Initial condition perturbations in a convective scale ensemble prediction system

Keresturi, Endi

University of Zagreb
FACULTY OF SCIENCE
DEPARTMENT OF GEOPHYSICS

Endi Keresturi

INITIAL CONDITION PERTURBATIONS IN A CONVECTIVE SCALE ENSEMBLE PREDICTION SYSTEM

DOCTORAL DISSERTATION

Zagreb, 2021

University of Zagreb
FACULTY OF SCIENCE
DEPARTMENT OF GEOPHYSICS

Endi Keresturi

INITIAL CONDITION PERTURBATIONS IN A CONVECTIVE SCALE ENSEMBLE PREDICTION SYSTEM

DOCTORAL DISSERTATION

Supervisors:
dr. sc. Kristian Horvath
dr. sc. Yong Wang

Zagreb, 2021

Sveučilište u Zagrebu
PRIRODOSLOVNO-MATEMATIČKI FAKULTET
GEOFIZIČKI ODSJEK

Endi Keresturi

PERTURBACIJE POČETNIH UVJETA U PROGNOSTIČKOM ANSAMBL-SUSTAVU KONVEKTIVNE SKALE

DOKTORSKI RAD

Mentori:
dr. sc. Kristian Horvath
dr. sc. Yong Wang

Zagreb, 2021

This dissertation is elaborated on Zentralanstalt für Meteorologie und Geodynamik (ZAMG) and Croatian Meteorological and Hydrological Service (DHMZ) under the supervision of dr. sc. Yong Wang and dr. sc. Kristian Horvath

I would like to thank dr. sc. Yong Wang, dr. sc. Kristian Horvath and dr. sc. Branka Ivačan-Picek without whom this long and prosperous journey wouldn't even start. They have all put their faith in me and granted me the possibility to go to a PhD research stay at ZAMG where I have spent three wonderful years. Special thanks go to my colleagues at ZAMG - Christoph Wittmann, Florian Weidle and Florian Meier for providing unconditional help with my research and assistance in finding my way in a foreign country.

I wish to thank Jelena Bojarova for providing helpful comments during *Jk* blending implementation, prof. dr. sc. Branko Grisogono whose suggestions have improved this manuscript and RC LACE for financial support

Finally, special thanks go to my family and friends whose support, during all these years, has played an important role in my success. My deepest gratitude goes to Dijana who believed in me, supported me and kept encouraging me from the very beginning.

Live long and prosper!

Table of Contents

ABSTRACT

One of the main challenges presented by a limited area model ensemble prediction system (LAMEPS) concerns the limited capacity for its initial condition (IC) perturbations to correctly represent large-scale flow uncertainties due to its limited-size domain, deficiencies in formulating lateral boundary conditions and inadequate availability of observations. In addition, a mismatch between LAMEPS IC and host EPS lateral boundary perturbations can form spurious waves at the boundaries which spread through the domain, induce noise to the meteorological fields and render inoperative frequent assimilation cycles.

In the present work, an ensemble *Jk* blending method is proposed for improving representation of large-scale uncertainties and for addressing consistent initial conditions and lateral boundary perturbations. Our approach involves employing *Jk* blending within a framework of 3-dimensional variational (3D-Var) ensemble data assimilation (EDA). In such a system, small-scale perturbations are generated from 3D-Var EDA, while large-scale perturbations are generated from the host ensemble via *Jk* blending. We hypothesize that final analyses are optimal, and contain perturbed small and large scales which are, at the same time, consistent with each other and with perturbations coming from lateral boundaries.

The ensemble *Jk* method is implemented to the C-LAEF (Convection-permitting Limited-Area Ensemble Forecasting) system and is compared to the standard perturbed-observation EDA approach, i.e., perturbed-observation EDA without large-scale constraint. The comparison shows that the ensemble *Jk* method gives a more skillful and reliable EPS, especially for the upper-air variables. In addition, positive effects on the surface pressure and precipitation of large-scale perturbations are shown. The ensemble *Jk* method's capacity to alleviate perturbation mismatches is also assessed.

Additionally, two readily available techniques, i.e., neighborhood and lagging, to improve C-LAEF's IC perturbation sampling of the initial uncertainties and to address the problem of relatively low model effective resolution are evaluated. Both of them show significantly positive impact on ensemble forecast quality and on detection of extreme weather events.

Keywords: ensemble prediction system, initial condition perturbations, blending, data assimilation, limited-area model, neighborhood approach, lagged forecasting

LIST OF ABBREVIATIONS

3D-Var	3-dimensional variational assimilation
4D-Var	4-dimensional variational assimilation
ALADIN-LAEF	Aire Limitee Adaptation Dynamique Developpement InterNational – Limited-Area Ensemble Forecasting
AROME	Application of Research to Operations at Mesoscale
AT	Altitude tolerance
BS	Brier score
C-LAEF	Convection-permitting Limited-Area Ensemble Forecasting
CRPS	Continuous ranked probability score
DF	Digital filter
DWD	Deutscher Wetterdienst
ECMWF	European Centre for Medium-Range Weather Forecasts
EDA	Ensemble data assimilation
EPS	Ensemble prediction system
ET	Ensemble transform
ETKF	Ensemble transform Kalman filter
EnKF	Ensemble Kalman filter
FSS	Fractions skill score
GF08	Guidard and Fischer (2008)
HIRLAM	High Resolution Limited Area Model
IC	Initial condition
IFS HRES	Integrated Forecast System high resolution
INCA	Integrated Nowcasting through Comprehensive Analyses
L-S	Land-sea
LAF	Lagged averaged forecasting
LAM	Limited-area model
LAMEPS	Limited-area model ensemble prediction system
LBC	Lateral boundary condition
MSLP	Mean sea level pressure
NCEP	National Centers for Environmental Prediction
NEP	Neighborhood ensemble probability

NMEP	Neighborhood maximum ensemble probability
NWP	Numerical weather prediction
PDF	Probability density function
RC-LACE	Regional Cooperation for Limited Area Modelling in Central Europe
RH2M	Relative humidity at 2 m
RH500	Relative humidity at 500 hPa
RH850	Relative humidity at 850 hPa
RMSE	Root mean square error
ROC	Relative operational characteristics
SSETKF	Scale-selective ensemble transform Kalman filter
T2M	Temperature at 2 m
T500	Temperature at 500 hPa
T850	Temperature at 850 hPa
TKE	Turbulence kinetic energy
ULL	Upper-level low
W10M	Wind speed at 10 m
W500	Wind speed at 500 hPa
W850	Wind speed at 850 hPa
WG10M	Wind gusts speed at 10 m
ZAMG	Zentralanstalt für Meteorologie und Geodynamik

§ 1. INTRODUCTION

1.1. Theoretical background

Let us first introduce some basic concepts and define the most important variables used through the rest of this dissertation.

1.1.1. Numerical weather prediction

Numerical weather prediction (NWP; e.g., Kalnay, 2003; Coiffier, 2011; Bauer *et al.*, 2015) is a method of weather forecasting that employs a set of equations that describe flow of fluids. This set of equations, called the atmospheric or NWP model, is solved numerically for different moments in time. For this purpose, equations are discretized – the equations related to continuous variables are replaced by equations related to discrete variables. Discretization of equations means that they are solved on specific points in 3D space and time, i.e., grid. Physical distance between the two points in space (Δx) is usually referred to as the grid size. Finite grid size means that there always will be physical processes and scales of motion that cannot be represented by an NWP model. Those unresolved processes need to be accounted for because they affect the larger-scale fields and processes that are explicitly resolved. The procedure by which the important atmospheric processes that cannot be resolved directly by an NWP model are represented is called parameterization (e.g., Stensrud, 2007).

NWP models can be integrated over the whole globe (global models) or on limited size domains (limited area models; LAMs). LAMs have the advantage of using the finer grid size than global models because the available computational resources are focused on a specific area instead of being spread over the globe. This allows them to resolve explicitly smaller-scale meteorological phenomena that cannot be represented on the coarser grid of a global model. Because LAM domains have physical boundaries, they need to be provided with lateral boundary conditions (LBCs; e.g., Warner *et al.*, 1997). This is usually done by a technique called nesting, where LAM is *nested* into a global model or another LAM. In other words, another NWP model is used to provide the LBC data. This model is usually referred to as the host model.

Today (i.e., 2021), typical grid sizes of global models are in the range of 10-20 km, while for LAMs they are around 10 times smaller. We will denote N_m-dimensional NWP model state vector with $\boldsymbol{X}$. $\boldsymbol{X}$ is a column vector containing all model grid points in 3D space for all the physical variables (e.g., temperature, humidity, wind, etc.). Size of N_m is typically somewhere between 10^6 – 10^{10}. Capital *T*, as index in variable names, will denote the truth (the true state of the atmosphere - $\boldsymbol{X}_T$) which is unknown.

1.1.2. Data assimilation

The time integration of an NWP model is an initial-value problem, which means that the ability to make a skillful forecast requires both that the computer model be a realistic representation of the atmosphere, and that the initial conditions (ICs) be known accurately. The key part of the latter are observations of the current state of the atmosphere. Observations are made with different instruments at different locations in different ways (i.e., 2 m temperature observation using a common thermometer, satellite observations of clouds, radar observations of precipitation, etc.; see Montmerle (2018) for more information on observations for high resolution NWP). Each observation represents the information that is the closest to the true state of the atmosphere. $\boldsymbol{Y}$ denotes vector of observations which is usually, at least, two orders of magnitude smaller than $\boldsymbol{X}$. New observations are included into model state through a process known as data assimilation (e.g., Kalnay, 2003). The problem of data assimilation is to combine the model state with the observations to find the best possible description of the atmosphere at a given time, i.e., analysis $\boldsymbol{X^a}$:

$$\boldsymbol{X}_T = \boldsymbol{X}^a + \varepsilon_a \tag{1.1}$$

Where ε_a is the analysis error which is unknown. In other words, we need to find an optimum analysis of a field of model variables $\boldsymbol{X^a}$, given a background field $\boldsymbol{X^b}$ (past model forecast) available at grid points in three dimensions, and a set of observation $\boldsymbol{Y}$ available at irregularly spaced points. Similar to $\boldsymbol{X^a}$, $\boldsymbol{X^b}$ and $\boldsymbol{Y}$ also contain errors which we denote ε_b and ε_o, respectively. Through the years, many methods have been developed for that purpose (Bannister, 2017; Gustafsson *et al.*, 2018). Most of them are based on the idea of least-squares estimation or minimization of variance of residual. From the well-known linear theory of least-squares estimation (e.g., Kalnay, 2003), the optimal least-squares estimator (analysis in our case) is defined by the following interpolation equation (i.e., analysis equation):

$$\boldsymbol{X^a} = \boldsymbol{X^b} + W\left[\boldsymbol{Y} - H\left(\boldsymbol{X^b}\right)\right] \tag{1.2}$$

Where W is weight matrix, and H is a nonlinear observation operator which maps model forecast to observation locations. Analysis error covariance matrix (**A**; i.e., variance of residual) should be minimized and is calculated from:

$$\boldsymbol{A} = E(\varepsilon_a \varepsilon_a^T) = \boldsymbol{B} - WH\boldsymbol{B} \tag{1.3}$$

Where $\boldsymbol{B} = E(\varepsilon_b \varepsilon_b^T)$ is the covariance matrix of the background errors ε_b. $E()$ is the expected value and it is assumed that $E(\varepsilon_a) = E(\varepsilon_b) = 0$ ($\boldsymbol{X^a}$ and $\boldsymbol{X^b}$ are unbiased).

Today, the most popular methods are the, so-called, variational methods (Le Dimet and Talagrand, 1986) that use observations and past model forecasts (background) to find the state of the atmosphere closest to the truth. The most used methods in this category are 3D-Var (Gustafsson *et al.*, 2001) and 4D-Var (Rabier *et al.*, 2000). Variational methods are equivalent to the least-square estimation, but the approach is different as they find the analysis that minimizes a cost function measuring its distance to the background and to the observations:

$$J(\mathbf{X}) = J_b(\mathbf{X}) + J_o(\mathbf{X}) \tag{1.4}$$

$$J(\mathbf{X}) = \frac{1}{2}\left(\mathbf{X} - \mathbf{X^b}\right)^{\mathrm{T}} \mathbf{B}^{-1}\left(\mathbf{X} - \mathbf{X^b}\right) + \frac{1}{2}\left(\mathbf{Y} - H\mathbf{X^b}\right)^{\mathrm{T}} \mathbf{R}^{-1}\left(\mathbf{Y} - H\mathbf{X^b}\right) \tag{1.5}$$

where $\boldsymbol{R} = E(\varepsilon_o \varepsilon_o^T)$ is the covariance matrix of observation errors ε_o which are also assumed to be unbiased. In (1.4), it is assumed that background and observation errors are uncorrelated. The cost function *J(**X**)* is then minimized, using iterative minimization algorithms (e.g., Kalnay, 2003). The minimum of *J(**X**)* is attained for $\boldsymbol{X} = \boldsymbol{X^a}$, i.e., the analysis is given by the solution of:

$$\boldsymbol{\nabla}_X J(\boldsymbol{X}^a) = 0 \tag{1.6}$$

where $\boldsymbol{\nabla}$ is the gradient operator with respect to **X**. Solving (1.6) for $\boldsymbol{X^a}$, one gets the (1.2). Once the initial state of the atmosphere with sufficient accuracy is known, NWP model is propagated forward in time to obtain atmospheric state at some latter time – a forecast.

1.2. Ensemble prediction systems

We will elaborate the purpose of, illustrate the basic principles and the assumptions behind the concepts of ensemble forecasting.

1.2.1. Why do we need ensembles?

Propagating a set of equations forward in time to obtain a single forecast at some latter time is generally known as deterministic forecasting. In the perfect linear world without any errors, a single deterministic forecast is all we need. However, in the real world there are many sources of errors (Lorenz, 1969; Buizza *et al.*, 2005; Leutbecher and Palmer, 2008) and the NWP research community has been aware of them since the very beginning of NWP (Charney, 1951). These include the following:

a) Errors in ICs. These are primarily due to inadequate data assimilation procedures, insufficient number of and errors in observations. As thoroughly described in Janjić *et al.* (2018), observation errors consist of measurement errors and representativeness errors, where the latter can be divided even further into three components: i) errors due to unresolved scales and processes (model deficiencies), ii) observation operator errors in data assimilation and iii) pre-processing and quality control errors.
b) Model errors. These include equation simplification and truncation mostly due to a limited computer power which, in turn, result in our models being unable to resolve all the scales and phenomena that occur in the real world. Those unresolved atmospheric processes are represented by parametrizations which are an additional source of errors.
c) Errors in formulation of lower, upper and, in case of LAMs, lateral boundary conditions.

Existence of these errors necessarily render our deterministic forecast inaccurate (Charney, 1951).

These shortcomings became even more apparent after Lorenz (1963b) demonstrated that numerical models exhibit chaotic behavior, i.e., high sensitivity to errors in ICs. This means that difference of any magnitude at initial time will lead to a completely different state of the system at some latter time. Chaos was summarized by Lorenz as: “Chaos: When the present determines the future, but the approximate present doesn’t approximately determine the future.” (e.g., Kalnay, 2019). The chaotic nature and ever-present errors of the initial state necessarily lead to a complete loss of forecasting skill after a finite period of time, which Lorenz (1965) estimated to be about two weeks, an estimation that still holds today (e.g., Buizza and Leutbecher, 2015). Moreover, it is routinely known that predictability of the forecast is flow-dependent, as it was shown by Lorenz (1963a). In other words, predictability depends on the current state of the atmosphere – on some days, accurate 10-day forecast could be possible, while on the others, only three-day forecast could be possible. The question then

emerged – can we know, a priori, the uncertainty of the forecast? In addition, as people become aware of these facts, they realized that it would be more accurate to issue a forecast in terms of probabilities. For example, yes/no forecast of snowfall three days from today, could be more beneficial if it were formulated like this: there is 80 % probability for a snowfall three days from today. One deterministic forecast cannot provide any of those information - we need something better.

1.2.2. Ensemble forecasting

Recognizing the fact that initial state can only be known with some uncertainty leads to specifying the initial state, at initial time t_0, not as a point in the model phase space, but as a probability density function (PDF), $p(\boldsymbol{X})$ (Ehrendorfer, 1997). The time evolution of $p(\boldsymbol{X})$ can then be achieved by Liouville (perfect model) or Fokker-Planck equation (model error accounted for; Ehrendorfer, 1994a). Furthermore, Epstein (1969) proposed to evolve only first and second moment of $p(\boldsymbol{X})$. However, both approaches, where N_m is as high as 10^{10}, are computationally prohibitive (Ehrendorfer, 1994b; Palmer and Hagedorn, 2006).

The first computationally feasible method was proposed by Leith (1974) where he uses Monte Carlo technique to randomly sample the initial PDF and evolve each initial state with the forecast model. Today, this approach is known as ensemble forecasting (e.g., Leutbecher and Palmer, 2008) where multiple forecasts (ensemble members) are obtained, each started from a slightly different IC. This slight differentiation between ensemble members is produced by adding a small, but different type of perturbation to each member separately. If the initial sample (ensemble) is large enough, first and higher moments of $p(\boldsymbol{X})$, at some latter time t, can be accurately reconstructed from ensemble of forecasts. Leith (1974) showed that ensemble size can be as small as 8 members for accurately obtaining the first moment of $p(\boldsymbol{X})$. Ma *et al*. (2012) argue that ensemble size of 20-40 is enough for operational purposes today, while Leutbecher (2018) shows that ensemble size of more than 50 is relevant only when forecasting extreme events. Thus, this approach can provide the information needed to address the before mentioned problems. A more comprehensive history of ensemble prediction systems (EPSs) development can be found in Lewis (2005) and Kalnay (2019).

The three main goals of ensemble forecasting are:

a) Increase of the forecast accuracy by using the ensemble mean (averaging over all members). Ensemble mean benefits from error cancelation and it is, thus, more accurate than a single deterministic forecast (e.g., Bengtsson *et al.,* 2008).

b) Provide a priori information about predictability of which ensemble spread (standard deviation of ensemble members) can be a good indicator (Hopson, 2014)
c) Provide a probabilistic information.

1.2.3. Perturbations

To achieve these goals, an optimal EPS needs to account for all sources of uncertainties. These include uncertainties in IC, NWP model and boundary conditions (e.g., Houtekamer *et al.*, 1996; Nutter *et al.*, 2004b; Buizza *et al.*, 2005; Vié *et al.*, 2011; Wang *et al.*, 2011; Bouttier *et al.*, 2012; Nuissier *et al.*, 2012; Romine *et al.*, 2014; Wang *et al.*, 2019; Zhang, 2019). In other words, perturbations added to the EPS cannot be random, but need to follow some rules. What is the best possible way to add perturbations to the EPS is still not clear and it is an active field of research. This lack of consensus on the best way of perturbing an EPS, led to the development of numerous perturbation methods in different meteorological centers around the world.

At first, only uncertainties in ICs were considered as perturbing ICs is the most important part of a global EPS. Analysis contains both random errors introduced by the most recent observations, and growing errors associated with the instabilities of the evolving flow, dynamically generated (from random errors introduced in earlier analyses) by the repeated use of the forecast first guess. While the main goal of data assimilation is to produce $\boldsymbol{X^a}$ as close to the true state of the atmosphere $\boldsymbol{X_T}$ as possible, the main goal of IC perturbation methods is to accurately estimate analysis errors, so that second moment of those perturbations is close to **A** (1.3). If those two conditions are met, EPS will, assuming model errors are ignored and that errors are Gaussian in nature, give the correct estimation of first and second moment of the $p(\boldsymbol{X})$ at some latter time. At first, no explicit method to perturb the ICs were used. Instead, Hoffman and Kalnay (1983) developed lagged average forecasting (LAF) as an alternative to Monte Carlo forecasting, in which the forecasts initialized at the current initial time, as well as at previous times are combined to form an ensemble (see also Branković *et al*., 1990).

The first two methods used to explicitly generate IC perturbations operationally were the *breeding method* (Toth and Kalnay, 1993, 1997), developed by the National Centers for Environmental Prediction (NCEP), and *singular vectors* (Buizza and Palmer, 1995; Molteni *et al.*, 1996), developed by the European Centre for Medium-Range Weather Forecasts (ECMWF). They do not directly try to estimate analysis errors, but, although different, are based on the same idea – perturbations should be added only in these directions of phase

space in which errors will amplify the most. In theory, this fact should allow the ensemble to better sample the initial PDF because the forecast from these ICs will span the most important directions in the phase space (e.g., Vannitsem, 2017).

Today, more advanced methods exist, for example: *Ensemble Transform* (ET; Wei *et al.*, 2008), *Ensemble Transform Kalman Filter* (ETKF; Wang and Bishop, 2003), *Ensemble Kalman Filter* (EnKF; Evensen, 2003; Houtekamer and Mitchell, 2005), *Ensemble of Data Assimilations* (EDA; Buizza *et al*., 2008), etc. The main advantage of these methods is the fact that the perturbations they produce are more consistent with analysis errors and, thus, better sample the initial uncertainty and estimate **A** (1.3; Wei *et al.*, 2008).

Different approaches also exist when accounting for the model uncertainty. These include: multi-model (Du *et al.*, 2003; Garcia-Moya *et al.*, 2011; Iversen *et al.*, 2011; Hagedorn *et al.*, 2012, among others), multi-physics (Jankov *et al*., 2005; Berner *et al.*, 2011; Wang *et al.*, 2011, among others), multi-dynamics and stochastic physics (Berner *et al.*, 2009; Palmer *et al.*, 2009; Bouttier *et al.*, 2012; Wang *et al*., 2019; Wastl *et al*., 2019, among others). The simplest version of multi-model ensemble is the so-called *Poor-Man ensemble* (Bowler *et al.*, 2008a) where multiple single forecasts from various available models are just pulled together to form an ensemble if one cannot afford to run its own ensemble. The more complicated version is, obviously, if one runs many different models at the same institute. A disadvantage of multi-model approach is the cost to develop and maintain many models. Multi-physics ensembles are formed by alternating physics schemes or parameter value within a physics scheme. For example, one can vary convective or turbulence parameterization schemes or just one parameter within those schemes. Multi-dynamics ensembles, on the other hand, have different dynamical cores or different configurations of one. Stochastic physics approach targets uncertainties arising from parameterizations of sub-grid physical processes, truncation or missing processes. Several stochastics-physics methods exist, the most popular being stochastically perturbed physics tendencies scheme (Palmer *et al.*, 2009) which adds random number to physical tendencies of model variables in $\boldsymbol{X}$. Stochastic physics schemes are the most popular and are widely used due to their good performance, especially in increasing model spread and low maintenance costs.

1.2.4. Future prospects

Today, almost 30 years after the first EPSs became operational at NCEP and ECMWF, ensemble forecasting remains the preferred forecasting technique and is widely used for the

short (up to three days), medium (up to two weeks) and extended (sub-seasonal and seasonal) forecast ranges (e.g., Buizza, 2019). Among all of the before mentioned perturbation methods, answering the question which one is the best is far from trivial. There were many studies done which compared different methods (Bowler, 2006; Descamps and Talagrand, 2007; Wei *et al*., 2008; Magnusson *et al*., 2009; Zhou *et al*., 2016, among others). Regarding the IC perturbations, the results from those studies indicate that the second-generation methods like EnKF, ET and ETKF perform better than the first-generation ones like breeding or singular vectors, while EnKF outperforms ET and ETKF.

A look into the future reveals a tendency toward a unified procedure of data assimilation and EPS in an NWP system: ensemble forecast variance will provide background error covariance information for data assimilation, while data assimilation will give an ensemble of analyses to initiate an ensemble of forecasts. In such a coupled system, not only EPS can be improved by having more realistic IC perturbations reflecting true error of the day, but also the quality of analysis is improved by using flow-dependent background error information (Bannister, 2017). Furthermore, Palmer (2014) and Váňa *et al.* (2017) discuss that complete determinism of NWP equations should be abandoned altogether and probabilistic approach should be incorporated into the equations and the computer architecture itself. Recently, however, techniques to statistically predict the weather, based on machine learning, are rising in popularity. Currently, the accuracy of such methods is on the level NWP was in the late 1950s, but the prospect for such methods appears excellent (see Weyn *et al*., 2019 and references therein).

1.3. Motivation

In recent years, limited area model ensemble prediction systems (LAMEPSs) have been developed to benefit from ensemble approach applied at the mesoscale and convective scale (e.g., Xue *et al.*, 2007; Bowler *et al.*, 2008b; Clark *et al.*, 2009; Vié *et al.*, 2011; Wang *et al.*, 2011; Peralta *et al.*, 2012; Wang *et al.*, 2012; Schellander-Gorgas *et al.*, 2017; Wastl *et al.*, 2021, among others). Such ensembles are of particular importance due to a rapid loss of predictability observed at these scales (Hohenegger and Schär, 2007; Zhang *et al.*, 2007; Judt, 2018). It has long been known that by decreasing the model grid size, i.e., by increasing its resolution, forecast errors grow faster (Weyn and Durran, 2018). For a convection-permitting grid sizes of ~1 km, the small-scale errors saturate in a few hours (Hohenegger and Schär,

2007; Vannitsem, 2017; Weyn and Durran, 2017; Flora *et al*., 2018) after which the structures on those scales can be regarded as noise. Convective processes that dominate those scales are non-linear and strongly affected by uncertainties. Those facts are a serious obstacle to the successful atmospheric modeling of the smaller and smaller scales. Thus, probabilistic interpretation and ensemble forecasting in the form of LAMEPSs are a necessity. However, LAMEPSs have problems of their own which need to be addressed before an operational implementation. Here, we will focus on three problems, first two being LAMEPS specific, while the last one affects NWP models in general.

1.3.1. Problem 1: IC and LBC perturbation mismatch

When using a LAMEPS, one must address a new source of uncertainty – the specification of LBCs. This is typically done by nesting each LAMEPS member to a different host EPS member. Saito *et al.* (2012) discuss the importance of LBC perturbation in LAMEPSs. However, applying this procedure can lead to the following issue: when the method used to generate perturbations in a host EPS is different and independent from the method used in an LAMEPS, a conflict between such perturbations at the lateral boundaries can be observed. This problem has been recognized by many authors (e.g., Bowler and Mylne, 2009; Brousseau *et al.*, 2011; Wang *et al.*, 2011; Caron, 2013; Kühnlein *et al.*, 2014; Wang *et al.*, 2014; Davies, 2014; Skamarock *et al.*, 2018). Kühnlein *et al.* (2014) demonstrated improved precipitation forecasts when IC perturbations were more consistent with LBC perturbations.

Caron (2013) employed hourly cycling 3D-Var combined with ETKF to generate IC perturbations for a 1.5-km LAMEPS nested inside another LAMEPS. He demonstrated what can happen if IC perturbations are inconsistent with LBC perturbations - spurious waves are generated at the boundaries and quickly spread to the rest of the domain, resulting in excessive surface pressure spread in a 1.5-km LAMEPS; almost needless to say, this is a classical numerical problem for all LAMs. Figure 1 shows surface pressure perturbations for one single member of 1.5-km LAMEPS. Anomaly is still present after 1 h of integration. This poses a serious problem for 1-h data assimilation cycling as this anomaly will be transferred to the next cycle. Caron (2013) showed that is exactly what happens and, as a consequence, correlations between the host LAMEPS and 1.5-km LAMEPS decrease with time and the perturbation amplitudes between 1.5-km LAMEPS and host LAMEPS differ significantly (Fig. 11 in Caron, 2013). This shows that perturbation mismatching can pose a serious problem for LAMEPSs and needs to be solved before an operational implementation.

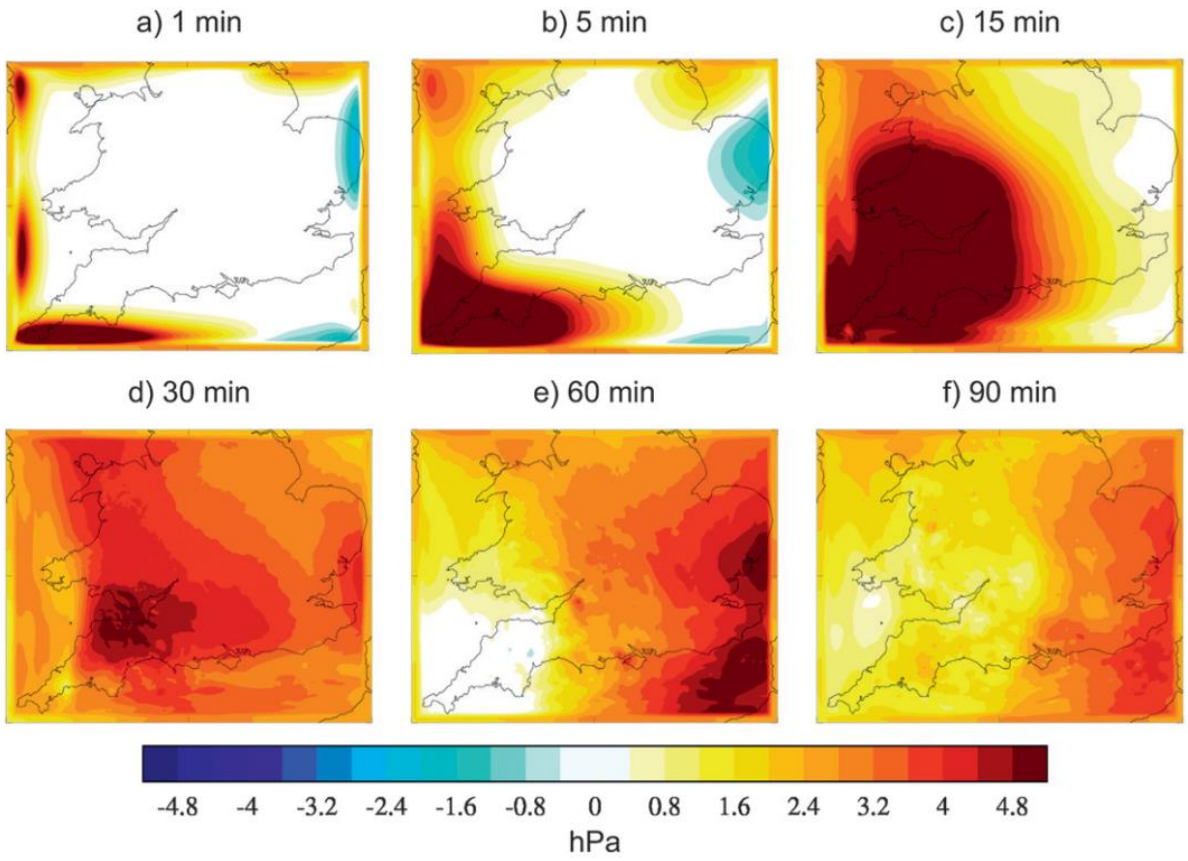

Fig. 1. Figure 9 from Caron (2013), the domain pertains to England and Wales and shows a time series of surface pressure perturbation (hPa) from ensemble member 8. The minutes shown above each panel refer to the forecast lead time with respect to 0430 UTC 5 Dec 2009.

1.3.2. Problem 2: large-scale representation deficiency

In addition to the problem 1, as discussed in Guidard and Fischer (2008; GF08 hereafter) another challenge of limited area modeling concerns the fact that LAMs are less effective at representing large-scale (e.g., synoptic) flow than global models (see also Hong and Kanamitsu, 2014; Hsiao *et al*., 2015; Kretschmer *et al.*, 2015; Verdesco *et al.*, 2016; Schwartz *et al.*, 2020). There are several reasons for this:

a) LAM domain is finite in size which limits its capability to capture and fully resolve large-scale phenomena (synoptic systems, Rossby waves, etc.).
b) LAM assimilation systems are often in a less advanced state of development then the global counterparts (e.g., Gustafsson *et al.*, 2018).

c) LAMs are designed for smaller-scale applications, with a greater emphasis on the mesoscale which can result in accumulation of a large-scale bias in time due to significant distortion of large-scale wave propagation (e.g., Feng *et al.*, 2020).
d) LBC formulation limitations cause atmospheric phenomena to be poorly propagated through the LAM domain (e.g., Warner *et al.*, 1997).
e) LAM analysis will not be able to use observational information that is located outside the domain (a fraction of the available observations is assimilated; see Gustafsson *et al*., 2018).

Issues of large-scale representation also reflect LAMEPSs in the sense that their IC perturbations, which are produced through techniques such as EDA, cannot correctly account for large-scale uncertainties (e.g., Bölöni, 2011). The simultaneous presence of small- and large-scale perturbations is beneficial for EPS (Wang *et al.*, 2014; Johnson and Wang, 2016; Raynaud and Bouttier, 2016). Furthermore, Schwartz (2019) demonstrated that having poorly resolved large scales can completely counteract the benefit of small grid size of a convection-permitting LAMEPS.

1.3.3. Solution: blending approach

Problem 2 can partially be addressed by applying a blending technique (Brožkovà *et al.*, 2001; Yang, 2005; Wang *et al*., 2014; Hsiao *et al*., 2015; Vendrasco *et al.*, 2016; Müller *et al.*, 2017) where large-scale information from the global model is merged with LAM small-scale information. Different authors have developed various approaches to the blending problem (see Feng *et al*., 2020 for a review of different blending approaches). Brožkovà *et al.* (2001) and Wang *et al.* (2014) proposed the use of a digital filter (DF) in the LAM and host model fields to obtain blended analysis. This approach can be further extended to involve variational data assimilation where blending is performed before or after variational data assimilation.

Caron (2013) and Wang *et al.* (2014) showed that DF-blending techniques can also be used to address the problem 1. Caron (2013) replaced the ETKF with, so-called, scale-selective ETKF (SSETKF). In the SSETKF, only small-scale perturbations are produced with ETKF, while large-scale perturbations are taken directly from the host EPS and blended together by using the DF, thus doing the DF-blending. It was demonstrated that spurious surface pressure perturbations are completely gone after the ETKF was replaced by the SSETKF (Fig. 7 in Caron, 2013) and that blending of only the largest wavelength was enough to achieve that result. Wang *et al.* (2014) used the DF-blending method under the framework

of their ALADIN-LAEF (Aire Limitee Adaptation Dynamique Developpement InterNational – Limited-Area Ensemble Forecasting) system (Wang *et al.*, 2011) to combine IC perturbations generated through the breeding method with global ECMWF-EPS (Buizza *et al*., 2008) IC perturbations. DF-blending was used to blend small-scale ALADIN-LAEF and large-scale ECMWF-EPS IC perturbations to simultaneously introduce them into perturbed ICs. They achieved better upper-air scores and increased ensemble spread by applying the blending method rather than the breeding method without blending.

Johnson and Wang (2016) argue that it is important for large-scale perturbations to be consistent with LAMEPS analysis errors, while Vendrasco *et al.* (2016) show that small and large scales need to be consistent in the final analysis. This is not the case with the DF-blending since large-scales come from a completely different model and are independent of LAMEPSs small scales or analysis errors. Furthermore, as GF08 discusses, DF-blending approach presents unique problems. Most notably, it has no relation to data assimilation theory, and there have been concerns regarding its optimality in the sense of minimum variance estimate. Thus, they proposed a method which includes global model information directly into a limited area variational analysis (sub-section 2.1.) and implemented it in ALADIN (Aladin IT, 1997; Termonia *et al.* 2018) and Action de Recherche Petite Echelle et Grande Echelle (ARPEGE) framework within ALADIN 3D-Var. We refer to this as the *Jk* blending method. Doing the blending inside the 3D-Var will result in the optimality of the blended analysis where small and large scales will be consistent. GF08 demonstrated a small but positive impact on objective and subjective scores. Dahlgren and Gustafsson (2012) explored this method further within a HIgh Resolution Limited Area Model (HIRLAM) and ECMWF framework of the HIRLAM 3D-Var and 4D-Var. They demonstrated a positive impact on forecast accuracy of surface and upper-air variables (Fig. 2).

Large-scale information can also be included through the use of dual-resolution "hybrid" variational-ensemble data assimilation methods (e.g., Schwartz, 2016; Wu *et al.*, 2017; Hu *et al.*, 2017). Such methods incorporate ensemble-derived, flow-dependent background error covariances drawn from a coarser resolution model into a high-resolution deterministic analysis. As opposed to the *Jk* blending method, which includes large-scale information as a new member in the 3D-Var, hybrid variational-ensemble data assimilation methods execute this through a background error covariance matrix. Schwartz (2016) demonstrated improved forecasts started from hybrid analyses.

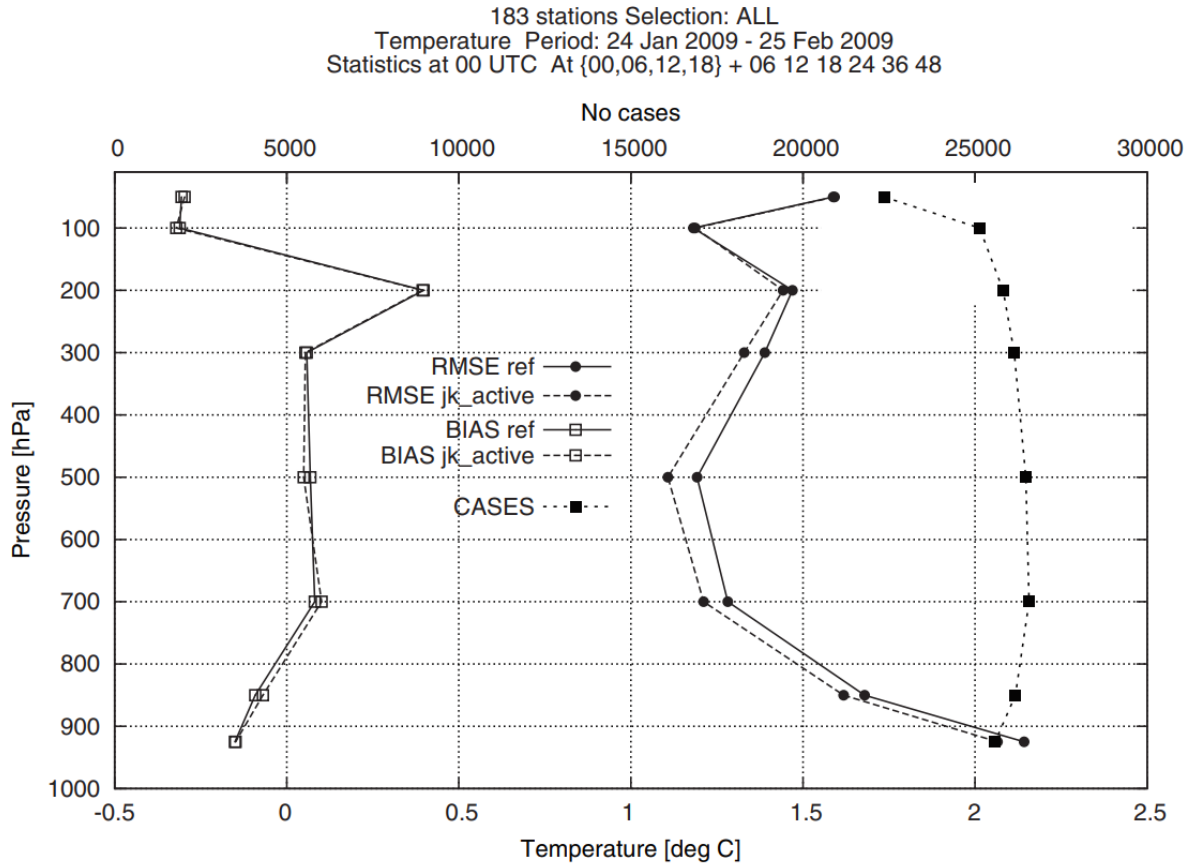

Fig. 2. Figure 12. from Dahlgren and Gustafsson (2012). Verification of temperature forecasts compared with observations from radiosondes. To the right is the root mean square error and to the left the mean error (bias) averaged over all forecast ranges, valid at 00 UTC. Full line: reference experiment (i.e., without *Jk*). Dotted line: reference with *Jk* blending. The number of observations used in the statistics is also shown using the upper x-axis.

1.3.4. Problem 3: grid size vs model's true resolution

It is important to understand that the model grid size is not the same as the model resolution (e.g., Grasso, 2000). The second is sometimes referred to as the model effective resolution and is generally, at least, 5 times lower than the first (e.g., Pielke, 2002; Skamarock, 2004; Horvath *et al*., 2011; Mittermaier, 2014). Therefore, all point predictions within that area (i.e., neighborhood of $5\Delta x \times 5\Delta x$) should be considered equally likely and the output of the model should be viewed as the spatial and (or) temporal function of that neighborhood. Adding this to the fact that small scale errors saturate in a few hours, we realize that one cannot expect that a convection-permitting model forecast exactly matches the observations on the grid scale

(Theis *et al.*, 2005; Mittermaier, 2014). Furthermore, double-penalty effect (Mass *et al.*, 2002; Mittermaier, 2014) makes this problem worse, because it penalizes the forecast twice - displacement errors get penalized (increasing the false alarm rate), and closeness is not rewarded (decreasing the hit rate).

1.3.5. Solution: neighborhood approach

In order to alleviate difficulties regarding the rapid error growth at the small scales and the problem 3, neighborhood methods (Theis *et al.*, 2005; Ebert, 2008; Ben Bouallegue *et al.*, 2013; Mittermaier, 2014; Schwartz and Sobash, 2017, among others) were developed for:

a) The use in the forecast verification as spatial verification methods (see sub-section 3.3.) where they generally share a common trait of relaxing the traditional requirement that forecast and observed events exactly match at the grid scale to account for observation and model uncertainties (Schwartz and Sobash, 2017). For example, looking at the neighborhood around a desired point, we can easily spot a convective storm that has passed a few kilometers away and we would not notice it if we only look at the forecast for the desired point.
b) To extend an EPS by increasing the number of its members and (or) to provide a way to calculate ensemble probabilities which better reflect the model's true resolution (sub-section 8.1.1.).

The concept of the neighborhood is illustrated in Fig. 3. following Theis *et al.* (2005). Left-hand side of the figure shows (x, y)-plane of the model grid. Shaded area denotes a 5 × 5 neighborhood of the point (x_0, y_0) (shown in red), while Δx and Δy denote the model grid size in x and y direction, respectively. Total number of grid points inside a neighborhood (N_b) is obtained by choosing a neighborhood length scale r which can be either a number of grid boxes or physical distance and can be applied using square or circular neighborhood geometry. For example, left-hand side of Fig. 3 shows a square $r \times r$ neighborhood where N_b ($= r^2$) = 25 grid points. In addition, N_b can also be extended to a time dimension (right-hand side of the Fig. 3), i.e., (x, t)-plane where Δt denotes the time step between successive model output times. How probabilities are calculated from the neighborhood, or how to combine an EPS with the neighborhood approach, is described in the sub-section 8.1.1. In contrast to all the studies cited here, we will not increase r beyond the theoretical values for the true model resolution (i.e., 5-7Δx), because we want to maximize the resolution of the output forecasts and keep them close to the convection-permitting range.

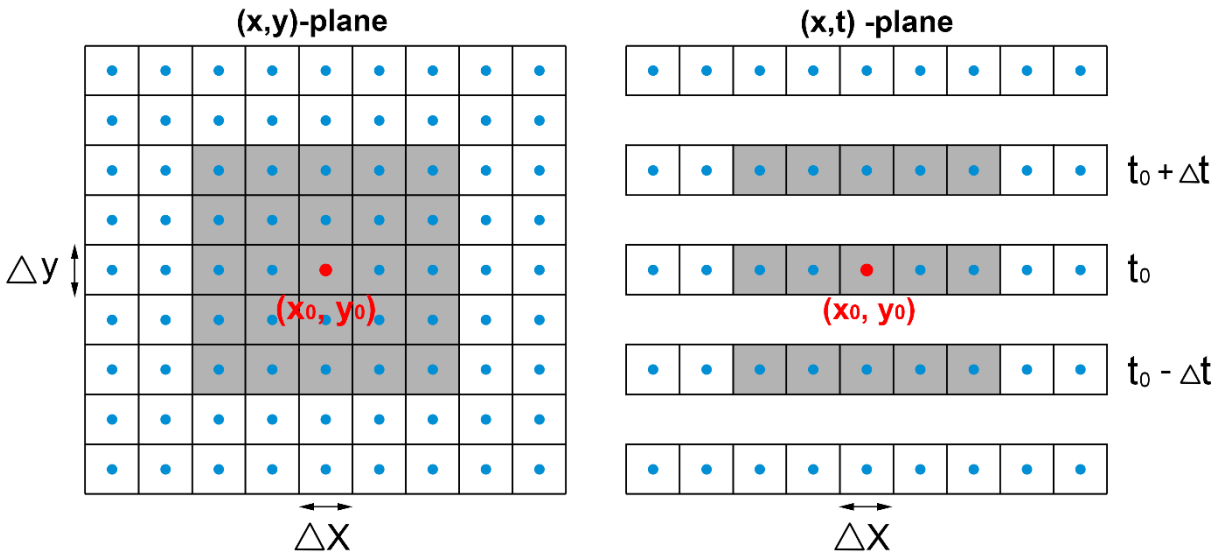

Fig. 3. Example of a square spatio-temporal neighborhood of a given grid point (red) at location (x_0, y_0) and forecast lead time t_0. Left: the spatial neighborhood in the (x, y)-plane. Right: the spatio-temporal neighborhood in the (x, t)-plane. Δx and Δy denote the size of a grid box and Δt denotes the time step between successive model output times. Shaded grid boxes belong to the neighborhood.

1.4. Objectives

At Zentralanstalt für Meteorologie und Geodynamik (ZAMG), the LAMEPS ALADIN-LAEF has been developed within the framework of Regional Cooperation for Limited Area Modelling in Central Europe (RC-LACE; Wang *et al.*, 2018) and has been used in operations since 2009. It is based on ALADIN model and is currently run at 11 km horizontal grid size with a planned upgrade to 5 km in early 2020. However, with ever-increasing levels of computer power, it is now possible to run LAMEPSs at convection-permitting resolutions corresponding to a grid size of 1-4 km (e.g., Vié *et al.*, 2011; Bouttier *et al.*, 2012; Nuissier *et al.*, 2012; Romine *et al.*, 2014; Kühnlein *et al.*, 2014; Schwartz *et al.*, 2015; Johnson and Wang, 2016; Hagelin *et al.*, 2017; Schwartz and Sobash, 2019). For this reason, at ZAMG, a new convection-permitting LAMEPS named C-LAEF (Convection-permitting Limited Area Ensemble Forecasting) based on the Application of Research to Operations at Mesoscale (AROME) model (Seity *et al.*, 2011; Termonia *et al.*, 2018) is being developed.

