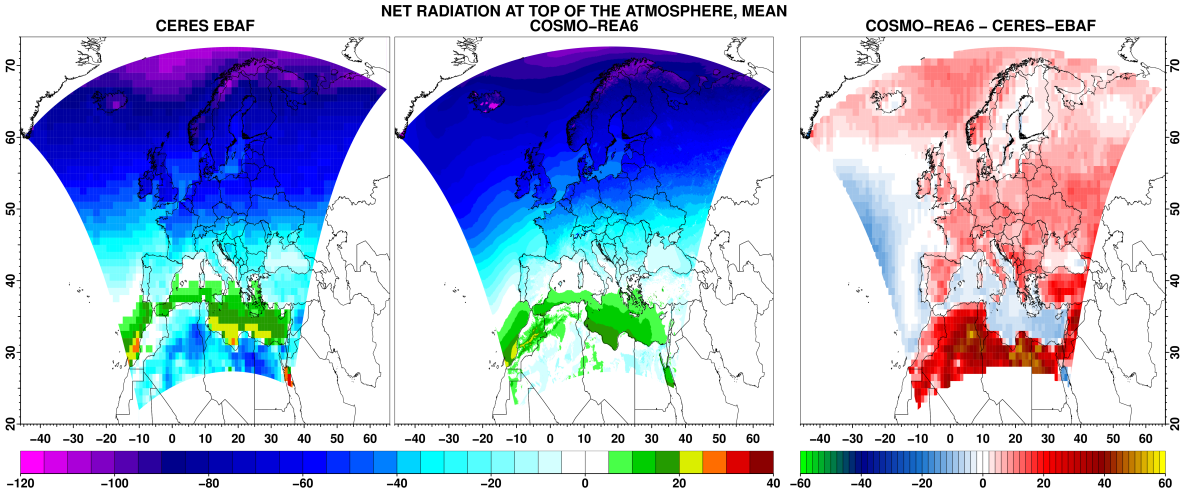
As mentioned by Loeb *et al.*, 2009, a solar constant of 1365 W m^{-2} is assumed in CERES-EBAF whereas in COSMO-REA6 the solar constant is set to 1368 W m^{-2}, which contributes to the differences in the TOA radiation. Kopp *et al.*, 2005 revised the value of the solar constant as being 1361 W m^{-2}, which should be accounted for in the COSMO model in the future. Loeb *et al.*, 2009 further compare CERES-EBAF with other satellite-based data products and show imbalances in the global net TOA radiation of -3 to 7 W m^{-2}, which are even more pronounced on a regional domain. The reason for this are uncertainties in the absolute calibration of the measurements. Another explanation for the discrepancies between CERES-EBAF and COSMO-REA6 in the longwave part of the spectrum is that the COSMO model produces too high clouds. Böhme *et al.*, 2011 found out that the COSMO model produces high clouds too often and thus lower temperatures of the cloud tops than observed by MSG satellites. The same result was found by Bollmeyer *et al.*, 2015 for the actual model version used for producing the reanalysis. Another matter found by Crewell *et al.*, 2008 is the dry bias of radiosonde ascents. Since radiosonde measurements are the main data source for the assimilation in the upper troposphere they could introduce a dry bias in this part of the atmosphere, leading to higher temperatures as observed by Böhme *et al.*, 2011. Despite the deficiencies of the COSMO model in those regards, the differences in TOA radiation are in the range of $10-15$ W m^{-2} which is again in the range of the uncertainty of CERES-EBAF. Only the Africa-related data introduce large errors



(a)



(b)



(c)

Figure 4.15.: Mean net radiation at the top of the atmosphere for a) shortwave radiation, b) longwave radiation and c) net, i.e. shortwave+longwave, radiation, for the years 2001-2013 for CERES-EBAF and COSMO-REA6 and the difference from CERES-EBAF in W m^{-2}.

which need a deeper analysis.

4.4. Climate classification using Köppen-Geiger maps

In Section 2.4, the climate classifications by Köppen were introduced. The climate classifications were calculated for COSMO-REA6, ERA-Interim and the E-OBS dataset together for the overlapping years 1997-2013. They are presented along with the map by Kottek *et al.*, 2006 in Figure 4.16. The maps are bounded by the COSMO-REA6 area. In the Köppen-Geiger maps from Kottek *et al.*, 2006, shown at the top left, warm temperate climates (C) can be found in Mid- and South-Europe, except for the Alps, and in mediterranean coastal regions of Africa. The rest of Africa shows arid climates (B). The continental land mass east of about 20°E and north of about 45°N as well as Scandinavia have snow climate conditions (D), except for high mountainous regions in Scandinavia where there are polar conditions (E). Iceland also shows polar climate conditions except for some points at the coast with warm temperate conditions. The precipitation regimes can be roughly divided into three parts: All regions north of about 45°N are fully humid (f), regions adjacent to the Mediterranean sea are summer dry regions (s) and Africa is dominated by desert conditions (W). The temperature conditions are more distinct: In latitude regions between 45° and 60°N warm summer conditions (b) apply. North of about 60° the warm summer conditions change to cool summer and polar tundra conditions can be found in high mountainous regions of Scandinavia and on Iceland. East of Ukraine regions with hot summers (a) can be found as well as in the mediterranean countries. Some regions around the Black Sea are classified as hot summer regions as well. Hot arid conditions can be found almost everywhere in North-Africa with some cold arid regions in Morocco and Algeria. Some small scale features can be found, like warm summer regions in northern Spain and Portugal instead of hot summer regions in the south.
Comparing the COSMO-REA6 conditions, taken from 17 years of reanalysis output, with the 50 year record, it is found that the overall climate conditions in Europe are very well reproduced in COSMO-REA6. The boundaries of the climate conditions are very well met, especially at 60°N between the warm and cool summer conditions and at approximately 20°E between the warm temperate and snow climate. Even small-scale features like the conditions in mountainous regions in Scandinavia and in the Alps can be found in COSMO-REA6. Of course, some differences can still be found. E.g. in eastern Spain, COSMO-REA6 shows large regions with cold arid steppe climates. ERA-Interim shows very similar features but due to the coarser resolution small-scale patterns like the polar regions in the Alps and in Scandinavia are not reproduced. Only the E-OBS dataset shows very different features, especially in the precipitation conditions. Between 45°N and 60°N, winter dry conditions instead of fully humid conditions can be found for the precipitation conditions whereas the temperature conditions are the same as in the 50 year record, COSMO-REA6 and ERA-Interim. This could point to a problem in E-OBS with the precipitation measurements or the timing of the observations. Haylock *et al.*, 2008 note that in the observations underlying the E-OBS dataset it was often not clear whether a measurement of 24h precipitation recorded as e.g. 1960-01-02 corresponded to the precipitation that fell on that date or that the observation was taken on that date, mostly at 09 UTC. They tried to correct for this fact by shifting the dates by ± 1 day and finding the date with the highest correlation with ERA-40 data. Another aspect could be that in E-OBS the gridded data sets of precipitation and temperature were created independently from each other. ERA-Interim and COSMO-REA6 are consistent between all output parameters due to the producing

models and this consistency between precipitation and temperature is lacking in E-OBS. This comparison of the climate conditions can be viewed as an independent verification of the performance of the reanalysis over a long time period and it becomes clear that the model is performing reasonable and reproduces the correct climate conditions for the European region.

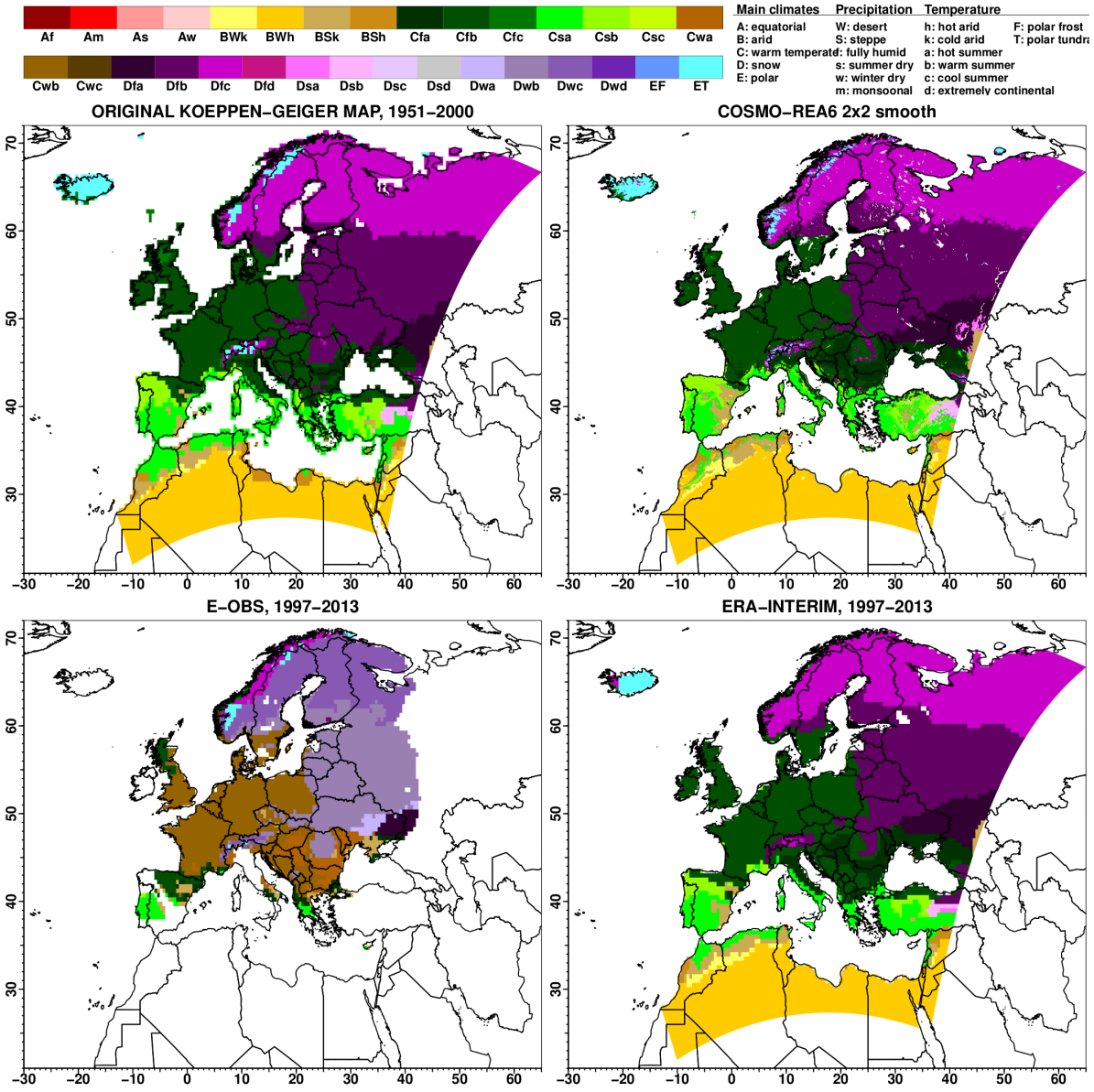


Figure 4.16.: Climate classification using the criteria proposed by Köppen, 1918. Top left: The original Köppen-Geiger map from Kottek *et al.*, 2006, based on the years 1951-2000. Top right: The Köppen-Geiger map from COSMO-REA6 output, based on the years 1997-2013. Bottom left: The Köppen-Geiger map from E-OBS data, based on the years 1997-2013. Bottom right: The Köppen-Geiger map from ERA-Interim data, based on the years 1997-2013.

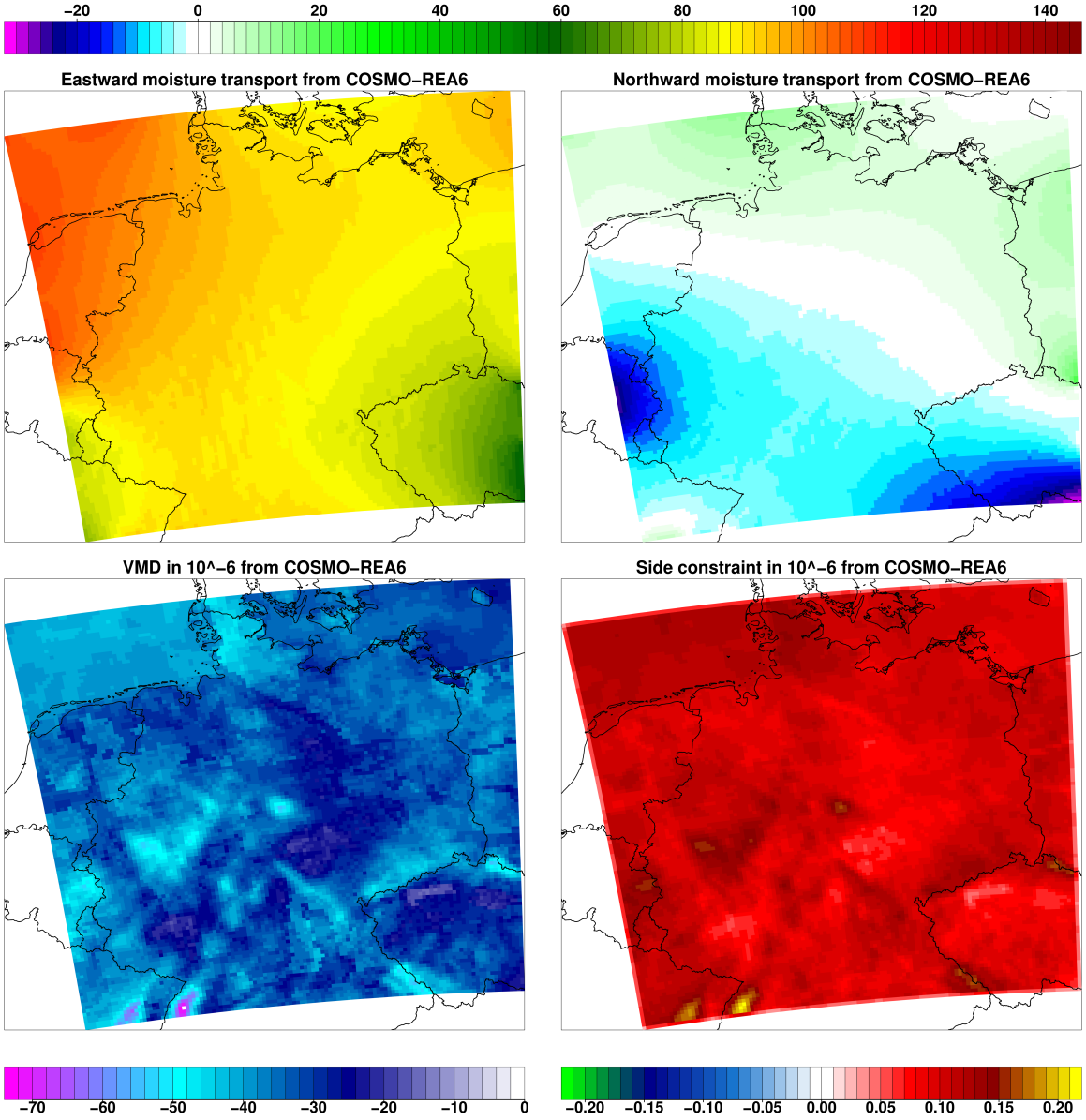


Figure 4.17.: The observed eastward and northward moisture transports (top left and top right) in kg m^{-1} s^{-1}, the observed vertically integrated moisture flux divergence in 10^{-6} kg m^{-2} s^{-1} (bottom left) and the observed difference between the divergence of the moisture transports and the VMD in 10^{-6} kg m^{-2} s^{-1} (bottom right).

