BONNER METEOROLOGISCHE ABHANDLUNGEN

Heft 83 (2018) (ISSN 0006-7156)
Herausgeber: Andreas Hense

Liselotte Bach

TOWARDS A PROBABILISTIC REGIONAL REANALYSIS FOR EUROPE

BONNER METEOROLOGISCHE ABHANDLUNGEN

Heft 83 (2018) (ISSN 0006-7156)
Herausgeber: Andreas Hense

Liselotte Bach

TOWARDS A PROBABILISTIC REGIONAL REANALYSIS FOR EUROPE

Towards a probabilistic regional reanalysis for Europe

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
MSc.-Physik der Erde und der Atmosphäre
Liselotte Bach
aus
Hamburg

Bonn, Februar, 2017

Diese Arbeit ist die ungekürzte Fassung einer der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn im Jahr 2017 vorgelegten Dissertation von Liselotte Bach aus Hamburg.

This paper is the unabridged version of a dissertation thesis submitted by Liselotte Bach born in Hamburg to the Faculty of Mathematical and Natural Sciences of the Rheinische Friedrich-Wilhelms-Universität Bonn in 2017.

Anschrift des Verfassers: Address of the author:

Liselotte Bach
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: PD. Dr. Silke Trömel, Rheinische Friedrich-Wilhelms-Universität Bonn

Tag der Promotion: 30. Oktober 2017

How little things can make a big difference
(Malcolm Gladwell)

Abstract

A new development in the field of reanalyses is the incorporation of uncertainty estimation capabilities. In the course of this work, a new probabilistic regional reanalysis system for the CORDEX-EUR11 domain has been developed. It is based on the numerical weather prediction model COSMO at a 12 km grid spacing. The lateral boundary conditions of all ensemble members are provided by the global reanalysis ERA-Interim. In the basic implementation of the system, uncertainties due to observation errors are estimated. Atmospheric assimilation of conventional observations perturbed by means of random samples of observation error yields estimates of the reanalysis uncertainty conditioned to observation errors.

The data assimilation employed is a new scheme based on observation nudging that is denoted ensemble nudging. The lower boundary of the atmosphere is regularly updated by external snow depth, sea surface temperature and soil moisture analyses. One of the most important purposes of reanalyses is the estimation of so-called essential climate variables. For regional reanalyses, precipitation has been identified as one of the essential climate variables that are potentially better represented than in other climate data sets. For that reason, the representation of precipitation in the system is assessed in a pilot study. Based on two experiments, each of which extends over one month, a preliminary comparison to the global reanalysis ERA-Interim, a dynamical downscaling of the latter and the high-resolution regional reanalysis COSMO-REA6 is conducted. In a next step, the probabilistic capabilities of the reanalysis system are assessed versus the ECMWF-EPS in terms of six-hourly precipitation sums. The added value of the probabilistic regional reanalysis system motivates the current production of a 5-year long test reanalysis COSMO-EN-REA12 in the framework of the FP7-funded project Uncertainties in Ensembles of Regional Reanalyses (UERRA).

To provide an indication of how the probabilistic reanalysis system would be configured ideally for future long-term reanalyses, it is updated to a new model version of COSMO and extended by further uncertainty estimation capabilities. Now, model error can be accounted for by the method of stochastic perturbation of physical tendencies. Further, uncertainties in the lateral boundary conditions are incorporated by utilizing a global ensemble of the new numerical weather prediction model ICON. A comparative verification of screen-level temperature as second important essential climate variable and precipitation in a range of numerical experiments with different configurations of the reanalysis system indicates that the best probabilistic capabilities are achieved by accounting for as many uncertain components as possible.

Contents

1 Introduction

Reanalyses of the past evolution of the earth system or its components are the most frequently used long-term climate data sets. Reanalyses also provide a fundamental basis for the development of adaptation and mitigation strategies (Dee et al. 2014). At the time of writing, the paper that introduces the global reanalysis ERA-Interim (Dee et al. 2011a) has a number of 6429 citations.

Atmospheric reanalyses emerge from a reconstruction of the four-dimensional record of the past atmospheric state, where a high degree of spatio-temporal consistency and accuracy is strived for. They have first been proposed for climate-change studies by Trenberth and Olson (1988) and Bengtsson and Shukla (1988). The central component of reanalysis systems is a data assimilation algorithm that feeds historical observations of the atmospheric state into a numerical weather prediction (NWP) model. Through the model equations, the information content of the observations is spread in space and time and passed from the observed to the non-observed variables. By the atmospheric flow, it is transported to observation-sparse regions. From this process, complete and physically consistent four-dimensional fields evolve. Data assimilation ensures that the model state stays close to the true state of the atmosphere.

Due to these properties, reanalyses have many advantages over related climate data sets. Their physical consistency between the variables is a major advantage over gridded observation climate data sets, e.g. the E-OBS data (Haylock et al. 2008) or HadCRUT (Brohan et al. 2006, Jones et al. 2012, Morice et al. 2012). Via their spatio-temporal completeness reanalyses add value to heterogeneously distributed observations which in turn add realistic detail such as extremes on the resolved scales to the reanalysis. The latter is beneficial compared to dynamical downscalings and hindcasts which are likewise spatio-temporally complete climate data sets, but due to the absence of data assimilation rather tend towards the model climate. Therefore, they cannot provide the same accurate information about the evolution of weather and climate.

Finally, the meta data (observation feedback) evolving from the quality and consistency checks that the observations are subjected to even helps to improve the instrumental record itself. The most important components of reanalysis systems and the resulting added value is illustrated in Figure 1.1.

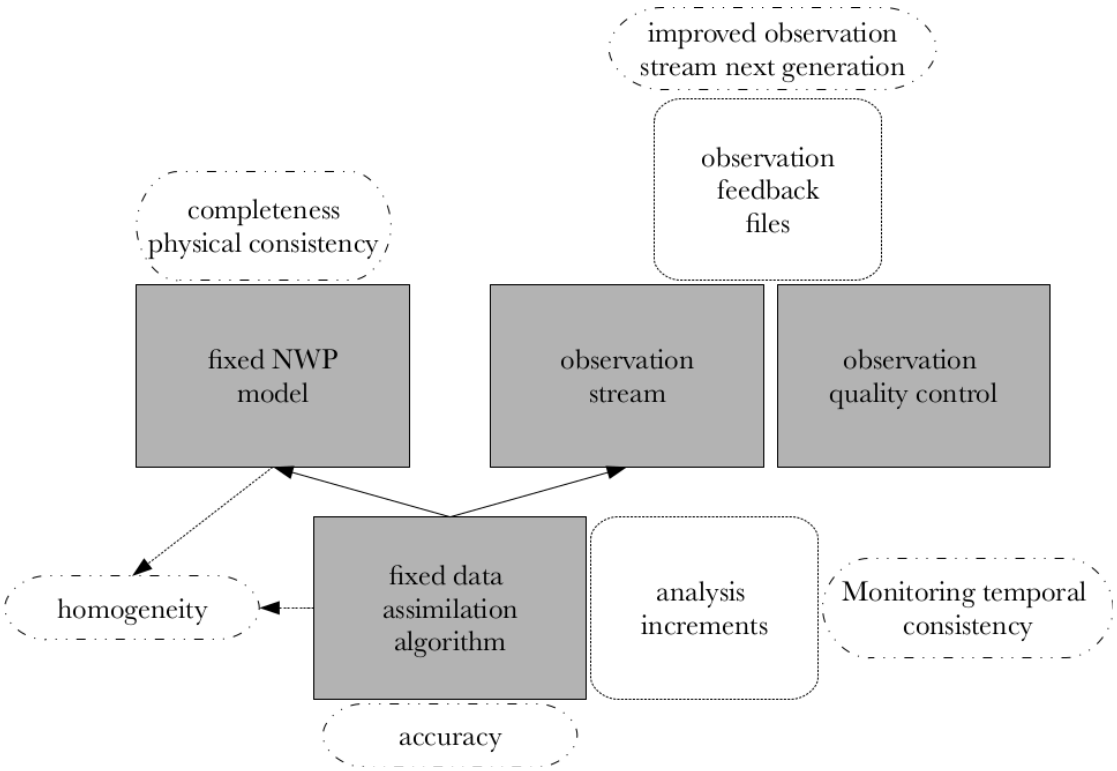

Figure 1.1: Components of a reanalysis system leading to an added value of reanalyses. Concept: The key components are an NWP model, a data assimilation algorithm, observations and an observation quality control. Monitoring the analysis increments[1], temporal consistency can be controlled and biases can be detected on time during the production. Added value: Fixed model and data assimilation versions increase homogeneity. The NWP model enables completeness and physical consistency. Data assimilation adds accuracy.

The pioneering reanalyses ERA-15 (Gibson et al. 1997), NASA/DAO (Bony et al. 1997a,b) and NCEP/ NCAR (Kalnay et al. 1996) have been global data sets with low spatial resolution. Analysing the large and synoptic atmospheric scales, their main objective has been a correct representation of the general atmospheric circulation and global averages.

Towards probabilistic regional reanalyses

In this decade, climate change and extremes on the local scale are attracting increasing notice (Thorne and Vose 2010). This has brought forward the idea of *regional reanalyses* (Mesinger et al. 2006) that zoom in to a certain geographic region of the earth, providing data at a significantly increased spatial sampling. Due to a smaller grid spacing that allows for a better representation of model orography, land use, soil, vegetation, sea surface temperature and land-sea contrasts the influence and interaction with the lower

boundary conditions can be accounted for much better in regional reanalyses. Also, scale interactions within the atmospheric flow which are important for the representation of extremes can be described better. Dynamically, abandoning the assumption of incompressibility and choosing a non-hydrostatic formulation enables the representation of mesoscale processes. For these reasons, the evolution of the atmosphere at a certain point in space can be represented much more accurately than in global reanalyses. This allows e.g. for the regional monitoring of precipitation (Jermey and Renshaw 2016).
A further novel idea that is picked up in the thesis on hand is to generate *ensembles of reanalyses* from which a probability density function (PDF) of the analysed state can be derived. It has only been in the 1990s, that the idea of computing ensembles has been introduced to operational numerical weather prediction in the big operational centres (Ehrendorfer 1997). However, applied to reanalysis computing ensembles is a very new approach, overall because only recent progress in super-computing and data banks allow for the immense computational efforts and storage and retrieval of data required.

Why is computing ensembles of reanalyses essential?

To provide an answer to this question, the importance of the probabilistic approach to numerical weather prediction has to be reviewed first. In the 1960s, the mathematician and meteorologist Edward N. Lorenz found evidence of the deterministic chaos underlying atmospheric prediction (Lorenz 1963). In a numerical experiment, Lorenz observed that trajectories evolving from infinitesimally differing initial conditions diverge until, after a certain integration time, they are fully uncorrelated (Lorenz 1982, 1993). Fundamental aspects of numerical weather prediction became clear: Numerical weather prediction is sensitively dependent to the initial conditions and the model formulation (Lorenz 1990). Further, due to the error growth following from by definition non-perfect initial conditions and non-perfect models, the atmosphere can only be predicted over a limited time span. Note that it can be differentiated between two kinds of uncertainties underlying numerical weather prediction: Aleatoric uncertainty, which is a fundamental property of the atmosphere as non-linear dynamical system that cannot be avoided and epistemic uncertainty, which results from technical short-comings in the practice of NWP that can be reduced in principle, e.g. by improving the model formulation.
Since non-linearities and scale interaction increase towards smaller atmospheric length scales (Lorenz 1969), the predictability of the atmosphere is shortest at the small scales

[1]Note that analysis increments are mostly weighted differences between the prior distribution and the observations. The nudging technique applied in this work corrects the model state continuously during forward integration. Therefore, the corresponding analysis increments represent aggregates of all changes that have been applied to the model state at a certain grid point over the duration of an analysis cycle.

and improves towards the large and synoptic scales. The quality of the initial conditions depends on the capabilities of the data assimilation algorithm estimating the analysis, the density and quality of the assimilated observations, the formulation of the numerical weather prediction model and on the weather regime or the *errors of the day* (Corazza et al. 2003).
It was soon recognized that it would not be feasible to find a fully accurate solution to NWP given these limitations. Therefore, Epstein (1969) suggested to describe the problem of weather forecasting as the evolution of a PDF rather than as deterministic. Thereby, the objective is to predict the predictability of the forecasted evolution of the atmospheric state (see e.g. Ehrendorfer (1997)).

How can ensembles be generated?

Both initial condition and model uncertainty limit atmospheric predictability. Therefore, ensemble forecasts can be generated by perturbing either the initial conditions or the NWP model itself or both. Using initialization techniques like breeding of growing modes (Toth and Kalnay 1997) or singular vectors (Palmer et al. 1997), the directions of strongest error growth in phase space can be estimated and utilized to generate sets of physically meaningful initial conditions perturbing a deterministic analysis. Examples for ensemble generation methods that incorporate model error are perturbed physics ensembles, e.g. employed in Stensrud et al. (2000) or stochastic perturbation of physical tendencies (Buizza et al. 1999). Finally, the lower boundary conditions can be perturbed or in the case of limited-area models probabilistic lateral boundary conditions can be employed to estimate the forecast uncertainty with respect to the uncertainty introduced by the steering model.
A related approach is ensemble data assimilation which provides ensembles of analyses that can directly be employed as initial conditions. The basic principle of ensemble data assimilation methods based on ensemble Kalman filtering is to generate a background forecast ensemble (by means of some of the above explained perturbation strategies) and to forward-integrate it over a short time span, to estimate a flow-dependent background error covariance matrix that incorporates the structure of the error of the day (based on the ensemble perturbations) and to correct the ensemble by combining it with observations (Houtekamer and Mitchell 1998). Ensemble Kalman filtering techniques are advantageous since the resulting analysis ensemble directly reflects the analysis uncertainty that depends on the errors of the day and the density and quality of the assimilated observations. Traditional (non-hybrid) variational techniques employ a climatological forecast error structure (Talagrand and Courtier 1987, Courtier et al. 1998). In ensemble mode every member is updated independently. Such techniques allow e.g. for

the incorporation of observational uncertainty by assimilating perturbed observations. A similar approach is chosen in this work using observation nudging as data assimilation algorithm.

What does this imply for ensembles of reanalyses?
As discussed, reanalyses are estimates of the atmospheric phase space trajectory evolving from data assimilation cycles that combine short-range forecasts with observations. Due to the absence of better estimates, reanalyses are often used as equivalent to truth (Dee et al. 2011b). However, central to the problem of predictability, resulting from a combination of erroneous forecasts and erroneous observations, analyses are likewise only estimates of truth, even though analysis errors are typically significantly smaller than forecast errors. Therefore, it is of great importance to estimate the magnitude and distribution of the reanalysis uncertainty and to provide the resulting uncertainty estimates to users. Just as for forecasts, the objective is to resemble the analysis error of the ensemble mean by means of the ensemble spread to obtain a reliable uncertainty estimate. Ensembles of reanalyses can be generated by means of the same techniques as forecast ensembles. The estimated uncertainty structures are e.g. useful for the renewable energy sector which depends on spatial and situation-dependent inter-variable error correlations of variables like wind and solar irradiation to optimally configure the supply of power to the electricity network.

Scope and outline of this thesis

Reanalysis requires the availability of complex operational production suites whose meaningful and efficient design is a key challenge. Once that a basic suite is established, it can be used to conduct research and to advance the system stepwise. In the framework of the FP-7 funded project Uncertainties in Ensembles of Regional Reanalyses (UERRA), a probabilistic regional reanalysis system has been developed whose domain extends over the European-Atlantic region (CORDEX-EUR11 domain). It is based on the limited-area model COSMO of the COnsortium for Small-scale MOdeling. The thesis on hand is focussed on

- the design of this new probabilistic regional reanalysis suite with dedicated development of a new ensemble data assimilation technique

- an extension of the suite by further ensemble generation methods to enhance the uncertainty estimation.

Initially, the suite will be utilized to produce the test reanalysis COSMO-EN-REA12 that will span the years 2006 to 2010. The main objective of UERRA is the development and preparation of regional reanalysis systems that will be operationally applicable by the Copernicus Climate Change Service (C3S) that is currently built up by the European Center for Medium-Range Weather Forecasting (ECMWF) and the European Commission. The production system has been advanced from a suite developed in the German Hans-Ertel-Centre for Weather Research (Bollmeyer et al. 2015).
The structure of this thesis is as follows: In chapter 2, the state of the art of the field of reanalyses is reviewed. The configuration of the new probabilistic reanalysis suite is outlined in chapter 3. An important point is the introduction of the aforementioned new ensemble data assimilation technique denoted ensemble nudging which allows for estimating the uncertainty of a regional reanalysis with respect to observation uncertainty. This technique will be employed for the production of the aforementioned test reanalysis in UERRA. Chapter 4 is devoted to an evaluation of numerical experiments with the new system. In the first part, experiments conducted with the basic implementation which makes use of ensemble nudging are evaluated with a focus on precipitation. Firstly, it is tested if data sets produced with the new suite bear comparison with related deterministic climate data sets, e.g. the global reanalysis ERA-Interim. Secondly, the probabilistic capabilities of the new system are verified. Then, further ensemble generation techniques are incorporated to assess how the uncertainty estimation capabilities of the suite can be enhanced. The central achievement of this work is a probabilistic regional reanalysis suite for Europe based on the newest model version of COSMO which allows for estimating reanalysis uncertainty given uncertainty in the assimilated observations, in the NWP model and in the lateral boundary conditions. In the final chapter, the main results of this thesis are discussed and ideas for further work are outlined. Note that parts of the abstract as well as all chapters are published in Bach et al. (2016),

2 State of the art

This chapter gives an overview of the state of the art of the field of atmospheric reanalysis. In a first step, an overview of global, regional and probabilistic reanalyses is given. Then, challenges of reanalysis related to climate quality and violation of conservation laws are discussed. In the project UERRA, in which this work has been conducted, intensive discussion on applications and user interaction has taken place. These are briefly discussed for the sake of completeness.

2.1 Global reanalyses

The first generation of global reanalyses including ERA-15 (Gibson et al. 1997), NASA/ DAO (Bony et al. 1997a,b) and NCEP/NCAR (Kalnay et al. 1996) succeeded in representing the major modes of climate variability that are associated with large-scale phenomena on planetary or synoptic scales like the southern oscillation and monsoons (Lahoz et al. 2010). Major problems encountered were the representation of the general circulation, the hydrological cycle and the global observing system. It was soon recognized that the evolution of the global observing system imposes a strong limit on the usefulness of raw reanalysis data for the estimation of climate trends (WCRP 1997, 2000, Kistler et al. 2001). In the second generation, including ERA-40 (Uppala et al. 2005), JRA-25 (Onogi et al. 2007) and NCEP/DOE (Kanamitsu et al. 2002), most of these problems remained unsolved (Lahoz et al. 2010).
The third generation of global reanalyses has shown major improvements, particularly due to a strongly improved and extended observation stream, but also due to advanced model versions (see e.g. Dee et al. (2011b)) and the introduction of variational bias correction (Dee 2005). It comprises ERA-Interim (Dee et al. 2011b), JRA-55 (Kobayashi et al. 2015), NASA/MERRA (Rienecker et al. 2011) and NOAA/CFSR (Saha et al. 2010). Currently, a fourth generation of reanalyses has started to be produced or is planned for the near future. MERRA-2, starting from 1979 is updated in near-real time production. The production of ERA5 has started at ECMWF/Copernicus Climate Change Service (C3S) (Hersbach and Dee 2016). A new Japanese reanalysis JRA-3Q is planned for 2018 (see Simmons (2016)) (see summary in Table 2.1).

Table 2.1: Overview of current global reanalysis data sets

Name	Grid spacing	Time span	Reference
ERA-15	190 km	1979 - 1993	Gibson et al. (1997)
NASA/DAO	210 km	1980 - 1993	Bony et al. (1997a,b)
NCEP/NCAR	210 km	1948 - present	Kalnay et al. (1996)
ERA-40	125 km	1957 - 2002	Uppala et al. (2005)
JRA-25	120 km	1979 - 2014	Onogi et al. (2007)
NCEP/DOE	210 km	1979 - present	Kanamitsu et al. (2002)
ERA-Interim	80 km	1979 - present	Dee et al. (2011b)
JRA-55	60 km	1958 - present	Kobayashi et al. (2015)
NASA/MERRA		1979 - 2016	Rienecker et al. (2011)
NOAA/CFSR	38 km	1979 - 2011	Saha et al. (2010)
MERRA2	50 km	1979 - present	
ERA5	80 km / 30 km	1979 - present	Hersbach and Dee (2016)
JRA-3Q			in planning

Table 2.2: Global reanalyses extending over the whole 20th century

Name	Grid spacing	Time span	Reference
20CR	190 km	1871 - 2012	Compo et al. (2011)
ERA-20C	125 km	1900 - 2010	Poli et al. (2016)
CERA-20C	210 km	1900 - 2010	Laloyaux et al. (2016)

Understanding events like the 1877/1878 El Niño that came along with the Indian famine (Compo et al. 2011), but also trend estimation and understanding of climate variability has raised the need for very-long term global reanalyses. Surface pressure observations have been found to be most suitable for assimilation. On the one hand, the observations are available and recoverable for a very long time span (Cram et al. 2015). This is important in view of the rapid growth of the global observing system. On the other hand, surface pressure provides essential information about the dynamics of the atmosphere approximating the barotropic part of the flow through geostrophy (Compo et al. 2011), while surface pressure tendencies provide information about the divergent part of the flow.

The pioneering century reanalysis has been 20CR of NOAA-CIRES extending from 1871 to 2012 (Compo et al. 2011). The long-term global reanalysis of ECMWF comprises both pure model simulations ERA-20CM with external forcings (Hersbach et al. 2015) and a genuine reanalysis data set ERA-20C (Poli et al. 2016). Different from 20CR, ERA-20C also incorporates marine wind observations. ECMWF has recently replaced ERA-20C by CERA-20C which is based on a fully-coupled ocean-atmosphere system (Laloyaux et al. 2016). First evaluation has shown pronounced added value (Laloyaux 2016) over

Table 2.3: Overview of current regional reanalysis data sets

Name	Grid spacing	Time span	Reference
NARR	32 km	1979 - present	Mesinger et al. (2006)
Arctic system reanalysis	30 km	2010 - 2011	Bromwich et al. (2016)
SMHI Hirlam	22 km	1989 - 2010	Landelius et al. (2016)
	5 km surface		Dahlgren et al. (2016)
COSMO-REA6	6 km	1995 - 2014	Bollmeyer et al. (2015)
COSMO-REA2	2 km	2007 - 2013	Wahl et al. (2017) (accepted)
Met Office	24 km	2008 - 2009	Jermey and Renshaw (2016)
Safran	8 km	1979 - present	Vidal et al. (2010)
Mera	2.5 km	1981 - 2015	Whelan and Gleeson (2016)

ERA-20C (Poli et al. 2016). The currently available reanalyses extending over the whole 20th century are summarized in Table 2.2.

2.2 Regional reanalyses

The field of regional reanalysis (see Table 2.3) has been pioneered by the North American Regional Reanalysis (Mesinger et al. 2006). An important reanalysis project is the Arctic system reanalysis (Bromwich et al. 2016). In the framework of the EU-funded project European Reanalysis and Observations for Monitoring (EURO4M), regional reanalysis systems have been set up by the Met Office (Jermey and Renshaw 2016) and by the Swedish Meteorological and Hydrological Institute (SMHI) (Landelius et al. 2016, Dahlgren et al. 2016). Additionally, the Hirlam reanalysis by SMHI has been downscaled and combined with surface observations by means of optimal interpolation using MESAN-Safran (Soci and Bazile 2013).

A reanalysis effort for Europe based on COSMO has been undertaken in the Hans-Ertel-Centre for Weather Research (Bollmeyer et al. 2015, Simmer et al. 2016). Concurrently, high-resolution regional reanalyses have been produced for single countries, including a very-high resolution regional reanalysis for Germany based on COSMO (Wahl et al. (2017), accepted) and for France (Vidal et al. 2010) using Safran. Met Eireann has produced a regional reanalysis for Ireland and Great Britain using Hirlam (Whelan and Gleeson 2016). At the moment, a project for a regional reanalysis of similarly high resolution is set up for Scandinavia (personal communication, Ole-Einar Tveito, Met Norway). The Copernicus Climate Change Service will commission the production of regional reanalyses to European met services in the foreseeable future.

2.3 Probabilistic reanalyses

To estimate the uncertainties underlying reanalyses, the idea of drawing multiple realizations of reanalyses has been adapted from ensemble forecasting in NWP in recent years. The essential epistemic uncertainties (i.e. uncertainties that result from short-comings in the NWP practice, but can be eliminated in principle) that can directly be accounted for by operation of different perturbation schemes are errors in the assimilated observations, model uncertainty arising for example from physical parameterization of the sub-grid scales and errors in the lower and possibly lateral boundary conditions (in the case of regional reanalyses).
Currently, the global ensemble reanalysis ERA5 is produced by ECMWF/Copernicus Climate Change Service (Hersbach and Dee 2016). It incorporates uncertainties in sea surface temperature, observations as well as in model physics (personal communication, Andras Horanyi and Hans Hersbach, ECMWF).
In UERRA, regional probabilistic reanalysis systems are set up, some of which are intended to be further developed and employed for long-term productions that will be ordered by the Copernicus Climate Change Service. The ensemble systems are run for a test period of at least 2006 and 2010 to learn about the individual strengths and weaknesses in a comparison. The developed systems comprise an 8-member system based on ALADIN that aims to characterize the impact of perturbed observations, observation density and physical parameterizations. Additionally, a mini ensemble comprising two members is produced at SMHI, one of which makes use of ALADIN and one of ALARO physics. At the Met Office, a 10 member ensemble system is set up based on UKMO and 4D-Var. It accounts for uncertainties in the lower boundaries, model error and for observational uncertainty. A further reanalysis system developed in UERRA is introduced in this thesis and in Bach et al. (2016). For an overview of the probabilistic reanalyses that are currently produced in UERRA see http://uerra.eu/work-packages/wp2/23-description.html.

2.4 Challenges in reanalysis

Despite the obvious benefits of reanalyses, the developments in the field are far from being concluded. As mentioned in the introduction, reanalyses have been proposed for climate change studies initially (Trenberth and Olson 1988, Bengtsson and Shukla 1988). This requires so-called climate quality, i.e. long-term homogeneity of the data set to allow for a reliable estimation of climate trends (Dee et al. 2011b). A first step towards

this goal is the use of a fixed NWP system. The rapid growth of the global observing system, however, is what turns the reanalysis problem complicated (WCRP 1997, 2000, Newson 1998, Kistler et al. 2001).

Along the instrumental record, the number of observations has grown by four orders of magnitude. The evolution of the observing system is schematically illustrated in Figure 2.1. A tremendous jump occurs at the beginning of the satellite era in 1979. In ERA-40, it manifests as an observational shock triggered by a systematic error reduction in the presence of model bias, see Bengtsson et al. (2004) and Figure 1 in Dee (2005). In NWP, this effect, denoted positive observation impact, is the fundamental criterium for the operational assimilation of a new observation type. In reanalysis, however, the repetitive introduction of new observational sources leads to superimposing systematic changes in the error characteristics of the data set. To avoid jumps at the beginning of the satellite era, many reanalyses start only in 1979 (see Tables 2.1 and 2.3). At ECMWF, the strategy is full transparency (Dee et al. 2011a). Any change in error characteristics induced by new sources of uncertainties is quantified in observing system experiments. Moreover, all available meta-data related to the observations is provided to the users in observation feedback data to allow for post-processing such as subdividing the data in quasi-homogeneous subsets. In the long-term reanalyses 20CR (Compo et al. 2011), ERA-20C (Poli et al. 2016) and CERA-20C (Laloyaux et al. 2016), the exclusive assimilation of surface-observations to increase homogeneity has proven feasible (see section 2.1). However, due to the strongly reduced number of observations, the accuracy of the data sets can be expected to be decreased compared to other reanalyses (Dee et al. 2011a).

Obviously, intensive work on the observation input stream is required if climate quality is strived for. Important methods are variational bias correction (Dee and Uppala 2008) and homogenization using innovation statistics. A prominent example is the spatio-temporal homogenization of the radiosonde input stream for ERA-Interim using observation feedback data from ERA-40 (Haimberger 2007). These homogenized radiosonde data have in turn provided the benchmark for a variational bias-correction of satellite radiance data during the production of ERA-Interim.

A further challenge of reanalysis is the violation of conservation laws through data assimilation (Lahoz et al. 2010). In the presence of model bias, data assimilation introduces source and sink terms for energy, mass and momentum. These propagate through the whole model system (Dee 2005), e.g. leading to spin-up effects. The energy budgets and deficient representation of the hydrological cycle in global reanalyses has e.g. been found by Trenberth et al. (2009), Trenberth, K. E. Smith (2008), Betts et al. (2006), Bengtsson et al. (2007). While conservation is less important for NWP, climate studies depend

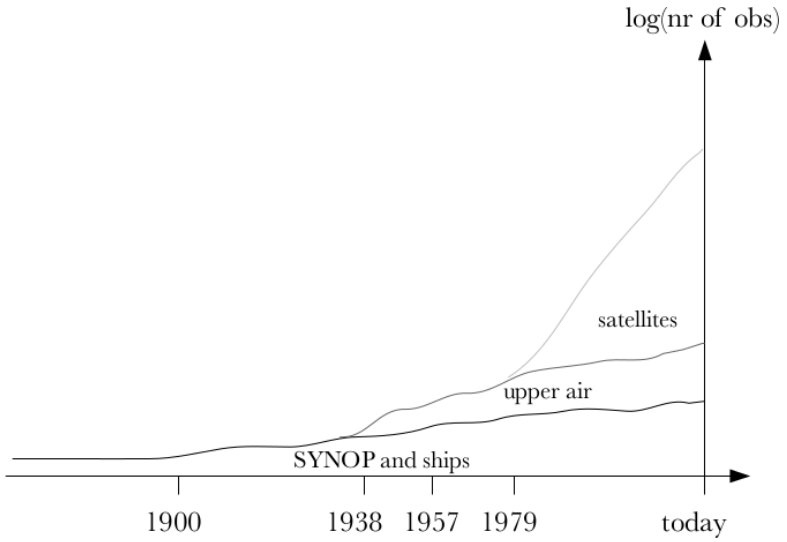

Figure 2.1: Temporal evolution of the global observing system, adapted from a graphic of Dick Dee in the presentation of Paul Poli in ECMWF data assimilation course 2015. 1938: First radiosonde networks. 1957: extension of radiosonde network especially on southern hemisphere in the framework of the International Geophysical Year. 1979: Beginning of the satellite era. Today: High density of data such as modern remote sensing observations, aircraft data, buoys.

heavily on conserved quantities (Lahoz et al. 2010). A new approach to improve conservation is to develop earth system models that fully describe all climate sub-systems including atmosphere, ocean, trace gase, aerosols and the land surface. Through a more realistic description of fluxes at the interfaces between the systems, this has the potential to enable conservation (Laloyaux et al. 2016). Additionally, fully-coupled data assimilation methods have to be developed. Ideally, these ensure that any change applied in one climate sub-system entails a change in the others. However, such an approach requires the estimation of accurate flow-dependent cross-correlations between the systems which is only at an early stage of research.