The main objective of this dissertation is:

- Develop and implement an IC perturbation method that will try to solve problems 1 and 2. This method is planned to be used in a first operational configuration of C-LAEF. The main idea is as follows. We propose employing the *Jk* blending method within the C-LAEF, e.g., within the framework of its 3D-Var EDA system (Bouttier *et al.*, 2012) to generate an ensemble of analyses. We refer to this as an ensemble *Jk* method. In such a system, small-scale perturbations are generated through 3D-Var EDA, while large-scale perturbations are generated from the host EPS via *Jk* blending. In turn, we hypothesize that final analyses are optimal, contain perturbed small and large scales which are, at the same time, consistent mutually and with perturbations coming from lateral boundaries. In this way, one can alleviate problems of large-scale forcing, inconsistencies between ICs and LBCs and inconsistencies between small and large scales.

The secondary objective of this dissertation is to utilize relatively cheap and readily available techniques to improve C-LAEF performance:

- Implement LAF to extend C-LAEF IC perturbations with the information from past forecasts in order to improve initial uncertainty estimate.
- Implement neighborhood approach within C-LAEF system to account for the problem 3 and as an additional tool to counter the issue of rapid error growth at the small scales.

Successful completion of these objectives will result in the improvement of convection-permitting LAMEPS's forecasts, detection of extreme weather events and a potential benefit for issuing different types of warnings. Additionally, LAF, neighborhood approach and *Jk* blending can also be employed to deterministic models, such as ALADIN which is used at Croatian Meteorological and Hydrological Service (Tudor *et al.*, 2013).

§ 2. A NOVEL METHOD OF PERTURBING INITIAL CONDITIONS IN A LAMEPS

2.1. *Jk* blending method

We now present basic formalisms of the *Jk* blending method. For a more detailed description, the reader is referred to GF08. 3D-Var is a variational method where, in order to obtain optimal analysis $\boldsymbol{X^a}$, a so-called cost function $J(\mathbf{X})$ that measures the distance of the LAM state vector $\boldsymbol{X}$ to observations $\boldsymbol{Y}$, and to the LAM background $\boldsymbol{X^b}$, is minimized (sub-section 1.1., equations (1.4) - (1.6)). When we assume that host model errors are uncorrelated with observation and background errors, an extra term $J_k(\mathbf{X})$, measuring the distance from large-scale information in $\boldsymbol{X}$ to the host model large-scale information $\boldsymbol{X^k}$, can be added to equation (1.4):

$$J(\mathbf{X}) = J_b(\mathbf{X}) + J_o(\mathbf{X}) + J_k(\mathbf{X}) \tag{2.1}$$

$$J_k(\mathbf{X}) = \frac{1}{2}\left(H_1\mathbf{X} - H_k\mathbf{X^k}\right)^{\mathrm{T}}\mathbf{V}^{-1}\left(H_1\mathbf{X} - H_k\mathbf{X^k}\right) \tag{2.2}$$

where $\boldsymbol{V} = E(\varepsilon_k \varepsilon_k^T)$ is the host model large-scale error (ε_k) covariance matrix, H_1 is the operator which fetched $\boldsymbol{X}$ from a nominal high-resolution to a low-resolution LAM space, and H_k is the operator which fetches $\boldsymbol{X^k}$ from the host model low-resolution space to the same low-resolution LAM space as H_1. This is done to include only large scales form the host model (e.g., large-scale constraint) as discussed in the sub-section 1.3., H_k consists of two steps: the interpolation of host model fields to a high-resolution LAM geometry and its truncation to a lower resolution (see sub-section 4.2.), such that only large scales are affected during analysis.

We applied the *Jk* blending method to the AROME 3D-Var system (Fischer *et al.*, 2005), in which $\boldsymbol{X}$ is the state vector of 5 analysis variables (as is $\boldsymbol{X^k}$): temperature, vorticity, divergence, specific humidity and the logarithm of surface pressure. A large-scale constraint can be applied to all of these or to only some. Our choices regarding this option are discussed in sub-section 4.2.

As stated in GF08, correlations between host model and background errors are different from zero as a result of lateral boundary coupling. However, when one accepts

inaccuracies of roughly 10 %, such correlations can be disregarded, and the total cost function can be represented as the sum of three parts as shown by (2.1). In this study, it is assumed that this assumption also holds for our model configuration.

2.2. Ensemble *Jk* method

For the IC perturbation method of C-LAEF, we introduce a new method - the ensemble *Jk* method which combines the ensemble of 3D-Var with the *Jk* blending method. As discussed in the introduction, one must obtain consistent IC and LBC, which is crucial if LAMEPS is integrated over a small domain. To combine large-scale perturbations drawn from the host EPS (the global EPS in this study), which provides the LBC for the LAMEPS, with the small-scale perturbations from LAMEPS, we propose an ensemble data assimilation blending technique. The approach is described as follows:

a) 3D-Var EDA is used to generate perturbed analyses by:
    i) Random observation perturbations designed to simulate observation errors. For each observation, its error is taken from an observation database and is multiplied by a normally distributed random number with a zero mean and a unit standard deviation (white noise). This value is then added to the original observation value to obtain a perturbed observation. Perturbed-observation EDA perturbations are advantageous in that they are designed to estimate analysis uncertainty (e.g., Wei *et al.*, 2008) which is a desirable property of IC perturbations. A theoretical justification of this approach is given in Žagar *et al.* (2005), while Burgers *et al.* (1998) show that perturbing observations does sample from the correct distribution in the case that observation errors are Gaussian and the system is linear.
    ii) Background perturbations from short-range forecasts of the LAMEPS. Each EDA analysis uses a forecast from different member of LAMEPS to provide a background information.

    For the time being, model errors inside the EDA itself are not considered, but they can have a beneficial impact on ensemble spread and reliability (e.g., Cardinali *et al.*, 2014).
b) To generate correct large-scale information in perturbed analyses that are consistent with perturbed LBCs, we introduce a large-scale constraint by applying *Jk* blending to

3D-Var EDA. In other words, perturbed global analyses were incorporated as a new term of the 3D-Var cost function and were assimilated with observations.

In brief, each EDA member assimilates the following three components:

a) Perturbed observations
b) Perturbed background
c) Perturbed global analysis

Mathematically, the cost function of one ensemble member of the ensemble *Jk* method can be written as:

$$J_{\mathrm{i}}(\mathbf{X}_{\mathrm{i}}) = \frac{1}{2}\left(\mathbf{X}_{\mathrm{i}} - \boldsymbol{X}_i^b\right)^{\mathrm{T}}\mathbf{B}^{-1}\left(\mathbf{X}_{\mathrm{i}} - \boldsymbol{X}_i^b\right) + \frac{1}{2}(\mathbf{y}_{\mathrm{i}} - H\mathbf{X}_{\mathrm{i}})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}_{\mathrm{i}} - H\mathbf{X}_{\mathrm{i}}) + \frac{1}{2}\left(H_1\mathbf{X}_{\mathrm{i}} - H_k\boldsymbol{X}_i^k\right)^{\mathrm{T}}\mathbf{V}^{-1}\left(H_1\mathbf{X}_{\mathrm{i}} - H_k\boldsymbol{X}_i^k\right) \tag{2.3}$$

where subscript *i* denotes the ensemble member. Final perturbed analyses are then found by minimizing these cost functions as in the standard 3D-Var. The resulting blended analyses include small-scale perturbations (LAMEPS) and large-scale perturbations (global EPS) which are consistent with one another, while also being consistent with LBC perturbations. We believe that the ensemble *Jk* method can alleviate the problem of IC and LBC perturbation mismatch and improve the representation of large-scale uncertainties in ensemble systems. In addition, it should alleviate the problem of inconsistencies between small and large scales usually present in DF-blending approaches.

More information on **V**, resulting perturbations, truncation and tuning of the ensemble *Jk* method is given in section 4.

§ 3. EXPERIMENTAL DESIGN

3.1. C-LAEF configuration

The C-LAEF system is based on the convection-permitting AROME model as it was operationally configured at ZAMG. AROME is a high-resolution limited area spectral non-hydrostatic model. It solves non-hydrostatic fully compressible Euler equations system (Bubnova *et al.,* 1995). A two-time level, semi-implicit, semi-Lagrangian discretization scheme on Arakawa A-grid (Arakawa and Lamb, 1977) is used. The vertical discretization is based on finite differences with the use of hybrid pressure terrain-following coordinate (Simmons and Burridge, 1981). Most of the prognostic variables have a spectral representation based on a double Fourier decomposition (Haugen and Machenhauer, 1993). It takes most of the ALADIN code on the adiabatic part while its physics package is mainly an adaptation of that used in Mesoscale Non-Hydrostatic research model Meso-NH (e.g., Lafore *et al.*, 1998), and it is well adapted to a small grid size of roughly 1-2 km (Termonia *et al.*, 2018). In addition, AROME 3D-Var system is almost identical to that developed for the ALADIN system (Fischer *et al.*, 2005). A more detailed description of AROME dynamics, physics and assimilation properties can be found in Seity *et al.* (2011), Vié *et al.* (2011) and Brousseau *et al.* (2016).

In the present study, C-LAEF is configured by AROME model with a horizontal grid size of 2.5 km and 90 vertical levels with 17 ensemble members (16 perturbed plus control). Table 1 summarizes the general C-LAEF settings. The integration domain used is shown in Fig. 4. Due to high computational costs related to C-LAEF, the domain is relatively small at 1080 km in the north/south direction and 1500 km in the east/west direction. The LBCs are provided by the ECMWF-EPS, and lateral boundary coupling is conducted using Davies method (Davies, 1976) with an 8-grid point relaxation zone every 3 h. Our choices regarding the IC perturbations are described in the next sub-section, and for LBC perturbations, 17 members of the ECMWF-EPS are used. Due to technical limitations, the same ECMWF-EPS members used for ALADIN-LAEF (the first 16 plus control members) are also used for C-LAEF. As this study's emphasis is on IC uncertainties, model errors are not considered and no model error representation scheme is used. The same is true for lower boundary conditions (i.e., surface variables; e.g., land albedo, soil moisture, soil temperature, etc.) which are not

perturbed. We would like to emphasize that impact of model perturbations on C-LAEF was explored in Wastl *et al.* (2019) who reported a beneficial impact on ensemble spread.

Here, 3D-Var is used for data assimilation. Conventional observations including AMDAR, SYNOP, PILOT, TEMP, SHIP and EUROPROFILERS plus two satellite products GEOWIND atmospheric motion vectors and 25 km advanced Scatterometer ocean winds, are assimilated. In this testing phase, 6-h continuous assimilation cycles applied at 0000, 0600, 1200 and 1800 UTC are performed.

C-LAEF prognostic variables, parameterization schemes and diffusion used in this study are as follows:

a) *Prognostic variables*: 12 3D prognostic variables. Two components of the horizontal wind, temperature, specific humidity, rain, snow, graupel, cloud droplets, ice crystals, turbulence kinetic energy (TKE), two non-hydrostatic variables – pressure departure and vertical divergence and one 2D variable – hydrostatic surface pressure.
b) *Microphysics*: an upgraded 3-class ice parameterization (Pinty and Jabouille, 1998) which is coupled to a Kessler scheme for warm processes.
c) *Turbulence*: prognostic TKE combined with a diagnostic mixing length. TKE scheme was developed by Cuxart *et al.* (2000) and uses the Bougeault and Lacarrere (1989) mixing length.
d) *Surface*: an externalized version of the Meso-NH surface scheme, called Externalized Surface (SURFEX) model (Le Moigne *et al.*, 2018). SURFEX simulates the exchange of energy and water between the atmosphere and surface. 4 types of parameterizations, depending on surface tiles, are used: ISBA (Noilhan and Planton, 1989) over land tiles with 1-layer snow scheme (Douville *et al.*, 1995), TEB model (Masson, 2000) over urban areas, iterative ECUME algorithm (Belamari and Pirani, 2007) over oceans and for inland water, classic Charnock's (1955) formulation is used.
e) *Radiation*: the shortwave radiation scheme (Fouquart and Bonnel, 1980) uses six spectral bands. Cloud optical properties are derived from Morcrette and Fouquart (1986) for liquid clouds and Ebert and Curry (1992) for ice clouds. Longwave radiation is computed by the Rapid Radiative Transfer Model (Mlawer *et al.*, 1997).
f) *Convection*: deep convection is assumed to be explicitly resolved by the model's dynamics. For shallow convection the Eddy-Diffusivity-Kain-Fritsch scheme developed by Pergaud *et al*. (2009) is used.

g) *Diffusion*: the spectral part of the diffusion is a fourth-order linear scheme, with the same strength for each spectral prognostic variable. For non-spectral variables, a so-called semi-Lagrangian horizontal diffusion nonlinear scheme (Váňa *et al.*, 2008) is applied. The water vapor and TKE variables are not diffused at all.

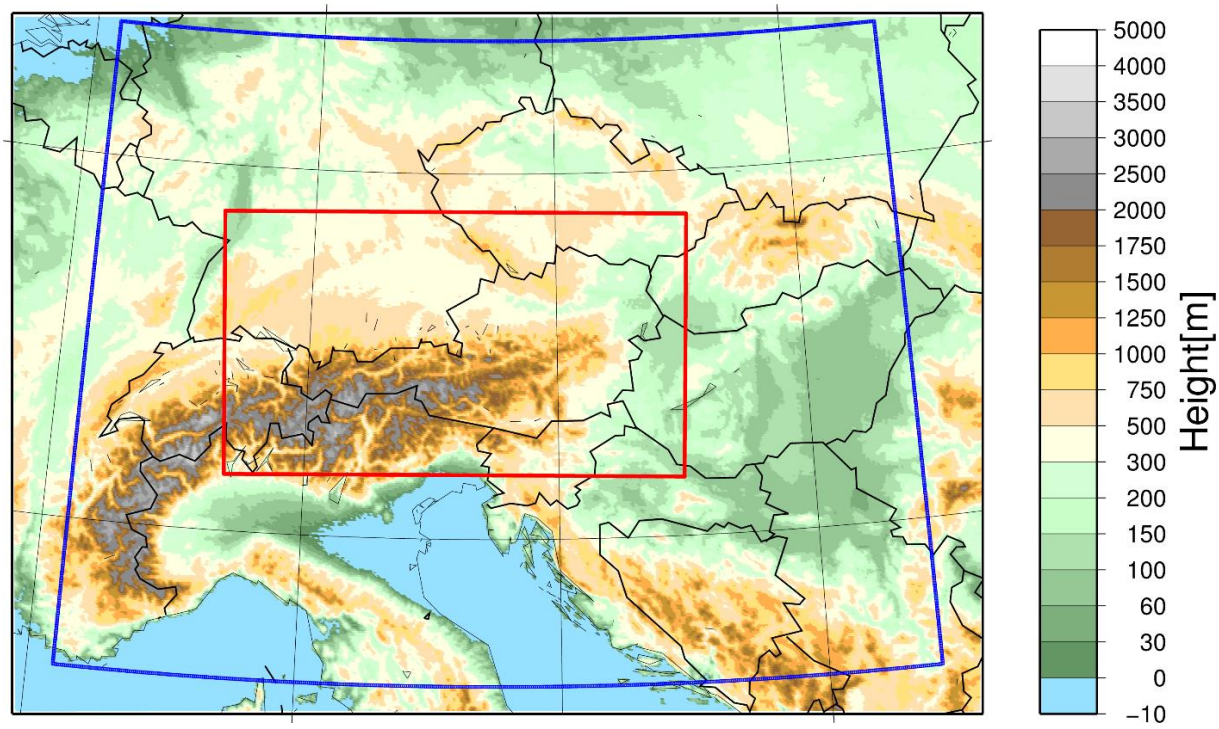

Fig. 4. C-LAEF computational domain and orography with INCA domain outlined in red and the verification domain shown in blue.

3.2. Experiments

The aim here is to evaluate the ensemble *Jk* method and to assess its added value to the standard perturbed-observation EDA (e.g., Houtekamer *et al.*, 1996; Bouttier *et al.*, 2012), i.e., EDA without the J_k term in (2.3). For this reason, two experiments are conducted:

a) REF – a reference experiment in which C-LAEF is using perturbed-observation EDA as an IC perturbation method.

b) JK – experiment where C-LAEF uses the ensemble *Jk* method for the same purpose.

Table 1. C-LAEF settings used in this study. See text for more details.

OPTION	VALUE
Grid size	2.5 km
Number of vertical levels	90
Number of members	16 + control
Data assimilation	3D-Var, 6-h cycling
Forecasts	24-h, 1200 UTC only
IC perturbations	Ensemble *Jk* or EDA
LBC perturbations	ECMWF-EPS
Model perturbations	None
Surface perturbations	None

Both experiments use the same configuration as the one described in sub-section 3.1. and shown in Table 1. Assimilation is performed on the full model resolution for 17 different ensemble members. No re-centering of EDA or ensemble *Jk* perturbations is performed as this has a negative impact on the ensemble performance (Lang *et al.*, 2015) and all forecasts are started directly from EDA/ensemble *Jk* members.

An additional experiment (DOWN) is performed for the period studied in section 7. In DOWN, perturbed analyses were not generated by EDA or by the ensemble *Jk* method but by the downscaling of ECMWF-EPS analyses. This makes them fully consistent with LBC perturbations and DOWN serves as a reference experiment in section 7. Comparison of perturbed-observation EDA against the downscale experiment in terms of traditional scores is presented in Raynaud and Bouttier (2016) and will not be repeated here. They concluded that EDA "significantly improves the AROME EPS performance for surface weather variables at early ranges, namely up to 12 h depending on the variable, as measured by the spread/skill relationship and the CRPS over a one-month period."

In JK experiment, global ECMWF-EPS analyses are used as a source of large-scale information in the ensemble *Jk* method. Horizontal resolution of the ECMWF-EPS in our experiments is spectral cubic $T_{CO}640$ with 91 vertical levels. The ECMWF-EPS uses singular vectors in combination with EDA for IC perturbation generation (e.g., Buizza and Leutbecher, 2015) and resulting perturbations are completely independent from ours.

It might be necessary to emphasize that the same host EPS (ECMWF-EPS in this study) must be used to provide all the necessary information for LAMEPS, i.e., LBC information, global analyses for the ensemble *Jk* method and **V** matrix calculation (see sub-section 4.1. for details on how **V** is calculated).

3.3. Verification

The JK and REF experiments described above are verified in 3-h intervals for surface variables, i.e., 2 m temperature (T2M), 2 m relative humidity (RH2M), 10 m wind speed (W10M) and mean sea level pressure (MSLP) against observations drawn from 832 available surface stations within the verification domain (blue area in Fig. 4). Model fields are interpolated via bilinear interpolation to observation locations. For upper-air variables, verification is done every 6 h against high-resolution deterministic ECMWF (0.1° × 0.1°) and deterministic NCEP analyses (0.5° × 0.5°). This approach is applied because the ensemble *Jk* method uses perturbed ECMWF analyses; thus, verification against deterministic ECWMF analyses can be biased toward JK experiment. Verifying variables include temperature, relative humidity and wind speed at 500 and 850 hPa pressure levels (T500, T850, RH500, RH850, W500 and W850, respectively).

Verifying precipitation forecasts in convection-permitting models has its difficulties and traditional verification methods may not be appropriate (Ebert, 2008; Mittermaier, 2014). The main reasons for this are: a) grid averaged values (model forecast) cannot simply be compared to the point values (observations; Göber *et al.*, 2008), as representativeness error can be as high as 50 % of the total error (Haiden *et al.*, 2012), especially given the fact that model effective resolution is at least 5 times higher than its grid size (sub-section 1.3.4.); b) it has been shown in the past that point verification is unable to show benefit of increasing model resolution for grid sizes below about 10 km (Mass *et al.*, 2002) and c) it is very sensitive to the double-penalty effect. This had led to the development of so-called spatial verification techniques (Gilleland *et al.*, 2009). The spatial evaluation accounts for structure errors, spatial displacements and field deformations. Through different approaches (i.e., scale separation, neighborhood, object- or feature-based approach, etc.) spatial verification reduces the double-penalty effect and is able to more correctly compare different resolution models. Although, most of them assumes gridded observations are available, techniques to verify against point observations have also been developed (Ebert, 2008; Mittermaier, 2014).

In this study, we use Fractions Skill Score (FSS; Roberts and Lean, 2008), which is a form of a neighborhood approach (see sub-section 8.1.), and it is computed following the ensemble formulation proposed by Duc *et al.* (2013). INCA (Integrated Nowcasting through Comprehensive Analyses; Haiden *et al.*, 2011, Wang *et al.*, 2017) analyses are used as a reference for precipitation verification. The INCA system is currently being developed at ZAMG, and it uses data drawn from automatic weather stations, remote sensing data (radar, satellite), forecast fields of NWP models, and high-resolution topographic data (Haiden *et al.*, 2011) to generate a high-resolution precipitation analysis field for a domain covering Austria and surroundings (red quadrangle on Fig. 4). C-LAEF forecasts are interpolated to the high-resolution INCA grid (1 km grid size) as this is the recommended practice for our purpose (see Wolff *et al.,* 2014).

Good EPSs should generate superior ensemble mean forecasts over the control, high spread-skill relations and reliable probability forecasts (Du, 2007). Furthermore, Murphy (1993) described aspects or attributes that contribute to forecasting quality. We will now give a brief description of attributes used in this study and the reader is referred to Murphy (1993) or Wilks (2006) for a more detailed description.

a) *Accuracy* – the correspondence between individual forecasts and the events they predict.
b) *Bias* – the correspondence between the average forecast and the average observed value of the predictand.
c) *Reliability* – the average agreement between the forecasted probability of an event and the observed frequency of that event. In other words, if an event was forecasted with an average probability of 40%, the observed probability of this event should also be 40 %.
d) *Resolution* – the degree to which the forecasts sort the observed events into groups that are different from each other. This means that the observed distribution when "A" was forecast is different from the observed distribution when "B" is forecast. It is conditioned on forecasts. For example, if observed temperature distribution following forecasts of, say, -10 °C and -20 °C are very different, the forecasts can resolve these different temperature outcomes, and are said to exhibit resolution.
e) *Discrimination* – ability of the forecast to distinguish between occurrences and non-occurrences by forecasting different probabilities before occurrence of an event than before non-occurrence of this event. It is the opposite to the resolution

in the sense that it is conditioned on observations. For example, we compare forecasted distributions for all observed events of -10 °C and -20 °C. Discrimination is the measure of difference between those two forecasted distributions.

f) *Uncertainty* – the variability of the observations. Does not depend on forecasts.

g) *Sharpness* – the variability of the forecasts. Does not depend on observations. Forecasts that rarely deviate much from the climatological value of an event exhibit low sharpness. By contrast, forecasts that are frequently much different from the climatological value of an event are sharp.

By considering the following example, it is easy to see why only one attribute is not enough to characterize a probabilistic forecast. If we forecast 10 events whose climatological frequency of occurrence is 30 %, we do not want to give 10 forecast all with 30 % probability of occurrence (perfect reliability, but no resolution), but we want to give 7 forecasts with 0 % probability and 3 forecasts with 100 % probability (perfect reliability and resolution).

Given these sources of complexity, no single score can assess all attributes of probabilistic forecasts at once. For this reason, we use various scores to assess different aspects of forecasts and to draw conclusions about the EPSs under consideration. These include a) the root mean square error (RMSE) of ensemble mean for assessing ensemble mean accuracy, b) the continuous rank probability score (CRPS) for assessing overall EPS skill, c) ensemble RMSE/spread relation and outlier statistics for assessing reliability, d) decomposition of the Brier score for assessing accuracy of ensemble probabilities, reliability and resolution, and e) relative operational characteristics (ROC) for assessing discrimination. Before each score is presented in section 5., a brief description and mathematical formulation is provided. However, more detailed descriptions of these scores can be found in Talagrand *et al.* (1997) and Wilks (2006).

To determine if the difference in scores between the experiments is statistically significant, the moving-block bootstrap technique, following the procedure of Wilks (1997) and using 5000 resamples at a confidence level of 90 %, was applied. Block length (4 in our case) was defined as a closest integer to a cube root of the sample length (Hall *et al.*, 1995). Observation errors are not considered, as this extends beyond the scope of this study. Nevertheless, as observation errors can have significant effect on final scores (e.g., Bowler, 2008; Candille and Talagrand, 2008; Hacker *et al.*, 2011; Duc and Saito, 2018; Ben Bouallegue *et al.*, 2020); we discuss this limitation in section 5.7.

§ 4. DIAGNOSTICS AND TUNING

4.1. V matrix

As illustrated by equations (2.2) and (2.3), in addition to **B** and **R** error covariance matrices, a **V** error covariance matrix must also be provided. This is an important step because **V** controls the impact of large-scale constraint on the final analyzed state. To see how this is the case, suppose we do not have any observations. Gradient of the cost function (equation 2.2), in that case, takes the following form:

$$\nabla J(\boldsymbol{X}^a) = \boldsymbol{B}^{-1}(\boldsymbol{X}^a - \boldsymbol{X}^b) + \boldsymbol{V}^{-1}(\boldsymbol{X}^a - \boldsymbol{X}^k) = 0 \tag{4.1}$$

Where we assumed that $H_1 = H_k = 1$ because there are no observations. Further, by rearranging:

$$\boldsymbol{X}^a - \boldsymbol{X}^k = \boldsymbol{V}\boldsymbol{B}^{-1}(\boldsymbol{X}^b - \boldsymbol{X}^a) \tag{4.2}$$

Now, let us assume that we only take background information at one model grid point *j*. In this case $\boldsymbol{B}^{-1}(\boldsymbol{X}^b - \boldsymbol{X}^a)$ is just one number, say *S*. Thus:

$$\boldsymbol{X}^a - \boldsymbol{X}^k = S \begin{bmatrix} V_{1j} \\ \cdots \\ V_{N_m j} \end{bmatrix} \tag{4.3}$$

Therefore, analysis increment is proportional to a column vector of **V**. In other words, the impact of the large-scale information is spread from the *j*th point to all other N_m model points by **V**. **V** assures that statistically consistent increments are also created at the neighboring grid points and levels of the model for this and all other model variables. The same is true for **B**, which can be seen if this exercise is repeated, so that large-scale information is taken only for one model grid point, while keeping vector $\boldsymbol{X}^b$ complete.

In calculating **V**, we used the same procedure normally applied to **B**, for example, the ensemble simulation method (Berre *et al.*, 2006), resulting in **V** which is climatological and does not contain any flow dependent information. To obtain climatological host model large-scale error covariances using the ensemble simulation method, analyses differences between 8 pairs of ECMWF-EPS members need to be calculated. This calculation is done for 0000 and 1200 UTC runs over two-week periods in January, April, July and October of 1 year to include seasonal variability of error covariances, as this can have a significant impact on

forecast quality (Storto and Randriamampianina, 2010). As in GF08, **V** is formulated in a univariate manner and the cross-covariances among errors of different model variables are left for **B** to handle.

To check if **V** is calculated correctly and behaves expectedly, we present a few basic diagnostics. Fig. 5 shows the horizontal error variance spectrum (Berre, 2000) of **V** and **B** matrices for two levels (roughly 500 and 920 hPa) and variables (divergence and specific humidity) as a function of a total wave number k^* (Berre, 2000):

$$k^* = N_x \sqrt{\left(\frac{k}{N_x}\right)^2 + \left(\frac{l}{N_y}\right)^2} \tag{4.4}$$

where k (l) is a wavenumber in x (y) direction and N_x (N_y) domain size in x (y) direction in km. Note that **V** comes from ECMWF-EPS and **B** comes from C-LAEF. It can be seen that, after $k^* \sim 15$, **V** spectra starts to decrease exponentially. Values beyond that are unphysical and are just an artifact of interpolation of ECMWF-EPS analyses to the C-LAEF geometry. Given the fact that grid size of ECMWF-EPS analyses used to calculate **V** was 32 km, *Nx* is equal to 1080 km and *Ny* is 1500 km, a value $k^* \sim 15$ is correct.

Additionally, Fig. 6 shows horizontally averaged vertical correlations of temperature and vorticity for the model level 32 (around 500 hPa). Those correlations tell us how an analysis increment is spread to the other model levels by **V**. Correlations in Fig. 6 can be compared to the correlations for **B** computed over Europe for ECMWF IFS HRES (Integrated Forecast System High RESolution; ECMWF, 2019) model in Pereira and Berre (2006; their Fig. 14) or for LAM in Stanešić *et al.* (2019; their Fig. 5). The comparison reveals that **V** is calculated correctly. Negative temperature correlations can be explained by the vertical structure of the upper-level troughs in the mid-latitudes, where upper-level tropospheric cold air is associated with a decrease of the tropopause and a warming of the air above it (Stanešić *et al.*, 2019), in accordance with conceptual models od upper-level dynamical processes (e.g., Hoskins, 1985).

4.2. Tuning and truncation of the ensemble *Jk* method

In the ensemble *Jk* method, we test several means of tuning the impact of large-scale constraint:

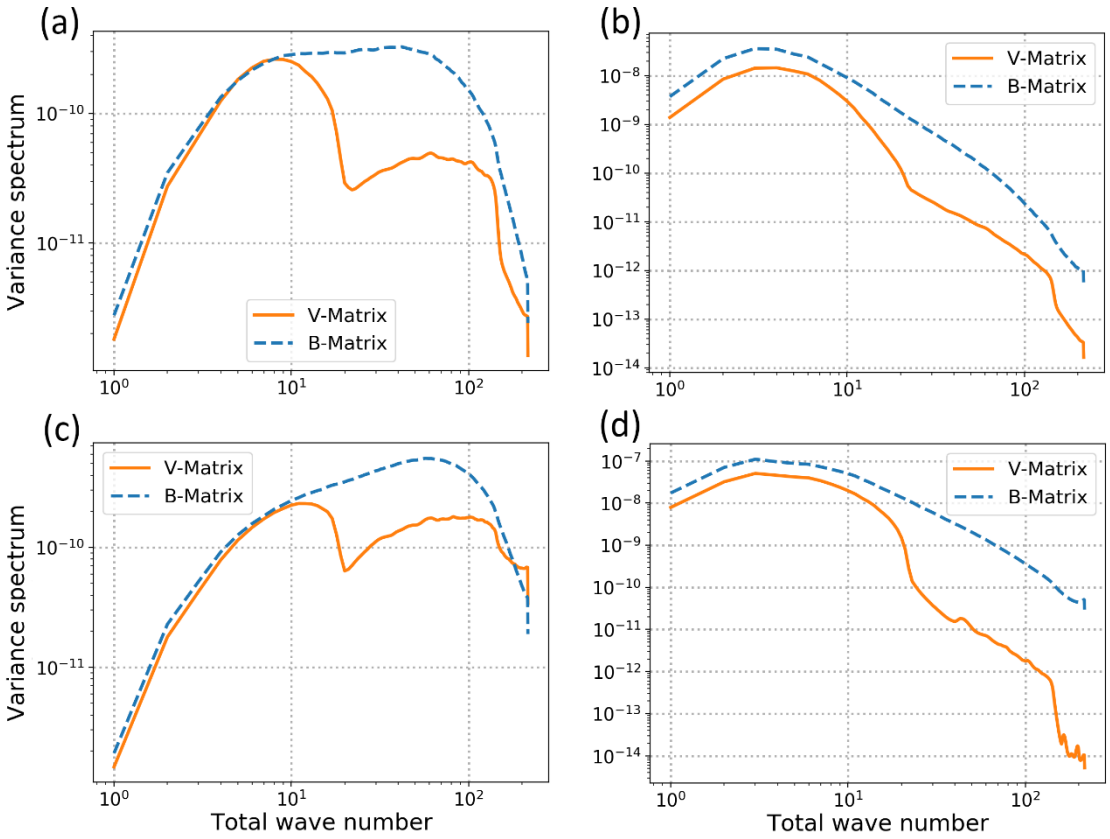

Fig. 5. Horizontal variance spectrum of C-LAEF background errors (dashed blue) and host model analysis errors (solid orange) for (a) divergence (s^{-2}) at approximately 500 hPa, (b) specific humidity (g^2kg^{-2}) at approximately 500 hPa, (c) divergence (s^{-2}) at approximately 920 hPa and (d) specific humidity (g^2kg^{-2}) at approximately 920 hPa.

a) The selection of variables included in $\boldsymbol{X}^k$. Multiple tests are performed and model behavior studied from the use of different variables. We ultimately use all analysis variables with the exception of surface pressure, as this is the best performing option.

b) A scaling factor (for each analysis variable separately) controls the relative weight of the *Jk* term relative to *Jb* and *Jo* terms of the cost function in (2.3), i.e., it can be used to fine tune the impact of all variables separately. This is a value that multiplies standard deviations within the **V**. A similar scaling factor for background error standard deviations is commonly used in ALADIN/AROME variational assimilation (i.e., REDNMC; Desroziers *et al.*, 2005). Regarding the magnitude of this scaling factor, a probabilistic approach is used here. A scaling factor for each variable is randomly selected from between two reasonable empirically predetermined values,

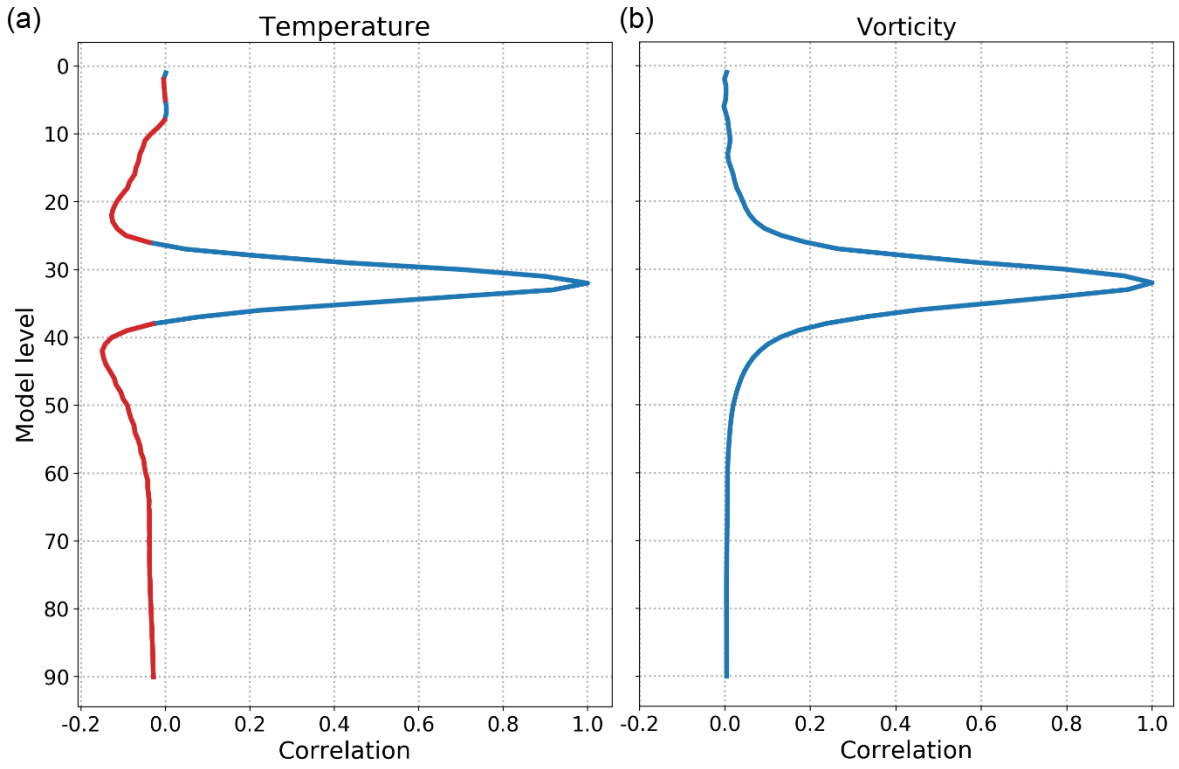

Fig. 6. Vertical profile of the vertical correlations at level 32 (approximately 500 hPa) averaged over horizontal wavenumbers for (a) temperature and (b) vorticity. Correlations above (below) zero are colored blue (red); see the text for details.

so that each member presents a slightly different combination. Each random number is determined from a seed, so that experiments can be reproduced exactly.

c) The inflation of ECMWF-EPS analyses perturbations (calculated by subtracting control analysis from each ensemble member) by a factor of 2 (0) at the lowest (highest) model level with exponential change between is performed before their use in ensemble *Jk* method. After this procedure, newly calculated perturbations are again added to the ECMWF control to create inflated analyses. When this inflation procedure is not applied, the ensemble spread is reduced in JK relative to REF. Inflation restores the spread to similar values as in REF while not affecting other scores such as the RMSE.

As already explained in the sub-section 2.1., the J_k term in (2.3) needs to be truncated such that only large scales are kept from the host EPS. To determine the exact scale of truncation, we adopted the same procedure as that applied in GF08. From Fig. 5, it can be observed that beyond $k^* \sim 10$, two spectra begin to diverge, and little energy is contained in

mesoscale spectra of the host EPS. We thus do not expect to derive any useful information from the host EPS at these scales, and the truncation was set at $k^* = 8$ which corresponds to a wavelength of roughly 135 km. In other words, any information from the host EPS on the scales below 135 km is discarded. This value of the truncation was kept constant on all model levels for all variables and all assimilation cycles. Recently, however, Feng *et al.* (2020) developed a method to dynamically select the truncation wavenumber computed according to the spectral characteristics of the LAM forecast quality and the spectral distribution of errors in the host model. The method results in less model bias and less disturbance to the LAM analysis balance. Implementing it in the C-LAEF is left for a future study.

4.3. Perturbation spectrum

To check if the ensemble *Jk* method behaves expectedly (i.e., acting on large scales), we look at the perturbation spectrum. By constraining only the large scales, it is expected to see an impact on the large-scale part of the perturbation spectrum. Spectral analysis is done as described in Denis *et al.* (2002) where discrete cosine transform is used to obtain spectra from 2D meteorological fields. Here, we apply this technique to two sets of IC perturbations (not the full fields).

Perturbations are calculated by taking the difference between an ensemble member and a control (unperturbed) member at initial time. First set is obtained by using EDA with observation perturbations only (no large-scale information from the ECMWF-EPS) and second by using ensemble *Jk* method. Figure 7. shows kinetic energy and temperature perturbation spectra for ensemble member 6 at the model level 50 (around 800 hPa) and model level 80 (around 990 hPa) averaged over the period of one week and two times per day (0000 and 1200 UTC). EDA only perturbations are shown in blue, while ensemble *Jk* perturbations are shown in orange. Denoted wave numbers 10 and 100 correspond to wavelengths of about 150 km and 15 km, respectively. The difference is obvious – ensemble *Jk* method gives perturbations with more energy at large scales, while keeping the same amount of energy at small scales. Spectra starts to differ somewhere between wavenumbers 7 and 10 confirming that our truncation is functioning properly. This is true for both kinetic energy and temperature. Regarding the different model levels – difference for kinetic energy is smaller at level 80 because ensemble *Jk* method has a weaker influence closer to the ground as ECMWF-EPS's wind perturbations are more active in the upper atmosphere. Reverse is

true for the temperature, since ECMWF-EPS's temperature perturbations are more active near the ground (e.g., Isaksen *et al.*, 2010) and this is clearly seen from Fig. 7.

Therefore, we confirm that the ensemble *Jk* method behaves as expected and obtained perturbations have more energy at largest scales.

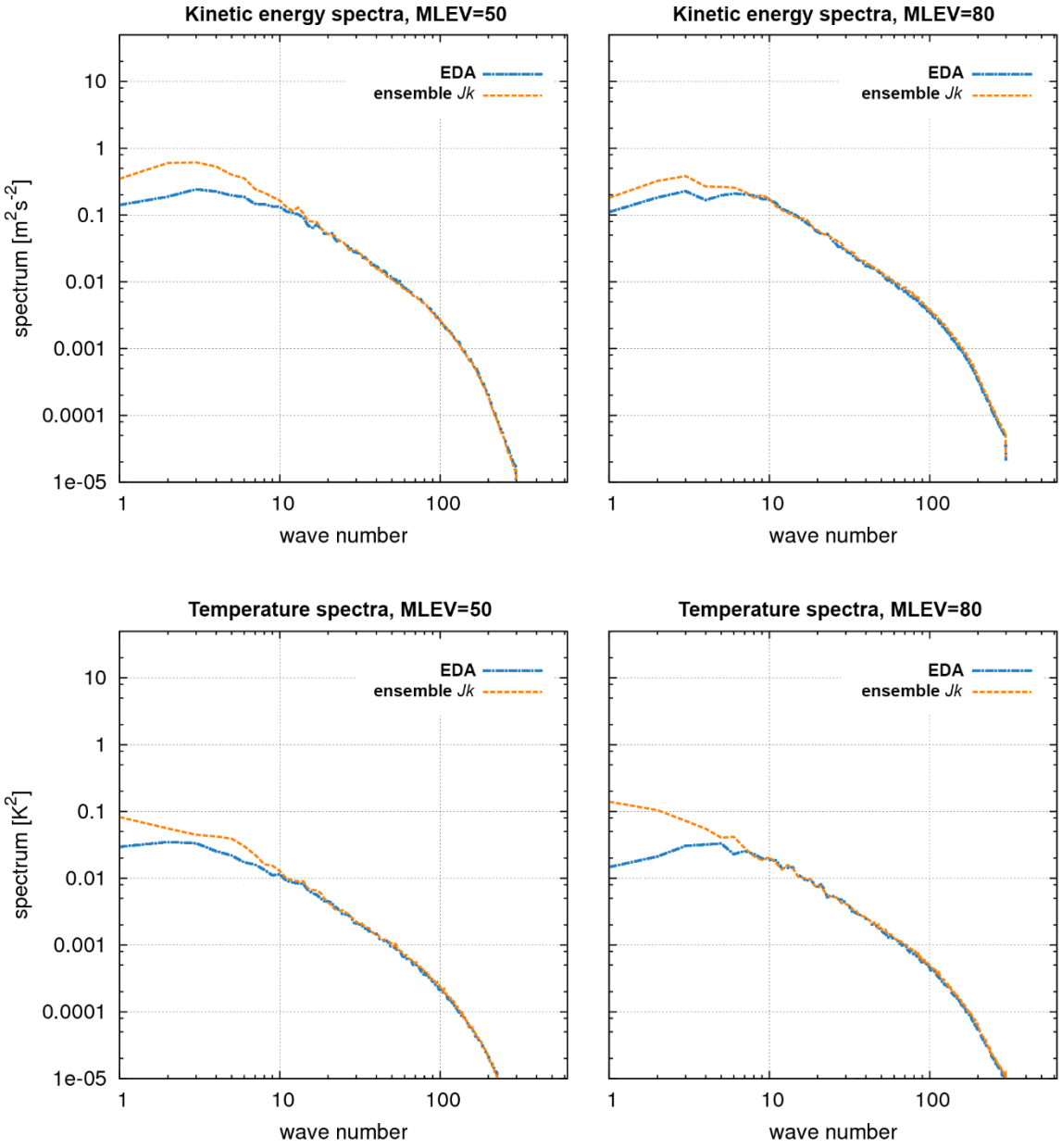

Fig. 7. Kinetic energy (upper row) and temperature (lower row) spectra at model levels 50 (left column) and model level 80 (right column) averaged over the period of 7 days, two times per day. EDA perturbations (blue) and ensemble *Jk* perturbations (orange). Denoted wave numbers 10 and 100 correspond to wavelengths of about 150 km and 15 km, respectively.

§ 5. LONG-TERM VERIFICATION

In this section, we present the results of a two-month integration of REF and JK experiments to illustrate differences in average model behavior. Experiments were performed for 62 days from 1 July until 31 August 2016 with 1200 UTC runs with a 24 h forecast range. These months involved several episodes of strong convection occurring over central Europe, rendering them appropriate periods for the verification of a convection-permitting EPS.

Due to the limited domain used, a forecast range of 24 h proved more than sufficient for our IC perturbation impact study, as LBC perturbations dominate even before this time is reached. For upper-air variables, only verification results done against independent NCEP analyses are presented, but conclusions remain the same if verification is done against ECMWF analyses. However, verifying against high resolution ECMWF analyses yields even better JK results with an increased percentage of statistically significant results relative to those presented in the following sub-sections (not shown).

5.1. RMSE of ensemble mean and spread

A well-known and understood measure of accuracy is RMSE while ensemble spread, defined as standard deviation of ensemble members, measures deviation of ensemble members from the ensemble mean.

$$RMSE = \sqrt{\frac{1}{M}\sum_{i}^{M}(\bar{F}_i - O_i)^2} \tag{5.1}$$

$$Spread = \sqrt{\frac{1}{N-1}\sum_{i}^{N}(F_i - \bar{F})^2} \tag{5.2}$$

where $\bar{F}$ is the ensemble mean forecast, O is the observation, M is the number of forecast-observation pairs and N is the ensemble size. In a statistically consistent or reliable EPS, the RMSE of ensemble mean should match its spread multiplied by a correction factor for finite ensemble size (see Fortin *et al.*, 2014). This follows from the fact that in a perfect ensemble (i.e., a perfect sampling of true state uncertainty), truth should be statistically

indistinguishable from any ensemble member (Buizza *et al.*, 2005). Thus, the RMSE/spread relation is a good measure of reliability. Lower (higher) spread than RMSE indicates under-(over-) dispersion.

Fig. 8 shows RMSE and spread for T500, T850, RH500, RH850, W500 and W850. Positive effects of the ensemble *Jk* method on the accuracy of the ensemble mean can be clearly identified because RMSE is reduced for almost all variables and at both levels in JK. However, differences observed are mostly statistically significant within early forecast ranges (up to 12 h). Spread remains similar in JK and REF at 850 hPa but is reduced at 500 hPa in JK, which is a favorable result because JK would be over-dispersive otherwise. As REF is under-dispersive, reducing RMSE more than spread enhances the reliability of JK.

