

Towards a probabilistic regional reanalysis system for Europe: evaluation of precipitation from experiments

By LISELOTTE BACH^{1*}, CHRISTOPH SCHRAFF², JAN D. KELLER^{2,3} and ANDREAS HENSE¹, ¹*Meteorological Institute of the University of Bonn, Auf dem Hügel 20, Bonn, Germany;* ²*Deutscher Wetterdienst, Offenbach, Germany;* ³*Hans-Ertel-Centre for Weather Research, Germany*

(Manuscript received 10 May 2016; in final form 16 October 2016)

ABSTRACT

A new development in the field of reanalyses is the incorporation of uncertainty estimation capabilities. We have developed a probabilistic regional reanalysis system for the CORDEX-EUR11 domain that 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 we denote 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, we assess the representation of precipitation in our system in a pilot study. Based on two experiments, each of which extends over one month, we conduct a preliminary comparison to the global reanalysis ERA-Interim, a dynamical downscaling of the latter and the high-resolution regional reanalysis COSMO-REA6. In a next step, we assess our reanalysis system's probabilistic capabilities versus the ECMWF-EPS in terms of six-hourly precipitation sums. The added value of our 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 Re-Analyses (UERRA).

Keywords: climate dynamics, essential climate variables, ensemble data assimilation, CORDEX, uncertainty estimation, precipitation

1. Introduction

An atmospheric reanalysis is a weather and climate data set that emerges from a reconstruction of the four-dimensional record of the past atmospheric phase space trajectory, striving for a high degree of spatio-temporal consistency and accuracy. The key ingredient of reanalysis systems is a data assimilation algorithm that feeds historical observations of the atmospheric state into a numerical weather prediction (NWP) model (Bengtsson and Shukla, 1988). Through the model equations, the information content of the observations is spread in space and time. Additionally, it is transferred from the observed to the non-observed

variables and transported to observation-sparse regions. This yields complete and physically consistent four-dimensional fields. Data assimilation is essential to keep the model close to the true atmospheric trajectory.

Even though data assimilation, observation quality control and NWP model are kept fix over the reanalysis time span, reanalysis time-series mostly contain non-climatic signals; that is, they do not provide *climate quality* (e.g. Bengtsson et al., 2004; Thorne and Vose, 2010). This is due to the spatio-temporal evolution of the global observing system that features strong improvements in measurement quality over time, changes in the spatial distribution of observing systems and an over-whelming temporal increase of the number of observations available for data assimilation (e.g. Compo et al., 2011). For that reason, to date most reanalyses do not claim to provide climate quality which would be essential for long-term trend climate monitoring (Thorne

*Corresponding author.
email: libach@uni-bonn.de

and Vose, 2010). Rather, by tendency most of the reanalysis data sets [except for ERA-20C described in Poli et al. (2016), CERA-20C described in Laloyaux et al. (2016) and 20-CR introduced in Compo et al. (2011), which try to achieve climate quality] offer the best analysis quality at any point of time.

Many applications of reanalysis data (e.g. near-real-time monitoring) would benefit from a high degree of accuracy at the local scales as impacts, extremes and adaptation are increasingly attracting notice (Thorne and Vose, 2010). Recently, this appraisal has put forward the idea of regional reanalyses additionally to the global ones [for global reanalyses see e.g. Kalnay et al., (1996); Onogi et al. (2007); Saha et al. (2010); Rienecker et al. (2011); Dee et al. (2011); Compo et al. (2011); Kobayashi et al. (2015); Poli et al. (2016)]. Regional reanalyses make use of limited-area models with high grid resolution and thus ‘zoom in’ to a certain geographical region. Through a better representation of soil–atmosphere interactions, orographic effects, land-use effects and land–ocean contrasts in the model and assimilation of observations on smaller length-scales, enhanced accuracy on the mesoscale can be achieved. Also, the spatial sampling of information is strongly increased compared to global reanalyses.

The pioneering regional reanalyses have been the North American Regional Reanalysis (Mesinger et al., 2006) and the Arctic System Reanalysis (Bromwich et al., 2016). In recent years, European activities are being conducted in the FP-7 funded projects EURO4M¹ (2010–2013) and Uncertainties in Ensembles of Regional Re-Analyses (UERRA)² (2014–2017). These aim at the preparation of regional reanalysis systems that will be operationally applicable by the Copernicus Climate Change Service³ that is currently built up by the European Commission and the European Centre for Medium Range Weather Forecasting (ECMWF). This service will provide consistent estimates of essential climate variables (Bojinski et al., 2014) to support policy and decision makers and different sectors of public. The regional reanalysis systems currently available for the European-Atlantic region include a Met Office system (Jermy and Renshaw, 2016) at 12-km grid spacing, a system of the Swedish Hydrological and Meteorological Institute at 22-km resolution (Dahlgren et al., 2016; Landelius et al., 2016) and a high-resolution system of the German Hans-Ertel Centre for Weather Research (Simmer et al., 2016) at a 6-km grid spacing (Bollmeyer et al., 2015).

Albeit regional reanalyses can be expected to be more accurate, there remain uncertainties. This motivates the additional production of probabilistic regional reanalyses to estimate uncertainties. Currently, a global ensemble reanalysis system for the reanalysis ERA5 is implemented at the ECMWF and regional ones at the Met Office and the Swedish Meteorological and Hydrological Institute (SMHI).

In the course of this work, we introduce a further new ensemble regional reanalysis system based on the NWP model COSMO of the CONsortium for Small scale MOdeling (Schättler et al., 2011). This system makes use of a new ensemble generation technique that we denote *ensemble nudging*.

It allows for estimating the uncertainty of a regional reanalysis arising from uncertainties in the assimilated observations. The purpose of this article is to give a first impression of the usefulness and added value of our regional ensemble reanalysis system that will be initially employed for the production of a 5-year-long test reanalysis in the framework of UERRA. It will be called COSMO-EN-REA12 and comprise 21 ensemble members. The next version of the system will be enhanced by further uncertainty estimation capabilities that will account additionally for uncertainties in the lateral boundary conditions as well as errors in the model formulation.

At this stage, we show the ability of our basic system for probabilistically representing precipitation as essential climate variable. Note that at this point of time only short experiments extending over two 1-month periods are available. These have been conducted to provide an indication of the system’s probabilistic capabilities and its potential added value over a coarser-resolution global reanalysis before a longer time span is produced. Further note that we do not strive for a system capable of producing local long-term climate-quality time-series yet. The text is organised as follows. In section 2, the reanalysis framework is introduced. In section 3, we show basic diagnostics of the ensemble. In section 4, we evaluate precipitation from experiments. We close by a summary.

2. The reanalysis framework

The reanalysis system to be introduced has been advanced from a deterministic nudging regional reanalysis system that employs the limited-area model COSMO (Bollmeyer et al., 2015).

2.1. The numerical weather prediction model

COSMO is an NWP model that is targeted at the representation of atmospheric meso- and meso- α -scale processes. To allow for direct comparison to other European reanalyses or downscaling experiments, the model is set up for the

¹See: euro4m.eu

²See: uerra.eu

³See: <http://www.ecmwf.int/en/about/what-we-do/copernicus/copernicus-climate-change-service>

CORDEX-EUR11 domain (Giorgi et al., 2009) at a horizontal grid spacing of 12 km on a rotated latitude-longitude grid. The domain is shown in Fig. 1 and its specifications are summarised in Table 1.

2.1.1. Dynamics and numerics. The dynamical core of the model 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 (Schättler et al., 2011) which are solved using a two time-level, third-order Runge-Kutta time-split scheme (Wicker and Skamarock, 2002). This is complemented by an explicit vertical advection scheme. For advection, a second-order finite-volume scheme with strangle-splitting is employed (Bott, 1989). Horizontally, COSMO utilises a staggered Arakawa-C grid (Arakawa and Lamb, 1981), while a terrain-following Gal-Chen hybrid vertical coordinate (Gal-Chen and Somerville, 1975) is used in the vertical. For details on the dynamics and numerics please be referred to Doms and Baldauf (2015).