2.5 Applications and user interaction

Reanalysis data provide a fundamental basis for climate research, but are alike needed in many different sectors of public, e.g. in renewable energies, agriculture, risk management, the water sector and insurance companies (see report[1] on user requirements by CORE-CLIMAX[2] (COordinating earth observation data validation for RE-analysis for

[1]http://www.coreclimax.eu/sites/coreclimax.itc.nl/files/documents/Deliverables/WP_Reports/Deliverable-D552-CORECLIMAX.pdf)
[2]http://www.coreclimax.eu/

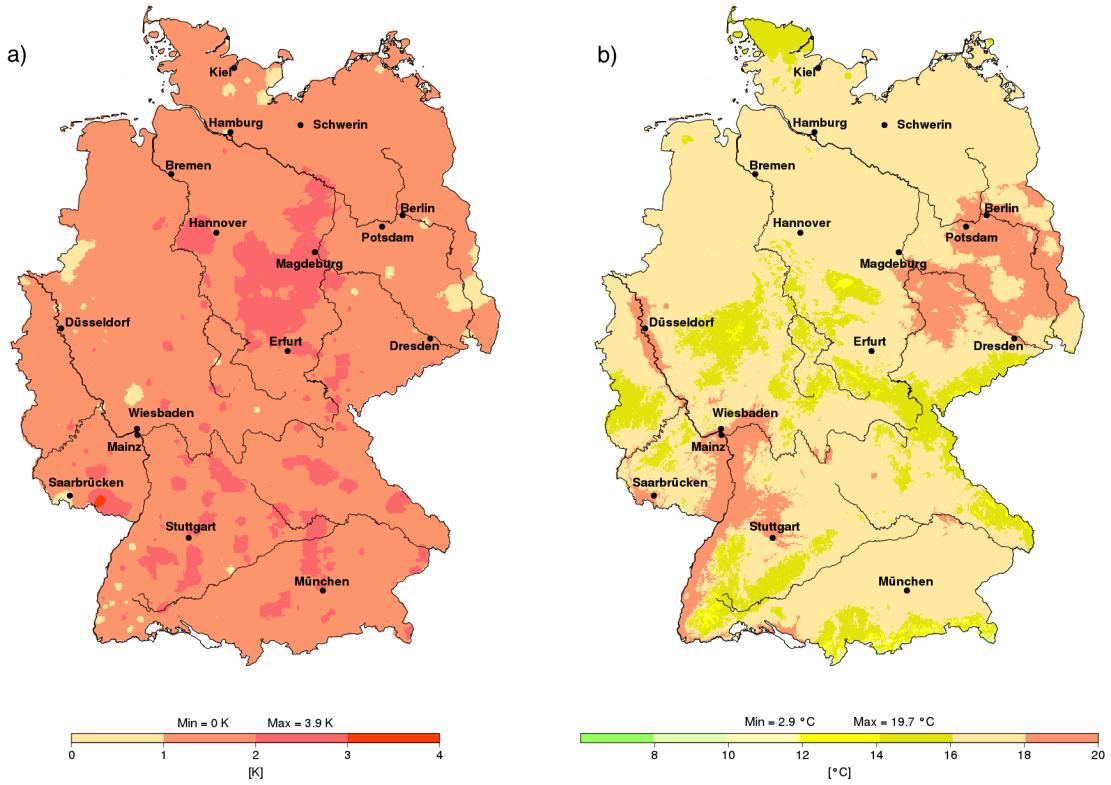

Figure 2.2: Temperature maps from Deutscher Klimaatlas for month of July. a) Long time average (1961 - 1990), b) anomaly of July 2016. Based on observation data gridded on 1-km Gauss-Krüger grid under consideration of orography. This may be based on high-resolution regional reanalysis data in the future.

CLIMAte ServiceS)). In climate research, global reanalyses are prominently used for the investigation of low-frequency variability, trends and the large-scale atmospheric circulation (Mitas and Clement 2006, Wright et al. 2013, Marshall 2003). Verification of climate models (Hodges et al. 2011, Davini et al. 2012), downscaling experiments or downstream applications like hydrological modeling benefit from reanalysis data, see e.g. Soares and Cardoso (2012), Maurer et al. (2001).

In climate monitoring, adaptation and mitigation reanalysis data play an increasingly important role. Climate services disseminate climate information to users on a national (National Climate Monitoring, DWD), regional (Copernicus Climate Change Service, coordinated by the European Commission) and global (Global Framework for Climate Services) level. They provide data sets together with quality estimates as well as climate indices[3] and climate impact indicators[4] based on essential climate variables (Bojinski

[3]http://climatedataguide.ucar.edu/climate-data/overview-climate-indices
[4]http://eea.europa.eu/data-and-maps/indicators/

et al. 2014). Essential climate variables are a set of physical, biological and chemical variables defined by the Global Climate Observing System (GCOS) that ought to help "to understand and predict the evolution of climate, to guide mitigation and adaptation measures, to assess risks and enable attribution of climatic events to underlying causes, and to underpin climate services." (Bojinski et al. 2014). Climate indices are derived from meteorological variables, e.g. European temperature, heat waves, droughts, hail events and extreme precipitation. Climate impact factors relate essential climate variables to socio-economic variables to provide information about health impact due to ozone, water-limited crop productivity or heat stress in urban areas. Figure 2.2 shows an example from the Klimaatlas of the National Climate Monitoring of Deutscher Wetterdienst. Illustrated are the long-time average of screen-level temperature for Germany in July as well as the anomaly of July 2016. Currently, this is based on traditional in-situ observations. However, the potential of reanalyses for such applications is being examined (personal communication, Frank Kaspar, DWD).

In recent years, users are getting increasingly involved in reanalysis efforts. In UERRA, a whole working group is dedicated to user interaction, verification and outreach. A user workshop[1] has taken place in February, 2016. Such meetings are particularly important to orientate the developments more towards the user requirements. At a regional reanalysis workshop[1] hosted by the Copernicus Climate Change Service[2], the idea of involving specific user groups in tentative verification of preliminary studies ahead of production has been raised.

[1] http://uerra.eu/project-meetings/user-worshop-1.html

[1] https://climate.copernicus.eu/events/regional-reanalysis-workshop of the Copernicus Climate Change Service

[2] http://climate.copernicus.eu/what-copernicus

3 The probabilistic regional reanalysis suite

In this chapter, the newly developed probabilistic reanalysis suite based on COSMO is introduced. Note that it has been advanced from a deterministic nudging regional reanalysis system (Bollmeyer et al. 2015). The chapter is structured as follows. Firstly, the numerical weather prediction model COSMO is introduced, where special emphasis is put to the dynamics and numerics as well as parameterizations. Further, the model domain and the lateral boundary conditions are discussed. An overview of the atmospheric data assimilation is followed by a discussion of ensemble generation techniques that account for different uncertainty sources relevant to regional reanalyses. In the subsequent section, the analysis of the lower boundary condition including soil moisture and temperature, sea surface temperature and snow is described, followed by an overview of the process cycle of the newly implemented reanalysis suite integrating the aforementioned components. Note that two versions of the probabilistic regional reanalysis suite have been implemented. In the last section of this chapter, the differences between the two are outlined.

3.1 The numerical weather prediction model

COSMO is the limited-area numerical weather prediction model developed and maintained by the Consortium for Small-Scale Modelling. It is targeted at the representation of atmospheric meso-β and meso-α-scale processes. COSMO is run in daily operations of the operational centers that take part in the COSMO consortium, but can also be run in climate mode (COSMO-CLM). To allow for direct comparison to other European reanalyses or downscaling experiments, the model is set up for the CORDEX-EUR11 domain (Giorgi et al. 2009) at a horizontal grid spacing of 12 km. The domain is shown in Figure 3.1 and its specifications are summarized in Table 3.1.

Table 3.1: Domain specification

	CORDEX-EUR11
Rotated North pole	−162.0, 39.25
Lower left corner in rotated coordinates	−23.375, −28.375
Grid spacing	0.11°
Number of grid points	424 × 412
Number of vertical levels	40
Top of the model atmosphere	22700 m

3.1.1 Dynamics and numerics

The dynamical core of COSMO comprises prognostic equations for temperature, pressure deviation from a reference profile, the three-dimensional wind vector, turbulent kinetic energy, and specific contents of water vapour, cloud water, cloud ice, rain, snow and graupel. Their derivation is detailed in Schättler et al. (2011). Firstly, basic budget equations are formulated for momentum, mass and heat. In a next step, they are Reynolds-averaged, i.e. split into a mean flow and zero-mean deviations. Thereby, it is accounted for the fact that the horizontal atmospheric scales that describe mesoscale atmospheric motion range from 100 m to 100 km. The resulting Reynolds fluxes and stresses, e.g. turbulent fluxes of momentum, heat and constituents that describe the feedback of the sub-grid scales on the resolved scales, are parameterized (see 3.1.2). To allow for an application of the model to a wide range of scales, a non-hydrostatic form of the equations is used. Apart from the assumption of a shallow atmosphere, no scale-dependent approximations are applied. The thermodynamic variables are formulated as sums of base-state variables and deviations from the base state, where the latter is horizontally homogeneous, time invariant, hydrostatically balanced and only depends on the height above the surface. Since a linearization based on the anelastic approximation is not considered reasonable due to the large extension of the European COSMO domain, the main reasoning is that the horizontal base-state pressure gradient terms can be removed in the equation of motion which reduces the error arising in the computation of the pressure gradient force in terrain-following coordinates.

Horizontally, the set of equations is transformed to a spherical rotated latitude-longitude grid which prevents numerical problems related to a convergence of the meridians. To include surface terrain, the equations are vertically formulated in a terrain-following generalized Gal-Chen hybrid coordinate (Gal-Chen and Somerville 1975). Discretization on a staggered Arakawa-C grid (Arakawa and Lamb 1981) using finite differences yields a rectangular and regular computational grid. To improve the numerical efficiency in view

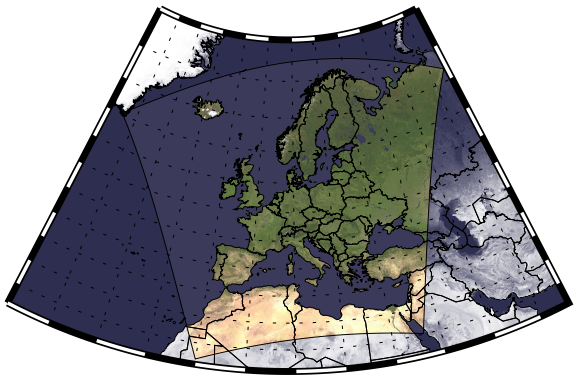

Figure 3.1: CORDEX-EUR11 domain.

of fast sound waves that are part of the solution due to the compressibility of a non-hydrostatic model atmosphere, mode-splitting is applied for the temporal integration of the equations (Klemp and Wilhelmson 1978). Thereby, the equation terms related to sound waves are integrated in small time steps while the stable slow terms are integrated in bigger time steps. The numerical integration scheme is a two time-level, third-order Runge-Kutta time-split scheme (Wicker and Skamarock 2002). For advection, a second-order finite-volume scheme with strang-splitting is employed (Bott 1989). Further details on the dynamics and numerics can be found in Doms and Baldauf (2015).

3.1.2 Physical parameterizations

Reynolds-averaging of the governing equations yields sub-grid scale terms which need to be parameterized. These include turbulent mixing terms, clouds and precipitation related to microphysical processes, moist convection, radiative processes, effects of sub-grid scale orography and fluxes at the soil-atmosphere and ocean-atmosphere interface. Optionally, the effects of lakes and sea ice are parameterized.
In the employed configuration of COSMO, cloud microphysics are parameterized by a bulk-water continouity model. For the turbulent fluxes of momentum, heat and humidity, a 1-D diagnostic closure is employed in the horizontal. It utilizes the boundary layer approximation that assumes horizontal homogeneity so that all horizontal turbulent fluxes vanish. The vertical turbulent transport is computed using a second-order closure by Mellor and Yamada (1974) that is nothing else than a K-closure. Prandtl layer surface fluxes of sensible heat, turbulent momentum and water vapour are formulated in dependency of stability and roughness-length based on Louis (1979). The turbulent fluxes of liquid and solid water are neglected.

The radiative transfer scheme is based on the δ-two-stream solution of the radiative transfer equation for plan-parallel horizontally homogeneous atmospheres (Ritter and Geleyn 1992). It provides the heating rate due to radiative effects to the prognostic equation of temperature and is updated in 15-minute intervals.
To account for a form drag of low-level flow due to blocking through pronounced sub-grid scale orography and a generation of gravity waves, the sub-grid scale orography scheme by Lott and Miller (1997) is employed.
Subgrid convection is parameterized by the Tiedtke scheme (Tiedtke 1989) and the vertical turbulent diffusion is computed based on a prognostic equation for turbulent kinetic energy utilizing a level 2.5 closure scheme (Doms et al. 2011).
The lower boundary condition, i.e. soil and vegetation whose presentation is crucial for the soil-atmosphere coupling is computed by the model TERRA_ML (Schrodin and Heise 2001) which employs 7 vertical layers. It provides surface temperature and specific humidity at the ground based on the equation of heat conduction. Using Richard's equation, the soil water is predicted. As diagnostic variables, evaporation and transpiration of plants are derived from soil water content and in parts from radiation and temperature.
To represent fluxes between lakes and atmosphere, the fresh-water lake parameterization scheme FLake described in Mironov (2008) can be switched on. Application of the sea ice scheme developed by Mironov and Ritter (2004) allows to account for changed sea-atmosphere interactions related to the interface's albedo and roughness length in presence of sea ice. For details on the parameterization schemes see Doms et al. (2011).

3.1.3 Lateral boundary conditions

The lateral boundaries and initial conditions are provided by the global ECMWF reanalysis ERA-Interim (Dee et al. 2011b). 0.5° ERA-Interim analysis fields from 00 and 12 UTC are used, as well as +03-h, +06-h and +09-h reforecasts which comprise the three-dimensional fields specific humidity, temperature, pressure, wind, cloud liquid and ice water content. Additionally, the surface variables skin temperature, 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 are included in the lateral boundary data.
At the lateral boundary conditions, a one-way nesting following Davies (1983) is applied. The relaxation zone near the boundaries is configured such that it consists of 7 grid points which is approximately 85 km. Additionally, the model fields are drawn towards the boundary data sets of the driving model within an upper boundary Rayleigh

relaxation zone reaching from the model top down to approximately 235 hPa.
The regional reanalysis system does not yet provide the opportunity to assimilate satellite data. However, through the lateral boundary conditions, it is provided with valuable synoptic-scale information from various satellite products (Dee et al. 2011b) that force the mesoscale dynamics.

3.2 Analysis of atmospheric variables

As discussed, reanalyses evolve from a physically consistent reprocessing of past observations of the atmospheric state through a numerical weather prediction model using a data assimilation algorithm. The latter is presented here. Further, the observation stream as well as observation pre-processing and quality control are discussed.

3.2.1 Observation nudging

The data assimilation scheme incorporated in the COSMO code is observation nudging (Schraff 1996, 1997, Schraff and Hess 2012). It has been used in daily operations of DWD since 1999 and has proven to be useful for reanalysis purposes (Bollmeyer et al. 2015). In general, nudging has been used for many applications up to present including research and development of NWP (Stauffer and Seaman 1990, 1991, Seaman et al. 1995, Schraff 1997, Schroeder et al. 2006, Deng et al. 2004, Leidner et al. 2001), research in the area of hybrid data assimilation methods (Lei et al. 2012b,a), initialization of climate runs (Otte et al. 2012, Baehr et al. 2014) and in ocean data assimilation (Chen et al. 2013). Even though it is not mathematically optimal in a least-squares or maximum-likelihood sense, at least in its current implementation in COSMO, nudging has a good performance-cost ratio that argues for its usability for the purpose of ensemble reanalyses.
Observation nudging performs a continuous relaxation of the model variables towards observations. This is realized through introduction of tendency terms proportional to observation-model equivalent departures to the prognostic equations of the model, so that the model is continuously nudged towards the observations during its forward integration. The update equation for any prognostic variable ψ in model space is given by

$$\frac{\partial \Psi(x,t)}{\partial t} = F(\Psi, x, t) + G_y \cdot \sum_k W_k(x,t) \cdot [\Psi^k - \Psi(x_k, t)], \tag{3.1}$$

where F represents the tendencies from the resolved and the parameterized sub-grid processes and the right-hand term the analysis increment. It compounds of observation increments being deviations between k observations Ψ^k that influence a model grid point

Table 3.2: Observation stream for reanalysis

Observing system	Report type	Observed variable
Radiosondes	PILOT	Upper-air wind
	TEMP	Upper-air wind, temperature, humidity
		Screen-level wind, humidity, geopotential
Aircraft	AIREP	Wind, temperature
	AMDAR	Wind, temperature
	ACARS	Wind, temperature
Wind profiler		Upper-air wind
Surface systems	SYNOP	Surface pressure, wind, humidity
	SHIP	Surface pressure, wind, humidity
	DRIBU	Surface pressure, wind, humidity

x and the model equivalents $\Psi(x_k, t)$, i.e. the model state interpolated to the observation locations. The observation increments are weighted according to their horizontal and vertical distance from the observation location using autoregressive structure functions. Moreover, a quality weight and a temporal weight with a maximum value at the observational time are applied. The sums of the weighted observation increments are then multiplied by the nudging coefficient G_y to obtain the analysis increments. G_y has units of inverse time and determines the characteristic relaxation time scale which is chosen to be about 30 minutes. Before the resulting analysis increments are added to the prognostic variables, explicit balancing is applied between the analysis increments that correspond to different prognostic variables. This includes a hydrostatic temperature correction balancing the near-surface pressure increments and a geostrophic wind correction.

A main advantage of nudging is the continuous and slow correction of the model that allows for maintaining the dynamical balances between the variables through adaption of the model state to the corrections. This is different from methods in which the analysis increments are impressed within one time step. These often suffer from substantial spin-up effects. For further details on the implementation of nudging in COSMO please refer to Schraff (1997) or Schraff and Hess (2012).

3.2.2 Observation stream for atmospheric data assimilation

The observation stream that is assimilated consists basically of conventional observations. The observation types and corresponding variables are summarized in Table 3.2. The upper-air observations comprise radiosonde ascents (TEMP) and pilot balloon ascents (PILOT) as well as aircraft observations which can be classified into 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 include manual and automatic reports from SYNOP stations, manual and automatic SHIP reports and drifting buoys (DRIBU). Additionally, wind profiler data are assimilated.

3.2.3 Observation pre-processing and quality control

The implementation of the new reanalysis suite benefits from the quality-controlled observation stream of COSMO-REA6 (Bollmeyer et al. 2015). The observations assimilated there and corresponding quality flags are directly accessed from feedback observation files. Thereby, the largest part of the observation pre-processing and quality control can be skipped which saves a significant amount of computational time. Some steps, like the assignment of the observations to the model grid due to a different grid spacing have to be repeated. For the sake of completeness, the most important steps of pre-processing and quality control are reviewed in this section.

The pre-processing before data assimilation includes a retrieval of reports that provide the observations, reading them, assembling different reports, checking the quality of the observations, checking their position and assigning them to model grid points and finally providing all information to the data assimilation.

Firstly, the reports are assigned to grid points where particular emphasis is placed to the vertical representativeness. This is e.g. important for humidity observations from mountain stations in presence of low-level inversions which is complicated by significant differences between the true terrain and model orography. Also, sea and land observations are assigned to appropriate model grid points and surface observations are checked for their usefulness in terms of orography differences. Observations that are assimilated in different physical units than reported are converted. Further, radiosonde and aircraft reports, which are messaged in four unconnected reports are assembled and superadiabatic lapse rate checks are performed. Aircraft reports are thinned and their vertical correlation scale is scaled down to reduce the smoothing effect of the dense observational information on the vertical profile. Humidity observations from radiosondes are bias-corrected. The assignment of certain quality flags (active, passive or rejected) is

explained which is encoded as numbers, e.g. a report is rejected because its height is below the lowest model level. Further, observations that lie within 5 grid points from the lateral boundary conditions are rejected. The quality control includes consistency checks that are performed between individual observations and the surrounding model grid points using prescribed observation error standard deviations to compute thresholds. Neighboring observations are used to perform buddy or spatial consistency checks which is particularly important for surface pressure and integrated water vapor (precipitable water). Gross error checks are performed to find observations that exceed reasonable limits (e.g. pressure more than 1060 hPa). In regions or reports with increased data density redundancy checks determine observations that do not contain complementing information and thus can be sorted out. The observation pre-processing is detailed in Schraff and Hess (2012).

3.3 Ensemble generation methods

Uncertainties in regional reanalyses arise from nonlinear scale-interaction of the atmospheric dynamics and from errors in the assimilated observations, errors in the model formulation, numerical errors and errors in the lower and lateral boundary conditions. In the work on hand, ensemble generation methods that yield equally likely ensemble members are chosen. The objective is that the members exhibit temporal deviations that reflect random error, but are equally accurate on average. In the implementation of the reanalysis suite that will be employed for the production of a test reanalysis in UERRA (denoted *basic reanalysis suite*), an ensemble nudging technique that has been newly developed in the course of this work accounts for errors in the assimilated observations. In an experimental version of the suite that may be employed for future long-term regional reanalyses (denoted *extended reanalysis suite)*, options to stochastically perturb the physical tendencies and to use an ensemble of lateral boundary conditions have been implemented. Thereby, model error and uncertainties in the lateral boundary conditions can be accounted for.

3.3.1 Ensemble nudging

The most obvious way for generating a nudging ensemble is to nudge the ensemble members towards perturbed observations. This ensemble generation technique yields an estimation of the uncertainty of a nudging reanalysis given observation uncertainties. A perturbed observation is obtained by perturbing the original observation o by means of a perturbation o' sampled from a normal distribution $o' \sim N(0, \sigma_o^2)$ with zero mean

Table 3.3: Observation error standard deviations for sea observing systems; abbreviations for variables: U/V=wind components, Φ=geopotential

Obs type	Variable	Error standard deviation
DRIBU	U/V[m/s]	5.4
DRIBU	Φ[m]	14.0
SHIP	Φ[m]	14.0

and an estimated observation error variance σ_o^2. This approach has first been chosen by Environment Canada to incorporate observation uncertainty in the analysis based on an Ensemble Kalman Filter (Houtekamer et al. 1996). An observation error as defined in the context of data assimilation consists of a measurement component and a representativity component, where the latter includes an error of the observation operator computing the model equivalents at the observation locations (Hollingsworth and Lönnberg, 1986). The employed observation error standard deviations are summarized in Tables 3.3 and 3.4. They have been adopted from the operational nudging scheme which relies on retuned estimates from former ECMWF and DWD global data assimilation systems.
As is the state-of-the-art procedure for perturbation of observations that exclude satellite data (which are correlated between the channels) normally distributed, unbiased, stationary in time as well as spatio-temporally uncorrelated observation errors are assumed. Possible spatial correlations are reduced by observation thinning. Vertical profiles from upper-air systems are checked for super-adiabatic lapse rates inadvertently generated by perturbation and are corrected accordingly before assimilation. The observations are retrieved from feedback observation files from the reanalysis COSMO-REA6. The quality flags are retained and no further quality control is performed after perturbation so that the uncertainty estimated by the ensemble arises only from observation error and not from sets of observations differing between the ensemble members. The ensemble members are perturbed and run in parallel independent streams that do not exchange information.

3.3.2 Stochastic perturbation of physical tendencies

Stochastic perturbation of physical tendencies (SPPT) allows for representing model error due to uncertainties in the sub-grid scale parameterized processes. SPPT perturbs the sum of all parameterized tendencies in the prognostic equations by multiplicative noise that varies smoothly in space and time. The method has first been introduced to the medium range ensemble prediction system of the ECMWF (Buizza et al. 1999). It has proven to have a positive impact on both reduction of analysis errors when applied

Table 3.4: Observation error standard deviations for upper-air systems; abbreviations for variables: U/V = wind components, Φ = geopotential, T = temperature, RH = relative humidity; abbreviations for observation types: T/P/W=TEMP/PILOT/wind profiler, AI=AIREP, SY=SYNOP

Lev[hPa]	U/V[m/s]	Φ[m]	T[K]	RH[]	T[K]	UV[m/s]	Φ[m]	U/V[m/s]
	T/P/W	T	T	T	AI	AI	SY	SY
1000	2.0	4.3	1.2	0.09	1.2	2.5	7.0	3.6
850	2.4	4.4	1.0	0.13	1.0	2.5	8.0	3.6
700	2.5	5.2	0.7	0.12	0.7	3.0	8.6	5.8
500	3.4	8.4	0.4	0.13	0.5	3.5	12.1	6.8
400	3.5	9.8	0.4	0.13	0.5	4.0		
300	3.7	10.7	0.5	0.14	0.6	4.0		
250	3.5	11.8	0.5	0.14	0.6	4.0		
200	3.5	13.2	0.6	0.14	0.7	4.0		
150	3.4	15.2	0.7	0.14	0.8	4.0		
100	3.3	18.1	0.8	0.14	0.9	4.0		
70	3.2	19.5	0.8	0.14	1.0	4.0		
50	3.2	22.5	0.9	0.14	1.1	4.0		
30	3.3	25.0	0.9	0.14	1.1	4.0		
20	3.6	32.0	1.0	0.14	1.2	4.0		
10	4.5	40.0	1.2	0.14	1.4	4.0		

in parallel with ensemble data assimilation (Schraff et al. 2016), but also to forecast skill. Further, it has a positive impact on the reduction of model biases (Jung et al. 2005, Palmer et al. 2004, Berner et al. 2009). The implementation of SPPT in COSMO follows closely the one in IFS, however there are small deviations. Therefore, both are summarized briefly. In IFS, the perturbations are applied collinearly to the original tendencies X_c so that the perturbed tendency reads

$$X_P = (1 + r\mu)X_c. \tag{3.2}$$

Perturbed are the zonal and meridional velocity components, temperature and specific humidity. All variables are perturbed by means of the same random number r so that the perturbations of all variables are perfectly correlated. By the factor μ the perturbation amplitude is tapered to zero near the earth's surface and in the stratosphere to maintain physical realism and reduce the risk of numerical instability. To introduce spatio-temporal correlation, the random numbers are drawn using a spectral pattern generator (Berner et al. 2009, Li et al. 2008). The spectral coefficients of the pattern are described by an autoregressive process of first order AR(1), which employs random numbers whose real and imaginary parts are independent Gaussian random numbers

with unit variance and zero mean. Further, they are white in time and independent between different spherical harmonics. The temporal correlation between two time steps is formulated as exponential of an inverse decorrelation time scale. For details be referred to Palmer et al. (2004).
In COSMO, the same form of perturbed tendency equation is utilized and the same variables are perturbed. The random number generation, however, is simplified. In a first step, random numbers are drawn for each grid point of a coarse grid from a normal distribution with prescribed standard deviation and maximum amplitude. To introduce spatial correlation and obtain smooth fields, the random numbers are then interpolated to the original grid. The same field is used for all vertical levels, however, it is tapered to zero towards the upper and lower boundaries of the model. Every N time steps a new field is used for perturbation. In between the fields are interpolated linearly. For further details see Torrisi (2012).

3.3.3 Probabilistic lateral boundary conditions

In the extended set-up of the reanalysis suite for COSMO-EN-REA12 an ensemble of lateral boundary conditions from the ICOsahedral Non-hydrostatic model (ICON) of DWD and the Max-Planck-Institute for Meteorology (MPI) is used. ICON is a unified modeling system for both numerical weather prediction and climate applications. Its numerical core is formulated on an icosahedral-triangular Arakawa-C grid. This allows for a straight-forward nesting-based mesh refinement. For further information be referred to Zängl et al. (2015). Here, an ensemble produced with the ICON ensemble data assimilation (ICON-EDA) is employed. The latter is a hybrid combination of 3D-VAR and a local ensemble transform Kalman filter (LETKF). In the utilized experiment, the ensemble has been initialized by random perturbations of the initial conditions and perturbations applied to the fields of SST (personal communication, Alex Cress, DWD).

3.4 Analysis of sea and land-surface variables

The surface of the earth interacts with the atmosphere. It represents the lower boundary condition to all atmospheric processes and thus has to be updated on a regular basis to improve the turbulent fluxes at the surface- atmosphere interface. For that purpose, different offline schemes are applied to COSMO, namely an analysis of snow depth, a sea surface temperature (SST) analysis and a variational soil moisture analysis. Each member is updated separately by these external analyses so that the ensemble members do not exchange any information and are fully independent.