RMSE and spread for surface variables are shown in Fig. 9, in which results are more neutral. There is a slight improvement in RMSE of JK for T2M, RH2M and W10M, mostly over the first 3-9 h and for MSLP from 9 to 18 h. However, none of the differences observed are significant. Spread remains almost the same for T2M and RH2M, while it is slightly higher in the case of MSLP and lower for W10M in JK.

Both REF and JK are under-dispersive for both upper-air variables and especially for surface variables. These results are consistent with those of numerous other studies (e.g., Wang *et al.*, 2011; Bouttier *et al.*, 2015; Harnisch and Keil, 2015; Johnson and Wang, 2016, among others). Under-dispersion is a well-known problem of ensemble forecasting in general (e.g., Leutbecher and Palmer, 2008). Reasons for this are the deficiencies in methods that try to estimate the true uncertainties in various components of LAMEPS (i.e., IC, LBCs, model, etc.) and because observation errors are generally not accounted for (see sub-secion 8.2.4). In this study, under-dispersion is even more evident as model and surface perturbations are not included. In addition, due to the relatively large difference in grid size between C-LAEF and ECMWF-EPS, LBC "sweeping" effect (Nutter *et al.*, 2004a) is enhanced and plays a significant role in the longer forecast ranges, placing an additional constraint on the ensemble spread growth. The under-dispersion can be alleviated by implementing techniques described in section 8.

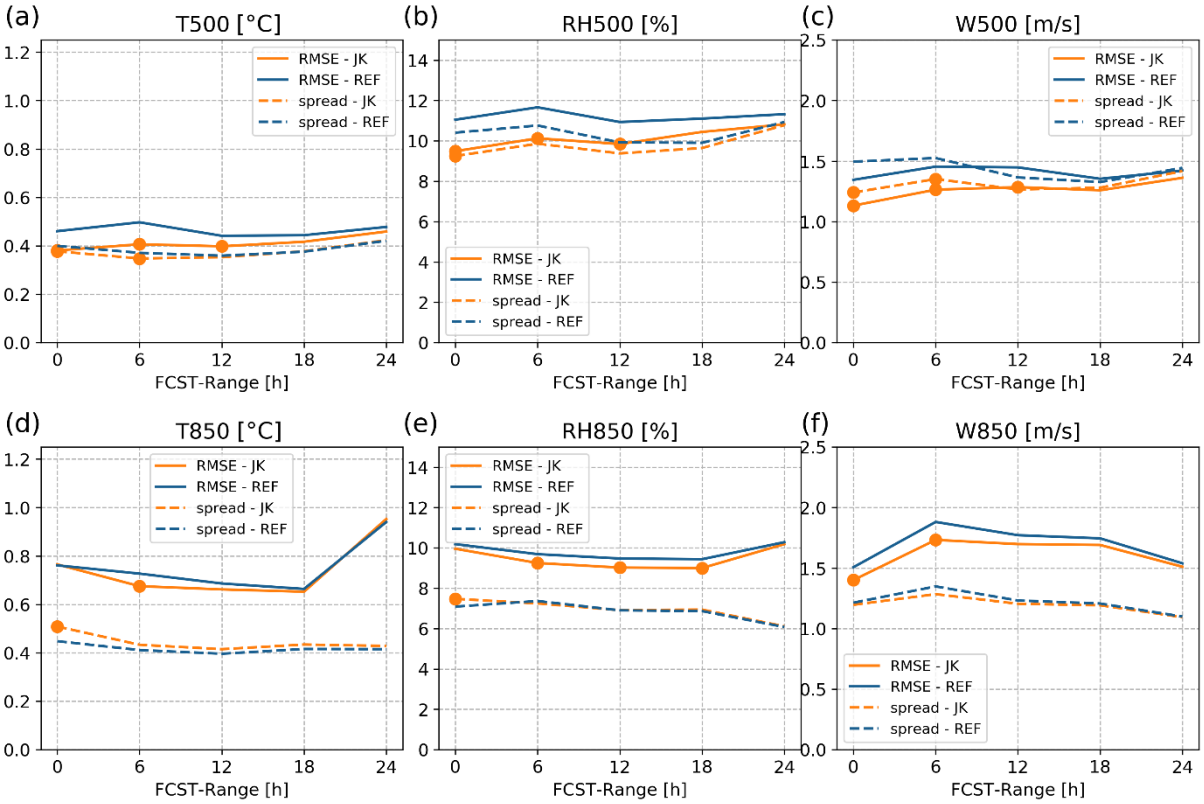

Fig. 8. RMSE of the ensemble mean and ensemble spread of REF (dash-dotted and dotted blue, respectively) and JK (solid and dashed orange, respectively) for (a) T500, (b) RH500, (c) W500, (d) T850, (e) RH850 and (f) W850 for the verification period. Forecast ranges with statistically significant differences are marked with bullets.

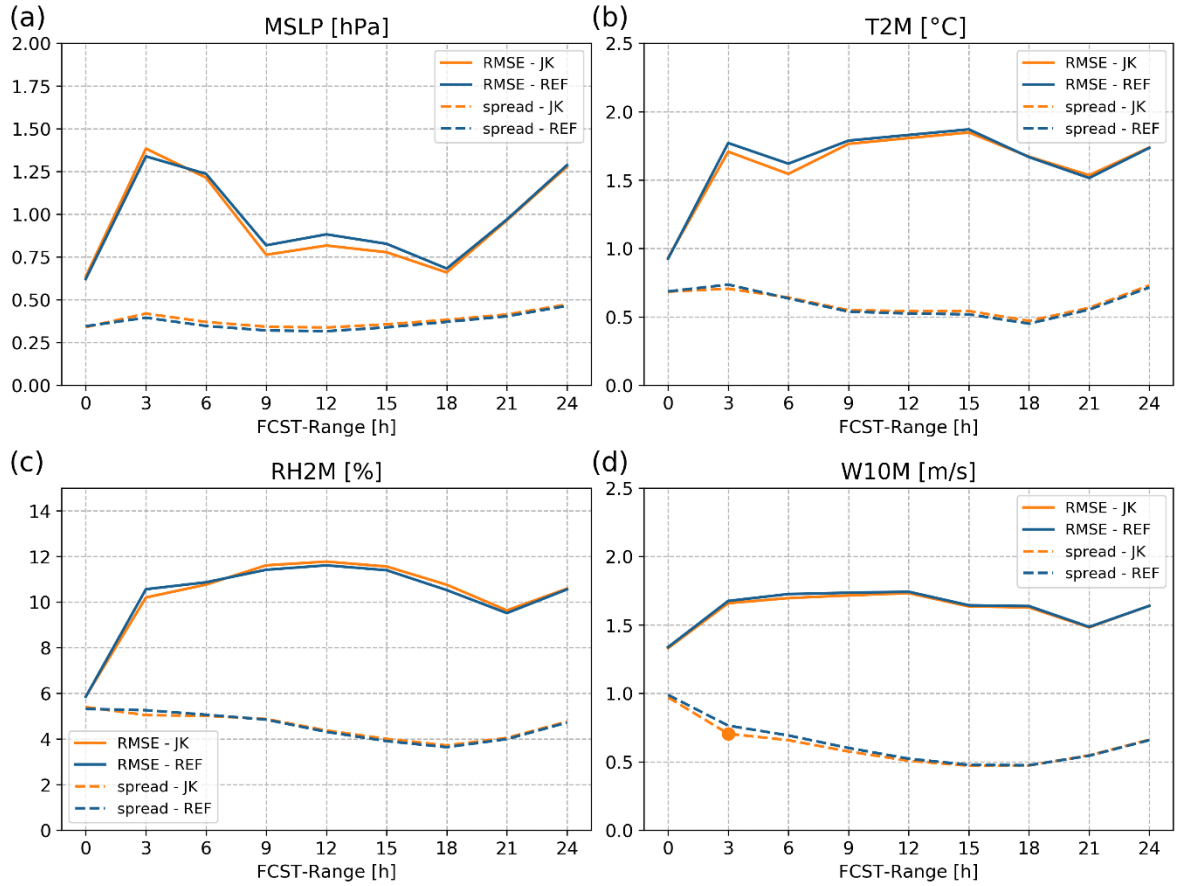

Fig. 9. As in Fig. 8, but for (a) MSLP, (b) T2M, (c) RH2M and (d) W10M.

5.2. Continuous ranked probability score

The CRPS is related to the rank probability score but for an infinite number of classes; thus, it compares the full predicted distribution with observations. The CRPS is a good measure of overall EPS skill. It is dependent on a predicted distribution's first (ensemble mean) and second moment (spread) and is negatively oriented, meaning that lower values are better.

$$CRPS = \int_{-\infty}^{\infty} [F(x) - O(x)]^2 dx \tag{5.3}$$

$$O(x) = \begin{cases} 0, & x < observed\ value \\ 1, & x \geq observed\ value \end{cases} \tag{5.4}$$

Where *F(x)* and *O(x)* are cumulative probability density functions of forecasts and observations, respectively.

The CRPS for upper-air and surface variables is shown in Figs. 10 and 11, respectively. Better CRPS is obtained for all upper-air variables for all levels for JK. Significance test results are similar to RMSE results – significant within early forecast ranges for upper-air (up to 12 h) and up to 18 h for RH850. For the surface, significant improvements in CRPS are obtained at 9 and 12 h for MSLP in JK. Insignificant improvements observed in the first forecast hours for T2M in JK are also visible here. Thus, the CRPS supports conclusions drawn from Fig. 8 and 9 that JK is significantly more skillful for upper-air variables and MSLP.

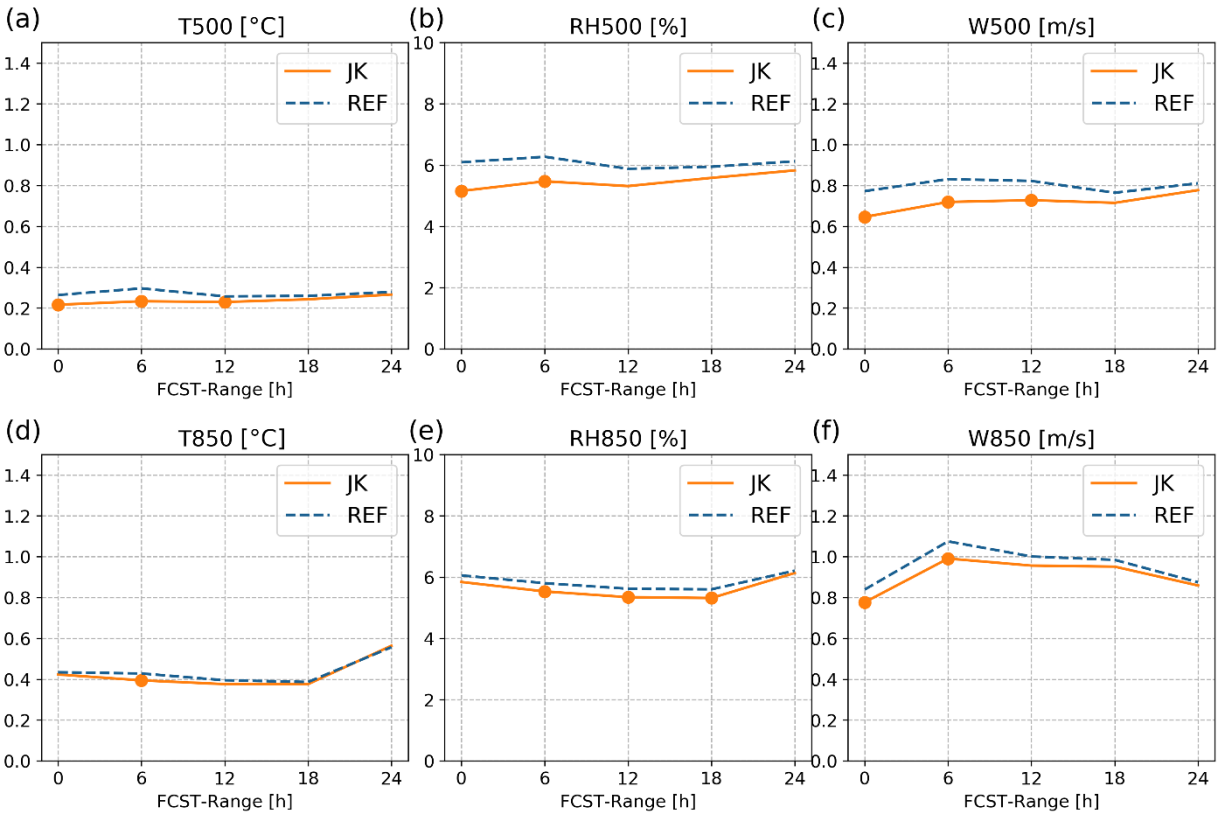

Fig. 10. CRPS of REF (dashed blue) and JK (solid orange) for (a) T500, (b) RH500, (c) W500, (d) T850, (e) RH850 and (f) W850 for the verification period. Forecast ranges with statistically significant differences are marked with bullets.

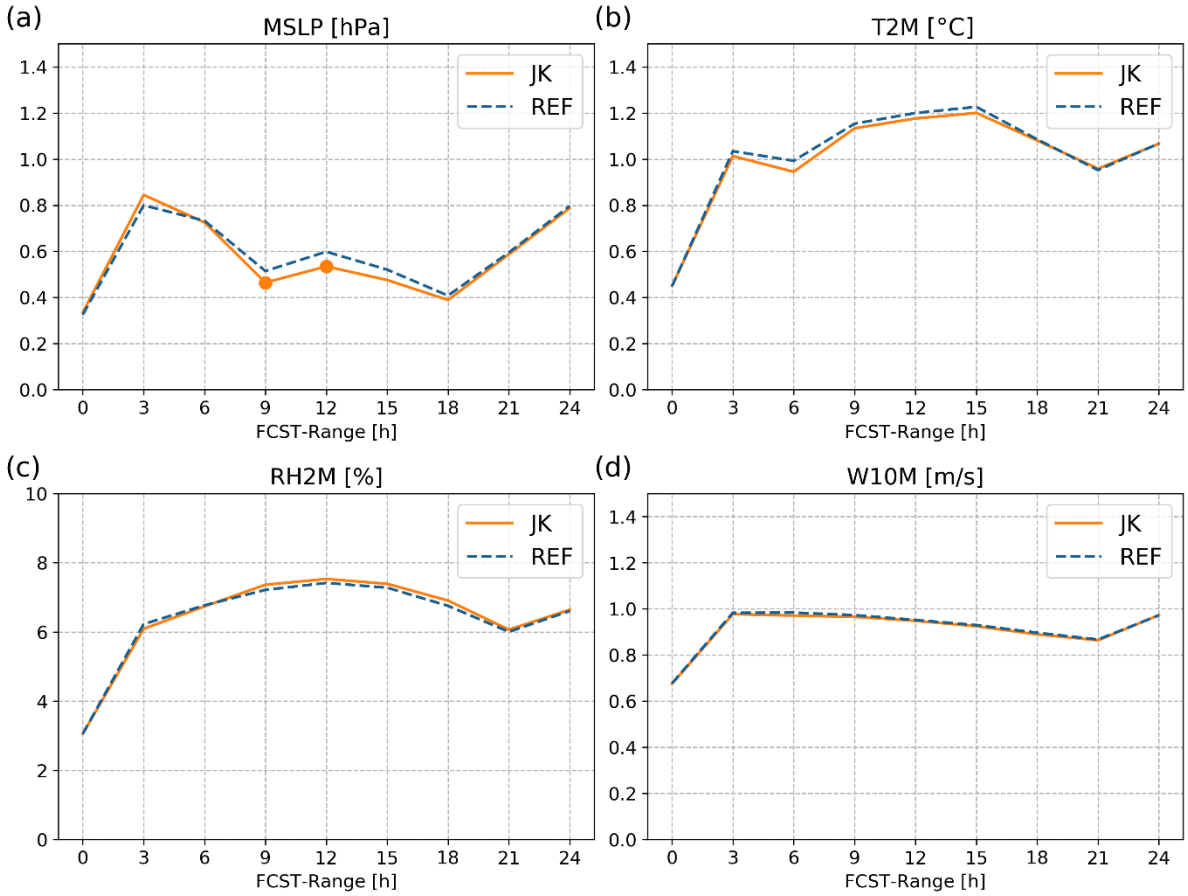

Fig. 11. As in Fig. 10, but for (a) MSLP, (b) T2M, (c) RH2M and (d) W10M.

5.3. Outlier statistics

Outliers are defined as the number of cases in which verifying observations lie outside of the whole ensemble and are typically expressed as percentages. The perfect outlier percentage is calculated as $2/(N+1) * 100\%$ (e.g., Wilks, 2006) where N is the ensemble size (thus, a perfect score is 11 % in our experiments). The percentage of outliers serves as a good measure of reliability. This follows from the fact that if the ensemble spread is consistently too small, then the observation will often be an outlier in the distribution of ensemble members, implying that ensemble relative frequency will be a poor approximation to probability

Fig. 12 shows the percentage of outliers for upper-air variables. The number of outliers is greatly reduced in JK for T500 and T850, with weaker effect found for RH500 and RH850 and with almost neutral effect found for W850 and W500. This means that reliability

is mostly improved for temperature and relative humidity. Significance test results support these conclusions.

Regarding the surface variables and results given in Fig. 13, the greatest reduction in the number of outliers in JK is achieved for MSLP, which is also significant at 9 and 12 h forecast range, resulting in a more reliable ensemble. Slight and insignificant reductions are also observed for T2M.

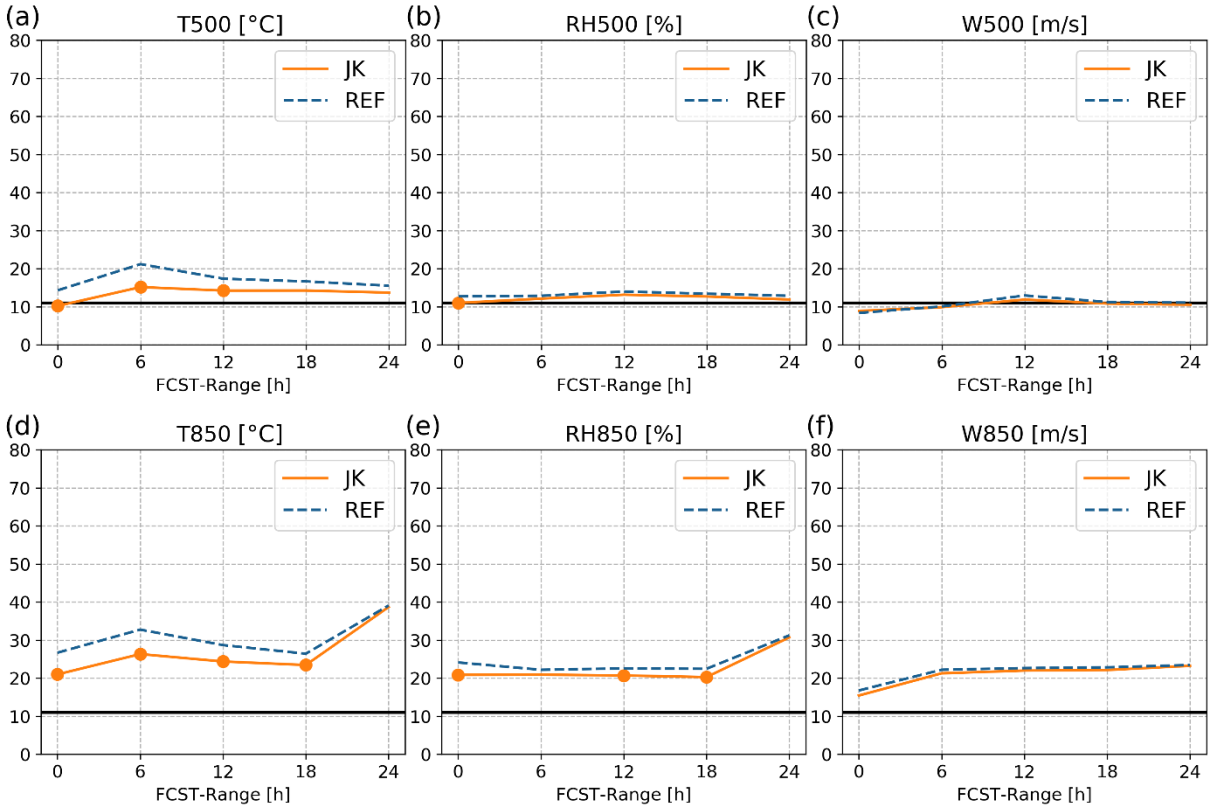

Fig. 12. Percentage of outliers of REF (dashed blue) and JK (solid orange) for (a) T500, (b) RH500, (c) W500, (d) T850, (e) RH850 and (f) W850 for the verification period. Forecast ranges with statistically significant differences are marked with bullets. Solid black lines denote an ideal value of 11 %.

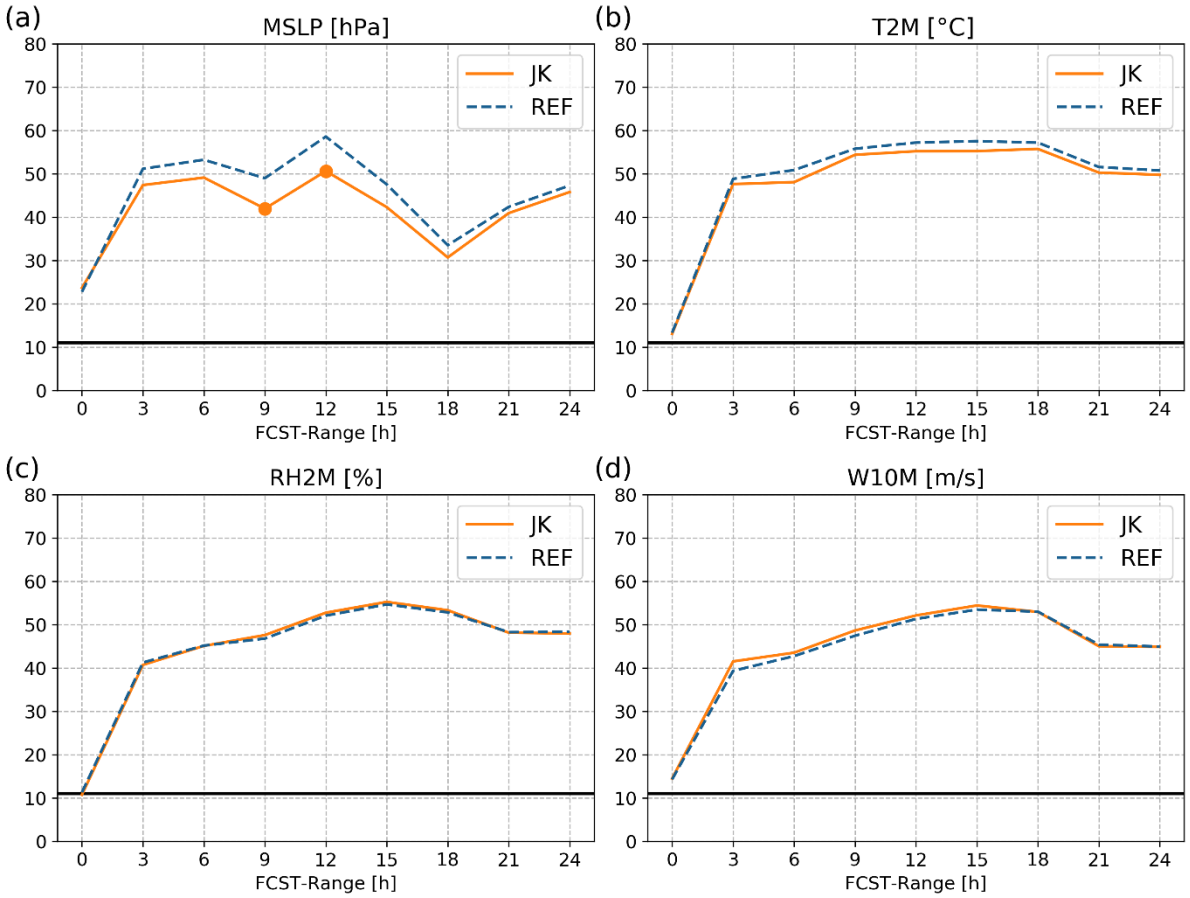

Fig. 13. As in Fig. 12, but for (a) MSLP, (b) T2M, (c) RH2M and (d) W10M.

5.4. Decomposition of the Brier score

The Brier score (BS) is the mean squared error of the probability forecasts, considering that the observation is 1 if the event occurs, and that the observation is 0 if the event does not occur (e.g., Wilks, 1997). Similar to the RMSE, the score averages the squared differences between pairs of forecast probabilities and the subsequent binary observations and it is negatively oriented. The BS can be decomposed in three terms: reliability, resolution and uncertainty (Wilks, 2006):

$$BS = \sqrt{\frac{1}{M}\sum_{i}^{M}(p_i - o_i)^2} \tag{5.5}$$

$$BS = reliability - resolution + uncertainty \tag{5.6}$$

where p_i and o_i denote ensemble and observation probability of occurrence, respectively. Reliability and resolution are the two main characteristics describing the quality of a probabilistic forecast (Murphy, 1993). Since more accurate forecasts are characterized by smaller values of BS, one would like the reliability term to be as small as possible, and the resolution term to be as large (in absolute value) as possible, while uncertainty depends only on the observations, and is unaffected by the forecasts.

Figure 14 shows the decomposition of the BS for upper-air variables averaged over the verification period and first 12 h of forecast range. Significant reduction of BS is visible for almost all variables for a range of different thresholds. More importantly, decomposition of BS shows that improvement comes not only from improved reliability, but also from improved resolution. Almost perfect reliability at 500 hPa is in the agreement with the results from sub-sections 5.1. and 5.3. For the surface (Fig. 15), mostly neutral results are obtained.

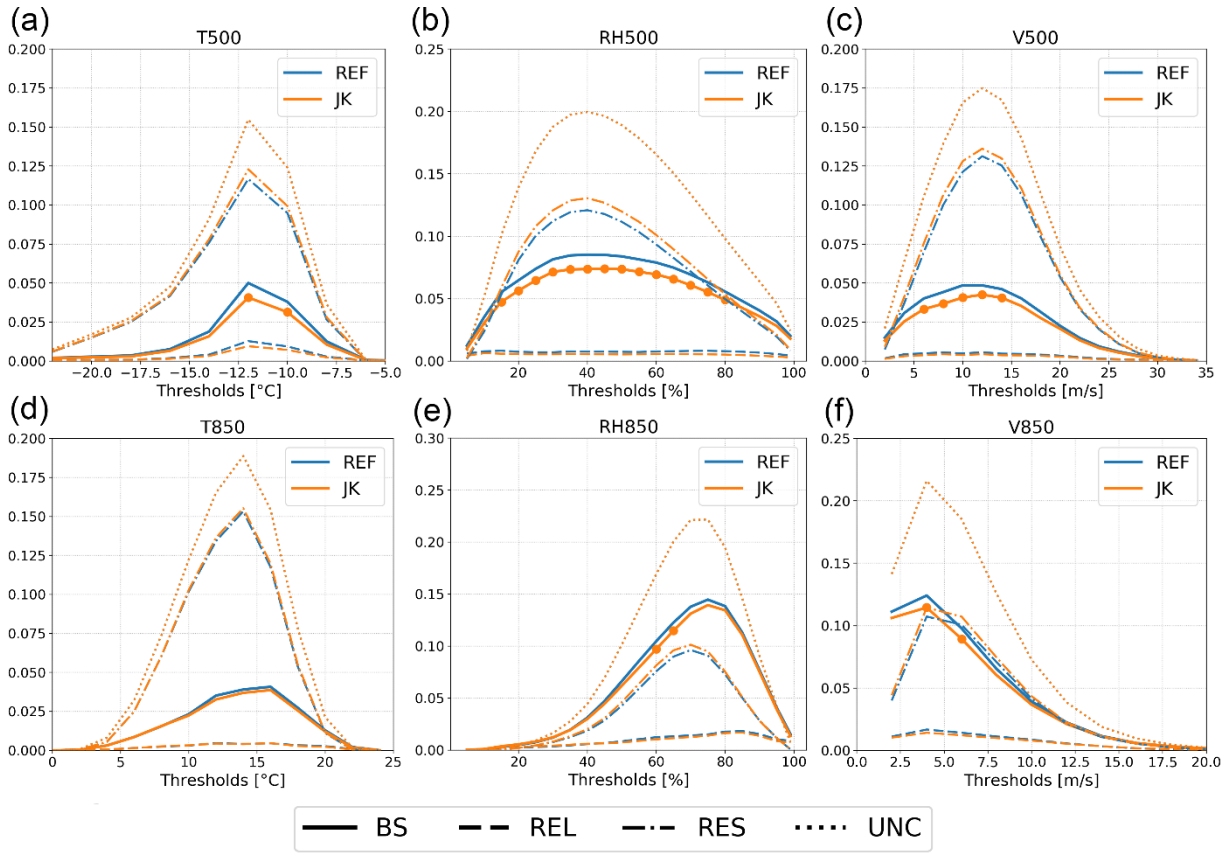

Fig. 14. Decomposition of the BS for REF (blue) and JK (orange) for (a) T500, (b) RH500, (c) W500, (d) T850, (e) RH850 and (f) W850 for the verification period and averaged over first 12 h. Thresholds with statistically significant differences are marked with bullets.

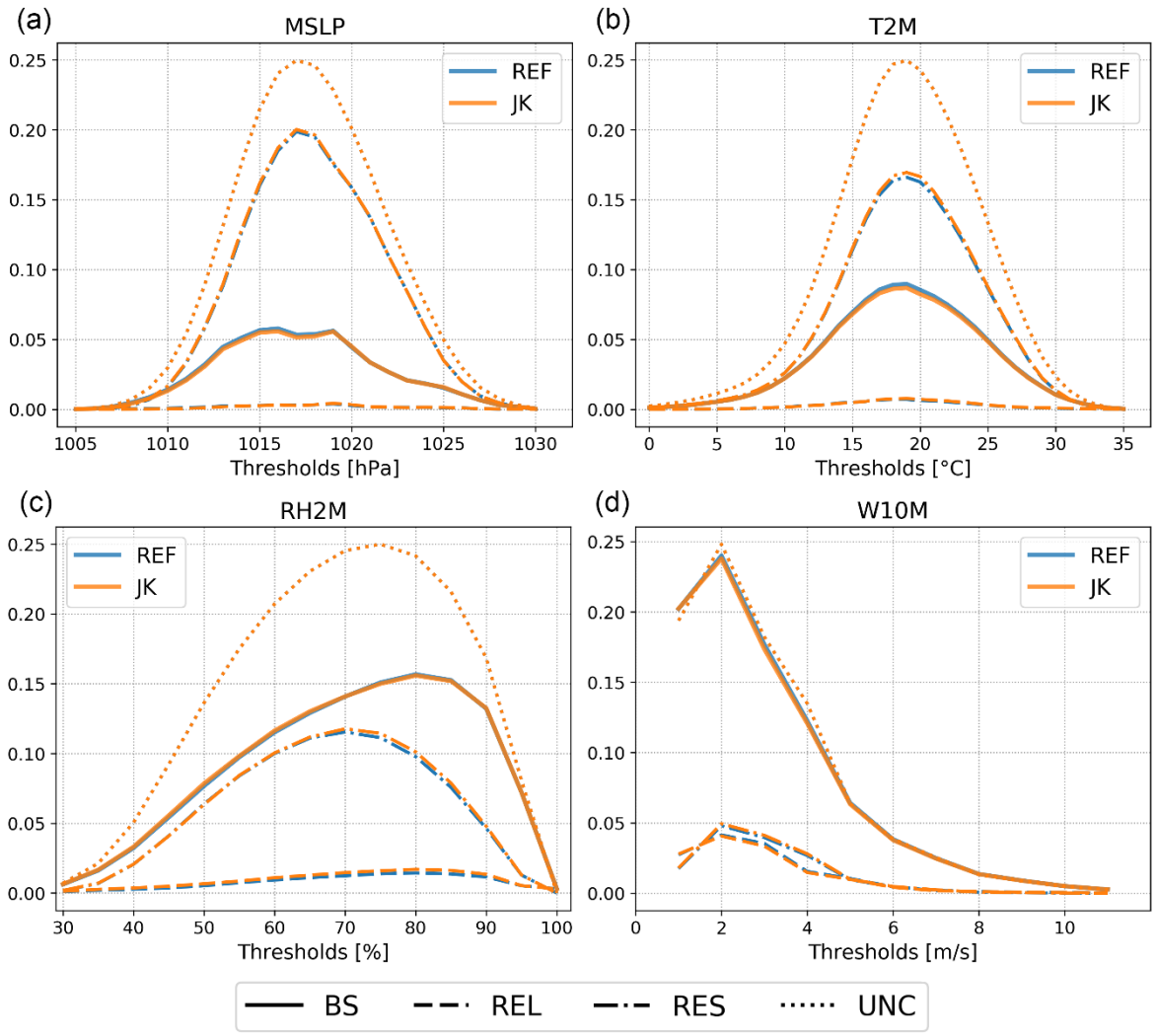

Fig. 15. As in Fig. 14, but for (a) MSLP, (b) T2M, (c) RH2M and (d) W10M.

5.5. Relative operating characteristics

The ROC curve plots the hit rate (hits divided by the sum of hits and misses) against false alarm rate (false alarms divided by the sum of false alarms and correct negatives) for a set of predefined probability thresholds (in our case, used probability thresholds are: 0-0.1, 0.1-0.2, …, 0.9-1.0). It compares two conditional distributions where the hit rate corresponds to occurrences and where the false alarm to non-occurrences of an event measuring the attribute of discrimination. Perfect EPS will have hit rate equal to 1 and false alarm rate equal to 0. Thus, perfect curve travels from bottom left to top left of diagram, then across to top right of diagram. Diagonal line indicates no skill. ROC is usually used to test whether the model is

able to detect occurrences of extreme events. Therefore, ROC was calculated over a range of different thresholds and the results presented here are for the thresholds for which JK performed the best with an emphasis on higher threshold values, while results for other thresholds will be mentioned in the text.

The ROC curve of upper-air variables averaged over the first 12 h of forecast range is shown in Fig. 16. Discrimination for wind speed is improved for all the tested thresholds (0 – 38 ms^{-1}) and both levels. Slight improvements of discrimination in relative humidity are visible for all thresholds up to 99 % (75) % at 500 hPa (850 hPa), while it is neutral for the rest. For temperature, mostly neutral results are obtained except for a slight improvement for more extreme values at 500 hPa. In short, the results show that JK can better distinguish between the occurrences and non-occurrences of an event for almost all variables, a range of different thresholds and it is better at detecting extreme events for W500, W850, T500 and RH500.

For the surface variables and extreme thresholds (Fig. 17), the situation is more neutral for RH2M and W10M, although a small improvement in JK is visible for W10M and thresholds above 4 ms^{-1}. For T2M, discrimination is visibly improved in JK for thresholds above 28 °C, while for MSLP, JK better discriminates low pressure events (below 1008 hPa which is an extreme value for summer over central Europe). This indicates that JK is slightly better at detecting extreme events than REF.

5.6. Precipitation

Figure 18 shows FSS for 3-h precipitation of JK versus REF as a function of the lead-time for the whole verification period. At a given lead-time, 62 skill scores (one for each day) are computed of the FSS of JK using the REF FSS as reference. This second normalization (of JK FSS in respect to REF FSS) of the score is computed to avoid the variability of the FSS score depending on the rainfall of both experiments. The left panel shows the median of the Skill Score of FSS of JK to FSS of REF illustrated as matrix of colors (red means JK better than REF and blue the opposite, white is for no difference) for different thresholds, scales and forecast ranges. The right panel in Fig. 8 is the significance level for the comparison (meaning the percentage of times FSS of JK is higher than the FSS of REF). At 6 h forecast range, precipitation forecast is improved for all thresholds (scales above 195 km) and for thresholds up to 5 mm (scales above 45 km). For 12 and 18 h forecast ranges, neutral results are

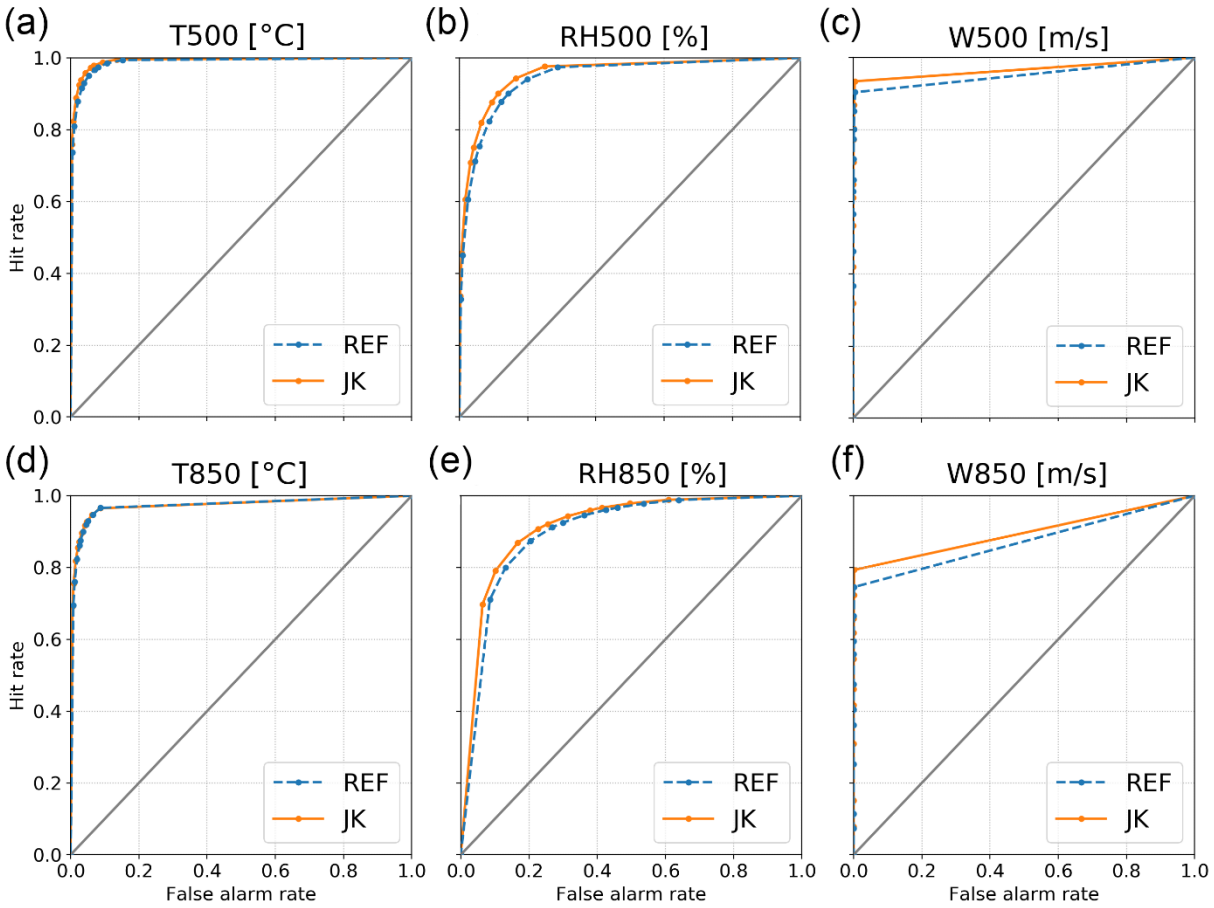

Fig. 16. ROC of REF (dashed blue) and JK (solid orange) for (a) T500 (threshold: > -10 °C), (b) RH500 (> 55 %), (c) W500 (> 30 ms^{-1}), (d) T850 (> 16 °C), (e) RH850 (> 55 %) and (f) W850 (> 24 ms^{-1}) for the verification period and averaged over first 12 h.

observed, while for 24 h forecast range, precipitation forecast is improved for all scales (threshold of 1 mm) and for higher thresholds (scales above 195 km). Results presented here indicate that, at early forecast ranges, mostly larger scales are affected (this is in agreement with the fact that only large scales are affected in the ensemble *Jk* method), while at later forecast ranges, nonlinear model integration spreads the influence to all scales. Improvements at higher thresholds are especially important as detecting extreme weather events is one of the main purposes of convection-permitting LAMPESs.

5.7. Summary and discussion

The results presented in this section show that JK is significantly better for all EPS attributes studied here, especially reliability, and for all upper-air variables. In addition, JK showed the ability to better detect extreme events. Regarding the surface variables, JK is significantly better at MSLP and precipitation which is especially important for a convection-permitting EPS, while is mostly neutral for the other surface variables. More neutral results for surface variables and for all scores presented in this section are in line with analysis of Dahlgren and Gustafsson (2012) regarding the *Jk* blending method and are as expected for a number of reasons. First, lower boundary conditions have a considerable impact on surface variables which we do not perturb and are the same for all members as described in sub-section 3.1., also see Table 1. Second, surface variables are sensitive to small scale variability and processes while the ensemble *Jk* method acts only on upper-air variables and at the largest scales.

We did not conduct observation error simulations of any kind. Though observation error can affect the outcome of forecast verification (see sub-section 8.2.4.), we believe that this was not the case in our study. First, we were only interested in relative performance, i.e., differences observed between experiments, and including the observation error would have had effect of the same direction for both experiments. Second, RMSE/spread relation and outlier statistics are scores for which the absolute performance of each system is also important. As shown by other works (e.g., Bouttier *et al.*, 2012; Romine *et al.*, 2014), accounting for observation error, would have increased ensemble spread given in Figs. 8 and 9. This could mean that an EPS using the ensemble *Jk* method could be declared over-dispersive in such case. However, a slight modification of the ensemble *Jk* method can be done to account for such over dispersion. The inflation factor discussed in sub-section 4.2. can be reduced to decrease the spread while keeping RMSE unchanged (not shown).

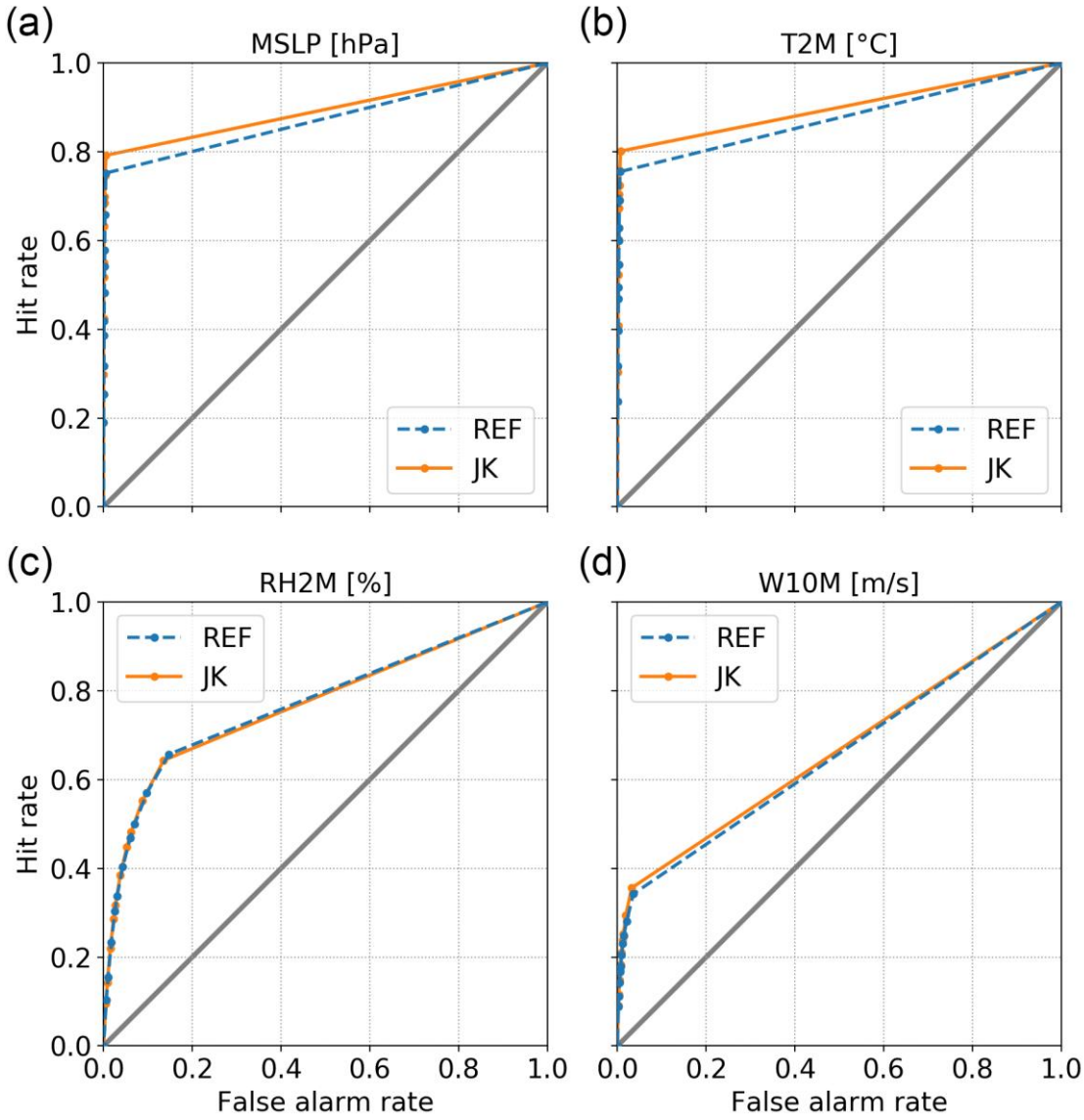

Fig. 17. As Fig. 16, but for (a) MSLP (< 1008 hPa), (b) T2M (> 30 °C), (c) RH2M (> 90 %) and (d) W10M (> 6 ms^{-1}).