2.1.2. Physical parameterisations. Processes acting on the subgrid scales are parameterised as follows: Cloud microphysics are parameterised by a bulk-water continuity model. 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. Subgrid convection is parameterised by the Tiedtke scheme (Tiedtke, 1989), and the vertical turbulent diffusion is computed based on a prognostic equation for turbulent kinetic energy utilising a level 2.5 closure scheme (Doms et al., 2011). Soil and vegetation are represented by the model TERRA (Schulz et al., 2016) which employs seven 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. For details on the parameterisation schemes, see Doms et al. (2011).

2.1.3. Lateral boundary conditions. The lateral boundaries and initial conditions are provided by the global ECMWF reanalysis ERA-Interim (Dee et al., 2011). We use 0.5° ERA-Interim analysis fields from 00 and 12 UTC, as well as +03-h, +06-h and +09-h reforecasts comprising the three-dimensional fields specific humidity, temperature,

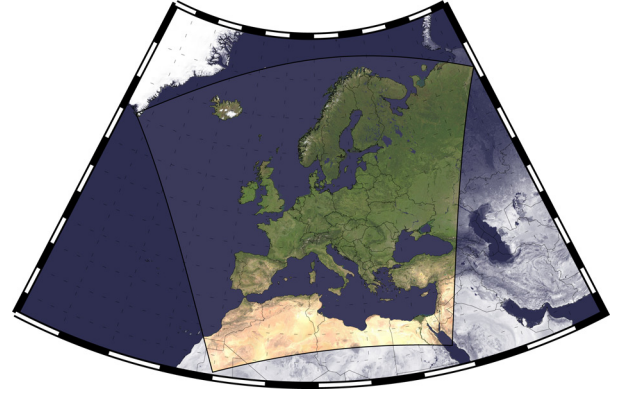


Fig. 1. CORDEX-EUR11 domain (colour-shaded area).

Table 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	22 700 m

pressure, wind, cloud liquid and ice water content. Additionally, the surface variables of 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. In our case, the relaxation zone near the boundaries consists of seven grid points which is about 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.

Note that our regional reanalysis system does not yet provide the opportunity to assimilate satellite data (see Section 2.2.1). However, through the lateral boundary conditions, it is provided with valuable synoptic-scale information from various satellite products (Dee et al., 2011) that force the mesoscale dynamics.

2.2. Analysis of atmospheric variables

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 at the German Weather Service (DWD) for many years 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; Leidner et al., 2001; Deng et al., 2004; Schroeder et al., 2006), research in the area of hybrid data assimilation methods (Lei et al., 2012a, 2012b), initialisation 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, 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 realised 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

$$\frac{\partial \Psi(x, t)}{\partial t} = F(\Psi, x, t) + G_y \cdot \sum_k W_k(x, t) \cdot [\Psi^k - \Psi(x_k, t)], \quad (1)$$

where F represents the model dynamics and physics 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 x and the model equivalent equivalents $\Psi(x_k, t)$; that is, 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 or Gaussian 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 increment and a geostrophic wind correction. For further details on the implementation of nudging in COSMO, please refer to Schraff (1997) or Schraff and Hess (2012).

2.2.1. Observation stream for atmospheric data assimilation. The observation stream that we assimilate consists basically of conventional observations. The obser-

vation types and corresponding variables that are assimilated are shown in Table 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). Moreover, wind profiler data are assimilated.

2.2.2. 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 $o' \sim N(0, \sigma_o)$ with zero mean and an estimated observation error standard deviation σ_o . 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 Lönnerberg, 1986). The employed observation error standard

Table 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

deviations are summarised in Tables 3 and 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 excludes satellite data (which are correlated between the channels), we assume normally distributed, unbiased, stationary in time as well as spatio-temporally uncorrelated observation errors. 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 ensemble members are perturbed and run in parallel independent streams that do not exchange information.

2.3. 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. This improves the representation of the turbulent fluxes. We apply different offline schemes to COSMO, namely an analysis of the snow depth, a sea surface temperature analysis and a variational soil moisture analysis. Note that each member is updated separately by these external analyses. Thus, the ensemble members do not exchange information so that they are fully independent realisations.

2.3.1. 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 minimisation problem.

Table 3. Observation error standard deviations for sea observing systems

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

U/V, wind components; ϕ , geopotential.

The cost function for the soil moisture η at any horizontal grid point is given by

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

where T is the screen-level temperature predicted for noon and T^o are observations of screen-level temperature and η and η^b are vectors whose dimension equals the number of analysed soil levels. They provide the moisture contents of the analysed and background soil layers. Both have a lower limit which is the air dryness point as well as an upper limit which is the pore volume. The observation error covariance matrix $\mathbf{R}$ is assumed to be stationary and diagonal, while the background error covariance matrix $\mathbf{B}$ is provided by an additional cycled Kalman filter analysis scheme. Further information can be found in Schraff and Hess (2012) and Hess (2001).

2.3.2. Analysis of sea surface temperature. Sea surface temperature strongly impacts the fluxes of sensible and latent heat over the ocean. Its spatial distribution influences the evolution of Rossby waves and cyclones. To capture not only the large-scale variations, but also relevant smaller-scale processes, we perform the sea surface temperature analysis (Schraff and Hess, 2012) once a day. 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 $0.17^\circ \times 0.1^\circ$ for the Baltic Sea, provided by the Federal Maritime and Hydrographic Agency of Germany (BSH), is used. A correction scheme is employed using the ERA-Interim sea surface temperature 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.

2.3.3. 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. We update the snow distribution and depth 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

Table 4. Observation error standard deviations for upper-air and surface systems

	U/V [m/s]	Φ [m]	T [K]	RH	T [K]	UV [m/s]	Φ [m]	U/V [m/s]
Lev [hPa]	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		

T/P/W = TEMP/PILOT/wind profiler and AI = AIREP and SY = SYNOP; U/V, wind components; ϕ , geopotential; T, temperature; RH, relative humidity.

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.

2.4. Process cycle

The individual components described above have been integrated into a system shown in Fig. 2. It 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

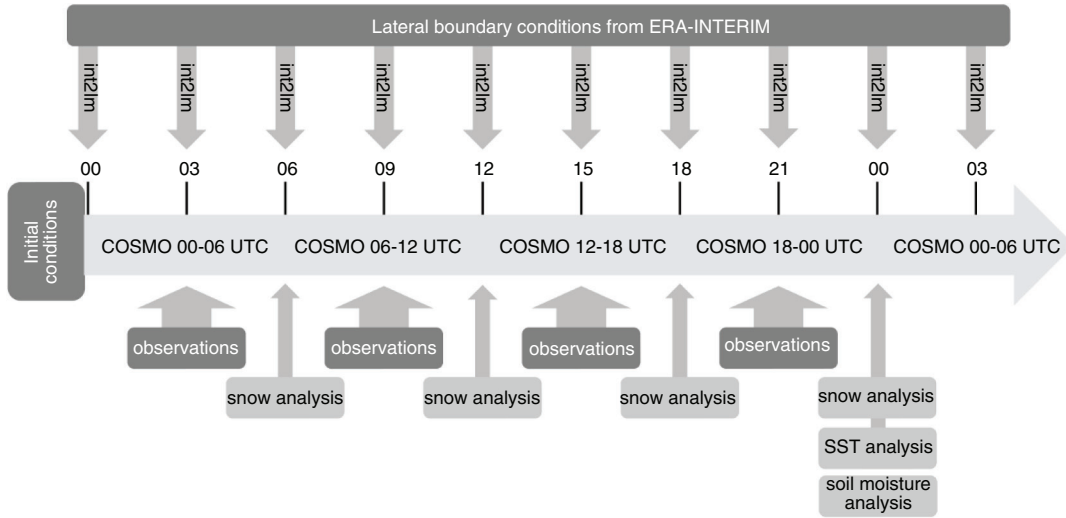


Fig. 2. Process cycle of the ensemble nudging reanalysis system for one exemplary ensemble member. The cycle is repeated for all members in parallel. All members are provided with the same lateral boundary conditions. The tool *int2lm* interpolates the fields from the steering model to the COSMO grid.

conducted for all ensemble members separately. 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 initialised in 6-hourly intervals and have twice a day lead times of 30 hours and twice a day of 6 hours. Feedback observation files summarise information on observations and observation-analysis departures in 6-hourly intervals.

3. Basic diagnostics

We start the evaluation by showing some basic diagnostics and properties of the regional ensemble reanalysis system. 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, we show the evolution of average spread in reforecasts.