3.4.1 Analysis of soil moisture

Soil moisture influences screen-level temperature and humidity, particularly at clear-sky days with strong soil-atmosphere coupling due to high radiative impact. Moreover, evaporation and thus precipitation depend on soil moisture. To avoid the evolution of systematic errors, the soil water content is analysed based on a variational assimilation of screen level temperature, whereby a horizontal decoupling is assumed to reduce the minimization problem. The cost function for the soil moisture η at any horizontal grid point is given by

$$J(\eta) = \frac{1}{2}(T^o - T(\eta))^T \mathbf{R}^{-1}(T^o - T(\eta)) + \frac{1}{2}(\eta - \eta^b)^T \mathbf{B}^{-1}(\eta - \eta^b), \tag{3.3}$$

where T is the screen level temperature predicted for noon, T^o are observations of screen-level temperature and η and η_b are vectors whose dimension equals the number of analysed soil levels. The gradient of the cost function 3.3 can be solved analytically if the linearization $T(\eta) \doteq T(\eta_b) + \Gamma(\eta - \eta_b)$ is applied which leads to

$$\eta_a = \eta_b + (\Gamma \mathbf{R}^{-1} \Gamma + \mathbf{B}^{-1})^{-1} \Gamma^T \mathbf{R}^{-1}(T^o - T(\eta_b)), \tag{3.4}$$

where Γ is the Jacobian of the relation between temperature and soil moisture contents, $\mathbf{B}$ the background covariance matrix and $\mathbf{R}$ the observation error covariance matrix which is assumed to be stationary and diagonal. To compute the analysis η_a, the Jacobian Γ which measures the sensitivity of the screen level temperature to soil moisture has to be determined. For that purpose, three additional forecasts are conducted, one of which is a routine forecast and two are forecasts initialized with perturbed soil moisture fields. In the first perturbed forecast, the upper three soil levels are perturbed between pore volume and air dryness point, both of which depend on the soil type. In the second perturbed forecast, level 4 and 5 are provided with soil moisture perturbations.
If a cold start (first start of the system) is conducted, the background covariance matrix $\mathbf{B}$ is initialized with climatological values, afterwards it is updated by a cycled Kalman filter analysis. Further information on the soil moisture analysis can be found in Schraff and Hess (2012), Hess (2001).

3.4.2 Analysis of sea surface temperature

Sea surface temperature (SST) strongly impacts the fluxes of sensible and latent heat over the ocean. Its spatial distribution co-determines the evolution of Rossby waves and cyclones. To capture not only the large-scale variations, but also relevant smaller-

scale processes such as anomalies due to up- and downwelling of the thermohaline ocean circulation or fast temperature increases in shallow coastal areas during clear-sky days, the SST analysis (Schraff and Hess 2012) is performed once a day at 00 UTC. Initially, the extent of the sea ice cover is analysed by interpolating an ERA-Interim sea ice analysis to the model grid. Furthermore, a weekly high-resolution sea ice analysis at a resolution of $0.17° \times 0.1°$ for the Baltic Sea is used, which is provided by the Federal Maritime and Hydrographic Agency of Germany (BSH). A correction scheme is employed using the ERA-Interim SST field as first-guess which is corrected by means of ship and buoy observations from the five foregoing days. The observation increments are weighted according to the distance between analysis and observation time, observation type as well as the spatial distance between observation and grid point to be corrected.

3.4.3 Analysis of snow depth

The distribution of snow cover on the ground modifies surface albedo and turbulent fluxes and co-determines the screen-level temperature through modification of the surface albedo that impacts the absorption of short-wave radiance. Moreover, it affects the characteristics of large-scale air masses. The snow distribution and depth are updated every 6 hours (Schraff and Hess 2012). The observations used for data assimilation are SYNOP observations of total snow depth or 6-hourly precipitation sums should the observed surface temperature fall below 0 °C. To identify permanently ice-covered regions, a monthly snow depth climatology provided by ECMWF is additionally taken into account. Depending on the horizontal and vertical distances between the locations of the grid points to be corrected and observations, the observations within a radius of influence are weighted to form the snow depth increments. These are added to the field that was beforehand obtained by the forward integration of the model employing observation nudging in the atmosphere.

3.5 Process cycle

The individual components described above have been integrated into a production suite (using the workflow package ecflow) sketched in Figure 3.2. It is operated at the super-computing facilities of ECMWF and works as follows: The initial conditions at the beginning of the reanalysis period (and of the reanalysis streams, respectively) and 3-hourly boundary conditions are provided by ERA-Interim. Every 6 hours, the nudging runs are interrupted to perform a snow analysis. Once a day at 00 UTC the sea surface temperature and the soil moisture undergo an update. This is conducted

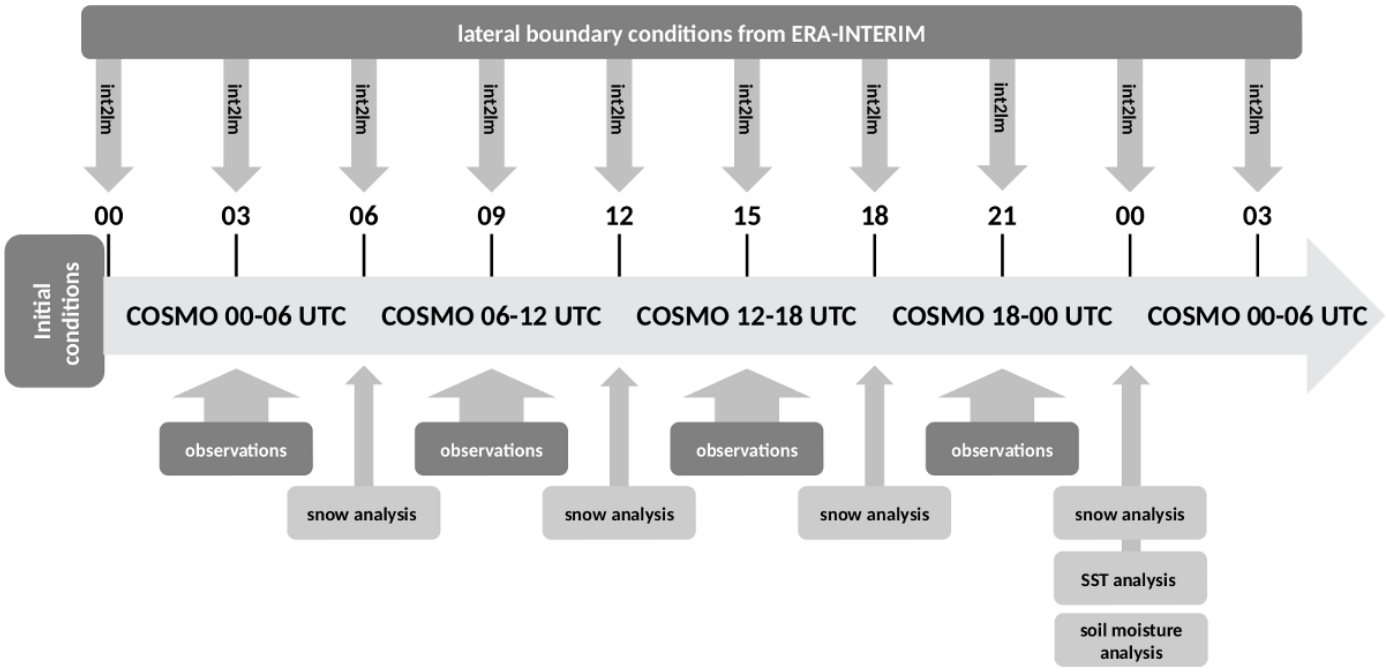

Figure 3.2: Process cycle of the probabilistic regional reanalysis system for one exemplary ensemble member. The cycle is repeated for all members in parallel. All members are provided with the same lateral boundary conditions. The tool *int2lm* interpolates the fields from the steering model to the COSMO grid.

for all ensemble members separately. Note that ensemble nudging is the only ensemble generation technique that is operationally applied for the production of the reanalysis COSMO-EN-REA12. External parameters including leaf area index, plant cover, root depth, carbon dioxide concentrations and an ozone maximum are updated once a day according to a prescribed annual cycle. The three-dimensional fields of the dynamically relevant quantities on model levels are stored in 6-hourly intervals while the surface fields and fields on pressure and height levels are archived at an hourly frequency. The system additionally incorporates reforecasts that are initialized in six-hourly intervals and have twice a day lead times of 30 hours and twice a day of 6 hours. Feedback observation files summarize information on observations and observation-analysis departures in six-hourly intervals. Rather technical components that are not depicted in the process cycle include the retrieval of ERA-Interim data from the Mars archive of ECMWF, the retrieval of data needed for the analysis of the lower boundary conditions, the extraction of observation feedback files of COSMO-REA6 from the ECFS archive, linking of files, post-processing of the output variables to the physical units agreed in UERRA, transformation from the data format wgrib1 to wgrib2, archiving in ECFS and archiving in Mars.

3.6 Basic and extended reanalysis suites

Note that two versions of the probabilistic regional reanalysis suite have been implemented. Here, the differences between the two are outlined. The first suite is operationally used for the production of the reanalysis COSMO-EN-REA12 comprising 21 ensemble members in UERRA and referred to as *basic reanalysis suite*. It makes use of ensemble nudging and thus measures the uncertainty of the reanalysis given observation error estimates. In the second version of the system, possibilities for accounting for uncertainties due to model error using SPPT and errors in the lateral boundary conditions have been implemented (ICON ensemble). In the basic system, ERA-Interim is employed as lateral boundary condition for all ensemble members. In the extended version, one ICON member can be chosen as control lateral boundary condition. Alternatively, 20 ICON members can be utilized as probabilistic boundary conditions. Note that for the basic reanalysis suite, COSMO version V5_0 has been chosen. Reasoning for this has been the wish for consistency with the high-resolution regional reanalyses COSMO-REA6 and COSMO-REA2. Only late it turned out, that there had already been an update from REA6 to REA2. For the development of the extended reanalysis suite, however, a further update of COSMO has been indispensable to allow for the use of SPPT. Moreover, to allow for processing the data format grib2 of ICON, int2lm (which interpolates the initial and boundary data and updates external data) has been updated from V2_0 to V2_02. However, the quality of prospective reanalyses will benefit from use of these more recent and further-developed versions.

4 Performance of the reanalysis suite

In this chapter, the capabilties of the newly developed probabilistic reanalysis suite are demonstrated. For this purpose, a range of numerical experiments has been conducted with the *basic* and *extended reanalysis suites* introduced in chapter 3. Initially, the experiments are described. Then, basic diagnostics including analysis increments, horizontal kinetic energy spectra, potential spin-up effects and the evolution of spread in reforecasts are shown. The *basic reanalysis system* using ensemble nudging that has been developed for the production of a test reanalysis COSMO-EN-REA12 in UERRA is evaluated subsequently.

Regional reanalyses are supposed to yield consistent estimates of essential climate variables. The most important essential climate variables whose representation in regional reanalyses potentially has an added value are low-level winds, screen-level temperature, precipitation, surface radiation budget components and cloud properties. Since in UERRA they have been identified as the variables for which an added value should be demonstrable compared to global reanalyses as a start (personal communication Dale Barker, Met Office; 2nd UERRA General Meeting, Tortosa, Spain), it has been chosen to approach the evaluation of the system by means of precipitation and screen-level temperature in the first instance. Therefore, to justify its use for production, an important question is to which extent the new system can yield data that are competitive with related climate data sets. This is assessed by means of a comparison study in which a comprehensive verification of precipitation against rain gauge observations is conducted. The particularly new aspect of this work is the generation of an ensemble of reanalyses. Therefore, the probabilistic capabilities of the new suite with respect to precipitation are demonstrated conducting a comparison to the ECMWF-EPS (Palmer et al. 1997).

In production mode, the new reanalysis suite deploys ensemble nudging as ensemble generation technique. However, uncertainty in the assimilated observations is not the only source of error. Therefore, the suite has been extended by the possibility to account for model error and for errors in the lateral boundary conditions (LBCs). To give an indication of how a reanalysis ensemble would be ideally generated in the future, four numerical experiments are comparatively verified by means of precipitation and screen-level temperature.

Table 4.1: Experiments conducted with basic and extended reanalysis suites

Name	Period	COSMO	Int2lm	Suite	Method	LBCs
Summer	01.06.-30.06.2011	V5_0	V2_0	basic	ensemble nudging	ERA-Interim
Winter	01.12.-31.12.2011	V5_0	V2_0	basic	ensemble nudging	ERA-Interim
EN	15.05.-15.06.2014	V5_04	V2_02	extended	ensemble nudging	ICON-ctrl
SPPT	15.05.-15.06.2014	V5_04	V2_02	extended	SPPT	ICON-ctrl
ICON	15.05.-15.06.2014	V5_04	V2_02	extended	ICON	ICON-ens
ESI	15.05.-15.06.2014	V5_04	V2_02	extended	EN+SPPT+ICON	ICON-ens

4.1 Numerical experiments

To develop the new probabilistic regional reanalysis system and tune ensemble nudging, but also to demonstrate the system's capabilities ahead of the production, two numerical experiments have been conducted with the *basic reanalysis suite* during the initial development phase of UERRA. Each of the experiments extends over one month and comprises of 20 perturbed ensemble members and one control run which assimilates the original unperturbed observations. The global reanalysis ERA-Interim is used as lateral boundary conditions. To test the performance in substantially different weather regimes, a winter (December) and a summer month (June) of 2011 have been chosen as experimental periods. The experiments are referred to as *summer* and *winter* experiments.
The *extended reanalysis suite* is able to account for model error using stochastic perturbation of physical tendencies (see section 3.3.2). Further, probabilistic LBCs can be employed. Here, a global ICON ensemble of DWD (see section 3.3.3) is used. It is conceivable to replace it with the new global probabilistic reanalysis ERA5 of the ECMWF once that it is disseminated. The experiments conducted extend from 15/05/2014 to 15/06/2015 and comprise 20 ensemble members. Three account for only one source of uncertainty. They are referred to as *EN* (ensemble nudging), *SPPT* (stochastic perturbation of physical tendencies) and *ICON* (ICON ensemble as LBCs). For consistency, one ICON member has been defined as control run and is employed as LBCs for *EN* and *SPPT*. Further, an experiment combining all three error sources denoted *ESI* (Ensemble Nudging + SPPT + ICON) has been conducted. The numerical experiments are summarized in Table 4.1.

4.2 Basic diagnostics

To gain first insight into the performance of the suite for COSMO-EN-REA12, basic diagnostics and properties of the regional ensemble reanalysis system are shown. This

includes an investigation of analysis increments, spin-up effects and an estimation of the effective resolution of the reanalysis system using spatial kinetic energy spectra. To provide insight into the uncertainties that are indicated by the ensemble, the evolution of average spread in reforecasts is assessed. The evaluation shown here is based on the experiment that spans June 2011 (denoted summer experiment, see Table 4.1). It comprises 20 ensemble members and one control run which assimilates the original unperturbed observations. Additionally, reforecasts with a forecast horizon of 30 hours, initialized at 00 UTC, are evaluated.

4.2.1 Analysis increments

Analysis increments are the adjustments made to the model state by data assimilation. They provide important diagnostics of the performance of a system. Generally, under the assumption of a constantly dense observation network, analysis increments can be used to reveal systematic errors in the system. Systematic behavior of analysis increments hints at biases in the NWP model or the observations or combinations. However, reanalyses conducted over long periods are exposed to a strong increase in the density of observing systems. Major changes here can lead to systematic changes in the behavior of the analysis increments. Therefore, it is important to monitor analysis increments during the production of a reanalysis, both to achieve the best possible climate quality if this is strived for and to detect technical problems in time (Dee et al. 2011b). Usually, monthly or annual averages are examined to filter out the effect of model biases depending on the weather regime.

Since nudging is applied continuously during the forward integration of the model, the analysis increments shown here represent aggregates of all changes that have been applied to the model state at a certain grid point over 6-hourly analysis cycles. Since there is no longer time span available, a horizontally averaged time-sequence of analysis increments of temperature and specific humidity is shown for an arbitrarily chosen member. The vertical structure in temperature displayed in Figure 4.1 exhibits a diurnal cycle and is essentially persistent. A warm bias of COSMO is visible in the middle part of the troposphere and a cold bias near the surface. During day, warm biases occur on the lower levels between 15th and 25th of June. The amplitude of the analysis increments of maximally ± 0.5 K is relatively low compared to the standard deviation that reaches maximum values of 1.5 K. The variability of the analysis increments is highest near the surface and decreases with height.

The averaged analysis increments for specific humidity also have an essentially persistent vertical structure. They indicate a moist bias in all atmospheric levels except for the

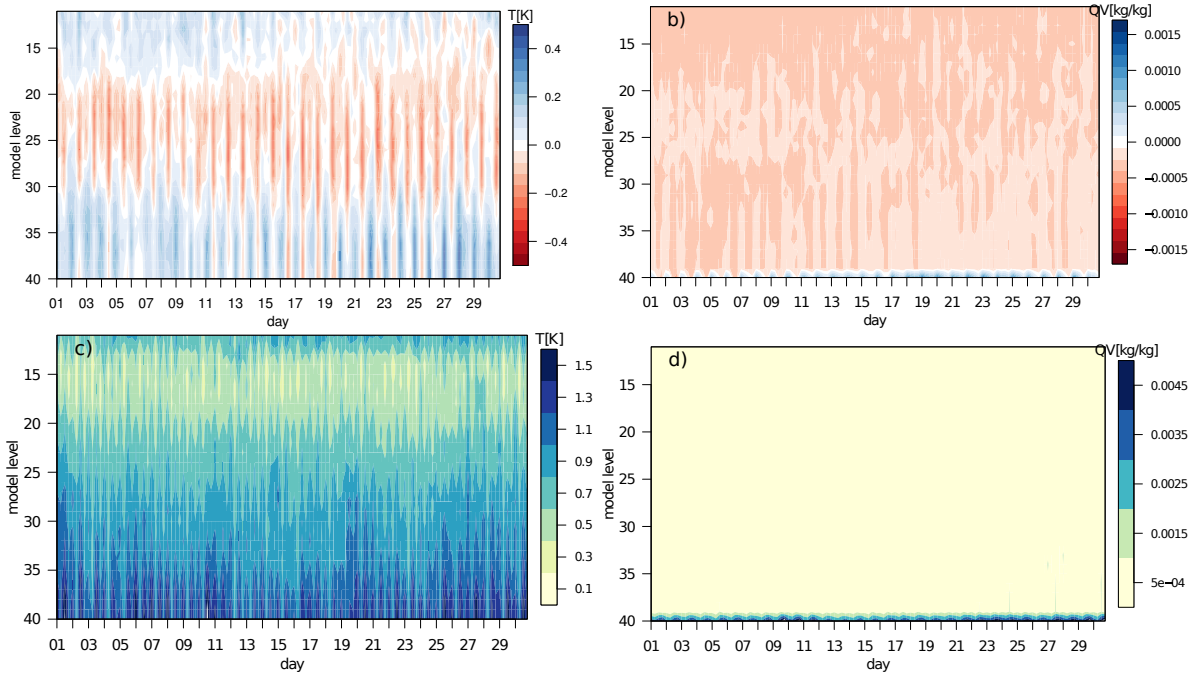

Figure 4.1: Analysis increments for June 2011. Spatially averaged and in dependency of model level and time of day for (a) temperature and (b) specific humidity. The analysis increments are aggregated over 6-hourly analysis cycles from 00, 06, 12 and 18 UTC. Model levels 20 and 30 are located at about 550 hPa and 100 hPa above the surface. For temperature red color indicates a too warm model, blue color a too cold model. For specific humidity red color means that the model is too moist while blue means that it is too dry. Further, spatial standard deviations of analysis increments are shown for (c) temperature and (d) specific humidity in dependency of model level and time of day.

lowest one, which has a dry bias following a diurnal cycle. Near the surface, the bias is maximally one third of the standard deviation whereas in the levels above the mean analysis increments have approximately the same amplitude as the standard deviation. Note that the small analysis increments in the upper levels possibly originate from a relatively low data density, since only humidity from radiosondes is assimilated.

4.2.2 Spin-up

Spin-up effects of precipitation occur during the first few hours after the initialization of forecasts (Betts and Ball 1999, Arpe 1991). They become apparent as an under- or over-estimation of precipitation which occurs as a consequence of pronounced dry or moist model biases, but also biases in temperature or due to the occurrence of gravity waves leading to non-zero analysis increments. Since NWP models tend towards their

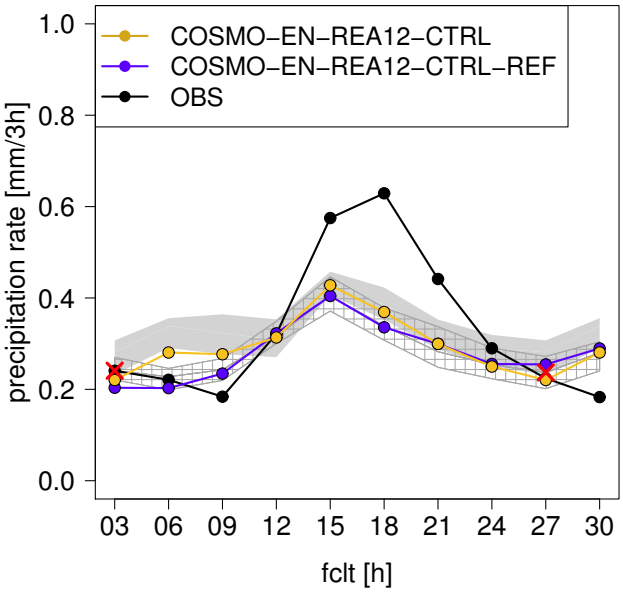

Figure 4.2: Diurnal cycle of 3-hourly precipitation rates in ensemble of reforecasts of COSMO-EN-REA12 for June 2011, initialized at 00 UTC (shown as hatched area, 5 to 95 % percentiles, control run blue points), reanalysis ensemble (shown as shaded area, 5 to 95 % percentiles, control run yellow points) and corresponding rain gauge observations (black dots). Shown for German subdomain, see Figure 4.6. The red crosses mark the median of the reforecast ensemble at 03 and 27 hours lead time.

model climate and physical balance between the variables, moistening the model by data assimilation often leads to an over-estimation of precipitation in the first few hours of lead time or reversely to an under-estimation if the model has a moist humidity bias. Since this impacts the representation of the hydrological cycle in reanalyses, it is important to be investigated and communicated to users.

Figure 4.2 shows the diurnal cycle of 3-hourly precipitation rates in reforecasts as a function of forecast lead time. The employed reforecasts are initialized at 00 UTC, so that the lead time coincides with the time of day. Additionally, the diurnal cycle of precipitation rates from analysis data is shown. The graphic illustrates that the reforecast ensemble does not exhibit pronounced spin-up effects. Firstly, the amplitude of the precipitation rate measured by the observations falls within the range of the ensemble in the first 6 hours. Secondly, the median of COSMO-EN-REA12 after three hours lead time agrees with the median of the ensemble after 27 hours which is the same time of day. Both findings indicate that the ensemble simulates the right amount of precipitation in the first hours of lead time. This is different for the control run which slightly underestimates precipitation in the beginning. The analysis ensemble slightly overestimates precipitation between 03 and 09 UTC so that in reforecasts initialized at

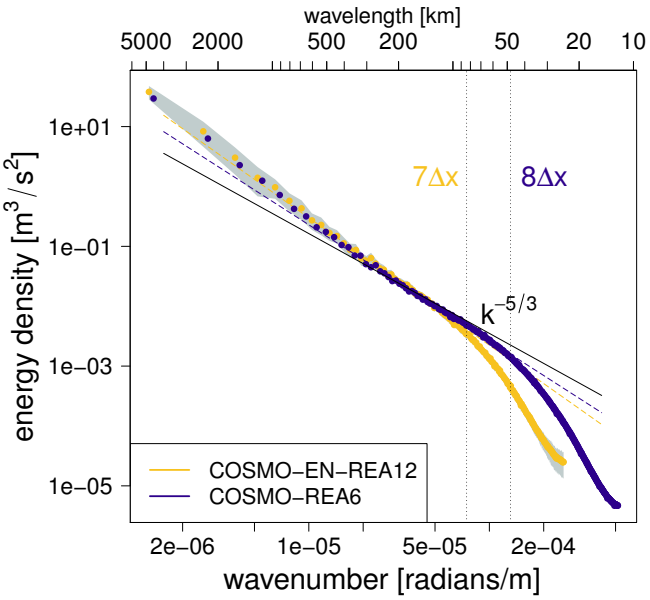

Figure 4.3: Horizontal kinetic energy spectra for COSMO-EN-REA12 (yellow points) and COSMO-REA6 (blue points) in dependence of wavelength and wave number for June 2011. The grey area shows the uncertainty estimated from the ensemble. The vertical lines mark the effective resolution of the reanalysis systems. The continuous line shows $k^{-5/3}$ line. The dashed yellow and blue lines show slopes of spectra at the mesoscale, both of which are slightly steeper than $k^{-5/3}$.

03, 06 or 09 UTC small spin-down effects may occur. Since the overestimation is much more pronounced in the perturbed ensemble members than in the control run it must be an effect of the observation perturbations. These possibly cause a stochastic drift shifting the ensemble mean. From 12 hours of lead time onwards, no significant difference can be identified between reanalysis and reforecast ensemble.
Figure 4.2 further reveals that COSMO-EN-REA12 underestimates the precipitation rate during the afternoon and peaks too early. The peak placed too early is a well-known problem in models with parameterized deep convection that diagnose the convective precipitation from the grid scale environment (for COSMO see Baldauf et al. (2011)).

4.2.3 Spatial kinetic energy spectra

Horizontal kinetic energy spectra have been found to have a $k^{-5/3}$ slope on the mesoscale, see for example Skamarock (2004). Theoretically, this slope extends down to the beginning of the microscale. However, in real world models a lower limit is imposed by the grid spacing. The wave length at which the spectrum leaves the $k^{-5/3}$ slope marks the point at which smaller spatial processes are no longer fully represented, whereby the

effective resolution of the model can be measured. Both the question if an NWP model reproduces the $k^{-5/3}$ slope and the effective resolution are interesting to investigate. Following Bierdel et al. (2012), the effective resolution is estimated as the wavelength at which the spectrum begins to fall steeply under the $k^{-5/3}$ slope. In Figure 4.3, the horizontal kinetic energy spectrum of the new system averaged over the whole experimental time span is compared to the one of COSMO-REA6 at 6 km grid spacing (Bollmeyer et al. 2015) which has provided the basis for the development of the system. The energy spectra for the horizontal velocity components at a height of approximately 5 km are computed separately and averaged afterwards. For a detailed description of the computation be referred to Bierdel et al. (2012).
At the mesoscale, linear trends are fit to both spectra in the log-log representation. These show a slightly steeper slope than $k^{-5/3}$, where the spectrum of COSMO-REA6 has a slightly smaller deviation. Both spectra agree well with each other at the synoptic scales and the meso-α-scale and do not have a pronounced transition to a slope of k^{-3} that has been found for the larger scales. The effective resolution is about $7\Delta x$ for COSMO-EN-REA12 and $8\Delta x$ for COSMO-REA6. This corresponds to approximately 85 km for COSMO-EN-REA12 and 50 km for COSMO-REA6 and agrees with values described in literature (Skamarock 2004).
Further, the horizontal kinetic energy spectra of analysis data from COSMO-EN-REA12 have been compared to the spectra of reforecasts initialized at 00 UTC and valid at 3, 10, 24 and 30 hours lead time. All agree with each other (therefore not shown), so that the data assimilation does not have a distorting impact on the spectrum. Thus, the absence of a spin-up effect found in the foregoing section is complemented by the absence of a spin-up effect in the dynamics. This may be a positive effect of nudging which applies the corrections to the model state very slowly and over a long time so that the model dynamics maintain a high degree of balance.

4.2.4 Evolution of spread

The quality of the uncertainty that is quantified by an ensemble can only be assessed in comparison with observations. Such a verification is conducted in the section following this one. Here, qualitative insight into the behavior of ensemble spread over a subdomain of Germany is provided. Figure 4.4 shows the mean temporal evolution of horizontally averaged spread (measured in terms of standard deviation) in reforecasts, initialized at 00 UTC for temperature, zonal wind, relative humidity and geopotential. In the vertical, the spread is largest at the lower levels for relative humidity and between 400 and 600 hPa for the other variables. Above 400 hPa, it decreases rapidly with increasing height which

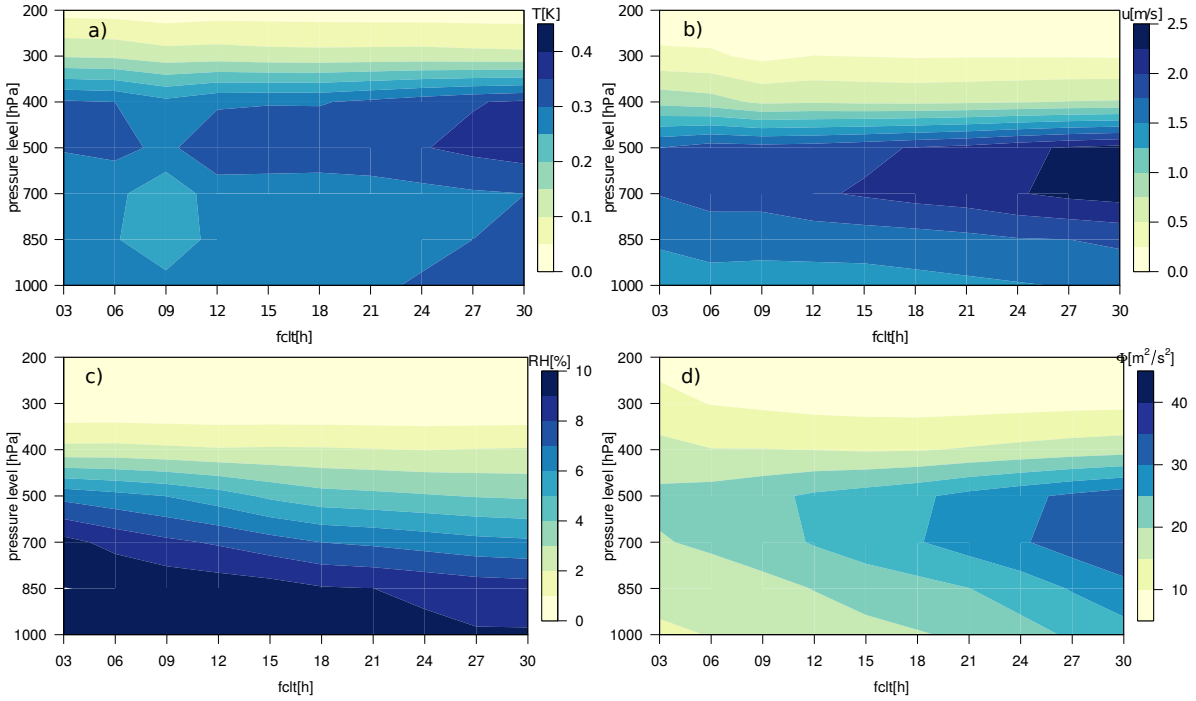

Figure 4.4: Monthly mean evolution of horizontally averaged spread (over German subdomain) as a function of pressure levels and forecast lead time. Data is from reforecasts initialized at 00 UTC for experiment for June 2011. (a) Temperature, (b) zonal wind, (c) relative humidity and (d) geopotential.

is likely related to the Rayleigh relaxation above 235 hPa towards the steering model (IFS/ERA-Interim) being identical for all ensemble members. As a function of lead time, the spread decreases for relative humidity and increases for zonal velocity, geopotential and slightly for temperature. The increase with lead time in three of the shown variables indicates that the pure perturbation of the initial conditions by perturbation of the assimilated observations is sufficient to allow for a reasonable development of spread.