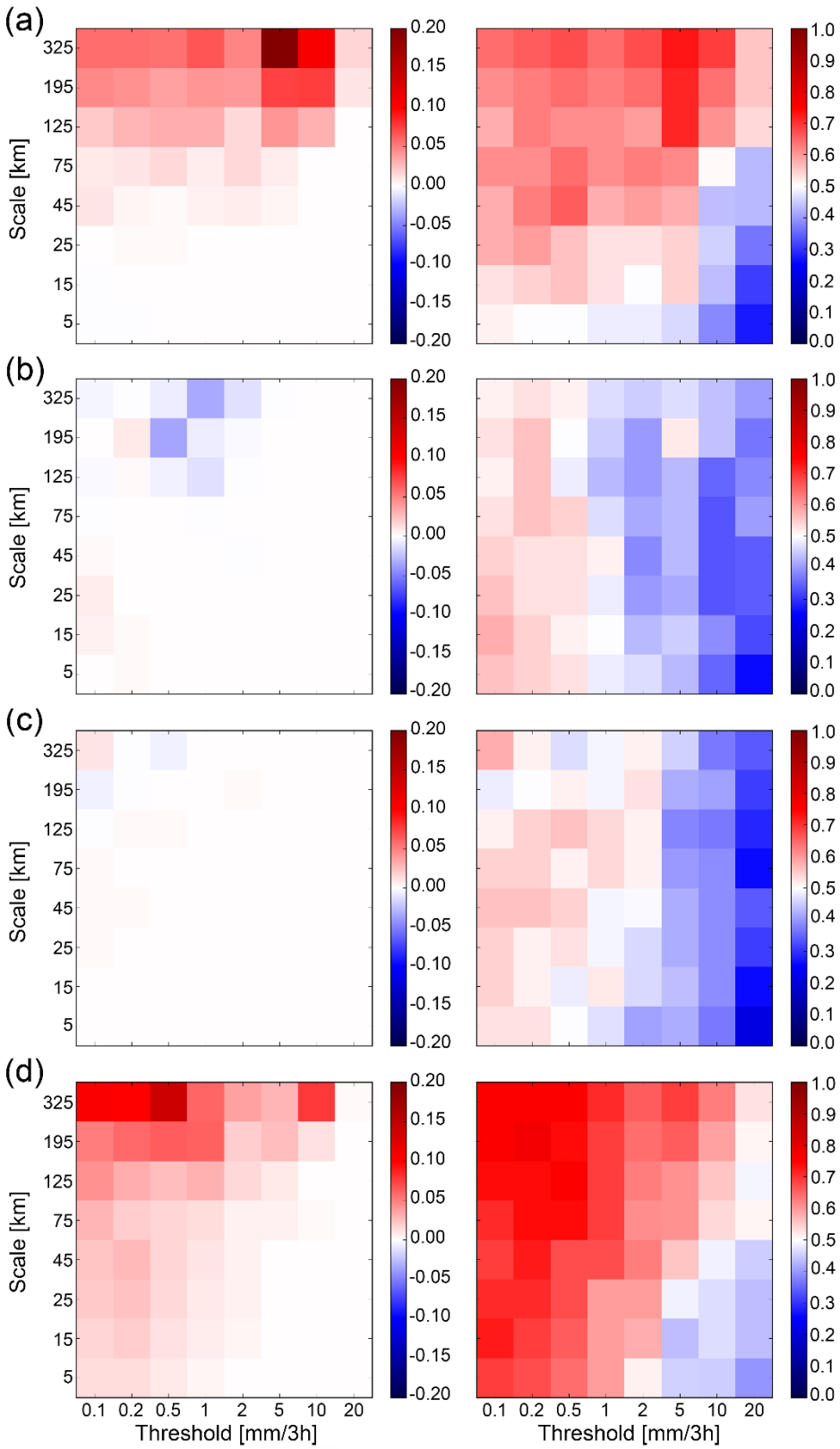

Fig. 18. The left panel is the median Skill Score of FSS of JK to FSS of REF (red means JK is better than REF and blue the opposite) as a function of scale and threshold. The right panel is the significance level for the comparison (percentage of times FSS of JK is higher than the FSS of REF). The plots are shown for (a) 6 h, (b) 12 h, (c) 18 h and (d) 24 h forecast ranges.

§ 6. CASE STUDIES

We expect to observe benefits of including large-scale perturbations in cases where strong synoptic forcing is located near the borders of the computational domain. Any information outside of the model domain is propagated inside through LBCs. However, due to well-known deficiencies related to lateral boundary formulation in LAMs (e.g., Warner *et al.*, 1997; Nutter *et al.*, 2004a; Termonia *et al.*, 2009) and due to LAM orientation toward the mesoscale, this information can be deficient. We believe that introducing large-scale information in an additional way - through ensemble *Jk* method can alleviate such problems. We now illustrate benefits of including large-scale perturbations from the host model through two case-studies. One of them is characterized by a cold front entering and moving through the C-LAEF domain and the other one is characterized by a cyclone which is located just outside the C-LAEF domain, but impacts the weather inside.

6.1. 11 July 2016

Figure 19 shows a synoptic situation observed on 11 July 2016 at 0000 and 1200 UTC. As is shown, a cold front of a cyclone over the North Sea is moving over the C-LAEF domain. At the same time, strong low-level strong advection of warm and moist air from the Mediterranean area is observed. This synoptic setup has a great potential to support pre-frontal convection. This is exactly what happened as there were many reports of strong rain and winds, large hail over south and east Germany, Austria and Poland and two tornadoes over Belarus.

Figure 20 (left column) illustrates ensemble mean MSLP for REF and JK experiments started on 11 July 2016 at 0000 UTC from 3 and 9 h forecasts. The overall forecasted spatial distribution in pressure is similar across the whole domain. The biggest difference, considering all forecast hours is visible over the upper half of the domain where lower pressure, related to an incoming cold front, propagates further to the south in JK (note how 1011 and 1014 hPa contours are displaced more to the south in the left column in Fig. 20). The right column in Fig. 20 shows the difference between ensemble mean absolute error (forecast – observation) for JK and REF plotted on surface station locations. As the front

enters and moves through the domain, an interesting phenomenon is observed; improvements in MSLP forecast for JK follow the progression of the frontal system. Smaller errors of up to 1.5 hPa are observed in the northern part of the domain from which the cold front progresses towards inland Europe. Pressure distribution is clearly better represented in JK. Although surface pressure large-scale information was not used in the ensemble *Jk* method directly, improvements in surface pressure forecast come from the other variables as surface pressure reflects the entire column of the atmosphere.

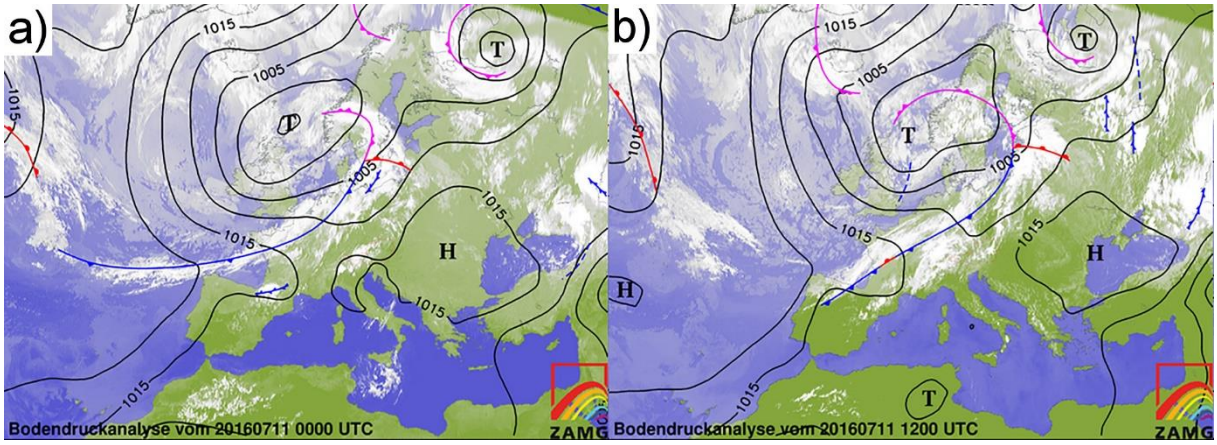

Fig. 19. Synoptic situation observed on 11 July 2016 at (a) 0000 UTC and (b) 1200 UTC. Black lines denote mean sea level pressure where H (T) denotes the high (low) pressure centre; blue (red) line denotes cold (warm) fronts and purple lines denote occluded fronts.

Figure 21 (Fig. A1 in Appendix) shows the probability of 3-h precipitation above 1 mm (10 mm) calculated from 3 - 6 h and 6 - 9 h forecasts for REF and JK experiments started at 0000 UTC on 11 July 2016. Given that no precipitation is observed (Fig. A2 in Appendix) for these hours, our REF experiment generates excessively high probability values (up to 60 %) for precipitation over southern Germany and western Austria. For the JK, area affected and probabilities are reduced with the latter remaining a below 20 % for 3 – 6 h and at 30 % for 6 – 9 h forecasts, thus lowering the possibility of a false alarm. Later during that day, convection developed as the front moved through the domain. Figure 22 (Fig. A3 in Appendix) shows the probability of 3-h precipitation above 1 mm (10 mm) calculated from 9 – 12 h forecasts for REF and JK experiments started at 1200 UTC on 11 July 2016 and INCA analysis. JK experiment better captured overall precipitation distribution, especially over

northern Austria and southern Czech Republic and a dry area over southeastern Germany which are both missed in REF. Intense precipitation in northern Italy is better forecasted in JK with above 90 % probability for rain above 10 mm compared to around 40 % in REF (Fig. A3). However, both experiments missed the intense precipitation over central Austria. Area over southern Switzerland is not considered in the analysis because, in INCA, observation data is scarce there and the quality of the radar data is suboptimal.

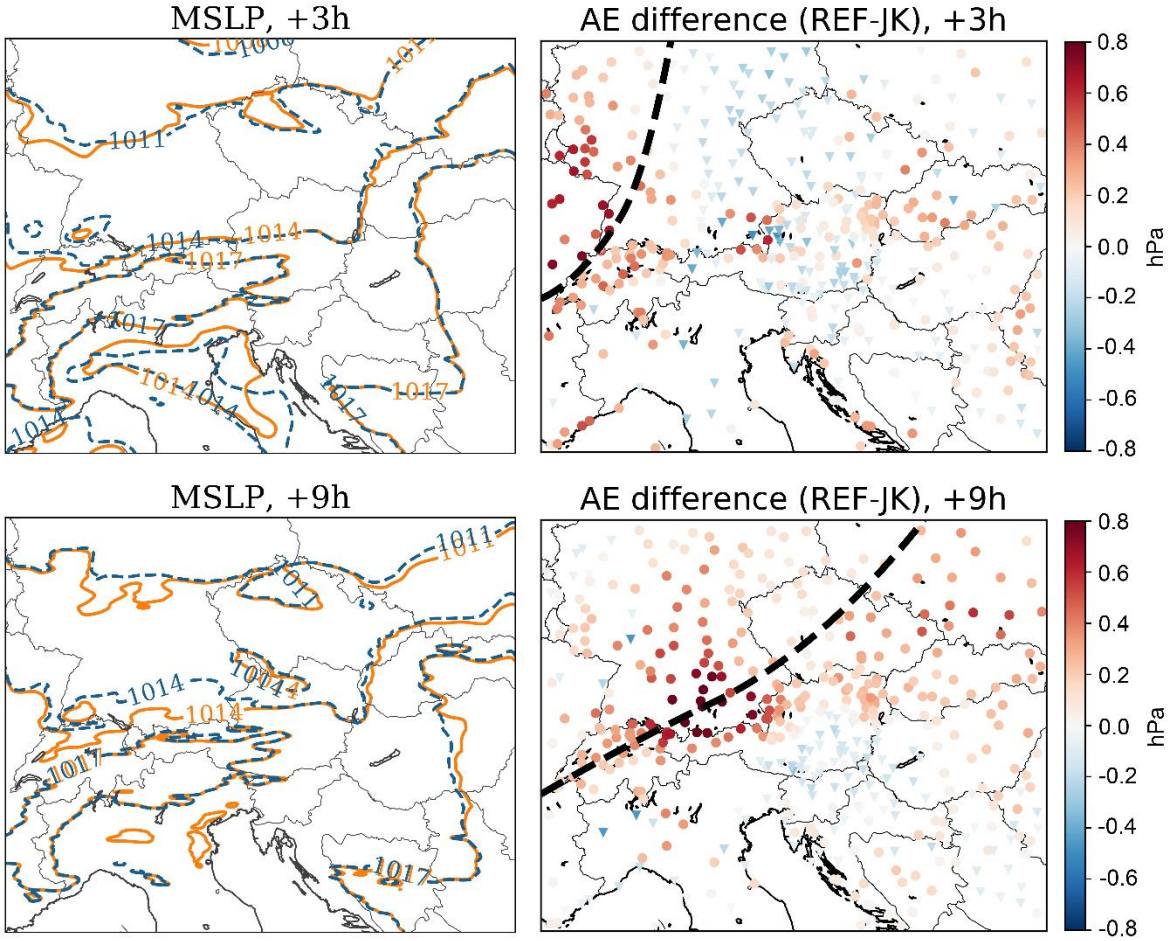

Fig. 20. Left column, ensemble mean MSLP for REF (dashed blue) and JK (solid orange) contoured every 3 hPa. Right column, absolute error difference (error(REF) – error(JK)) on the surface station locations denoted by red circles (blue triangles) if difference positive (negative). Black dashed lines denote the approximate position of incoming cold front. Forecasts are started on 11 July 2016 at 0000 UTC. The upper and lower rows denote 3 and 9 h forecast ranges, respectively.

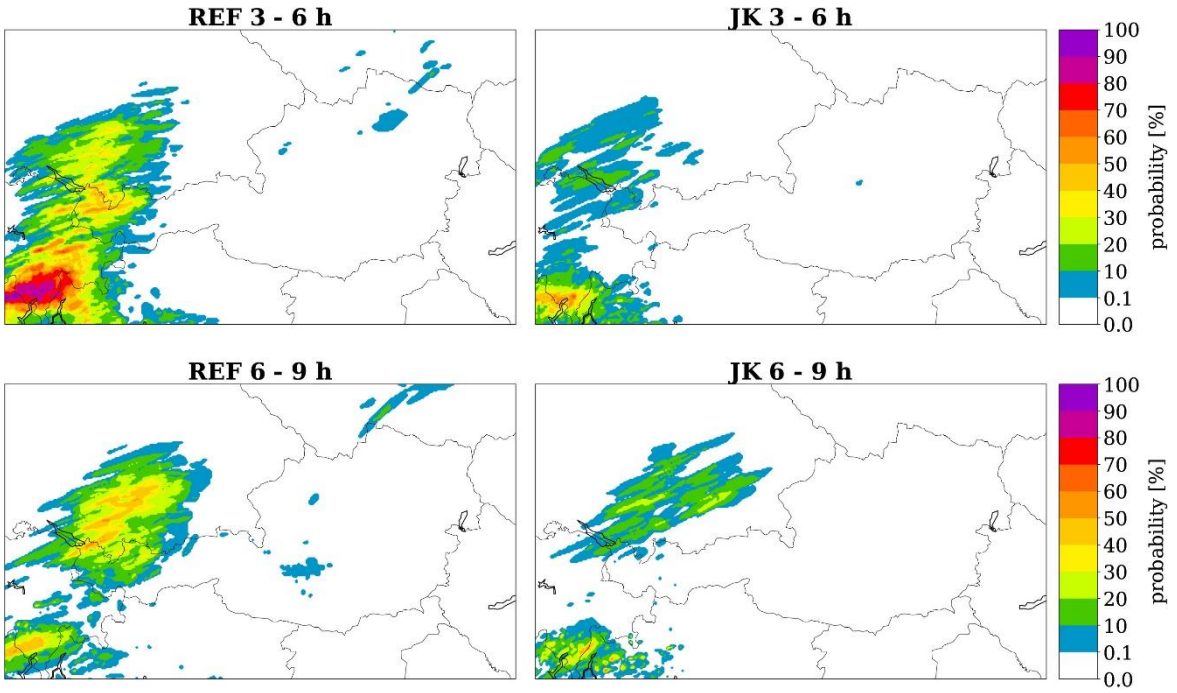

Fig. 21. Probability of 3-h precipitations exceeding 1 mm for REF (left column) and JK (right column) forecasts started on 11 July 2016 at 0000 UTC. Upper and lower rows show probabilities between forecast ranges of 3 and 6 h and between 6 and 9 h, respectively.

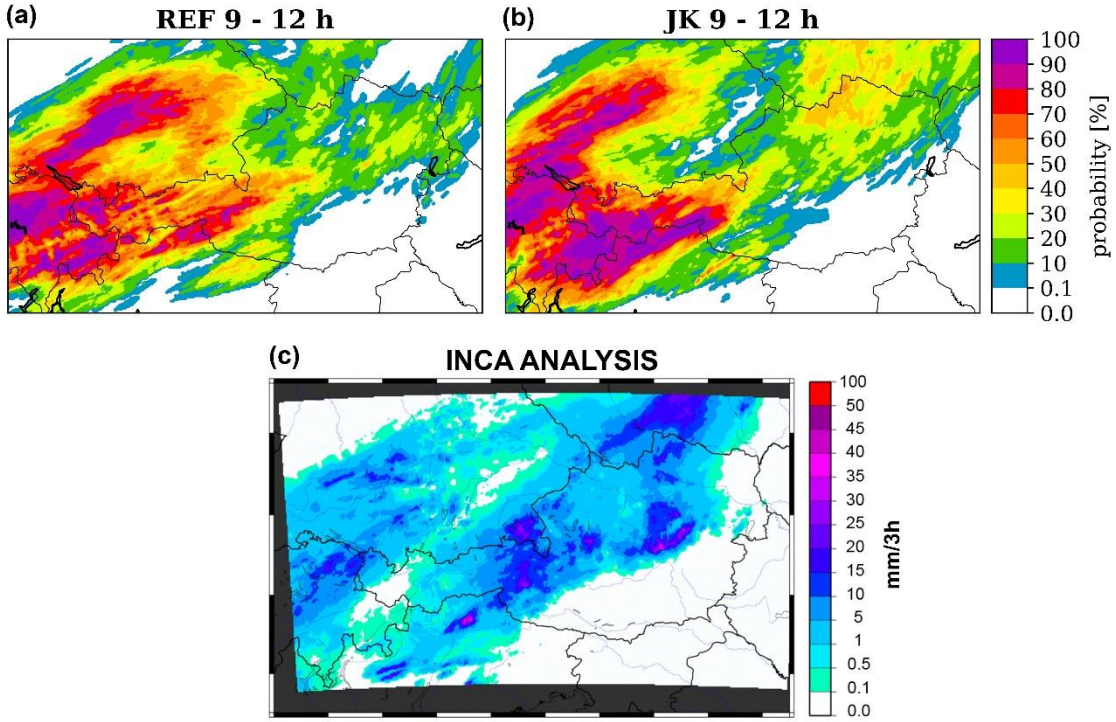

Fig. 22. Probability of 3-h precipitations exceeding 1 mm for (a) REF and (b) JK forecasts started on 11 July 2016 at 1200 UTC. INCA analysis (mm/3h) is shown in (c).

6.2. 28 August 2016

Figure 23 shows synoptic situation observed on 27 August 2016 at 1800 and 28 August 2016 at 0000, 0600 and 1200 UTC. A dominant feature is an eastern European high which encompasses almost entire C-LAEF domain. However, a cyclone located over the British Isles is moving eastward while, in front of it, a convergence zone is forming at the northwestern corner of the C-LAEF domain. Later, a secondary low forms at the northern boundary of the C-LAEF domain with a cold front moving through the domain. This is the kind of situation that interests us – synoptic forcing is located outside the domain, but it impacts the weather inside it. Normally, LBCs are the only source of information about the events outside the computational domain for LAMEPS, but this time, the ensemble *Jk* method provides additional source.

Similar to the right column in Fig. 20., Fig. 24. shows the difference between ensemble mean absolute error (forecast – observation) for JK and REF plotted on surface station locations. Again, as the front enters and moves through the domain, improvements in MSLP forecast for JK follow the propagation of the front. Smaller errors of up to 2 hPa are observed in the northwestern part of the domain from which the cold front progresses towards eastern Europe. As in the case of 11 July, pressure distribution is clearly better represented in JK. In addition, forecasts for T2M, RH2M and W10M are also improved in the northwestern part of the domain (Fig. A4 in Appendix).

To see how the convergence zone was represented in REF and JK, Fig. 25 shows 6-h ensemble median precipitation forecasts and respective observations. Precipitation related to the convergence zone was observed mostly in the northwest corner of the C-LAEF domain. In JK, precipitation pattern resembles observation consistently, while in REF, there is excessive precipitation along the western domain boundary. Neither experiment captured precipitation over Switzerland, but this event was caused by the orographic forcing of the Swiss Alps and not related to the convergence zone.

Overall, the results presented in this section demonstrate the benefits of applying large-scale constraints on C-LAEF perturbations. Similar to the results from section 5., impact on surface variables is mostly obvious on MSLP and precipitation, but it is also observed for the other variables. Improvement to MSLP forecast can be an indicator of the overall forecast quality as MSLP reflects the entire model column. As a better illustration of MSLP improvement, Fig. 26. shows RMSE of ensemble mean of MSLP of REF and JK over the verification period averaged over first 15 forecast hours. Again, MSLP is, almost constantly,

better forecasted in JK. Improved pressure distribution can lead to a better large-scale feature placement, such as synoptic systems, fronts, etc. Small errors in the placement of such features, at analysis time, can then lead to a completely different synoptic system setup at a later time (Dahlgren and Gustafsson, 2012).

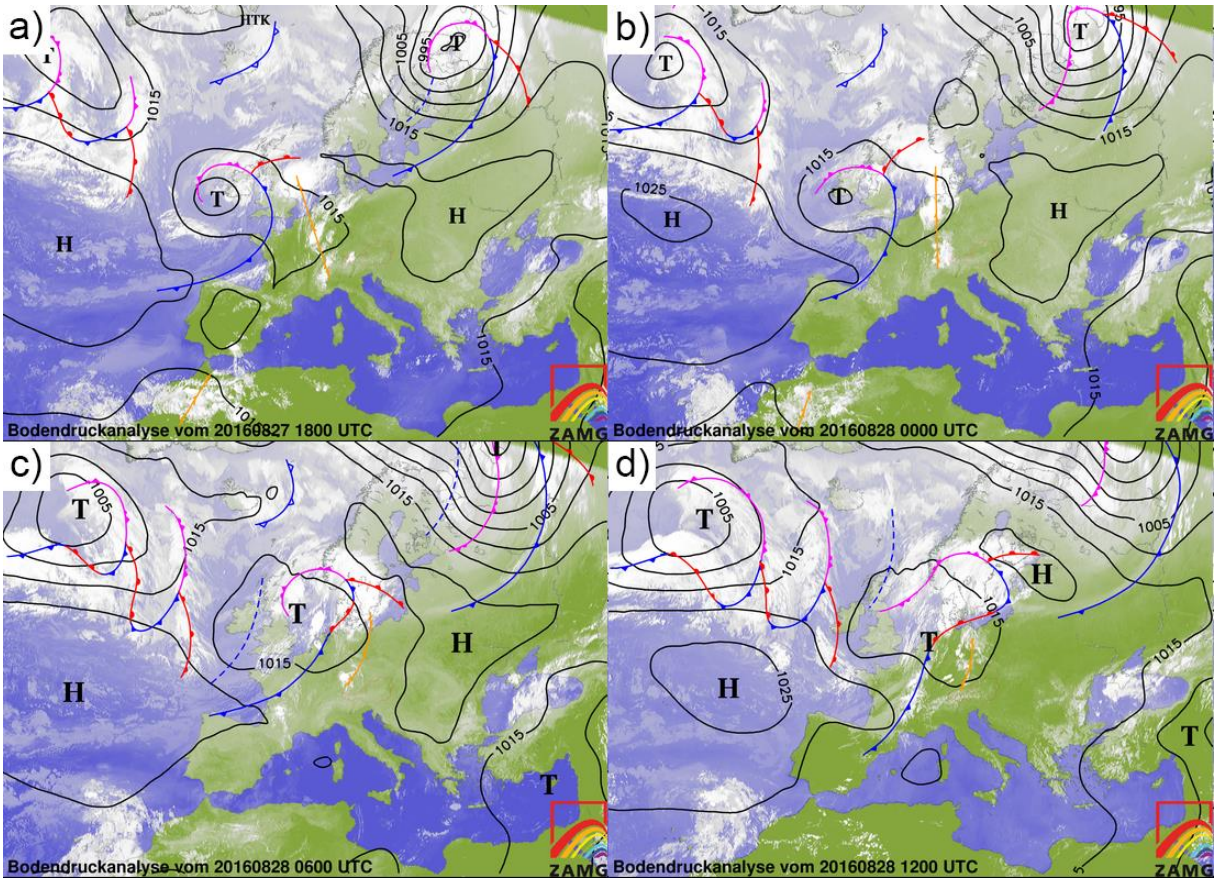

Fig. 23. Synoptic situation observed on 27 August 2016 at (a) 1800 UTC and 28 August 2016 at (b) 0000 UTC, (c) 0600 UTC and (d) 1200 UTC. Black lines denote mean sea level pressure where H (T) denotes the high (low) pressure centre; blue (red) line denotes cold (warm) fronts, purple lines denote occluded fronts and yellow lines denote convergence zones.

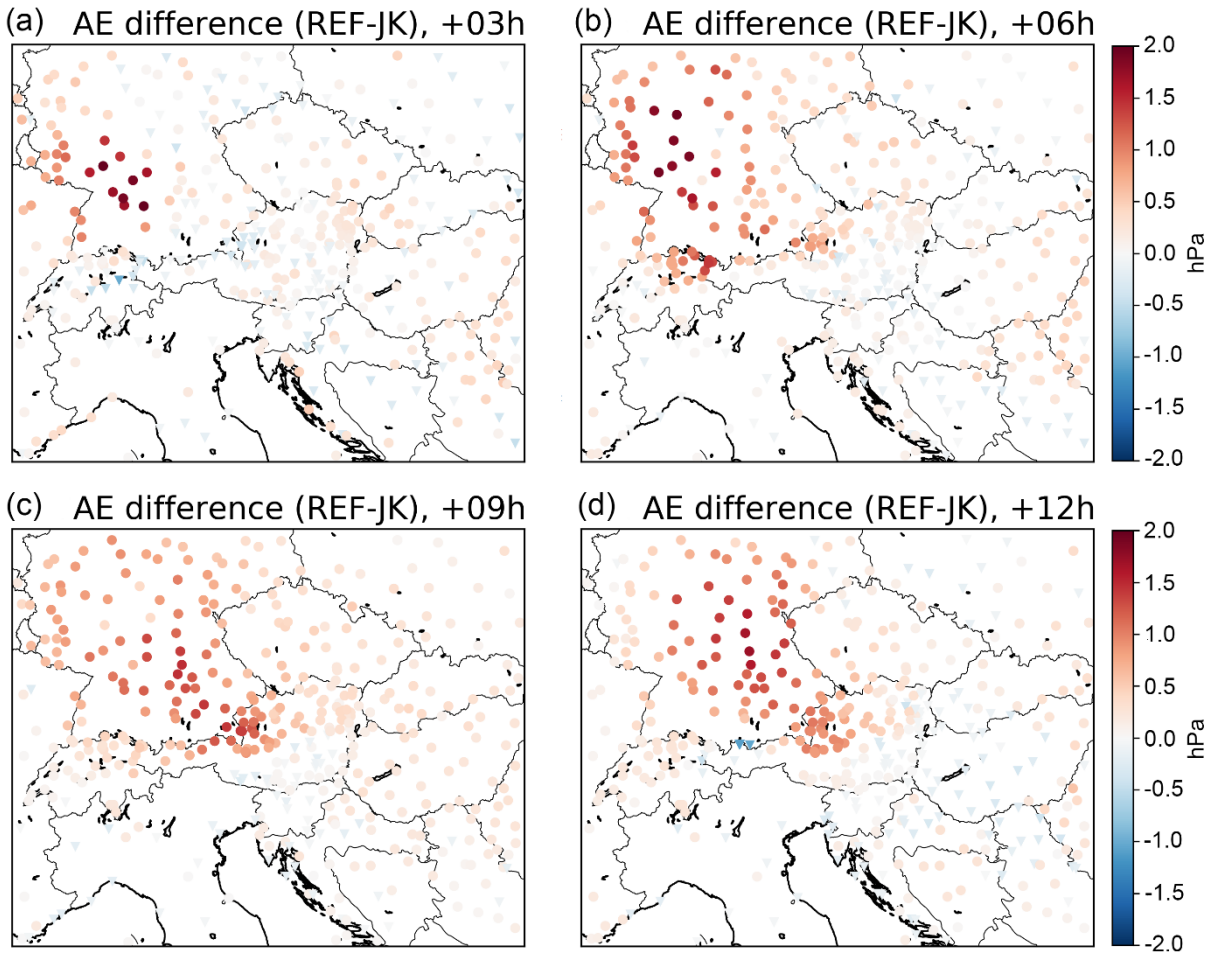

Fig. 24. Absolute error difference (error(REF) – error(JK)) on the surface station locations denoted by red circles (blue triangles) if difference positive (negative). Forecasts are started on 28 August 2016 at 0000 UTC and are valid on (a) 0300 UTC, (b) 0600 UTC, (c) 0900 UTC and (d) 1200 UTC.

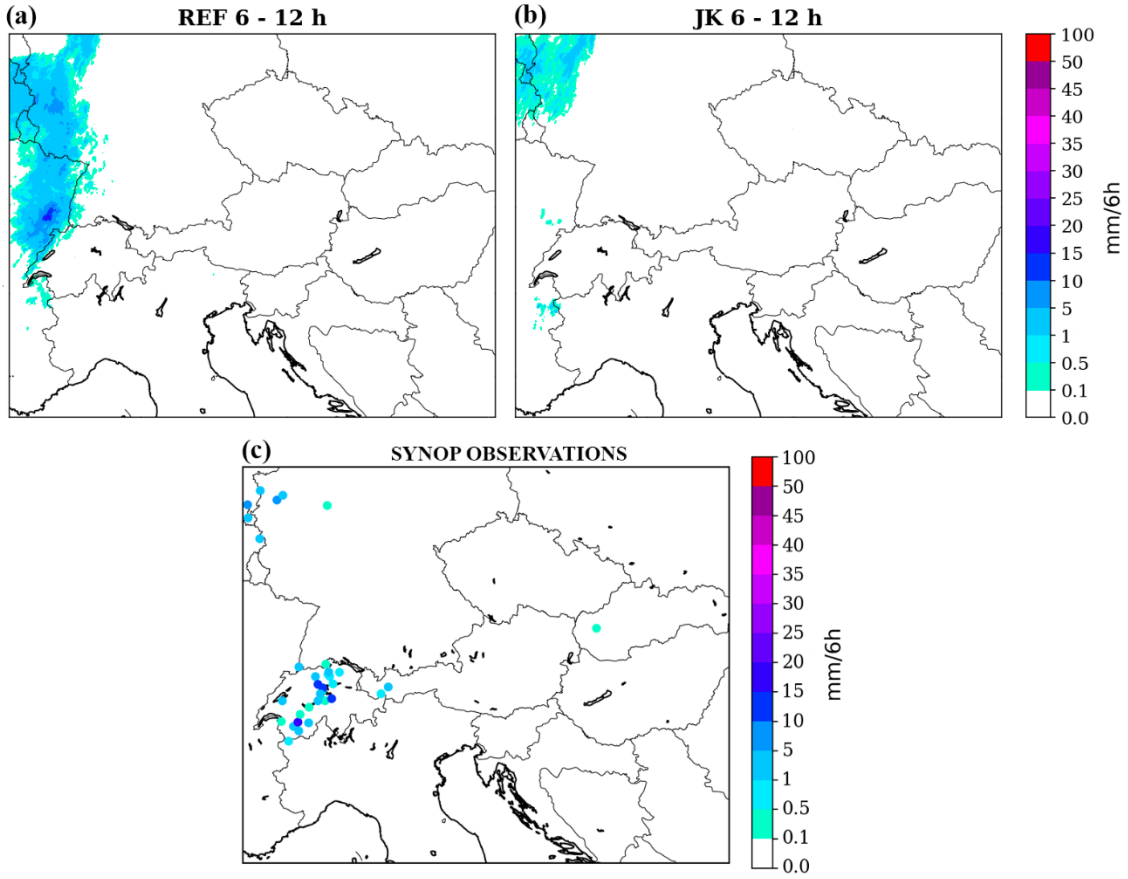

Fig. 25. Ensemble median of 6-h accumulated precipitation for (a) REF and (b) JK of a forecast started on 27 August at 1200 UTC. SYNOP observations are shown on (c).

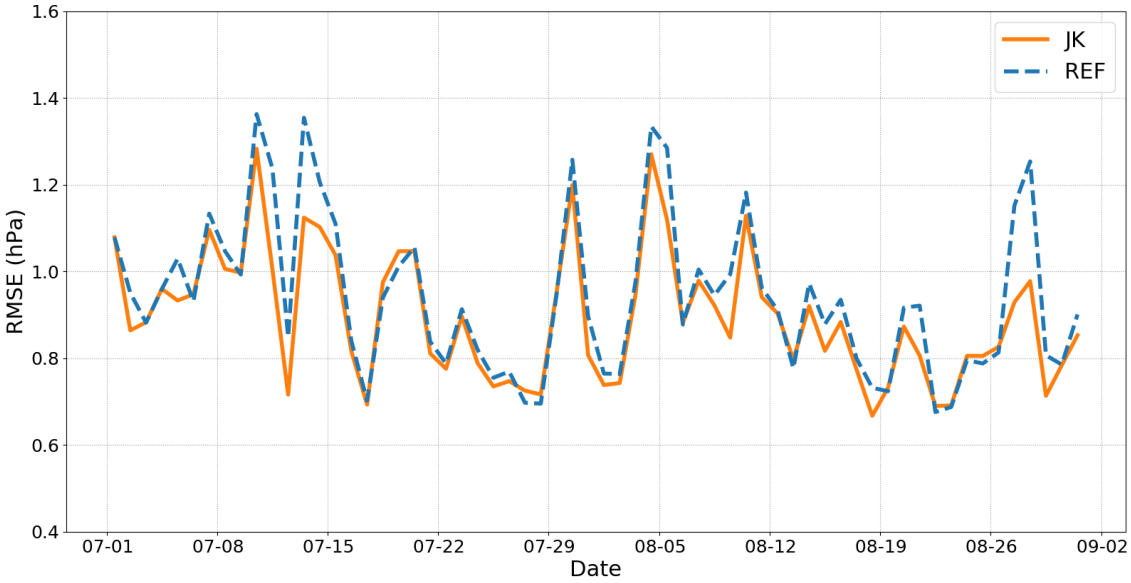

Fig. 26. RMSE of the ensemble mean of REF (dashed blue) and JK (solid orange) for MSLP over verification period averaged over first 15 forecast hours.

§ 7. PERTURBATION MISMATCHING

We now turn to the second main focus of this dissertation – mismatched IC and LBC perturbations. An additional experiment (DOWN) was performed whereby IC perturbations are fully consistent with LBC perturbations. For this reason, the DOWN experiment serves as a new reference experiment. In addition, all experiments are performed using 1 h LBC update frequency.

As stated by Warner *et al.* (1997), when solutions of the LAM and host model for areas close to lateral boundaries are different, such differences may generate spurious gradients and feedback between the two grids, which can influence the interior of the LAM domain. Given that EDA IC perturbations observed in REF experiment are completely independent of LBC perturbations originating from ECMWF-EPS, such differences can arise.

7.1. Case 1

One such event was observed for REF run started on 17 July 2016 at 0000 UTC. Figure 27a shows synoptic situation on 17 July 2016 at 0000 UTC. An upper-level low (ULL) is located over eastern Europe (not shown) and it is almost entirely inside of the C-LAEF domain. In situations like this, if the host EPS and LAMEPS differently place the ULL, a significant perturbation mismatch is expected to occur.

Figure 28 shows the MSLP spread at the analysis time and for the first hour of model integration. After only 5 min (Fig. 28b), a large spread (with maximum value of roughly 3.5 hPa) can be observed close to the northeastern and along the southern domain boundary. The place of origin of those spread anomalies approximately corresponds to the location where the northern and southern domain borders intersect the ULL. As integration continues (Figs. 28c-d), the anomalies in spread advance further into the domain. After 1 h (Fig. 28d), the most intensive spread is located near the western and eastern lateral boundaries. Comparing the spread relative to DOWN result (Fig. 29a) it may be inferred that the spread is clearly excessive. The fact that excessive spread is still present after 1 h of integration can create serious issues for hourly assimilation cycles (e.g., Caron, 2013) and it needs to be solved.

Given that anomalies form at the boundaries and they are not present in DOWN (Fig. A5 in Appendix) or the host model at all (not shown), it is clear that they are unrealistic. It is well known, in NWP, that initial imbalances in different meteorological fields (i.e., force imbalances) necessarily lead to accelerations which in turn generate spurious waves (e.g., artificial gravity waves and sound waves; Holton, 2004). From Fig. 28., it can be seen that it takes about 1 h for the anomaly to cross the domain in north/south direction. Given the domain size of about 1100 km, the speed of the anomaly is approximately the speed of sound. This means that the likely cause of this phenomena are the horizontally propagating sound waves (Lamb waves) generated by an IC and LBC perturbation mismatch event. These waves are essentially the noise that dominate the forecast fields in the first hour.

Potential solution to this problem could come from so-called space consistent coupling, whereby LBCs at initialization time do not come from the host model but from the LAM analysis (Brousseau *et al.*, 2016). In turn, LBC perturbations are not imposed with full weight from the start and are instead slowly introduced over the first forecast hour to reduce the risk of perturbation conflict. Figure 29b shows the results of such integration after 1 h. Excessive spread is still observed close to the northwestern and eastern lateral boundaries and does not seem to be reduced at all (compared to Fig. 28d), leading to conclude that space consistent coupling is unable to solve this issue for such an event.

However, by using 3-h LBC update frequency, space consistent coupling can solve this problem (not shown) because in this case, LBC perturbations are introduced even more slowly. Nevertheless, given the benefits of higher LBC update frequency (e.g., Nutter *et al.*, 2004a; Termonia *et al.*, 2009), the recent availability of global EPS 1 h files (e.g., ECMWF) and continuing efforts made to improve LAMEPSs in general, we assume that results derived from a 1 h LBC update frequency are more significant.

An alternative approach involves using the ensemble *Jk* method, for which LAM IC perturbations are more consistent with LBC perturbations by design. The results of integration using the ensemble *Jk* method for the same period are shown in Fig. 29c. Excessive spread is largely reduced after 1 h of integration, and overall spread distribution resembles the DOWN pattern most consistently (although the spread is still slightly larger than in DOWN). Increased spread to the east of the domain (also found in later forecast ranges) is also found in DOWN and can be explained by the presence of a ULL over this part of the domain (Fig. 27a).

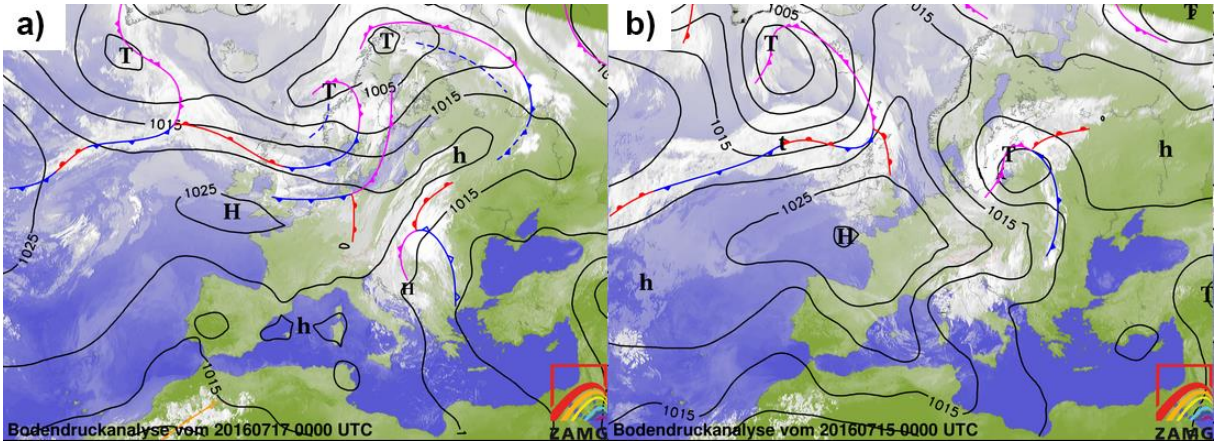

Fig. 27. As Fig. 19., but for (a) 17 July 2016 at 0000 UTC and (b) 15 July 2016 at 0000 UTC.

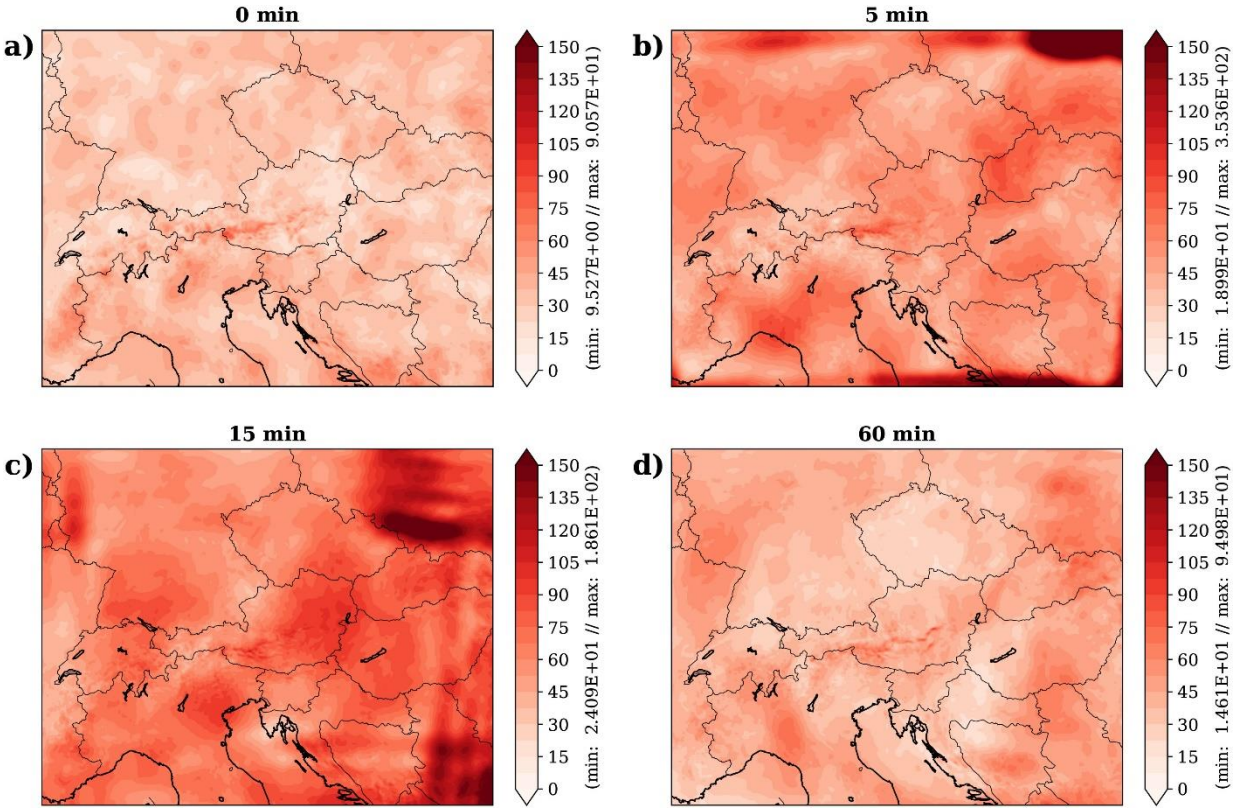

Fig. 28. Surface pressure spread in Pa for REF started on 17 July 2016 at 0000 UTC and valid for (a) analysis time, (b) 0005 UTC, (c) 0015 UTC and (d) 0100 UTC.

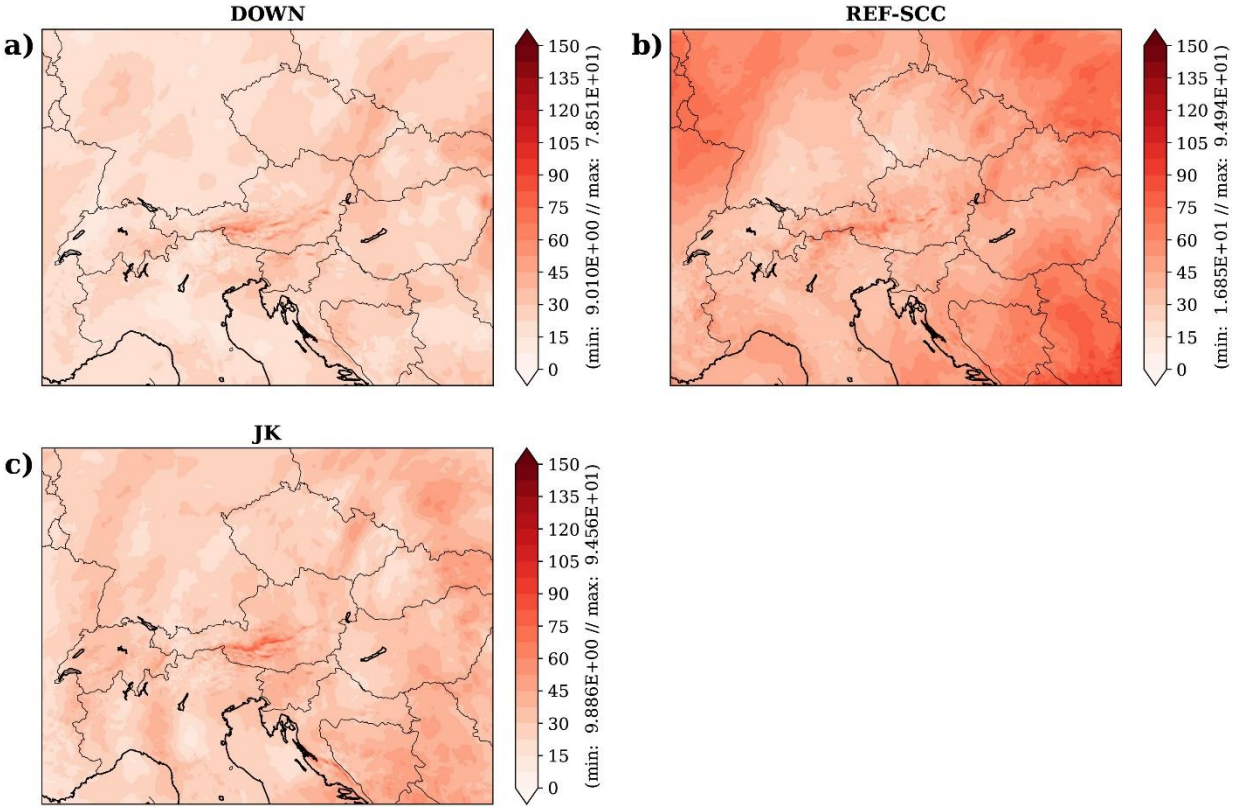

Fig. 29. As in Fig. 28 but after 1 h of model integration for (a) DOWN, (b) REF with space consistent coupling and (c) JK.