The evaluation shown here is based on an experiment that spans June 2011. It comprises 20 ensemble members and one control run which assimilates the original un-

perturbed observations. Additionally, reforecasts with lead times of 30 hours, initialised at 00 UTC are evaluated.

3.1. 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 behaviour of analysis increments hints at biases in the NWP model or the observations or combinations. However, as aforementioned reanalyses of long periods are exposed to a strong increase in the density of observing systems. Major changes here can lead to systematic changes in the behaviour 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 (see introduction) and to detect technical problems in time. 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, our analysis increments represent aggregates of all changes that have been applied to the model state at a certain grid point over 6-hourly analysis cycles. Since we do not yet have a longer time span

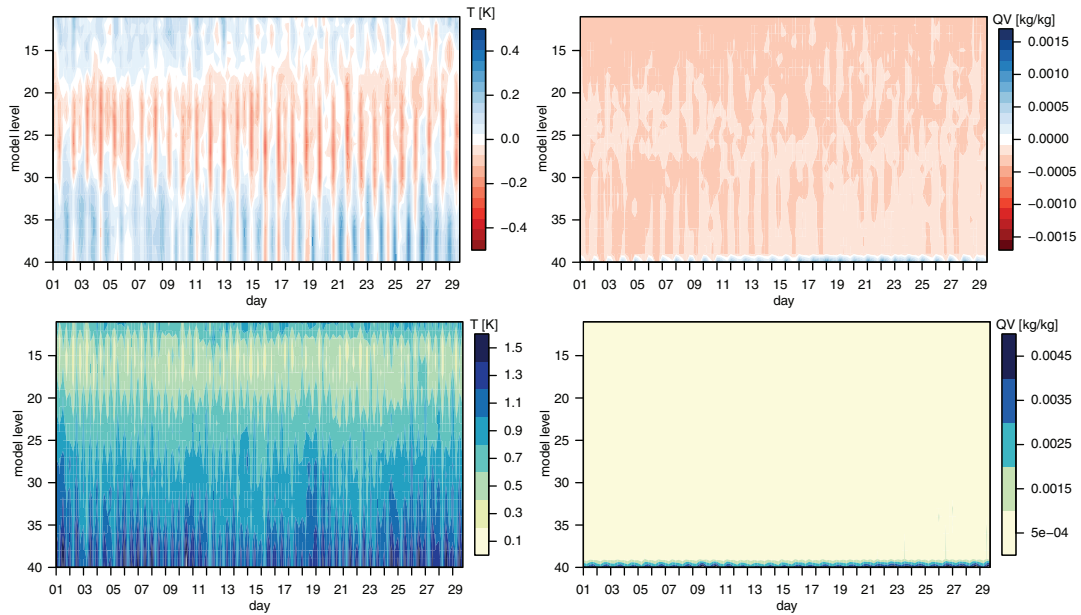


Fig. 3. Analysis increments for June 2011. The upper panels show spatially averaged analysis increments in dependency of model level and time of day for temperature (left) and specific humidity (right). The analysis increments are aggregated over 6-hourly analysis cycles from 00, 06, 12 and 18 UTC. Model levels 20 and 30 are located at about 550 hPa and 100 hPa above the surface, respectively. For temperature red colour indicates a too warm model and blue colour indicates a too cold model. For specific humidity red colour means that the model is too moist while blue colour means that it is too dry. The lower panels show the horizontal variability in terms of standard deviations of analysis increments for temperature (left) and specific humidity (right) in dependency of model level and time of day.

available, we show exemplarily a horizontally averaged time-sequence of analysis increments of temperature and specific humidity for an arbitrarily chosen member. The vertical structure in temperature displayed in Fig. 3 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 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.

3.2. Spin-up

Spin-up effects of precipitation occur during the first few hours after the initialisation of forecasts (Arpe, 1991; Betts and Ball, 1999). They become apparent as an under- or

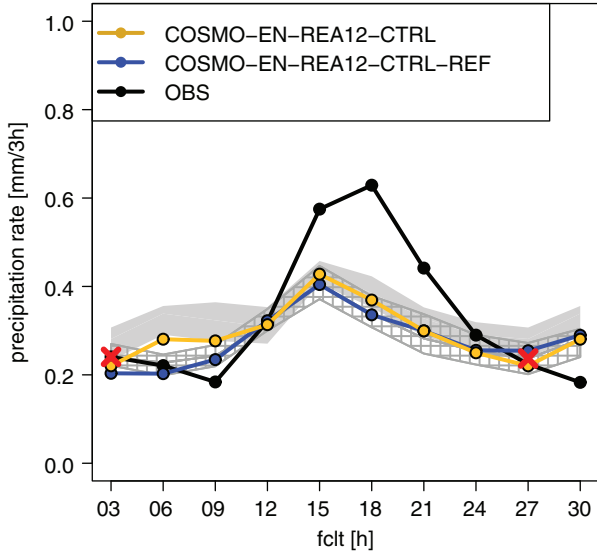


Fig. 4. Diurnal cycle of 3-hourly precipitation rates in ensemble of reforecasts of COSMO-EN-REA12, initialised at 00 UTC (shown as hatched area, 5 % to 95 % percentiles, control run blue points), reanalysis ensemble (shown as shaded area, 5 % to 95 % percentiles, control run yellow points) and corresponding rain gauge observations (black dots). Shown for German stations, see Fig. 8. The red crosses mark the median of the reforecast ensemble at 03 and 27 hours lead time.

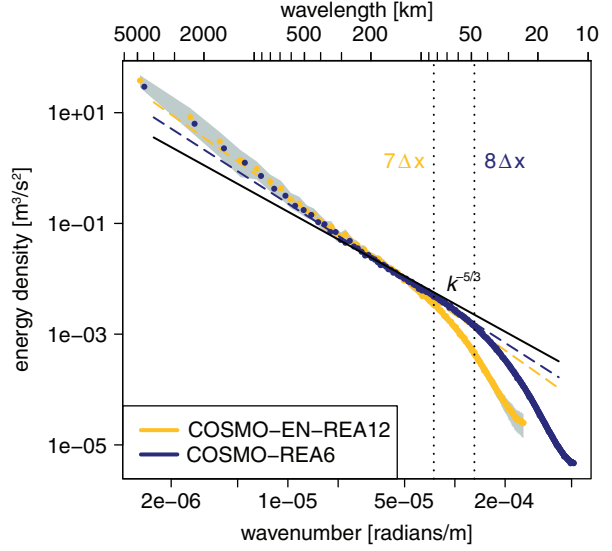


Fig. 5. Horizontal kinetic energy spectra for COSMO-EN-REA12 (yellow points) and COSMO-REA6 (blue points) in dependence of wavelength and wave number. The grey area shows the uncertainty estimated from the ensemble. The vertical lines mark the effective resolution of the reanalysis systems. The continuous 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 $k^{-5/3}$.

overestimation 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 model climate and physical balance between the variables, moistening the model by data assimilation often leads to an overestimation of precipitation in the first few hours of lead time or reversely to an underestimation 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 shows the diurnal cycle of 3-hourly precipitation rates in reforecasts as a function of forecast lead time. The employed reforecasts are initialised 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 3 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 initialised at 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. From 12 hours of lead time onwards, no significant difference can be identified between reanalysis and reforecast ensemble.

Figure 4 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 parameterised deep convection that diagnose the convective precipitation from the grid scale environment [for COSMO see Baldauf et al. (2011)].