4.3 Evaluation of the basic reanalysis suite

The goal of this section is to demonstrate the capabilities of the *basic reanalysis suite* in terms of a verification against independent observations. As discussed in the introduction to this chapter, in UERRA precipitation has been identified as the variable (next to screen-level temperature) for which an added value should be demonstrable compared to global reanalyses.

To show why this can be expected let us compare the monthly precipitation climatologies of June 2011, derived from ERA-Interim and from the ensemble mean of COSMO-EN-REA12 (see Figure 4.5). As anticipated, the regional reanalysis is capable of representing

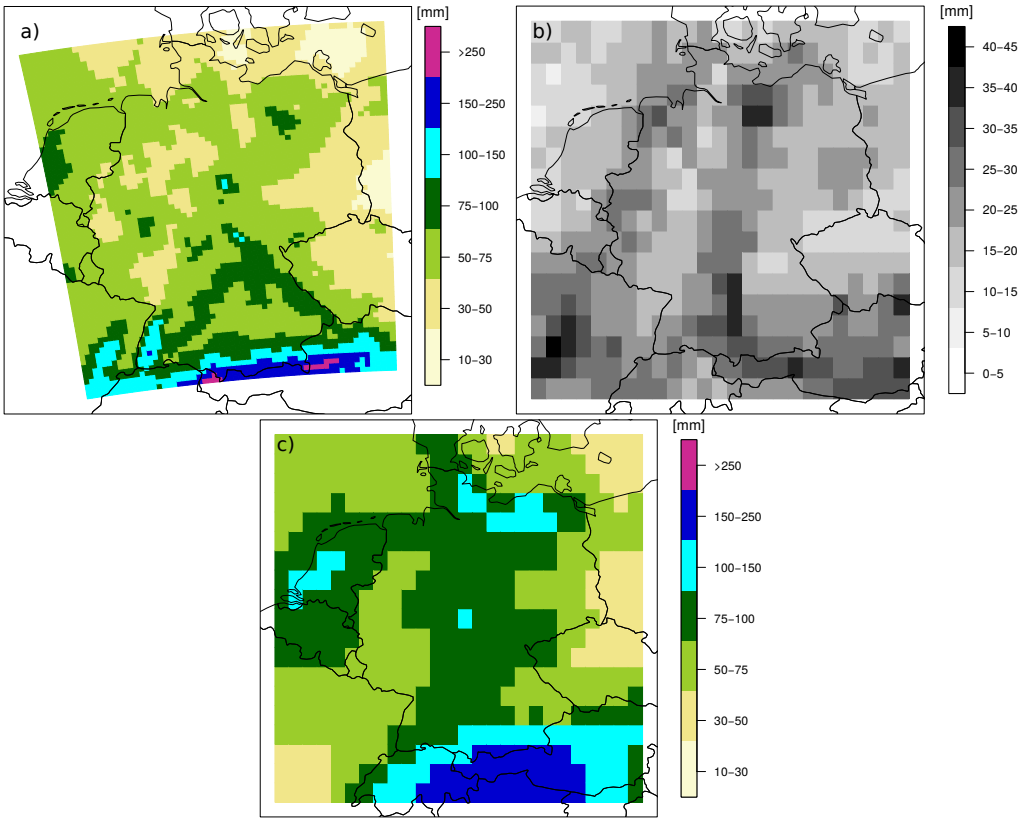

Figure 4.5: Monthly precipitation climatologies for Germany for June 2011. Based on ensemble mean of (a) COSMO-EN-REA12 and (c) ERA-Interim. (b) Uncertainty (spread) of the monthly integrated precipitation estimated by COSMO-EN-REA12.

mesoscale variability, while ERA-Interim rather shows large-scale patterns. Therefore, the precipitation fields of COSMO-EN-REA12 can be expected to agree much better with observations which would offer the possibility to monitor precipitation on a much more local scale. This is investigated in the next sections. The probabilistic reanalysis system further allows for the estimation of the uncertainties in such precipitation climatologies. The spread of the monthly integrated precipitation resulting from the summer experiment (see Table 4.1) is shown in Figure 4.5. The estimated uncertainty has a pronounced spatial variability and a large maximum value of 40 to 45 mm. Regions of large spread coincide with regions of large integrated precipitation amounts. Here, these regions are located in Eastern Germany as well as over the Alps. The large spread of up to 45 mm substantiates the value of uncertainty estimation in reanalysis, e.g. for application in monthly climatologies as provided by near-real time climate monitoring (for an example see the *Deutscher Klimaatlas*, described in Kaspar et al. (2013) or 2.5).

Figure 4.6: Rain gauges used for verification. The grey stations are used for the probabilistic verification, the white ones are additionally used for the deterministic verification.

The experiments from which precipitation is evaluated in the following comprise 20 ensemble members and one nature or control run that assimilates the original observations and extend over the months of June 2011 and additionally December 2011. They are referred to as *summer* and *winter experiments* (see Table 4.1) of COSMO-EN-REA12. Precipitation is not assimilated so that the analysis is independent of the verifying observation. Due to a slow and continuous correction by nudging it can be assumed that the fields are in dynamical balance. This is supported by the absence of spin-up effects shown in section 4.2.2. Therefore, the reanalysed precipitation fields can directly be used so that an evaluation of them is shown instead of precipitation from reforecasts. Note that the verifying rain gauge observations are regarded as free of errors which can be seen as contradictory to the assumption of error prone observations for ensemble generation. However, due to a lack of knowledge about the distribution of observation errors of rain gauges it is assumed that they are negligible compared to the errors of the model-based analysis.

The subsequent evaluation of precipitation focuses on two research questions:

1. Does the ensemble nudging reanalysis system at 12 km resolution have an added value?

2. What are the probabilistic capabilities of the ensemble?

The goal that is pursued in the first part of the evaluation is to compare the basic system to the global reanalysis ERA-Interim and downscaling data (that are used for many of the same applications) and to demonstrate how its probabilistic capabilities score for a short time span. For this purpose, precipitation from all involved data sets is verified on the original grids. This is considered useful since users will most probably use the highly-resolved data rather than interpolate it to a coarse grid.
In the second part, the probabilistic capabilities of COSMO-EN-REA12 are assessed. Due to the lack of a global reanalysis ensemble at the time of writing, the regional reanalysis data are compared to forecasts from the ECMWF-EPS. Again, both systems are compared on the original grids (ECMWF-EPS interpolated to a regular grid of approximately 25 km grid spacing).

4.3.1 Deterministic performance

Initially, each ensemble member is verified separately. It is assessed if the ensemble members from the ensemble nudging experiments bear comparison with their corresponding control run, the global reanalysis ERA-Interim, the high-resolution regional reanalysis COSMO-REA6 (Bollmeyer et al. 2015) and a dynamical downscaling from ERA-Interim using COSMO at 6 km grid spacing, COSMO-DOWN6 (denoted "COSMO-DS" in Bollmeyer et al. (2015)). In this context, dynamical downscaling means that a free model run is started from initial conditions provided by COSMO-REA6. This run is provided with three-hourly lateral boundary conditions from ERA-Interim. Since no observations are assimilated the model state will approach its model climate in the interior of the domain after a certain time. The downscaling is useful to show the difference to the reanalysis data sets in which data assimilation adds realistic mesoscale information. ERA-Interim (ECMWF) is based on the spectral model IFS with a grid spacing of about 0.7° and a 12-hourly 4D-Var cycle (Dee et al. 2011b). Analysis data from 00 and 12 UTC as well as +03-h, +06-h and +09h reforecasts are employed. The nudging regional reanalysis COSMO-REA6 is based on the COSMO model at 6 km grid spacing and 40 vertical levels. It makes use of ERA-Interim as lateral boundary conditions and is involved in the evaluation to see the difference in performance at a bisection of the grid spacing which is required to effort the computation of a 20 member ensemble.

Due to difficulties related to the international exchange of precipitation observations with high temporal resolution, the verification is confined to a German subdomain which has a very dense network comprising 1034 rain gauge stations (see Figure 4.6). To each of the stations the nearest neighbor from the model grid points is assigned. It happens that one grid point of the coarser resolved reanalysis ERA-Interim is assigned to multiple rain gauge stations. The verification is based on 3-hourly precipitation sums.
Figure 4.7 shows the frequency bias, the log odds ratio as well as the equitable threat score for the summer experiment in the left column and the winter experiment in the right column. All scores are computed based on a contingency table for binary events. As thresholds, 0.1 mm/3h, 1 mm/3h, 2.5 mm/3h and 5 mm/3h have been chosen which exhibit high-enough base rates (shown in Figure 4.8) to be considered fair for a comparison including ERA-Interim and the short experimental periods that extend only over 30 days. To estimate the sampling uncertainty of the scores 1000 bootstrap samples are performed based on days as entities.

4.3.1.1 Frequency bias

The frequency bias (Donaldson et al. 1975) shown in Figure 4.7 a) and b) together with the base rates shown in comparison with the observations in Figure 4.8 is useful to assess systematic errors. It tests the agreement of the marginal distributions of precipitation events in the reanalysis compared to the observed events

$$FB = \frac{a+b}{a+c} \in [0, \infty] \tag{4.1}$$

with a hits, b false alarms, c misses (and d correct negatives, see scores hereafter). ERA-Interim heavily overestimates the number of precipitation events at 0.1 mm/3h and 1 mm/3h and underestimates it at 2.5 mm/3h and 5 mm/3h. This is also revealed by the base rate of ERA-Interim which is over-rated at 0.1 mm/3h and too small at 5 mm/3h. However, at 1 mm/3h in winter, the bias is almost perfect. In summer, the control run of COSMO-EN-REA12 and COSMO-REA6 have a virtually perfect frequency bias at the threshold 1 mm/3h. The boxplot shows that the ensemble members scatter evenly around the control. Going to higher thresholds the bias for both COSMO-REA6 and COSMO-EN-REA12 reveals an increasing underestimation of precipitation, whereby the observation perturbations have a positive impact improving the bias of the ensemble members compared to the one of the control run. In winter, COSMO-EN-REA6 and COSMO-EN-REA12 have a frequency bias close to 1, whereby the latter is the best system regarded over all shown decision thresholds. The observation perturbations have

a positive impact at the higher thresholds.

4.3.1.2 Log odds ratio

The log odds ratio (Stephenson 2000) displayed in Figure 4.7 c) and d) gives insight into the spatio-temporal coherence of reanalyses and observations. It measures the ratio of the odds of making a hit (hits compared to misses) to the odds of making a false alarm (false alarms compared to correct negatives)

$$LOR = \log\left(\frac{ad}{bc}\right) \in [-\infty, +\infty]. \tag{4.2}$$

It is positively orientated and gives better scores to rarer events due to a high weight of the correct negatives d. The score is best for COSMO-REA6 and clearly worst for the downscaling. The superiority of COSMO-REA6 compared to the control run of COSMO-EN-REA12 shows the loss of accuracy due to the bisection of the grid size. The inferiority of the downscaling confirms the added value in accuracy of the reanalysis data sets due to data assimilation. In summer, there is no significant difference between ERA-Interim and ensemble nudging. In winter, ERA-Interim has a better log odds ratio than ensemble nudging.

4.3.1.3 Equitable threat score

The equitable threat score shown Figure 4.7 e) and f) measures the fraction of hits of the sum of all precipitation events in reanalysis and observations. Supplementary, it accounts for hits due to random chance. It is defined by

$$ETS = \frac{a - a_{random}}{a + b + c} \tag{4.3}$$

with

$$a_{random} = \frac{(a + c)(a + b)}{n}, \tag{4.4}$$

where n is the total number of events. Again, COSMO-REA6 performs best for 1 mm/3h while the downscaling is the worst system. COSMO-EN-REA12 and ERA-Interim exhibit no significant difference except for 1 mm/3h where ERA-Interim outperforms the new system in the winter experiment. In the summer experiment, nudging has a more positive influence on accuracy going from 6 (downscaling) to 12 km grid spacing (nudging) than in winter.

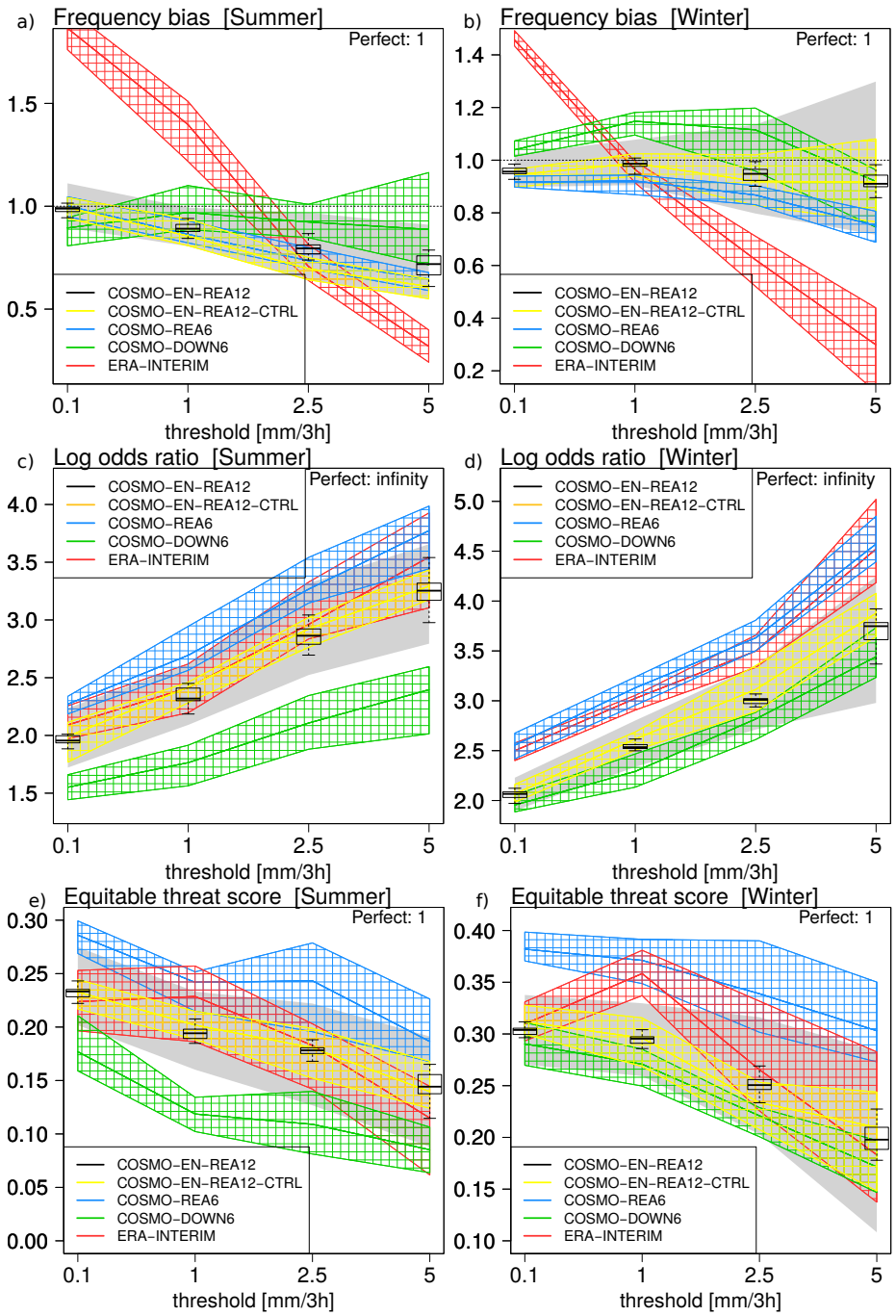

Figure 4.7: Verification measures (a) and (b) frequency bias, (c) and (d) log odds ratio, (e) and (f) equitable threat score based on the (2 × 2) contingency table for binary events as a function of threshold for 3-hourly accumulated precipitation. The left panels show the summer experiment (June 2011), the right ones the winter experiment (December 2011). The control run of the ensemble nudging experiments is depicted in yellow and the nudging ensemble as boxplots. ERA-Interim is presented in red, COSMO-REA6 in blue and COSMO-DOWN6 in green. The solid lines represent the median of 1000 bootstrap samples while the hatched areas are 95 % confidence intervals. For each of the ensemble members, 1000 bootstrap samples have been drawn. The grey shaded area presents the 95 % confidence interval of the resulting 20000 bootstrap samples.

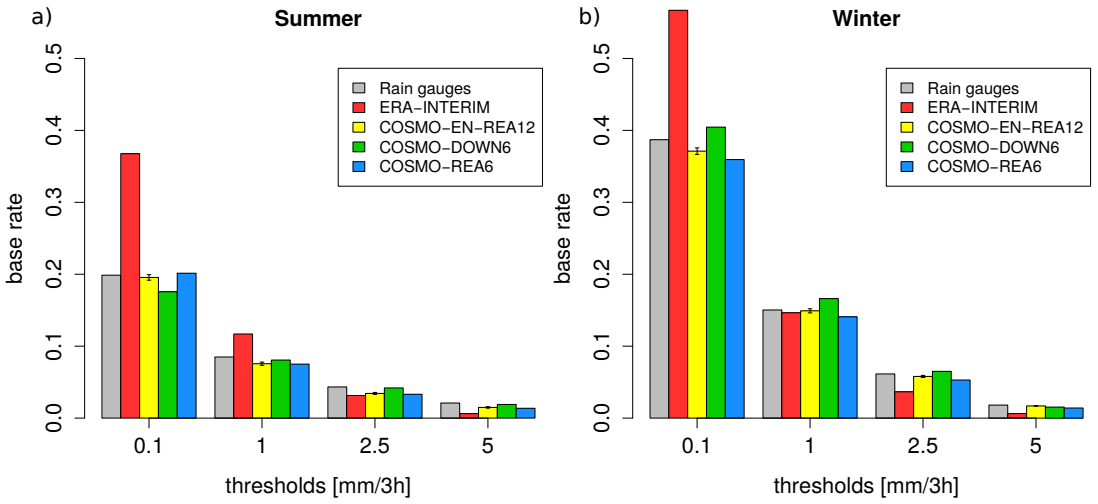

Figure 4.8: Base rates of threshold exceedance for 3-hourly accumulated precipitation for the (a) summer and (b) winter experiments. Sample size approximately 255000.

4.3.1.4 Discussion

The first research question posed above *does the ensemble nudging reanalysis system at 12 km resolution have an added value* compared to global reanalyses and dynamical downscalings has guided the evaluation in this section. As preliminary answer to this question it is found that the accuracy of ensemble nudging is comparable to the one of ERA-Interim. Thus, there is no added value in accuracy at the small precipitation thresholds chosen to have a fair amount of events in the relatively short experimental periods. A comparison to COSMO-REA6 yields that the system looses the significant added value going from 6 km to 12 km grid spacing for the sake of a 20 plus 1 member ensemble. Compared to the dynamical downscaling COSMO-DOWN6, there is still an added value in accuracy (except for the ETS in winter) even though the system has only 12 instead of 6 km grid size.

The findings regarding accuracy of precipitation on small thresholds are different from the ones of Jermey and Renshaw (2016), who also find an added value for their regional reanalysis system at these ranges. This may be enabled through a significantly longer time span of 2 years and may also arise from assimilation of remote sensing and satellite data. One of the major applications of regional reanalysis data will certainly be monitoring of extreme events. However, to show an added value in accuracy at very high precipitation thresholds, a much longer data set must be available to obtain reliable statistics. Potentially, also the grid spacing and ensemble size have to be significantly increased for that purpose. Still, higher precipitation amounts are assessed in the probabilistic verification where the ECMWF-EPS at a resolution higher than ERA-Interim is chosen

for comparison. Different from the measures of accuracy, the frequency bias reveals a clear added value of the new regional reanalysis system compared to ERA-Interim. The frequency bias of all regional systems is nearly perfect and best for COSMO-EN-REA12 in winter.

4.3.2 Probabilistic performance

The probabilistic capabilities of an ensemble include consistency, accuracy, reliability, resolution and sharpness as well as absence of conditional and unconditional biases (Anderson 1997, Murphy 1973b, Wilks 1995). As there is no global ensemble reanalysis data available at the moment these properties are compared to 6-hourly accumulated precipitation sums from +06-hour ECMWF-EPS forecasts based on IFS (Palmer et al. 1997). The used reanalysis data from COSMO-EN-REA12 and the ECMWF-EPS forecasts are valid at 06 and 18 UTC. Note that there is no ECMWF-EPS data with higher temporal resolution available. However, for the evaluation of the deterministic performance of our system shown in the foregoing section all data is available at 3-hourly resolution which limits allowable spatio-temporal displacements of precipitating systems compared to 6-hourly accumulations. An improved positioning of precipitation would be beneficial for regional reanalyses. For these reasons, 3-hourly precipitation sums have been used in section 4.3.1 and 6-hourly ones are employed here.
In a first step, for most of the scores computed in the following sections probabilities have to be estimated from the ensemble. The event probability $p(y_i)$ is estimated using a beta-binominal model with a flat beta prior (Agresti and Hitchcock 2005) which leads to

$$p(y_i) = \frac{i+1}{N_{ens}+2}, \tag{4.5}$$

where i is the number of the ensemble member (the members are sorted in ascending order) and N_{ens} the total number of ensemble members. This relation has the advantage that the estimated probability can neither become 0 nor 1 which would be equal to the statement of a deterministic forecast or analysis. As a start, the consistency of COSMO-EN-REA12 is assessed making use of analysis rank histograms.

4.3.2.1 Consistency

Consistency means that the ensemble members and the observations are drawn from the same PDF so that all members are equally likely to represent truth (Anderson 1996, Hamill and Colucci 1997). This is the case if the analysis rank histogram corresponding

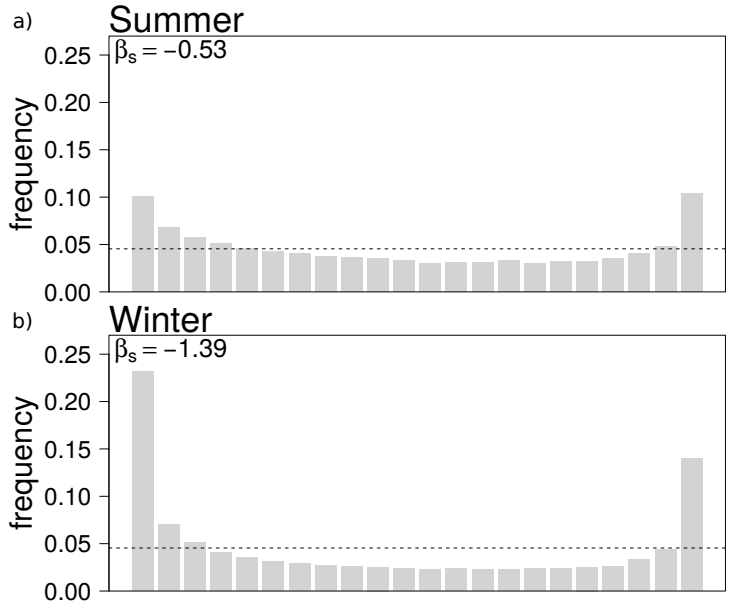

Figure 4.9: Analysis rank histograms for 6-hourly accumulated precipitation for (a) the summer experiment and (b) the winter experiment of COSMO-EN-REA12.

to the ensemble is approximately flat. The analysis rank histograms shown in Figure 4.9 are computed based on 6-hourly precipitation sums of the ensemble nudging summer and winter experiments. The dashed lines indicate the frequency at which the histograms would be perfectly flat. To avoid a distortion through a random distribution of no-precipitation observations between a no-precipitation ensemble, the events at which both the ensemble and the observation indicate "no precipitation" are omitted.

For the summer experiment shown in a), slight overestimation of the lowest rank and the highest rank can be observed which indicates a too narrow ensemble PDF and thus under-dispersiveness. The overweighting of the lowest rank arises from events at which the whole ensemble overestimates precipitation. Observations ranked to the highest bin result from the opposite. The under-dispersiveness is reflected by the negative β-score of -0.53 (Keller and Hense 2011), which is computed based on a fit of a β-distribution to the bin values of the analysis rank histogram.

The analysis rank histogram for the winter experiment displayed in Figure 4.9 b) reveals a stronger underestimation of uncertainties through a more pronounced u-shaped form, quantitatively measured by a more negative beta score of -1.39. Obviously, it occurs quite frequently that the whole ensemble over-estimates precipitation. This can e.g. arise from the ensemble being too confident about the position of a frontal system so that all members misplace precipitation.

4.3.2.2 Skill measured by Brier score and CRPS

The Brier score (Brier 1950) is a scalar summary measure of accuracy for dichotomous quantities. The score is equal to the mean squared deviation between n pairs of ensemble probability p_k for the exceedance of a threshold c given by

$$p_k = \frac{\#\{y_i,\, 1 \leq i \leq N | y_i \geq c\}}{N_{ens}} \tag{4.6}$$

and binary observations o_k with

$$o_k = \begin{cases} 1 & \text{if } o \geq c \\ 0 & \text{else.} \end{cases} \tag{4.7}$$

It is given by

$$BS = \frac{1}{n} \sum_{k=1}^{n} (p_k - o_k)^2 \tag{4.8}$$

with $0 \leq BS \leq 1$ and $BS = 0$ for a perfect ensemble system. The Brier skill score given by

$$BSS = 1 - \frac{BS}{BS_{ref}} \tag{4.9}$$

is useful to compare the accuracy of COSMO-EN-REA12 and the ECMWF-EPS as reference system. In that constellation, COSMO-EN-REA12 has skill or an added value if $BSS > 0$. Figure 4.10 shows the Brier skill score for precipitation from both COSMO-EN-REA12 experiments accumulated over 6 hours with +06-hour ECMWF-EPS forecasts valid at 06 and 18 UTC as a reference. For the summer experiment, COSMO-EN-REA12 has unconditionally more accuracy than the ECMWF-EPS. At 0.1 mm/6h, the percentage improvement of COSMO-EN-REA12 is about 38 % for the median of 1000 bootstrap samples. For the other thresholds it is about 18 %. In winter ensemble nudging performs poorer than the ECMWF-EPS except for the 0.1 mm/6h threshold and is approximately comparable at 0.5 mm/6h. For both seasons, the Brier skill score declines with increasing threshold. In winter, its relative lack of accuracy increases with increasing threshold. The Brier score indicates that the regional ensemble is capable of a better probabilistic representation of summer precipitation for the chosen thresholds. To take into account the whole range of precipitation amounts, the continuous ranked

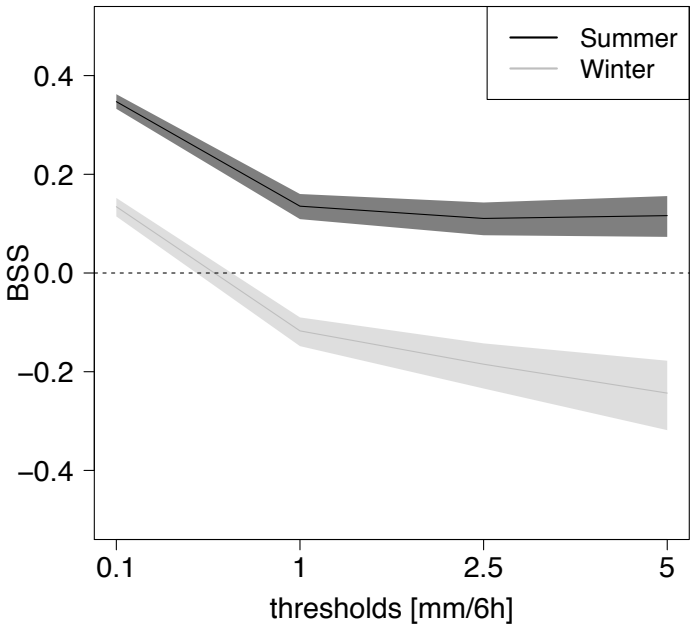

Figure 4.10: Brier skill score of COSMO-EN-REA12 versus ECMWF-EPS for 6-hourly accumulated precipitation. The summer experiment is illustrated in dark grey and the winter experiment in light grey. The solid lines represent the medians and the shading the sampling uncertainty as given by the 95% quantile of 1000 bootstrap samples.

probability score CRPS (Hersbach 2000) is employed which is given by

$$CRPS = \int_{-\infty}^{\infty} [P_f(x) - P_o(x)]^2 dx \tag{4.10}$$

where

$$P_o(x) = \begin{cases} 0, & x < o \\ 1, & x \geq o \end{cases} \tag{4.11}$$

is the Heaviside function that goes from 0 to 1. The predicted variable x equals the observation o and P_f is the cumulative distribution function of the forecast or analysis probability (Wilks 1995). Analogously to the BSS, a $CRPSS$ skill score can be defined. 1000 bootstrap samples are drawn and it is obtained $CRPSS \in [-0.01, 0.00, 0.012]$ for the winter and $CRPSS \in [-0.02, 0.00, 0.016]$ for the summer experiment, where the triple represents the 5 %-percentile, the median and the 95 %-percentile. These results show that under consideration of the whole range of precipitation amounts COSMO-EN-REA12 and the ECMWF-EPS have comparable probabilistic accuracy.

The CRPSS is most suitable to obtain an indication if the regional ensemble has an added value in convective regimes. To quantify that, a subset from the 5th to the 8th of June 2011 is chosen from the summer experiment. During these days, strong small-scale convective activity has been observed over large parts of Germany. Drawing 1000 bootstrap samples, it is obtained $CRPSS \in [0.0032, 0.43, 0.53]$. Even though the large spread between the quantiles shows a pronounced uncertainty due to a strongly reduced sampling size, this result indicates a clear potential for an added value of COSMO-EN-REA12 over the ECMWF-EPS in convective regimes.