7.2. Case 2

Another event where perturbation mismatch had a great impact was observed for REF run started on 15 July 2016 at 0000 UTC. Figure 27b shows synoptic situation on 15 July 2016 at 0000 UTC. Two ULLs are positioned over Europe causing strong pressure gradients (not shown). One of them is located over the Baltic sea and it partially enters northern part of the C-LAEF domain. The other one is located over the bay of Genoa and it is partially contained in the southern part of the C-LAEF domain.

As Fig. 28, Fig. 30 shows the MSLP spread for the first hour of model integration. After 5 min (Fig. 30b), a large spread (with maximum value of roughly 2.6 hPa) can be observed along both the northern and southern domain boundaries. As was in the Case 1, locations of the maximum spread anomalies approximately correspond to the location where the northern and southern domain borders intersect the ULLs. As integration continues (Figs. 29c-d), the anomalies in spread advance further into the domain at a speed close to the speed of sound. After 1 h (Fig. 30d), excessive spread (relative to DOWN result shown in Fig. 31a)

has not left the domain and approaches the western, eastern and southern lateral boundaries. Figure A6 in Appendix shows that no such anomalies are present in DOWN.

Space consistent coupling is, again, unable to resolve this issue (Fig. 31b) as excessive MSLP spread is still present in the larger part of the domain. The result of integration using the ensemble *Jk* method for the same period is shown in Fig. 31c. As before, ensemble *Jk* method improves results - excessive spread is largely reduced after 1 h of integration, and overall spread distribution resembles the DOWN pattern most consistently.

In addition to the two case studies described above, analysis of perturbation mismatching is performed over a longer time period, specifically 12. 7. – 17. 7. 2016. This particular period includes both of our case studies and was characterized by multiple frontal passages and heavy cyclonic activity inside the domain. For this reason, perturbation mismatches were a frequent phenomenon. Figure 32 shows MSLP spread difference (REF – DOWN in blue and JK – DOWN in orange) for 0000 UTC model runs for different forecast ranges over the 6-day period. Excessive MSLP spread is heavily reduced in JK after 15 min, and by about 50 % after 1 h of integration making it very close to the MSLP spread of DOWN.

It needs to be mentioned that one should not expect and aim for JK and DOWN MSLP spread to match exactly. JK experiment generates small-scale perturbations from the initial time, while DOWN downscales ECMWF-EPS large-scale perturbations which does not contain small-scale information. This results in more difference between ensemble members in JK, i.e., more spread. Thus, JK and REF experiments are expected to have more spread than DOWN, but this spread needs to be realistic and not a result of spurious waves.

From all of the results presented in this chapter, we conclude that the ensemble *Jk* method can indeed be used to alleviate the problem of mismatching perturbations and that constraining only the largest scales (above 135 km as discussed in sub-section 4.2.) can achieve this goal.

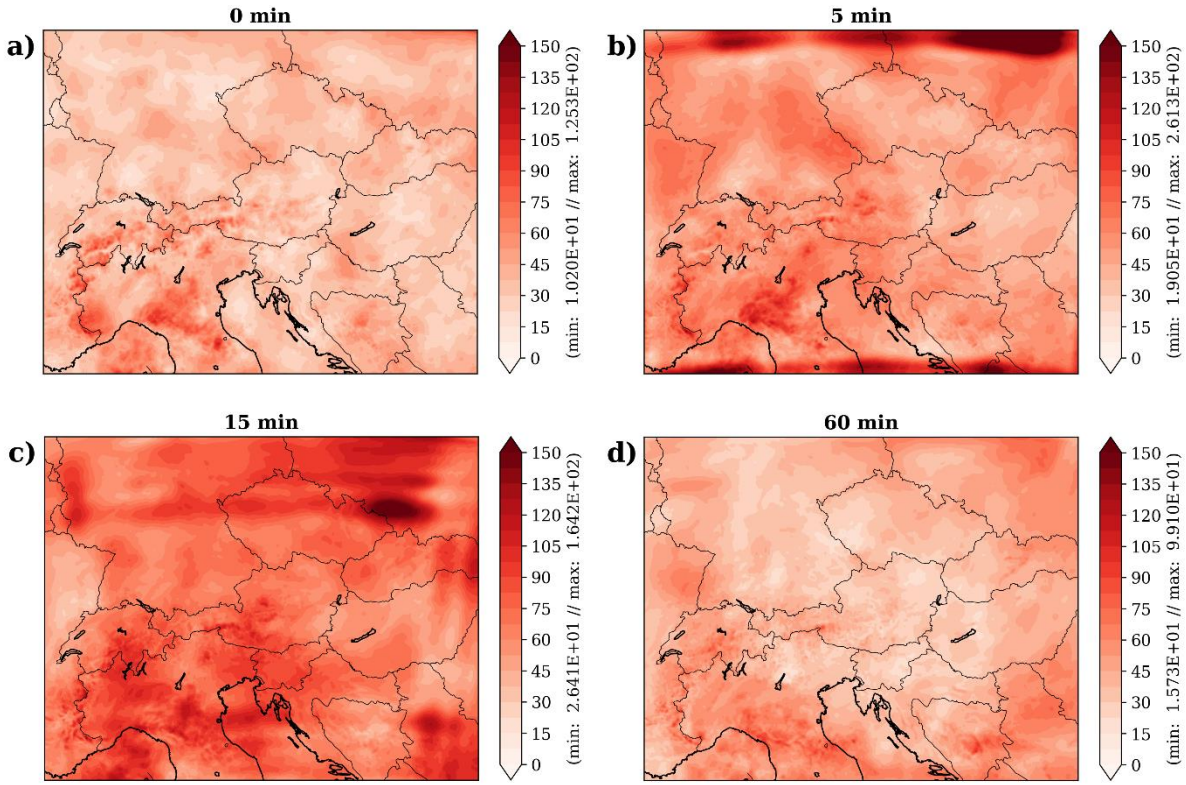

Fig. 30. As Fig. 28. but for REF run started at 15 July 2016 at 000 UTC.

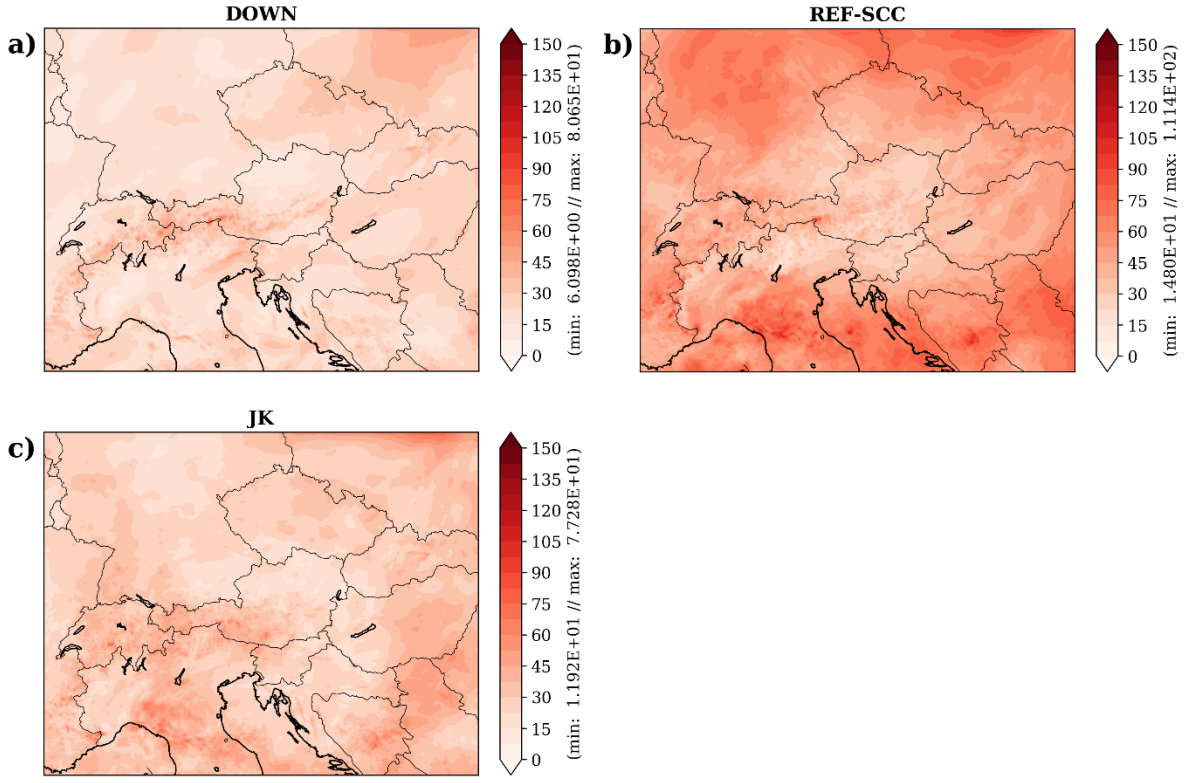

Fig. 31. As Fig. 29. but for 15 July at 0100 UTC.

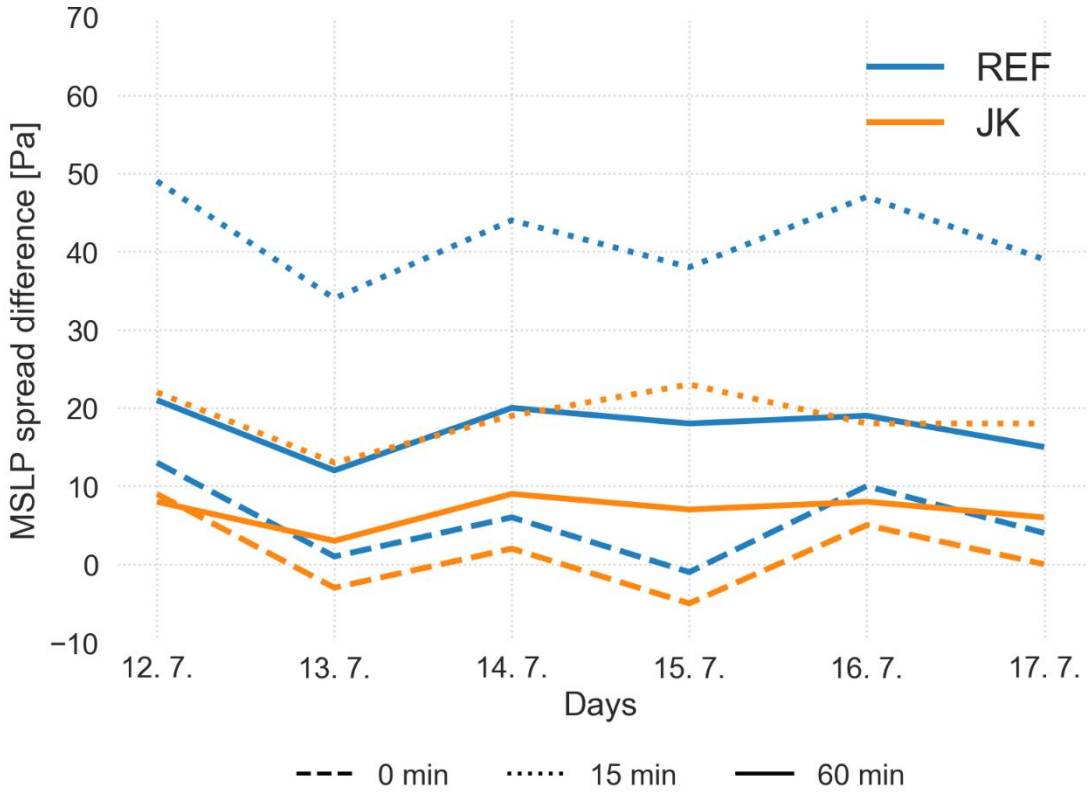

Fig. 32. MSLP spread difference (REF – DOWN in blue and JK – DOWN in orange) for different forecast ranges in minutes: 0 (dashed), 15 (dotted) and 60 (solid) over 6-day period: 12. 7. – 17. 7. 2016.

§ 8. TECHNIQUES TO IMPROVE ENSEMBLE PERFORMANCE

Many studies in the past have dealt with the idea of extending an EPS or even a deterministic system by cheap techniques like LAF (Mittermaier, 2007; Ben Bouallegue *et al*., 2013; Raynaud *et al.*, 2015), neighbourhood approaches (Theis *et al.*, 2005; Ben Bouallegue *et al.*, 2013; Mittermaier, 2014; Schwartz and Sobash, 2017, among others) or combination of many different deterministic models/EPSs (Bowler *et al.*, 2007; Buizza, 2014). Primary goal of those studies was to address the problem of under-dispersion which is also true for C-LAEF as discussed in sub-section 5.1. All mentioned studies reported benefits when using before mentioned techniques.

Here, we will focus on two techniques which can be easily implemented and do not require any extra information or processing. First, neighborhood approach will be explored for the use in C-LAEF. This approach combined with ensemble of forecasts was already used in Ben Bouallegue and Theis (2014) and Schwartz and Sobash (2017) who reported positive effect on precipitation forecast, i.e., increase in ensemble reliability and resolution. Here, we will focus more on the other surface variables. Second, LAF will be explored for use in C-LAEF. This technique was studied in Ben Bouallegue *et al.* (2013), Raynaud *et al.* (2015) and Raynaud and Bouttier (2017) who focused on extending LAMEPS by adding lagged members from the same EPS, while Mittermaier (2007) explored possibility of creating an LAMEPS from lagged runs of a deterministic model. They reported increased reliability and resolution when using LAF. Here, we will take a slightly different approach and combine lagged deterministic AROME forecasts with the C-LAEF ensemble.

8.1. Neighborhood ensemble

8.1.1. Creating a neighborhood ensemble and calculating probabilities

The most straightforward way to extend an EPS by applying the neighborhood approach is to assume that model forecasts at grid points inside the neighborhood constitute a sample from the unknown probability density function of the forecast at the location (x_0, y_0) and the

forecast lead time t_0. In other words, the model forecasts within the neighborhood are assumed to be independent and identically distributed according to this probability density function (Theis *et al.*, 2005). This is a valid assumption considering the arguments given in the sub-section 1.3.4. So, if we have an EPS with N = 10 ensemble members, and choose N_b to be 25, our total number of ensemble members will be N x N_b = 250. In other words, we have increased our ensemble size 25 times with almost no additional computations required. We now proceed to describe how probabilities can be calculated from such an EPS. Let q denote an event threshold (e.g., q = 1,0 mm/h) and f_{ij} forecasts at i = 1, ..., M grid points for j, …, N ensemble members. Then, the binary probability (BP) of event occurrence at the ith point for the jth member (BP_{ij}) is simply:

$$BP(q)_{ij} = \begin{cases} 1 & if \quad f_{ij} \geq q \\ 0 & if \quad f_{ij} < q \end{cases} \tag{8.1}$$

Now, BP_{ij} can be transformed into a neighborhood probability (NP) at i for the jth ensemble member (NP_{ij}) by dividing the number of points within the neighborhood of the i where the event occurs by total number of points in the neighborhood (N_b):

$$NP(q)_{ij} = \frac{1}{N_b} \sum_{k=1}^{N_b} BP(q)_{kj} \tag{8.2}$$

Averaging NP_{ij} over all ensemble members yields the NEP (neighborhood ensemble probability) as defined by Schwartz *et al*. (2010) and called *fuzzy probabilities* by Ben Bouallegue and Theis (2014):

$$NEP(q)_i = \frac{1}{N} \sum_{j=1}^{N} NP(q)_{ij} \tag{8.3}$$

NEP at the ith point can be interpreted as the ensemble mean probability of event occurrence at i given a length scale r. In other words, NEP is the probability calculated from combination of all ensemble members with their neighborhood. We see that calculation of NEP consists of two averaging steps: neighborhood averaging and ensemble averaging. In this case, r can be interpreted as a smoothing length scale. Because of this, NEPs do not possess good reliability or resolution for rare events because of sharpness loss. However, averaging can be very helpful in alleviating the problem of fast error saturation and inherited uncertainty of the small scales. In addition, it enables us to present model information on the scales closer to its true resolution.

There is no particular reason why, in (8.2), we should average all binary probabilities to obtain NP. As an alternate procedure, we can search the neighborhood of *i* to determine whether the event has occurred at any grid point within that neighborhood of *i* to redefine event occurrence at *i*. In other words, event occurrence at *i* can be determined by searching for the maximum value within the neighborhood of *i* and comparing it to *q*. This is somewhat relaxed condition for an event occurrence, because it needs be forecasted anywhere within the neighborhood. Thus, when these searching approaches are applied, NP will always be 1 or 0 and can be redefined to a binary neighborhood probability (BNP):

$$BNP(q)_{ij} = \max\left(\sum_{k=1}^{N_b} BP(q)_{kj}\right) \tag{8.4}$$

Then, by averaging over all ensemble members, neighborhood maximum ensemble probability (NMEP) is obtained:

$$NMEP(q)_i = \frac{1}{N}\sum_{j=1}^{N} BNP(q)_{ij} \tag{8.5}$$

NMEP is the probability of event occurrence within the *r* of *i* and, thus, is a probability defined over a spatial scale larger than a grid size. Unlike the NEP, production of NMEPs does not contain spatial smoothing. As a consequence, it does not suffer from a lack of sharpness for rare events. Using this analogy, we can create a new *N*-member EPS by applying any function f (i.e., maximum, minimum, median, etc.) to an N_b-member neighborhood before forming an EPS. In contrast to the original *N*-member EPS, this new one will have members with the 'best' information extracted from their respective neighborhoods.

To recapitulate, we have created an *N* x N_b-member EPS by combining *N* member EPS with its N_b -neighborhood. Probabilities calculated from that EPS, by averaging, are called the NEP. Alternatively, probabilities can be calculated by searching the neighborhood for occurrence of an event. Those probabilities are called the NMEP. To create a NMEP-like EPS, we take a function f of each member's neighborhood separately to obtain a new *N*-member EPS. More detailed application of NEP and NMEP for precipitation forecasts can be found in Ben Bouallegue and Theis (2014) or in Schwartz and Sobash (2017). They found that NEP is more applicable to light and moderate precipitation due to its smoothing nature, while NMEP is significantly better at detecting high precipitation events. In this study, we will focus our attention to other surface variables (MSLP, T2M, RH2M and W10M), try to

assess the benefits and adopt the NEP and NMEP-like approaches to the continuous variables, such as T2M.

8.1.2. Model setup and experiments

Our goal is to assess the added value of the neighborhood approach on the C-LAEF ensemble. For this reason, we define two experiments:

a) REF – represents C-LAEF raw ensemble (N = 17)
b) NGH – represents REF with the neighborhood approach.

Except for the neighborhood, both experiments are configured as the REF experiment from sub-section 3.2.

First, we will configure NGH as NEP. This means that we will increase our ensemble size from 17 to 17 x N_b and each member will be treated as independent and equally likely. In such a configuration, probabilities, ensemble mean and spread will be calculated from 17 x N_b members. Different values for r were tested (and thus for N_b), as previous studies reported that no universally optimal value for neighborhood size exists, but depend on variable and threshold (e.g., Mittermaier, 2014). Table 2 shows chosen neighborhood sizes for all the variables. In addition, attention must be paid to the orography and land-sea boundary around ith point. For example, temperature and humidity can vary greatly between land and sea points. The same is true for the points with greatly different altitude. We do not want, for example, to include mountainous model point from Medvednica to the neighborhood of Zagreb when forecasting T2M. For those reasons, filtering of the model points based on altitude and land-sea (L-S) mask was performed. Table 2 shows values of altitude tolerance (AT) which were applied to all the variables. Any point whose altitude differed by more than AT from the altitude of i was discarded from the neighborhood. Similarly, if L-S filtering was True, all sea points were discarded from the neighborhood. This means that total ensemble size varies for each location and will be less or equal to 17 x N_b. Table 2 shows the maximum possible ensemble size for each variable. Thus, it is necessary to make the neighborhood a function of a specific location and variable for an optimal performance.

Unfortunately, only 3-h model output data were available for this study. This prevents us from studying the expansion of the neighborhood into the time dimension as 3-h gap between two successive model output times is too big. This is unfortunate because the temporal uncertainty of the model forecasts can be of a significant magnitude (see Mittermaier and Stephenson, 2015).

Table 2. Chosen values for neighborhood length scale (*r*), altitude tolerance (AT) and land-sea (L-S) filtering.

Variable	r [grid points]	AT [m]	L-S filtering	Max. ensemble size
MSLP	7	100	False	833
T2M	7	100	True	833
RH2M	7	100	True	833
W10M	5	200	True	425

8.1.3. Verification

Differences between REF and NGH will be demonstrated for surface variables, namely MSLP, T2M, RH2M and W10M. The verification methodology and the period used are the same as described in sub-section 3.3. with the only difference being number of surface stations used. Due to the very high number of ensemble members when neighborhood is used, performing any calculation requires a high amount of computer memory. For this reason, we used a lower number of surface stations to keep the amount of data manageable - 30 surface stations within Croatia are used to perform the verification.

8.1.4. Results - NEP

a) *RMSE of ensemble mean and spread*

Figure 33 shows RMSE of ensemble mean and spread for all variables. Accuracy of the ensemble mean is increased for RH2M and T2M (although without passing a significance test) and remained almost the same for W10M and MSLP. Ensemble spread is significantly increased for T2M, RH2M and W10M, while remaining the same for MSLP. MSLP is a type of variable that mostly reflects large-scale features and we do not expect noticeable pressure variations inside of the local neighborhood which is consistent with results obtained here. Other variables are much more local in nature, and thus, the impact of neighborhood is obvious. Ensemble spread is significantly increased as variations between neighboring model points are much larger. Combination of increased spread and slightly reduced RMSE leads to a more reliable EPS and reduced under-dispersion.

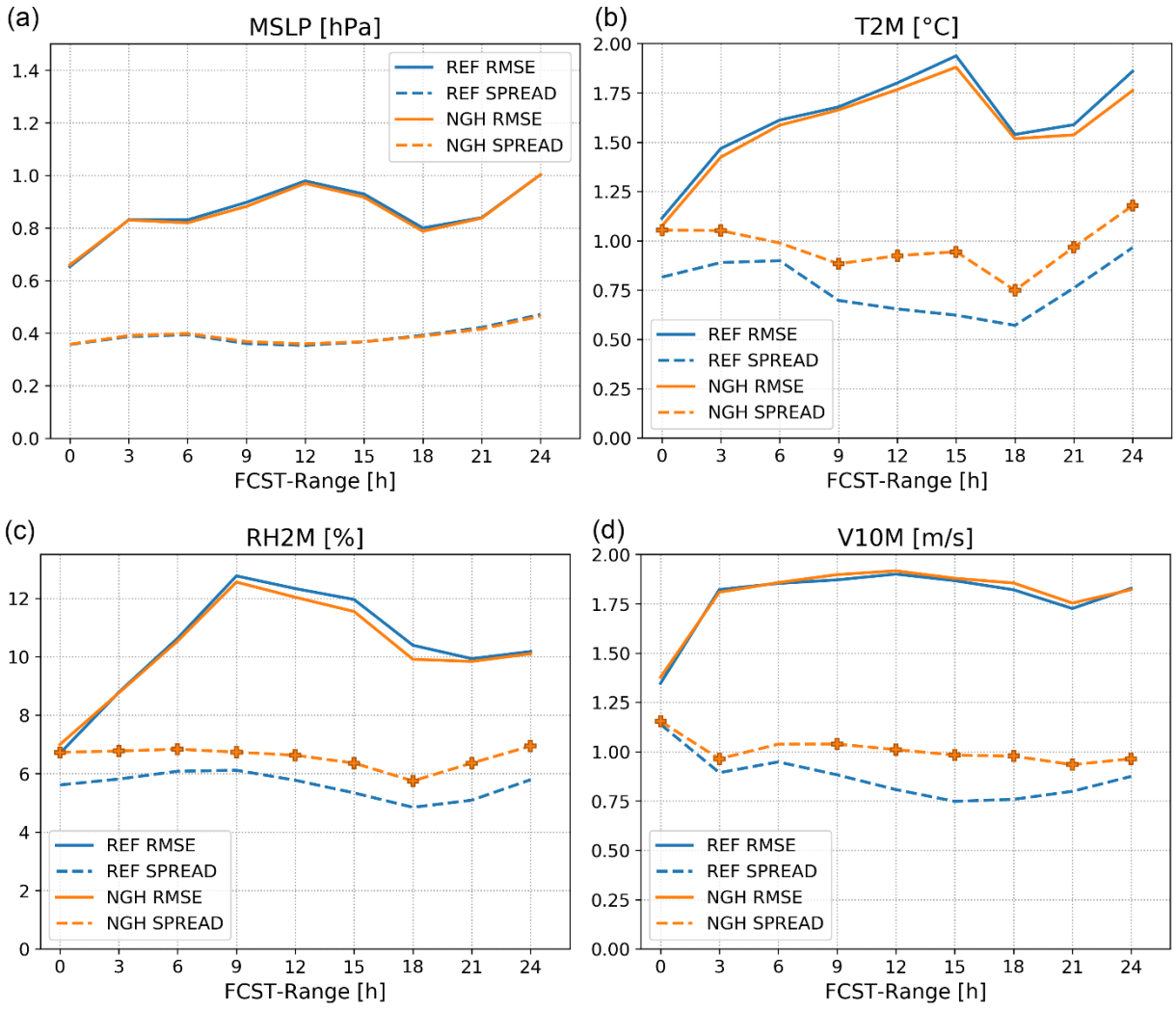

Fig. 33. RMSE of the ensemble mean (solid) and spread (dashed) of REF (blue) and NGH (orange) for (a) MSLP, (b) T2M, (c) RH2M and (d) W10M for the verification period. Forecast ranges with statistically significant difference are marked with bullets (RMSE) and crosses (spread).

b) *CRPS*

CRPS, for all variables, is shown in Fig. 34. In consistency with RMSE/spread relation, reduction of CRPS is visible for all variables in NGN, mostly for T2M and RH2M, while for W10M is visibly reduced but not statistically significant.

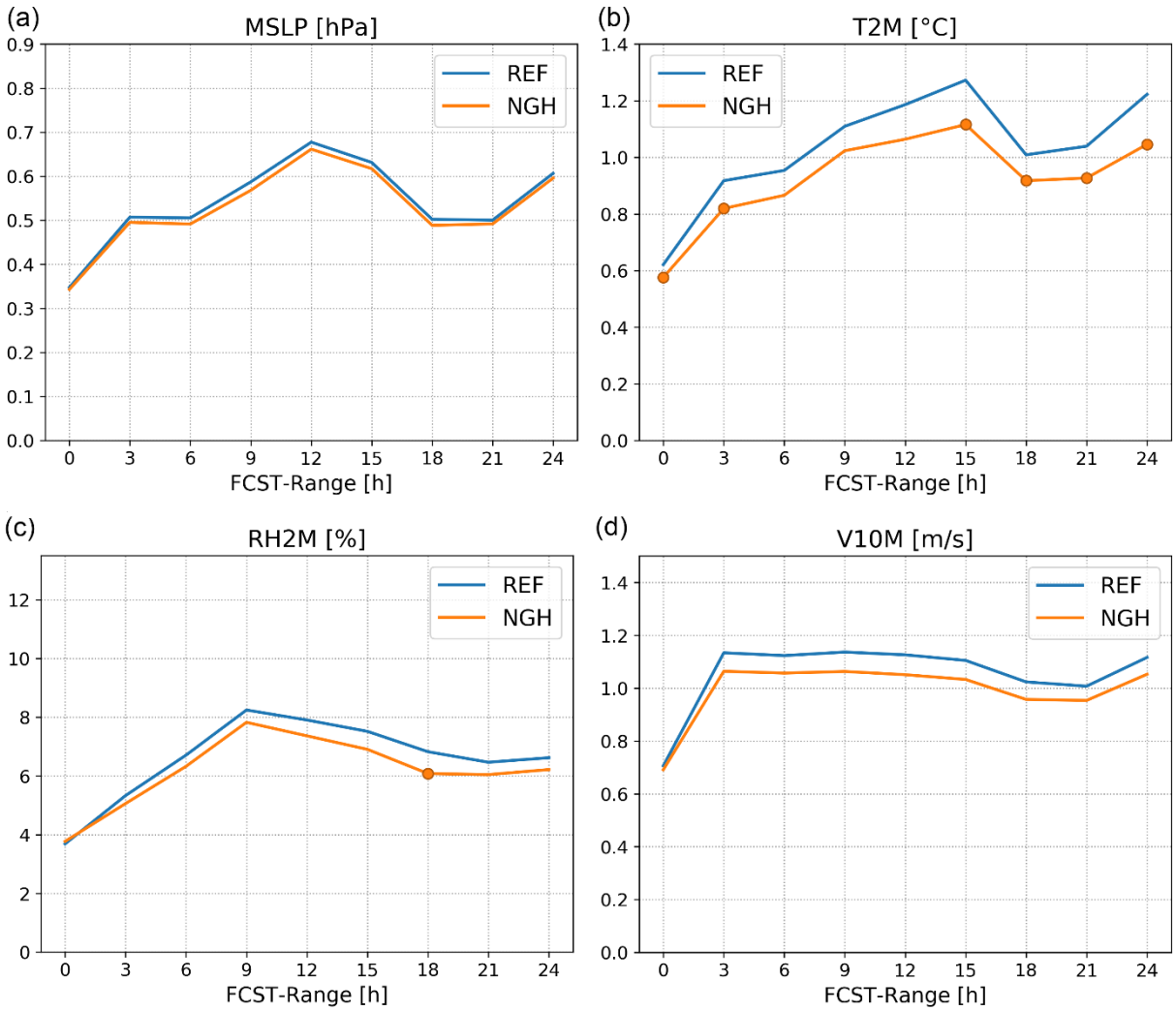

Fig. 34. CRPS of REF (blue) and NGH (orange) for (a) MSLP, (b) T2M, (c) RH2M and (d) W10M for the verification period. Forecast ranges with statistically significant difference are marked with bullets.

c) *Outlier statistics*

Figure 35 shows outlier statistics for all variables. As described in section 5., perfect value of outliers depends on the ensemble size. For REF, perfect value is 11 % (dashed blue) and for NGH (dashed orange) it is approximately zero. Nevertheless, percentage of outliers is greatly reduced in NGH for all variables. As a consequence of the larger ensemble size, more observations will be contained within the EPS.

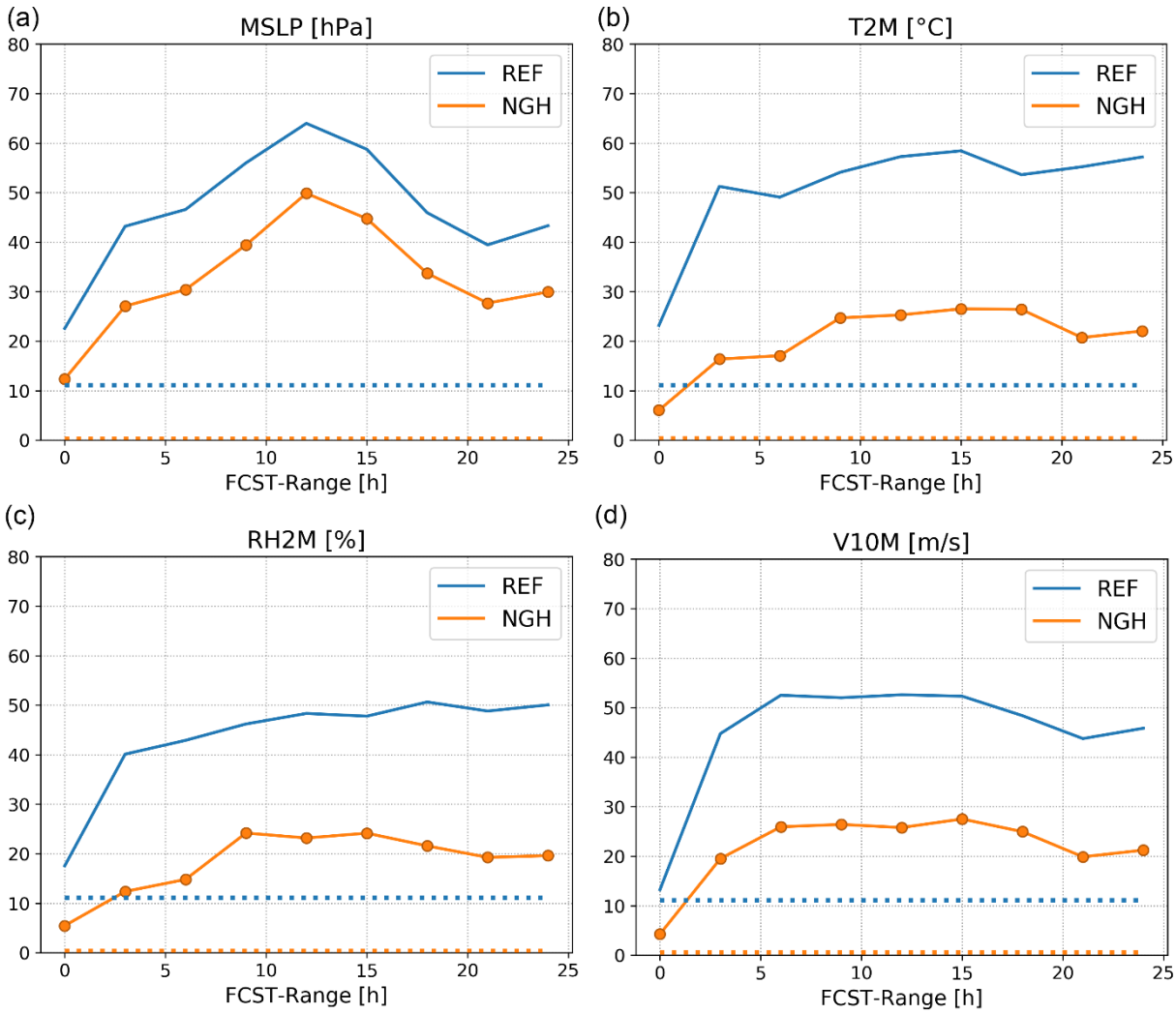

Fig. 35. Percentage of outliers of REF (blue) and NGH (orange) for (a) MSLP, (b) T2M, (c) RH2M and (d) W10M for the verification period. Forecast ranges with statistically significant difference are marked with bullets. Dashed blue (orange) line denotes an ideal value for REF (11 %) and NGH (~ 0) %.

d) *Brier score decomposition*

To check the accuracy of NGH obtained probabilities, as well as its reliability and resolution, BS decomposition was calculated for the verification period and averaged over all forecast ranges as shown on Fig. 36. As expected, no significant improvement is obtained for MSLP, while for the other variables, BS is decreased for majority of the thresholds. Moreover, those improvements are due to the improved, both, reliability and resolution.

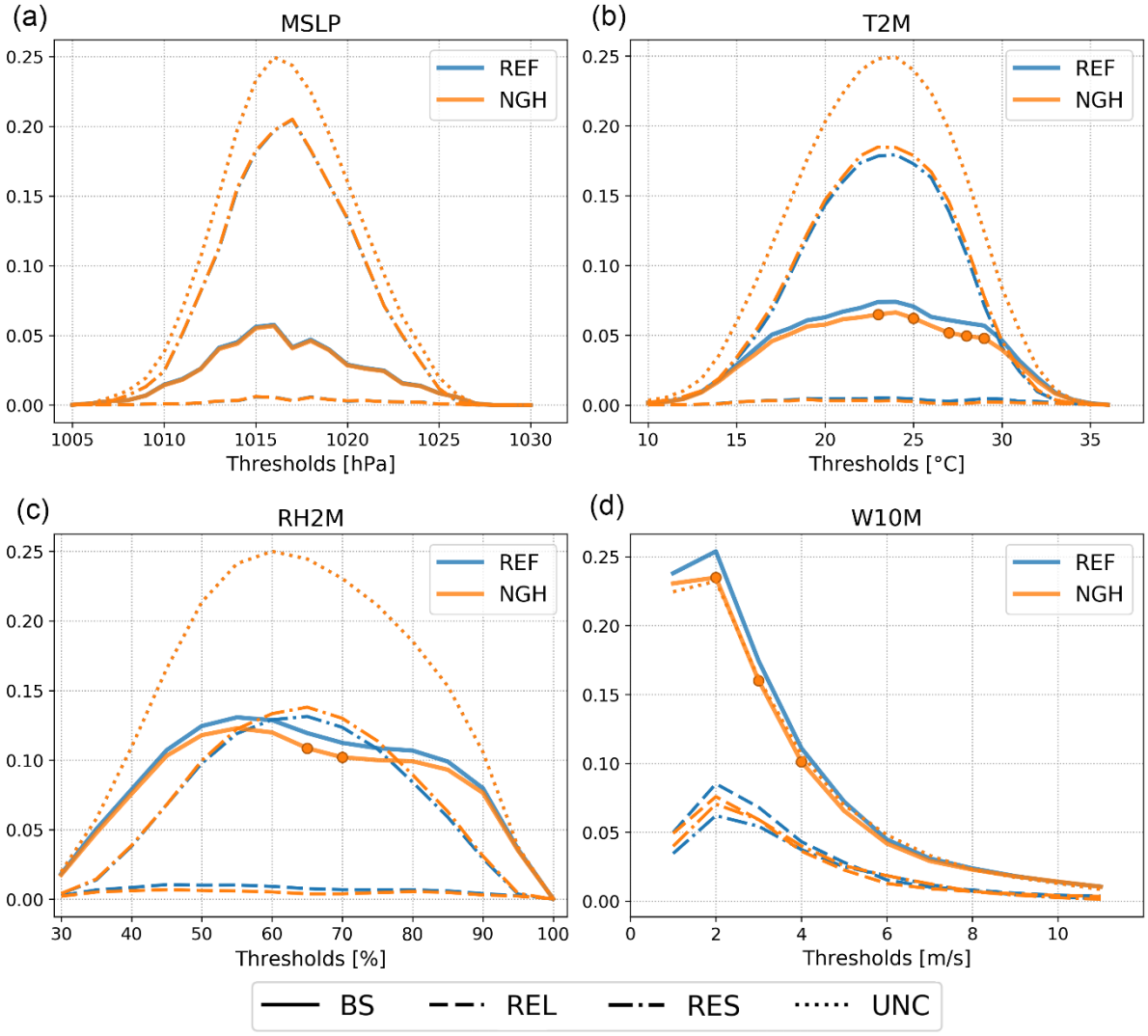

Fig. 36. Decomposition of the BS for REF (blue) and JK (orange) for (a) MSLP, (b) T2M, (c) RH2M and (d) W10M for the verification period and averaged over all forecast ranges. Thresholds with statistically significant differences (only for BS) are marked with bullets.

e) *ROC*

To check the attribute of discrimination and the ability of NGH to detect extreme values, ROC was calculated for different thresholds. Figure 37, shows ROC curve for T2M, RH2M and W10M and for relatively high thresholds (above 32 °C for T2M, above 90 % for humidity, above 2 and 8 ms^{-1} for W10M) averaged over all forecast ranges. Setting probability thresholds (the same as in sub-section 5.5.) to constant values has to be done in order to properly compare EPSs of different size. NGH has an improved discrimination

for all variables, except MSLP, and all but the most extreme thresholds. For the most extreme thresholds, improvements are visible only for T2M.

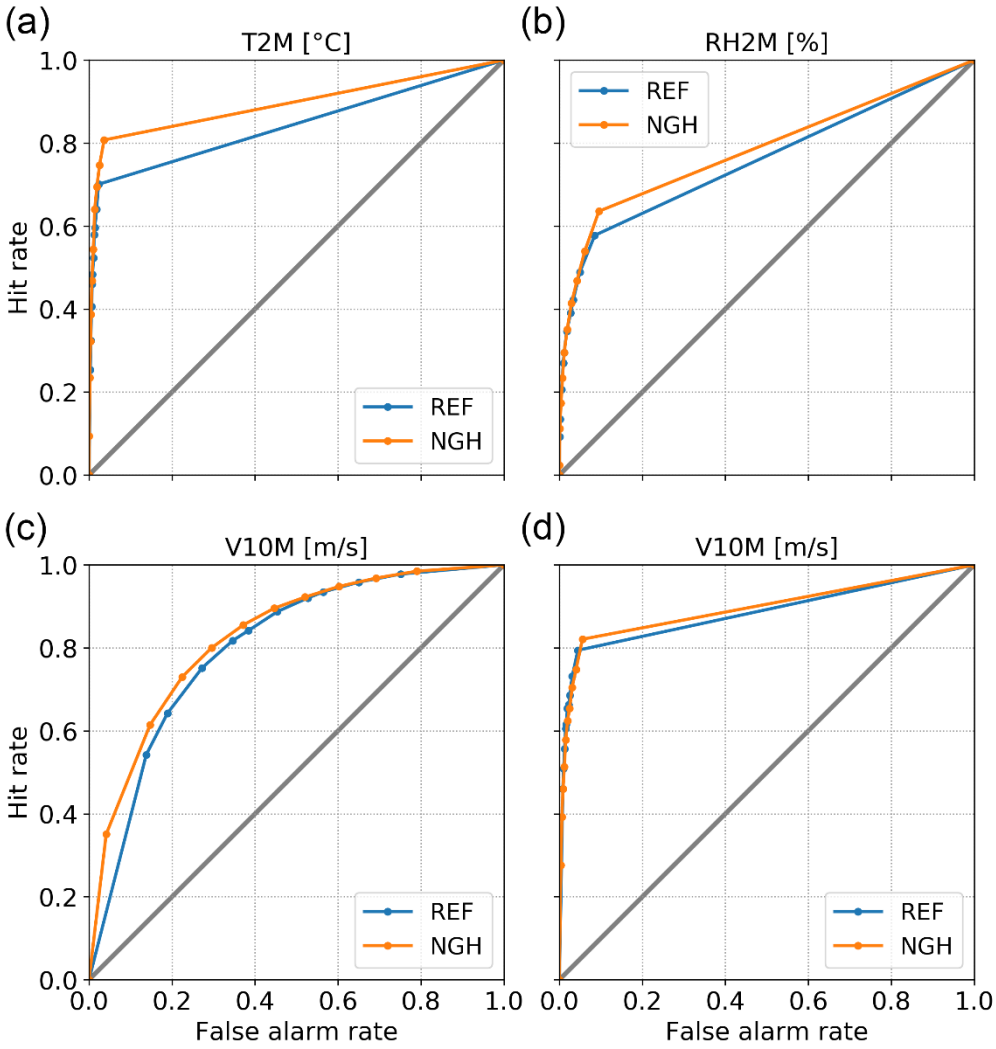

Fig. 37. ROC curves averaged over all forecast ranges of REF (blue) and NGH (orange) for (a) T2M (threshold: > 32 °C), (b) RH2M (threshold: > 90 %), (c) W10M (threshold: > 2 ms^{-1}) and (d) W10M (threshold: > 8 ms^{-1}) for the verification period.

8.1.5. Results - NMEP

As already mentioned in the sub-section 8.1.1., NMEPs are better at predicting the extreme events (e.g., extreme precipitation). To check the usefulness of the NMEP-like approach for another variable for which predicting extremes is very important - wind speed, NGH was

configured as NMEP and relabeled to NGH_NMEP. This time, instead of doing both, neighborhood and ensemble averaging when calculating probabilities, we take the maximum value of the neighborhood and then do the ensemble averaging (equations (8.4) and (8.5)). In other words, we search the neighborhood of a specific point *i* to determine whether the event has occurred at any grid point within the neighborhood of *i* to redefine event occurrence at *i*. Figure 38 shows the ROC curves of REF, NGH and NGH_NMEP for wind speed above 10 ms^{-1} averaged over all forecast ranges and the verification period. NGH_NMEP demonstrates significantly improved discrimination for high wind speeds. This means that NGH_NMEP is better at resolving events from non-events, a property which can be of a significant importance when issuing warnings.

In the former sub-sections, we saw that neighborhood contains valuable information that, when extracted properly, can hugely benefit a forecast. Is there any other way to extract even more information from the neighborhood? The answer is yes, there is. A NMEP-like approach can also be adopted for continuous variables like T2M and predicting extremes like minimum and maximum temperatures. As described in sub-section 8.1.1., NEMP-like EPS with 17 members is created by applying any function *f* to the neighborhood of each member separately. Here, *f* is the maximum function (8.4), but applied to absolute forecast values instead of BPs.