3.3. 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 wavelength 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), we estimate the effective resolution as the wavelength at which the spectrum begins to fall steeply under the $k^{-5/3}$ slope. In Fig. 5, we compare the horizontal kinetic energy spectrum of our system averaged over the whole experimental time span to the one of COSMO-REA6 at 6-km grid spacing (Bollmeyer et al., 2015), which has provided the basis for the development of our 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, we fit linear trends to both spectra. 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. We find an effective resolution of 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, we have compared the horizontal kinetic energy spectra of analysis data from COSMO-EN-REA12 to the spectra of reforecasts initialised at 00 UTC and valid at 3, 10, 24 and 30 hours lead time. All agree with each other (therefore

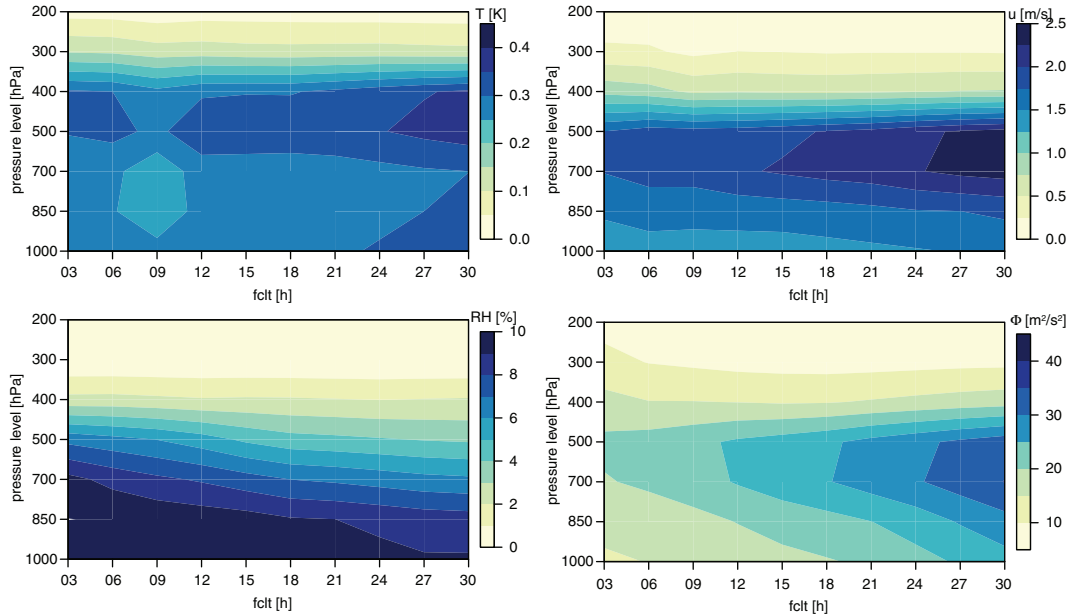


Fig. 6. Monthly mean evolution of horizontally averaged spread (over German subdomain) as a function of pressure levels and forecast lead time. Data are from reforecasts initialised at 00 UTC for experiment for June 2011. Upper panels are for temperature (left) and zonal wind (right), lower panels for relative humidity (left) and geopotential (right).

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.

3.4. Uncertainty

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, we provide qualitative insight into the behaviour of ensemble spread over a subdomain of Germany. Figure 6 shows the mean temporal evolution of horizontally averaged spread in reforecasts, initialised 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 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. Evaluation of precipitation

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. We have chosen to approach the evaluation of our system by means of precipitation in the first

instance. 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 as a start (personal communication Dale Barker, Met Office; 2nd UERRA General Meeting, Tortosa, Spain).

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 Fig. 7). As anticipated, the regional reanalysis is capable of representing 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. We will investigate this in the next sections. Our system further allows for the estimation of the uncertainties in such precipitation climatologies. The spread of the monthly integrated precipitation resulting from our experiment is shown in Fig. 7. The estimated uncertainty has a pronounced spatial variability and a large maximum value of 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, for example, 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)].

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 compared to global reanalyses and dynamical downscalings from global reanalyses?
- (2) What are the probabilistic capabilities of the ensemble?

The experiments we evaluate comprise 20 ensemble members and one nature or control run that assimilates the

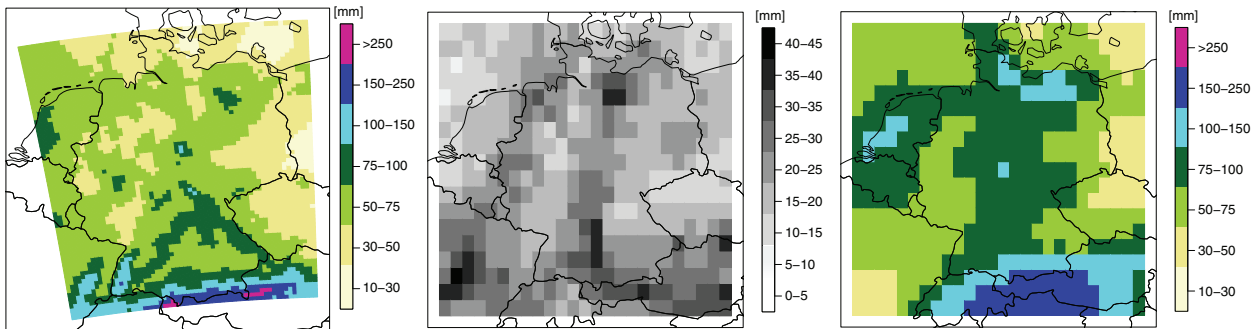


Fig. 7. Monthly precipitation climatologies for Germany for June 2011 based on ensemble mean of COSMO-EN-REA12 (left) and ERA-Interim (right). In the middle, the uncertainty (spread) of the monthly integrated precipitation estimated by COSMO-EN-REA12 is shown.

original observations and extend over the months of June 2011 and additionally December 2011. We refer to them as *summer* and *winter experiments* 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 3.2. Therefore, our reanalysed precipitation fields can directly be used so that we show an evaluation of them 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, we lack knowledge about the distribution of observation errors of rain gauges and assume that they are negligible compared to the errors of the model-based analysis.

The goal we pursue in the first part of the evaluation is to compare our 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. We are aware that this approach is not common practice. However, for our purpose we consider it the appropriate choice, because users will most probably use the highly-resolved data rather than interpolate it to a coarse grid.

In the second part, we test the probabilistic capabilities of COSMO-EN-REA12. Due to the lack of a global reanalysis ensemble at the time of writing, we compare to forecasts from the ECMWF-EPS. Again, both systems are compared on the original grid resolutions (ECMWF-EPS interpolated to a regular grid of approximately 25-km grid spacing).

4.1. Deterministic performance

Initially, we test whether 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 3-hourly lateral boundary conditions from ERA-Interim. Since no observations are assimilated the model trajectory 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., 2011). We make use of analysis data from 00 and 12 UTC as well as +03-h, +06-h and +09-h reforecasts. 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 resolution 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, we confine the verification to a German subdomain which has a very dense network comprising 1034 rain gauge stations (see Fig. 8). To each of the stations the nearest neighbour 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. We employ 3-hourly accumulated precipitation sums for the verification.

Figure 9 shows the frequency bias, the log odds ratio as well as the equitable threat score for the summer experiment in the upper row and the winter experiment in the lower row. All scores are computed based on a contingency table for binary events. As thresholds, we have chosen 0.1 mm/3 h, 1 mm/3 h, 2.5 mm/3 h and 5 mm/3 h which exhibit high-enough base rates (shown in Fig. 10) to be considered fair for a comparison including ERA-Interim

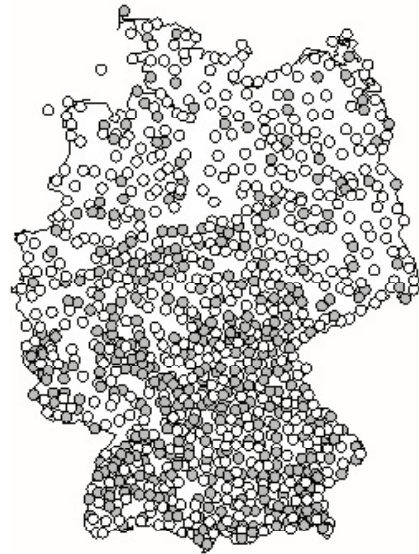


Fig. 8. Rain gauges used for verification. The grey stations are used for the probabilistic verification, and the white ones are additionally used for the deterministic verification.