4.5. Variational approach for the moisture budget

Now the results from the variational approach will be presented. As mentioned in Chapter 3, the computation was done on a (120×120) domain which was chosen to cover the north of Germany. Figure 4.17 shows the eastward and northward moisture transports together with the vertically integrated moisture flux divergence and the difference between the moisture transport divergence and the VMD. For convenience, only the divergent part of the horizontal moisture transports is shown, since the variational approach only modifies and returns

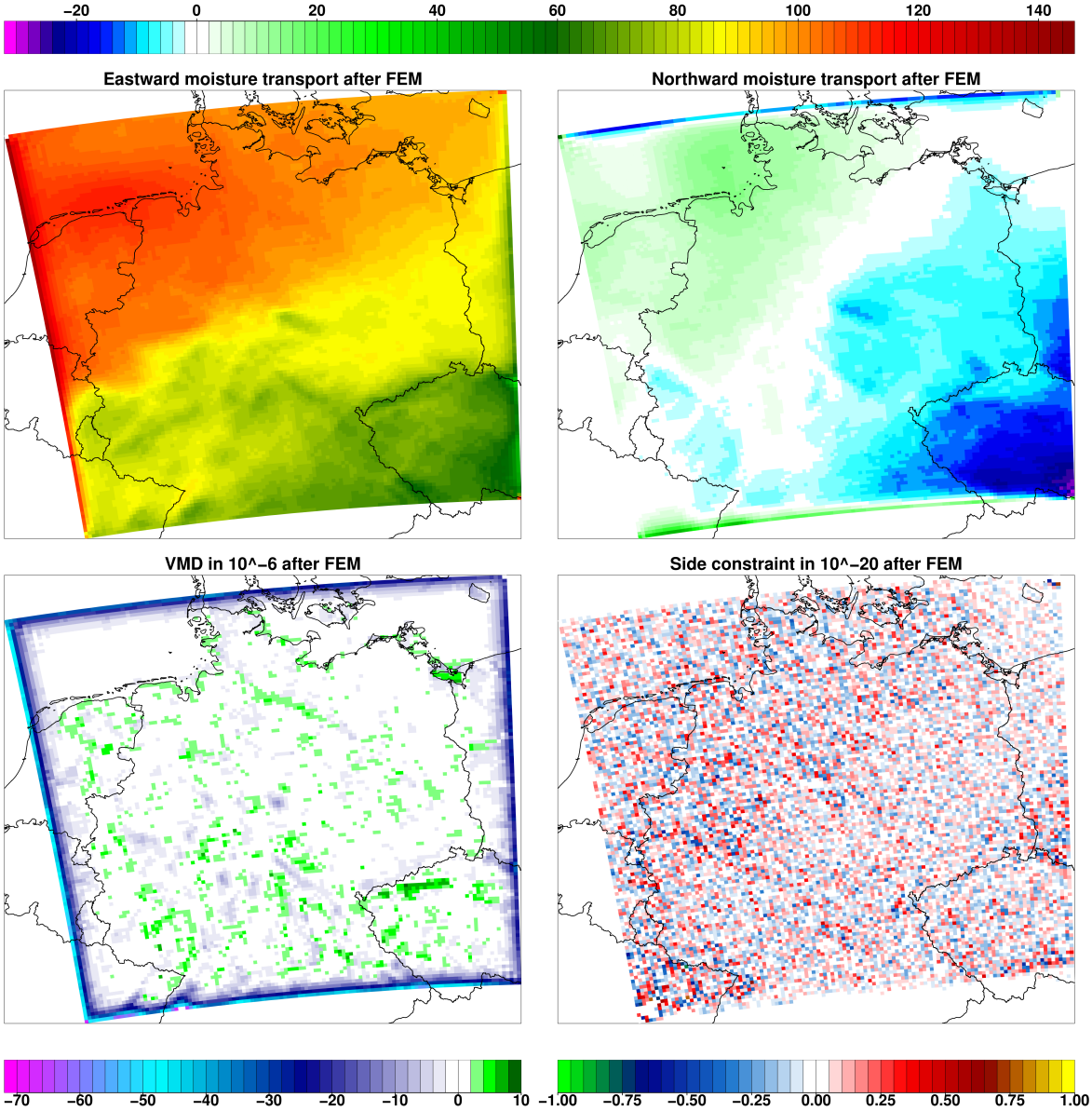


Figure 4.18.: The modified eastward and northward moisture transports (top left and top right) in kg m^{-1} s^{-1}, the modified vertically integrated moisture flux divergence in 10^{-6} kg m^{-2} s^{-1} (bottom left) and the modified difference between the divergence of the moisture transports and the VMD in 10^{-20} kg m^{-2} s^{-1} (bottom right).

the divergent part of the transports. The eastward moisture flux is positive throughout the domain and ranges from 57.64 to 109.99 kg m^{-1} s^{-1} with higher values over northwest Germany, the Netherlands and the Northern Sea and lesser values towards the southeast and the Czech Republic. The northward moisture transport is weaker with values between -30.20 and 21.11 kg m^{-1} s^{-1}. The positive values are located to the north and east of the domain and negative values are present in the west and southeast of the domain. The VMD is negative throughout, i.e. there is more precipitation than evaporation in the chosen domain which is plausible since the domain is mainly over land. The values range from -74.28 to

$-15.36 \cdot 10^{-6}$ kg m^{-2} s^{-1}. The difference between the divergence of the moisture budget and the VMD, i.e. the moisture budget from Eq. (3.113) shows still some recognizable structures similar to the VMD and is nowhere close to zero. This is the case although the moisture budget is two orders of magnitude smaller than the VMD with values between 0.01 and $0.21 \cdot 10^{-6}$ kg m^{-2} s^{-1}. Therefore, the moisture budget in the observed data is not fulfilled and needs to be corrected.
The fields after application of the variational approach are shown in Figure 4.18. The modified moisture transports show a similar structure and order as the observed ones. The modified eastward moisture transport ranges from 35.08 to 144.91 kg m^{-1} s^{-1} with the maximum over the North Frisian Islands and decreases from the northwest to the southeast. The modified northward moisture transport ranges from -28.04 to 56.67 kg m^{-1} s^{-1} with highest values over the Northern Sea and lowest values over the Czech Republic. The gradient in the modified northward moisture transport is rotated compared to the observed one and the values now decrease from the northwest to the southeast. In both transports, boundary effects are still visible at the west and east (north and south) boundaries in the eastward (northward) moisture transports. The same is true for the VMD. The boundary conditions are implemented into the approach via the last row and column of the left-hand side matrix in (3.136). Nevertheless, those boundary effects should be further investigated in future studies, because they possibly result from the matrices $\boldsymbol{T}$, $\boldsymbol{D_\lambda}$ and $\boldsymbol{D_\varphi}$. The structure of the modified VMD is basically the same as in the original VMD but has been shifted to higher values and now ranges between -70.86 and 8.69 kg m^{-2} s^{-1}. Now, the balance between evaporation and precipitation is much closer to zero. But the most important result is the modified moisture budget now showing values in the order of 10^{-20} kg m^{-2} s^{-1} and no structure at all. So the variational approach fulfilled the moisture budget in Eq. (3.113) up to numerical precision.
The results presented here are very sensitive to the choice of the errors σ_S and σ_M. As mentioned in chapter 3, σ_S was chosen as $5 \cdot 10^{-6}$ kg m^{-2} s^{-1} and $\sigma_M = a \cdot \sigma_S \cdot \kappa$ with $a = 6370$ km the earth´s radius and the scaling parameter κ set to 0.10. This results in $\sigma_M = 3.185$ kg m^{-1} s^{-1}. Preliminary tests with four different values for κ (0.10, 0.25, 0.50, 1.00) have been carried out and reveal that the influence of the boundary values becomes more dominant with increasing κ and thus with increasing error of the moisture transports. Similar plots as in Figure 4.18 for different κ can be found in the Appendix D. Further sensitivity studies would be desirable as would be a correct specification of the errors of both the moisture transports and the VMD. It is possible that the introduction of the hard constraint side condition that the divergence of the horizontal moisture transports should balance the VMD exactly is too strong. A weak constraint side condition where the squared difference between the horizontal moisture transports and the VMD should be minimized would be worth investigating.

5. Summary and Conclusions

During this work two high-resolution regional reanalyses have been created. Both reanalyses were produced using the operational NWP system of DWD with the COSMO model and additional modules. The first reanalysis, called COSMO-REA6, matches the European CORDEX-EURO-11 domain at a resolution of 0.055° with 40 vertical levels. The second reanalysis, COSMO-REA2, was produced with a horizontal resolution of 0.025° and 50 vertical levels and covers Mid-Europe. Both reanalyses provide a large output of almost 100 two-dimensional variables with an output frequency of 15 minutes and almost 30 three-dimensional variables with an output frequency of one hour. A portable, script-based system for the production of the reanalysis was developed, tested and implemented at the supercomputers at ECMWF and generated COSMO-REA6 for the years 1997-2014 and COSMO-REA2 for the years 2007-2013. For comparisons, a dynamical downscaling simulation (COSMO-DS) has been carried out for the year 2011 and some experiments on the influence of the assimilation of radar rain rates via latent heat nudging in both COSMO-REA6 and COSMO-REA2 for the summer months of 2011 were performed. One focus in the evaluation of both reanalyses was on the performance of the model and on precipitation which is one of the few variables which can be independently evaluated since it is not assimilated into the model. The global ERA-Interim reanalysis has been used as a reference as it is typically used for a large number of applications.
As a standard measure in data assimilation, the analysis increments in COSMO-REA6 for a one year period were evaluated, showing only small changes in two main prognostic variables, temperature and wind speed, indicating that the model is reasonably stable. Systematic deviations are found with the specification of the boundary layer height and the vertical position of the jet stream, where the standard deviation of the analysis increments in both temperature and wind speed increases. However, the patterns and order of magnitude are similar compared to the ERA-Interim reanalysis (see Dee *et al.*, 2011).
Comparing mean yearly precipitation sums over Europe using the GPCC data set, COSMO-REA6 and ERA-Interim slightly underestimate precipitation in mid and south Europe and overestimate precipitation in Scandinavia and Russia with comparably high overestimates at the Norwegian shore. An observed diurnal cycle of precipitation intensities during summer 2011 is reproduced by all COSMO simulations with a time shift from late afternoon towards night-time in COSMO-REA6 and COSMO-DS. ERA-Interim shows almost constant precipitation intensities throughout the whole day. The shift in the COSMO simulations is a known limitation of the convective scheme, which is why the shift is not reproduced by COSMO-REA2 where only shallow convection is parametrized and deep convection is resolved in the model. The assimilation of radar rain rates in COSMO-REA6 improves the representation of the diurnal cycle by shifting the diurnal cycle back towards late evening and dampens the results of COSMO-REA2. When precipitation sums are considered, the LHN improves both COSMO reanalyses, in the case of COSMO-REA6 considerably. COSMO-DS is showing almost no diurnal cycle in precipitation sums and ERA-Interim reveals a diurnal cycle with the peak in the morning and midday hours. Repeating this evaluation for the years 2007-2013, the same results are reproduced, showing the shift in the diurnal cycle of precipitation intensities in COSMO-REA6 and the better performance of COSMO-REA2 during summer. No diurnal

