HarmonEPS—The HARMONIE Ensemble Prediction System

Inger-Lise Frogner,[a] Ulf Andrae,[b] Jelena Bojarova,[b] Alfons Callado,[c] Pau Escribà,[c]
Henrik Feddersen,[d] Alan Hally,[e] Janne Kauhanen,[f] Roger Randriamampianina,[a]
Andrew Singleton,[a] Geert Smet,[g] Sibbo van der Veen,[h] and Ole Vignes[a]

[a] Norwegian Meteorological Institute (Met Norway), Oslo, Norway
[b] Swedish Meteorological and Hydrological Institute, Norrköping, Sweden
[c] Spanish Meteorological Agency (AEMET), Barcelona, Spain
[d] Danish Meteorological Institute, Copenhagen, Denmark
[e] Irish Meteorological Service (Met Éireann), Dublin, Ireland
[f] Finnish Meteorological Institute, Helsinki, Finland
[g] Royal Meteorological Institute of Belgium, Brussels, Belgium
[h] Royal Netherlands Meteorological Institute, De Bilt, Netherlands

ABSTRACT

HarmonEPS is the limited-area, short-range, convection-permitting ensemble prediction system developed and maintained by the HIRLAM consortium as part of the shared ALADIN–HIRLAM system. HarmonEPS is the ensemble realization of HARMONIE–AROME, used for operational short-range forecasting in HIRLAM countries. HarmonEPS contains a range of perturbation methodologies to account for uncertainties in the initial conditions, forecast model, surface, and lateral boundary conditions. This paper describes the state of the system at the version labeled cycle 40 and highlights some directions for further development. The different perturbation methods available are evaluated and compared where appropriate. Several institutes have operational or preoperational implementations of HarmonEPS, such as MEPS (Finland, Norway, and Sweden), COMEPS (Denmark), IREPS (Ireland), KEPS (the Netherlands), AEMET-γSREPS (Spain), and RMI-EPS (Belgium), and these systems are briefly described and compared with the ensemble prediction system (IFS ENS) from the European Centre for Medium-Range Weather Forecasts (ECMWF).

1. Introduction

To account for the inherently chaotic nature of the atmosphere (Lorenz 1963) ensemble prediction systems (EPSs), where a set of forecasts are provided instead of one deterministic forecast, have become the most commonly used method. EPSs were first introduced in global predictions (Toth and Kalnay 1993; Molteni et al. 1996; Houtekamer et al. 1996). Later they were also introduced in limited-area models (LAMs) (e.g., Du and Tracton 2001; Marsigli et al. 2005; Frogner et al. 2006; Bowler et al. 2008b; Iversen et al. 2011; Aspelien et al. 2011; Wang et al. 2011; García-Moya et al. 2011). EPSs have also been introduced at convection-permitting scales (e.g., Gebhardt et al. 2008; Hacker et al. 2011; Clark et al. 2011; Marsigli et al. 2014a; Romine et al. 2014; Bouttier et al. 2012, 2016; Frogner et al. 2016; Schwartz et al. 2017; Hagelin et al. 2017; Klasa et al. 2018).

Uncertainties exist in all parts of the forecasting system, and it is important to account for them to have a reliable and skillful EPS. Ideally uncertainties in initial conditions, surface and forecast model physics and dynamics all need to be addressed. For LAMs lateral boundary conditions (LBCs) are important. How to best account for these inherent uncertainties has been the topic in many studies, and several methods have been developed over the years. The method of singular

---

Corresponding author: Inger-Lise Frogner, i.l.frogner@met.no

vectors (Buizza and Palmer 1995) and the breeding of growing modes (Toth and Kalnay 1993) are examples of some of the pioneering work to account for initial state uncertainties. Other methods are, for example, ensemble Kalman filtering (EnKF; Evensen 2003) and its variations (e.g., Bishop et al. 2001; Hunt et al. 2007) and ensembles of data assimilations (EDAs; e.g., Buizza et al. 2008). There is a great variety of methods to account for model error, ranging from multiphysics and multimodels, where different parameterization schemes within one model (Wang et al. 2011), or different models (Iversen et al. 2011) are used, to stochastic model error schemes like stochastically perturbed physics tendencies scheme (SPPT; Buizza et al. 1999) and stochastically perturbed parameterizations scheme (SPP; Ollinaho et al. 2017). The importance of perturbing the surface was demonstrated in Bouttier et al. (2016) and for the lateral boundaries in Frogner and Iversen (2002) and Romine et al. (2014).

The international research program High Resolution Limited Area Model (HIRLAM) presently consists of 10 countries: Denmark, Estonia, Finland, Iceland, Ireland, Lithuania, the Netherlands, Norway, Spain, and Sweden, with France as an associated member. HIRLAM has a tradition of running EPSs, and in the years 2011–19 the multimodel, 8-km, 52-member pan-European Grand Limited Area Ensemble Prediction system (GLAMEPS) ran as a time-critical facility at ECMWF on behalf of all the HIRLAM countries and the Belgian Meteorological Institute (RMI) (Iversen et al. 2011). However, in recent years the focus has shifted from mesoscale systems to convection-permitting systems, and so also for EPS. This resulted in GLAMEPS being terminated in June 2019. Instead HIRLAM now devotes research and development into the limited-area, short-range, convection-permitting ensemble prediction system HarmonEPS. HarmonEPS was first run in an operational environment for the Sochi Winter Olympic games in 2014 (Frogner et al. 2016; Kiktev et al. 2017). HarmonEPS aims to describe uncertainty in all parts of the system. However, HarmonEPS is under development, and many sources of uncertainty are not yet taken into account, or are not yet fully known or understood.