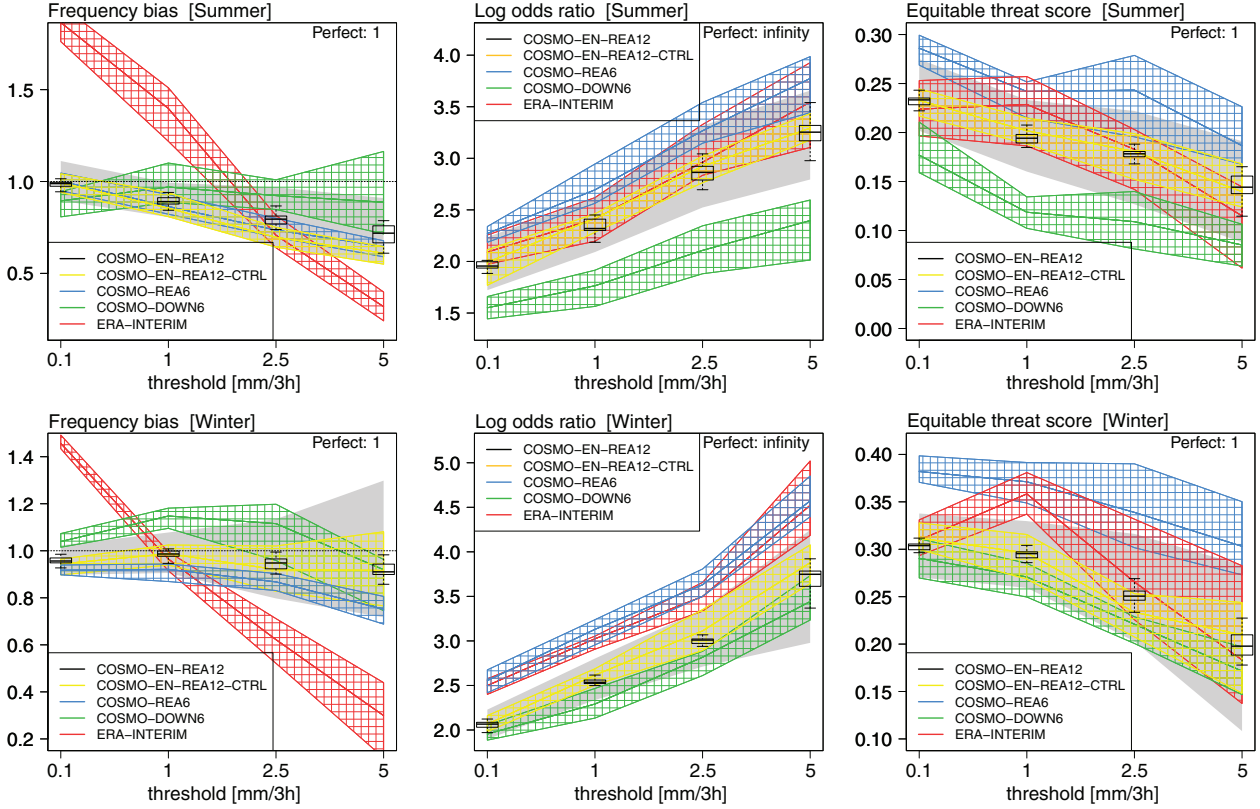


Fig. 9. Verification measures (from left to right: frequency bias, log odds ratio, equitable threat score) based on the (2×2) contingency table for binary events as a function of threshold for 3-hourly accumulated precipitation. The upper panels show the summer experiment (June 2011), and the lower ones show 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 20 000 bootstrap samples of the ensemble members.

and our short experimental periods that extend only over 30 d. To estimate the sampling uncertainty of the scores we perform 1000 bootstrap samples based on days as entities.

4.1.1. Frequency bias. The frequency bias (Donaldson et al., 1975) in the first column of Fig. 9 together with the base rates shown in comparison with the observations in Fig. 10 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

$$FB = \frac{a + b}{a + c} \in [0, \infty] \quad (3)$$

with a hits, b false alarms, c misses and d correct negatives. ERA-Interim heavily overestimates the number of precipitation events at 0.1 mm/3 h and 1 mm/3 h and underestimates it at 2.5 mm/3 h and 5 mm/3 h. This is also revealed by the base rate of ERA-Interim which is over-rated at 0.1 mm/3 h and

too small at 5 mm/3 h. However, at 1 mm/3 h in winter, the bias is almost perfect. In summer, the ensemble nudging control run and COSMO-REA6 have a virtually perfect frequency bias at the threshold 1 mm/3 h. The boxplot shows that the ensemble members scatter evenly around the control. Going to higher thresholds the bias for both COSMO-REA6 and ensemble nudging 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 ensemble nudging have a frequency bias close to 1, whereby the latter is the best system regarded over all shown decision thresholds. The observation perturbations again have a positive impact at the higher thresholds.

4.1.2. Log odds ratio. The log odds ratio (Stephenson, 2000) in the second column of Fig. 9 gives insight into the spatio-temporal coherence of reanalyses and observations.

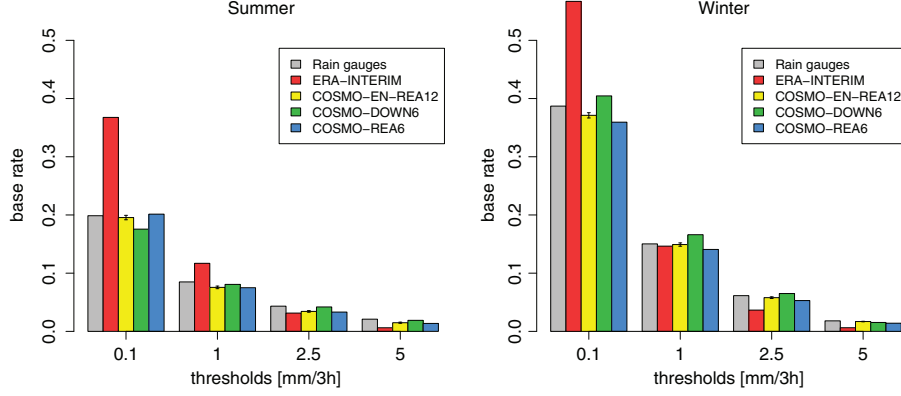


Fig. 10. Base rates of threshold exceedance for 3-hourly accumulated precipitation in the summer and winter experiments. Sample size approximately 255 000.

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)

$$LOR = \log\left(\frac{ad}{bc}\right) \in [-\infty, +\infty]. \quad (4)$$

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 ensemble nudging control run 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.

4.1.3. Equitable threat score. The equitable threat score shown in the third column of Fig. 9 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

$$ETS = \frac{a - a_{random}}{a + b + c} \quad (5)$$

with

$$a_{random} = \frac{(a + c)(a + b)}{n}, \quad (6)$$

where n is the total number of events. Again, COSMO-REA6 performs best for 1 mm/3 h while the downscaling is the worst system. Ensemble nudging and ERA-Interim exhibit no significant difference except for 1 mm/3 h where ERA-Interim outperforms our system in the winter experiment. In the summer experiment, nudging has a more

positive influence on accuracy going from 6-km (downscaling) to 12-km grid spacing (nudging) than in winter.

4.1.4. Discussion. As preliminary answer to the first research question posed above we find that the accuracy of ensemble nudging is comparable to the one of ERA-Interim. Thus, we do not have an added value in accuracy at the small precipitation thresholds chosen to have a fair amount of events in our relatively short experimental periods. Examination of COSMO-REA6 yields that our system loses the added value compared to ERA-Interim 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, we find that there is still an added value in accuracy (except for the ETS in winter) even though our system has only 12 instead of 6-km grid size.

Our findings regarding accuracy of precipitation on small thresholds are different from the ones of Jerney 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, we take a look at higher precipitation amounts 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 our system compared to ERA-Interim. The frequency bias of all regional systems is nearly perfect and best for ensemble nudging in winter.

4.2. Probabilistic performance

The probabilistic capabilities of an ensemble include consistency, accuracy, reliability, resolution and sharpness as well as absence of conditional and unconditional biases (Murphy, 1973b; Wilks, 1995; Anderson, 1997). As there is no global ensemble reanalysis data available at the moment we compare these properties 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 we do not have 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 are 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, we have used 3-hourly precipitation sums in section 4.1 and now use 6-hourly ones.

As a first step, for most of the scores computed in the following sections probabilities have to be estimated from the ensemble. We estimate the event probability $p(y_i)$ using a beta-binominal model with a flat beta prior (Agresti and Hitchcock, 2005) which leads to

$$p(y_i) = \frac{i + 1}{N_{ens} + 2}, \quad (7)$$

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. We start by examining the ensemble nudging's consistency making use of analysis rank histograms.

4.2.1. Consistency and uncertainty estimation capabilities. 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 to the ensemble is approximately flat.

The analysis rank histograms in the upper row of Fig. 11 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, we omit the events at which both the ensemble and the observation indicate 'no precipitation'. We do not show results for the ECMWF-EPS here as our sample size is too small for judging the performance of a 50-member ensemble using an analysis rank histogram.