cycle in either precipitation intensities or sums can be observed during the winter months and all models follow the observations closely with ERA-Interim showing too small values in the precipitation intensities. Histograms of precipitation events above different thresholds show the good ability of the COSMO reanalyses to reproduce the occurrence and intensities of precipitation up to high rates of 50 mm per hour and beyond. ERA-Interim is producing too often precipitation and at the same time too many small precipitation events. ERA-Interim is also not able to produce precipitation above 10 mm per hour. But it is clearly pointed out that the characteristics evaluated here relate to spatio-temporal scales ERA-Interim had not been designed for. The introduction of LHN to the models potentially improves the frequency of the precipitation events in the different categories. The results indicate that a grid spacing of 6 km already produces precipitation characteristics that are comparable to point observations and a grid spacing of 2 km does so even more.
Another relevant characteristic for reanalyses is the coherence with independent observations. It is commonly known that the objective scores can be generally worse for high resolution models than for low resolution models, i.e. the so-called double-penalty problem. Thus, inevitable space and timing errors in the dynamical downscaling lead to larger errors. Measures such as the log-odds ratio for different precipitation threshold indicate that the dynamical downscaling is clearly outperformed by the regional reanalysis (which has been shown by Bollmeyer *et al.*, 2015), whereas for low thresholds, ERA-Interim is capable of compensating this problem by a good hit rate, taking a loss at the false alarm rate. The best agreement with observations can be found in COSMO-REA2 which is a combined effect of both the resolution of the model and the assimilation of radar rain rates, improving the occurrence of precipitation in both space and time.
Further evaluation was done on the radiative fluxes in the model. Both shortwave and longwave radiation as well as the net radiation, i.e. the sum of shortwave and longwave radiation, at the surface and the top of the atmosphere were compared between COSMO-REA6 and data from CERES-EBAF. The overall structure of the radiation fluxes is very similar in COSMO-REA6 compared to CERES-EBAF. The main differences are found in the southern part of the model domain. In shortwave radiation at the surface, the values over the Mediterranean Sea are remarkably higher in CERES-EBAF than in COSMO-REA6. This is probably a combined result of the different albedo values and an incorrect sea surface temperature representation in COSMO-REA6 which is only analysed every 24 hours by the SST analysis using ERA-Interim SST values and is kept constant in between. A deeper analysis of the sea surface temperature in COSMO-REA6 in comparison with ERA-Interim and other independent data sets like e.g. the Hadley Centre Sea Ice and Sea Surface Temperature data set (HadISST, Rayner *et al.*, 2003) would be desirable. Over Africa large absolute differences between 30 and 40 W m^{-2} in both shortwave and longwave radiation can be found in contrast to values of $\pm 10 - 20$ W m^{-2} in the remainder of the model domain. In addition, distinct structures in shortwave radiation at the surface in COSMO-REA6 can be found over Africa which stem from albedo values for different soil types. Differences in radiation fluxes at the top of the atmosphere do not exceed approximately ± 15 W m^{-2}, except for Africa and the Middle East where the differences are in the range of $30 - 60$ W m^{-2}. Apparently, there seems to be a problem in the radiative fluxes over Africa which could result e.g. from a wrong specification of albedo and other so far unknown error sources. Smiatek *et al.*, 2008 presented a new data set for soiltypes and albedo values in the climate version of COSMO (COSMO-CLM) and a preprocessor for making this data available for COSMO-CLM. A change to such a new data set or a comparable is desirable if it shows more realistic albedo and soiltype values. Nevertheless, a thorough evaluation of the

problems over Africa is needed, especially in view of considerations for an African reanalysis. Besides those problems, the radiation fluxes in COSMO-REA6 are in the range of uncertainty estimates of the CERES-EBAF data sets (Kato *et al.*, 2013; Loeb *et al.*, 2009).
The reproduction of the overall climate conditions is an important issue in the production of a reanalysis spanning several years. These have been evaluated using Köppen-Geiger maps and comparing COSMO-REA6 with a 50-year climatology, E-OBS observations over Europe and ERA-Interim. COSMO-REA6 and ERA-Interim are clearly able to reproduce the climate conditions in Europe and Northern Africa whereas E-OBS shows some deficiencies which probably stem from the not-consistent way of producing the gridded data set.
Bollmeyer *et al.*, 2015 have performed evaluations on several other parameters such as occurrence and structure of clouds or integrated water vapour. Throughout, the results reveal similar patterns as have been discussed before: both COSMO-REA6 as well as COSMO-DS better resolve small-scale structures than ERA-Interim but without the use of any data assimilation the model shows larger errors in e.g. water vapour fields. As a conclusion, the dynamical downscaling shows a similar performance compared to the reanalysis and may thus be sufficient on climatological time-scales. Nevertheless, the use of data assimilation considerably improves the representation of the actual state of the atmosphere, which is a key criterion for reanalyses. Current global reanalyses, here represented by ERA-Interim, are naturally limited in resolution. However, concerning large-scale dynamics, the advantages of a nested, high-resolution reanalysis become less prominent — especially, when considering that the lateral boundary conditions are actually provided by ERA-Interim.
There are a number of applications where the high-resolution regional reanalysis represents a valuable addition to global reanalysis and an alternative to dynamical downscaling approaches. The results therefore encourage the further exploitation of the high-resolution regional reanalysis in order to continuously reprocess and evaluate longer time periods.
Finally, a variational approach for the moisture budget was carried out. The balance between the divergence of the vertically integrated horizontal moisture transports and the vertically integrated moisture flux divergence, i.e. the difference between evaporation and precipitation, is not fulfilled in the COSMO-REA6 reanalysis. Applying a calculus of variations based on finite elements modified both the transports and the flux divergence in a consistent way. The calculated balance between the divergence of the modified transports and the modified VMD fulfills the balance up to numerical precision. The presented results could only be applied to a small subdomain of COSMO-REA6 due to technical issues. For the future, a migration to large supercomputers would be advantageous in order to carry out the presented approach for the complete domain.
The presented results are very sensitive to the assumed errors of both the horizontal moisture transports and the VMD. Therefore, a realistic assessment of the errors is needed to ensure realistic results for the variational approach. Apart from that, the exact fulfillment of the moisture balance in the side condition may be a too strong constraint. A weak constraint side condition could be exploited where only the squared difference between the moisture transports and the VMD is minimized. Investigating this approach could provide useful insights on the variational approach. Another problem in the variational approach could be the assumption of stationarity. In the derivation of the moisture budget equations it was assumed that the storage term is small enough in a long-term average that it can be neglected. Here, only a one year average was used, which could be problematic for this assumption, i.e. the storage term could be small, but not negligible and could thus affect the moisture budget. To be sure that the assumption holds, the variational approach should be applied to the average

over the complete eighteen years of the reanalysis. Furthermore, an implementation of complete covariance structures would complete the results and grant the possibility to calculate posterior covariance structures and thus give rise to an estimate of the error introduced by the variational approach (Bollmeyer and Hense, 2014).
The large output of both COSMO-REA6 and COSMO-REA2 offers a great potential for many applications. To prove the benefit of using the regional reanalysis data sets, the evaluation presented in this work should be extended to stations all across Europe to not only focus on Germany. Therefore, getting access to rain gauge measurements of high quality and temporal coverage from other European countries is an effort already being taken. The biggest problem in evaluating reanalysis is that most conventional observations have been assimilated into the model. Thus, the model is dependent on them and makes those observations useless for the evaluation of the model. That is why new measurements have to be explored such as the GPS measurements of integrated water vapour which were used by Bollmeyer *et al.*, 2015. An important focus in this regard is the highly resolved coverage of such independent observations. When comparing the highly resolved reanalyses with only 6-hourly or daily observations on coarse grids or only few stations, the added value of the high temporal and horizontal resolution in both COSMO-REA6 and COSMO-REA2 is lost. One possibility is to use radar and satellite data which have a good temporal coverage of several minutes to one hour, high horizontal resolutions and are not assimilated into the model (except for the LHN in COSMO-REA2). The high resolution of COSMO-REA6 and COSMO-REA2 allows to identify distinct objects such as precipitation fields. Those can be evaluated using object-based verification methods like e.g. the "Structure-Amplitude-Location" score SAL by Wernli *et al.*, 2008. Object-based verification methods are a fairly new developing field and an overview of several of these methods is provided by the Spatial Forecast Verification Methods Inter-Comparison Project (ICP). For satellite data, new tools like the Passive and Active Microwave radiative TRAnsfer model PAMTRA (Mech *et al.*, submitted) can be used to evaluate hydrometeor contents of clouds in the atmosphere. Furthermore, the data set could be exploited for new studies, e.g. in the emerging energy sector. The reanalysis data are suited to provide the energy sector with information on areas with high potential for wind or solar power production. Another issue that has to be adressed is the handling of the output. One year of the complete COSMO-REA6 output is about 25 TB big which is why only some selected variables could be evaluated so far. Currently, new servers are being installed allowing storage of more output. The evaluation can be thus expanded to more variables and larger time periods.

Appendix

A. Numerical implementation of the finite element method

For the numerical implementation of the inverse method presented in chapter 3, some special treatment is needed which is described in the following. Further details on the finite element methods and their implementation can be found in Schäfer, 1999 and Göber, 1997.

A.1. Discretization in finite elements

For the application of the finite element discretization, triangle elements are needed. Therefore, the model grid is separated into triangles by dividing one model box (with four grid points as corner points) into two triangles. To then compute the equations in chapter 3 in finite elements one needs to transform every triangle to an unitary triangle. This is done in the following way. Consider a triangle with corner points $P_1'(x_1,y_1)$, $P_2'(x_2,y_2)$ and $P_3'(x_3,y_3)$. The transformation equations onto the unitary triangle then read

$$\begin{aligned} x &= x_1+(x_2-x_1)\xi+(x_3-x_1)\eta \\ y &= y_1+(y_2-y_1)\xi+(y_3-y_1)\eta \end{aligned} \tag{A.1}$$

with ξ and η being the two coordinates for the unitary triangle. To complete the transformations, the metric coefficients h_ξ and h_η as well as the Jacobi matrix are needed. The metric coefficients are defined by

$$h_\xi = \left|\frac{\partial \vec{r}}{\partial \xi}\right|, h_\eta = \left|\frac{\partial \vec{r}}{\partial \eta}\right| \tag{A.2}$$

where $\vec{r}$ is the location vector

$$\vec{r} = \begin{pmatrix} x_1+(x_2-x_1)\xi+(x_3-x_1)\eta \\ y_1+(y_2-y_1)\xi+(y_3-y_1)\eta \end{pmatrix}. \tag{A.3}$$

The metric coefficients then become

$$h_\xi = \sqrt{(x_2-x_1)^2+(y_2-y_1)^2} \tag{A.4}$$

$$h_\eta = \sqrt{(x_3-x_1)^2+(y_3-y_1)^2}. \tag{A.5}$$

The Jacobi matrix reads

$$J = \begin{pmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial x}{\partial \eta} \\ \frac{\partial y}{\partial \xi} & \frac{\partial y}{\partial \eta} \end{pmatrix} = \begin{pmatrix} (x_2-x_1) & (x_3-x_1) \\ (y_2-y_1) & (y_3-y_1) \end{pmatrix} \tag{A.6}$$

and its determinant is defined by

$$det(J) = (x_2-x_1)\cdot(y_3-y_1)-(x_3-x_1)\cdot(y_2-y_1). \tag{A.7}$$

With the help of the determinant, the area element $dxdy$ can be transformed into $det(J)d\xi d\eta$. To complete the transformation we need to express the metric coefficients of the back transformation, h_x and h_y, in terms of h_ξ and h_η. This is done via the chain rule and one obtains

$$h_x = h_\xi \xi_x + h_\eta \eta_x \tag{A.8}$$
$$h_y = h_\xi \xi_y + h_\eta \eta_y . \tag{A.9}$$

Here, an abbreviation for the partial derivatives of ξ and η has been introduced:

$$\xi_x = \frac{\partial \xi}{\partial x}, \xi_y = \frac{\partial \xi}{\partial y}, \eta_x = \frac{\partial \eta}{\partial x}, \eta_y = \frac{\partial \eta}{\partial y} . \tag{A.10}$$

The partial derivatives are obtained by deriving the transform equations (A.1) with respect to x respectively y and solve for the wanted terms. One obtains

$$\frac{\partial \xi}{\partial x} = \frac{y_3 - y_1}{det(J)} \quad , \frac{\partial \xi}{\partial y} = -\frac{(x_3 - x_1)}{det(J)} \tag{A.11}$$
$$\frac{\partial \eta}{\partial x} = -\frac{y_2 - y_1}{det(J)}, \frac{\partial \eta}{\partial y} = \frac{(x_2 - x_1)}{det(J)} . \tag{A.12}$$

For the forming of the mass-matrix $\boldsymbol{\mathcal{T}}$ and the differential matrices $\boldsymbol{\mathcal{D}}_\lambda$ and $\boldsymbol{\mathcal{D}}_\varphi$ described in chapter 3, the general approach shall be made clear in a simple 3-by-3 example. Consider a model grid consisting of 9 model grid points ($\phi_1 \ldots \phi_9$), which is divided into 8 unitary triangles (T1...T8) as shown in Figure A.1. The mass-matrix $\boldsymbol{\mathcal{T}}$ in the x-y coordinate system for one

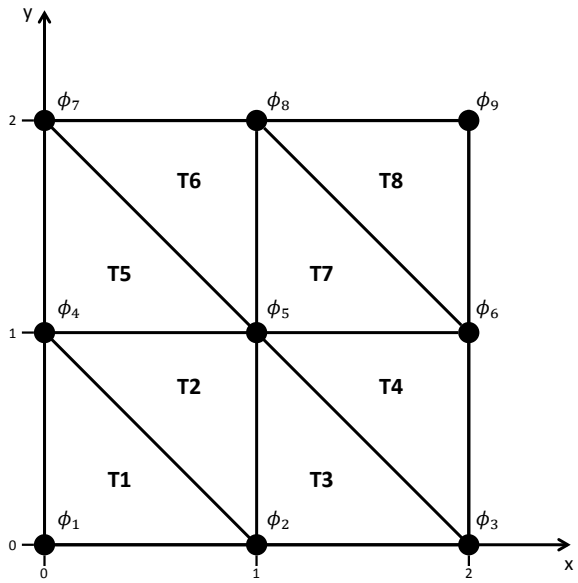


Figure A.1.: Discretization in finite elements of a 3-by-3 example model grid.

triangle element k would be

$$\boldsymbol{\mathcal{T}}^k = \int_{T_k} \vec{h}\vec{h}^T dxdy . \tag{A.13}$$

Using the transformation equations from above, the mass-matrix becomes

$$\boldsymbol{\mathcal{T}}^k = det(J)^k \underbrace{\int_{UT_k} \vec{h}\vec{h}^T d\xi d\eta}_{\widetilde{\boldsymbol{\mathcal{F}}_M}} . \tag{A.14}$$