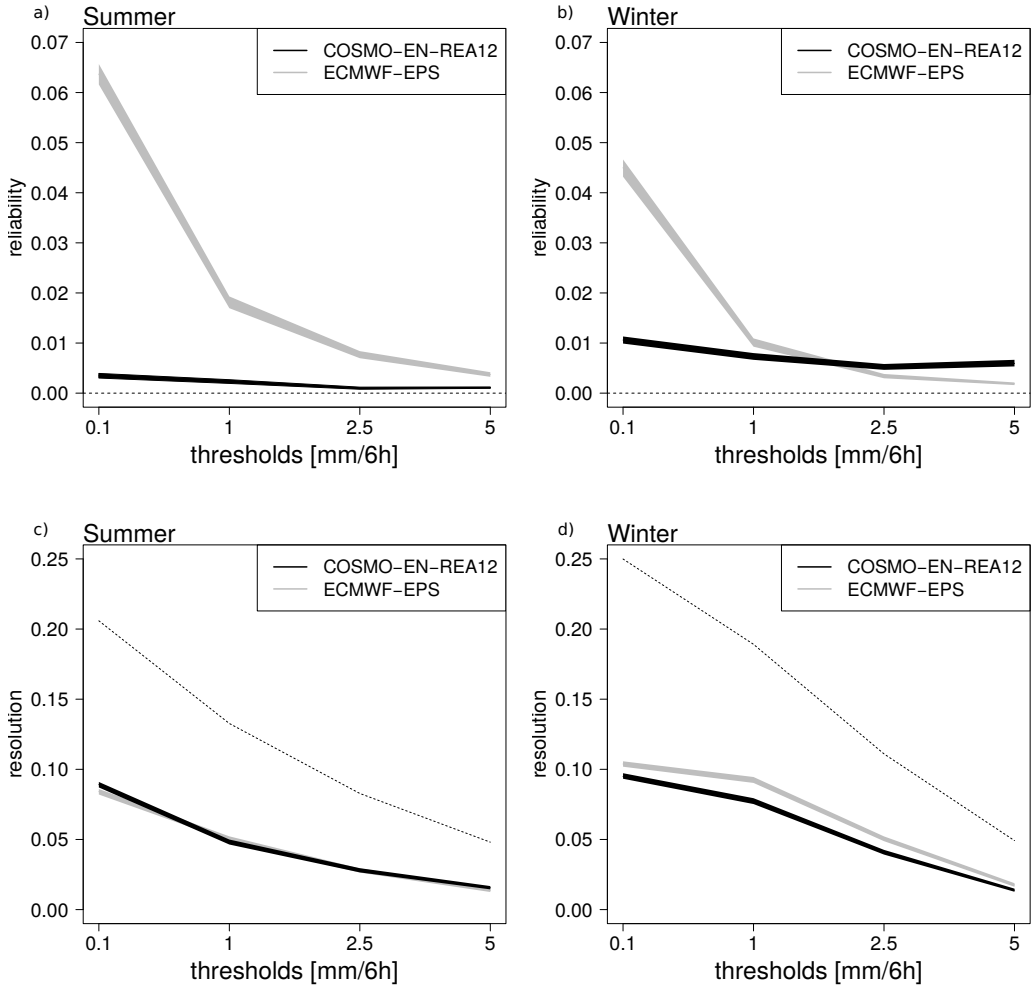

Figure 4.11: Decomposition of the Brier score: 5 to 95% quantiles of 1000 bootstrap samples of reliability [a) and b)] and resolution [c) and d)] for the summer (left) and winter (right) experiments. The thin dotted lines indicate the maximally achievable resolution if the reliability was perfect (uncertainty component).

A decomposition of the Brier score into a reliability, a resolution and an uncertainty component (Murphy 1973a) is depicted in Figure 4.11. It is

$$BS = \underbrace{\frac{1}{n}\sum_{i=1}^{n} N_i(y_i - \overline{o}_i)^2}_{Reliability} - \underbrace{\frac{1}{n}\sum_{i=1}^{n} N_i(\overline{o}_i - \overline{o})^2}_{Resolution} + \underbrace{\overline{o}(1-\overline{o})}_{Uncertainty}, \tag{4.12}$$

where n is the total sample size subdivided in I subsamples of size N_i with $n = \sum_{i=1}^{I} N_i$. $\overline{o}_i$ is the relative frequency of precipitation observations conditioned to the ensemble probabilities y_i in the subsamples I. It is given by

$$\overline{o}_i = p(o_i = 1|y_i) = \frac{1}{N_i}\sum_{k \in N_i} o_k, \tag{4.13}$$

where $\overline{o}$ is the sample climatology of the observations

$$\overline{o} = \frac{1}{n}\sum_{k=1}^{n} o_k = \frac{1}{n}\sum_{i=1}^{I} N_i \overline{o}_i, \tag{4.14}$$

which is a weighted average of the conditional relative frequency of observations assigned to I subsamples. The decomposition including estimates of sampling uncertainty based on 1000 bootstrap samples shows that the superiority of COSMO-EN-REA12 in summer can be traced back to a significantly better reliability component while the resolution of both systems differs hardly within the uncertainty intervals. Thus, the agreement between the observed frequencies conditioned to the ensemble probabilities and the ensemble probabilities themselves is much better for COSMO-EN-REA12 (reliability), whereas both systems are similarly able to issue probabilities that deviate significantly from the climatological base rate (resolution). In winter, the superiority of the ECMWF-EPS arises due to a better resolution component while the reliability is only better at the higher thresholds.

4.3.2.3 Reliability diagrams

Reliability diagrams display the observed frequencies conditioned to the ensemble probabilities versus the latter (Wilks 1995). Theoretically, the ensemble probability and the conditional frequency of occurrence are equivalent for a reliable ensemble system. Then, the reliability curve agrees with the diagonal of the diagram.
However, due to limited sample sizes reliability diagrams cannot be expected to be exactly diagonal. To account for that, Bröcker and Smith (2007) have developed *consistency resampling*. This method estimates intervals of observed frequencies for which

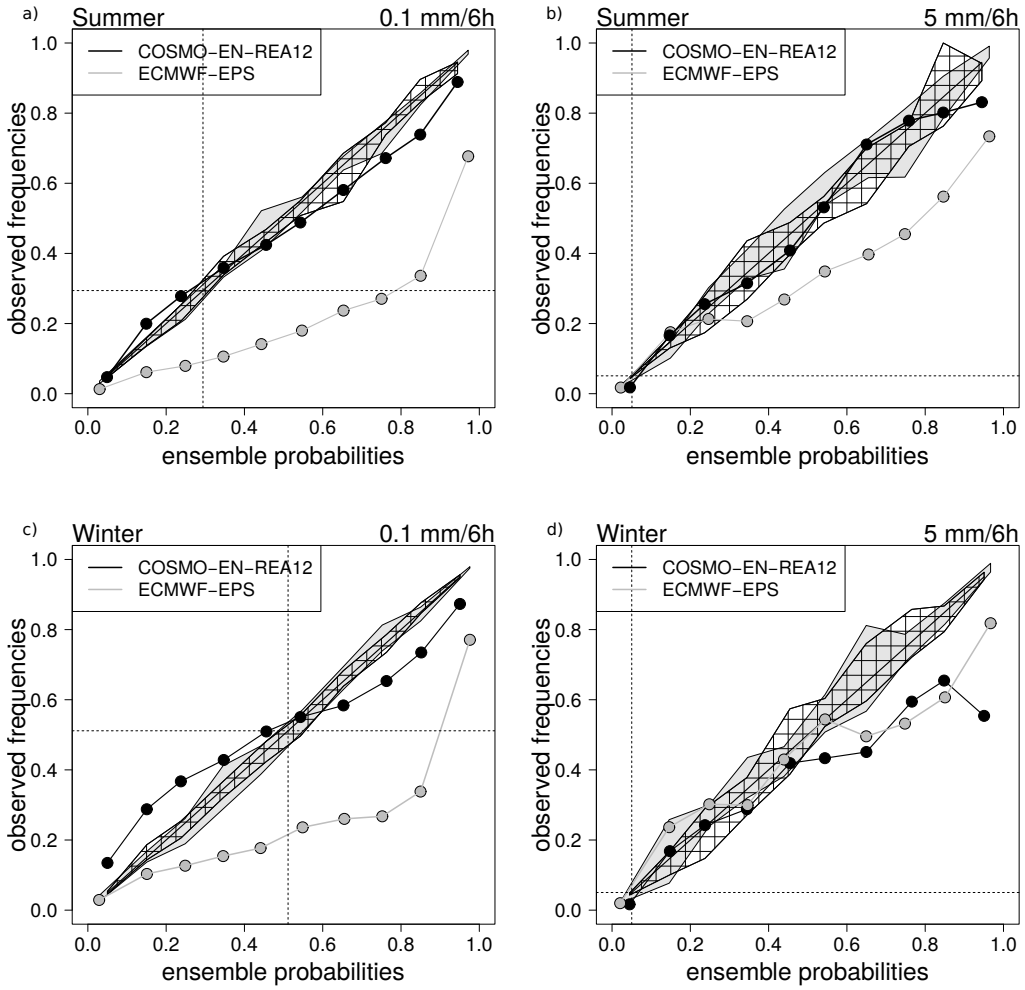

Figure 4.12: Reliability diagrams for 0.1 mm/6h [a) and c)] and 5 mm/6h decision thresholds [b) and d)] for the summer (upper row) and winter experiments (lower row). The grey shaded area represents consistency intervals for the ECMWF-EPS and the black hatched one for COSMO-EN-REA12 . These represent the areas within which the ensemble is reliable. The dashed vertical and horizontal lines represent the climatological observed frequency of the events.

the ensemble system is still reliable given sampling uncertainty. The estimated intervals are plotted over the bin means ranging from the 5 % to the 95 % percentiles. Reliability is then shown by the extent to which the observed relative frequencies fall within the consistency bars.
In a so-called resampling cycle, the whole vector of reanalysis probabilities having sample size N is sampled into a new order. In a next step, a corresponding set of binary observations is generated. For that purpose, an independent uniformly distributed random variable of sample size N is drawn. Running through all indices, the random variable is assigned 1 where it is smaller than the resampled reanalysis probability and 0 in all other places. By definition, the resampled probability vector is reliable for the new binary observations. This cycle is repeated 1000 times.
Figure 4.12 shows reliability diagrams for 6-hourly precipitation sums from COSMO-EN-REA12 and the ECMWF-EPS. Instead of bars, the consistency intervals are illustrated as polygonal lines. For all experiments and thresholds, the ECMWF-EPS except 5 mm/6h in winter exhibits over-forecasting of the observed frequencies. In the summer experiment, COSMO-EN-REA12 is well-calibrated at both thresholds.
In winter at 0.1 mm/6h, small conditional biases are observable. The small ensemble probabilities are associated with slight under-forecasting while the higher ones are associated with slight over-forecasting. Such a form of a reliability curve indicates an under-dispersive ensemble that is too confident about specific events so that similar errors occur in many of the ensemble members. In winter, at the 5 mm/6h threshold the ensemble has a conditional bias at the higher ensemble probabilities. Here, the ensemble is again over-confident so that it over-estimates the observed frequencies. A possible reason for that could be spatio-temporal displacement errors of frontal systems that are similar in most of the members. Just like the reliability component of the decomposed Brier score, the reliability curves show an added value of COSMO-EN-REA12 compared to the ECMWF-EPS.

4.3.2.4 Resolution and discrimination

The ROC (Receiver Operating Characteristic) curve is a signal detection curve for binary data whereby the probability of detection is displayed versus the false alarm rate for probabilistic decision thresholds which are illustrated as points (Mason 1982). In a perfect ensemble system the curve would run from (0,0) to (0,1) to (1,1), i.e. low decision thresholds correspond to high probabilities of detection and high false alarm rates whereas higher decision thresholds should come along with lower probabilities of detection and lower false alarm rates. The closer the curve is to the diagonal the less

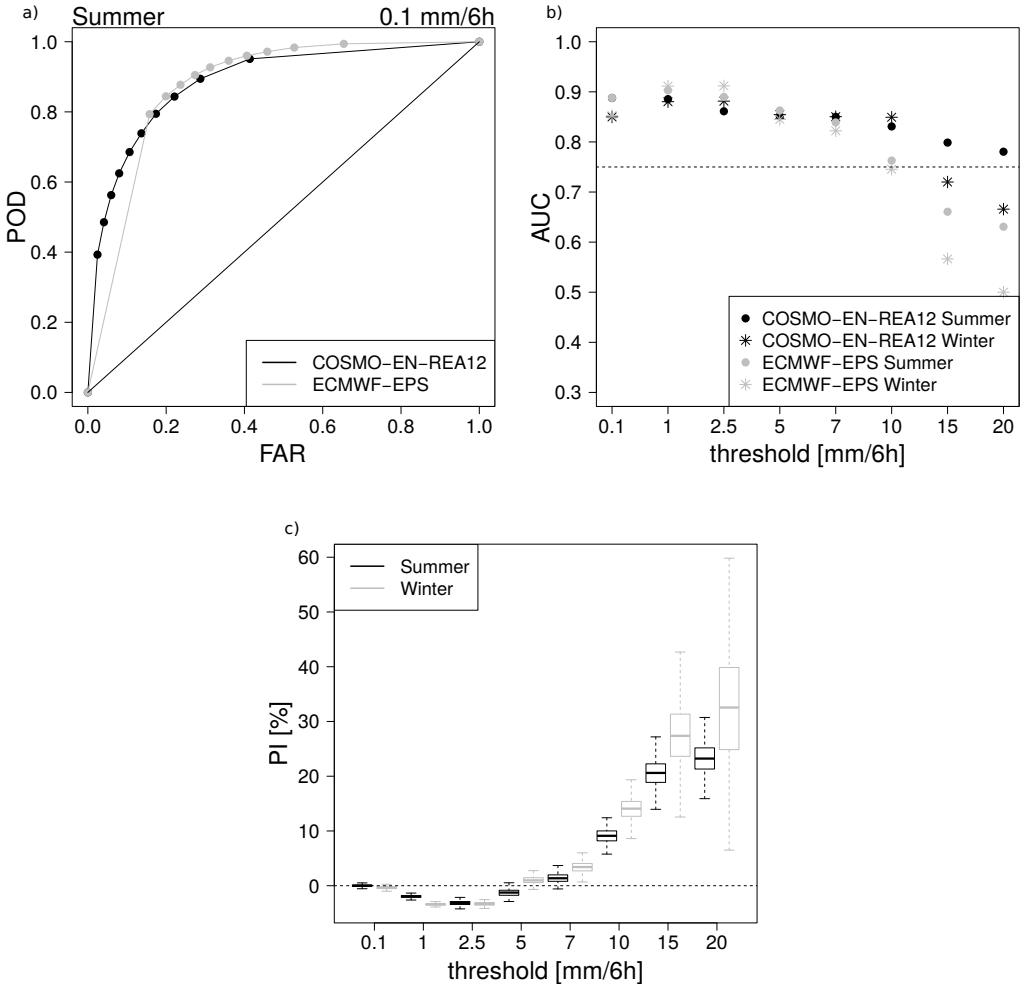

Figure 4.13: ROC curve measures. (a) Exemplary ROC curves for 0.1 mm/6h. (b) Absolute values for the area under the ROC curve (AUC) for different thresholds. (c) Percentage improvement (PI) in terms of AUC of COSMO-EN-REA12 over the ECMWF-EPS for the summer (black boxplots) and winter experiments (grey boxplots). COSMO-EN-REA12 is depicted in black and the ECMWF-EPS in grey in the upper two pictures. In c) the summer experiment is illustrated in black and the winter experiment in grey.

the ensemble system can discriminate between events and the less resolution it has.
An exemplary ROC curve is displayed in Figure 4.13 a). It shows the resolution of the probabilistic regional reanalysis system in comparison to the ECMWF-EPS based on 6-hourly precipitation sums accumulated to 06 and 18 UTC for a threshold of 0.1 mm/6h. On first sight, the ROC curves COSMO-EN-REA12 and the ECMWF-EPS seem to be comparable. However, for the higher decision thresholds, the ECMWF-EPS is shifted to higher probabilities of detection and higher probabilities of false detection compared to COSMO-EN-REA12. This is probably rooted in a much higher base rate of the IFS (see Figure 4.8) at this small threshold which has been observed by means of the frequency bias of ERA-Interim and might be similar for the IFS version that the used ECMWF-EPS is based on.
Figure 4.13 b) shows the area under the ROC curve (AUC) which allows for an easier comparison for different experiments and thresholds. Optimally, the AUC is 1. Below a value of 0.75 depicted as dashed line the discriminative capabilities of an ensemble system can be regarded as poor. It can be observed that both systems can discriminate between events that happen and events that do not happen up to a threshold of 10 mm/6h. Here, the ECMWF-EPS falls under the 0.75 line while COSMO-EN-REA12 retains an area under the ROC curve of about 0.85. Above that threshold, COSMO-EN-REA12 keeps discrimination for the summer experiment.
To quantify the quality of COSMO-EN-REA12 relative to the ECMWF-EPS, a percentage of improvement is computed for the AUC including 1000 bootstrap samples to incorporate sampling uncertainty. This is illustrated as black boxplots for summer and grey boxplots for winter in Figure 4.13 c). As discussed, the two systems are comparable for the smallest threshold. At 1 mm/6h and 2.5 mm/6h COSMO-EN-REA12 is slightly worse while from 5mm/6h upwards it is increasingly superior which is consistently more pronounced in winter. This is one of the most important features that argues for using estimates of precipitation as essential climate variables from regional reanalyses.

4.3.2.5 Discussion

Summarising the evaluation of the basic probabilistic performance measures applied to a verification of precipitation discussed in this section, a tentative answer can be given to the second research question posed above. *What are the probabilistic capabilities of the ensemble?* The analysis rank histograms show that the ensemble is quite well calibrated for the summer experiment. In winter, a more pronounced under-dispersiveness can be observed. A comparative verification to 6-hourly accumulated precipitation sums from the ECMWF-EPS (due to a lack of global reanalysis ensemble data that will soon be

available from ERA5, ECMWF) reveals that the overall probabilistic accuracy of the two systems is more or less the same for the experimental periods. This is a good result, since the ECMWF-EPS is one of the world-wide gold-standard ensemble prediction systems. However, computed for a subset of the data that is dominated by small-scale convection, the CRPSS indicates even a clear added value of the regional ensemble.
The resolution is comparable at the lower precipitation thresholds, while an added value can be observed for COSMO-EN-REA12 at higher thresholds as shown by the percentage improvement of the area under the ROC curve. Also, a clear added value in reliability becomes apparent. All in all, for the experimental period, the probabilistic regional reanalysis system demonstrates to yield an ensemble with good probabilistic capabilities.

4.4 Evaluation of the extended reanalysis suite

In the foregoing section, experiments with the *basic reanalysis suite* have been evaluated. There, observation error is accounted for by employing the newly-developed technique of ensemble nudging. Here, experiments with an *extended reanalysis suite* incorporating further uncertainty sources are assessed. Additionally to ensemble nudging, stochastic perturbation of physical tendencies (see sections 3.3.2 and 4.1) is utilized to stochastically account for model error due to the parameterized terms in the prognostic equations. Further, an ensemble of the global model ICON is engaged as probabilistic lateral boundary conditions (see sections 3.3.3 and 4.1).
In total, four experiments including reforecasts initialized at 00 UTC have been conducted. All comprise of 20 ensemble members and assimilate conventional observations and wind profilers using nudging. Further, one ICON run has been chosen as deterministic lateral boundary condition that is employed instead of ERA-Interim in the reanalysis system configurations that do not use perturbed LBCs. The experiments extend from 15/05/2014 to 15/06/2014. During this time span, various weather situations occur, including isolated strong convective cells, dry days and widespread convective and frontal precipitation. For an overview of the experiments see section 4.1.
The goal of this section is to provide an indication of which configuration of the reanalysis system is most suitable for future probabilistic reanalyses based on COSMO and nudging. For UERRA, 5 years of a test reanalysis are produced using ensemble nudging. However, it is conceivable to employ the system for a longer-term reanalysis in the future so that it is worth knowing how an improved reanalysis ensemble can be obtained. In the subsequent evaluation, the probabilistic capabilities of the different system configurations are assessed with a focus on two research questions:

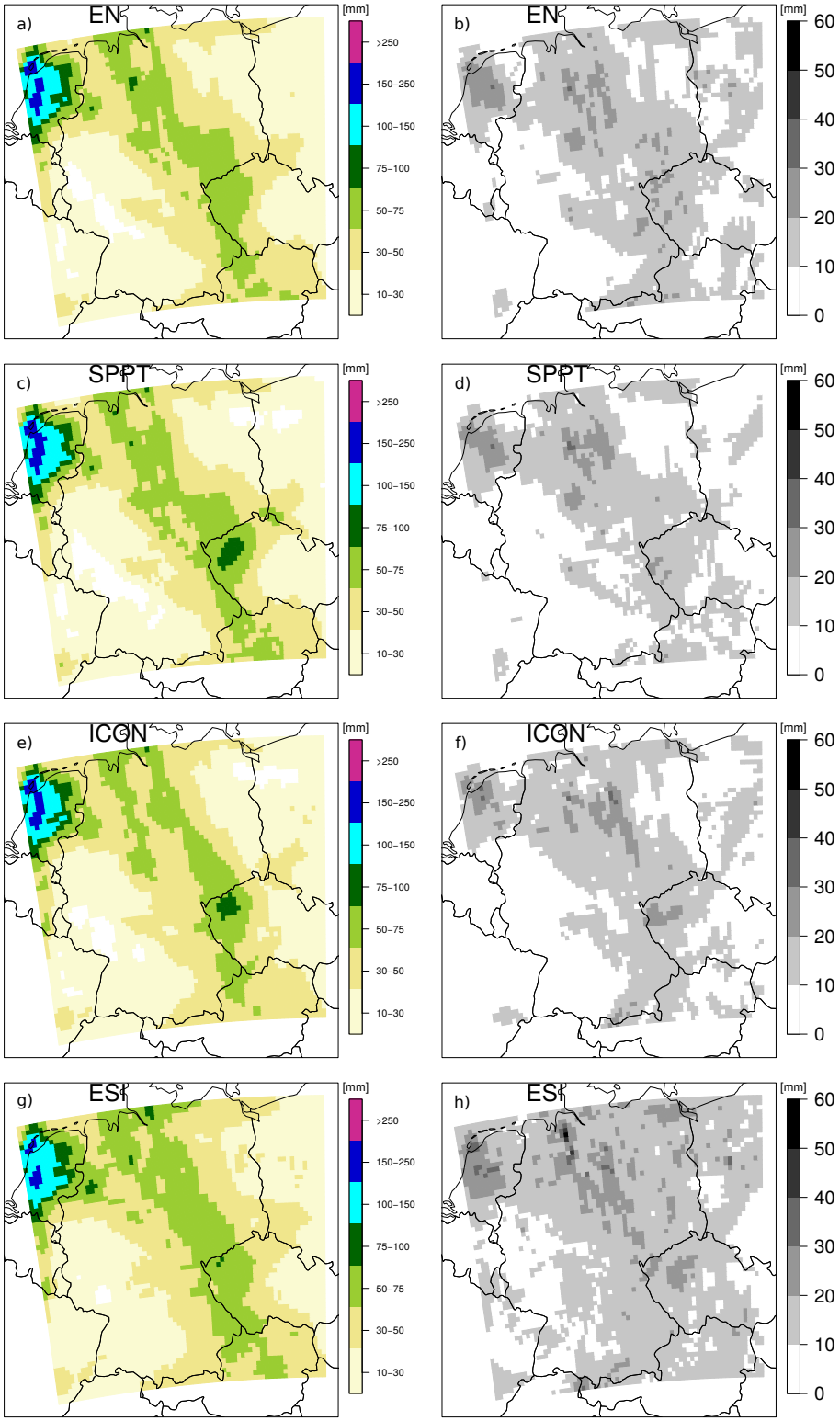

Figure 4.14: Monthly precipitation climatologies for Germany for May 15, 2014 to June 15, 2014. Left panels: ensemble mean of precipitation accumulated over experimental period in (a) *EN*, (c) *SPPT*, (e) *ICON*, (g) *ESI*. Right panels: corresponding spread of (b) *EN*, (d) *SPPT*, (f) *ICON*, (h) *ESI*.

1. How would a reanalysis ensemble be generated ideally with respect to precipitation?
2. Which reanalysis system shows the best probabilistic capabilities with respect to screen-level temperature?

As discussed, precipitation is one of the two essential climate variables for which an improvement is anticipated in regional reanalyses. Therefore, the future probabilistic regional reanalysis system should represent it as good as possible. In the second part, a comparative verification of screen-level temperature as second important essential climate variable is conducted. Note that the results are discussed with a focus on improvement over the *basic reanalysis system* using ensemble nudging.

4.4.1 Evaluation of precipitation

For the verification of precipitation, analysis data accumulated over 6 hours is used. Just as before, 1034 rain gauge stations over a German subdomain provide an observational reference (see Figure 4.6). To each station, a unique nearest neighbor grid point is assigned. Again, it is not accounted for observation errors of the rain gauges.
Figure 4.14 depicts monthly precipitation climatologies for all grid points based on the ensemble means of the experiments. All show similar precipitation patterns. However, the spread of the monthly integrated precipitation sums reveals pronounced differences. The experiment combining all uncertainty sources *ESI*, for example, shows a region with a significantly more pronounced spatial extension of 10 to 20 mm uncertainty. Also, the spread reaches values up to 60 mm while it takes maximum values of 40 mm in the other experiments. In the following, the differences between the system configurations will be observed in more detail. First, the consistency of the different ensemble reanalysis systems is tested using analysis rank histograms. This is followed by an assessment of the probabilistic accuracy based on the Brier score. Moreover, resolution and reliability are investigated in terms of a decomposition of the Brier score as well as reliability diagrams and ROC curve measures.

4.4.1.1 Consistency

Figure 4.15 compares analysis rank histograms of the four experiments conducted with the *extended reanalysis suite*. As discussed in section 4.3.2.1, analysis rank histograms provide information about the consistency of an ensemble system. This means that they reflect if the ensemble members and observations are drawn from the same PDF so that all members are equally likely. Further, rank histograms indicate under- or over-dispersiveness and under- or over-forecasting biases. Events at which both ensemble and

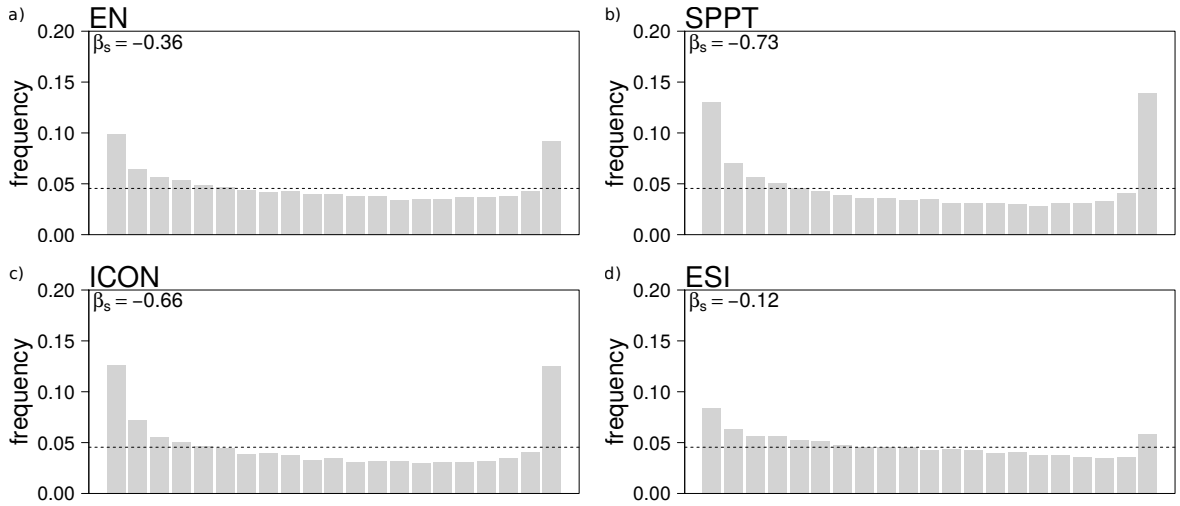

Figure 4.15: Analysis rank histograms for 6-hourly accumulated precipitation from experiments verified by means of 1034 German rain gauges depicted in Figure 4.6. (a) *EN*, (b) *SPPT*, (c) *ICON*, (d) *ESI*.

observations indicate "no precipitation" are omitted. All ensembles show a u-shaped form. The β-score (Keller and Hense 2011) reveals that the *ICON*-experiment is slightly better ($\beta = -0.66$) than the *SPPT*-experiment ($\beta = -0.73$). *EN* is significantly less under-dispersive having a β-score of -0.36 and *ESI* shows a nearly flat analysis rank histogram ($\beta = -0.12$). Reason for the under-and over-forecasting biases that most probably lead to the u-shaped form may be spatial displacements of precipitating systems that the ensemble as a whole is too confident of. A possible explanation for the superiority of the experiments incorporating observation uncertainty may be that the observation perturbations induce a better spatial spread of the position of convective cells or fronts. This may originate from perturbations in the assimilated observations that trigger or dampen convection in single ensemble members so that wider PDFs are estimated. Such effects are also supposable for *SPPT*, however, the variance of the perturbations is possibly too weakly tuned compared to the magnitude of the observation perturbations. In the experiment combining all uncertainty sources *ESI*, a projection of the perturbations on each other may potentially be the cause leading to the best consistency.