For the summer experiment, 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 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 overestimates precipitation. This can, for example, arise from the ensemble being too confident about the position of a frontal system so that all members misplace precipitation.

4.2.2. 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 the mean squared

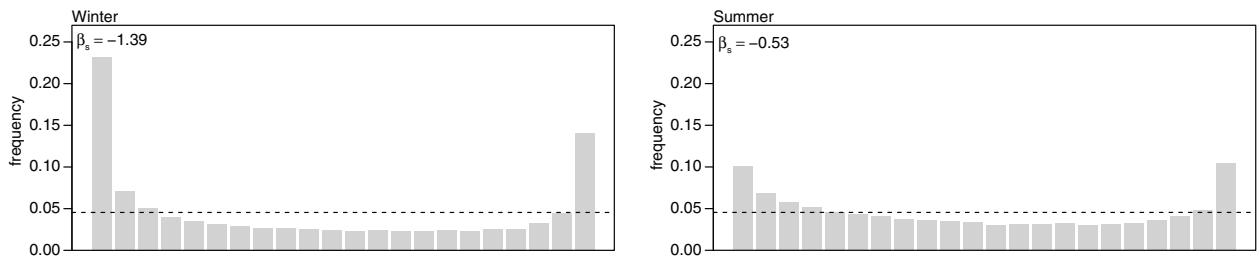


Fig. 11. Analysis rank histograms for 6-hourly accumulated precipitation in the summer experiment (left panel) and the winter experiment (right panel) with the probabilistic regional reanalysis system.

deviation between n pairs of ensemble probability y_k and binary observations o_k

$$BS = \frac{1}{n} \sum_{k=1}^n (p(y_k) - o_k)^2 \quad (8)$$

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

$$BSS = 1 - \frac{BS}{BS_{ref}} \quad (9)$$

is useful to compare the accuracy of ensemble nudging and the ECMWF-EPS as reference system. In that constellation, ensemble nudging has skill or an added value if $BSS > 0$. Figure 12 shows the Brier skill score for precipitation from both ensemble nudging experiments accumulated over 6 hours with +06-hour ECMWF-EPS forecasts valid at 06 and 18 UTC as a reference. For the summer experiment, ensemble nudging has unconditionally more accuracy than the ECMWF-EPS. At 0.1 mm/6 h, the percentage improvement of ensemble nudging is about 38 %. For the other thresholds it is about 18 % for the median. In winter, ensemble nudging performs poorer than the ECMWF-EPS except for the 0.1 mm/6 h threshold and is approximately comparable at 0.5 mm/6 h. 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

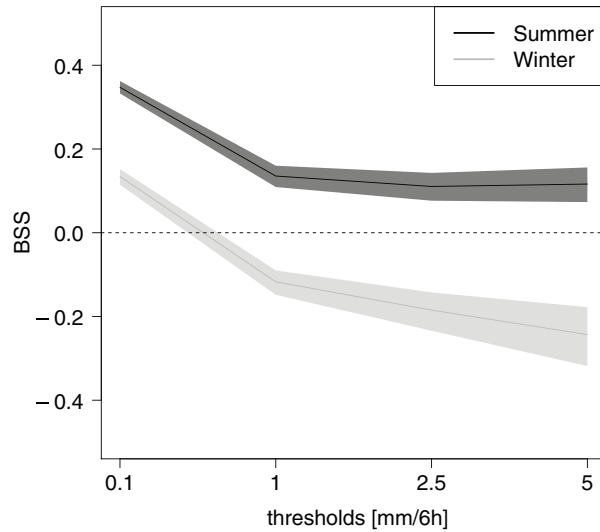


Fig. 12. Brier skill score of ensemble nudging versus ECMWF-EPS for 6-hourly accumulated precipitation. The summer experiment is illustrated in dark grey and the winter experiment in light grey. The solid lines represent the medians and the shading represents the sampling uncertainty as given by the 95 % quantile of 1000 bootstrap samples.

probabilistic representation of summer precipitation. However, this conclusion holds only for the chosen thresholds.

To take into account the whole range of precipitation amounts, we make use of the continuous ranked probability score CRPS (Hersbach, 2000) which is given by

$$CRPS = \int_{-\infty}^{\infty} [P_f(x) - P_o(x)]^2 dx \quad (10)$$

where,

$$P_o(x) = \begin{cases} 0, & x < o \\ 1, & x \geq o \end{cases} \quad (11)$$

is the Heaviside function that goes from 0 to 1 where 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. We draw 1000 bootstrap samples and obtain $CRPSS \in [-0.01, 0.00, 0.012]$ 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 ensemble nudging 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, we choose a subset of our data from the 5th to the 8th of June 2011. During these days, strong small-scale convective activity has been observed over large parts of Germany. Drawing 1000 bootstrap samples, we obtain $CRPSS \in [0.32, 40.7, 50.3]$. 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.

A decomposition of the Brier score into a reliability, a resolution and an uncertainty component (Murphy, 1973a) is depicted in Fig. 13. The decomposition including estimates of sampling uncertainty based on 1000 bootstrap samples shows that the superiority of ensemble nudging 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 ensemble nudging (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.

4.2.3. Reliability. Reliability diagrams show the conditional event relative frequencies conditioned to the ensemble probabilities (Wilks, 1995). Theoretically, the ensemble probability and the 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 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. We repeat this cycle 1000 times.

Figure 14 shows reliability diagrams for 6-hourly precipitation sums from ensemble nudging and the ECMWF-EPS. Instead of bars, the consistency intervals are illustrated as polygonal lines. For all experiments and thresholds, the ECMWF-EPS except for 5 mm/6 h in winter exhibits over-forecasting of the observed frequencies. In the summer experiment, ensemble nudging is well calibrated at both thresholds. In winter, at 0.1 mm/6 h, 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/6 h threshold, the ensemble has a conditional bias at the higher ensemble probabilities.

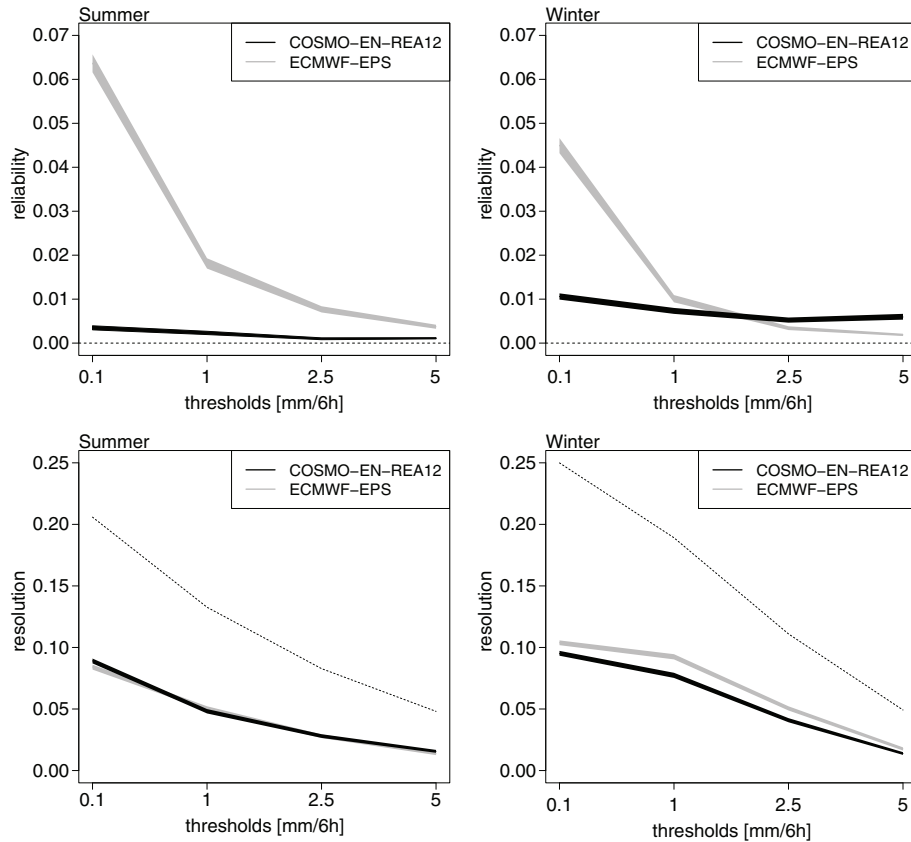


Fig. 13. Decomposition of the Brier score: 5 % to 95 % quantiles of 1000 bootstrap samples of reliability (upper panels) and resolution (lower panels) for the summer (left) and winter (right) experiments. The thin dotted lines indicate the maximally achievable resolution if the reliability was perfect.