Here, $\int_{T_k}$ is the integral over the triangle k whereas $\int_{UT_k}$ is the integral over the transformed unitary triangle k and $det(J)^k$ is the Jacobi-determinant for that triangle. Similarly, the differential matrices can be transformed into

$$\begin{aligned}
\boldsymbol{\mathcal{D}}^k_\lambda &= \int_{T_k} \vec{h}\vec{h}^T_x dxdy \\
&= \det(J)^k \int_{UT_k} \xi_x \vec{h}\vec{h}^T_\xi + \eta_x \vec{h}\vec{h}^T_\eta d\xi d\eta \\
&= \det(J)^k \xi_x \underbrace{\int_{UT_k} \vec{h}\vec{h}^T_\xi d\xi d\eta}_{\widetilde{\boldsymbol{\mathcal{F}}_\xi}} + det(J)^k \eta_x \underbrace{\int_{UT_k} \vec{h}\vec{h}^T_\eta d\xi d\eta}_{\widetilde{\boldsymbol{\mathcal{F}}_\eta}}
\end{aligned} \tag{A.15}$$

$$\begin{aligned}
\boldsymbol{\mathcal{D}}^k_\varphi &= \int_{T_k} \vec{h}\vec{h}^T_y dxdy \\
&= \det(J)^k \int_{UT_k} \xi_y \vec{h}\vec{h}^T_\xi + \eta_y \vec{h}\vec{h}^T_\eta d\xi d\eta \\
&= \det(J)^k \xi_y \underbrace{\int_{UT_k} \vec{h}\vec{h}^T_\xi d\xi d\eta}_{\widetilde{\boldsymbol{\mathcal{F}}_\xi}} + det(J)^k \eta_y \underbrace{\int_{UT_k} \vec{h}\vec{h}^T_\eta d\xi d\eta}_{\widetilde{\boldsymbol{\mathcal{F}}_\eta}} .
\end{aligned} \tag{A.16}$$

What becomes clear from these transformations is, that the geometrical dependence appears in a factor which is outside of the integral, making it easier to compute the transformations. What is needed now is a approach for the the function $h(\xi,\eta)$, which is chosen to be linear in both directions:

$$h_i(\xi,\eta) = \alpha_i^1 + \alpha_i^2 \xi + \alpha_i^3 \eta . \tag{A.17}$$

The function h_i has the value 1 at the corresponding point $P_i(\xi_i,\eta_i)$ and is linearly falling off to zero towards the other corner points of the triangle. Using the above demands on the coefficients, one comes up with

$$\begin{aligned}
h_1(\xi,\eta) &= 1 - \xi - \eta \\
h_2(\xi,\eta) &= \xi \\
h_3(\xi,\eta) &= \eta .
\end{aligned} \tag{A.18}$$

With these basis vectors one obtains

$$\vec{h}_\xi = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \quad , \quad \vec{h}_\eta = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} . \tag{A.19}$$

Table A.1.: Overview of the assignment of corner points for the triangles in the example Figure A.1.

		triangle number							
		1	2	3	4	5	6	7	8
Corner point	1	ϕ_1	ϕ_2	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_5	ϕ_6
	2	ϕ_2	ϕ_5	ϕ_3	ϕ_6	ϕ_5	ϕ_8	ϕ_6	ϕ_9
	3	ϕ_4	ϕ_4	ϕ_5	ϕ_5	ϕ_7	ϕ_7	ϕ_8	ϕ_8

For the computation of $\widetilde{\boldsymbol{\mathcal{F}}_{\boldsymbol{M}}}$, $\widetilde{\boldsymbol{\mathcal{F}}_{\boldsymbol{\xi}}}$ and $\widetilde{\boldsymbol{\mathcal{F}}_{\boldsymbol{\eta}}}$ there exists a useful formula which is valid for cases of p,q being non-negative integers and an integral over a unitary triangle:

$$I^U_{pq}T = \int_0^1 \int_0^{1-\eta} \xi^p \eta^q d\xi d\eta = \frac{p!q!}{(p+q+2)!}\,. \tag{A.20}$$

With the help of A.20 and the A.19 one obtains:

$$\widetilde{\boldsymbol{\mathcal{F}}_{\boldsymbol{M}}} = \frac{1}{24}\begin{pmatrix} 1 & 2 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \;,\; \widetilde{\boldsymbol{\mathcal{F}}_{\boldsymbol{\xi}}} = \frac{1}{6}\begin{pmatrix} -1 & 1 & 0 \\ -1 & 1 & 0 \\ -1 & 1 & 0 \end{pmatrix} \;,\; \widetilde{\boldsymbol{\mathcal{F}}_{\boldsymbol{\eta}}} = \frac{1}{6}\begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}. \tag{A.21}$$

A.2. Building the complete matrices

What was presented so far were the matrices for a single triangle. To extent this approach to every triangle and one needs the corner points for each triangle. Therefore, for each triangle the corner points are counted from one to three, starting at the bottom left point and then counting counterclockwise. Using the example from above (Figure A.1), this results in the assignment of corner points summarized in Table A.1. This assignment table is then needed for the building of the complete matrix, e.g. $\widetilde{\boldsymbol{\mathcal{F}}_{\boldsymbol{M}}}$, in the following way: if the i-th and j-th corner point of the k-th element equals n and m, then add the component $(\widetilde{\boldsymbol{\mathcal{F}}_{\boldsymbol{M}}})^k_{ij}$ of the k-th element to the component $(\boldsymbol{\mathcal{F}}_{\boldsymbol{M}})_{nm}$ of the complete matrix. For the example, this then looks as

follows:

$$\boldsymbol{\mathcal{F}_M} = \begin{pmatrix}
(\widetilde{\boldsymbol{\mathcal{F}_M}})^1_{11} & (\widetilde{\boldsymbol{\mathcal{F}_M}})^1_{12} & 0 \\
(\widetilde{\boldsymbol{\mathcal{F}_M}})^1_{21} & (\widetilde{\boldsymbol{\mathcal{F}_M}})^1_{22} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^2_{11} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^3_{11} & (\widetilde{\boldsymbol{\mathcal{F}_M}})^3_{12} \\
0 & (\widetilde{\boldsymbol{\mathcal{F}_M}})^3_{21} & (\widetilde{\boldsymbol{\mathcal{F}_M}})^3_{22} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^4_{11} \\
(\widetilde{\boldsymbol{\mathcal{F}_M}})^1_{31} & (\widetilde{\boldsymbol{\mathcal{F}_M}})^1_{32} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^2_{31} & 0 \\
0 & (\widetilde{\boldsymbol{\mathcal{F}_M}})^2_{21} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^3_{31} & (\widetilde{\boldsymbol{\mathcal{F}_M}})^3_{32} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^4_{31} \quad \cdots \\
0 & 0 & (\widetilde{\boldsymbol{\mathcal{F}_M}})^4_{21} \\
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix}$$

$$\cdots \begin{pmatrix}
(\widetilde{\boldsymbol{\mathcal{F}_M}})^1_{13} & 0 & 0 \\
(\widetilde{\boldsymbol{\mathcal{F}_M}})^1_{23} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^2_{13} & (\widetilde{\boldsymbol{\mathcal{F}_M}})^2_{12} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^3_{13} & 0 \\
0 & (\widetilde{\boldsymbol{\mathcal{F}_M}})^3_{23} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^4_{13} & (\widetilde{\boldsymbol{\mathcal{F}_M}})^4_{12} \\
(\widetilde{\boldsymbol{\mathcal{F}_M}})^1_{33} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^2_{33} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^5_{11} & (\widetilde{\boldsymbol{\mathcal{F}_M}})^2_{32} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^5_{12} & 0 \\
(\widetilde{\boldsymbol{\mathcal{F}_M}})^2_{23} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^5_{21} & (\widetilde{\boldsymbol{\mathcal{F}_M}})^2_{22} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^3_{33} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^4_{33} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^5_{22} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^6_{11} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^7_{11} & (\widetilde{\boldsymbol{\mathcal{F}_M}})^4_{32} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^7_{12} \quad \cdots \\
0 & (\widetilde{\boldsymbol{\mathcal{F}_M}})^4_{23} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^7_{21} & (\widetilde{\boldsymbol{\mathcal{F}_M}})^4_{22} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^7_{22} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^8_{11} \\
(\widetilde{\boldsymbol{\mathcal{F}_M}})^5_{31} & (\widetilde{\boldsymbol{\mathcal{F}_M}})^5_{32} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^6_{31} & 0 \\
0 & (\widetilde{\boldsymbol{\mathcal{F}_M}})^6_{21} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^7_{31} & (\widetilde{\boldsymbol{\mathcal{F}_M}})^7_{32} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^8_{31} \\
0 & 0 & (\widetilde{\boldsymbol{\mathcal{F}_M}})^8_{21}
\end{pmatrix}$$

$$\cdots \begin{pmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
(\widetilde{\boldsymbol{\mathcal{F}_M}})^5_{13} & 0 & 0 \\
(\widetilde{\boldsymbol{\mathcal{F}_M}})^5_{23} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^6_{13} & (\widetilde{\boldsymbol{\mathcal{F}_M}})^6_{12} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^7_{13} & 0 \\
0 & (\widetilde{\boldsymbol{\mathcal{F}_M}})^7_{23} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^8_{13} & (\widetilde{\boldsymbol{\mathcal{F}_M}})^8_{12} \\
(\widetilde{\boldsymbol{\mathcal{F}_M}})^5_{33} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^6_{33} & (\widetilde{\boldsymbol{\mathcal{F}_M}})^6_{32} & 0 \\
(\widetilde{\boldsymbol{\mathcal{F}_M}})^6_{23} & (\widetilde{\boldsymbol{\mathcal{F}_M}})^6_{22} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^7_{33} + (\widetilde{\boldsymbol{\mathcal{F}_M}})^8_{33} & (\widetilde{\boldsymbol{\mathcal{F}_M}})^8_{32} \\
0 & (\widetilde{\boldsymbol{\mathcal{F}_M}})^8_{23} & (\widetilde{\boldsymbol{\mathcal{F}_M}})^8_{22}
\end{pmatrix}$$

$$= \begin{pmatrix}
2 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 6 & 1 & 2 & 2 & 0 & 0 & 0 & 0 \\
0 & 1 & 4 & 0 & 2 & 1 & 0 & 0 & 0 \\
1 & 2 & 0 & 6 & 2 & 0 & 1 & 0 & 0 \\
0 & 2 & 2 & 2 & 12 & 2 & 2 & 2 & 0 \\
0 & 0 & 1 & 0 & 2 & 6 & 0 & 2 & 0 \\
0 & 0 & 0 & 1 & 2 & 0 & 4 & 1 & 0 \\
0 & 0 & 0 & 0 & 2 & 2 & 1 & 6 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 2
\end{pmatrix} . \tag{A.22}$$

The building of the complete matrices $\boldsymbol{\mathcal{F}_\xi}$ and $\boldsymbol{\mathcal{F}_\eta}$ is done accordingly. Note that this example is only for a 3-by-3 grid, the grid which is used in this study is potentially larger. Therefore, numerical solver suited for these kinds of problems need to be applied.

A.3. Solving the system

When trying to solve Eq. (3.136), some problems arose in inverting the DGL-matrix. The numerical solver was not able to solve the system due to the large differences (8 orders of magnitude) between the terms in $\frac{1}{\sigma_M^2}\boldsymbol{\mathcal{T}}$ and $\frac{1}{\sigma_S^2}\boldsymbol{\mathcal{T}}$, resulting from the σ-Terms. Therefore, the

system had to be slightly changed in order for the numerical solver to work properly. Eq. (3.136)

$$\begin{pmatrix} \frac{1}{\sigma_M^2}\boldsymbol{\mathcal{T}} & \mathbf{0} & \mathbf{0} & \boldsymbol{\mathcal{D}}_\lambda^T & \vec{1}|_{\lambda_B} \\ \mathbf{0} & \frac{1}{\sigma_M^2}\boldsymbol{\mathcal{T}} & \mathbf{0} & \boldsymbol{\mathcal{D}}_\varphi^T & \vec{1}|_{\varphi_B} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\sigma_S^2}\boldsymbol{\mathcal{T}} & -\boldsymbol{\mathcal{T}} & \vec{0} \\ \boldsymbol{\mathcal{D}}_\lambda & \boldsymbol{\mathcal{D}}_\varphi & -\boldsymbol{\mathcal{T}} & \mathbf{0} & \vec{0} \\ \vec{1}|_{\lambda_B}^T & \vec{1}|_{\varphi_B}^T & \vec{0}^T & \vec{0}^T & 0 \end{pmatrix} \cdot \begin{pmatrix} \vec{m}_\lambda \\ \vec{m}_\varphi \\ \vec{s} \\ \vec{\gamma} \\ \mu \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma_M^2}\boldsymbol{\mathcal{T}}\vec{m}_{\lambda,0} \\ \frac{1}{\sigma_M^2}\boldsymbol{\mathcal{T}}\vec{m}_{\varphi,0} \\ \frac{1}{\sigma_S^2}\boldsymbol{\mathcal{S}}\vec{s}_0 \\ \vec{0} \\ \sum_{\lambda_B}\vec{m}_{\lambda,B}|_{\lambda_B} + \sum_{\varphi_B}\vec{m}_{\varphi,B}|_{\varphi_B} \end{pmatrix}. \quad \text{(A.23)}$$

will be abbreviated in the following as $\boldsymbol{A}\cdot\vec{x} = \vec{b}$. In order to get rid of the σ-Terms one left-multiplies the system by $\boldsymbol{\mathcal{C}}^{-1}$, where