Here, an application of NMEP on predicting the next day minimum temperatures is demonstrated. Figure 39a shows 12 UTC ensemble mean temperature forecast of REF (blue) and corresponding observations (grey) averaged over the verification period. A significant warm bias is evident for the temperatures at almost all forecast ranges, especially at the morning of the next day. The next day minimum temperature RMSE is equal to 1.95 °C. A new NGH_NMEP is configured so that *f* is equal the minimum value in the neighborhood. This new NGH_NMEP is shown as an orange curve in Fig. 39a. In NGH_NMEP, warm bias is eliminated and RMSE for the next day minimum temperature is 1.55 °C which is a reduction of about 20 %. In addition, Fig. 39b shows ROC curves of REF and NGH_NMEP for temperatures below 15 °C (a very low value for a summer morning in Croatia), 12-18 h forecast range (corresponds to 0200-0800 h local time) and averaged over verification period. NGH_NMEP, again, visibly improves the ability of discrimination for a low T2M threshold, meaning that it is more capable in detecting extreme events.

8.1.6. Summary and discussion

As already noted in Ben Bouallegue and Theis (2014), apparently, forecast quality benefits from shifting the focus away from the specific location towards a somewhat broader region. This, of course, comes at the cost of omitting fine-scale information. However, a theoretical justification for this approach lies in the two theses established in sections 1.3. and 1.3.4.: grid size vs resolution and a quick predictability loss which renders small-scale information useless anyway. We did not increase the *r* beyond 7 because we wanted to stay close to the model true resolution and, also, maximize the output resolution. However, previous studies have reported positive impact on precipitation forecasts of increasing the *r* up to 40 times the grid size (e.g., Ben Bouallegue and Theis, 2014).

Results presented here indicate that a cheap technique like the neighborhood approach can greatly benefit an EPS. Accuracy, reliability, resolution and discrimination are increased for T2M, RH2M and W10M. The problem of under-dispersion is greatly reduced. Impact is smallest on MSLP, because MSLP is mostly large-scale variable and does not vary much within the neighborhood. In addition, the ability of EPS to detect extreme events is also improved. However, detection of the extreme events is greatly improved only when NGH is configured as NMEP. Based on the results presented here, we can make the following recommendations. Using EPS as NEP and increasing its size by N_b is useful for general applications. For example, forecasting light to medium rain or weak to moderate wind speeds, hourly temperature or humidity forecasts, etc. For predicting the extremes, for example, high rain and wind speeds, minimum or maximum temperatures, EPS as NMEP is recommended. Ben Bouallegue and Theis (2014) noted that DWD (Deutscher Wetterdienst) forecasters favour NEMP when issuing warnings. Finally, making the neighborhood a function of a specific location and variable is highly advisable.

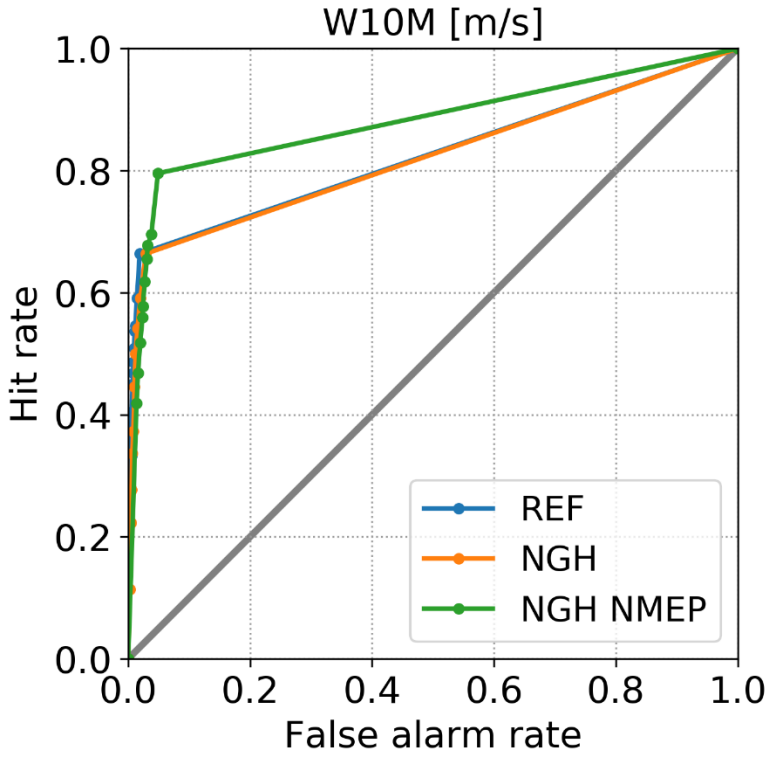

Fig. 38. ROC curves averaged over all forecast ranges of REF (blue), NGH (orange) and NGH NMEP (green) for W10M (threshold: > 10 ms^{-1}) and the verification period.

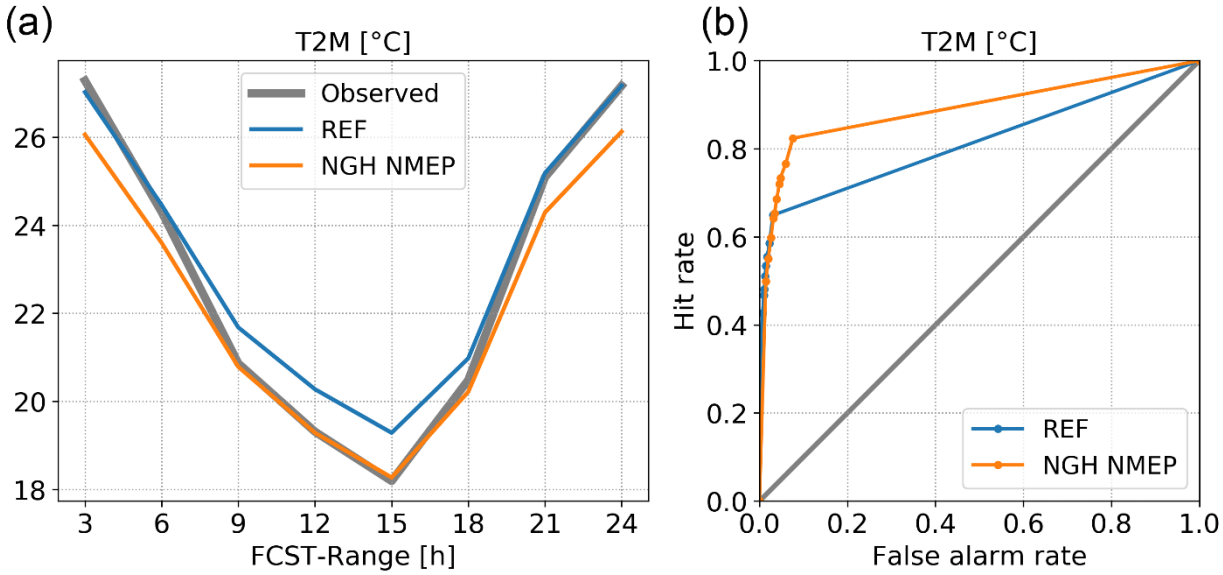

Fig. 39. (a) 12 UTC ensemble mean temperature forecast of REF (blue), NGH NMEP (orange) and corresponding observations (grey) averaged over the verification period. (b) ROC curve for T2M (threshold: < 15 °C), 12-18 h forecast range and averaged over the verification period.

8.2. Adding lagged deterministic forecasts

In this sub-section, we explore a unique possibility – extending an LAMEPS with lagged deterministic forecasts. This setup is possible because, at ZAMG, operational deterministic AROME system is configured the same (except for 3-h cycling) as C-LAEF (C-LAEF control to be precise). AROME forecast range is 60 h that enables us to use up to 4 older AROME runs and combine them with C-LAEF to create a new 21-member ensemble and keep 48 h forecast range that will be used in operations. Because of this, we expect to have an interchangeable 21 member EPS where all members are equally likely.

LAF approach can improve EPS's IC uncertainty estimate because the additional members from LAF are generated from different initializations, which represent the atmospheric state in a flow-dependent and time-evolving manner. Thus, IC perturbations are extended by the flow-dependant information from the past forecasts. Under more "predictable" regimes, one would expect the differences between successive forecasts (realizations) to be small, while under more "volatile" flow regimes, one would expect the spread in the forecasts to be larger (Mittermaier, 2007). Raynaud and Bouttier (2017) showed that LAF did significantly increase ensemble spread. As it was shown in section 5., larger spread is beneficial for C-LAEF.

Special weighting strategy of LAF members will not be applied and all members will be weighted equally as previous studies have already shown its negligible impact on time-lagged ensembles (Ben Bouallegue *et al.,* 2013; Raynaud *et al.,* 2015). In our case, this is even more valid as we are adding only 4 additional members.

8.2.1. Model setup and experiments

Our goal is to assess the added value of including 4 lagged AROME members to the C-LAEF ensemble. For this reason, we define two experiments:

a) REF – C-LAEF raw ensemble (17 members).
b) LAG – C-LAEF plus 4 lagged AROME runs (-3, -6, -9 and -12 h; 21 members).

Both AROME and C-LAEF are configured as described in the sub-section 3.1. Lagged members and C-LAEF control are coupled to the high-resolution ECMWF IFS HRES.

8.2.2. Verification

Methodology used here is similar to the one described in sub-section 3.3., C-LAEF 00 UTC hourly runs were archived for the period 19. 5. – 30. 6. 2019. with two missing days (26. 5. and 2. 6.). This gives us the total of 41 days for which to perform a verification. Forecasts for variables were archived – precipitation, W10M, wind gusts at 10 m (WG10M) and surface global short-wave radiation. Verification was performed against 320 automatic surface stations within Austria by matching nearest model grid point to the observation location. For precipitation, the same methodology as described in sub-section 3.3. is used, i.e., verification against high resolution INCA analyses.

Decomposition of CRPS to a term related to spread and a term related to mean absolute error (Leutbecher, 2018), and decomposition of RMSE to bias, bias of standard deviation and dispersion (Horvath *et al.*, 2012) will help us take a deeper insight into the reasons behind the obtained results.

8.2.3. Results

Verification for surface global short-wave radiation gave neutral results for all scores and will not be discussed further. Also, traditional verification for precipitation yielded similarly neutral results for most of the scores and will not be discussed further. Precipitation results will be discussed in more detail in terms of FSS. We will now present results obtained by averaging over the entire verification period of 41 days.

a) *RMSE of ensemble mean and spread*

Figure 40 shows RMSE and spread of REF and LAG for W10M and WG10M. As we can see, there is hardly an impact on accuracy of the ensemble mean for W10M and WG10M, while spread is significantly increased in both cases implying better reliability and reduced under-dispersion of LAG. Decomposition of RMSE revealed that bias of W10M (WG10M) is slightly reduced (increased) in LAG. This increase in bias is countered by lower dispersion error in LAG (not shown).

Looking more thoroughly at WG10M plots, one can observe strange behavior (peaks) of RMSE curve at +0 and +1 h, but also at +6, +12 … at 6-h intervals. Those were identified as problems in the pre-operational C-LAEF forecasts that need further investigation.

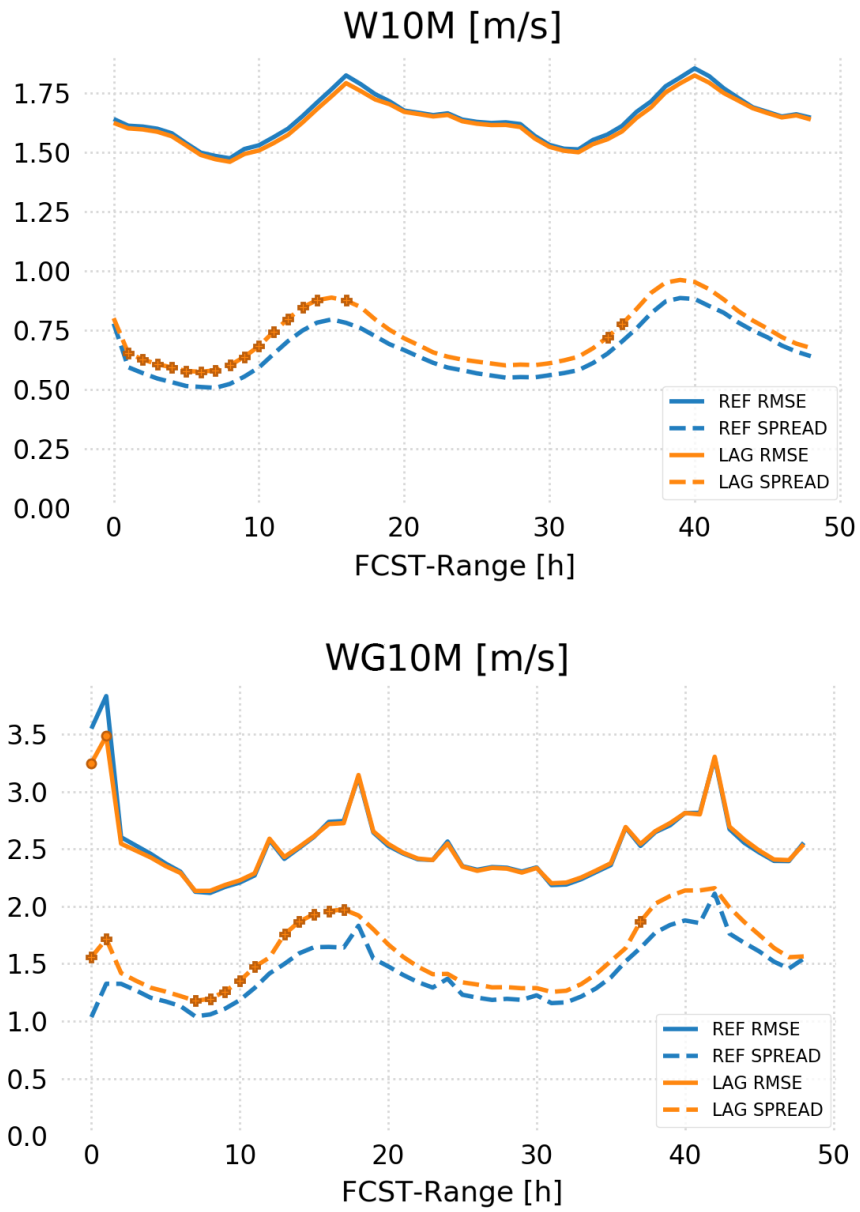

Fig. 40. RMSE of the ensemble mean and ensemble spread for W10M (up) and WG10M (down). Forecast ranges with statistically significant differences are marked with bullets.

b) *CRPS*

Figure 41 shows CRPS and its decomposition for W10M and WG10M. We see a slight reduction of CRPS in LAG for both variables, although not statistically significant. CRPS decomposition shows that this reduction of CRPS comes from significant increase in ensemble spread in LAG.

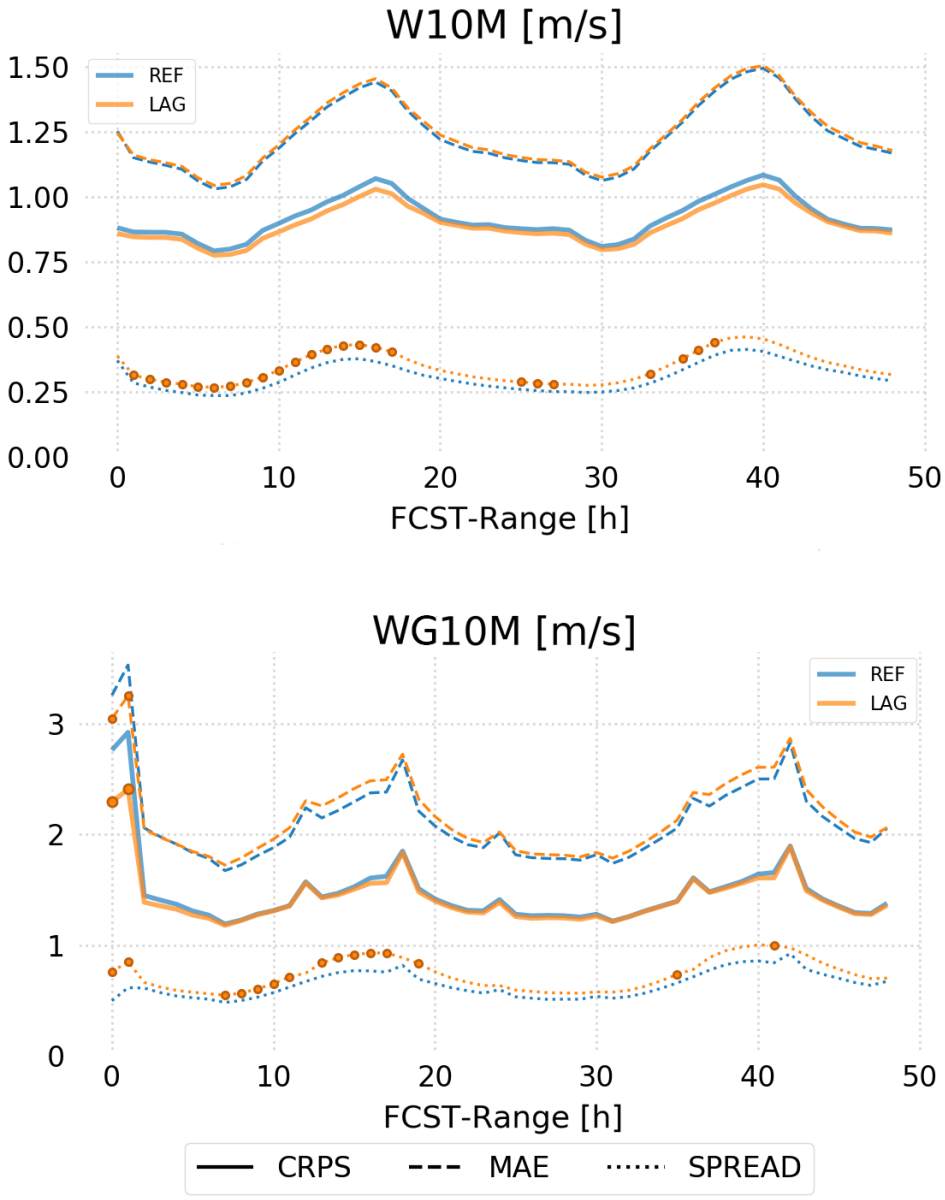

Fig. 41. Decomposition of CRPS for W10M (up) and WG10M (down). Forecast ranges with statistically significant differences are marked with bullets.

c) *Outlier statistics*

Percentage of outliers is given in Fig. 42. Horizontal dashed lines denote ideal value for each experiment. It is very clear that the number of outliers is reduced in LAG meaning reliability is increased. This is consistent with the findings in 8.2.3 a).

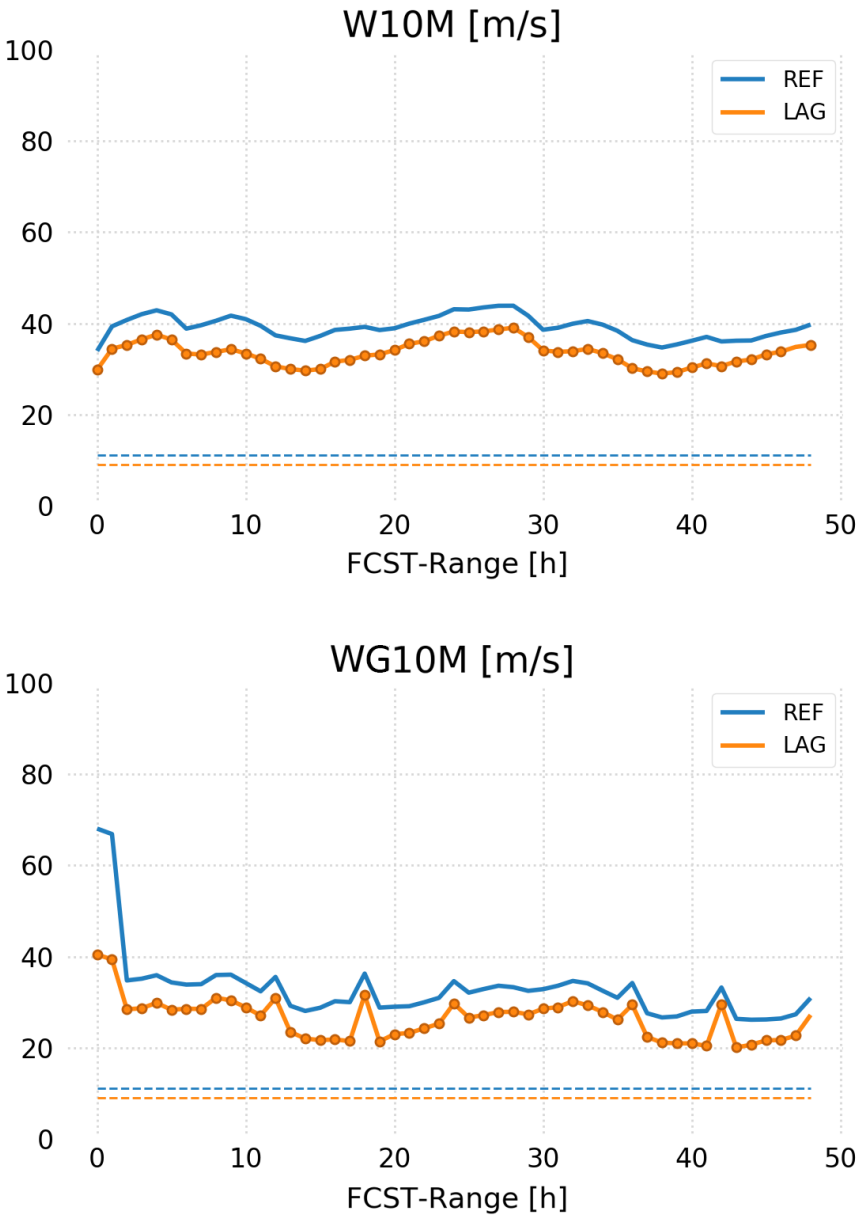

Fig. 42. Percentage of outliers for W10M (up) and WG10M (down). Forecast ranges with statistically significant differences are marked with bullets. Dashed horizontal lines denote ideal values of 11 % (9 %) for REF (NGH).

d) *ROC*

Figure 43 shows ROC curve and ROC area for thresholds of 3 and 5 ms^{-1} (10 and 15 ms^{-1}) for W10M (WG10M). For all thresholds and variables, discrimination is slightly improved in LAG.

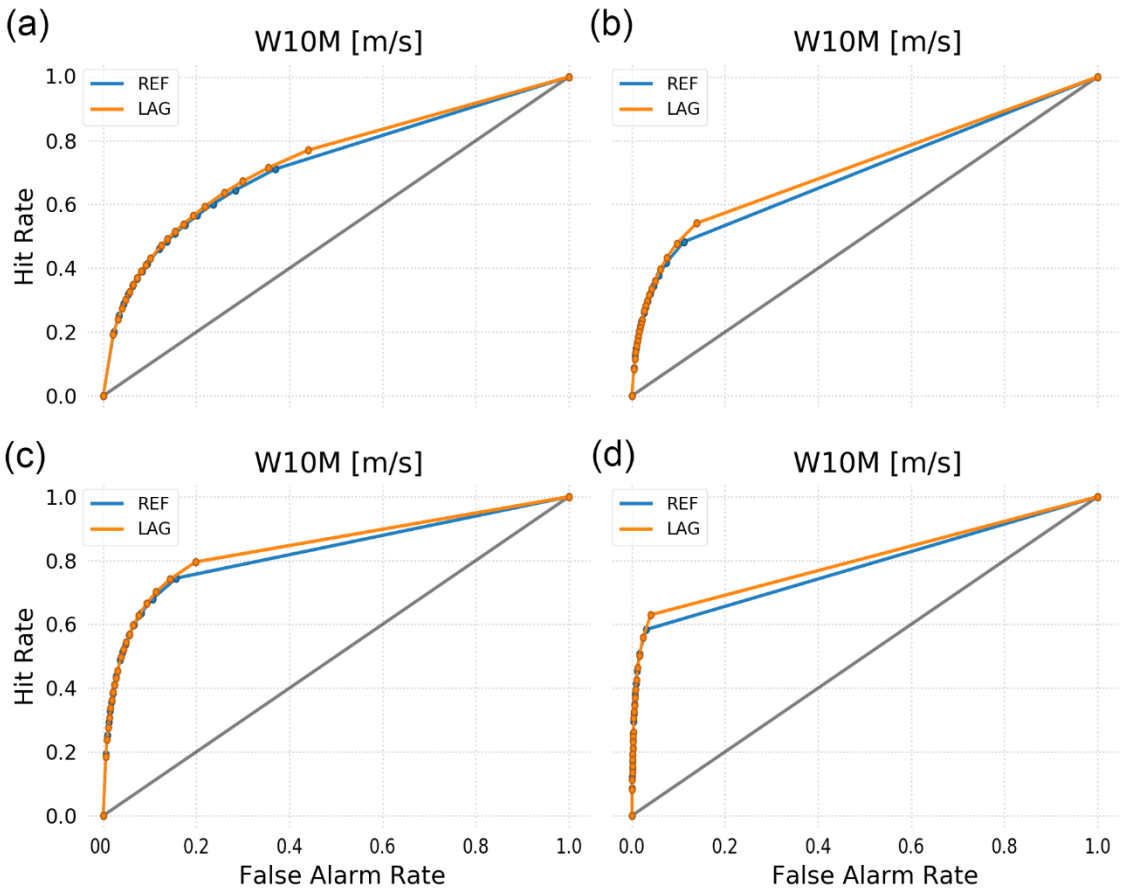

Fig. 43. ROC curve and ROC area for: (a) W10M (3 ms^{-1}), (b) W10M (5 ms^{-1}), (c) WG10M (10 ms^{-1}) and (d) WG10M (15 ms^{-1}).

e) *Brier score decomposition*

Decomposition of BS for different thresholds confirms previous conclusions that LAG is more reliable ensemble, while resolution is similar for both experiments (not shown).

f) *Precipitation*

The left panel in Fig. 44 (similar to Fig. 18) shows the median of the Skill Score of FSS of LAG to FSS of REF illustrated as a matrix of colours (red means LAG is better than REF and blue the opposite, white is for no difference) for different thresholds, scales and forecast ranges. The right panel shows the percentage of times (days) FSS of LAG is higher than the FSS of REF. FSS is visibly improved in LAG for the first 5 h of integration mostly for scales higher than 25 km and thresholds up to 2 mm. Small improvements are still visible up to 20 h of integration, after which results become mostly neutral.

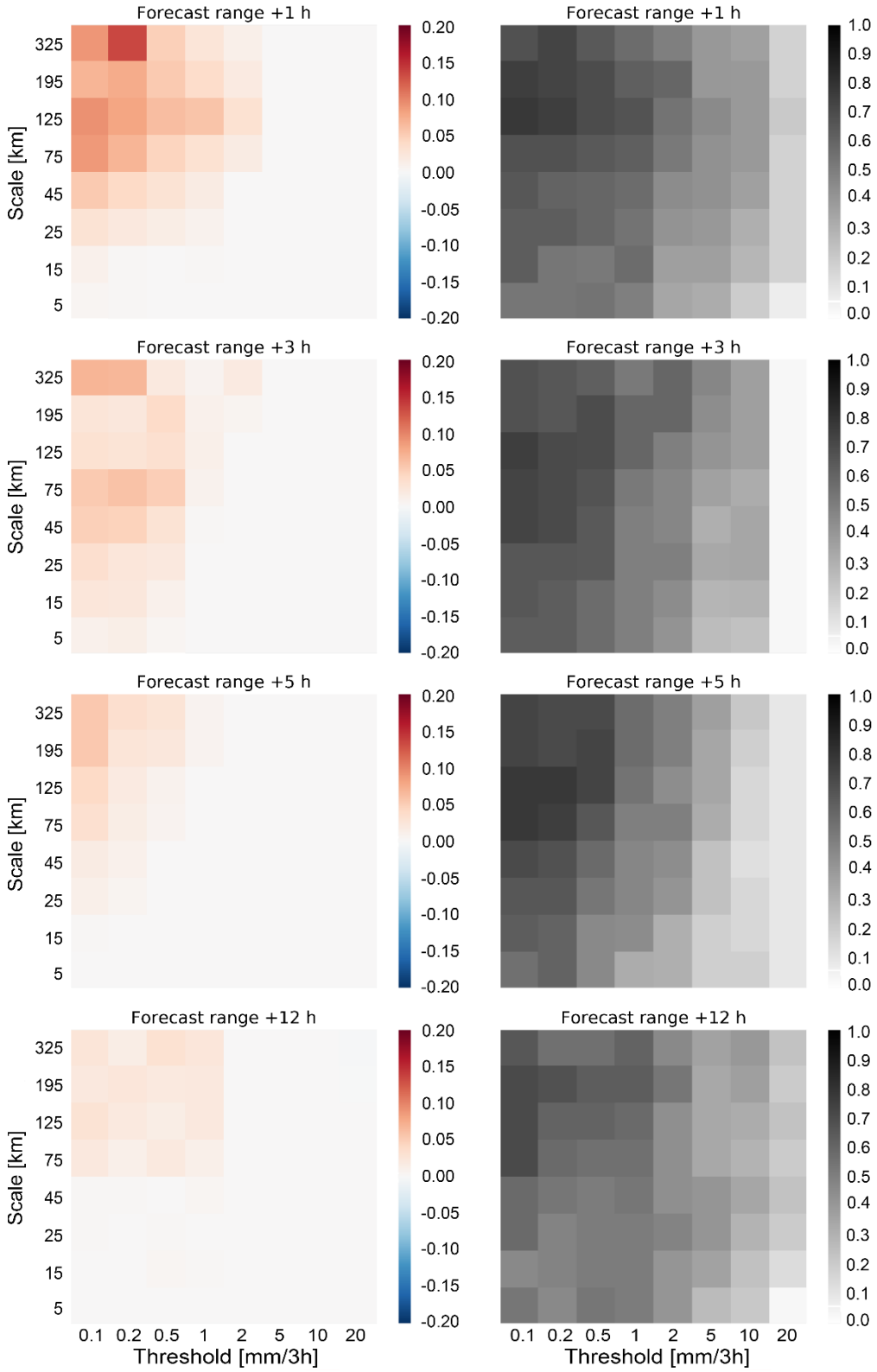

Fig. 44. Similar to Fig. 18, the left panel is the median Skill Score of FSS of LAG to FSS of REF (red means LAG is better than REF and blue the opposite) as a function of scale and threshold. The right panel is the significance level for the comparison (percentage of times FSS of LAG is higher than the FSS of REF).

8.2.4. Summary and discussion

It is interesting to ask wheatear the observed differences between the experiments are only due to the larger ensemble size of LAG? To answer this question, verification was performed again, but this time 4 C-LAEF members were removed in LAG, so that both experiments have 17 members. All of the results and conclusions remained similar (not shown) meaning that pure ensemble size does not have a significant impact. However, we saw that most of the improvements came from increased ensemble spread, how can we reconcile this?

Further inspection showed that AROME members tend to cluster together which significantly benefits the ensemble spread because it increases the difference between ensemble members. This means that there were some accidental differences in the AROME vs C-LAEF model configuration that needs investigation. Nevertheless, having more different members is known to benefit the ensemble performance (e.g., Johnson and Swinbank, 2009). Even if some members are consistently worse than others, the effect of error cancelation and filtering of less predictable scales when doing averages can still improve the accuracy of the ensemble mean as shown in Hagedorn, *et al.* (2005).

The goal of this sub-section was to assess the impact of adding lagged deterministic model forecast (AROME) to convection-permitting LAMEPS (C-LAEF). In other words, to study the added value of past forecasts to extend the C-LAEF IC perturbations with a new flow-dependant information. Practical benefits of this configuration were clearly visible in improved ensemble reliability, spread, and slightly higher accuracy for W10M and WG10M forecasts. Detection of extreme wind gust speeds is improved and the impact on precipitation is also positive with the highest impact in the first 5 hours of integration.

Lastly, we will reflect on the observation error impact on the problem of under-dispersion. As can be seen from almost all surface verification plots presented in this dissertation, all experiments are heavily under-dispersive, which is a known problem in the field of ensemble forecasting and it is regularly the case in LAMEPS studies (Keresturi *et al.*, 2019). However, most of the studies, including this one, neglect observation errors. This should be avoided as it makes our ensembles look worse than they actually are (see Bowler *et al.*, 2015). As an example, Fig. 45 shows W10M RMSE and spread relation from Fig. 40 but with observation error accounted for by the method recommended by Bowler *et al.* (2015). Now, under-dispersion is significantly reduced, and this is plotted for a rather small value of W10M observation error – 1.2 ms^{-1}.

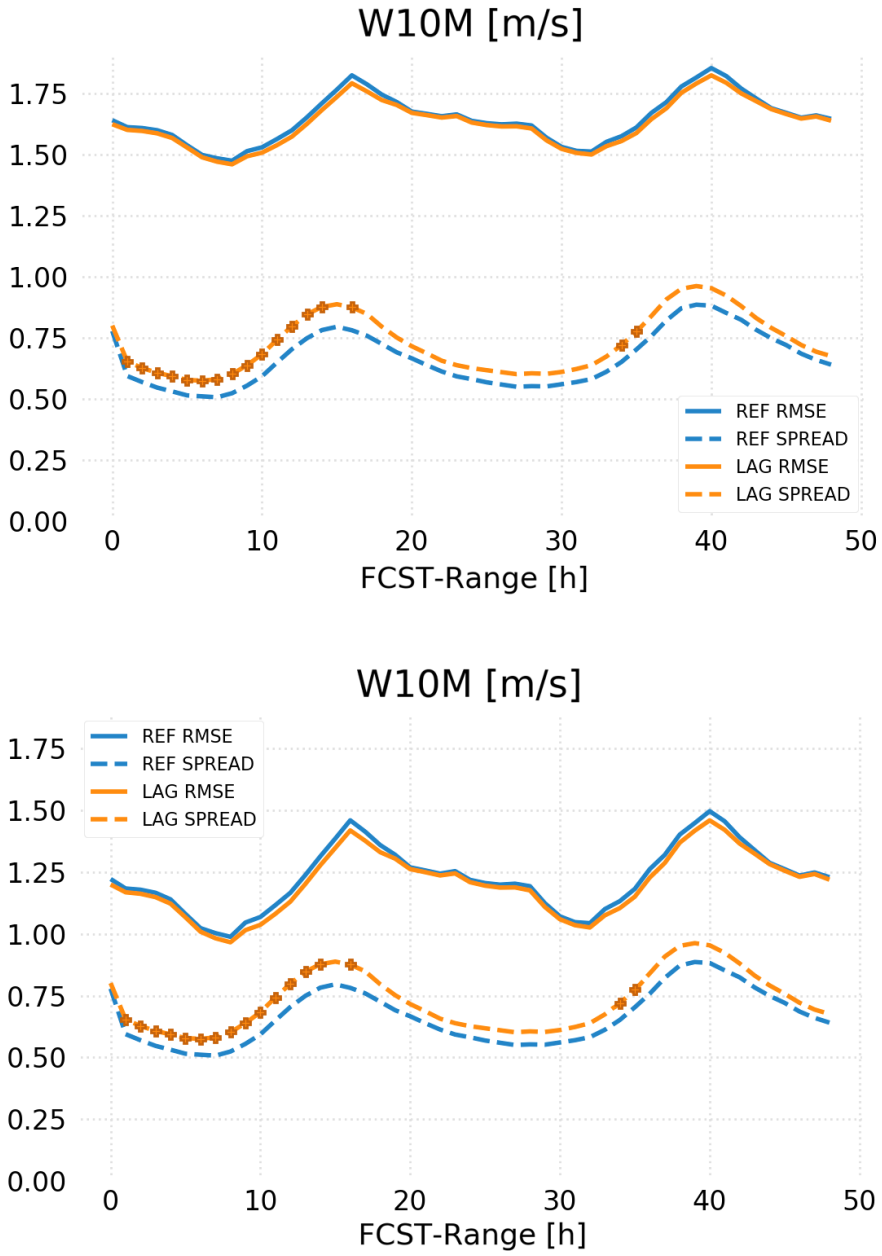

Fig. 45. Up (down) RMSE of the ensemble mean and ensemble spread for W10M without (with) observation error accounted for.

§ 9. CONCLUSION

Probabilistic approaches to numerical weather prediction are widely recognized as superior to deterministic and ensemble prediction systems are becoming more widely favored forecast tools. This is especially true for meso and convective scales where rapid loss of predictability, i.e., very fast error growth, after which the structures on those scales can be regarded as noise, may be mitigated by utilizing ensemble forecasts. For these reasons, probabilistic approach through LAMEPSs have been used more frequently over the last 10 years and motivation to resolve problems related to LAMEPSs is strong.

One of the most principal questions with LAMEPSs is how to correctly account for uncertainties of large-scale flow. Those uncertainties arise due to the typically small computational domains, deficiencies of LBC formulations, inferior assimilation procedures, focus on the mesoscale and unavailability of observations outside the domain. Furthermore, mismatches between LAMEPS's initial condition and host EPS lateral boundary perturbations can result when they are obtained independently. As a result, spurious waves are generated at the domain boundary which then spread through the rest of the domain inducing unwanted noise to the meteorological fields what considerably deteriorates forecast skill.

In this dissertation, we introduce an ensemble *Jk* method as a new idea for improving LAMEPS forecasts, whereby host model large-scale perturbations are introduced into LAMEPS IC perturbations using the *Jk* blending method. *Jk* blending is applied to the 3D-Var EDA system so that perturbed analyses are optimal and small scales are consistent with large scales, as opposed to using the blending method developed by Wang *et al.* (2014). As a result, analyses contain perturbed small scales (LAMEPS EDA) and large scales (host EPS). Furthermore, when including host model perturbations in LAMEPS initial perturbations, final perturbations are more consistent with perturbations coming from lateral boundaries, thus minimizing the risk of potential mismatch.

The ensemble *Jk* method is implemented into convection-permitting ensemble system C-LAEF with ECMWF-EPS providing LBCs and large-scale perturbations. To assess the effect of the large-scale constraint on IC perturbations, two experiments are conducted, one using an ensemble *Jk* and the other using standard perturbed-observation EDA, i.e., EDA without *Jk* blending. Impacts on average model performance are measured over 62

convectively active days of the summer of 2016. The results show that using the ensemble *Jk* method provides a more accurate and reliable EPS. This is mostly the case due to a reduction of RMSE, thus due to increased accuracy. RMSE of ensemble mean is strongly reduced for upper-air variables, while spread is maintained on similar or slightly lower values. The lower percentage of outliers supports these conclusions as does the CRPS, which is reduced for all variables. Furthermore, ROC score shows that discrimination is improved, while decomposition of the Brier score demonstrates the improvement in resolution. In other words, benefits of using the ensemble *Jk* method are visible in all the attributes of a probabilistic forecast tested here.

Performance in terms of surface variables is more neutral, as was expected due to its high sensitivity to small scales and the forcing effects of lower boundary conditions, both of which are unaffected by the ensemble *Jk* method. Nevertheless, slight but insignificant improvements resulting from using the ensemble *Jk* method are observed for early forecast ranges (up to 9 h) for temperature and humidity, and for surface pressure, improvements are visible up to 18 h. The strongest effect is observed for mean sea level pressure forecasts, which are now more reliable and accurate. Discrimination is slightly improved for extreme thresholds for MSLP and T2M indicating that detection of extreme events is also improved. Precipitation forecasts are verified by a different methodology and by using the FSS. Improvements for precipitation thresholds up to 20 mm and for scales ranging from the highest (325 km) down to the smallest (5 km) appear when using the ensemble *Jk* method.

To further illustrate benefits of considering large-scale perturbations from the host model, two case-studies of strong synoptic forcing observed near the borders of the computational domain are presented. It is demonstrated that using the ensemble *Jk* method improves the surface pressure distribution in areas adjacent to incoming front, in turn minimizing the possibility of the false detection of precipitation and improving precipitation forecast during a period of convective activity over Austria and western Germany.

To alleviate the problem of mismatched perturbations, the impact of ensemble *Jk* method is analyzed in the form of two case studies of strong perturbation mismatch. Both events are characterized by upper-level lows intersecting the C-LAEF domain. Strong spatial anomalies in surface pressure spread are observed close to the northeastern and southern domain boundaries after only 5 min of model integration. Those are most likely the result of spurious Lamb waves. After 1 h, excessive spread is still present close to the edges of the domain. Introducing LBC perturbations gradually over the first hour of integration, i.e., space

consistent coupling, did not remedy this problem. However, when ensemble *Jk* method is used and when large-scale IC perturbations were made more consistent with LBC perturbations, even only for the largest scales, excessive spread is mitigated after 1 h. From these findings, it is clear that future implementations of C-LAEF with hourly assimilation cycles, must address the problem of perturbation mismatch, which is the leading cause of excessive surface pressure spread observed here. The ensemble *Jk* method can be used to alleviate such problems.

Overall, using the ensemble *Jk* method benefits the ensemble performance of the upper-air variables, precipitation and mean sea level pressure, while effects on the other surface variables are less significant and mostly reserved for special cases, such as those discussed in section 6. Observed impacts are strongest within the early forecast ranges (0 - 12 h), which is to be expected from an IC perturbation study (e.g., Vié *et al.*, 2011). Both of our experiments are under-dispersive, particularly for lower model levels. This is mostly the case due to an absence of model and surface perturbations and to a lack of consideration of observation errors which was demonstrated in sub-section 8.2.4.

To address under-dispersion and additionally improve C-LAEF's IC perturbation sampling of the initial uncertainties, LAF approach is tested. Coherent operational setup of C-LAEF and operational deterministic convection-permitting model AROME at ZAMG, enabled us a unique possibility to test expansion of C-LAEF members with lagged AROME forecasts. All members are equally weighted. Practical benefits of this configuration are clearly visible in improved ensemble reliability, spread, and slightly higher accuracy for W10M and WG10M forecasts with neutral impact on solar radiation forecasts. Detection of extreme wind gust speeds is also improved. Impact on precipitation is positive with the highest impact in the first 5 hours of integration.

Finally, a well-known fact in NWP is that model grid size is not synonymous with the concept of model resolution. The latter is at least 5 times lower. To account for this fact, the problem of rapid error growth at the small scales and to alleviate under-dispersion, a neighborhood approach can be used. Here, we combine the LAMEPS with the neighborhood approach. C-LAEF forecasts for a specific point were extended with the forecasts for its neighbouring points by assuming they are equally likely to occur at that specific point. The result of such an approach is shifting the focus away from the specific location towards a somewhat broader region which can alleviate the beforementioned problems. Different configurations of neighborhood were tested. The best approach is to make the neighborhood a

function of a location and a variable. This is advisable because the orography and land-sea boundary can have significant impacts on the forecast. The results indicate that accuracy, reliability and resolution are increased for T2M, RH2M and W10M. Impact is the smallest on MSLP, because MSLP is mostly large-scale variable and does not vary much within the neighborhood. In addition, C-LAEFs ability to detect extreme events is also improved.

The main scientific contributions of this dissertation are as follows. First, an innovative method, named ensemble *Jk* method, for generation of IC perturbations for LAMEPSs was developed. Second, as a part of the development process of ensemble *Jk* method, the *Jk* blending method was implemented into AROME model for the first time. Third, variations of LAF and neighborhood approaches were applied and tested as a relatively cheap techniques to improve the LAMEPSs performance. Lastly, this work was a part of the development process of a new convection-permitting LAMEPS named C-LAEF and ensemble *Jk* method is currently used in its operational configuration (Wastl *et al*., 2021).

§ 10. BIBLIOGRPAHY

1. Aladin IT., 1997: The ALADIN project mesoscale modelling seen as basic tool for weather forecasting and atmospheric research. *WMO Bull*., **46**, 317–324.
2. Arakawa, A. and Lamb, V. R., 1977: Computational design of the basic dynamical processes of the UCLA general circulation model. *Methods in Computational Physics: Advances in Research and Applications*, **17**, 173–265.
3. Bannister, R. N., 2017: A review of operational methods of variational and ensemble-variational data assimilation. *Q. J. R. Meteorol. Soc.*, **143**, 607-633.
4. Bauer, P., Thorpe, A. and Brunet, G., 2015: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55.
5. Belamari, S. and Pirani, A., 2007: Validation of the optimal heat and momentum fluxes using the ORCA2-LIM global ocean-ice model. Marine environment and security for the European area. Integrated Project (MERSEA IP), Deliverable D4.1.3, 88 pp.
6. Ben Bouallegue, Z., Theis, S. E. and Gebhardt, C., 2013: Enhancing COSMO-DE ensemble forecasts by inexpensive techniques. *Meteorol. Zeitschrift*, **22**, 49–59.
7. Ben Bouallègue, Z. and Theis, S. E., 2014: Spatial techniques applied to precipitation ensemble forecasts: From verification results to probabilistic products. *Meteor. Appl*., **21**, 922–929.
8. Ben Bouallegue, Z., Haiden, T., Weber, N. J., Hamill, T. M. and Richardson, D. S., 2020: Accounting for representativeness in the verification of ensemble precipitation forecasts. *Mon. Wea. Rev.*, **148**, 2049–2062.
9. Bengtsson, L. K., Magnusson, L. and Källén, E., 2008: Independent estimations of the asymptotic variability in an ensemble forecast system. *Mon. Wea. Rev.*, **136**, 4105–4112.
10. 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.
11. Berner, J., Ha, S.-Y., Hacker, J., Fournier, A. and Snyder, C., 2011: Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Mon. Wea. Rev.*, **139**, 1972–1995.