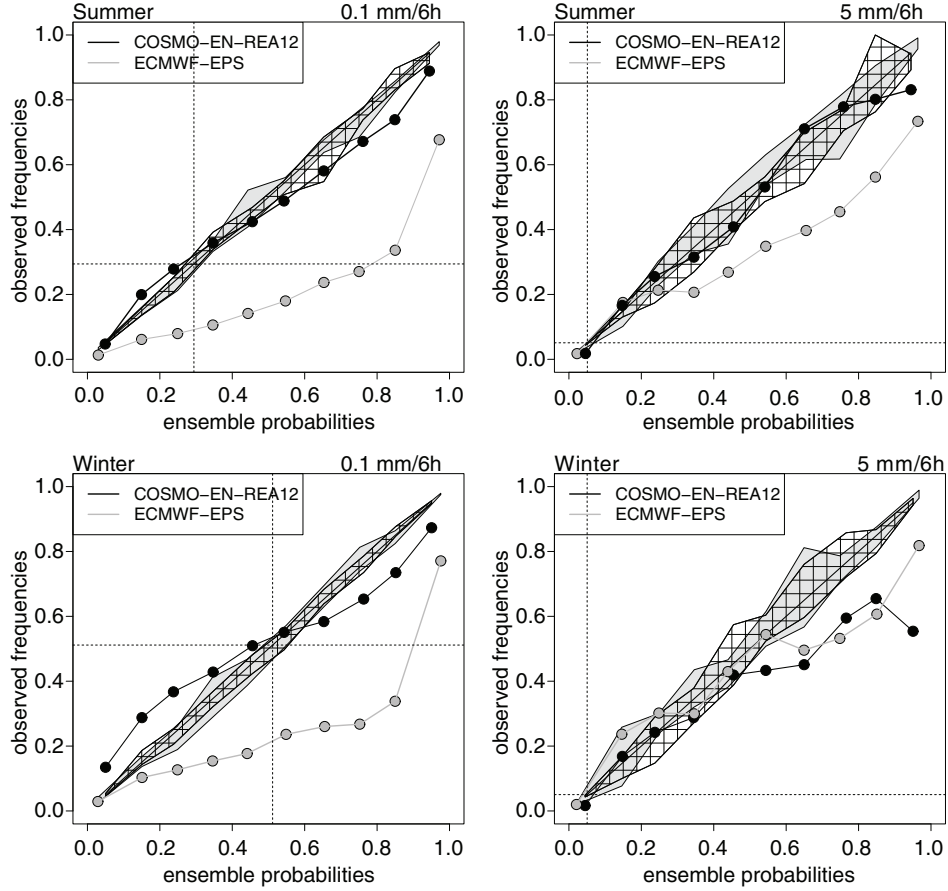


Fig. 14. Reliability diagrams for 0.1 mm/6 h (left panels) and 5 mm/6 h (right panels) decision thresholds for the summer and winter experiments. The grey shaded area represents consistency intervals for the ECMWF-EPS and the black hatched one for ensemble nudging. These represent the areas within which the ensemble is reliable. The dashed vertical and horizontal lines represent the climatological observed frequency of the events.

Here, the ensemble is again over-confident so that it overestimates 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 ensemble nudging compared to the ECMWF-EPS.

4.2.4. Resolution and discrimination. The Receiver Operating Characteristic (ROC) 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); that is, 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 the ensemble system can discriminate between events and the less resolution it has.

An exemplary ROC curve is displayed in the uppermost panel of Fig. 15. It shows the resolution of the ensemble nudging 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/6 h. On first sight, the ROC curves of 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 ensemble nudging. This is probably rooted in a much higher base rate of the IFS (see Fig. 10) 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 employed ECMWF-EPS data are based on.

The plot in the middle of Fig. 15 shows the area under the ROC curve which allows for an easier comparison for different experiments and thresholds. Optimally, the area under the ROC curve is 1. Below a value of 0.75 depicted as

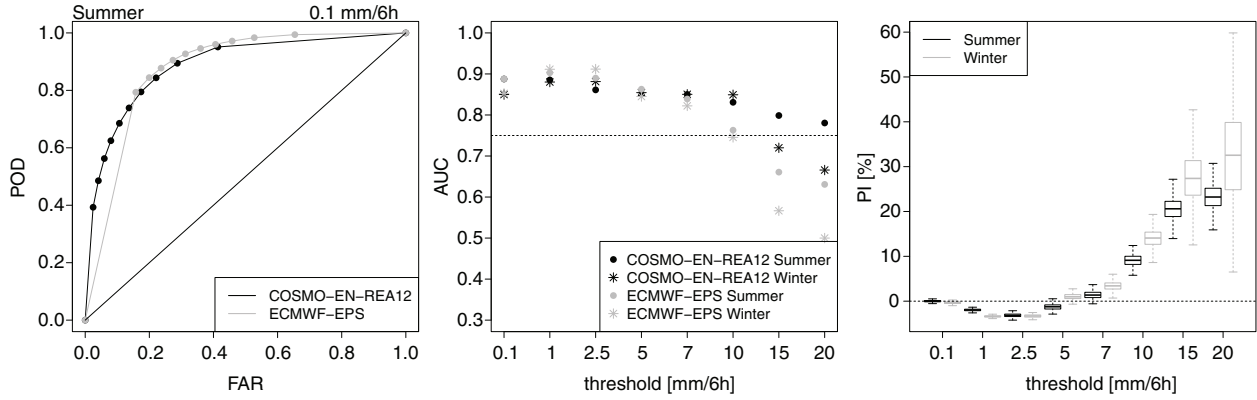


Fig. 15. ROC curve measures. The left panel shows exemplary ROC curves for 0.1 mm. The panel in the middle shows absolute values for the area under the ROC curve (AUC) for different thresholds. In both panels, COSMO-EN-REA12 is depicted in black and the ECMWF-EPS in grey. The right panel depicts the percentage improvement (PI) in terms of AUC of ensemble nudging over the ECMWF-EPS for the summer (black boxplots) and winter experiments (grey boxplots).

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/6 h. Here, the ECMWF-EPS falls under the skill line while ensemble nudging retains an area under the ROC curve of about 0.85. Above that threshold, ensemble nudging keeps discrimination for the summer experiment.

To quantify the quality of ensemble nudging relative to the ECMWF-EPS, we compute a percentage of improvement for the area under the ROC curve including 1000 bootstrap samples to incorporate sampling uncertainty. This is illustrated as black boxplots for summer and grey boxplots for winter in the lowermost panel of Fig. 15. As discussed, the two systems are comparable for the smallest threshold. At 1 mm/6 h and 2.5 mm/6 h ensemble nudging is slightly worse while from 5 mm/6 h upwards ensemble nudging is increasingly superior which is consistently more pronounced in winter. This is one of the most important features that speak for using estimates of precipitation as essential climate variables from regional reanalyses.

4.2.5. Discussion. Summarising the evaluation of the basic probabilistic performance measures discussed in this section, we can give a tentative answer with respect to precipitation to the second research question posed initially. 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 our experimental

period. However, computed for a subset of the data that is dominated by small-scale convection, the CRPSS indicates a clear added value of the regional ensemble.

The resolution is comparable at lower precipitation thresholds, while an added value can be observed for ensemble nudging 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, our probabilistic regional reanalysis system demonstrates to yield an ensemble with good probabilistic capabilities.

5. Conclusion and outlook

The regional reanalysis system introduced here will be employed to produce a test ensemble reanalysis extending over 5 years in the framework of UERRA. The suite is based on the limited-area model COSMO and set up for the CORDEX-EUR11 domain at a grid spacing of 12 km. Using ensemble nudging as data assimilation scheme, it allows for estimating the sensitivity of the reanalysis to uncertainties in the observations, that is, representativity errors, errors in the observation operators as well as measurement errors and their projection on the model dynamics. Major arguments for the production of regional reanalyses are the better spatial sampling and an added value in accuracy compared to global reanalyses and dynamical downscalings of the latter.