$$\boldsymbol{\mathcal{C}} = \begin{pmatrix} \frac{1}{\sigma_M} & 0 & \cdots & 0 & 0 & & & & & & & & & & 0 \\ 0 & \ddots & & & 0 & & & & & & & & & & \vdots \\ \vdots & & \frac{1}{\sigma_M} & & \vdots & & & \mathbf{0} & & & & \mathbf{0} & & & \vdots \\ 0 & & & \ddots & 0 & & & & & & & & & & \\ 0 & 0 & \cdots & 0 & \frac{1}{\sigma_M} & & & & & & & & & & \vdots \\ & & & & & \frac{1}{\sigma_S} & 0 & \cdots & 0 & 0 & & & & & \\ & & & & & 0 & \ddots & & & 0 & & & & & \vdots \\ & & \mathbf{0} & & & \vdots & & \frac{1}{\sigma_S} & & \vdots & & & \mathbf{0} & & \\ & & & & & 0 & & & \ddots & 0 & & & & & \vdots \\ & & & & & 0 & 0 & \cdots & 0 & \frac{1}{\sigma_S} & & & & & \\ & & & & & & & & & & 1 & 0 & \cdots & 0 & 0 & \vdots \\ & & & & & & & & & & 0 & \ddots & & & 0 & \vdots \\ & & \mathbf{0} & & & & & \mathbf{0} & & & \vdots & & 1 & & \vdots & \vdots \\ & & & & & & & & & & 0 & & & \ddots & 0 & \vdots \\ & & & & & & & & & & 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & & \cdots & & \cdots & & \cdots & & \cdots & & \cdots & \cdots & \cdots & \cdots & 0 & 1 \\ 1 & \cdots & \cdots & \cdots & 2N_g & 2N_g+1 & \cdots & \cdots & \cdots & 3N_g & 3N_g+1 & \cdots & \cdots & \cdots & 4N_g & 4N_g+1 \end{pmatrix} \begin{matrix} 1 \\ \vdots \\ \vdots \\ \vdots \\ 2N_g \\ 2N_g+1 \\ \vdots \\ \vdots \\ \vdots \\ 3N_g \\ 3N_g+1 \\ \vdots \\ \vdots \\ \vdots \\ 4N_g \\ 4N_g+1 \end{matrix}$$

is a diagonal matrix with σ_M on the diagonal for the first $2\cdot N_g$ rows, with N_g the dimension of the matrix $\boldsymbol{\mathcal{T}}$, and σ_S for the next N_g rows. The last N_g+1 rows are filled with ones. This results in

$$\begin{aligned} \boldsymbol{\mathcal{C}}^{-1}\boldsymbol{A}\vec{x} &= \boldsymbol{\mathcal{C}}^{-1}\vec{b} \\ \underbrace{\boldsymbol{\mathcal{C}}^{-1}\boldsymbol{A}\boldsymbol{\mathcal{C}}}_{\tilde{\boldsymbol{A}}}\boldsymbol{\mathcal{C}}^{-1}\vec{x} &= \boldsymbol{\mathcal{C}}^{-1}\vec{b} \end{aligned}$$

$$\begin{pmatrix} \boldsymbol{\mathcal{T}} & \mathbf{0} & \mathbf{0} & \sigma_M\boldsymbol{\mathcal{D}}_\lambda^T & \sigma_M\cdot\vec{1}|_{\lambda_B} \\ \mathbf{0} & \boldsymbol{\mathcal{T}} & \mathbf{0} & \sigma_M\boldsymbol{\mathcal{D}}_\varphi^T & \sigma_M\cdot\vec{1}|_{\varphi_B} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\mathcal{T}} & -\sigma_S\boldsymbol{\mathcal{T}} & \vec{0} \\ \sigma_M\boldsymbol{\mathcal{D}}_\lambda & \sigma_M\boldsymbol{\mathcal{D}}_\varphi & -\sigma_S\boldsymbol{\mathcal{T}} & \mathbf{0} & \vec{0} \\ \sigma_M\cdot\vec{1}|_{\lambda_B}^T & \sigma_M\cdot\vec{1}|_{\varphi_B}^T & \vec{0}^T & \vec{0}^T & 0 \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\sigma_M}\vec{m}_\lambda \\ \frac{1}{\sigma_M}\vec{m}_\varphi \\ \frac{1}{\sigma_S}\vec{s} \\ \vec{\gamma} \\ \mu \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma_M}\boldsymbol{\mathcal{T}}\vec{m}_{\lambda,0} \\ \frac{1}{\sigma_M}\boldsymbol{\mathcal{T}}\vec{m}_{\varphi,0} \\ \frac{1}{\sigma_S}\boldsymbol{\mathcal{S}}\vec{s}_0 \\ \vec{0} \\ \sum_{\lambda_B}\vec{m}_{\lambda,B}|_{\lambda_B} + \sum_{\varphi_B}\vec{m}_{\varphi,B}|_{\varphi_B} \end{pmatrix}.$$

This system of equations could then be solved by the numerical solver and the resulting transports and VMD needed to be multiplied by σ_M or σ_S to obtain the final results.

B. Reanalysis output

The folowing table shows a list of all output variables produced by both COSMO-REA6 and COSMO-REA2 along with their output frequencies.

Output field	Field type	Output freq	add. info
Albedo	2D	15-min	
Analysis increment for pressure	3D	Hourly	
Analysis increment for specific cloud water	3D	Hourly	
Analysis increment for specific humidity	3D	Hourly	
Analysis increment for temperature	3D	Hourly	
Analysis increment for wind direction	3D	Hourly	
Analysis increment for wind speed	3D	Hourly	
Cloud base mass flux	2D	15-min	
Cloud base of convective clouds	2D	15-min	
Cloud ceiling height above MSL	2D	15-min	
Cloud cover	3D	Hourly	
Cloud depth	2D	15-min	
Cloud top of convective clouds	2D	15-min	
Coefficient of vertical diffusion of heat	3D	Hourly	
Coefficient of vertical diffusion of momentum	3D	Hourly	
Convective available potential energy	2D	15-min	
Convective base index	2D	15-min	
Convective cloud cover	3D	Hourly	
Convective top index	2D	15-min	
Convective turbulent energy	2D	15-min	
Convective turbulent kinetic energy	2D	15-min	
Dew point in 2-m height	2D	15-min	
Diffuse downward shortwave radiation at the surface	2D	15-min	
Diffuse upward shortwave radiation at the surface	2D	15-min	
Direct downward shortwave radiation at the surface	2D	15-min	
Downward longwave radiation at the surface	2D	15-min	
Drag coefficient CD	2D	15-min	
Duration of sunshine	2D	15-min	
Eddy diffusivity rate	3D	Hourly	
Evaporation	2D	15-min	
Fresh snow factor	2D	15-min	
High cloud cover	2D	15-min	
Latent heat flux from bare soil	2D	15-min	

Latent heat flux from plants	2D		15-min		
Low cloud cover	2D		15-min		
Maximum 10-m convective gust	2D		15-min		
Maximum 10-m dynamical gust	2D		15-min		
Maximum 10-m wind speed without gusts	2D		15-min		
Maximum wind velocity	2D		15-min		
Mean sea level pressure	2D		15-min		
Medium cloud cover	2D		15-min		
Meridional component of momentum flux	2D		15-min		
Meridional wind speed		3D		Hourly	
Meridional wind speed in 10-m height	2D		15-min		
Net shortwave radiation at the surface	2D		15-min		
Net shortwave radiation at TOA	2D		15-min		
Net thermal radiation at the surface	2D		15-min		
Net thermal radiation at TOA	2D		15-min		
Pressure anomaly	2D		15-min		
Pressure perturbation		3D		Hourly	
Pressure tendency	2D		15-min		
Relative humidity in 2-m height	2D		15-min		
Sea ice cover	2D		15-min		
Specific humidity at the surface	2D		15-min		
Snow density	2D		15-min		
Snow height	2D		15-min		
Snow temperature	2D		15-min		
Soil ice content	2D		15-min		
Soil temperature	2D		15-min		
Soil water content	2D		15-min		
Solar downward radiation at top	2D		15-min		
Specific cloud ice content		3D		Hourly	
Specific cloud water content		3D		Hourly	
Specific cloud water content of convective clouds		3D		Hourly	
Specific graupel content		3D		Hourly	only REA2
Specific humidity		3D		Hourly	
Specific humidity in 2-m height	2D		15-min		
Specific rain water content		3D		Hourly	
Specific snow water content		3D		Hourly	
Surface flux of water vapour	2D		15-min		
Surface latent heat flux	2D		15-min		
Surface pressure	2D		15-min		
Surface temperature	2D		15-min		
Surface precipitation amount of convective rain	2D		15-min		
Surface precipitation amount of convective snow	2D		15-min		
Surface precipitation amount of grid scale rain	2D		15-min		
Surface precipitation amount of grid scale snow	2D		15-min		

Surface precipitation rate for convective rain	2D		15-min		
Surface precipitation rate for convective snow	2D		15-min		
Surface precipitation rate for grid scale rain	2D		15-min		
Surface precipitation rate for grid scale snow	2D		15-min		
Surface precipitation rate for grid scale graupel	2D		15-min		only REA2
Surface sensible heat flux	2D		15-min		
Synthetic satellite images	2D		15-min		REA6 only 2007-2012
Temperature anomaly	2D		15-min		
Temperature		3D		Hourly	
Temperature increment due to latent heat		3D		Hourly	
Temperature in 2-m height	2D		15-min		
Temperature of sea ice	2D		15-min		
Temperature of soil layers	2D		15-min		
Tendency of specific humidity		3D		Hourly	
Tendency of turbulent kinetic energy		3D		Hourly	
Transfer coefficient CH	2D		15-min		
Total cloud ice content	2D		15-min		
Total cloud cover	2D		15-min		
Total cloud water content	2D		15-min		
Total precipitable water	2D		15-min		
Total precipitation	2D		15-min		
Turbulent kinetic energy		3D		Hourly	
Upward longwave radiation at the surface	2D		15-min		
Vertical integral of divergence of total water content	2D		15-min		
Vertical integral of humidity, cloud water and ice	2D		15-min		
Vertical wind speed		3D		Hourly	
Water content of interception store	2D		15-min		
Water equivalent of accumulated snow depth	2D		15-min		
Water run-off	2D		15-min		
Zonal component of momentum flux	2D		15-min		
Zonal wind speed		3D		Hourly	
Zonal wind speed in 10-m height	2D		15-min		

C. Ecflow

"ecFlow is a work flow package that enables users to run a large number of programs (with dependencies on each other and on time) in a controlled environment. It provides reasonable tolerance for hardware and software failures, combined with good restart capabilities. It is used at ECMWF to around half our operational suites across a range of platforms. ecFlow submits tasks (jobs) and receives acknowledgments from tasks when they change status and when they send events. It does this using using child commands embedded in the scripts. ecflow stores the relationship between tasks and is able to submit tasks dependent on triggers."[1]

The above quotation from ECMWF is a very compact, to-the-point description of ecflow. In the following, the migration of the RPS described in section 2.2 to ecflow together with some technical details on possible implementations of other systems in ecflow is described.
At ECMWF, there are two systems open for registered users. One of them is the supercomputer CRAY XC-30[2], which is abbreviated as *cca* at ECMWF, where the large models can be run. The other system is called *ecgate*[3], which is a Linux cluster for processing computing jobs which are not suitable for the supercomputer. Ecflow is run on ecgate and submits all jobs to the CRAY. For the purpose of monitoring the process of the tasks in ecflow, ECMWF provides the monitoring tool *ecflowview*. A snapshot of ecflowview is shown in Figure C.1. Ecflow operates via an ecflow-server, which is called "testserver" in Figure C.1. There, the different projects are logged onto, called *suite*. In the following, the description is concentrated on the suite `reana_cycle_2002`, which produced the COSMO-REA6 stream for 2002-2006. Every suite can be divided into *families* and every family can again be subdivided into *tasks*. The different colors show the status of the different suites, families and tasks, yellow showing that a job (suite, family or task) is finished, red showing abort, cyan showing that the job waits in the queue, blue indicating that is waiting for another to finish and green telling that a job is running at the moment.
When starting `reana_cycle_2002`, the family `preparations` with its task `mkdirs` is conducted first, which creates all the necessary directories needed for the production. The main family is called `reana_loop`, which is repeated several times and counted via the variable `ITIME`, shown to the right of the family in Figure C.1. `reana_loop` is supposed to do all tasks every 6 hours but the repeating variable is an integer and therefore not well suited for cycling over dates, because it can not keep track of the 24 hour cycle of a day, even less of leap years. Therefore, dates in the form yyyyMMddhhmmss, e.g. 20110601000000 for the 1st June 2011, 00-00-00 UTC, are internally converted into seconds since the 1st January 1970 00-00-00 UTC and counted forward in `ITIME`. E.g. the shown time 1148860800 corresponds to the date 20060529000000. This converting of dates makes it very easy to jump back and forth in dates which is needed for several tasks. Families and tasks in ecflow can be started conditionally, which is shown by the grey boxes. `reana_loop` for example is only started when preparations is complete. The first task in `reana_loop` is the task `get_boundaries`.