12. Berre, L., 2000: Estimation of synoptic and mesoscale forecast error covariances in a limited area model. *Mon. Wea. Rev.*, **128**, 644– 667.
13. Berre, L., Ştefănescu, S. E. and Belo-Pereira, M., 2006: The representation of the analysis effect in three error simulation techniques. *Tellus A*, **58**, 196–209.
14. Bougeault, P. and Lacarrere, P., 1989: Parameterization of orography induced turbulence in a meso-beta-scale model. *Mon. Wea. Rev.*, **117**, 1870–1888.
15. Bouttier, F., Vié, B., Nuissier, O. and Raynaud, L., 2012: Impact of stochastic physics in a convection-permitting ensemble. *Mon. Wea. Rev.*, **140**, 3706–3721.
16. Bouttier, F., Raynaud, L., Nuissier, O. and Ménétrier, B., 2015: Sensitivities of the AROME ensemble to initial and surface perturbations during HyMeX. *Q. J. R. Meteorol. Soc.,* **142 (Suppl 1)**, 390–403.
17. Bowler, N. E., 2006: Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model. *Tellus A*, **58:5**, 538-548.
18. Bowler, N. E., Arribas, A. and Kenneth, R. M., 2007: The benefits of multi-analysis and poor-man's ensembles. *Mon. Wea. Rev.*, **136**, 4113–4129.
19. Bowler, N. E., 2008: Accounting for the effect of observation errors on verification of MOGREPS. *Meteorol. Appl.*, **15**, 199–205.
20. Bowler, N. E., Arribas, A. and Mylne, K. R., 2008a: The Benefits of Multianalysis and Poor Man's Ensembles. *Mon. Wea. Rev.*, **136**, 4113–4129.
21. Bowler, N. E., Arribas, A., Mylne, K. R., Robertson, K. B. and Beare, S. E., 2008b: The MOGREPS short-range ensemble prediction system. *Q. J. R. Meteorol. Soc.*, **134**, 703–722.
22. Bowler, N. E. and Mylne, K. R., 2009: Ensemble transform Kalman filter perturbation for a regional ensemble prediction system. *Q. J. R. Meteorol. Soc.*, **135**, 757–766.
23. Bowler, N. E., Cullen, M. J. P. and Piccolo, C., 2015: Verification against perturbed analyses and observations. *Nonlin. Processes Geophys.*, **22**, 403–411.
24. Bölöni, G., 2011: Background error estimation in an atmospheric limited area model. Ph.D. dissertation, pp 119.
25. Branković, Č., Palmer, T. N., Molteni, F., Tibaldi, S. and Cubasch, U., 1990: Extended-range predictions with ECMWF models: Time-lagged ensemble forecasting. *Q. J. R. Meteorol. Soc.,* **116**, 867–912.

26. Brousseau, P., Berre, L., Bouttier, F. and Desroziers, G., 2011: Background-error covariances for a convective-scale data-assimilation system: AROME–France 3DVar. *Q. J. R. Meteorol. Soc.*, **137**, 402–422.
27. Brousseau, P., Seity, Y., Ricard, D. and Léger, J., 2016: Improvement of the forecast of convective activity from the AROME-France system. *Q. J. R. Meteorol. Soc*., **142**, 2231–2243.
28. Brožkovà, R., Klarić, D., Ivatek-Šahdan, S., Geleyn JF, Casse, V., Siroka, M., Radnoti, G., Janoušek, M., Stadlbacher, K. and Seidl H., 2001: DFI blending: An alternative tool for preparation of the initial conditions for LAM. Research Activities in Atmospheric and Oceanic Modelling, H. Ritchie, Ed., WMO CAS/JSC WGNE Rep. 31.
29. Bubnova, R., Hello, G., Benard, P. and Geleyn, J.-F., 1995: Integration of the fully elastic equations cast in the hydrostatic pressure terrain-following in the framework of the ARPEGE/ ALADIN NWP system. *Mon. Wea. Rev.*, **123**, 515–535.
30. Buizza, R. and Palmer, T. N., 1995: The singular vector structure of the atmospheric general circulation. *J. Atmos. Sci*., **52**, 1434–1456.
31. Buizza, R., Houtekamer, P. L., Toth, Z., Pellerin, G., Wei, M. and Zhu, Y., 2005: A comparison of the ECMWF, MSC and NCEP global ensemble prediction systems. *Mon. Wea. Rev.*, **133**, 1067–1097.
32. Buizza, R., Leutbecher, M. and Isaksen, L., 2008: Potential use of an ensemble of analyses in the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc*., **134**, 2051–2066.
33. Buizza, R., 2014: The TIGGE global medium-range ensembles. *ECMWF Tech. Mem.* 739.
34. Buizza, R. and Leutbecher, M., 2015: The forecast skill horizon. *Q. J. R. Meteorol. Soc.*, **141**, 3366-3382.
35. Buizza, R., 2019: Introduction to the special issue on "25 years of ensemble forecasting". *Q. J. R. Meteorol. Soc.*, **145 (Suppl. 1)**, 1-11.
36. Burgers, G., VanLeeuwen, P. and Evensen, G. 1998. Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.
37. Candille, G. and Talagrand, O., 2008: Impact of observational error on the validation of ensemble prediction systems. *Q. J. R. Meteorol. Soc.*, **134**, 959-971.
38. Cardinali, C., Žagar, N., Radnóti, G. and Buizza, R., 2014: Representing model error in ensemble data assimilation. *Nonlinear Processes in Geophysics*, **21**, 971-985.

39. Caron, J. F., 2013: Mismatching perturbations at the lateral boundaries in limited area ensemble forecasting: A case study. *Mon. Wea. Rev.*, **141**, 356–374.
40. Charney, J. G., 1951: Dynamical forecasting by numerical process. Compendium of meteorology. American Meteorological Society, Boston, MA.
41. Charnock, H., 1955: Wind stress over a water surface. *Q. J. R. Meteorol. Soc.*, **81**, 639–640.
42. Clark, A. J., Gallus, W. A., Xue, M. and Kong, F., 2009: A comparison of precipitation forecast skill between small convection-allowing and large convection parameterizing ensembles. *Weather Forecast.*, **24**, 1121–1140.
43. Coiffier, J., 2011: Fundamentals of Numerical Weather Prediction. Cambridge University Press, pp 340.
44. Cuxart, J., P. Bougeault, and J.-L. Redelsberger, 2000: A turbulence scheme allowing for mesoscale and large-eddy simulations. *Q. J. R. Meteorol. Soc.*, **126**, 1–30.
45. Dahlgren, P. and Gustafsson, N., 2012: Assimilating host model information into a limited area model. *Tellus A*, **64**, 15836.
46. Davies, H. C., 1976: A lateral boundary formulation for multi-level prediction models. *Q. J. R. Meteorol. Soc.*, **102**, 405–418.
47. Davies, T., 2014: Lateral boundary conditions for limited area models. *Q. J. R. Meteorol. Soc.*, **140**, 185–196.
48. Denis, B., J. Côté, and R. Laprise, 2002: spectral decomposition of two-dimensional atmospheric fields on limited-area domains using the discrete cosine transform (DCT). *Mon. Wea. Rev.*, **130**, 1812–1829.
49. Descamps, L. and Talagrand, O., 2007: On some aspects of the definition of initial conditions for ensemble prediction. *Mon. Wea. Rev.*, **135**, 3260–3272.
50. Desroziers, G., Berre, L., Chapnik, B. and Poli, P., 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.*, **131**, 3385–3396.
51. Douville, H., Royer, J. F. and Mahfouf, J. F., 1995: A new snow parameterization for the Meteo-France climate model. *Climate Dynamics 12*, **1**, 21–35.
52. Du, J., G., DiMego, M., Tracton, S. and Zhou, B., 2003: NCEP short-range ensemble forecasting (SREF) system: multi-IC, multi-model and multi-physics approach. Research Activities in Atmospheric and Oceanic Modelling. Report 33, CAS/JSC Working Group Numerical Experimentation (WGNE), WMO/TD-No. 1161, 5.09-5.10.

53. Du, J., 2007: Uncertainty and ensemble forecast. National Weather Service, Office of Science & Technology, Science & Technology Infusion Lecture Series, 42 pp.
54. Duc, L., Saito, K. and Seko, H., 2013: Spatial-temporal fractions verification for high-resolution ensemble forecasts. *Tellus A*, **65**, 1–22.
55. Duc, L. and Saito, K., 2018: Verification in the presence of observation errors: Bayesian point of view. *Q. J. R. Meteorol. Soc*., **144**, 1063–1090.
56. Ebert, E., 2008: Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteorol. Appl.,* **15**, 51–64.
57. ECMWF, 2019: IFS documentation. CY46R1, Parts III–VI, ECMWF. [Available online at https://www.ecmwf.int/en/publications/ifs-documentation.]
58. Ehrendorfer, M., 1994a: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I: Theory. *Mon. Wea. Rev.*, **122**, 703–713.
59. Ehrendorfer, M., 1994b: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part II: Applications. *Mon. Wea. Rev*., **122**, 714–728.
60. Ehrendorfer, M., 1997: Predicting the uncertainty of numerical weather forecasts: a review. *Meteorol. Zeitschrift*, **6**, 147–183.
61. Epstein, E. S., 1969: Stochastic dynamic predictions. *Tellus*, **21**, 739–759
62. Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367.
63. Feng, J., Sun, J. and Zhang, Y., 2020: A dynamic blending scheme to mitigate large-scale bias in regional models. *Journal of Advances in Modeling Earth Systems*, **12**, e2019MS001754.
64. Fischer, C., Montmerle, T., Berre, L., Auger, L. and Ştefănescu, S. E., 2005: An overview of the variational assimilation in the Aladin/France NWP system. *Q. J. R. Meteor. Soc.*, **131**, 3477–3492.
65. Flora, M. L., Potvin, C. K., and Wicker, L. J., 2018: Practical predictability of supercells: exploring ensemble forecast sensitivity to initial condition spread. *Mon. Wea. Rev.*, **146**, 2361-2379.
66. Fortin, V., Abaza, M., Anctil, F. and Turcotte, R., 2014: Why should ensemble spread match the RMSE of the ensemble mean? *J. Hydrometeor.*, **15**, 1708–1713.
67. Fouquart, Y. and Bonnel, B., 1980: Computations of solar heating of the earth's atmosphere: A new parameterization. *Beitr. Phys. Atmos.*, **53**, 35–62.

68. Garcia-Moya, J.-A., Callado, A., Escribà, P., Santos, C., Santos-Muñoz, D. and Simarro, J., 2011: Predictability of short-range forecasting: a multimodel approach. *Tellus A*, **63(3)**, 550–563.
69. Gilleland, E., Ahiyevych, D., Brown, B. G., Casati, B. and Ebert, E., 2009: Intercomparison of spatial forecast verification methods. *Wea. Forecasting*, **24**, 1416-1430.
70. Grasso, L. D., 2000: The differentiation between grid spacing and resolution and their application to numerical modeling. *Bull. Amer. Meteor. Soc*., **81**, 579–580.
71. Göber, M., Zsótér, E. and Richardson, D. S., 2008: Could a perfect model ever satisfy a naïve forecaster? On grid box mean versus point verification. *Meteor. Appl.*, **15**, 359–365.
72. Guidard, V. and Fischer, C., 2008: Introducing the coupling information in a limited area variational assimilation. *Q. J. R. Meteorol. Soc.*, **134**, 723–736.
73. Gustafsson, N., Berre, L., Hörnquist, S., Huang, X.-Y., Lindskog, M., Navascués, B., Mogensen, K. S. and Thorsteinsson, S., 2001: Three-dimensional variational data assimilation for a limited area model. *Tellus A*, **53**, 425–446.
74. Gustafsson, N. and coauthors, 2018: Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Q. J. R. Meteorol. Soc*., **144**, 1218– 1256.
75. Hacker, J. P., Ha, S.-Y., Snyder, C., Berner, J., Eckel, F. A., Kuchera, E., Pocernich, M., Rugg, S., Schramm, J. and Wang, X., 2011: The U.S. Air Force Weather Agency's mesoscale ensemble: scientific description and performance results, *Tellus A*, **63**, 625–641.
76. Hagedorn, R., Doblas-Rexes, F. J. and Palmer, T. N., 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting – I. Basic concept. *Tellus A*, **57**, 219–233.
77. Hagedorn, R., Buizza, R., Hamill, T. M., Leutbecher, M. and Palmer, T. N., 2012: Comparing TIGGE multimodel forecasts with reforecast-calibrated ECMWF ensemble forecasts. *Q. J. R. Meteorol. Soc.*, **138**, 1814–1827.
78. Hagelin, S., Swinbank, J. Son, R., McCabe, A., Roberts, N. and Tennant, W., 2017: The MetOffice convective-scale ensemble, MOGREPS-UK. *Q. J. R. Meteorol. Soc*., **143**, 2846–2861.

79. Haiden, T., Kann, A., Wittmann, C., Pistotnik, G., Bica, B. and Gruber, C., 2011: The Integrated Nowcasting through Comprehensive Analysis (INCA) System and Its Validation over the Eastern Alpine Region. *Wea. Forecasting*, **26**, 166–183.
80. Haiden, T., Rodwell, M. J., Richardson, D. S., Okagaki, A., Robinson, T. and Hewson, T., 2012: Intercomparison of global model precipitation forecast skill in 2010/11 using the SEEPS score. *Mon. Wea. Rev.*, **140**, 2720–2733.
81. Hall, P., Horowitz, J. L. and Jing, B.-Y., 1995: On blocking rules for the bootstrap with dependent data. *Biometrika*, **82**, 561–574.
82. Harnisch, F. and Keil, C., 2015: Initial conditions for convective-scale ensemble forecasting provided by ensemble data assimilation. *Mon. Wea. Rev.*, **143**, 1583–1600.
83. Haugen, J. and Machenhauer, B., 1993: A spectral limited-area model formulation with time-dependent boundary conditions applied to shallow-water equations. *Mon. Wea. Rev.*, **121**, 2618–2630.
84. Hoffman, R. N. and Kalnay, E., 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus A*, **35**, 100-118.
85. Hohenegger, C. and Schär, C., 2007: Atmospheric predictability at synoptic versus cloud-resolving scales. *Bull. Amer. Meteor. Soc.*, **88**, 1783–1793.
86. Holton, J. R., 2004: An Introduction to Dynamic Meteorology. Elsevier Academic Press, pp 553.
87. Hong, S. and Kanamitsu, M., 2014: Dynamical downscaling: Fundamental issues from an NWP point of view and recommendations. *Asia-Pacific J. Atmos. Sci.*, **50**, 83–104.
88. Hopson, T. M., 2014: Assessing the ensemble spread–error relationship. *Mon. Wea. Rev.*, **142**, 1125–1142.
89. Horvath, K., Bajić, A. and Ivatek-Šahdan, S., 2011: Dynamical downscaling of wind speed in complex terrain prone to bora-type flows. *JAMC*, **50**, 1676–1691.
90. Horvath, K., Koračin, D., Vellore, R. K., Jiang, J. and Belu, R., 2012: Sub-kilometer dynamical downscaling of near-surface winds in complex terrain using WRF and MM5 mesoscale models. *J. Geophys. Res.*, **117**, D11111, 19pp.
91. Hoskins, B. J., McIntyre, M. E. and Robertson, A. W., 1985: On the use and significance of isentropic potential vorticity maps. *Q. J. R. Meteorol. Soc*., **111**, 877–946.
92. Houtekamer, P. L., Lefaivrem, L., Derome, J., Ritchie, H. and Mitchell, H. L., 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225–1242.

93. Houtekamer, P. L. and Mitchell, H. L., 2005: Ensemble Kalman filtering. *Q. J. R. Meteorol. Soc.*, **131**, 3269–3289.
94. Hsiao, L. F., Huang, X. Y., Kuo, Y. H., Chen, D. S., Wang, H., Tsai, C. C., Yeh, T. C., Hong, J. S., Fong, C. T. and Lee, C. S., 2015: Blending of global and regional analyses with a spatial filter: Application to typhoon prediction over the western North Pacific Ocean. *Wea. Forecasting*, **30**, 754–770.
95. Hu, M., Benjamin, S. G., Ladwig, T. T., Dowell, D. C., Weygandt, S. S., Alexander, C. R. and Whitaker, J. S., 2017: GSI three-dimensional ensemble-variational hybrid data assimilation using a global ensemble for the regional Rapid Refresh model. *Mon. Wea. Rev.*, **145**, 4205–4225.
96. Isaksen, L., Bonavita, M., Buizza, R., Fisher, M., Haseler, J., Leutbecher, M. and Raynaud, L., 2010: Ensemble of Data Assimilations at ECMWF. *ECMWF Tech. Mem. 636.*
97. Iversen, T., Deckmyn, A., Santos, C., Sattler, K., BjoRnar Bremnes, J., Feddersen, H and Frogner, I-L., 2011: Evaluation of 'GLAMEPS'—a proposed multimodel EPS for short range forecasting. *Tellus A*, **63**, 513–530.
98. Jankov, I., Gallus, W. A., Segal, M., Shaw, B. and Koch, S. E., 2005: The Impact of Different WRF Model Physical Parameterizations and Their Interactions on Warm Season MCS Rainfall. *Wea. Forecasting*, **20**, 1048–1060.
99. Janjić, T. and Coauthors, 2018: On the representation error in data assimilation. *Q. J. R. Meteorol. Soc.*, **144**, 1257–1278.
100. Johnson, C. and Swinbank, R., 2009: Medium-range multimodel ensemble combination and calibration. *Q. J. R. Meteorol. Soc.*, **135**, 777–794.
101. Johnson, A. and Wang, X., 2016: A Study of multiscale initial condition perturbation methods for convection-permitting ensemble forecasts. *Mon. Wea. Rev.*, **144**, 2579–2604.
102. Judt, F., 2018: Insights into atmospheric predictability through global convection-permitting model simulations. *J. Atmos. Sci.*, **75**, 1477–149.
103. Kalnay, E., 2003: Atmospheric modelling, data assimilation and predictability. Cambridge University Press, 341 pp.
104. Kalnay, E., 2019: Historical perspective: earlier ensembles and forecasting forecast skill. *Q. J. R. Meteorol. Soc.*, **145**, 25–34.

105. Keresturi, E., Wang, Y., Meier, F., Weidle, F. and Wittman, C., 2019: Improving initial condition perturbations in a convection-permitting ensemble prediction system. *Q. J. R. Meteorol. Soc.*, **145**, 993–1012.
106. Kretschmer, M., Hunt, B. R., Ott, E., Bishop, C. H., Rainwater, S. and Szunyogh, I., 2015: A composite state method for ensemble data assimilation with multiple limited-area models. *Tellus A*, **67**, 1–17.
107. Kühnlein, C., Keil, C., Craig, G. C. and Gebhardt, C., 2014: The impact of downscaled initial condition perturbations on convective-scale ensemble forecasts of precipitation. *Q. J. R. Meteorol. Soc.*, **140**, 1552–1562.
108. Lafore, J.-P., Stein, J., Asencio, N., Bougeault, P., Ducrocq, V., Duron, V., Fischer, C., Héreil, P., Mascart, P., Masson, V., Pinty, J. P., Redelsperger, J. L., Richard, E., Vilà-Guerau de Arellano, J., 1998: The Meso-NH atmospheric simulation system. Part I: Adiabatic formulation and control simulations. *Ann. Geophys.*, **16**, 90–109.
109. Lang, S. T. K., Bonavita, M. and Leutbecher, M., 2015: On the impact of re-centring initial conditions for ensemble forecasts. *Q. J. R. Meteorol. Soc.*, **141**, 2571–2581.
110. Le Dimet, F-X. and Talagrand, O., 1986: Variational algorithms for analysis and assimilation of meteorological observations. Theoretical aspects. *Tellus A*, **38**, 97–110.
111. Le Moigne, P. and Coauthors, 2018: SURFEX scientific documentation. Available at: https://www.umr-cnrm.fr/surfex/IMG/pdf/surfex_scidoc_v8.1.pdf
112. Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev*., **102**, 409–418.
113. Leutbecher, M. and Palmer, T. N., 2008: Ensemble forecasting. *J. Comp. Phy.*, **227**, 3515–3539.
114. Leutbecher, M., 2018: Ensemble size: How suboptimal is less than infinity? *Q. J. R. Meteorol. Soc.*, **145**, 107–128.
115. Lewis, J. M., 2005: Roots of ensemble forecasting. *Mon. Wea. Rev.*, **133**, 1865–1885.
116. Lorenz, E. N., 1963a: The predictability of hydrodynamic flow. *Trans. N.Y. Acad. Sci.*, **Ser. II, 25, 4**, 409-432.
117. Lorenz, E. N., 1963b: Deterministic nonperiodic flow. *J. Atmos Sci.*, **20**, 130–141.
118. Lorenz, E. N., 1965: A study of the predictability of a 28-variable atmospheric model. *Tellus*, **17**, 321–333.
119. Lorenz, E. N., 1969: How much better can weather prediction become? *Technology Rev*., July/August, 39–49.

120. Ma, J., Zhu, Y., Wobus, R. and Wang, P., 2012: An effective configuration of ensemble size and horizontal resolution for the NCEP GEFS. *Adv. Atmos. Sci.,* **29,** 782–794.
121. Magnusson, L., Nycander, J. and Källén, E., 2009: Flow-dependent versus flow-independent initial perturbations for ensemble prediction. *Tellus A*, **61**, 194–209.
122. Mass, C. F., Ovens, D., Westrick, K. and Colle, B. A., 2002: Does increasing horizontal resolution produce more skillful forecasts? *Bull. Amer. Meteor. Soc*., **83**, 407–430.
123. Mlawer, E. J., Taubman, S. J., Brown, P., Iacono, M. J. and Clough, S. A., 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682.
124. Mittermaier, M. P., 2007: Improving short-range high-resolution model precipitation forecast skill using time-lagged ensembles. *Q. J. R. Meteorol. Soc.*, **133**, 1487–1500.
125. Mittermaier, M. P., 2014: A Strategy for verifying near-convection-resolving model forecasts at observing sites. *Wea. Forecasting*, **29**, 185–204.
126. Mittermaier, M. P. and Stephenson, D. B., 2015: Inherent bounds on forecast accuracy due to observation uncertainty caused by temporal sampling. *Mon. Wea. Rev.*, **143**, 4236–4243.
127. Montmerle, T., 2018: Statement of guidance for high-resolution numerical weather prediction. WMO Statement of Guidance, World Meteorological Organization, 10 pp.
128. Molteni, F., Buizza, R., Palmer, T. N. and Petroliagis, T., 1996: The ECMWF Ensemble Prediction System: Methodology and validation. *Q. J. R. Meteorol. Soc.*, **122**, 73–119.
129. Morcrette, J.-J. and Fouquart, Y., 1986: The overlapping of cloud layers in shortwave radiation parameterizations. *J. Atmos. Sci.*, **43**, 321–328.
130. Murphy, A. H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281–293.
131. Müller, M. and Coauthors, 2017: AROME-MetCoOp: A Nordic convective-scale operational weather prediction model. *Wea. Forecasting*, **32**, 609–627.
132. Noilhan, J. and Planton, S., 1989: Simple parameterization of land surface processes for meteorological models. *Mon. Wea. Rev.*, **117**, 536–549.
133. Nuissier, O., Joly, B., Vié, B. and Ducrocq, V., 2012: Uncertainty on lateral boundary conditions in a convection-permitting ensemble: A strategy of selection for Mediterranean heavy precipitation events. *Nat. Hazards Earth Syst. Sci.* **12**, 2993–3011.

134. Nutter, P., Stensrud, D. and Xue, M., 2004a: Effects of coarsely resolved and temporally interpolated lateral boundary conditions on the dispersion of limited-area ensemble forecasts. *Mon. Wea. Rev.*, **132**, 2358–2377.
135. Nutter, P., Xue, M. and Stensrud, D., 2004b: Application of lateral boundary condition perturbations to help restore dispersion in limited-area ensemble forecasts. *Mon. Wea. Rev.* **132**, 2378–2390.
136. Palmer, T. and Hagedorn, R., 2006: Predictability of Weather and Climate. Cambridge University Press. 124–156.
137. Palmer, T. N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G. J., Steinheimer, M. and Weisheimer, A., 2009: Stochastic parametrization and model uncertainty. *ECMWF Tech. Mem. 598*.
138. Palmer, T. N., 2014: More reliable forecasts with less precise computations: a fast-track route to cloud-resolved weather and climate simulators? *Phil. Trans. R. Soc. A.*, **372**, 20130391.
139. Peralta, C., Ben Bouallégue, Z., Theis, S. E., Gebhardt, C. and Buchhold, M., 2012: Accounting for initial condition uncertainties in COSMO-DE-EPS. *J. Geophys. Res.* **117**, 1–13.
140. Pereira, M. B. and Berre, L., 2006: The Use of an ensemble approach to study the background error covariances in a global NWP model. *Mon. Wea. Rev.*, **134**, 2466–2489.
141. Pergaud, J., Masson, V., Malardel, S. and Couvreux, F., 2009: A parameterization of dry thermals and shallow cumuli for mesoscale numerical weather prediction. *Bound.-Layer Meteor.*, **132**, 83–106.
142. Pielke, R. A., 2002: Mesoscale meteorological modelling. *Academic press*, pp 676.
143. Pinty, J.-P. and Jabouille, P., 1998: A mixed-phased cloud parameterization for use in a mesoscale non-hydrostatic model: Simulations of a squall line and of orographic precipitation. Conf. on Cloud Physics, Everett, WA, *Amer. Meteor. Soc*., 217–220.
144. Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F. and Simmons, A., 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Q. J. R. Meteorol. Soc.*, **126**, 1143–1170.
145. Raynaud, L., Pannekoucke, O., Arbogast, P. and Bouttier, F., 2015: Application of a Bayesian weighting for short-range lagged ensemble forecasting at the convective scale. *Q. J. R. Meteorol. Soc*., **141**, 459–468.

146. Raynaud, L. and Bouttier, F., 2016: Comparison of initial perturbation methods for ensemble prediction at convective scale. *Q. J. R. Meteorol. Soc.*, **142**, 854–866.
147. Raynaud, L. and Bouttier, F., 2017: The impact of horizontal resolution and ensemble size for convective-scale probabilistic forecasts. *Q. J. R. Meteorol. Soc.*, **143**, 3037–3047.
148. Roberts, N. M. and Lean, H. W., 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97.
149. Romine, G. S., Schwartz, C. S., Berner, J., Fossell, K. R., Snyder, C., Anderson, J. L. and Weisman, M. L., 2014: Representing forecast error in a convection-permitting ensemble system. *Mon. Wea. Rev.*, **142**, 4519–4541.
150. Saito, K., Seko, H., Kunii, M. and Miyoshi, T., 2011: Effect of lateral boundary perturbations on the breeding method and the local ensemble transform Kalman filter for mesoscale ensemble prediction. *Tellus A*, **64**, 11594.
151. Schellander-Gorgas, T., Wang, Y., Meier, F., Weidle, F., Wittmann, C. and Kann, A., 2017: On the forecast skills of a convection permitting ensemble. *Geosci. Model Dev.*, **10**, 35–56.
152. Schwartz, C. S. and Coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25**, 263–280.
153. Schwartz, C. S., Romine, G. S., Sobash, R. A., Fossell, K. R. and Weisman, M. L., 2015: NCAR's experimental real-time convection-allowing ensemble prediction system. *Wea. Forecasting*, **30**, 1645–1654.
154. Schwartz, C. S., 2016: Improving large-domain convection-allowing forecasts with high-resolution analyses and ensemble data assimilation. *Mon. Wea. Rev.*, **144**, 1777–1803.
155. Schwartz, C. S. and Sobash, R. A, 2017: generating probabilistic forecasts from convection-allowing ensembles using neighborhood approaches: a review and recommendations. *Mon. Wea. Rev.*, **145**, 3397–3418.
156. Schwartz, C. S. and Sobash, R. A., 2019: revisiting sensitivity to horizontal grid spacing in convection-allowing models over the central and eastern United States. *Mon. Wea. Rev.*, **147**, 4411–4435.

157. Schwartz, C. S., 2019: Medium-range convection-allowing ensemble forecasts with a variable-resolution global model. *Mon. Wea. Rev.*, **147**, 2997–3023.
158. Schwartz, C. S., Wong, M., Romine, G. S., Sobash, R. A. and Fossell, K. R., 2020: Initial conditions for convection-allowing ensembles over the conterminous United States. *Mon. Wea. Rev*., **148**, 2645–2669.
159. Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C. and Masson, V., 2011: The AROME-France convective scale operational model. *Mon. Wea. Rev.*, **139**, 976–991.
160. Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.*, **132**, 3019–3032.
161. Skamarock, W. C., Duda, M. G., Ha, S. and Park, S.-H., 2018: Limited-area atmospheric modeling using an unstructured mesh. *Mon. Wea. Rev.*, **146**, 3445–3460.
162. Stanešić, A., Horvath, K. and Keresturi, E., 2019: Comparison of NMC and ensemble-based climatological background-error covariances in an operational limited-area data assimilation system. *Atmosphere*, **10**, p. 570.
163. Stensrud, D., 2007: Parameterization schemes: Keys to Understanding Numerical Weather Prediction Models. Cambridge University Press, 459 pp.
164. Storto, A. and Randriamampianina, R., 2010: Ensemble variational assimilation for the representation of background error covariances in a high-latitude regional model. *J. Geophys. Res.*, **115**, D17204.
165. Talagrand, O., Vautard, R. and Strauss, B., 1997: Evaluation of probabilistic prediction systems. Proc. ECMWF Workshop on Predictability, Reading, United Kingdom, ECMWF, 1–26.
166. Termonia, P., Deckmyn, A. and Hamdi, R., 2009: Study of the lateral boundary condition temporal resolution problem and a proposed solution by means of boundary error restarts. *Mon. Wea. Rev.*, **137**, 3551–3566.
167. Termonia, P., Fischer, C., Bazile, E., Bouyssel, F., Brožková, R., Bénard, P., Bochenek, B., Degrauwe, D., Derková, M., El Khatib, R., Hamdi, R., Mašek, J., Pottier, P., Pristov, N., Seity, Y., Smolíková, P., Španiel, O., Tudor, M., Wang, Y., Wittmann, C. and Joly, A., 2018: The ALADIN System and its canonical model configurations AROME CY41T1 and ALARO CY40T1, *Geosci. Model Dev.*, **11**, 257–281.
168. Theis, S. E., Hense, A. and Damrath, U., 2005: Probabilistic precipitation forecasts from a deterministic model: A pragmatic approach. *Meteor. Appl.*, **12**, 257–268.

169. Toth, Z. and Kalnay, E., 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.
170. Toth, Z. and Kalnay, E., 1997: Ensemble Forecasting at NCEP: the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3318.
171. Tudor, M., Stanešić, A., Horvath, K. and Bajić, A., 2013: Forecasting weather in Croatia using ALADIN numerical weather prediction model. *Climate Change and Regional/Local Responses*, P. Rayand and Y. Zhang, Eds., InTech, 59–88,
172. Váňa, F., Bénard, P., Geleyn, J.-F., Simon, A. and Seity, Y., 2008: Semi-Lagrangian advection scheme with controlled damping: An alternative to nonlinear horizontal diffusion in a numerical weather prediction model. *Q. J. R. Meteorol. Soc.*, **134**, 523–537.
173. Váňa, F., Düben, P., Lang, S., Palmer, T., Leutbecher, M., Salmond, D. and Carver, G., 2017: single precision in weather forecasting models: an evaluation with the IFS. *Mon. Wea. Rev.*, **145**, 495–502.
174. Vannitsem, S., 2017: Predictability of large-scale atmospheric motions: Lyapunov exponents and error dynamics. *Chaos*, **27**, 032101.
175. Vendrasco, E. P., Sun, J., Herdies, D. L. and Frederico de Angelis, C., 2016: Constraining a 3DVAR radar data assimilation system with large-scale analysis to improve short-range precipitation forecasts. *J. Appl. Meteor. Climatol.*, **55**, 673–690.
176. Vié, B., Molinié, G., Nuissier, O., Vincendon, B., Ducrocq, V., Bouttier, F. and Richard, E., 2011: Hydro-meteorological evaluation of a convection-permitting ensemble prediction system for Mediterranean heavy precipitating events. *Nat. Hazards Earth Syst. Sci.*, **12**, 2631–2645.
177. Wang, X. and Bishop, C., 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci*., **60**, 1140–1158.
178. Wang, Y., Belluš, M., Wittmann, C., Steinheimer, M., Weidle, F., Kann, A., Ivatek-Šahdan, S., Tian, W., Ma, X., Tascu, S. and Bazile, E., 2011: The Central European limited-area ensemble forecasting system: ALADIN-LAEF. *Q. J. R. Meteorol. Soc.*, **137**, 483–502.
179. Wang, Y., Tascu, S., Weidle, F. and Schmeisser, K., 2012: Evaluation of the added value of regional ensemble forecasts on global ensemble forecasts. *Wea. Forecasting*, **27**, 972–987.

180. Wang, Y., Belluš, M., Geleyn, J.-F., Ma, X., Tian, W. and Weidle, F., 2014: A new method for generating initial condition perturbations in a regional ensemble prediction system: Blending. *Mon. Wea. Rev.*, **142**, 2043–2059.
181. Wang, Y., Meirold-Mautner, I., Kann, A., Slak, A., Simon, A., Vivoda, J., Bica, B., Böcskör, E., Brezková, L., Dantinger, J., Giszterowicz, M., Heizler, G., Iwanski, R., Jachs, S., Bernard, T., Kršmanc, R., Merše, J., Micheletti, S., Schmid, F. and Vadislavsky, E., 2017: Integrating nowcasting with crisis management and risk prevention in a transnational and interdisciplinary framework. *Meteorol. Zeitschrift.*, **26**, 459–473.
182. Wang, Y., Belluš, M., Ehrlich, A., Mile, M., Pristov, N., Smolíková, P., Španiel, O., Trojáková, A., Brozkova, R., Cedilnik, J., Klarić, D., Kovačić, T., Mašek, J., Meier, F., Szintai, B., Tascu, S., Vivoda, J., Wastl, C. and Wittmann, C., 2018: 27 years of Regional Cooperation for Limited Area Modelling in Central Europe (RC LACE). *Bull. Amer. Meteor. Soc.*, **99**, 1415–1432.
183. Wang, Y., Bellus, M., Weidle, F., Wittmann, C., Tang, J., Meier, F., Xia, F. and Keresturi, E., 2019: Impact of land surface stochastic physics in ALADIN-LAEF. *Q. J. R. Meteorol. Soc.*, **145**, 3333–3350.
184. Warner, T. T., Peterson, R. A. and Treadon, R. E., 1997: A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. *Bull. Amer. Meteor. Soc.*, **78**, 2599–2617.
185. Wastl, C., Wang, Y. Atencia, A. and Wittmann, C., 2019: A hybrid stochastically perturbed parametrization scheme in a convection-permitting ensemble. *Mon. Wea. Rev.*, **147**, 2217–2230.
186. Wastl, C., Wang, Y., Atencia, A., Weidle, F., Wittmann, C., Zingerle, C. and Keresturi, E., 2021: C-LAEF: The convection permitting limited area ensemble forecasting system of Austria. Accepted for publication at *Q. J. R. Meteorol. Soc.*
187. Wei, M., Toth, Z., Wobus, R. and Zhu, Y., 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global ensemble forecast systems. *Tellus A*, **60**, 62–79.
188. Weyn, J. A. and Durran, D., 2017: The dependence of the predictability of mesoscale convective systems on the horizontal scale and amplitude of initial errors in idealized simulations. *J. Atmos. Sci*, **74**, 2191–2210.

189. Weyn, J. A. and Durran, D., 2018: Ensemble spread grows more rapidly in higher-resolution simulations of deep convection. *J. Atmos. Sci*., **75**, 3331–3345.
190. Weyn, J. A., Durran, D. R. and Caruana, R., 2019: Can machines learn to predict weather? Using deep learning to predict gridded 500-hPa geopotential height from historical weather data. *J. Adv. Model. Earth Syst.*, **11**, 2680–2693.
191. Wilks, D. S., 1997: Resampling hypothesis tests for autocorrelated fields. *J. Climate*, **10**, 65–82.
192. Wilks, D. S., 2006: Statistical Methods in the Atmospheric Sciences. 2nd ed. Academic Press, 627 pp.
193. Wolff, J. K., Harrold, M., Fowler, T., Gotway, J. H., Nance, L. and Brown, B. G., 2014: Beyond the basics: evaluating model-based precipitation forecasts using traditional, spatial, and object-based methods. *Wea. Forecasting*, **29**, 1451–1472.
194. Wu, W.-S., Parrish, D. F., Rogers, E. and Lin, Y., 2017: Regional ensemble–variational data assimilation using global ensemble forecasts. *Wea. Forecasting*, **32**, 83–96.
195. Xue, M. and Coauthors, 2007: CAPS real-time storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2007 Spring Experiment. Preprints, 22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction, Salt Lake City, UT, *Amer. Meteor. Soc.*, **3B**.
196. Yang, X., 2005: Analysis blending using a spatial filter in gridpoint model coupling. HIRLAM Newsletter, No. 48, Article 10, HIRLAM Programme, de Bilt, Netherlands, 49–55.
197. Zhang, F. Q., Bei, N. F., Rotunno, R., Snyder C. and Epifanio, C. C., 2007: Mesoscale predictability of moist baroclinic waves: Convection permitting experiments and multistage error growth dynamics. *J. Atmos. Sci*., **64**, 3579–3594.
198. Zhang, X., 2019: Multiscale characteristics of different-source perturbations and their interactions for convection-permitting ensemble forecasting during SCMREX. *Mon. Wea. Rev.*, **147**, 291–310.
199. Zhou, X., Y. Zhu, D. Hou, and D. Kleist, 2016: A Comparison of perturbations from an ensemble transform and an ensemble Kalman filter for the NCEP global ensemble forecast system. *Wea. Forecasting*, **31**, 2057–2074.
200. Žagar, N., Andersson, E. and Fisher, M., 2005: Balanced tropical data assimilation based on a study of equatorial waves in ECMWF short-range forecast errors. *Q. J. R. Meteorol. Soc.*, **131**, 987–1011.

§ 11. APPENDIX

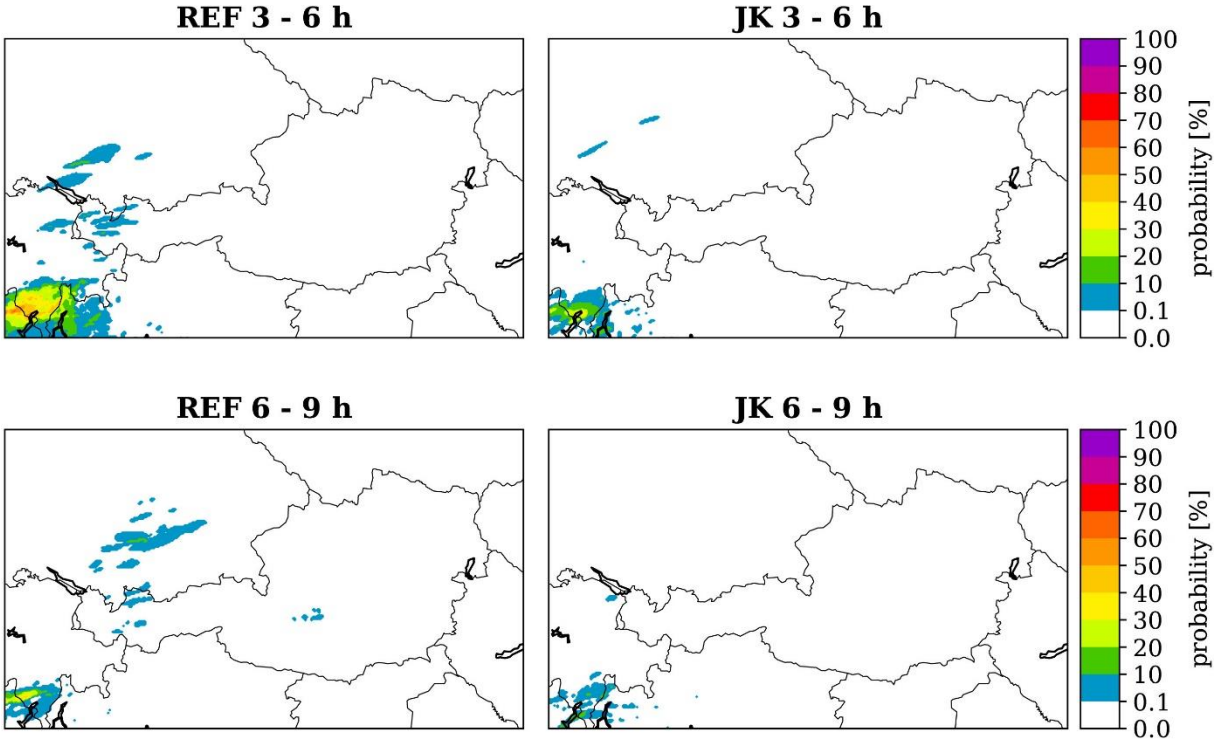

Fig. A1. As Fig. 21, but for 10 mm threshold.

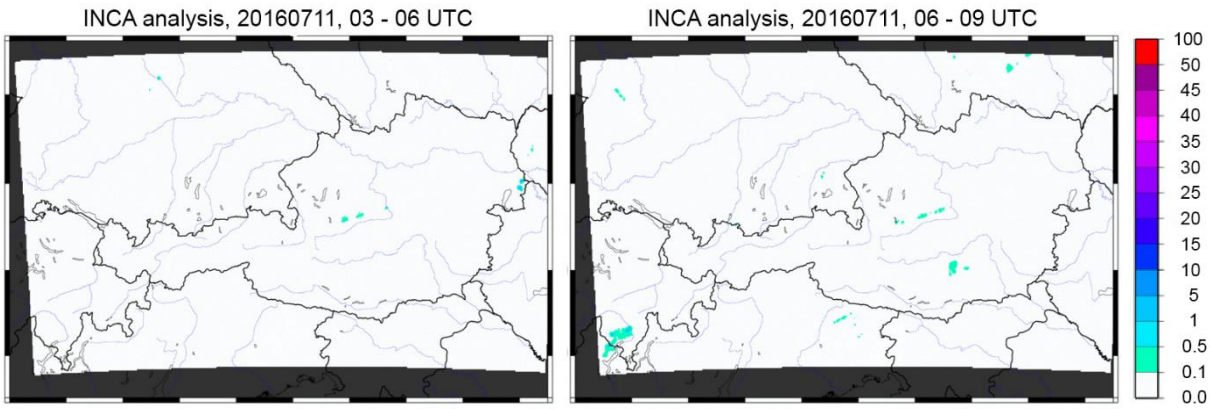

Fig. A2. Observed 3-h precipitation (mm/3h) on (a) 11 July 2016 at 0600 UTC and (b) 11 July 2016 at 0900 UTC.

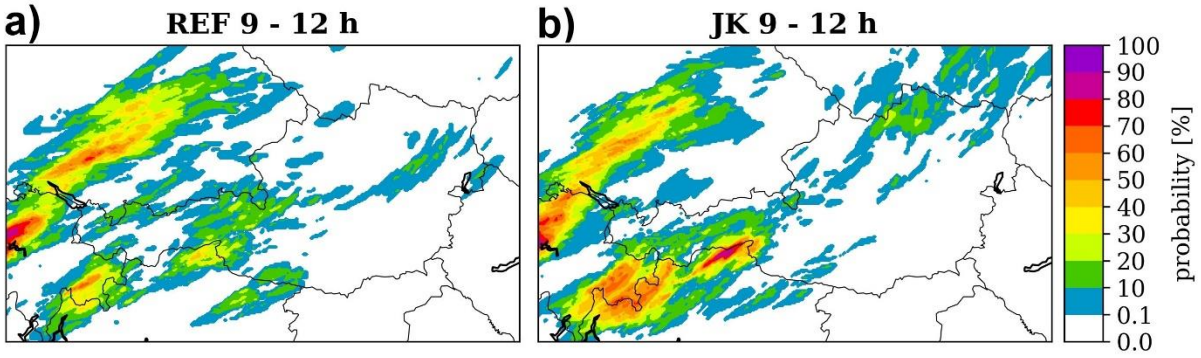

Fig. A3. As for (a) and (b) on Fig. 22, but for 10 mm threshold.

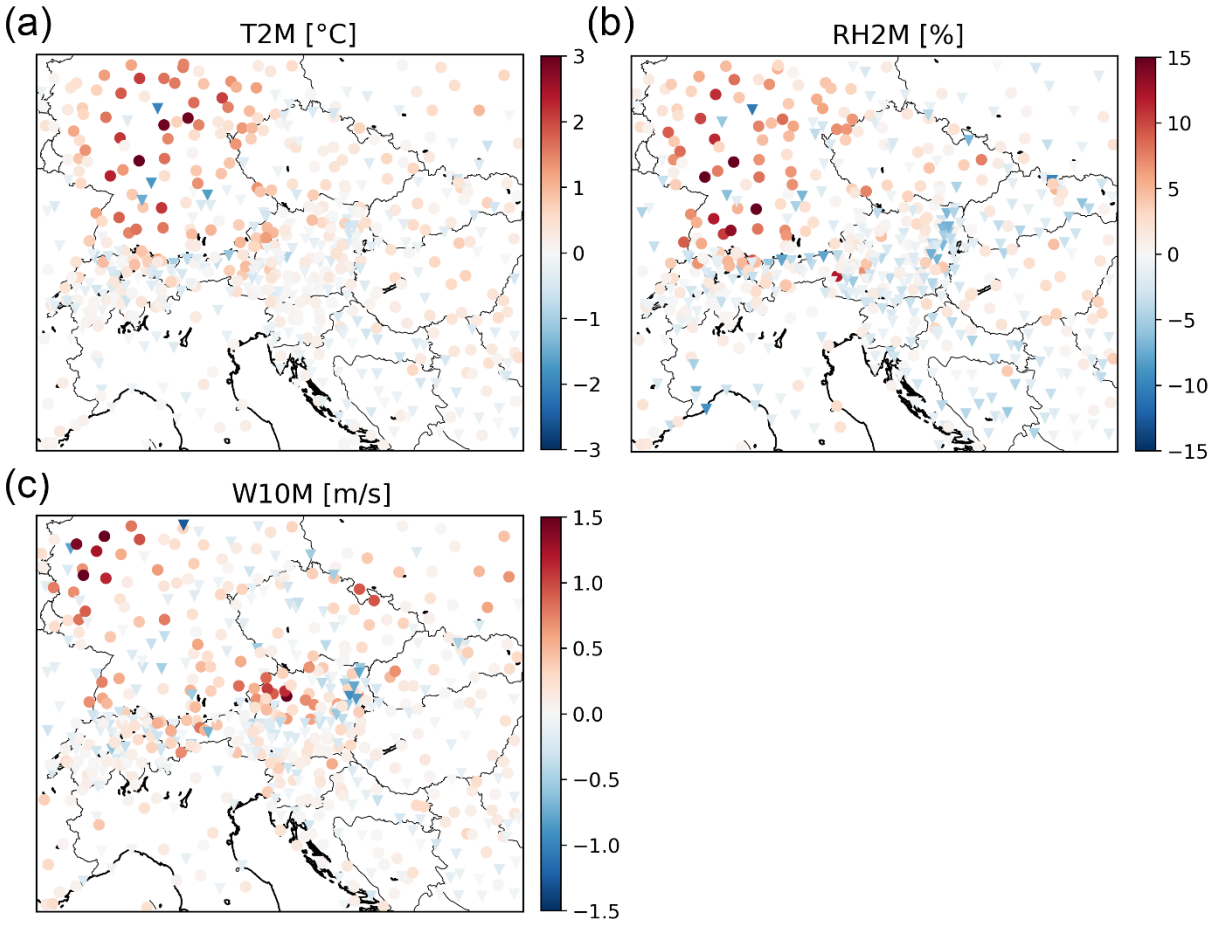

Fig. A4. As on Fig. 24 but for (a) T2M, (b) RH2M and (c) W10M. Averaged over all forecast ranges from Fig. 24.