We have shown basic diagnostics of the performance of our system followed by an evaluation of precipitation based on two experiments. Obviously, our evaluation period is too short for drawing decisive conclusions on the quality of our system in representing precipitation as essential climate variable. 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 evaluation period has been useful to observe tendencies for our reanalysis system's quality.

The deterministic verification shows that the accuracy 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. However, these results can still be changed through better statistics in a longer evaluation period with more precipitation events, particularly for strong precipitation. 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 6-km grid spacing.

While the bisection of the resolution relative to COSMO-REA6 leads to a degradation of accuracy, a clear benefit of our 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. In the future, we aim to improve this by incorporating model error and uncertainty in the lateral boundary conditions using an ensemble of a global reanalysis, for example, ERA5 which is currently prepared at ECMWF. 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 favourable 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 thresholds. The discrimination capabilities as summarised 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 summarised that our regional ensemble reanalysis system 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.

In the future, the reanalysis system will be enhanced by incorporating uncertainty in the lateral boundary conditions and model error. A second step would be a tuning of the system to be suitable for the production of reanalysis that will extend over many decades (instead of 5 years that the system in its configuration shown here is targeted at). Different from the spatio-temporal heterogeneity that global reanalyses face, we expect that regional reanalyses of the European climate will benefit from a density of conventional observations that is somewhat more homogeneous over time. However, this has to be investigated in more detail. To give an example, temperature observations from aircrafts exist only from the 1990s. This requires the quantification of their impact on reanalysis quality in observation simulation experiments. Potentially, it has to be considered to exclude aircraft temperature data from the assimilation to keep the number of non-climatic signals in the reanalysis small. A third aspect for continuing work in the future is that satellite radiances have shown to have a very positive impact, for example, in the global reanalysis ERA-Interim. We consider their incorporation in regional reanalysis systems important for the future. In the foreseeable future, we will replace the data assimilation by a local ensemble transform Kalman filter developed recently (Schraff et al., 2016). This new scheme will offer the possibility to include satellite radiances in the reanalysis.

6. Acknowledgement

This work has been conducted in the framework of the project UERRA, funded by the Seventh Framework Program of the European Commission under grant agreement no. 607193. We thank Christoph Bollmeyer, Ulrich Schättler and Martin Lange for their advice and technical support and Sabrina Wahl for providing us with an R-routine for computing the horizontal kinetic energy spectra.

References

- 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.

- Baehr, J., Fröhlich, K., Botzet, M., Domeisen, D. I. V., Kornblüeh, L. and co-authors. 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. and Raschendorfer, M. 2011. Operational convective-scale numerical weather prediction with the COSMO model: description and sensitivities. *Mon. Weather Rev.* **139**(12), 3887–3905.
- 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.
- 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.
- 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 co-authors. 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. and co-authors. 2015. Towards a high-resolution regional reanalysis for the european CORDEX domain. *Q. J. R. Meteorol. Soc.* **141**(686), 1–15.
- Bott, A. 1989. A positive definite advection scheme obtained by nonlinear renormalization of the advective fluxes. *Mon. Weather Rev.* **117**, 1006–1015.
- Brier, G. 1950. Verification of forecasts expressed in terms of probability. *Mon. Weather Rev.* **78**(1), 189–195.
- Bröcker, J. and Smith, L. 2007. Increasing the reliability of reliability diagrams. *Weather Forecast.* **22**(3), 651–661.
- 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**, 644–658.
- Chen, X., Liu, C., Driscoll, K. O., Mayer, B., Su, J. and co-authors. 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. and co-authors. 2011. The twentieth century reanalysis project. *Q. J. R. Meteorol. Soc.* **137**(654), 1–28.
- Dahlgren, P., Landelius, T. and Gollvik, S. 2016. A high resolution regional reanalysis for Europe Part I: 3-dimensional reanalysis with the regional High Resolution Limited Area Model (HIRLAM). *Q. J. R. Meteorol. Soc.* **142**(698), 2119–2131.
- Davies, H. C. 1983. Limitations of some common lateral boundary schemes used in regional NWP models. *Mon. Weather Rev.* **111**, 1002–1012.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P. and co-authors. 2011. 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. and co-authors. 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.
- 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.
- Giorgi, F., Jones, C. and Asrar, G. R. 2009. Addressing climate information needs at the regional level: the CORDEX framework. *WMO Bull.* **58**, 175–183.
- Hamill, T. M. and Colucci, S. J. 1997. Verification of EtaRSM short-range ensemble forecasts. *Mon. Weather Rev.* **125**(6), 1312–1327.
- Hersbach, H. 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather Forecast.* **15**, 559–570.
- Hess, R. 2001. Assimilation of screen-level observations by variational soil moisture analysis. *Meteorol. Atmos. Phys.* **77**(1–4), 145–154.
- Hollingsworth, A., Lnnberg, P. 1986. The statistical structure of short-range forecast errors as determined from radiosonde data Part II: The covariance of height and wind errors. *Tellus A.* **38A**, 137–161.
- Houtekamer, P. L., Lefavre, 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.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D. and co-authors. 1996. The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Am. Meteorol. Soc.* **77**(3), 437–471.
- Kaspar, F., Müller-Westermeier, G., Penda, E., Mächel, H., Zimmermann, K. and co-authors. 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.
- Kobayashi, S., Ota, Y., Harada, Y., Ebata, A., Moriya, M. and co-authors. 2015. The JRA-55 reanalysis: general specifications and basic characteristics. *J. Meteorol. Soc. Japan. Ser. II* **93**(1), 5–48.
- 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.
- Lönnberg, P. and Hollingsworth, A. 1986. The statistical structure of shortrange forecast errors as determined from radiosonde data Part II: The covariance of height and wind errors. *Tellus* **38A**, 137–161.
- Mason, I. 1982. A model for assessment of weather forecasts. *Aust. Meteorological Mag.* **30**, 291–303.
- Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C. and co-authors. 2006. North American regional reanalysis. *Bull. Am. Meteorol. Soc.* **87**(3), 343–360.
- 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.
- Onogi, K., Tsutsui, J., Koide, H., Sakamoto, M., Kobayashi, S. and co-authors. 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., Petroliaigis, T. and Park, S. 1997. The ECMWF ensemble prediction system. *Meteorol. Appl.* **4**(4), 301–304.
- Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J. and co-authors. 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. and co-authors. 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.
- Saha, S., Moorthi, S., Pan, H. L., Wu, X., Wang, J. and co-authors. 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. 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.
- Schraff, C. and Hess, R. 2012. *Consortium for Small-Scale Modeling 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. and co-authors. 2016. Kilometre-scale ensemble data assimilation for the COSMO model (KENDA). *Q. J. R. Meteorol. Soc.* **142**, 1453–1472.
- Schroeder, A. J., Stauffer, D. R., Seaman, N. L., Deng, A., Gibbs, A. M. and co-authors. 2006. An automated high-resolution, rapidly relocatable meteorological nowcasting and prediction system. *Mon. Weather Rev.* **134**, 1237–1265.
- Schulz, J.-P., Vogel, G., Becker, C., Kothe, S. and Rummel, U. 2016. Evaluation of the ground heat flux simulated by a multi-layer land surface scheme using high-quality observations at grass land and bare soil. *Meteorol. Zeitschrift* **25**, 607–620.
- 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 I: modeling design and basic performance characteristics. *J. Appl. Meteorol.* **34**, 1739–1761.
- Simmer, C., Adrian, G., Jones, S., Wirth, V., Göber, M. and co-authors. 2016. HErZ. The German Hans-Ertel Centre for Weather Research. *Bull. Am. Meteorol. Soc.* **97**, 1057–1068.
- Skamarock, W. C. 2004. Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Weather Rev.* **132**, 3019–3032.
- Stauffer, D. R. and Seaman, N. L. 1990. Use of four-dimensional data assimilation in a limited-area mesoscale model. Part I: experiments with synoptic-scale data. *Mon. Weather Rev.* **110**, 1250–1277.
- 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.
- Stephenson, D. B. 2000. Use of the odds ratio for diagnosing forecast skill. *Weather Forecast.* **15**(2), 221–232.
- 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**, 1779–1800.
- 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, VT.