[1] https://software.ecmwf.int/wiki/display/ECFLOW/What+is+ecFlow
[2] http://www.ecmwf.int/en/computing/our-facilities/supercomputer
[3] http://www.ecmwf.int/en/computing/our-facilities/ecgate

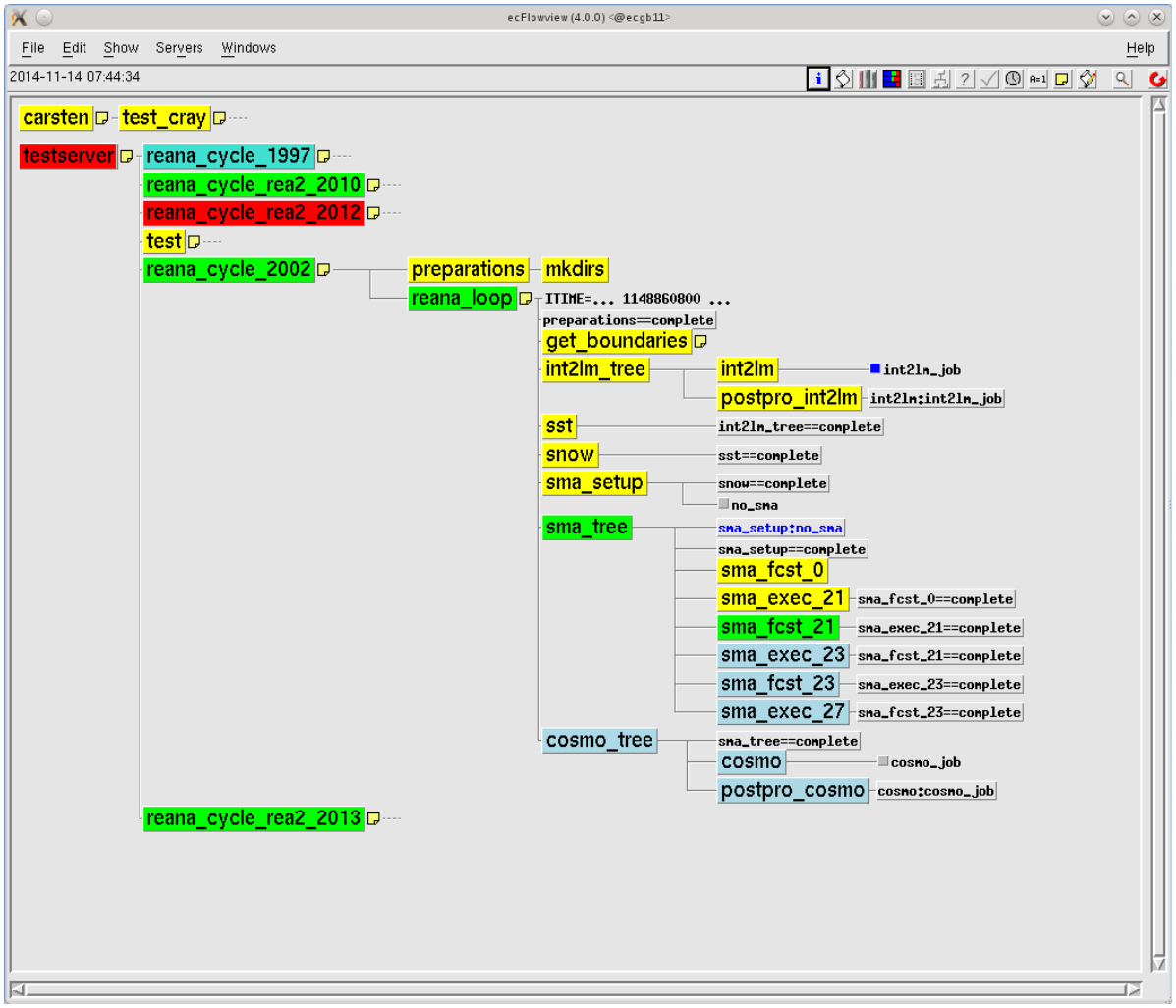


Figure C.1.: Snapshot of an ecflowview-Monitor

This task simply submits a job to cca which copies the observations for the COSMO run, for the snow and the SST-analysis to cca and extracts the boundary fields from ERA-Interim for some days in advance so that the data is always available in time. What follows is the family `int2lm_tree`, which consists of the actual `int2lm` task and a task for the postprocessing, `postpro_int2lm`. In `int2lm_tree`, a new feature of ecflow is introduced, called *event*. The `int2lm`-task contains an if-statement so that int2lm is run only every 24 hours and as last action the event `int2lm_job` is set. When int2lm is actually run, this happens after the execution, if not this happens immediately. The task `postpro_int2lm` is only executed when the event `int2lm_job` is set and essentially sets the correct links to the boundary files needed for the COSMO run.
When `int2lm` is finished, the `sst`-task with the SST-analysis is started (and executed only every 24 hours), followed by the Snow analysis in `snow`, executed after the SST analysis has finished. After this, the `sma_setup` is carried out, executing some preliminaries and checking whether the SMA is needed in this step. Remember this is done every 6 hours. If the SMA is not needed, the event `no_sma` is set, setting the following family `sma_tree` to “complete”

and thus skipping the SMA completely. Every 24 hours, sma_tree is carried out with the successive three forecasts sma_fcst_0, sma_fcst_21 and sma_fcst_23, each followed by the actual performance of the SMA, i.e. sma_exec_21, sma_exec_23 and sma_exec_27. The different numbers are historical and used here as in the operational use at DWD. The details are explained in 2.1.7.4. The last family is then cosmo_tree, consisting of cosmo, doing preparations and submitting the COSMO run and postpro_cosmo, which is doing some postprocessing steps and submits tasks for the automated archiving of the output. When cosmo_tree is finished, reana_loop starts again with ITIME counted 21600 seconds forward. The whole production of the reanalysis is done in this way at ecflow.
Every suite is controlled by a definition file in which variables can be defined that are evaluated by the different jobs. In this way, the definitions of those variables is outsourced into the definition-File, making it easier to keep track of all variables.

D. Sensitivity study for the variational approach

As a preliminary sensitivity study, the variational approach has been carried out with different prescribed error values for σ_S and σ_M. σ_S was chosen as $5 \cdot 10^{-6}$ kg m^{-2} s^{-1} and σ_M was set to $\sigma_S \cdot a \cdot \kappa$, with a the Earth´s radius and κ a scaling parameter. For the sensitivity study, κ was set to 0.1, 0.25, 0.50 and 1.00, resulting in errors of the moisture transport of 3.185, 7.9625, 15.925 and 31.85 kg m^{-1} s^{-1}. The results of the variational approach of this study are shown in the Figures D.1 to D.4 and show the increasing influence of the boundary values on the modified moisture transports with increasing κ.

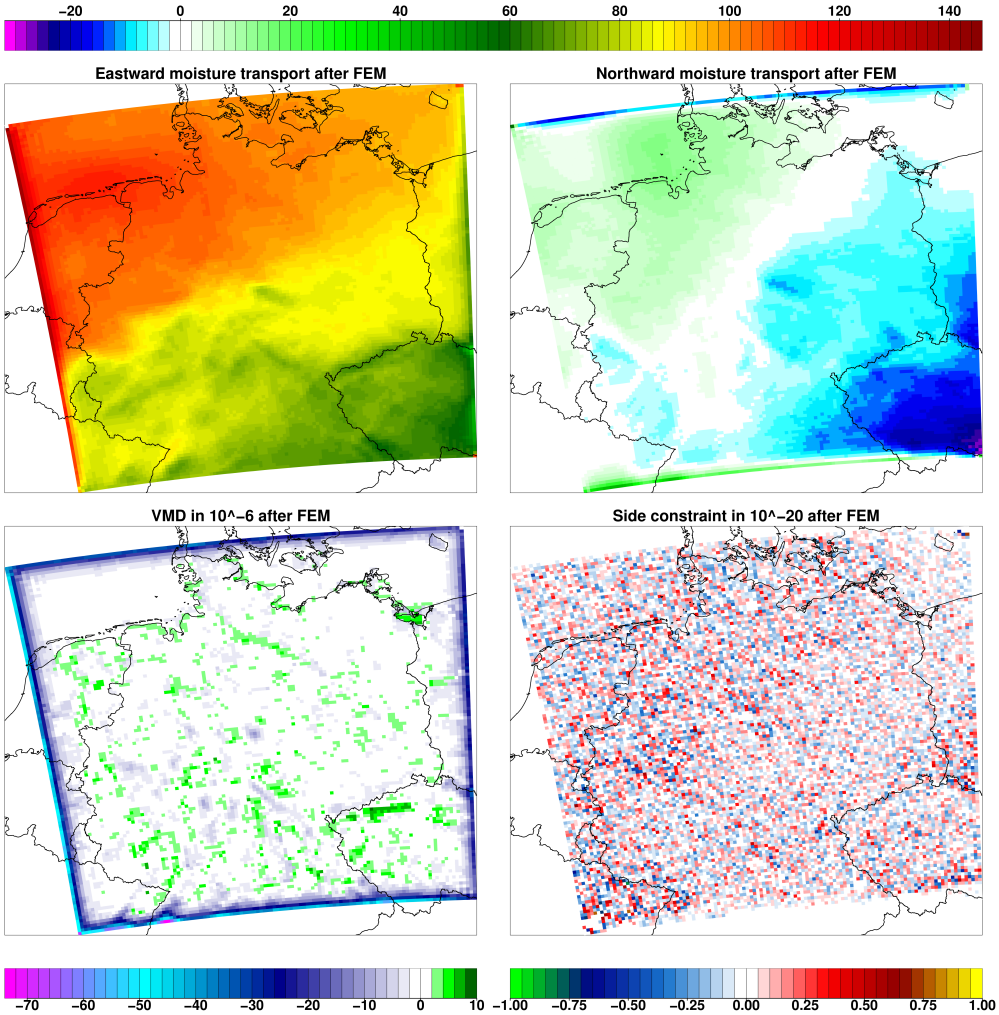


Figure D.1.: The modified eastward and northward moisture transports (top left and top right) in kg m^{-1} s^{-1}, the modified vertically integrated moisture flux divergence in 10^{-6} kg m^{-2} s^{-1} (bottom left) and the modified difference between the divergence of the moisture transports and the VMD in 10^{-20} kg m^{-2} s^{-1} (bottom right) for a scaling parameter of $\kappa = 0.1$.

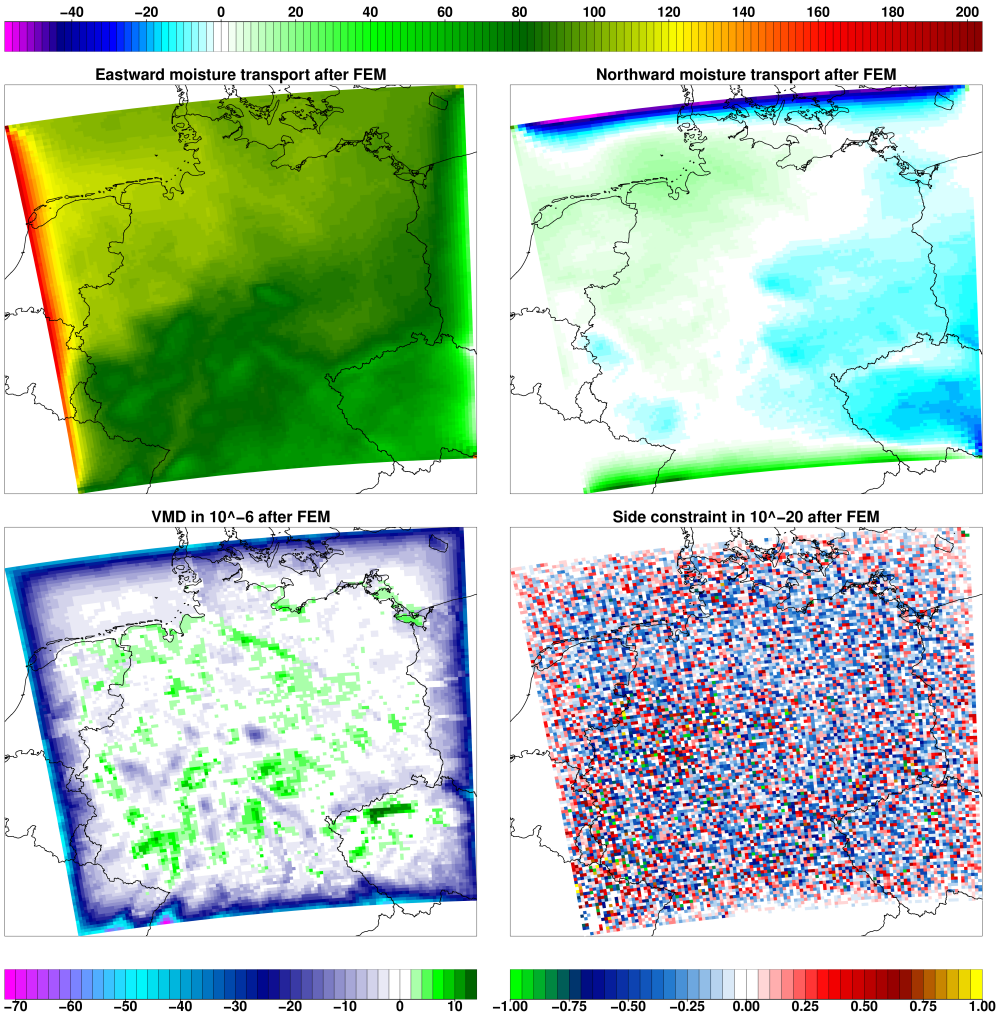


Figure D.2.: The modified eastward and northward moisture transports (top left and top right) in kg m^{-1} s^{-1}, the modified vertically integrated moisture flux divergence in 10^{-6} kg m^{-2} s^{-1} (bottom left) and the modified difference between the divergence of the moisture transports and the VMD in 10^{-20} kg m^{-2} s^{-1} (bottom right) for a scaling parameter of $\kappa = 0.25$.