4.4.1.2 Brier score and its decomposition

Figure 4.16 depicts the Brier score and its decomposition into reliability and resolution. The score measures the probabilistic accuracy depending on decision thresholds (see section 4.3.2.2 or Brier (1950)). The more ensemble members issue an event if it occurs, the

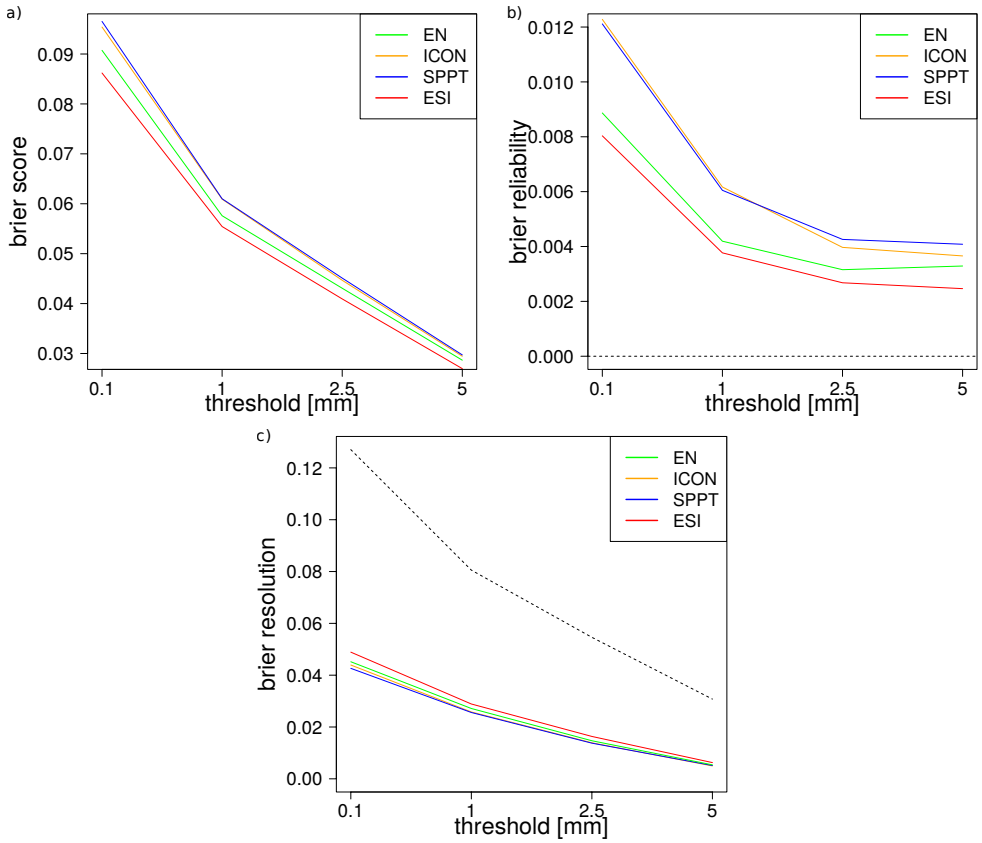

Figure 4.16: Brier score and its decomposition for 6-hourly accumulated precipitation from experiments: ensemble nudging (*EN*) in green, stochastic perturbation of physical tendencies (*SPPT*) in blue, probabilistic lateral boundary conditions (*ICON*) in orange and combined (*ESI*) in red. a) Brier score, b) Brier reliability, c) Brier resolution. The thin dashed black line in c) marks the maximally achievable resolution in a perfectly reliable ensemble system (uncertainty component).

better the Brier score will be. Being negatively oriented, it is best for the experiment combining all perturbation methods *ESI*, second best for *EN* and slightly worse for the *SPPT* and *ICON* experiments. This is reflected by the reliability component, which is significantly better for *ESI* and *EN*, but best for *ESI*. Again, the observation perturbations show to have a positive impact on the representation of the observed frequencies by the ensemble. The resolution component provides information about the capability of the ensemble to issue PDFs that significantly deviate from climatology. It is also by far best for *ESI*. Note that for this experiment the resolution reaches only approximately one third of the potential resolution that would be achieved if the reliability was perfect. Reason for that is that the reliability is strongly reduced compared to the *summer* and *winter* experiments evaluated before. The reliability diagrams shown next reveal similar.

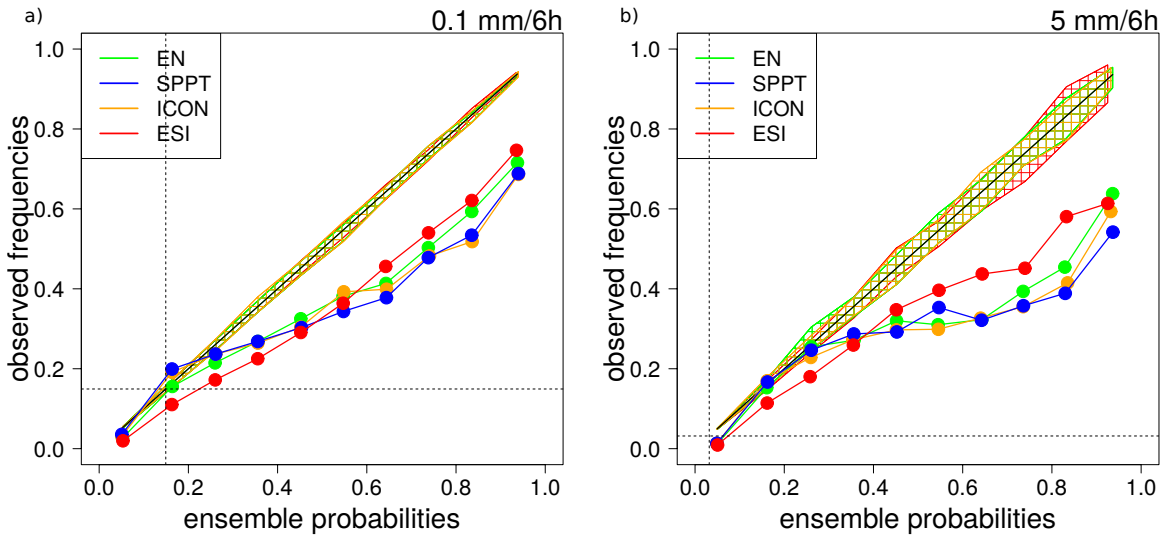

Figure 4.17: Reliability diagrams for a) 0.1 mm/6h and b) 5 mm/6h decision thresholds for experiments: ensemble nudging (*EN*) in green, stochastic perturbation of physical tendencies (*SPPT*) in blue, probabilistic lateral boundary conditions (*ICON*) in orange and combined (*ESI*) in red. The hatched areas represent consistency intervals within which the ensemble is reliable. The dashed vertical and horizontal lines represent the climatological observed frequency of the events.

4.4.1.3 Reliability

The reliability diagrams shown in Figure 4.17 indicate how well the different ensemble configurations are calibrated. As outlined in section 4.3.2.3, the observed frequencies conditioned to the ensemble probabilities would ideally be predicted by the latter. Again, 1000 cycles of consistency resampling are conducted to measure in which range of observed frequencies the systems are reliable given sampling uncertainty due to a limited sample size. At both the 0.1 mm/6h and the 5 mm/6h threshold, the ensemble systems tend to over-forecast the observed frequencies. While *ICON* and *SPPT* are reliable for the small observed frequencies at both thresholds and *EN* at the larger threshold, the combined experiment *ESI* suffers from an unconditional wet bias that expresses in an overestimation of the observed frequencies for all ensemble probabilities. However, at the higher ensemble probabilities the over-forecasting of *ESI* is less pronounced than in all other ensemble systems. Different from the 2011 experiments with the *basic reanalysis suite*, none of the experiments with the *extended reanalysis suite* indicates really good reliability properties at the chosen thresholds.

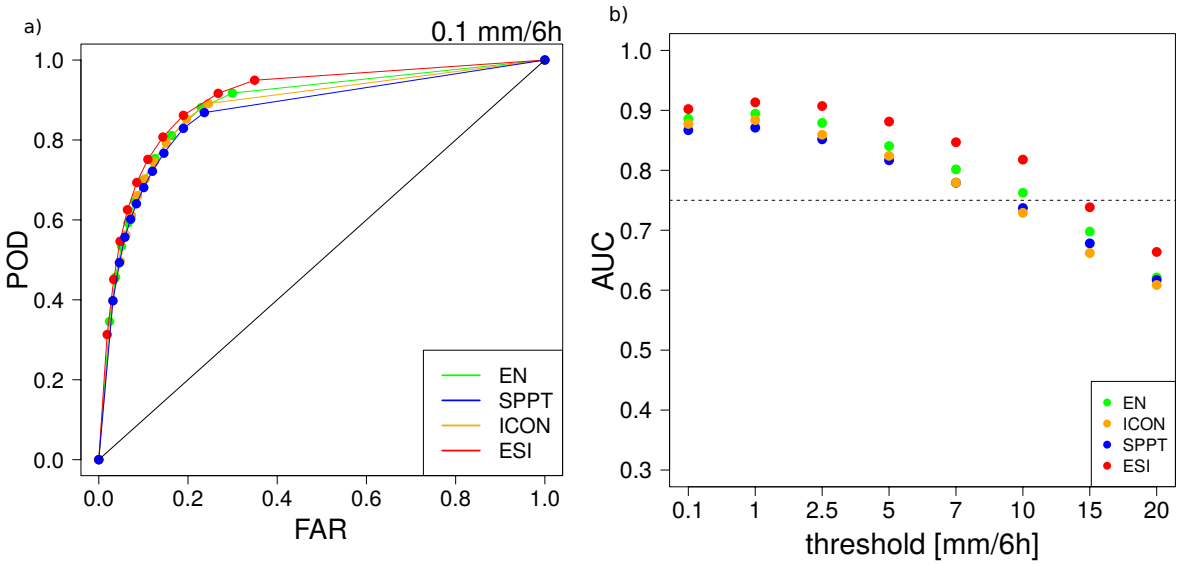

Figure 4.18: ROC curve measures. a) Exemplary ROC curves for 0.1mm, b) absolute values for the area under the ROC curve (AUC) for different thresholds. Experiments: ensemble nudging (*EN*) in green, stochastic perturbation of physical tendencies (*SPPT*) in blue, probabilistic lateral boundary conditions (*ICON*) in orange and combined (*ESI*) in red.

4.4.1.4 Resolution and discrimination

ROC curves plot the probability of detection versus the false alarm rate for probabilistic decision thresholds illustrated as points. As discussed, the curve runs ideally from (0,0) to (0,1) to (1,1), where low decision thresholds correspond to high probabilities of detection and high false alarm rates and higher decision thresholds correspond to lower probabilities of detection and lower false alarm rates. The closer the curve is to the diagonal the less the ensemble system can discriminate between events and the less resolution it has. See section 4.3.2.4 for more details.

Figure 4.18 a) compares ROC curves for all experiments conducted with the *extended reanalysis suite*. Just as before, a decision threshold of 0.1 mm/6h is chosen as example. The *SPPT* experiment is closest to the diagonal, followed by *ICON*, *EN* and *ESI*. Thus, the combined experiment can best discriminate between events that occur and ones that do not occur. The decision thresholds are shifted towards higher probabilities of detection and higher false alarm rates in the order *ESI*, *EN*, *ICON*, *SPPT*.

Figure 4.18 b) shows the area under the ROC curve (AUC) which allows for an easier comparison for different experiments and thresholds. For an ensemble system with perfect resolution, the AUC is 1 while a value of 0.75 indicates poor discriminative capabilities. The hierarchy between the experiments is for all thresholds just as observed

for 0.1 mm/6h. *ESI* is best and *SPPT* is worst. *ESI* and *EN* retain discrimination to a threshold of 10 mm/6h. At this threshold, the AUCs of *ICON* and *SPPT* fall under the no-skill line.

4.4.1.5 Discussion

The question guiding the comparative examination of the experiments conducted with the *extended reanalysis suite* in this section is how a regional ensemble reanalysis system would be ideally configured in the future. Here, this has been assessed with respect to the probabilistic representation of precipitation as essential climate variable.
Even though, due to the limited experimental time span of just one month length, a decisive conclusion cannot be drawn, the results give a relatively clear indication. In nearly all respects, the configuration of the *extended reanalysis suite* combining observation, model and LBC uncertainty (*ESI*) yields the best results. Firstly, it has the best probabilistic accuracy. Also, the system is most consistent between observations and ensemble PDF having the flattest analysis rank histogram. Further, it shows to be most reliable and to have best discriminative capabilities. This is both confirmed by the decomposition of the Brier score as well as by reliability diagrams and ROC curve measures. However, the reliability diagrams reveal an unconditional over-forecasting bias of the observed frequencies. This is due to a systematic over-estimation of the base rates at the chosen decision thresholds (not shown) and even though it is less pronounced for the *ESI* configuration, it is a weakness of the system. Since this has not been observed for the *basic reanalysis suite*, it should be investigated if it is a problem of the new COSMO model version or ICON replacing the ERA-Interim boundary conditions or due to the weather regime.

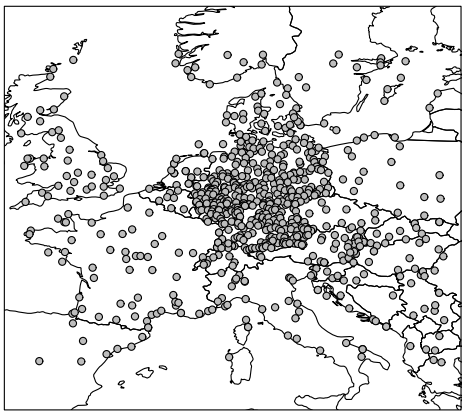

Figure 4.19: Approximately 800 European SYNOP stations used for the verification of screen-level temperature in the experiments conducted with the extended reanalysis suite.

4.4.2 Evaluation of screen-level temperature

To assess the representation of screen-level temperature in the different configurations of the *extended reanalysis suite*, hourly data from reforecasts initialized at 00 UTC is verified. To avoid a distortion of the scores due to orography differences, the four grid points nearest to each SYNOP station are searched, where the one with the smallest altitude difference to the model orography is assigned to the station. Further, only stations whose altitude is less than 150 m above or below the model orography are included. Additionally, a height correction is conducted assuming a moist adiabatic lapse rate. The stations are depicted in Figure 4.19. Even though it is questionable to make use of the mean of ensemble systems, it is common practice. Therefore, the accuracy of the ensemble mean is assessed first. This is followed by an evaluation of the probabilistic accuracy making use of the CRPS. A decomposition of the latter gives an indication of the reliability and resolution of the systems for screen-level temperature. Finally, the uncertainty estimation capabilities of the system are investigated by means of a newly developed form of the spread-skill ratio. Since the central question is how an improvement over the *basic reanalysis suite* using ensemble nudging as only ensemble generation technique can be achieved, skill scores are employed for all measures to assess a potential added value.

4.4.2.1 Accuracy of the ensemble mean

The root-mean squared error measures the spatially and temporally averaged squared difference between forecasts and observations. Its advantage is that it retains the units of the forecast variable, so that it is easily interpretable as forecast error magnitude. However, due to the squares it does not indicate the direction of deviations and gives greater weight to greater deviations. RMSE is defined as

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - o_i)^2}, \tag{4.15}$$

where N is the sample size including all grid and time points, y_i is the ith model value (here ensemble mean) and o_i the corresponding observation. For a perfectly accurate forecast, the RMSE is zero. A skill score can be formulated based on MSE such that

$$MSESS = 1 - \frac{MSE}{MSE_{ref}}. \tag{4.16}$$

The mean-squared error MSE which is the square of the RMSE can be decomposed into the forecast error variance, an observation error variance and a squared bias between

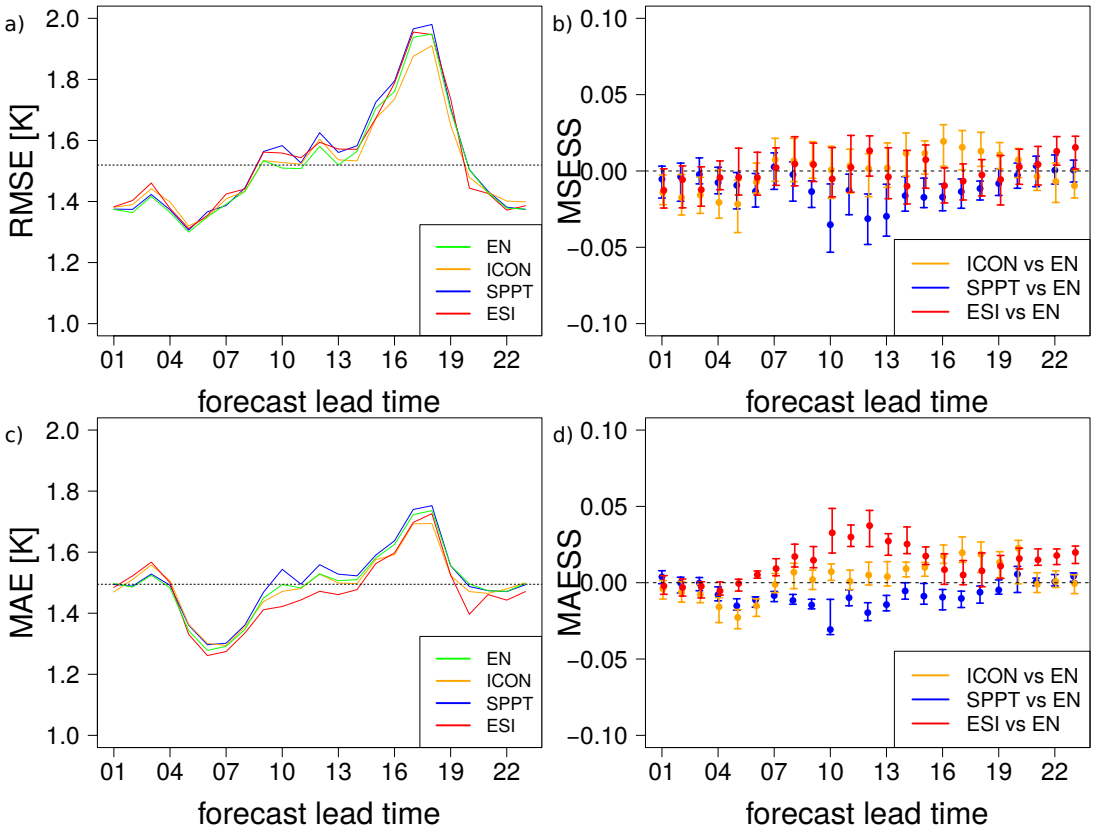

Figure 4.20: Measures of accuracy of the ensemble mean of screen-level temperature from reforecasts initialized at 00 UTC, verified against European SYNOP observations. a) RMSE, b) skill score based on MSE, c) MAE, d) skill score based on MAE. The thin black dashed lines in the left panels represent the temporal average of the accuracy measures of the *EN* experiment. Experiments: ensemble nudging (*EN*) in green, stochastic perturbation of physical tendencies (*SPPT*) in blue, probabilistic lateral boundary conditions (*ICON*) in orange and combined (*ESI*) in red. The error bars in the right panels represent the 5 to 95 % percentiles of the skill scores resulting from 1000 bootstrap cycles.

forecast and observations (Murphy and Winkler 1987). Since an ensemble cannot account for systematic errors and observation errors, the MSE is cleaned from an estimated observation error variance and a squared bias before drawing its square root.

Figure 4.20 shows the cleaned RMSE of the ensemble for all experiments. The observation error variance is assumed to be 2.0 K while the bias is estimated from the data, depending on lead-time. The RMSE evolves similarly for all experiments. Its average value is approximately 1.5 K. The increase in forecast error with lead time naturally reflects the error growth during forward integration. Note that this evolution of RMSE has also been found for forecasts initialized from a deterministic analysis using KENDA at 2.8 km grid spacing. The corresponding experiments have been conducted for the

development of KENDA and extend over the same experimental period (Schraff et al. (2016); personal communication, Christoph Schraff, DWD).
To assess a potential added value of the different system configurations over ensemble nudging, a bootstrapped skill score as defined in equation 4.16 is shown in Figure 4.20. A systematic advantage of one of the system configurations over ensemble nudging is not apparent. The experiment combining all uncertainty sources *ESI* is comparable to *EN* except for 22 and 23 hours lead time where it is superior. Between 16 and 19 hours lead time the ensemble mean of *ICON* yields a small but significant improvement.
As second measure of accuracy of the ensemble mean, the mean absolute error MAE is applied. It quantifies the average absolute deviation between forecast and observation

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - o_i|. \tag{4.17}$$

The evolution and magnitude of the MAE shown in Figure 4.20 are very similar to the one observed for RMSE. However, giving equal weight to all magnitudes of forecast error, i.e. less to large errors than RMSE, the MAE of *ESI* is smaller at a range of lead times. This is revealed by the skill score given by

$$MAESS = 1 - \frac{MAE}{MAE_{ref}}, \tag{4.18}$$

using *EN* as reference. Beginning from 7 hours lead time, the bootstrapped skill score reveals a significant improvement of up to 5 % for *ESI* and a small improvement for *ICON* between 10 and 15 UTC. *SPPT* achieves maximally comparable skill.

4.4.2.2 Probabilistic accuracy

In a next step, the probabilistic accuracy of the experiments is assessed. Taking into account the values of all ensemble members, the continuous ranked probability score (Hersbach 2000) is most suitable for this purpose. As discussed, it measures the distance of the ensemble members from the observation in terms of cumulative distribution functions. The CRPS is negatively oriented and better the sharper the ensemble is distributed around the observation.
Figure 4.21 depicts the CRPS for all experiments. The evolution with forecast lead time is similar to the one of MSE and MAE displayed in Figure 4.20. However, the CRPS averaged over all lead times is approximately 1.0 K which is 0.5 K less than the average magnitude of MAE and RMSE. The fact that the CRPS equals the MAE for one realization (Hersbach 2000) shows the added value of an ensemble over ensemble means or

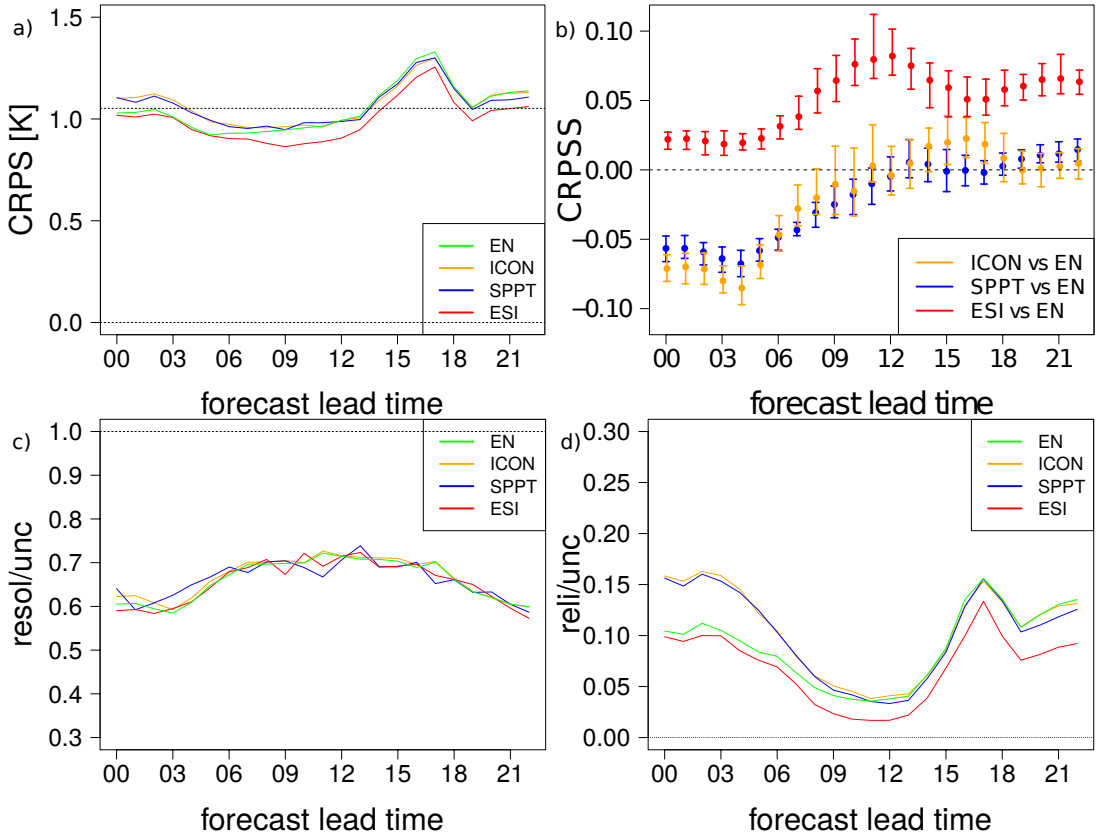

Figure 4.21: CRPS and its decomposition for reforecasts of screen-level temperature initialized at 00 UTC, verified against European SYNOP observations. a) CRPS, the thin black line represents the temporal average of the CRPS of the *EN* experiment, b) skill score based on CRPS, the error bars represent the 5 to 95 % percentiles of the skill scores resulting from 1000 bootstrap cycles, c) resolution component of the CRPS normalized by uncertainty, 1 for a perfectly reliable ensemble system, d) reliability component of the CRPS normalized by uncertainty, 0 for a perfectly reliable ensemble system. Experiments: ensemble nudging (*EN*) in green, stochastic perturbation of physical tendencies (*SPPT*) in blue, probabilistic lateral boundary conditions (*ICON*) in orange and combined (*ESI*) in red.

single realizations. The skill score based on CRPS is computed analogously to the ones based on MAE and RMSE and its temporal evolution depicted in Figure 4.21 is similar (see Figure 4.20). Differently, *ESI* has positive skill of up to 10 % for all forecast lead times when measured in terms of CRPS. The *ICON* and *SPPT* experiments, however, are inferior, particularly during the first hours of lead time. Afterwards, *ICON* is comparable to *EN*, while *SPPT* shows skill from +19 to +23 hours. The initial inferiority of *ICON* and *SPPT* may be rooted in the slow development of sufficient spread in these configurations.

Hersbach (2000) shows how the CRPS can be decomposed into reliability, resolution and uncertainty, which can be done for any proper scoring rule (Bröcker 2009). In Figure 4.21 the reliability and resolution components of the decomposed CRPS are shown. Both are normalized by the uncertainty component. The resolution component of the CRPS measures if the assessed ensemble system is capable of issuing case dependent probability forecasts. It is sensitive to the average ensemble spread and outliers. For a perfect ensemble system, the reliability component is zero and the resolution equals the uncertainty component. Thus, the resolution normalized by uncertainty is optimally 1. Figure 4.21 reveals that all experiments have similar resolution which reaches between 60 and 70 % of the uncertainty.
The reliability part of the CRPS is closely related to the consistency measured by analysis rank histograms. It tests, if the cumulative distribution function of an ensemble behaves such that the frequency of the observation falling below the middle of a bin confined by two ensemble members i and $i+1$ is equal to the frequency given by the number of member i and the total number of members N. Here, a large and significant improvement of *ESI* over *EN* can be observed which increases with lead time. From approximately 12 hours lead time *EN*, *ICON* and *SPPT* have comparable reliability while *ESI* is distinctly superior.

4.4.2.3 Uncertainty estimation

To conclude the evaluation of screen-level temperature, the uncertainty estimation capabilities of the different ensemble generation techniques employed in the *extended reanalysis suite* experiments are assessed. For this purpose, a spread-skill measure derived from a ratio of spread and mean-squared error is employed (personal communication, Rita Glowienka-Hense, Meteorological Institute of the University of Bonn). Note that the measure is not proper (Gneiting and Raftery 2007). However, it is able to measure if an ensemble yields a reliable uncertainty estimation. It is given by

$$ESS = \frac{1 - Anova}{1 + Anova - 2\rho(\bar{y}, o)\sqrt{Anova}}. \tag{4.19}$$

In this equation, the term *Anova* is obtained by an analysis of variance which decomposes the total variability of a set of probability forecasts into a variability part due to meteorological variability and one due to ensemble spread. *Anova* is equal to the explained variance of the meteorological variability so that $1 - Anova$ is the explained variance of the ensemble spread. $\rho(\bar{y}, o)$ is the correlation between the ensemble mean $\bar{y}$ and the observations o. The ESS is perfect, if the correlation between ensemble mean

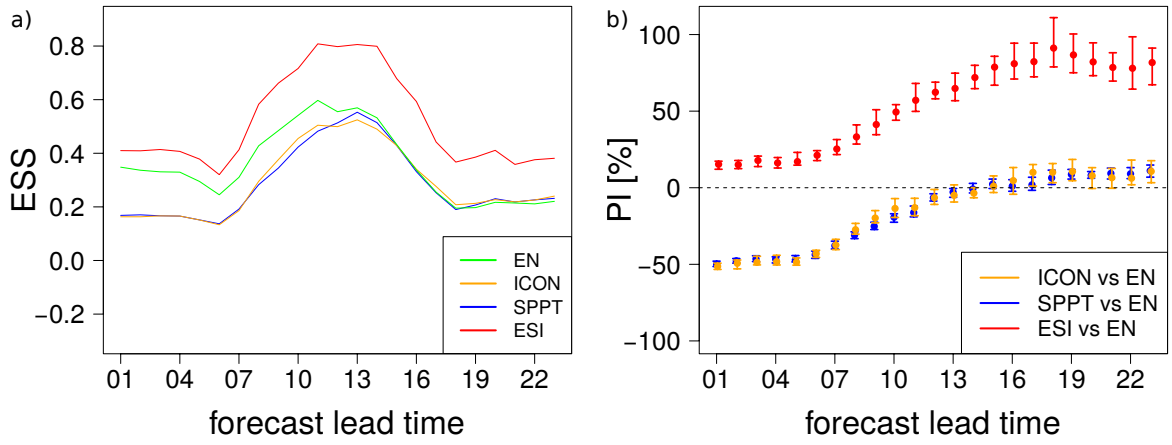

Figure 4.22: Ensemble spread score for screen-level temperature reforecasts initialized at 00 UTC verified against European SYNOP observations. a) Ensemble spread score, b) skill score based on ensemble spread score, the error bars in represent the 5 to 95 % percentiles of the skill scores resulting from 1000 bootstrap cycles. Experiments: ensemble nudging (*EN*) in green, stochastic perturbation of physical tendencies (*SPPT*) in blue, probabilistic lateral boundary conditions (*ICON*) in orange and combined (*ESI*) in red.

and observation is perfect and the ensemble is fully sharp so that there is zero spread. Note that both the reforecast data and the observations are standardized to zero mean and unit standard deviation. For details on the method of analysis of variance, be referred to von Storch and Zwiers (1999).

Figure 4.22 compares the evolution of the *ESS* with lead time for the *EN*, *SPPT*, *ICON* and *ESI* experiments. The *ESS* of all experiments improves to its best value between 10 and 13 UTC and declines afterwards. Its evolution reminds of a diurnal cycle. For the combined experiment it reaches a value of approximately 0.8 while it is only 0.5 for the other experiments. Initially, *EN* is superior to *SPPT* and *ICON*. However, towards the end it settles down to a similar level of *ESS*. The percentage improvement based on *ESS* using *EN* as reference underpins the clear advantage of the *ESI* experiment which increases up to 100 % with increasing lead time.