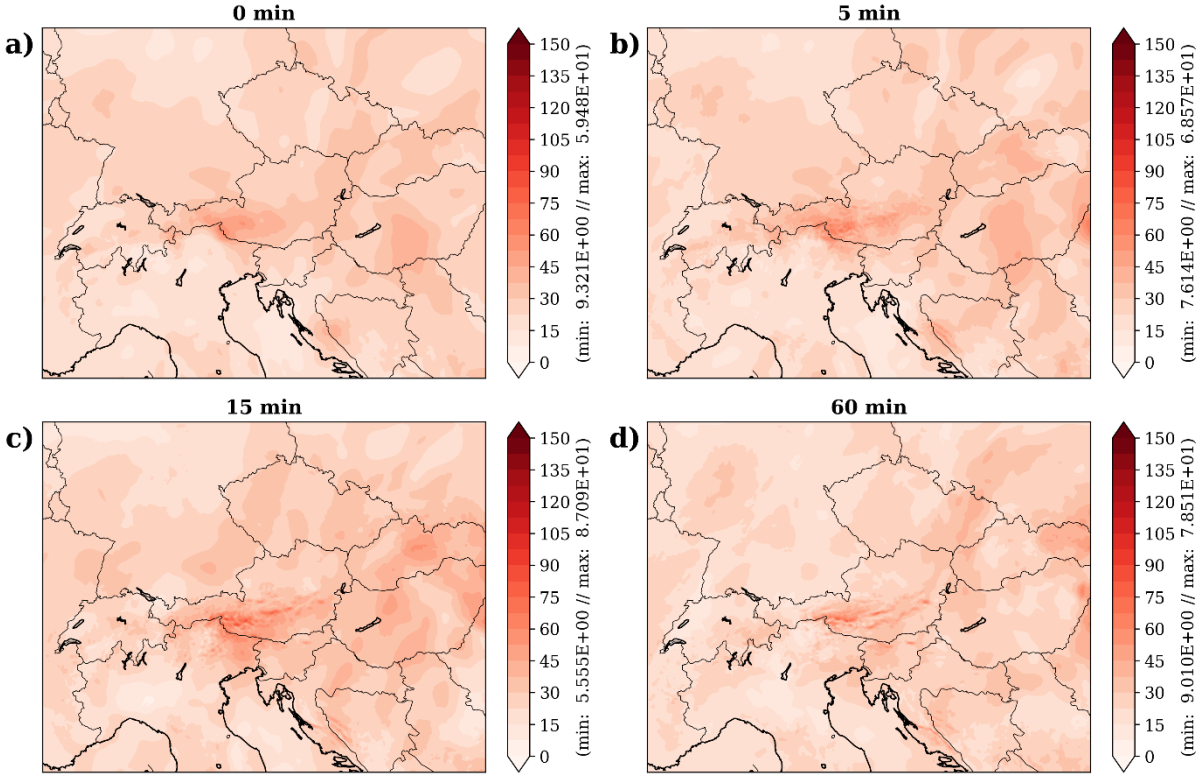

Fig. A5. As Fig. 28 but for DOWN.

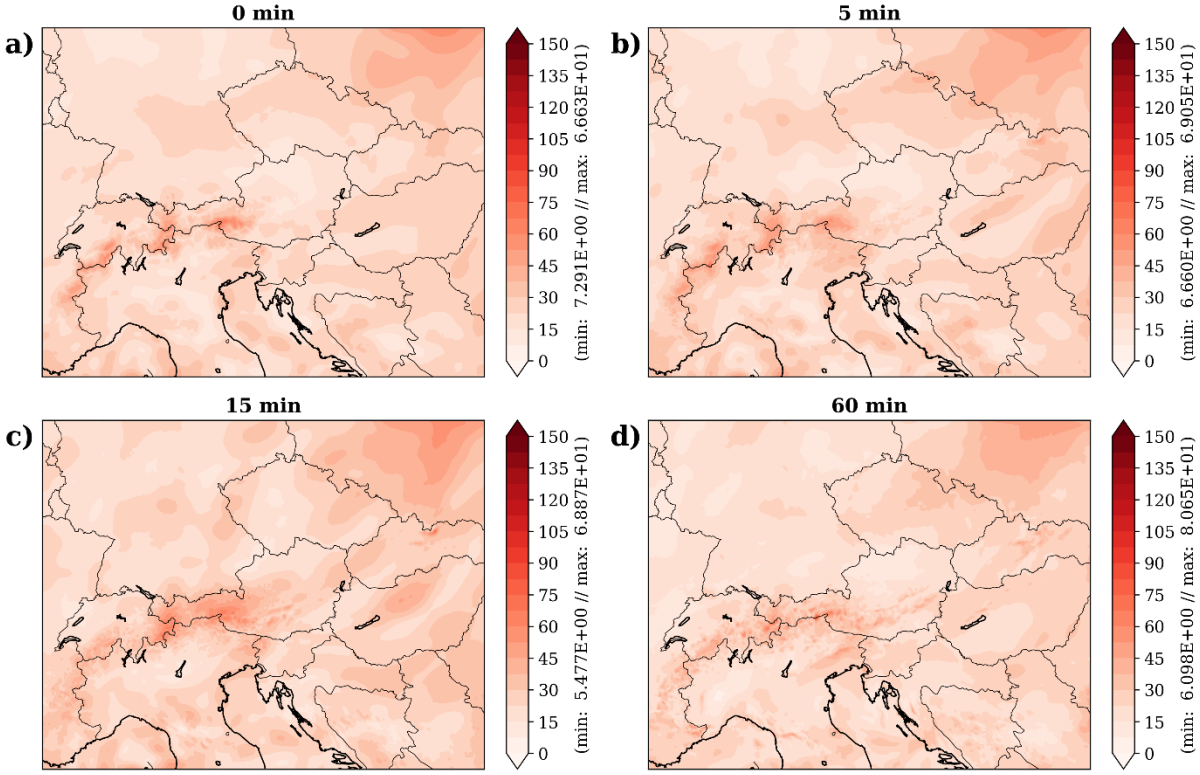

Fig. A6. As Fig. 30. But for DOWN.

§ 12. PROŠIRENI SAŽETAK

12.1. Uvod

Numerička prognoza vremena (npr., Kalnay, 2003; Coiffier, 2011; Bauer i sur., 2015) metoda je kojom se vrijeme prognozira pomoću fizikalnih jednadžbi rješavanih numerički. Skup takvih jednadžbi i svih pripadajućih alata potrebnih za uspješno izvođenje zovemo numeričkim modelom. Da bi se jednadžbe uspješno riješile na računalu, potrebno ih je diskretizirati, tj. prevesti u oblik pogodan računalu. Najpopularniji način je da ih se napiše na pravilnoj 3D mreži točaka. Razmak između tih točaka naziva se korak mreže. Tipičan korak mreže današnjih modela iznosi 1-10 km.

Model se može izvršavati na ograničenom području (regionalni model; LAM) ili na području cijele Zemlje (globalni model). Početni uvjeti od kojih počinje izvođenje modela dobivaju se metodama asimilacije podataka (npr. Kalnay, 2003). Asimilacija podataka posebna je disciplina kojoj je cilj, pomoću mjerenja i prošle prognoze modela, pronaći stanje atmosfere najbliže stvarnosti (tzv. analiza). S tom su svrhom razvijene brojne metode (npr., Bannister, 2017; Gustafsson i sur., 2018) i većina ih je bazirana na dobro poznatom principu minimizacije varijance ostatka. Danas su najpopularnije varijacijske metode (3D-Var ili 4D-Var; npr., Rabier i sur., 2000; Gustafsson i sur., 2004) koje traže analizu tako da minimiziraju funkciju troška (jednadžbe (1.4) i (1.5)). Osim početnih uvjeta, regionalni modeli trebaju i bočne rubne uvjete (npr., Warner i sur., 1997). Oni se uzimaju od nekog drugog modela veće domene (uglavnom je to globalni model).

Jednu integraciju modela u vremenu nazivamo determinističkom prognozom. Još od samih početaka numeričke prognoze vremena, ljudi su bili svjesni njenih brojnih nedostataka (Charney, 1951) kao npr. pogreške u procjeni početnih i rubnih uvjeta, pogreške u samim modelima, itd. Tomu nikako nije pogodovalo ni Lorenzovo otkriće kaosa (Lorenz, 1963). Postojanje brojnih izvora pogrešaka i kaotična priroda numeričkih metoda nužno vode do potpunog gubitka točnosti prognoze nakon nekog vremena koje je Lorenz procijenio na otprilike dva tjedna (Lorenz, 1965). Procjena je to koja stoji i danas (Buizza and Leutbecher, 2015).

Nemogućnost točne procjene početnog i budućeg stanja atmosfere vodi prema vjerojatnosnom pristupu prognozi vremena (npr., Ehrendorfer, 1997). Ansambl prognostički

sustavi (EPS; npr., Leutbecher i Palmer, 2008) posebna su tehnika prognoze vremena koja kombinira više različitih determinističkih izračuna (tzv. različiti članovi) stanja atmosfere u jedan zajednički skup, tzv. ansambl. Svaki član ansambla počinje iz neznatno izmijenjenih početnih i/ili rubnih uvjeta. Te neznatne izmjene nazivaju se perturbacije. Statističkim metodama iz takvog se ansambla može procijeniti točno stanje i dati mjera nesigurnosti procjene točnog stanja atmosfere (Leith, 1974) kao i izdati vjerojatnosna prognoza vremena.

Kako bi EPS postigao te ciljeve, potrebno je perturbirati sve izvore nesigurnosti – početne uvjete, rubne uvjete i sam numerički model. Danas postoje brojne metode razvijene posebno za svaki izvor nesigurnosti (Houtekamer i sur., 1996; Nutter i sur., 2004b; Buizza i sur., 2005; Vié i sur., 2011; Wang i sur., 2011; Bouttier i sur., 2012; Nuissier i sur., 2012; Romine i sur., 2014; Wang i sur., 2019).

Posljednjih se godina ansambl pristup sve češće koristi i u regionalnim modelima (LAMEPS; Xue i sur., 2007; Bowler i sur., 2008b; Clark i sur., 2009; Vié i sur., 2011; Wang i sur., 2011; Peralta i sur., 2012; Wang i sur., 2012; Schellander-Gorgas i sur., 2017). Razlog tomu je činjenica da pogreške u modelima rastu brže što je korak mreže manji (Hohenegger i Schär, 2007; Zhang *i sur.*, 2007; Judt, 2018). Kada te pogreške dosegnu svoju maksimalnu vrijednost, sve informacije s tih skala mogu se promatrati kao šum i deterministčke prognoze više nemaju smisla. Kod modela konvektivne skale to se događa već nakon nekoliko sati prognoze (Weyn i Durran, 2017). Zbog toga je vjerojatnosni pristup izrazito važan kod takvih modela.

Međutim, LAMEPS-ovi imaju i svojih problema koje je potrebno riješiti. U ovoj ćemo se disertaciji usredotočiti na tri problema:

1. Neusklađenost perturbacija početnih i bočnih rubnih uvjeta. Ako se perturbacije početnih uvjeta i bočnih rubnih uvjeta proizvode neovisno (što uvijek i je slučaj), na granicama domene mogu se pojaviti veliki gradijenti u meteorološkim poljima koji posljedično uzrokuju nerealistične težinske i zvučne valove (npr., Caron, 2013). Problem je to koji u potpunosti onemogućuje satne asimilacijske cikluse.
2. Procjena gibanja i nesigurnosti velike skale. Budući da su LAM-ovi ograničeni veličinom svoje domene, brojem mjerenja, nesavršenom formulacijom rubnih uvjeta itd., njihova informacija o gibanjima na velikim skalama nužno će biti nedostatna. LAMEPS-ovi imaju i dodatan problem loše procjene nesigurnosti na tim skalama.
3. Stvarna rezolucija modela. Važno je razumjeti da korak mreže i rezolucija modela nisu sinonimi (npr. Grasso, 2000). Stvarna rezolucija modela minimalno je 5 puta veća od

koraka njegove mreže (npr., Skamarock, 2004; Mittermaier, 2014). To znači da bismo sve točke mreže unutar te površine trebali promatrati jednako vjerojatnima i izlaze modela prilagoditi toj činjenici.

Rješenja spomenutih problema djelomično dolaze kroz sljedeće metode:

a) Metode miješanja. Ako znamo da procjena velikih skala regionalnog modela ne valja, informaciju o velikim skalama možemo uzeti iz onog modela za kojeg znamo da je bolji u toj procjeni (većinom je to globalni model). To se uglavnom radi metodama miješanja (miješaju se velike i male skale). Razni autori razvili su razne pristupe tom problemu (Brožkovà i sur., 2001; Yang, 2005; Wang i sur., 2014; Hsiao i sur., 2015; Verdasco i sur., 2016; Müller i sur., 2017). Vrlo je popularna metoda miješanja preko digitalnog filtera (Brožkovà i sur., 2001). Wang i sur. (2014) i Hsiao i sur. (2015) pokazali su kako su takve metode uspješne u smanjenu problema procjene gibanja velikih skala, dok je Caron (2013) pokazao da ona rješava i problem neusklađenosti perturbacija početnih i bočnih rubnih uvjeta. Međutim, Guidard i Fischer (2008; GF08) obrazlažu zašto ona nije u potpunosti ispravna i predlažu tzv. *Jk* metodu miješanja koja ima teorijski poželjnija svojstva. Dahlgren i Gustafsson (2012) pokazuju uspješnost *Jk* metode kod problema procjene gibanja velikih skala.

b) Metode susjedstva. Problem stvarne rezolucije modela može se ublažiti primjenama metoda susjedstva (Theis i sur., 2005; Ebert, 2008; Bouallegue i sur., 2013; Mittermaier, 2014; Schwartz i Sobash, 2017). Slika 3 prikazuje ideju susjedstva. Sivom bojom označeno je 5 × 5 susjedstvo oko točke (x_0, y_0). Ukupan broj točaka tog susjedstva je $N_b = 25$ što znači da više nemamo jednu prognozu za točku (x_0, y_0), nego njih 25. Mittermaier (2014) pokazuje kako primjena takve metode pomaže kod problema 3. Poglavlje 12.8 dodatno opisuje kako se metoda susjedstva može kombinirati s EPS-om.

Na ZAMG-u se trenutno razvija novi ansambl sustav konvektivne skale – C-LAEF. Glavni cilj ove disertacije je:

- Razviti novu metodu perturbacije početnih uvjeta koja će se koristiti u C-LAEF-u i koja će pokušati riješiti probleme neusklađenosti perturbacija i procjene gibanja i nesigurnosti velikih skala. Ova metoda kombinira perturbacije velike skale iz globalnog ansambla (koji ujedno daje i bočne rubne uvjete) preko *Jk* metode s perturbacijama male skale koje dolaze iz EDA-a regionalnog ansambla. Naša je hipoteza kako ćemo ovom metodom postići konzistentnije perturbacije početnih i

bočnih rubnih uvjeta kao i bolji opis gibanja i nesigurnosti velikih skala. Također, *Jk* metoda omogućuje da se spajanje malih i velikih skala odvije na dosljedan i optimalan način u smislu minimizacije varijance pogreške.

Sekundarni ciljevi ove disertacije su:

- Poboljšati procjenu nesigurnosti početnih uvjeta u C-LAEF-u dodavanjem prošlih deterministčkih prognoza.
- Implementirati metodu susjedstva kako bismo riješili problem stvarne rezolucije modela.

12.2. Nova metoda perturbacije početnih uvjeta

Ukratko ćemo predstaviti glavne značajke *Jk* metode miješanja (za mnogo detaljniji opis, čitatelj može pogledati GF08). 3D-Var varijacijska je metoda što znači da se, za pronalaženje optimalne analize, minimizira funkcija troška *J* (jednadžbe (1.4) – (1.6)). Kod *Jk* metode, funkciji troška dodaje se jedan dodatan član - $J_k(\boldsymbol{X})$ koji predstavlja informaciju velike skale globalnog modela (jednadžbe (2.1) i (2.2)) gdje **V** predstavlja kovarijacijsku matricu pogrešaka velike skale globalnog modela. U ovoj disertaciji, *Jk* metoda implementirana je unutar 3D-Var-a AROME modela (Fischer i sur., 2005) koji asimilira sljedećih 5 varijabli: temperaturu, specifičnu vlagu, vrtložnost, divergenciju i prizemni tlak.

U ovoj disertaciji uvodimo novu metodu perturbacija početnih uvjeta – ansambl *Jk* metodu koju ćemo implementirati u model C-LAEF. Ova metoda kombinira perturbacije velike skale iz globalnog ansambla (koji ujedno daje i bočne rubne uvjete) preko *Jk* metode s perturbacijama male skale koje dolaze iz ansambla asimilacije podataka (EDA-e) regionalnog ansambla. Glavne značajke te metode su sljedeće:

a) 3D-Var EDA koristi se za pronalaženje perturbiranih analiza tako što se perturbiraju:
    i) Mjerenja
    ii) Prošle prognoze
b) Kako bismo uključili i perturbacije velike skale globalnog modela, koristimo *Jk* metodu unutar 3D-Var EDA sustava.

Ukratko, svaki član ansambla asimilira sljedeće:

a) Perturbirana mjerenja
b) Perturbiranu prošlu prognozu
c) Perturbiranu globalnu analizu

Matematički, takva je funkcija troška dana jednadžbom (2.3). Vjerujemo da ansambl *Jk* metoda može riješiti problem nekonzistentnosti perturbacija bočnih rubnih i početnih uvjeta, poboljšati opis velikih skala u modelu te poboljšati njihovu dosljednost u odnosu na metode miješanja digitalnim filterom.

12.3. Model i eksperimenti

C-LAEF ansambl sustav baziran je na deterministčkom AROME modelu konvektivne skale. AROME je spektralni nehidrostatički model prilagođen za korake mreže od oko 1 km. 3D-Var AROME modela gotovo je identičan onom razvijenom za ALADIN (Ficher i sur., 2005). Više detalja o AROME modelu mogu se pronaći u Seity i sur. (2011), Vié i sur. (2011) ili Brousseau i sur. (2016).

U ovoj disertaciji, C-LAEF koristi korak mreže od 2,5 km, 90 vertikalnih nivoa i 17 članova ansambla (16 perturbiranih i jedan neperturbirani, tj. kontrolni član). Najvažnije postavke C-LAEF modela nalaze se u Tablici 1 dok je domena integracije prikazana na Slici 4. Bočni rubni uvjeti uzimaju se od ECMWF-EPS-a. Budući da je tema ovog rada perturbacije početnih uvjeta, perturbacije modela i donjeg rubnog uvjeta nisu korištene zbog toga što želimo staviti naglasak samo na perturbacije početnih uvjeta.

Glavni je cilj ove disertacije evaluacija ansambl *Jk* metode i zato ćemo definirati tri eksperimenta:

a) REF – referentni eksperiment u kojemu C-LAEF, za perturbiranje početnih uvjeta, koristi uobičajenu EDA metodu bez *Jk* člana.
b) JK – eksperiment u kojemu C-LAEF koristi ansambl *Jk* metodu za generiranje perturbacija početnih uvjeta.
c) DOWN – eksperiment u kojemu C-LAEF koristi dinamički adaptirane analize ECMWF EPS-a kao početne uvjete. Koristi se kao referentni eksperiment u potpoglavlju 12.7 jer ovakva konfiguracija nema problema s neusklađenošću perturbacija.

Osim navedene razlike, u oba eksperimenta C-LAEF konfiguriran je potpuno jednako. Možda je važno naglasiti kako se isti globalni model (ECMWF-EPS u ovom radu) mora koristiti za bočne rubne uvjete, perturbacije velike skale u ansambl *Jk* metodi kao i za računanje **V** matrice.

Verifikaciju dvaju navedenih eksperimenata provodit ćemo na sljedeći način. Površinske varijable verificirat ćemo svaka tri sata tako da ih usporedimo s mjerenjima na 832 lokacije unutar domene (plavi četverokut na Slici 4). One uključuju: temperaturu na 2 m (T2M), relativnu vlagu na 2 m (RH2M), brzinu vjetra na 10 m (W10M) i tlak zraka sveden na razinu mora (MSLP). Visinska polja verificirat ćemo u odnosu na analizu ECMWF-a (0,1° × 0,1°) i NCEP-a (0,5° × 0,5°), a ona uključuju: temperaturu, relativnu vlagu i brzinu vjetra na 500- i 850-hPa plohi (T500, T850, RH500, RH850, W500 i W850).

Verifikacija oborine mnogo je zahtjevniji zadatak i standardne metode nisu prikladne (Ebert, 2008; Mittermaier, 2014). Zbog toga su razvijene naprednije metode za verifikaciju oborine (Gilleland i sur., 2009), a jedna od njih je i FSS (Roberts i Lean, 2008) koja će biti korištena u ovom radu. Prostorno polje izmjerene oborine proizvedeno je pomoću INCA-e (Haiden i sur., 2011, Wang i sur., 2017) koja koristi radarske podatke, mjerenja i kratkoročne prognoze kako bi proizvela visokorezolucijsku analizu oborine.

Verifikacija ansambl sustava daleko je od jednostavnog zadatka budući da se uspješna vjerojatnosna prognoza mjeri preko raznih mjera: točnost, pristranost, pouzdanost, rezolucija, diskriminacija, nesigurnost i oštrina (Murphy, 1993; Wilks, 2006). Zbog takve razine kompleksnosti, ne postoji jedna verifikacijska mjera koja nam može dati informaciju o svim ovim svojstvima prognoze. Stoga ćemo koristiti više verifikacijskih mjera koje uključuju: a) korijen srednje kvadratne pogreške (RMSE) srednjaka ansambla kao mjera njegove točnosti, b) *Continuous rank probability score* (CRPS) za procjenu ukupne uspješnosti ansambla, c) odnos RMSE-a srednjaka ansambla i njegovog rasapa te statistike stršećih vrijednosti kao dvije mjere pouzdanosti, d) dekompoziciju Brierove mjere za procjenu točnosti prognoziranih vjerojatnosti, rezolucije i pouzdanosti i e) *relative operational characteristics* (ROC) kao mjeru diskriminacije. Svaka od ovih mjera bit će ukratko opisana u potpoglavljima 5. poglavlja. Za detaljniji matematički opis navedenih mjera, čitatelj može pogledati u Talagrand i sur. (1997) i u Wilks (2006).

Kako bismo odredili jesu li eksperimenti statistički značajno različiti jedan od drugog, koristili smo *bootstrap* metodu kako je opisano u Wilks (1997) pri čemu nivo značajnosti iznosi 90 %.

12.4. Dijagnostika i ugađanje

Kovarijacijska matrica **V** važna je jer određuje na koji se način utjecaj velike skale pojedine varijable širi na ostatak domene i ostale varijable. Ovo je jasno vidljivo iz jednadžbi (4.1) – (4.3). **V** je izračunata ansambl metodom kojom se standardno računa i **B** (Berre i sur., 2006, Stanešić i sur., 2019). Osnovna dijagnostika na slikama 5 i 6 pokazuje kako je **V** izračunata ispravno i ponaša se očekivano.

U ansambl *Jk* metodi postoji nekoliko načina na koji se utjecaj velike skale može prilagoditi:

a) Odabir varijabli na koje će *Jk* metoda utjecati. U ovom radu *Jk* metoda ima utjecaj na sve varijable koje se asimiliraju, osim prizemnog tlaka jer se to pokazalo kao najuspješnije rješenje.
b) Jačina utjecaja *Jk* metode na pojedinu varijablu. U ovom radu koristimo vjerojatnosni pristup prilikom odabira jačine tog utjecaja. Faktor skaliranja za svaku varijablu nasumično je odabran između dvije empirijski određene vrijednosti, tako da svaki član ansambla ima jedinstvenu konfiguraciju *Jk* metode.
c) Inflacija perturbacija globalnog ansambla. Ukoliko se pokaže da ansambl *Jk* metoda ima negativan utjecaj na rasap regionalnog ansambla, perturbacije globalnog modela mogu se uvećati prije upotrebe u ansambl *Jk* metodi. Takve perturbacije povećavaju rasap regionalnog ansambla.

Na kraju, potrebno je odrediti koje će se točno skale uzimati iz globalnog ansambla. U ovom je radu ta granica 135 km (totalni valni broj k* = 8 (4.4)) što znači da valne duljine globalnog ansambla ispod te vrijednosti nemaju nikakav utjecaj na regionalni ansambl.

Kao osnovna dijagnostika uspješnosti ansambl *Jk* metode, provedena je spektralna analiza na novonastalim perturbacijama početnih uvjeta. Sa slike 7 vidimo da novonastale perturbacije sadrže više energije na velikim skalama i da razlika počinje upravo oko totalnog valnog broja 8. Ovime potvrđujemo da se ansambl *Jk* metoda ponaša očekivano.

12.5. Dugoročna verifikacija

Kako bismo ocijenili uspješnost ansambl *Jk* metode, proveli smo dvomjesečnu 24-h integraciju JK i REF eksperimenata od 1. 7. do 31. 8. 2016. u 1200 UTC. Slijedi kratki opis dobivenih rezultata za svaku mjeru posebno.

a) RMSE srednjaka ansambla i rasap (slike 8 i 9)

RMSE statistički značajno je smanjen kod JK-ja unutar prvih 12 h prognoze za skoro sve varijable i oba nivoa. Rasap ansambla ostao je sličan što, uz manji RMSE, znači veću pouzdanost JK-ja. Rezultati za površinske varijable manje su različiti i JK je uglavnom nešto bolji za sve varijable unutar prvih 9 h prognoze.

b) CRPS (slike 10 i 11)

   CRPS je manji za sve varijable i nivoe visinskih polja kod JK-ja. Za površinske varijable, JK je bolji kod MSLP-a i T2M-a i neutralan za ostale varijable.

c) Statistika stršećih vrijednosti (slike 12 i 13)

   Statistički značajno smanjenje stršećih vrijednosti vidljivo je za T500, T850, MSLP, RH500 i RH850 što znači da je JK pouzdaniji ansambl.

d) Dekompozicija Brierove mjere (slike 14 i 15)

   Vjerojatnosti dobivene iz JK-eksperimenta statistički su značajno bolje za razne pragove i sve varijable i oba nivoa visinskih polja, dok su neutralne za površinske varijable. Također, dekompozicija pokazuje da poboljšanje ne dolazi samo zbog povećane pouzdanosti, nego i zbog povećane rezolucije.

e) ROC (slike 16 i 17)

   Rezultati testiranja za više različitih pragova pokazuju da JK ima bolju diskriminaciju za ekstremne događaje za W500, W850, T500, T2M, RH500 i MSLP, dok je za ostale varijable bolji za manje ekstremne pragove ili neutralan.

f) Oborina (slika 18)

   FSS pokazuje da je JK bolji u prognozi oborine za sve testirane pragove i skale iznad 195 km i za pragove do 5 mm za skale iznad 45 km unutar prvih 6 h prognoze. Za ostale prognostičke sate, JK i REF ne razlikuju se značajno osim za 24-h prognostički sat kada je JK bolji za sve skale i prag iznad 1 mm.

12.6. Pojedinačni slučajevi

Najveći utjecaj ansambl *Jk* metode očekujemo u situacijama u kojima postoji snažno sinoptičko forsiranje u blizini granica domene. Naime, kao što smo obrazložili u potpoglavlju 12.1, regionalni modeli imaju problem s točnim opisivanjem velike skale. Nadalje, svaka informacija izvan domene modela ulazi unutra preko bočnih rubnih uvjeta čija je formulacija problematična (Warner i sur., 1997; Nutter i sur., 2004a; Termonia i sur., 2009). Stoga, dobivanje informacije velike skale na dodatan način preko ansambl *Jk* metode može pomoći.

12.6.1. 11. Srpanj 2016.

Slika 19 pokazuje sinoptičku situaciju 11. srpnja 2016. u 0000 i 1200 UTC. Hladna fronta povezana s ciklonom iznad Sjevernog mora premješta se preko domene C-LAEF-a. Tu JK eksperiment točnije prognozira položaj nadolazeće fronte što se vidi iz polja prizemnog tlaka (slika 20). Bolja prognoza sinoptičkog sustava velike skale vodi do, u ovom slučaju, bolje prognoze oborine (slike 21 i 22).

12.6.2. 28. Kolovoz 2016.

Slika 23 pokazuje sinoptičku situaciju 27. kolovoza 2016. u 1800 UTC te 28. kolovoza 2016. u 0000, 0600 i 1200 UTC. Zona konvergencije formira se ispred ciklone iznad Britanskog otočja koja se premješta prema istoku. Nekoliko sati kasnije, druga se ciklona formira iznad Sjevernog mora, a hladna se fronta povezana s tom ciklonom premješta se preko domene C-LAEF-a. JK eksperiment opet točnije prognozira položaj hladne fronte i daje točniju raspodjelu prizemnog tlaka kao i ostalih varijabli (slike 24 i A4). Oborina povezana sa zonom konvergencije također je uspješnije prognozirana (slika 25).

12.7. Neusklađenost perturbacija

Drugi važan zadatak ove disertacije provjeriti je u kojoj mjeri ansambl *Jk* metoda pomaže kod problema neusklađenosti perturbacija početnih i bočnih rubnih uvjeta. U tu smo svrhu izdvojili dva slučaja u kojima je zabilježena nekonzistentnost perturbacija.

12.7.1. Slučaj 1

Slika 27a pokazuje sinoptičku situaciju 17. srpnja 2016. u 0000 UTC na kojoj se vidi visinska ciklona na istoku domene C-LAEF-a. U situacijama poput ove, gdje je ciklona djelomično obuhvaćena domenom modela, čak i male razlike u položaju ciklone između globalnog i regionalnog modela mogu uzrokovati značajne neusklađenosti na rubovima domene. Upravo je to i bio slučaj ovdje. Slika 28 pokazuje vremenski razvoj rasapa ansambla (eksperiment REF) za MSLP. Anomalija u tlaku brzo se širi domenom, te je rasap MSLP-a dosta veći u odnosu na DOWN. Uzimajući u obzir brzinu gibanja anomalije, lako se zaključuje da je riječ o horizontalno propagirajućim zvučnim valovima (Lambovi valovi). Pogledamo li rezultate

eksperimenta JK, vidimo da je rasap vrlo sličan DOWN-u i da je nerealističan rasap gotovo iščeznuo nakon prvog sata (slika 28c).

12.7.2. Slučaj 2

Slika 27b pokazuje sinoptičku situaciju 15. srpnja 2016. u 0000 UTC na kojoj se vide dvije visinske ciklone na sjeveru i jugu domene C-LAEF-a. Slično kao i u prošloj situaciji, anomalije u MSLP-u razvijaju se na rubovima domene te nastavljaju putovati prema suprotnim krajevima domene (slika 30). I ovdje JK opet uvelike smanjuje nerealističan rasap MSLP-a i čini ukupan rasap ansambla vrlo sličnim onim u DOWN-u (slika 31).

Iz rezultata ovog poglavlja možemo zaključiti kako ansambl *Jk* metoda predstavlja učinkovit način ublažavanja problema neusklađenosti perturbacija.

12.8. Tehnike za dodatno unaprjeđenje ansambla

12.8.1. Ansambl susjedstva

S obzirom na argumente dane u poglavlju 12.1, pojedini EPS možemo lako proširiti tako da postojećim članovima ansambla u svakoj točki modela dodamo i članove iz njenog susjedstva. Na primjer, ako imamo EPS od 17 članova i odaberemo N_b = 25 (broj točaka susjedstva), ukupan broj članova EPS-a porast će na 17 x 25 = 250. Računanje vjerojatnosti iz takvog EPS-a, može se provesti na više načina. Schwartz i Sobash (2017) definiraju dva najkorisnija načina: NEP i NMEP. NEP pretpostavlja da je svih 25 članova susjedstva jednako vjerojatno i tretira takav EPS kao EPS od 250 ravnopravnih članova. NMEP, s druge strane, uzima neku funkciju susjedstva f (npr. maksimum ili minimum) i onda ponovno formira EPS od 17 članova. Schwartz i Sobash (2017) primjenjuju NEP i NMEP pristup pri prognozi oborine. NMEP je bolji kod prognoze ekstrema, dok je NEP bolji kod prognoze klimatološki učestalijih događaja.

U ovoj disertaciji nećemo se detaljno baviti oborinom, nego ćemo ocijenit NEP i NMEP pristup pri prognozi ostalih površinskih varijabli (MSLP, T2M, RH2M i W10M) u usporedbi s eksperimentom REF. Važno je napomenuti da kod odabira susjednih točaka, pažnju treba obratiti na geografski položaj svake točke modela. Npr., ne želimo da susjedstvo neke nizinske kopnene točke sadrži planinske točke ili točke iznad mora. Preporuka je da se takve nereprezentativne točke izbace iz susjedstva. Također, vremenska nesigurnost prognoza

može biti značajna (Mittermaier i Stephenson, 2015), ali proširenje susjedstva u vremensku dimenziju nije bilo moguće u ovoj disertaciji zbog nedostupnosti satnih podataka.

Verifikacija (slike 33 – 37) je pokazala da NEP značajno povećava rasap i pouzdanost ansambla za sve varijable osim MSLP-a. Točnost i rezolucija malo su poboljšani za T2M i RH2M. Diskriminacija je poboljšana za T2M, RH2M i W10M, osim za ekstremne vrijednosti pragova gdje je poboljšanje vidljivo samo kod T2M.

NMEP pristup testirali smo na W10M (*f*=maksimum) i minimalnoj temperaturi na 2 m (*f*=minimum). Rezultati su pokazali da je NMEP pristup značajno bolji pri detekciji velikih brzina vjetra (slika 38) kao i točniji pri prognozi minimalnih temperatura (slika 39).

12.8.2. Dodavanje prošlih determinističkih prognoza

Budući da je, na ZAMG-u, deterministički prognostički model AROME konfiguriran jednako kao kontrolni član C-LAEF-a (osim 12 h dužeg prognostičkog razdoblja), C-LAEF ansambl možemo proširiti dodavanjem 4 prošle prognoze AROME-a i to bez gubitka na duljini prognostičkog razdoblja. Očekujemo da će dodavanje prošlih prognoza poboljšati procjenu nesigurnosti početnih uvjeta jer one sadrže dodatnu informaciju trenutnog stanja atmosfere.

Usporedit ćemo eksperiment REF s novim eksperimentom zvanim LAG u kojemu dodajemo 4 prošle prognoze AROME-a (-3, -6, -9 i -12 h) za W10M, WG10M i oborinu. Rezultati su pokazali da je LAG pouzdaniji i malo točniji EPS za W10M i WG10M. Prognoza oborine točnija s pozitivnim utjecajem vidljivim do čak 20. sata prognoze. Detekcija ekstremnih brzina vjetra i svojstvo diskriminacije također su poboljšani u odnosu na REF.

12.9. Zaključak

Danas je vjerojatnosni pristup prognozi vremena prepoznat kao superiorniji u odnosu na deterministički te postaje glavni alat za prognozu vremena većine svjetskih meteoroloških centara. To se posebno odnosi na mezo- i konvektivne skale gdje brzi gubitak predvidljivosti, tj. vrlo brz rast pogrešaka, posebno dolazi do izražaja. Nakon što se to dogodi, strukture na tim skalama mogu se smatrati šumom. Iz tih se razloga vjerojatnosni pristup putem LAMEPS-a sve češće koristi. Stoga su i motivi za rješavanje problema povezanih s LAMEPS-ovima snažni. Male računske domena, nedostatci u formulaciji rubnih uvjeta, inferiorne metode asimilacije podataka, usredotočenosti na mezoskalu i nedostupnosti mjerenja izvan domene, LAMEPS-ovi nisu u mogućnosti točno opisati nesigurnosti gibanja na velikim skalama.

Nadalje, neusklađenosti između perturbacija početnih i bočnih rubnih uvjeta LAMEPS-a mogu nastati ako su te perturbacije dobivene neovisno. Kao rezultat toga, na granici domene stvaraju se umjetni težinski i zvučni valovi koji se šire ostatkom domene i izazivaju neželjeni šum u meteorološkim poljima.

U ovoj smo disertaciji uveli novu metodu perturbacije početnih uvjeta u LAMEPS-u, tzv., ansambl *Jk* metodu. Ova metoda spaja perturbacije velike skale iz globalnog ansambla (koji ujedno daje i bočne rubne uvjete) preko *Jk* metode s perturbacijama male skale koje dolaze iz EDA-a regionalnog ansambla. Naša je hipoteza kako ćemo ovom metodom postići dosljednije perturbacije početnih i bočnih rubnih uvjeta kao i bolji opis gibanja i nesigurnosti velikih skala. Također, *Jk* metoda omogućuje da se spajanje malih i velikih skala odvije na dosljedan i optimalan način u smislu minimizacije varijance pogreške.

Ovu smo metodu implementirali u LAMEPS konvektivne skale – C-LAEF. Uspješnost metode provjerili smo kroz verifikaciju dobivenu na temelju dva mjeseca integracije modela (srpanj i kolovoz 2016.). Rezultati pokazuju kako ansambl *Jk* metoda daje točniji i pouzdaniji EPS kao i povećanje atributa rezolucije i diskriminacije, posebice za visinska polja. Utjecaj je slabiji na površinske varijable, ali poboljšanje je jasno vidljivo na prognozi oborine, tlaka i temperature. Neutralniji rezultati za površinske varijable očekivani su jer je kod njih jak utjecaj donjeg rubnog uvjeta kao i gibanja male skale na koje ansambl *Jk* metoda nema utjecaja. Dubljom analizom dva specifična meteorološka slučaja u kojima se ciklona nalazi na rubu domene, utvrđeno je da ansambl *Jk* metoda doprinosi boljoj prognozi položaja frontalnog sustava kao i drugim varijablama povezanim s njom.

Kako bismo ocijenili uspješnost nove metode pri smanjenju problema neusklađenosti perturbacija početnih i bočnih rubnih uvjeta, analizirali smo dva slučaja u kojima je došlo do izražene neusklađenosti i posljedičnog stvaranja umjetnih horizontalnih zvučnih valova u domeni. U oba je slučaja ansambl *Jk* metoda uspjela uvelike smanjiti nastali šum.

Kako bismo dodatno unaprijedili C-LAEF-ov opis nesigurnosti početnih uvjeta, dodali smo mu 4 prošle prognoze deterministčkog AROME modela kao 4 nova člana u ukupnom ansamblu. Pozitivni efekti ovakvog pristupa vidljivi su na povećanoj pouzdanosti, većem rasapu i većoj točnosti EPS-a.

Konačno, kao dodatan alat u borbi protiv brzo rastućih pogrešaka male skale kao i problema efektivne rezolucije modela, isprobali smo, tzv., metodu susjedstva. Rezultati pokazuju kako primjenom ove metode dobivamo osjetno veći rasap i pouzdanost ansambla.

Točnost i rezolucija također su povećani za neke varijable. No, možda i najbitnije, prognoza ekstremnih događaja poboljšana je.

Glavni znanstveni doprinosi ove disertacije su sljedeći. Prvo, razvijena je nova metoda perturbacija početnih uvjeta za LAMEPS – ansambl *Jk* metoda. Drugo, *Jk* metoda miješanja prvi je puta primijenjena u AROME modelu. Treće, dodavanjem prošlih deterministčkih prognoza LAMEPS-u kao i njegove kombinacije s metodom susjedstva, prognoze LAMEPS-a dodatno su unaprjeđenje. Četvrto, ovaj je rad dio razvojnog procesa novog LAMEPS-a konvektivne skale pod nazivom C-LAEF, a ansambl *Jk* metoda trenutno se koristi u njegovoj operativnoj konfiguraciji (Wastl i sur., 2021.).

§ 13. CIURRICULUM VITAE

PERSONAL INFORMATION

Name	Endi Keresturi
Date of birth	1. 1. 1990.
Place of birth	Virovitica, Croatia
Home address	Sv. Trojstva 115, 33000 Virovitica, Croatia
E-mail	endi.keresturi@gmail.com

WORK EXPERIENCE

1. 4. 2018. - Present	Head of the Applied Modelling Department Croatian Meteorological and Hydrological Service (DHMZ) Ravnice 48, 10000 Zagreb, Croatia E-mail: endi.keresturi@cirus.dhz.hr
1. 4. 2015. – 31. 3. 2018.	Department for Numerical Modelling Zentralanstalt für Meteorologie und Geodynamik (ZAMG) PhD research stay Hohe Warte 38, 1190 Vienna, Austria

EDUCATION

2015. – 2021.	Doctoral study in physics University of Zagreb, Faculty of Science, Zagreb, Croatia Department of Geophysics
2008. – 2014.	Master of Science in Physics-Geophysics University of Zagreb, Faculty of Science, Zagreb, Croatia Thesis title: *Osobine bure na zračnoj luci Dubrovnik*

HONOURS AND AWARDS

2012. – 2014.	Scholarship from the city of Virovitica

15. 2. 2017.	Best student oral presentation award - MetMed6 & MI5, Zagreb, Croatia.
4. 2020.	One of the most downloaded paper in QJRMS Keresturi, E., Wang, Y., Meier, F., Weidle, F. and Wittman, C., 2019: Improving initial condition perturbations in a convection-permitting ensemble prediction system. *Q. J. R. Meteorol. Soc.*, **145**, 993–1012.

CONFERENCES AND WORKSHOPS

21. – 22. 11. 2013.	Challenges in meteorology 3: Extreme weather and impact on society, Zagreb, Croatia
3. – 4. 11. 2014.	Workshop on advances in meso- and micro-meteorology, Gornja Stubica, Croatia
27. 11. 2014.	Application of atmospheric models in environmental protection
24. – 25. 11. 2015.	Challenges in meteorology 4: Climate change - responsibilities of the current generation, Zagreb, Croatia
20. – 22. 2. 2017.	MetMed6 & MI5, Zagreb, Croatia
19. – 21. 9. 2018.	LACE data assimilation working days, Bucharest, Romania
19. – 23. 10. 2018.	Linux system administrator I, SRCE, Zagreb, Croatia
15. – 16. 11. 2018.	Challenges in meteorology 6: Advanced technology for solving the meteorological challenges, Zagreb, Croatia
11. 2. – 15. 2. 2019.	ALADIN/HIRLAM data assimilation training, Budapest, Hungary
25. 2. – 1. 3. 2019.	Predictability and ensemble forecast systems, ECMWF, Reading, UK
15. 3. – 18. 3. 2019.	SQL, SRCE, Zagreb, Croatia
1. – 4. 4. 2019.	29th ALADIN Workshop, Madrid, Spain
20. – 23. 5. 2019.	HyMex, Split, Croatia
28. 11. – 30. 11. 2019.	Programming in Python, SRCE, Zagreb, Croatia

30. 3. – 2. 4. 2020.	30th ALADIN Workshop, Ljubljana, Slovenia
4. – 5. 11. 2020.	Challenges in meteorology 7: Meteorology supporting public authorities
27. – 30. 10. 2020.	Joint SRNWP-EPS and Post-processing workshop

RESEARCH ACTIVITIES

4. 2014.	University of Zagreb, Faculty of Science, Zagreb, Croatia Student research project Topic: *PM1 concentration in the building of Geophysics Department during 12. 11. 2012. - 26. 4. 2013.*
1. 4. 2015. – 31. 3. 2018.	Zentralanstalt für Meteorologie und Geodynamik (ZAMG) PhD research stay Topic: *improving initial condition perturbations for C-LAEF*
24. 6. – 19. 7. 2019.	Zentralanstalt für Meteorologie und Geodynamik (ZAMG) RC LACE research stay Topic: *adding lagged deterministic forecasts to C-LAEF*

§ 14. LIST OF PUBLICATIONS

1. Keresturi, E., Wang, Y., Meier, F., Weidle, F. and Wittman, C., 2019: Improving initial condition perturbations in a convection-permitting ensemble prediction system. *Q. J. R. Meteorol. Soc.*, **145**, 993–1012.
2. Odak Plenković, I., Schicker, I., Dabernig, M., Horvath, K. and Keresturi, E., 2020: Analog-based post-processing of the ALADIN-LAEF ensemble predictions in complex terrain. *Q. J. R. Meteorol. Soc.*, **146**, 1842– 1860.
3. Stanešić, A., Horvath, K. and Keresturi, E., 2019: Comparison of NMC and ensemble-based climatological background-error covariances in an operational limited-area data assimilation system. *Atmosphere*, **10**, p. 570.
4. Wang, Y., Bellus, M., Weidle, F., Wittmann, C., Tang, J., Meier, F., Xia, F. and Keresturi, E., 2019: Impact of land surface stochastic physics in ALADIN-LAEF. *Q. J. R. Meteorol. Soc.*, **145**, 3333– 3350.
5. Wastl, C., Wang, Y., Atencia, A., Weidle, F., Wittmann, C., Zingerle, C. and Keresturi, E., 2021: C-LAEF: The convection permitting limited area ensemble forecasting system of Austria. Accepted for publication at *Q. J. R. Meteorol. Soc.*