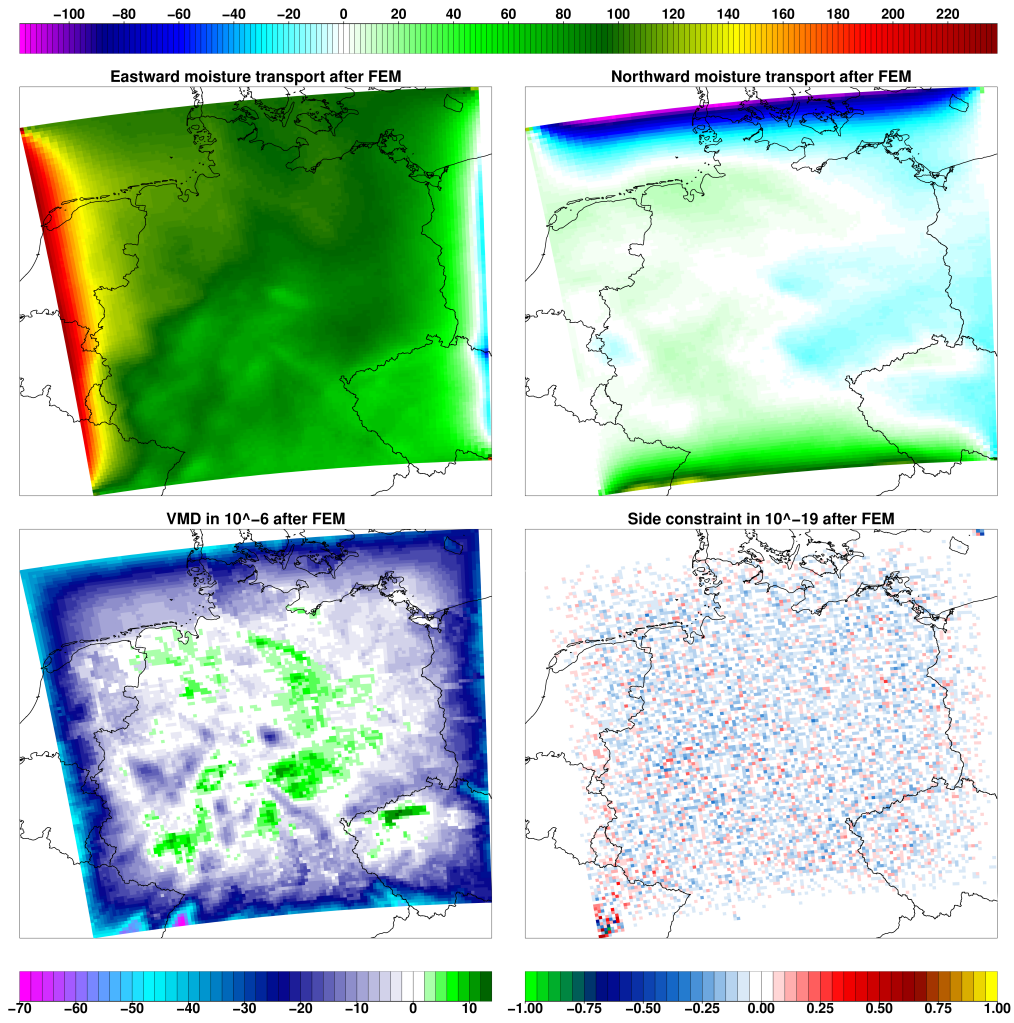


Figure D.3.: The modified eastward and northward moisture transports (top left and top right) in kg m^{-1} s^{-1}, the modified vertically integrated moisture flux divergence in 10^{-6} kg m^{-2} s^{-1} (bottom left) and the modified difference between the divergence of the moisture transports and the VMD in 10^{-19} kg m^{-2} s^{-1} (bottom right) for a scaling parameter of $\kappa = 0.5$.

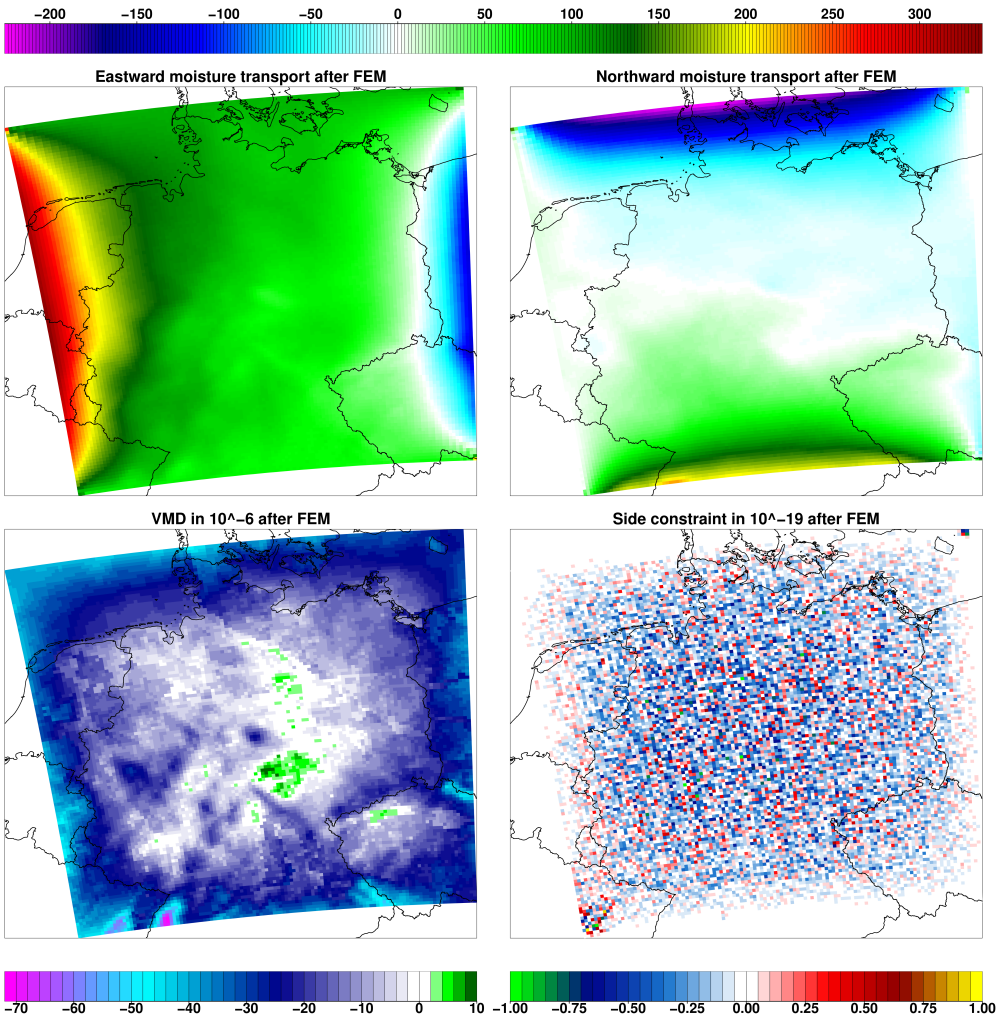


Figure D.4.: The modified eastward and northward moisture transports (top left and top right) in kg m^{-1} s^{-1}, the modified vertically integrated moisture flux divergence in 10^{-6} kg m^{-2} s^{-1} (bottom left) and the modified difference between the divergence of the moisture transports and the VMD in 10^{-19} kg m^{-2} s^{-1} (bottom right) for a scaling parameter of $\kappa = 1.0$.

Bibliography

Anthes, R. A. (1974). Data assimilation and initialization of hurricane prediction models. *Journal of the Atmospheric Sciences* **31**:3, 702–719.

Arakawa, A., Lamb, V. R. (1981). A potential enstrophy and energy conserving scheme for the shallow water equations. *Monthly Weather Review* **109**:1, 18–36.

Arakawa, A., Lamb, V. (1977). Computational design of the basic dynamical processes of the UCLA general circulation model. *Methods of Computational Physics* **17**: 173–265.

Beck, C., Grieser, J., Rudolf, B. (2004). A New Monthly Precipitation Climatology for the Global Land Areas for the Period 1951 to 2000. *Climate Status Report*. German Weather Service, Offenbach, Germany. Reprint available at http://gpcc.dwd.de, 181–190.

Bengtsson, L., Shukla, J. (1988). Integration of space and in situ observations to study global climate change. *Bulletin of the American Meteorological Society* **69**:10, 1130–1143.

Bergemann, K., Reich, S. (2012). An ensemble Kalman-Bucy filter for continuous data assimilation. *Meteorologische Zeitschrift* **21**:3, 213–219.

Bjerknes, V. (2009). The problem of weather prediction, considered from the viewpoints of mechanics and physics. *Meteorologische Zeitschrift*, 663–667.

Böhme, T., Stapelberg, S., Akkermans, T., Crewell, S., Fischer, J., Reinhardt, T., Seifert, A., Selbach, C., Van Lipzig, N. (2011). Long-term evaluation of COSMO forecasting using combined observational data of the GOP period. *Meteorologische Zeitschrift* **20**:2, 119–132.

Bollmeyer, C., Hense, A. (2014). Inverse modeling of energy transports and budgets of the atmosphere. *Climate Dynamics* **43**: 829–844.

Bollmeyer, C., Keller, J., Ohlwein, C., Wahl, S., Crewell, S., Friederichs, P., Hense, A., Keune, J., Kneifel, S., Pscheidt, I., Redl, S., Steinke, S. (2015). Towards a high-resolution regional reanalysis for the European CORDEX domain. *Quarterly Journal of the Royal Meteorological Society* **141**:686, 1–15.

Bouyssel, F., Cassé, V., Pailleux, J. (1999). Variational surface analysis from screen level atmospheric parameters. *Tellus A* **51**:4, 453–468.

Bromwich, D., Kuo, Y.-H., Serreze, M., Walsh, J., Bai, S., Barlage, M., Hines, K., Slater, A. (2010). Arctic System Reanalysis: Call for Community Involvement. *Eos, Transactions American Geophysical Union* **91**:2, 13–14.

Callies, U., Rhodin, A., Eppel, D. P. (1998). A case study on variational soil moisture analysis from atmospheric observations. *Journal of Hydrology* **212–213**: 95 –108.

Castro, C. L., Pielke, R. A., Leoncini, G. (2005). Dynamical downscaling: Assessment of value retained and added using the Regional Atmospheric Modeling System (RAMS). *Journal of Geophysical Research: Atmospheres (1984–2012)* **110**:D5.

Crewell, S., Mech, M., Reinhardt, T., Selbach, C., Betz, H.-D., Brocard, E., Dick, G., O'Connor, E., Fischer, J., Hanisch, T., et al. (2008). The general observation period 2007 within the priority program on quantitative precipitation forecasting: Concept and first results. *Meteorologische Zeitschrift* **17**:6, 849.

Davies, H., Turner, R. (1977). Updating prediction models by dynamical relaxation: An examination of the technique. *Quarterly Journal of the Royal Meteorological Society* **103**: 225–245.

Davolio, S., Buzzi, A. (2004). A Nudging Scheme for the Assimilation of Precipitation Data into a Mesoscale Model. *Weather and Forecasting* **19**: 855–871.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., Berg, L. v. d., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., Rosnay, P. d., Tavolato, C., Thépaut, J.-N., Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* **137**:656, 553–597.

Doms, G. (Sept. 2011). *A Description of the Nonhydrostatic Regional Model LM - Part I: Dynamics and Numerics*. Deutscher Wetterdienst, 147 pp.

Doms, G., Förstner, J., Heise, E., Herzog, H.-J., Mironov, D., Raschendorfer, M., Reinhardt, T., Ritter, B., Schrodin, R., Schulz, J.-P., Vogel, G. (Sept. 2011). *A Description of the Nonhydrostatic Regional Model LM - Part II: Physical Parameterization*. Deutscher Wetterdienst,

Dutton, J. A. (1986). *The ceaseless wind - An introduction to the theory of atmospheric motion*. Dover publications, Inc., 617 pp.

Ebita, A., Kobayashi, S., Ota, Y., Moriya, M., Kumabe, R., Onogi, K., Harada, Y., Yasui, S., Miyaoka, K., Takahashi, K., et al. (2011). The Japanese 55-year Reanalysis: An Interim Report. *SOLA* **7**: 149–152.

Ehrendorfer, M., Hantel, M., Wang, Y. (1994). A variational modification algorithm for three-dimensional mass flux non-divergence. *Quarterly Journal of the Royal Meteorological Society* **120**: 655–698.

Gal-Chen, T., Somerville, R. C. (1975). On the use of a coordinate transformation for the solution of the Navier-Stokes equations. *Journal of Computational Physics* **17**:2, 209–228.

Giorgi, F., Jones, C., Asrar, G. R., et al. (2009). Addressing climate information needs at the regional level: the CORDEX framework. *World Meteorological Organization (WMO) Bulletin* **58**:3, 175.

Göber, M. (1997). "Regionale Klimadiagnostik großskaliger Wirbel." PhD thesis. Meteorologisches Institut der Universitaet Bonn.

Hacker, J. M. (1981). Der Massen- und Energiehaushalt der Nordhemisphäre. *Bonner Meteorologische Abhandlungen* **27**: 102 pp.

Haltiner, G. J., Williams, R. T. (1980). *Numerical prediction and dynamic meteorology*. Vol. 2. Wiley New York, 477 pp.

Hantel, M., Haase, S. (1983). Mass consistent heat budget of the zonal atmosphere. *Bonner Meteorologische Abhandlungen* **29**: 84 pp.

Haylock, M., Hofstra, N., Klein Tank, A., Klok, E., Jones, P., New, M. (2008). A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research: Atmospheres (1984–2012)* **113**:D20.

Hess, R. (2001). Assimilation of screen-level observations by variational soil moisture analysis. English. *Meteorology and Atmospheric Physics* **77**:1-4, 145–154.

Holton, J. R. (1979). *An Introduction to Dynamic Meteorology*. Academic Press, 391 pp.

Inness, A., Baier, F., Benedetti, A., Bouarar, I., Chabrillat, S., Clark, H., Clerbaux, C., Coheur, P., Engelen, R., Errera, Q., et al. (2013). The MACC reanalysis: an 8 yr data set of atmospheric composition. *Atmospheric Chemistry and Physics* **13**: 4073–4109.

Jolliffe, I. T., Stephenson, D. B. (2012). *Forecast verification: a practitioner's guide in atmospheric science*. Wiley, 274 pp.

Jones, C. D., Macpherson, B. (1997). A latent heat nudging scheme for the assimilation of precipitation data into an operational mesoscale model. *Meteorological Application* **4**: 269–277.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., Joseph, D. (1996). The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society* **77**: (3), 437–471.

Kato, S., Loeb, N. G., Rose, F. G., Doelling, D. R., Rutan, D. A., Caldwell, T. E., Yu, L., Weller, R. A. (2013). Surface irradiances consistent with CERES-derived top-of-atmosphere shortwave and longwave irradiances. *Journal of Climate* **26**:9, 2719–2740.