4.4.2.4 Discussion

In the foregoing section, the question of how a prospective regional ensemble reanalysis system would be ideally generated has been assessed by means of precipitation. Here, this is complemented by a verification of screen-level temperature which is an in equal measure important essential climate variable. The evaluation yields similarly clear results as the evaluation of precipitation, stating that an improvement over pure ensemble

nudging as applied in the *basic reanalysis suite* is achieved by accounting for model error, observation error and uncertainties in the lateral boundary conditions.
Indicated by a skill score that is formulated in terms of mean absolut error, the accuracy of the ensemble mean is best for the *ESI* configuration. Measured in terms of root-mean squared error, no systematic improvement over ensemble nudging is found. The continuous ranked probability score, however, shows an even more pronounced improvement of the *ESI* configuration, particularly in terms of reliability. The resolution component is only improved in the first and last hours of lead time. Note that the score also shows the added value of an ensemble, since the probabilistic accuracy is reduced by about one third compared to the MAE of the ensemble mean which equals the CRPS for one realization. Finally, the uncertainty estimation as measured by a spread-skill ratio is strongly improved (up to 100 %) for the *ESI* experiment.

5 Summary, discussion and outlook

In this chapter, the most important points of this thesis are summarized and discussed. Afterwards, ideas for further developments are outlined in an outlook.

5.1 Summary and discussion

Regional reanalyses are data sets that are relatively new in climate research. They have promising attributes including spatio-temporal completeness, accuracy and consistency between the variables. Particularly appealing is their significantly increased spatial sampling and the fact that they analyse the atmospheric mesoscale additionally to the large scales which are represented in global reanalyses. Thereby, regional reanalyses newly offer the possibility to monitor climate on a local scale. Compared to the widely-used dynamical downscalings of global reanalyses using limited-area models, the observations assimilated in regional reanalyses add accurate mesoscale details to the 4-dimensional fields evolving from the reanalysis process. Therefore, they allow not only for the assessment of frequency distributions, but also for an examination of weather situations realistic in space and time.

In recent years, it has been recognized that uncertainty estimation is similarly important in reanalysis as in numerical weather prediction, even though analysis errors usually have much smaller amplitudes than forecast errors. Still, since users tend to use reanalysis data as equivalent to truth, it is of great importance to provide them with uncertainty estimates. In the framework of this thesis, a probabilistic regional reanalysis system for Europe based on the limited-area model COSMO has been implemented for the first time. For that purpose, a new ensemble technique based on observation nudging has been developed. The work has been conducted in the framework of the European project Uncertainties in Ensembles of Regional Reanalyses (UERRA), in which predominantly European met services have taken part. This has required to configure the reanalysis system in an operational framework with high demands to the quality of the implementation. Since the data will be made accessible to the public in the MARS archive of ECMWF, they have to fulfill high quality standards. Only close cooperation with DWD has allowed for developing a reanalysis system with sufficient capabilities.

The basic implementation of the probabilistic regional reanalysis system developed in the course of this work will be employed to produce a test ensemble reanalysis extending over 5 years in the framework of UERRA. The suite is set up for the CORDEX-EUR11 domain at a grid spacing of 12 km. Using the newly developed method of ensemble nudging as data assimilation scheme, the production configuration of the system allows for estimating the sensitivity of the reanalysis to uncertainties in the observations, i.e. representativity errors, errors in the observation operators as well as measurement errors and their projection on the model dynamics. An extended version of the system additionally provides capabilities for probabilistically accounting for model error and uncertainties in the LBCs.
For the evaluation of the new system, six numerical experiments have been conducted, each of which extends over one month. Initially, basic diagnostics of the performance of the reanalysis system have been assessed. This has been followed by an evaluation of the basic reanalysis suite conducting a comprehensive evaluation of precipitation by means of two numerical experiments. In a third step, four numerical experiments with the extended reanalysis suite have been evaluated to figure out how a prospective system would be ideally configured and which uncertainty sources should be accounted for. Here, precipitation and screen-level temperature have been assessed. Note that the evaluation periods of one month length are too short for drawing decisive conclusions on the quality of the new system. It can be expected that once that the 5-year reanalysis data set will be available the results might not anymore look entirely the same. Still, the pilot studies have been useful to observe tendencies for the reanalysis system's quality. In the next two sections, the results from the experiments with the *basic* and *extended reanalysis systems* are discussed.

5.1.1 Evaluation of the basic reanalysis suite

The deterministic verification of precipitation shows that the accuracy of COSMO-ENREA12 is comparable to ERA-Interim and thus suffers the loss of the clear added value that can be observed for the same system in higher resolution as shown by the reanalysis COSMO-REA6. By tendency, the accuracy of the ensemble nudging runs is slightly better in summer than in winter. However, different from ERA-Interim the frequency bias is nearly perfect. Not surprising, there is a strong added value observable in accuracy compared to a dynamical downscaling from ERA-Interim at a 6 km grid spacing.
While the bisection of the grid spacing relative to COSMO-REA6 leads to a degradation of accuracy, a clear benefit of the system is the ensemble mode. The analysis rank histograms show a nearly perfectly calibrated ensemble in summer. In winter, the under-

dispersiveness is more pronounced. As shown by the CRPS, the probabilistic accuracy taking into account the whole range of precipitation amounts and all meteorological situations having occurred in the analysed months is comparable to the ECMWF-EPS. However, with atmospheric conditions favorable for the development of smaller-scale convective regimes, a distinct added value could be found compared to ECMWF-EPS precipitation fields.
COSMO-EN-REA12 proves to have a significantly better reliability, while the resolution component of the Brier score is a bit worse in winter for the chosen decision thresholds. The discriminative capabilities as summarized by the area under the ROC curve show a percentage improvement of up to 50 % compared to the ECMWF-EPS which shows that the higher grid resolution of regional reanalyses has well an added value, particularly at the higher precipitation thresholds. This is required for monitoring of extreme events on increasingly short time scales (personal communication, Blaz Curnik, European Environmental Agency; 3rd UERRA general meeting, Meteo France, Toulouse). All in all it can be summarized that the basic implementation of the regional ensemble reanalysis system that will be employed in production mode shows good probabilistic and uncertainty estimation capabilities even though the results would be even better if a 6 km grid spacing could be employed analogously to COSMO-REA6. However, this trade-off has to be made to allow for a comprehensive uncertainty estimation using a certain number of ensemble members which is an essential new information for all users of data sets of essential climate variables.
Note that before the suite can be run operationally meeting the standards defined in UERRA, final steps have to be made, i.e. programs for post-processing of the output variables (e.g. compute horizontal wind speed from horizontal velocity components), transformation from data format wgrib1 to wgrib2 and archiving to the MARS archive have to be incorporated in the production system.

5.1.2 Evaluation of the extended reanalysis suite

The evaluation of the extended reanalysis suite with respect to the essential climate variables of precipitation and screen-level temperature has been guided by the question of how a prospective system could be improved by incorporating further uncertainty sources. To give a tentative answer to this question, four numerical experiments have been assessed, three of which just account for either observational uncertainty (ensemble nudging) or model uncertainty (stochastic perturbation of physical tendencies) or uncertainty in the LBCs (ICON ensemble). A further experiment combines all perturbation methods. Even though the experimental period is short, the results give a quite

clear indication. The probabilistic scores compared for the four numerical experiments by means of precipitation show that the system combining all perturbations methods is systematically superior in the experimental period. It is most consistent as measured by analysis rank histograms, has the best probabilistic accuracy as illustrated by the Brier score, is most reliable and can best issue case-dependent PDFs that significantly deviate from the climatology. Similar is shown by the verification of screen-level temperature. Both the skill scores formulated based on mean-absolute error and CRPS indicate the superiority of perturbing observations, model and LBCs. Thus, both the ensemble mean and the ensemble as a whole are improved. Very important, the uncertainty estimation capabilities measured by a new form of spread-skill ratio strongly gain reliability when accounting for the three employed sources of error. Since the ensemble nudging experiment scores second best in nearly every respect, a further important result of this evaluation is that in the basic reanalysis suite the most suitable single-perturbation method has been chosen. All in all, the results entail the recommendation to incorporate as many uncertainty sources in a prospective reanalysis system as possible.

5.2 Outlook

Compared to global reanalyses, the field of regional reanalyses is only in its beginning. So far, mainly basic configurations of regional reanalysis systems have been set up. A novel idea is the computation of probabilistic reanalyses. This has been picked up in the work on hand.

At the University of Bonn, the feasibility of technically setting up deterministic and probabilistic regional reanalysis suites has been demonstrated (Hans-Ertel Centre for Weather Research, 2011-2014; UERRA, 2014-now). This opens the possibility for heading towards suites of higher stage. To conclude this work, ideas for further advancement of the COSMO reanalysis systems are discussed. These relate to data assimilation and model development, the observation stream, ensemble generation techniques and verification of reanalysis data.

5.2.1 Data assimilation and model development

The currently available regional reanalysis systems based on COSMO employ observation nudging as data assimilation technique. Nudging has been used successfully in daily operations of DWD for nearly 20 years. However, it is an empirical method that is not mathematically optimal in a maximum-likelihood or least-squares sense, at least in the employed implementation. Rather, a time-constant nudging coefficient determines

how strongly the model is drawn towards the observations per time interval. Moreover, nudging uses temporally constant auto-regressive weighting functions for the spreading of observational information and can thus not sufficiently account for the error of the day. The Hans-Ertel-Centre has initialized the development of a new ensemble reanalysis system (personal communication, Roland Potthast, DWD). This system will be based on the new implementation of a local ensemble transform Kalman filter KENDA for the convective scale (Schraff et al. 2016). Since the LETKF requires the background ensemble to sample the subspace of uncertainty as efficiently as possible to obtain a meaningful analysis, diverse ensemble generation methods and covariance inflation techniques are employed with KENDA. Particularly the use of perturbed lateral boundary conditions has proven beneficial. Additionally, experiments have shown that a number of at least 40 ensemble members is required to obtain reasonable Gaussian statistics. Combined with the need for an ensemble of lateral boundary conditions, this poses a challenge to the applicability of the LETKF for the purpose of reanalysis. Initially, the idea of adding perturbations from a 40-member ICON-ensemble to the global reanalyses ERA-Interim (Dee et al. 2011a) or ERA5 (Hersbach and Dee 2016) will be tested. The latter is beneficial, since it will allow for retaining the high-quality global reanalysis data. These can certainly not be reproduced at similar quality with the resources that the Hans-Ertel Centre or DWD have on disposal. However, this approach will require an additional computation of 40 global ICON-members. Also, it has to be assessed if the ICON perturbations added to IFS/ERA-Interim will be instantly dampened by COSMO through gravity waves or strong precipitation due to dynamical imbalances. One of the most important advantages of replacing nudging by KENDA will be the positive effect on daily NWP operations that can be expected from testing the system over long time spans. Similarly important is the possibility of the LETKF to incorporate satellite data which has not proven feasible for nudging, related to the computation of the analysis in model space.

A further possibility for advancement is offered by the update of the lower boundary conditions through the external data assimilation schemes that are applied offline to the output of COSMO. In the future, satellite radiances may be incorporated in the assimilation for the analysis of snow distribution, soil moisture and sea surface temperature. This is done in the land data assimilation system LDAS for the land surface scheme H-Tessel of ECMWF (Rosnay et al. 2016).

An idea for the farer-distant future is the use of earth-system models that fully describe the dynamics of all climate sub-systems, i.e. atmosphere, ocean, cryosphere, biosphere, land-surface and aerosols and tracer gases. In recent years, intensive development of such coupled earth system models and coupled data assimilation has been initialized.

These models have the potential to improve the representation of the hydrological cycle and conservation in general as e.g. fluxes between the interfaces of the subsystems are better described. In the Transregional Collaborative Research Centre 32 *Patterns in Soil-Vegetation-Atmosphere-Systems*, under lead of the Meteorological Institute of the University of Bonn, a system including COSMO as atmospheric model, CLM as land surface model and Parflow as three-dimensional ground water flow model (TerrSysMP) is developed. Currently, KENDA is tested as atmospheric data assimilation component for the system. It is planned to perform a short test reanalysis with TerrSysMP-KENDA for a small sub-domain of Germany extending over North Rhine-Westphalia (personal communication, Clarissa Figura).
An aspect that is related to model development is concerned with the representation of different meteorological variables in COSMO. Since the data is planned to be disseminated for public use, the model might be tuned to achieve better accuracy for some essential climate variables that may be of less importance in numerical weather prediction or are post-processed before further use. In principle, such statistical post-processing steps correcting for biases or even calibrating the ensemble are also well conceivable within the reanalysis suite before the data are archived. However, since the dynamical balances between the variables being one of the major advantages of reanalyses are destroyed by post-processing, it is recommendable to offer both the original and the post-processed fields, so that the users can choose the ones fulfilling their requirements.

5.2.2 Incorporation of uncertainties

In the work on hand, methodologies to account for uncertainties of reanalyses have been implemented and initial experiments have been conducted. The methods include ensemble nudging, stochastic perturbation of physical tendencies and perturbed lateral boundary conditions. They account for observation error, model error and uncertainties introduced by the steering model. All techniques are suitable for the generation of equally-likely ensemble members. An experiment with a perturbed physics ensemble with 20 unique combinations of tuning parameters (personal communication, Christoph Gebhardt, DWD; Andrea Montani, Arpa Emilia-Romagna) has yielded ensemble members with very different biases and has therefore not been incorporated in the suite.
Different aspects of the ensemble generation methods are improvable. In its current implementation, ensemble nudging makes use of random samples from a full normal distribution with zero mean and an estimated observation error standard deviation. With low probability this can yield large observation perturbations. To prevent this, it is planned to draw random samples from a truncated normal distribution in the future.

Besides, the size of the observation error standard deviations has to be re-estimated. At present, rather old estimates that have been used originally for the observation quality control are employed. Further, possibilities for accounting for spatio-temporally correlated observation errors should be explored, at least for the upper-air observing systems in which the measurements do not have a sufficient spatial separation to be assumed uncorrelated.

The configuration of SPPT in the extended reanalysis suite makes use of relatively small variances for drawing the random numbers used for perturbation. Reason for that have been model aborts related to the appearence of $2\Delta x$-waves in COSMO. At DWD, these entailed the adjustment of a divergence dampening factor over steep terrain to improve numerical stability. For the development of the extended reanalysis suite, this resulted in a renewal of the COSMO version and replication of the performed experiments. At the same time, the variance of the perturbations was scaled down in the reanalysis system to improve stability (personal communication, Christoph Schraff, DWD). Should SPPT be utilized for the production of a longer-term reanalysis, the regulation of the corresponding tuning parameters should be assessed in a couple of sensitivity studies.

In the current configuration of the extended reanalysis system, ICON is employed as lateral boundary conditions. The suite offers the possibility to make use of either only one realization or of an ICON ensemble. In the future, the global reanalysis ERA5 of ECMWF (Hersbach and Dee 2016) may be incorporated as ensemble of LBCs. As discussed, it comprises ten members and is easily portable to the extended reanalysis suite.

This work has shown that the reanalysis ensemble and subsequent reforecasts are systematically underdispersive. A major problem that has been recognized is the insufficient representation of displacement errors of frontal or convective systems in mesoscale ensembles. The best results are obtained by combining ensemble nudging with stochastic perturbation of physical tendencies and an ensemble of lateral boundary conditions. It may be considered to include perturbation of even more components of the reanalysis system. To give an example, the uncertainty in the lower boundary conditions might be accounted for. At DWD, perturbation of soil moisture fields (Schraff et al. 2016) and sea surface temperature (personal communication, Alexander Cress, DWD) has turned out to be a valuable source of spread. In the current implementation, this is only implicitly accounted for by updating the lower boundary conditions of each member separately.

Finally, the observation density is a major factor of reanalysis uncertainty. In principle, computing ensembles of reanalyses offers the possibility to assimilate different subsets of the available observations in different ensemble members to quantify the related uncertainty. However, such techniques can also be employed for experiments in the development phases. Thereby, a most suitable observation stream with respect to accuracy and

long-term homogeneity could be identified and assimilated in each ensemble member to retain the objective of equally-likely ensemble members.

5.2.3 Observation stream

Up to present, not much attention has been paid to the observation stream of the COSMO reanalysis systems. As discussed in in chapter 2, the changing observation density and measurement accuracy have significant impact on the usefulness of reanalysis data for long-term trend monitoring. So far, the goal has been to implement basic reanalysis suites that can then be stepwise advanced. However, if the data are planned to be employed for trend analyses in the future, it is recommendable to invest work on the observations that are being assimilated.

Currently, the observations are retrieved from the data base of DWD. The observation quality control is conducted by the standard operational scheme. However, this quality control is not able to take into account temporally changing observing system characteristics. To give an example, Vaisala RS92 radiosondes have a dry bias near saturation. For many years, this bias has been corrected for in the pre-processing of the observation data. Recently, the data are directly corrected after measurement so that the bias-correction conducted in COSMO is not longer necessary (personal communication, Christoph Schraff, DWD). It has to be carefully checked that such double-corrections and related problems are excluded.

Further, the temporal availability and density of all incorporated observation types should be investigated. Temperature observations from aircrafts, for example, are only available from the early 1990s. Problems related to the explosion of the global observing system due to satellite data from 1979 have been discussed in section 2.4. Such a sudden large amount of additional observations can lead to a significant reduction of root-mean squared error in the presence of model bias, which introduces spurious signals to the time series. To quantify the impact of certain observational sources, observation simulation or data denial experiments could be conducted following the example of ECMWF. The potential of adaptive observation thinning to obtain an approximately constant observation density might be tested. In any case, all meta data should be made accessible for users to allow for full transparency. An idea that has been discussed is to use the quality-controlled observation input of the global reanalyses from the observational data base that is currently built up at ECMWF. Further, it may be examined if the homogenized radiosonde record by Haimberger (2007) can be used for assimilation in the future. Concluding, an improvement of the observation stream is considered to be a working step of great priority for the further development of the COSMO reanalysis suites.

5.2.4 Verification

A fundamental aspect of reanalysis projects is verification with independent observations. This gives beneficial feedback to the development of reanalysis systems and quantifies improvements. Further, it helps to inform users about strengths and weaknesses of the systems.
In the framework of this thesis, precipitation has been in the focus of verification. In the future, the employed deterministic and probabilistic scores may be complemented by use of scores that account for spatial characteristics of precipitating systems. Examples may be neighborhood methods, object-based techniques, scale separation and approaches that account for field deformation. Additionally, the set of verified variables will be extended, e.g. upper-air profiles should be included. Once that a longer time span of COSMO-EN-REA12 will be available, it will further be interesting to investigate how the uncertainty in the ensemble reflects in climate indices.
A major problem of reanalysis is that few independent data remain for verification since most of the data are being assimilated. Potentially, the feedback files of the reanalyses might be checked for observations that are set passive in redundancy checks and thus not assimilated, but actually have a good quality. Also, the potential of verification using modern high-density observation sources like MODE-S aircraft data (de Haan 2011) that are not yet assimilated can be explored.
A special aspect of reanalysis is the fact that analysis errors are typically smaller than forecast errors. Therefore, it is much more important to account for observation errors when verifying reanalysis data (personal communication, Christoph Schraff, DWD; Hans Hersbach, ECMWF). In principle, there exists a range of suitable methods in literature, see e.g. Röpnack et al. (2013), Saetra et al. (2004), Bowler (2006), Hamill (2001). However, these do not seem to have found their way to frequent application yet. Since the Hans-Ertel Centre for Weather Research currently focuses on evaluation of reanalysis data rather than on production, the development of further deterministic and probabilistic scores that take into account observation error may be interesting and valuable. Further, a more detailed investigation of the trade-off between the effective resolution of COSMO and the number of ensemble members is required. This may be investigated by producing experimental ensembles at different grid spacings and by quantifying the sensitivity of probabilistic scores to the number of ensemble members in bootstrapping experiments. In principal, the discussions in UERRA seem to have turned towards computing both deterministic high-resolution regional reanalyses and complementary probabilistic ones with coarser grids. At the Royal Netherlands Meteorological Institute (KNMI) it is planned to explore statistical techniques for estimating the uncertainties

of high-resolution reanalyses using coarser ensembles (personal communication, Maarten Pflieger, KNMI, Copernicus Workshop on Regional Reanalyses).

Finally, verification of regional reanalyses naturally comprises comparison to related data sets. It has emerged that taking a look at certain variables with a range of scores is rather impractical to obtain a simple picture of the performance of different systems. At the Met Office, a so-called NWP index[1] is computed in daily operations which takes into account the skill scores of a set of different variables (screen-level temperature, wind speed and direction, precipitation, total cloud amount, cloud base height, near-surface visibility) based on root-mean squared error with persistence as a reference. The skill scores of all parameters are then combined to a single value using weights that reflect the importance of the meteorological variables. Keune et al. (2014) show how spatio-temporal dependencies between different variables resulting from the atmospheric dynamics can be incorporated into such a verification to obtain a more realistic estimation of skill. Certainly, a score like the NWP index is not useful to obtain deep insight into the performance of a reanalysis system. However, complemented by the ability to account for correlations between the most important essential climate variables it might help to make decisions, e.g. which system should be run operationally by the Copernicus Climate Change Service.

[1] http://www.metoffice.gov.uk/research/weather/numerical-modelling/verification/uk-nwp-index-doc

Bibliography

Agresti, A. and Hitchcock, D. (2005). Bayesian inference for categorical data analysis. *Stat. methods Appl.*, 14:297–330.

Anderson, J. L. (1996). A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Clim.*, 9:1518–1530.

Anderson, J. L. (1997). The Impact of Dynamical Constraints on the Selection of Initial Conditions for Ensemble Predictions: Low-Order Perfect Model Results. *Mon. Weather Rev.*, 125(11):2969–2983.

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

Arpe, K. (1991). The hydrological cycle in the ECMWF short range forecasts. *Dyn. Atmos. Ocean.*, 16:33–59.

Bach, L., Schraff, C., Keller, J., and Hense, A. (2016). Towards a probabilistic regional reanalysis system for Europe : evaluation of precipitation from experiments. *Tellus A*, 68:1–21.

Baehr, J., Fröhlich, K., Botzet, M., Domeisen, D. I. V., Kornblueh, L., Notz, D., Piontek, R., Pohlmann, H., Tietsche, S., and Müller, W. A. (2014). The prediction of surface temperature in the new seasonal prediction system based on the MPI-ESM coupled climate model. *Clim. Dyn.*, 44(9-10):2723–2735.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., and Reinhardt, T. (2011). Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities. *Mon. Weather Rev.*, 139(12):3887–3905.

Bengtsson, L., Arkin, P., Berrisford, P., Bougeault, P., Folland, C. K., Gordon, C., Haines, K., Hodges, K. I., Jones, P., Kallberg, P., Rayner, N., Simmons, A. J., Stammer, D., Thorne, P. W., Uppala, S., and Vose, R. (2007). The need for a dynamical climate reanalysis. *Bull. Am. Meteorol. Soc.*, 88(4):495–501.

Bengtsson, L., Hagemann, S., and Hodges, K. I. (2004). Can climate trends be calculated from reanalysis data? *J. Geophys. Res. D Atmos.*, 109(11):1–8.

Bengtsson, L. and Shukla, J. (1988). Integration of space and in situ observations to study global climate change. *Bull. Am. Meteorol. Soc.*, 69(10):1130–1143.

Berner, J., Shutts, G. J., Leutbecher, M., and Palmer, T. N. (2009). A Spectral Stochastic Kinetic Energy Backscatter Scheme and Its Impact on Flow-Dependent Predictability in the ECMWF Ensemble Prediction System. *J. Atmos. Sci.*, 66:603–626.

Betts, A. K. and Ball, J. H. (1999). Basin-scale surface water and energy budgets for the Mississippi from the ECMWF reanalysis. *J. Geophys. Res.*, 104:19293–19306.

Betts, A. K., Zhao, M., Dirmeyer, P. A., and Beljaars, A. C. M. (2006). Comparison of ERA40 and NCEP / DOE near-surface data sets with other ISLSCP-II data sets. *J. Geophys. Res.*, 111(August):1–20.

Bierdel, L., Friederichs, P., and Bentzien, S. (2012). Spatial kinetic energy spectra in the convection-permitting limited-area NWP model COSMO-DE. *Meteorol. Zeitschrift*, 21(3):245–258.

Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., and Zemp, M. (2014). The concept of essential climate variables in support of climate research, applications, and policy. *Bull. Am. Meteorol. Soc.*, 95(9):1431–1443.

Bollmeyer, C., Keller, J. D., Ohlwein, C., Wahl, S., Crewell, S., Friederichs, P., Hense, A., Keune, J., Kneifel, S., Pscheidt, I., Redl, S., and Steinke, S. (2015). Towards a high-resolution regional reanalysis for the European CORDEX domain. *Q. J. R. Meteorol. Soc.*, 141(686):1–15.

Bony, S., Sud, Y., Lau, K. M., Susskind, J., and Saha, S. (1997a). Comparison and satellite assessment of NASA/DAO and NCEP-NCAR reanalyses over tropical ocean: Atmospheric hydrology and radiation. *J. Clim.*, 10(6):1441–1462.

Bony, S., Sud, Y., Lau, K. M., Susskind, J., and Saha, S. (1997b). Comparison and satellite assessment of NASA/DAO and NCEP-NCAR reanalyses over tropical ocean: Atmospheric hydrology and radiation. *J. Clim.*, 10(6):1441–1462.

Bott, A. (1989). A positive definite advection scheme obtained by nonlinear renormalization of the advective fluxes. *Mon. Weather Rev.*, 117:1006–1015.

Bowler, N. (2006). Explicitly Accounting for Observation Error in Categorical Verification of Forecasts. *Mon. Weather Rev.*, 134:1600–1606.

Brier, G. (1950). Verification of forecasts expressed in terms of probability. *Mon. Weather Rev.*, 78(1):189–195.

Bröcker, J. (2009). Reliability, sufficiency, and the decomposition of proper scores. *Q. J. R. Meteorol. Soc.*, 135(July):1512–1519.

Bröcker, J. and Smith, L. (2007). Increasing the Reliability of Reliability Diagrams. *Weather Forecast.*, 22(3):651–661.

Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B., and Jones, P. D. (2006). Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res. Atmos.*, 111(12):1–21.

Bromwich, D. H., Wilson, A. B., Bai, L.-S., Moore, W. K., and Bauer, P. (2016). A comparison of the regional Arctic System Reanalysis and the global ERA-Interim Reanalysis for the Arctic. *Q. J. R. Meteorol. Soc.*, 142(January):644–658.

Buizza, R., Miller, M., and Palmer, T. N. (1999). Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Q. J. R. Meteorol. Soc.*, 125:2887–2908.

Chen, X., Liu, C., Driscoll, K. O., Mayer, B., Su, J., and Pohlmann, T. (2013). On the nudging terms at open boundaries in regional ocean models. *Ocean Model.*, 66:14–25.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Bronnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, O., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J. (2011). The Twentieth Century Reanalysis Project. *Q. J. R. Meteorol. Soc.*, 137(654):1–28.

Corazza, M., Kalnay, E., Patil, D., Yang, S.-C., Morss, R., Cai, M., Szunyogh, I., Hunt, B., and Yorke, J. (2003). Use of the breeding technique to estimate the structure of the analysis "errors of the day". *Nonlinear Process. Geophys. Eur. Geosci. Union*, 10:233–243.

Courtier, P., Andersson, E., Heckley, W., Pailleux, J., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F., and Fisher, M. (1998). The ECMWF implementa-

tion of three-dimensional variational assimilation (3D-Var). I: Formulation. *Q. J. R. Meteorol. Soc.*, 124(1994):1783–1807.

Cram, T. a., Compo, G. P., Yin, X., Allan, R. J., McColl, C., Vose, R. S., Whitaker, J. S., Matsui, N., Ashcroft, L., Auchmann, R., Bessemoulin, P., Brandsma, T., Brohan, P., Brunet, M., Comeaux, J., Crouthamel, R., Gleason, B. E., Groisman, P. Y., Hersbach, H., Jones, P. D., Jónsson, T., Jourdain, S., Kelly, G., Knapp, K. R., Kruger, A., Kubota, H., Lentini, G., Lorrey, A., Lott, N., Lubker, S. J., Luterbacher, J., Marshall, G. J., Maugeri, M., Mock, C. J., Mok, H. Y., Nordli, Ø., Rodwell, M. J., Ross, T. F., Schuster, D., Srnec, L., Valente, M. A., Vizi, Z., Wang, X. L., Westcott, N., Woollen, J. S., and Worley, S. J. (2015). The International Surface Pressure Databank version 2. *Geosci. Data J.*, 2(1):31–46.

Dahlgren, P., Landelius, T., and K, P. (2016). A high-resolution regional reanalysis for Europe. Part 1: Three-dimensional reanalysis with the regional High -Resolution Limited-Area Model (HIRLAM). *Q. J. R. Meteorol. Soc.*, 142(July):2119–2131.

Davies, H. C. (1983). Limitations of some common lateral boundary schemes used in regional NWP models. *Mon. Weather Rev.*, 111:1002–1012.

Davini, P., Cagnazzo, C., Gualdi, S., and Navarra, A. (2012). Bidimensional Diagnostics, Variability, and Trends of Northern Hemisphere Blocking. *J. Clim.*, 25:6496–6509.

de Haan, S. (2011). HighâĂŘresolution wind and temperature observations from aircraft tracked by Mode S air traffic control radar. *J. Geophys. Res.*, 116(May):1–13.

Dee, D. (2005). Bias and data assimilation. *Q. J. R. Meteorol. Soc.*, 131(131):3323–3343.

Dee, D., Källén, E., and Simmons, A. (2011a). Comments on âĂIJReanalyses Suitable for Characterizing Long-Term TrendsâĂİ. *Bull. Am. Meteorol. Soc.*, 92(1):65–70.

Dee, D. and Uppala, S. (2008). Variational bias correction in ERA-Interim. Technical report, ECMWF, Reading.

Dee, D. P., Balmaseda, M., Balsamo, G., Engelen, R., Simmons, A. J., and Thépaut, J. N. (2014). Toward a consistent reanalysis of the climate system. *Bull. Am. Meteorol. Soc.*, 95(8):1235–1248.

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., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J.,

Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, L., Kallberg, P., Köhler, M., Matricardi, M., Mcnally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N., and Vitart, F. (2011b). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.*, 137(656):553–597.

Deng, A. J., Seaman, N. L., Hunter, G. K., and Stauffer, D. R. (2004). Evaluation of interregional transport using the MM5-SCIPUFF system. *J. Appl. Meteorol.*, 43(12):1864–1886.