Köppen, W. (1884). Die Wärmezonen der Erde, nach der Dauer der heissen, gemässigten und kalten Zeit und nach der Wirkung der Wärme auf die organische Welt betrachtet. *Meteorologische Zeitschrift* **1**: 215–226.

Köppen, W. (1918). Klassifikation der Klimate nach Temperatur, Niederschlag und Jahreslauf. *Petermanns Geographische Mitteilungen* **64**: 193–203.

Kopp, G., Lawrence, G., Rottman, G. (2005). "The total irradiance monitor (TIM): science results." *The Solar Radiation and Climate Experiment (SORCE)*. Springer, 129–139.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F. (2006). World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift* **15**:3, 259–263.

Loeb, N. G., Wielicki, B. A., Doelling, D. R., Smith, G. L., Keyes, D. F., Kato, S., Manalo-Smith, N., Wong, T. (2009). Toward optimal closure of the Earth's top-of-atmosphere radiation budget. *Journal of Climate* **22**:3, 748–766.

Mahfouf, J.-F. (1991). Analysis of soil moisture from near-surface parameters: A feasibility study. *Journal of Applied Meteorology* **30**:11, 1534–1547.

Majewski, D., Ritter, B. (2002). Das Global-Modell GME. *Promet* **27**:3/4, 111–122.

Mech, M., Maahn, M., Orlandi, E., Kneifel, S., Kollias, P., Redl, S., Crewell, S. (submitted). PAMTRA - passive and active microwave radiative transfer model: Basic concepts and applications for cloud research. *Atmospheric Chemistry and Physics*.

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., Jovic, D., Woollen, J., Rogers, E., Berbery, E. H., Ek, M. B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., Shi, W. (2006). North American Regional Reanalysis. *Bulletin of the American Meteorological Society* **87**: (3), 343–360.

Mieghem, J. v. (1973). *Atmospheric Energetics*. Oxford University Press, 306.

Mitchell, T. D., Jones, P. D. (2005). An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology* **25**:6, 693–712.

Onogi, K., Tsutsui, J., Koide, H., Sakamoto, M., Kobayashi, S., Hatsushika, H., Matsumoto, T., Yamazaki, N., Kamahori, H., Takahashi, K., et al. (2007). The JRA-25 Reanalysis. *Journal of the Meteorological Society of Japan. Ser. II* **85**:3, 369–432.

Rayner, N., Parker, D. E., Horton, E., Folland, C., Alexander, L., Rowell, D., Kent, E., Kaplan, A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres (1984–2012)* **108**:D14.

Reynolds, O. (1895). On the dynamical theory of incompressible viscous fluids and the determination of the criterion. *Philosophical Transactions of the Royal Society of London. A* **186**: 123–164.

Rhodin, A., Kucharski, F., Callies, U., Eppel, D., Wergen, W. (1999). Variational analysis of effective soil moisture from screen-level atmospheric parameters: Application to a short-range weather forecast model. *Quarterly Journal of the Royal Meteorological Society* **125**:559, 2427–2448.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, K., et al. (2011). MERRA: NASA's modern-era retrospective analysis for research and applications. *Journal of Climate* **24**:14, 3624–3648.

Rienecker, M., Suarez, M. J., Todling, R., Bacmeister, J., Takacs, L., Liu, H., Gu, W., Sienkiewicz, M., Koster, R., Gelaro, R., et al. (2008). *The GEOS-5 Data Assimilation System-Documentation of Versions 5.0.1, 5.1.0, and 5.2.0. NASA Technical Report Series on Global Modeling and Data Assimilation*. Vol. 27, NASA/TM-2008-105606, 97 pp.

Ritter, B., Geleyn, J.-F. (1992). A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations. *Monthly Weather Review* **120**:2, 303–325.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., et al. (2010). The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society* **91**:8, 1015–1057.

Saunders, R., Rayer, P., Blackmore, T., Matricardi, M., Bauer, P., Salmond, D. (2007). "A new fast radiative transfer model-RTTOV-9." *Joint 2007 EUMETSAT Meteorological Satellite Conference and the 15th Satellite Meteorology and Oceanography Conference of the American Meteorological Society, Amsterdam, The Netherlands*.

Saunders, R., Hocking, J., Rayer, P., Matricardi, M., Geer, A., Bormann, N., Brunel, P., Karbou, F., Aires, F. (2012). RTTOV-10 science and validation report. *EUMETSAT, NWPSAT-MO-TV-023*, 31.

Schäfer, M. (1999). *Numerik im Maschinenbau*. Springer Verlag, 294 pp.

Schättler, U., Doms, G., Schraff, C. (June 2011). *A Description of the Nonhydrostatic Regional Model LM - Part VII: User´s Guide*. Deutscher Wetterdienst, 192 pp.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., Rudolf, B. (2014). GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theoretical and Applied Climatology* **115**:1-2, 15–40.

Schraff, C., Hess, R. (June 2003). *A Description of the Nonhydrostatic Regional Model LM - Part III: Data Assimilation*. Deutscher Wetterdienst, 93 pp.

Simon, T., Wang, D., Hense, A., Simmer, C., Ohlwein, C. (2013). Generation and transfer of internal variability in a regional climate model. *Tellus A* **65**:

Smiatek, G., Rockel, B., Schättler, U. (2008). Time invariant data preprocessor for the climate version of the COSMO model (COSMO-CLM). *Meteorologische Zeitschrift* **17**:4, 395–405.

Stauffer, D. R., Seaman, N. L. (1990). Use of four-dimensional data assimilation in a limited-area mesoscale model. Part I: Experiments with synoptic-scale data. *Monthly Weather Review* **118**:6, 1250–1277.

Stephan, K., Klink, S., Schraff, C. (2008). Assimilation of radar-derived rain rates into the convective-scale model COSMO-DE at DWD. *Quarterly Journal Of The Royal Meteorological Society* **134**: 1315–1326.

Stephenson, D. B. (2000). Use of the "odds ratio" for diagnosing forecast skill. *Weather and Forecasting* **15**:2, 221–232.

Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, 342 pp.

Tiedtke, M. (1989). A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly Weather Review* **117**:8, 1779–1800.

Trenberth, K. E., Fasullo, J. T., Kiehl, J. (2009). Earth´s Global Energy Budget. *Bulletin of the American Meteorological Society*, 311–323.

Trenberth, K. E., Fasullo, J. T., Mackaro, J. (2011). Atmospheric moisture transports from ocean to land and global energy flows in reanalyses. *Journal of Climate* **24**:18, 4907–4924.

Trenberth, K. E., Koike, T., Onogi, K. (2008). Progress and prospects for reanalysis for weather and climate. *Eos, Transactions American Geophysical Union* **89**:26, 234–235.

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., Mcnally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., Woollen, J. (2005). The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society* **131**:612, 2961–3012.

Wernli, H., Paulat, M., Hagen, M., Frei, C. (2008). SAL-A novel quality measure for the verification of quantitative precipitation forecasts. *Monthly Weather Review* **136**:11, 4470–4487.

Wilby, R. L., Wigley, T. (1997). Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography* **21**:4, 530–548.

Zdunkowski, W., Bott, A. (2003). *Dynamics of the Atmosphere: A course in Theoretical Meteorology*. Cambridge University Press, 738 pp.

BONNER METEOROLOGISCHE ABHANDLUNGEN

Herausgegeben vom Meteorologischen Institut der Universität Bonn durch Prof. Dr. H. FLOHN (Hefte 1-25), Prof. Dr. M. HANTEL (Hefte 26-35), Prof. Dr. H.-D. SCHILLING (Hefte 36-39), Prof. Dr. H. KRAUS (Hefte 40-49), ab Heft 50 durch Prof. Dr. A. HENSE.

Heft 1-39: siehe `http://www2.meteo.uni-bonn.de/bibliothek/bma.html`

Heft 40: ***Hermann Flohn***: Meteorologie im Übergang Erfahrungen und Erinnerungen (1931-1991). 1992, 81 S. + XII. € 23

Heft 41: ***Adnan Alkhalaf and Helmut Kraus***: Energy Balance Equivalents to the Köppen-Geiger Climatic Regions. 1993, 69 S. + IX. € 19

Heft 42: ***Axel Gabriel***: Analyse stark nichtlinearer Dynamik am Beispiel einer rei- bungsfreien 2D-Bodenkaltfront. 1993, 127 S. + XIV. € 30

Heft 43: ***Annette Münzenberg-St.Denis***: Quasilineare Instabilitätsanalyse und ihre Anwendung auf die Strukturaufklärung von Mesozyklonen im östlichen Weddellmeergebiet. 1994, 131 S. + XIII. € 33

Heft 44: ***Hermann Mächel***: Variabilität der Aktionszentren der bodennahen Zirkulation über dem Atlantik im Zeitraum 1881-1989. 1995, 188 S. + XX. € 48

Heft 45: ***Günther Heinemann***: Polare Mesozyklonen. 1995, 157 S. + XVI. € 46

Heft 46: ***Joachim Klaßen:*** Wechselwirkung der Klima-Subsysteme Atmosphäre, Meereis und Ozean im Bereich einer Weddellmeer-Polynia. 1996, 146 S. + XVI. € 43

Heft 47: ***Kai Born***: Seewindzirkulationen: Numerische Simulationen der Seewind- front. 1996, 170 S. + XVI. € 48

Heft 48: ***Michael Lambrecht***: Numerische Untersuchungen zur tropischen 30-60-tägigen Oszillation mit einem konzeptionellen Modell. 1996, 48 S. + XII. € 15

Heft 49: ***Cäcilia Ewenz***: Seewindfronten in Australien: flugzeuggestützte Messungen und Modellergebnisse. 1999, 93 S. + X. € 30

Heft 50: ***Petra Friederichs***: Interannuelle und dekadische Variabilität der atmosphärischen Zirkulation in gekoppelten und SST-getriebenen GCM-Experimenten. 2000, 133 S. + VIII. € 25

Heft 51: ***Heiko Paeth***: Anthropogene Klimaänderungen auf der Nordhemisphäre und die Rolle der Nordatlantik-Oszillation. 2000, 168 S.+ XVIII. € 28

Heft 52: ***Hildegard Steinhorst***: Statistisch-dynamische Verbundsanalyse von zeitlich und räumlich hoch aufgelösten Niederschlagsmustern: eine Untersuchung am Beispiel der Gebiete von Köln und Bonn. 2000, 146 S. + XIV. € 25

Heft 53: ***Thomas Klein***: Katabatic winds over Greenland and Antartica and their interaction with mesoscale and synoptic-scale weather systems: three-dimensional numerical models. 2000, 146 S. + XIV. € 25

Heft 54: ***Clemens Drüe***: Experimentelle Untersuchung arktischer Grenzschichtfronten an der Meereisgrenze in der Davis-Straße. 2001, 165 S. + VIII. € 28

Heft 55: ***Gisela Seuffert***: Two approaches to improve the simulation of near surface processes in numerical weather prediction models. 2001, 128 S. + VI. € 25

Heft 56: ***Jochen Stuck***: Die simulierte axiale atmosphärische Drehimpulsbilanz des ECHAM3-T21 GCM. 2002, 202 S. + VII. € 30

Heft 57: ***Günther Haase***: A physical initialization algorithm for non-hydrostatic weather prediction models using radar derived rain rates. 2002, 106S. + IV. € 25

Heft 58: ***Judith Berner***: Detection and Stochastic Modeling of Nonlinear Signatures in the Geopotential Height Field of an Atmospheric General Circulation Model. 2003, 157 S. + VIII. € 28

Heft 59: ***Bernd Maurer***: Messungen in der atmosphärischen Grenzschicht und Validation eines mesoskaligen Atmosphärenmodells über heterogenen Landoberflächen. 2003, 182 S. + IX. € 30

Heft 60: ***Christoph Gebhardt***: Variational reconstruction of Quaternary temperature fields using mixture models as botanical – climatological transfer functions. 2003, 204 S. + VIII. € 30

Heft 61: ***Heiko Paeth***: The climate of tropical and northern Africa – A statistical-dynamical analysis of the key factors in climate variability and the role of human activity in future climate change. 2005, 316 S. + XVI. € 15

Heft 62: ***Christian Schölzel***: Palaeoenvironmental transfer functions in a Bayesian framework with application to Holocene climate variability in the Near East. 2006, 104 S. + VI. € 15

Heft 63: ***Susanne Bachner***: Daily precipitation characteristics simulated by a regional climate model, including their sensitivity to model physics, 2008, 161 S. € 15

Heft 64: ***Michael Weniger***: Stochastic parameterization: a rigorous approach to stochastic three-dimensional primitive equations, 2014, 148 S. + XV. open access[1]

Heft 65: ***Andreas Röpnack***: Bayesian model verification: predictability of convective conditions based on EPS forecasts and observations, 2014, 152 S. + VI. open access[1]

Heft 66: ***Thorsten Simon***: Statistical and Dynamical Downscaling of Numerical Climate Simulations: Enhancement and Evaluation for East Asia, 2014, 48 S. + LXXVII. open access[1]

Heft 67: ***Elham Rahmani***: The Effect of Climate Change on Wheat in Iran, 2014, 96 S. + XL open access[1]

Heft 68: ***Pablo A. Saavedra Garfias***: Retrieval of cloud and rainwater from ground-based passive microwave observations with the multi-frequency dual-polarized radiometer ADMIRARI, 2014, 168 S. + XIII open access[1]

Heft 69: ***Christoph Bollmeyer***: A high-resolution regional reanalysis for Europe and Germany - Creation and Verification with a special focus on the moisture budget, 2015, 103 S. + IX open access[1]

[1]Available at `http://hss.ulb.uni-bonn.de/fakultaet/math-nat/`

METEOROLOGISCHES INSTITUT
MATHEMATISCH NATURWISSENSCHAFTLICHE FAKULTÄT
UNIVERSITÄT BONN

Meteorologisches Institut
Universität Bonn