Doms, G. and Baldauf, M. (2015). A Description of the Nonhydrostatic Regional COSMO-Model Part I : Dynamics and Numerics. Technical report, Deutscher Wetterdienst, Offenbach am Main.

Doms, G., Förstner, J., Heise, E., Herzog, H.-J., Mironov, D., Raschendorfer, M., Reinhardt, T., Ritter, B., Schroding, R., Schulz, J.-P., and Vogel, G. (2011). A Description of the Nonhydrostatic Regional COSMO Model Part II : Physical Parameterization. Technical report, Deutscher Wetterdienst, Offenbach am Main.

Donaldson, R., Dyer, R., and Kraus, M. (1975). Objective evaluator of techniques for predicting severe weather events. *Bull. Am. Meteorol. Soc.*, 56(7):755–755.

Ehrendorfer, M. (1997). Predicting the uncertainty of numerical weather forecasts: a review. *Meteorol. Zeitschrift*, 183(August):147–183.

Epstein, E. (1969). Stochastic dynamic prediction. *Tellus*, XXI:739–759.

Gal-Chen, T. and Somerville, R. C. J. (1975). On the Use of a Coordinate Transformation for the Solution of the Navier-Stokes Equations. *J. Comput. Phys.*, 228:209–228.

Gibson, J., Kallberg, P., and Uppala, S. (1997). ERA Description. Technical report, ECMWF Re-analysis Final Report Series, Reading.

Giorgi, F., Jones, C., and Asrar, G. R. (2009). Addressing climate information needs at the regional level: the CORDEX framework. *WMO Bull.*, 58(July):175–183.

Gneiting, T. and Raftery, A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation Strictly Proper Scoring Rules, Prediction, and Estimation. *J. Am. Stat. Assoc.*, 1459(January):359–378.

Haimberger, L. (2007). Homogenization of Radiosonde Temperature Time Series Using Innovation Statistics. *J. Clim.*, 20:1377–1403.

Hamill, T. (2001). Interpretation of Rank Histograms for Verifying Ensemble Forecasts. *Mon. Weather Rev.*, 129:550–560.

Hamill, T. M. and Colucci, S. J. (1997). Verification of EtaâĂŞRSM Short-Range Ensemble Forecasts. *Mon. Weather Rev.*, 125(6):1312–1327.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M. (2008). A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006. *J. Geophys. Res. Atmos.*, 113(20).

Hersbach, H. (2000). Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Weather Forecast.*, 15:559–570.

Hersbach, H. and Dee, D. (2016). ECMWF Newsletter. Technical Report Spring 2015, ECMWF, Reading, UK.

Hersbach, H., Peubey, C., Simmons, A., Berrisford, P., Poli, P., and Dee, D. (2015). ERA-20CM: A twentieth-century atmospheric model ensemble. *Q. J. R. Meteorol. Soc.*, 141(691):2350–2375.

Hess, R. (2001). Assimilation of screen-level observations by variational soil moisture analysis. *Meteorol. Atmos. Phys.*, 77(1-4):145–154.

Hodges, K. I., Lee, R. W., and Bengtsson, L. (2011). A Comparison of Extratropical Cyclones in Recent Reanalyses ERA-Interim, NASA MERRA, NCEP CFSR, and JRA-25. *J. Clim.*, 24:4888–4906.

Houtekamer, P. and Mitchell, H. (1998). Data Assimilation Using an Ensemble Kalman Filter Technique. *Mon. Weather Rev.*, 126(1997):796–811.

Houtekamer, P. L., Lefaivre, L., Derome, J., Ritchie, H., and Mitchell, H. L. (1996). A system simulation approach to ensemble prediction. *Mon. Weather Rev.*, 124:1225–1242.

Jermey, P. M. and Renshaw, R. J. (2016). Precipitation representation over a two year period in regional reanalysis. *Q. J. R. Meteorol. Soc.*, 142:1300–1310.

Jones, P. D., Lister, D. H., Osborn, T. J., Harpham, C., Salmon, M., and Morice, C. P. (2012). Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. *J. Geophys. Res. Atmos.*, 117(5).

Jung, T., Palmer, T. N., and Shutts, G. J. (2005). Influence of a stochastic parameterization on the frequency of occurrence of North Pacific weather regimes in the ECMWF model. *Geophys. Res. Lett.*, 32:1–4.

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

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S. K., Hnilo, J. J., Fiorino, M., and Potter, G. L. (2002). NCEP-DOE AMIP-II reanalysis (R-2). *Bull. Am. Meteorol. Soc.*, 83(11):1631–1643.

Kaspar, F., Müller-Westermeier, G., Penda, E., Mächel, H., Zimmermann, K., Kaiser-Weiss, A., and Deutschländer, T. (2013). Monitoring of climate change in Germany - data, products and services of Germany's National Climate Data Centre. *Adv. Sci. Res.*, 10:99–106.

Keller, J. D. and Hense, A. (2011). A new non-Gaussian evaluation method for ensemble forecasts based on analysis rank histograms. *Meteorol. Zeitschrift*, 20(2):107–117.

Keune, J., Ohlwein, C., and Hense, A. (2014). Multivariate Probabilistic Analysis and Predictability of Medium-Range Ensemble Weather Forecasts. *Mon. Weather Rev.*, 142:4074–4090.

Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., Van Den Dool, H., Jenne, R., and Fiorino, M. (2001). The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bull. Am. Meteorol. Soc.*, 82(2):247–267.

Klemp, J. and Wilhelmson, R. (1978). The simulation of three-dimensional convective storm dynamics. *J. Atmos. Sci.*, 35:1070–1096.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K. (2015). The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *J. Meteorol. Soc. Japan. Ser. II*, 93(1):5–48.

Lahoz, W., Khattatov, B., and Menard, R., editors (2010). *Data assimilation. Making sense of observations*. Springer-Verlag Berlin Heidelberg, 1 edition.

Laloyaux, P. (2016). Status, plans and first results of CERA-20C.

Laloyaux, P., Balmaseda, M., Dee, D., Mogensen, K., and Janssen, P. (2016). A coupled data assimilation system for climate reanalysis. *Q. J. R. Meteorol. Soc.*, 142(694):65–78.

Landelius, T., Dahlgren, P., Gollvik, S., Jansson, A., and Olsson, E. (2016). A high resolution regional reanalysis for Europe. Part 2: 2D analysis of surface temperature, precipitation and wind. *Q. J. R. Meteorol. Soc.*, 142(698):2132–2142.

Lei, L., Stauffer, D., Haupt, S., and Young, G. (2012a). A hybrid nudging-ensemble Kalman filter approach to data assimilation . Part I : application in the Lorenz system. *Tellus*, 1(0):1–14.

Lei, L., Stauffer, D. R., and Deng, A. (2012b). A hybrid nudging-ensemble Kalman filter approach to data assimilation in WRF/DART. *Q. J. R. Meteorol. Soc.*, 138(669):2066–2078.

Leidner, S. M., Stauffer, D. R., and Seaman, N. L. (2001). Improving Short-Term Numerical Weather Prediction in the California Coastal Zone by Dynamic Initialization of the Marine Boundary Layer. *Mon. Weather Rev.*, 129(2):275–294.

Li, X., Charron, M., Spacek, L., and Candille, G. (2008). A Regional Ensemble Prediction System Based on Moist Targeted Singular Vectors. *Mon. Weather Rev.*, 136:443–462.

Lorenz, E. (1963). Deterministic nonperiodic flow. *J. Atmos. Sci.*, 20:130–141.

Lorenz, E. (1969). The predictability of a flow which possesses many scales of motion. *Tellus*, 21(3):289–307.

Lorenz, E. (1982). Atmospheric predictability experiments with a large numerical model. *Tellus*, 21:739–759.

Lorenz, E. (1990). Effects of analysis and model errors on routine weather forecasts. In *Proc. 1989 ECMWF Semin. ten years Mediu. Weather Forecast.*, pages 115–128.

Lorenz, E. (1993). *The essence of chaos*. University of Washington Press.

Louis, J.-F. (1979). A parametric model of vertical eddy fluxes in the atmosphere. Technical report, European Center for Medium-Range Weather Forecasting, Reading, UK.

Marshall, G. J. (2003). Trends in the Southern Annular Mode from Observations and Reanalyses. *J. Clim.*, 16(1999):4134–4143.

Mason, I. (1982). A model for assessment of weather forecasts. *Aust. Meteorlogical Mag.*, 30:291–303.

Maurer, E. P., Donnell, G. M. O., Lettenmaier, D. P., and Roads, J. O. (2001). Evaluation of the land surface water budget in NCEP / NCAR and NCEP / DOE reanalyses using an off-line hydrologic model. *J. Geophys. Res.*, 106(2000).

Mellor, G. L. and Yamada, T. (1974). A hierarchy of turbulence closure models for planetary boundary layers. *J. Atmos. Sci.*, 31:1791–1806.

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., Jović, 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., and Shi, W. (2006). North American regional reanalysis. *Bull. Am. Meteorol. Soc.*, 87(3):343–360.

Mironov, D. and Ritter, B. (2004). Testing the new ice model for the global NWP system GME of the German Weather Service. Technical report, Deutscher Wetterdienst, Offenbach am Main.

Mironov, D. V. (2008). Parameterization of Lakes in Numerical Weather Prediction. Description of a Lake Model. Technical report, Deutscher Wetterdienst, Offenbach am Main.

Mitas, C. M. and Clement, A. (2006). Recent behavior of the Hadley cell and tropical thermodynamics in climate models and reanalyses. *Geophys. Res. Lett.*, 33:1–4.

Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *J. Geophys. Res. Atmos.*, 117(8):1–22.

Murphy, A. (1973a). A new vector partition of the probability score. *J. Appl. Meteorol.*, 12:595–600.

Murphy, A. (1973b). Hedging and skill scores for probability forecasts. *J. Appl. Meteorol.*, 12:215–223.

Murphy, A. and Winkler, R. (1987). A general framework for forecast verification. *Mon. Weather Rev.*, 115:1330–1338.

Newson, R. (1998). Results of the WCRP First International Conference on Reanalysis. *GEWEX News*, 8:3–4.

Onogi, K., Tsutsui, J., Koide, H., Sakamoto, M., Kobayashi, S., Hatsushika, H., Matsumoto, T., Yamazaki, N., Kamahori, H., Takahashi, K., Kadokura, S., Wada, K., Kato, K., Oyama, R., Ose, T., Mannoji, N., and Taira, R. (2007). The JRA-25 Reanalysis. *J. Meteorol. Soc. Japan*, 85(3):369–432.

Otte, T. L., Nolte, C. G., Otte, M. J., and Bowden, J. H. (2012). Does nudging squelch the extremes in regional climate modeling? *J. Clim.*, 25(20):7046–7066.

Palmer, T. N., Barkmeijer, J., Buizza, R., Petroliagis, T., and Park, S. (1997). The ECMWF Ensemble Prediction System. *Meteorol. Appl.*, 4(4):301–304.

Palmer, T. N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G., Steinheimer, M., and Weisheimer, A. (2004). Stochastic parameterization and model uncertainty. Technical report, ECMWF, Reading, UK.

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., Laloyaux, P., Tan, D. G. H., Peubey, C., Thépaut, J.-N., Trémolet, Y., Hólm, E. V., Bonavita, M., Isaksen, L., and Fisher, M. (2016). ERA-20C: An atmospheric reanalysis of the 20th century. *J. Clim.*, 29:4083–4097.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M., Schubert, S. D., Takacs, L., Kim, G. K., Bloom, S., Chen, J., Collins, D., Conaty, A., Da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C., Reichle, R., Robertson, F. R., Ruddick, A., Sienkiewicz, M., and Woollen, J. (2011). MERRA: NASA's modern-era retrospective analysis for research and applications. *J. Clim.*, 24(14):3624–3648.

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

Röpnack, A., Hense, A., Gebhardt, C., Majewski, D., and Wetterdienst, D. (2013). Bayesian Model Verification of NWP Ensemble Forecasts. *Mon. Weather Rev.*, 141:375–387.

Rosnay, P. D., Balsamo, G., Sabater, J. M., and Dutra, E. (2016). High resolution land reanalysis.

Saetra, O., Hersbach, H., Bidlot, J., and Richardson, D. (2004). Effects of Observation Errors on the Statistics for Ensemble Spread and Reliability. *Mon. Weather Rev.*, 132:1487–1501.

Saha, S., Moorthi, S., Pan, H. L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y. T., Chuang, H. Y., Juang, H. M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., Van Den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J. K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C. Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R. W., Rutledge, G., and Goldberg, M. (2010). The NCEP climate forecast system reanalysis. *Bull. Am. Meteorol. Soc.*, 91(8):1015–1057.

Schättler, U., Doms, G., and Schraff, C. (2011). A description of the nonhydrostatic regional COSMO-model. Technical report, Deutscher Wetterdienst, Offenbach.

Schraff, C. and Hess, R. (2012). Consortium for Small-Scale Modelling A Description of the Nonhydrostatic Regional COSMO-Model Part III : Data Assimilation. Technical report, Deutscher Wetterdienst, Offenbach.

Schraff, C., Reich, H., Rhodin, A., Schomburg, A., Stephan, K., Perianez, A., and Potthast, R. (2016). Kilometre-scale ensemble data assimilation for the COSMO model (KENDA). *Q. J. R. Meteorol. Soc.*, 142(April):1453–1472.

Schraff, C. H. (1996). *Data assimilation and mesoscale weather prediction: A study with a forecast model for the alpine region.* PhD thesis, Swiss Federal Institute of Technology.

Schraff, C. H. (1997). Mesoscale data assimilation and prediction of low stratus in the Alpine region. *Meteorol. Atmos. Phys.*, 64(1-2):21–50.

Schrodin, R. and Heise, E. (2001). The multi-layer version of the DWD soil model TERRA_LM. Technical report, Deutscher Wetterdienst, Offenbach am Main.

Schroeder, A. J., Stauffer, D. R., Seaman, N. L., Deng, A., Gibbs, A. M., Hunter, G. K., and Young, G. S. (2006). An Automated High-Resolution, Rapidly Relocatable Meteorological Nowcasting and Prediction System. *Mon. Weather Rev.*, 134:1237–1265.

Seaman, N. L., Stauffer, D. R., and Lario-Gibbs, A. M. (1995). A multi-scale four-dimensional data assimilation system applied in the San Joaquin Valley during SARMAP. Part 1: Modeling design and basic performance characteristics. *J. Appl. Meteorol.*, 34:1739—-1761.

Simmer, C., Adrian, G., Jones, S., Wirth, V., Göber, M., Hohenegger, C., Janjic, T., Keller, J., Ohlwein, C., Seifert, A., Trömel, S., Ulbrich, T., Wapler, K., Weissmann, M., Keller, J., Masbou, M., Meilinger, S., Riß, N., Schomburg, A., Vormann, A., and Weingärtner, C. (2016). HErZ. The German Hans-Ertel Centre for Weather Research. *Bull. Am. Meteorol. Soc.*, 97(June):1057–1068.

Simmons, A. (2016). Reanalysis : the global context.

Skamarock, W. C. (2004). Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra. *Mon. Weather Rev.*, 132:3019–3032.

Soares, P. M. M. and Cardoso, R. M. (2012). WRF high resolution dynamical downscaling of ERA-Interim for Portugal. *Clim. Dyn.*, 8:2497–2522.

Soci, C. and Bazile, E. (2013). EURO4M project report D2.5 on the new MESANSAFRAN downscaling system. Technical Report April, Meteo France / UERRA, Toulouse.

Stauffer, D. and Seaman, N. (1991). Use of four-dimensional data assimilation in a limited-area mesoscale model Part II: Effects of data assimilation within the planetary boundary layer. *Mon. Weather Rev.*, 119:734–754.

Stauffer, D. R. and Seaman, N. L. (1990). Use of Four-Dimensional data assimilation in a limited-area mesoscale model. Part 1: Experiments with synoptic-scale data. *Mon. Weather Rev.*, 110:1250–1277.

Stensrud, D., Bao, J.-W., and Warner, T. (2000). Using Initial Condition and Model Physics Perturbations in Short-Range Ensemble Simulations of Mesoscale Convective Systems. *Mon. Weather Rev.*, 128:2077–2107.

Stephenson, D. B. (2000). Use of the âĂIJOdds RatioâĂİ for Diagnosing Forecast Skill. *Weather Forecast.*, 15(2):221–232.

Talagrand, O. and Courtier, P. (1987). Variational assimilation of meteorological observations with the adjoint vorticity equation . I : Theory. *Q. J. R. Meteorol. Soc.*, 113:1311–1328.

Thorne, P. and Vose, R. (2010). Reanalyses suitable for characterizing long-term trends. Are they really achievable? *Bull. Am. Meteorol. Soc.*, 91(3):353–361.

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

Torrisi, L. (2012). Implementation of stochastic physics in COSMO : recent tests.

Toth, T. and Kalnay, E. (1997). Ensemble Forecasting at NCEP and the Breeding Method. *Mon. Weather Rev.*, 125:3297–3319.

Trenberth, K., Fasullo, J., and Jeffrey, K. (2009). Earth's global energy budget. *Bull. Am. Meteorol. Soc.*, 90:311–324.

Trenberth, K. and Olson, J. (1988). An evaluation and intercomparison of global analyses from the National Meteorological Center and the European Centre for Medium Range Weather Forecasts. *Bull. Am. Meteorol. Soc.*, 69(9):1047–1057.

Trenberth, K. E. Smith, L. (2008). Atmospheric energy budgets in the Japanese Reanalysis: Evaluation and variability. *J. Meteorol. Soc. Japan*, 86:579–592.

Uppala, S. M., KâĎńllberg, 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., and Woollen, J. (2005). The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.*, 131(612):2961–3012.

Vidal, J. P., Martin, E., Franchistéguy, L., Baillon, M., and Soubeyroux, J. M. (2010). A 50-year high-resolution atmospheric reanalysis over France with the Safran system. *Int. J. Climatol.*, 30(11):1627–1644.

von Storch, H. and Zwiers, F. (1999). *Statistical Analysis in Climate Research*. Cambridge University Press, 2003 edition.

Wahl, S., Bollmeyer, C., Crewell, S., Figura, C., Friederichs, P., Hense, A., Keller, J. D., and Ohlwein, C. (2017). A novel convective-scale regional reanalyses COSMO-REA2 : Improving the representation of precipitation. *Meteorol. Zeitschrift*.

WCRP (1997). Proceedings of the first WCRP International Conference on Reanalyses.

WCRP (2000). Proceedings of the Second WCRP International Conference on Reanalyses.

Whelan, E. and Gleeson, E. (2016). A Mesoscale Regional Reanalysis for Ireland. Technical report, Met Eireann.

Wicker, L. J. and Skamarock, W. C. (2002). Time-Splitting Methods for Elastic Models Using Forward Time Schemes. *Mon. Weather Rev.*, 130(8):2088–2097.

Wilks, D. (1995). *Statistical methods in the atmospheric sciences*. Elsevier Academic Press, Burlington, USA.

Wright, J. S., Fueglistaler, S., and Sciences, O. (2013). Large differences in reanalyses of diabatic heating in the tropical upper troposphere and lower stratosphere. *Atmos. Chem. Phys.*, 13:9565–9576.

Zängl, G., Reinert, D., Pilar, R., and Baldauf, M. (2015). The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Q. J. R. Meteorol. Soc.*, 141(January):563–579.

Appendix

Reanalysis output

Table 1: Output variables and grib parameter id

Output field	**GRIB2 WMO**	**GRIB1 COSMO**
Model level parameters		
Cloud cover	260257	29.201
Pressure	54	1.2
Specific cloud liquid water content	246	31.201
Specific cloud ice water content	247	33.201
Specific humidity	133	51.2
Temperature	130	11.2
Zonal velocity	131	33.2
Vertical velocity	132	34.2
Pressure level parameters		
Cloud cover	260257	29.201
Geopotential height	156	6.2
Specific cloud liquid water content	246	31.201
Specific cloud ice water content	247	33.202
Relative humidity	157	52.2
Temperature	130	11.2
Zonal velocity	131	33.2
Vertical velocity	132	34.2
Height level parameters		
Cloud cover	260257	29.201
Pressure	54	1.2
Specific cloud liquid water content	246	31.201
Specific cloud ice water content	247	33.202
Relative humidity	157	52.2
Temperature	130	11.2
Wind speed	10	post-processed
Wind direction	3031	post-processed

Table 2: Output variables and grib parameter id

Output field	**GRIB2 WMO**	**GRIB1 COSMO**
Surface level parameters		
Precipitation and humidity		
Percolation	260430	90.2
Surface air relative humidity	260242	52.2
Surface runoff	174008	90.2
Total column water vapour	137	
Total precipitation	228228	61.2
Surface level parameters		
Accumulated radiation fluxes		
Albedo	260509	84.2
Evaporation	260259	57.2
Time-integrated surface latent heat flux	147	121.2
Time-integrated surface sensible heat flux	146	122.2
Time-integrated surface direct solar radiation	260264	22.01
Time-integrated surface net solar radiation	176	111.2
Time-integrated surface solar radiation downwards	169	post-processed
Time-integrated surface net thermal radiation	177	112.2
Time-integrated suface thermal radiation downwards	175	25.201
Temperature and wind speed		
10 m wind speed	207	32.2
10 m wind direction	260260	31.2
10 m wind gust speed	49	187.201
Surface air maximum temperature	201	15.2
Surface air minimum temperature	202	16.2
Surface air temperature	167	11.2
Skin temperature	235	11.2
Pressure		
Mean sea level pressure	151	2.2
Surface pressure	134	1.2
Cloud properties		
High cloud cover	3075	75.2
Low cloud cover	3073	73.2
Medium cloud cover	3074	74.2
Total cloud cover	228164	71.2
Snow		
Snow density	33	133.201
Snow depth	3066	66.2
Snow depth water equivalent	228141	
Snow fall water equivalent	228144	123.201
Soil		
Liquid non-frozen volumetic soil moisture	260210	220.201
Volumetric soil water	260199	198.201
Soil heat flux	260364	16.201
Soil temperature	260360	197.201

Levels, reforecast horizon and output frequencies

Table 3: Output levels of different level types

Model levels	**Pressure levels [hPa]**	**Height levels [m]**
0,..,40	100,150,200,250, 300,400,500,600, 700,750,800,825, 850,875,900,925, 950,975,1000	15,30,50,75, 100,150,200, 250,300,400, 500

Table 4: Reforecast horizon

00 UTC	**06 UTC**	**12 UTC**	**18 UTC**
+ 30h	+ 06h	+ 30h	+06h

Table 5: Output frequencies

Model levels	**Pressure levels**	**Height levels**	**Surface**
Reanalysis			
6h	1h	1h	1h
Reforecasts			
1h (+01,...,+06)	1h (+01,...,+06)	1h (+01,...,+06)	1h (+01,...,+06)
3h (+09,...,+30)	3h (+09,...,+30)	3h (+09,...,+30)	3h (+09,...,+30)

List of Figures

List of Tables

Acknowledgement

Many wonderful and inspiring people have accompanied me, supported me and enriched my life during the last three years.
I kindly thank my reviewers Andreas Hense, Silke Trömel, Roland Potthast and Peter Vöhringer. This work would not have been possible without the funding by the European Union's Seventh Framework Programme under grant agreement no. 607193.
I warmly thank my doctoral father Professor Andreas Hense. Your broad knowledge and your way of sharing any of your scientific ideas with your students, but likewise your great trust in us and our ideas and your relaxed attitude have made it great learning from you in three theses for so many years!
I'm deeply grateful to Christoph Schraff. Christoph, thank you for all your time and advise and the humor you've had with the PhD student having adopted you. Thank you for everything I've learned from you, for our lively debates, for the real help, for many ideas, for that you see even the little things, for having shown me how to get to the bottom of scientific problems and for being a great example in many respects.
Next I thank Uli Schättler and Dörte Liermann. Uli, thank you for your great help with improving and implementing the reanalysis system, for answering all of my technical questions and the two of you, thank you very much for all the work you have invested in the post-processing and archiving. Another big thank you is for Martin Lange, for your help with improving the external analyses. During my PhD time, DWD has become my second home and I would like to thank many more colleagues: Roland Potthast, for the general help and your ideas for my thesis; Alex Cress for the ICON ensemble; Klaus Stephan for your help during the first winter of my PhD thesis when we tested ensemble nudging with latent heat nudging; Hendrik Reich and Andreas Rhodin for the technical help during the days we tried to couple KENDA with ensemble nudging; Christoph Gebhardt and Susanne Theis for your ideas on reanalysis ensembles; Felix Fundel for some friendly hints on what to refrain from in verification and Uli Blahak for the infecting enthusiasm by which you answered all my int2lm and physics and tropical set up questions. Further, thanks a million to Frank Kaspar for the support in UERRA and the afternoons we have spent with Christoph discussing about applications of reanalysis data and how to extend the system reasonably for future reanalyses and

my dissertation. Another great thank you is for Andrea Kaiser-Weiss for your constant interest and our controversial discussions. I'm very grateful to Michael Borsche for having started to care for the reanalysis these days!
UERRA has given me the opportunity to meet many fantastic people from all over Europe. Comparatively, being hardly 24 years old initially, I have felt like a very young and unknowing member of the project. However, there has been Per Unden and I felt like Per asked me to lean in and that he trusted and encouraged me and involved me in discussions and thereby motivated me a lot to learn and ask questions and tame the reanalysis and get to know the real world outside university. Thank you very much, Per! Thanks to our work package leader, Richard Renshaw and also, thanks a lot to Peter Jermey! I thank Hans Hersbach and Christoph Frei for their helpful ideas on verification of probabilistic reanalyses. Also, thanks to all my UERRA colleagues for the nice international atmosphere and the great experience that my time in the project has been. Thanks to Richard Mladek for your patience and motivation related to chocolate concerning the archiving and to Carsten Maas for all your technical support with the ECMWF super computers and your friendly phone calls when something seemed suspicious. Finally, I thank Saki Uppala and Addie Simmons, our UERRA advisors, who know the whole story of reanalysis and whose advise and incredible love for the subject has infected and motivated me whenever I had the luck to meet them.
From my colleagues in Bonn I thank Chris Bollmeyer for having shared your reanalysis suite with me and for having answered all my questions on the phone when you had long left to Wetteronline and Clarissa Figura for your experience on the suite that helped me very much when I started working with it. Also, thanks to my colleagues, Benno Thoma, Mohammad, Chris Weijenborg for our lovely time in the office and for having invited me to sit in the men cave. Thank you for having endured my long and sometimes entertaining phone calls. Thank you very much to Thomas Burkhardt, who has taken care of me in MIUB since I had decided to be become a meteorologist at the age of 16 and all my other colleagues that have made the time in MIUB very enjoyable.
I thank Kerstin Hartung and Michael Ziegert and all my other friends for having turned on the sun again whenever I couldn't see the light at the end of the tunnel. I have a great family and I'm deeply grateful to my mum and my dad for their love and their real interest in everything we do and your support that we can always count on! The same thank you is for my dear sisters Elli and Karo and Thomas and Frank. I thank my uncle Ralf who has shown me C. elegans and raised my enthusiasm for science when I was a very young girl. My last thank you is for Rüdiger Hewer. It has been your backing and warm heart and your way of making me laugh and explaining me all the small and great questions of the world that has turned a good time of my life into a fantastic one.

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 1-63: siehe `http://www.meteo.uni-bonn.de/bibliothek/bma`

64-77: open access, verfügbar unter `https://uni-bn.de/kpSDaQffe1`

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

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

Heft 66: ***Thorsten Simon***: Statistical and Dynamical Downscaling of Numerical Climate Simulations: Enhancement and Evaluation for East Asia, 2014, 48 S. + VII. + Anhänge

Heft 67: ***Elham Rahmani***: The Effect of Climate Change on Wheat in Iran, 2014, [erschienen] 2015, 96 S. + XIII.

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, [erschienen] 2015, 168 S. + XIII.

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.

Heft 70: ***A S M Mostaquimur Rahman***: Influence of subsurface hydrodynamics on the lower atmosphere at the catchment scale, 2015, 98 S. + XVI.

Heft 71: ***Sabrina Wahl***: Uncertainty in mesoscale numerical weather prediction: probabilistic forecasting of precipitation, 2015, 108 S.

Heft 72: ***Markus Übel***: Simulation of mesoscale patterns and diurnal variations of atmospheric CO_2 mixing ratios with the model system TerrSysMP-CO_2, 2015, [erschienen] 2016, 158 S. + II

Heft 73: ***Christian Bernardus Maria Weijenborg***: Characteristics of Potential Vorticity anomalies associated with mesoscale extremes in the extratropical troposphere, 2015, [erschienen] 2016, 151 S. + XI

Heft 74: ***Muhammad Kaleem***: A sensitivity study of decadal climate prediction to aerosol variability using ECHAM6-HAM (GCM), 2016, 98 S. + XII

Heft 75: ***Theresa Bick***: 3D Radar reflectivity assimilation with an ensemble Kalman filter on the convective scale, 2016, [erschienen] 2017, 96 S. + IX

Heft 76: ***Zied Ben Bouallegue***: Verification and post-processing of ensemble weather forecasts for renewable energy applications, 2017, 119 S.

Heft 77: ***Julia Lutz***: Improvements and application of the STatistical Analogue Resampling Scheme STARS, 2016, [erschienen] 2017, 103 S.

Heft 78: ***Benno Michael Thoma***: Palaeoclimate Reconstruction in the Levant and on the Balkans, 2016, [erschienen] 2017, XVI, 266 S.

Heft 79: ***Ieda Pscheidt***: Generating high resolution precipitation conditional on rainfall observations and satellite data, 2017, V, 173 S.

Heft 80: ***Tanja Zerenner***: Atmospheric downscaling using multi-objective genetic programming, 2016, [erschienen] 2017, X, 191 S.

Heft 81: ***Sophie Stolzenberger***: On the probabilistic evaluation of decadal and paleoclimate model predictions, 2017, IV, 122 S.

Heft 82: ***Insa Thiele-Eich***: Flooding in Dhaka, Bangladesh, and the challenge of climate change, 2017 [erschienen] 2018, V, 158 S.

Heft 83: ***Liselotte Bach***: Towards a probabilistic regional reanalysis for Europe, 2017 [erschienen] 2018, VI, 114 S.

METEOROLOGISCHES INSTITUT
MATHEMATISCH NATURWISSENSCHAFTLICHE FAKULTÄT
UNIVERSITÄT BONN

Meteorologisches Institut
Universität Bonn