The purpose of this paper is to give an overview of the state of HarmonEPS at the version labeled cycle 40h1.1.1 and the choices that exist for describing the uncertainties in the lateral boundaries, initial conditions, surface, and forecast model. More than one option exists for some parts. Some components are operationally tested, some are still experimental, but common for all is that they are available in this version of HarmonEPS. Many studies have shown that convection-permitting models give better results for precipitation amounts,

structure, and scale than models with parameterized convection (e.g., Done et al. 2004). However, the chaotic nature of the atmosphere limits the ability to correctly predict location and intensity and a probabilistic approach is essential for such predictions. In Frogner et al. (2019) it was shown that for one of the operational implementations of HarmonEPS (MEPS, see section 5) not only the convection-permitting model and better horizontal resolution contributed to the added value over IFS ENS for precipitation, but also the ensemble itself. It was also demonstrated that the value of MEPS is largest in summer when predictability is lower than in winter. In this paper HarmonEPS configurations with and without a variety of perturbations are compared, one at a time in most cases. Based on the behavior of convection-permitting models and the findings in Frogner et al. (2019), improved probabilistic scores resulting from introduced perturbations are considered a general improvement. However, it is acknowledged that the verification metrics used in this paper do not specifically focus on small-scale phenomena. Investigating whether or not the perturbations introduced are the optimal perturbations for a convection-permitting ensemble will receive more attention in further studies where the HarmonEPS perturbations described here will serve as a reference. A more specific presentation of the perturbation strategies that are relevant for further HarmonEPS development is given in section 6. The basic configuration of HarmonEPS is described in section 2, the verification methodology used in this paper in section 3 and the perturbations available in HarmonEPS in section 4. Unlike GLAMEPS, HarmonEPS is not run as a common production for all HIRLAM countries over a common area, but with different configurations in different institutes or in a cooperation between institutes. In Fig. 1 the different areas used for HarmonEPS are shown. In section 5 the various operational and preoperational implementations of HarmonEPS are briefly described. The experiments described in this paper have served as guidance when the operational and preoperational versions have been constructed. Section 6 describes some suggested directions for further improvements of HarmonEPS. A list of acronyms used in this paper can be found in the appendix.

2. HarmonEPS system

HarmonEPS is the limited-area, short-range, convection-permitting ensemble prediction system developed and maintained by the HIRLAM consortium as part of the shared ALADIN–HIRLAM system (Termonia et al. 2018). The forecast model solves the nonhydrostatic
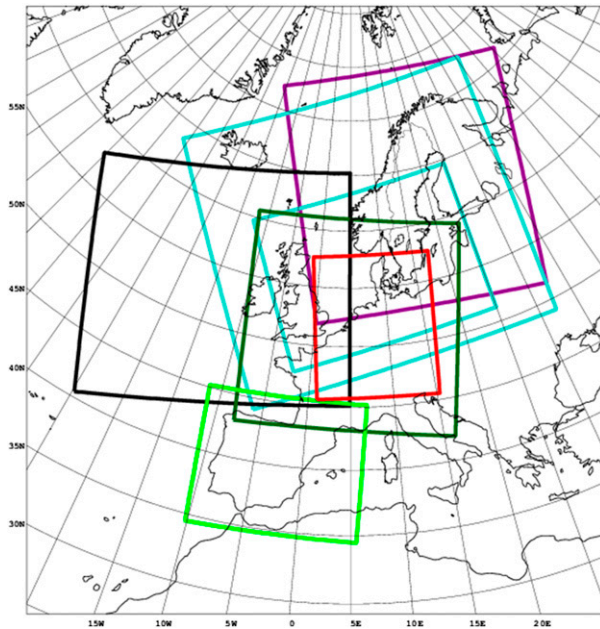
FIG. 1. HarmonEPS domains. Domain used by cooperation between Finland, Norway, and Sweden (MEPS, purple), domains used by Denmark (COMEPS, cyan), domain used by Belgium (RMI-EPS, red), domain used by Spain (AEMET-γSREPS, green), domain used by the Netherlands (KEPS, dark green), and domain used by Ireland (IREPS, black).

Eulerian equations in a mass-based vertical coordinate with semi-implicit time stepping and semi-Lagrangian advection (Bénard et al. 2010). There are two main HarmonEPS configurations, the most used is based on the HARMONIE–AROME configuration (Bengtsson et al. 2017). The physical parameterization comprises of prognostic equations for the cloud species and turbulent kinetic energy (TKE), a shallow convection scheme, and multiband radiation described in more detail in Bengtsson et al. (2017). The HarmonEPS system also has an option to run a forecast model configuration of the ALADIN–HIRLAM System that is based on a predecessor of the ALARO configuration recently described in Termonia et al. (2018). This version will be referred to as HARMONIE–ALARO in the present paper. The ALARO physics is developed with the aim of running at multiple resolutions across the gray zone where deep convection is partly resolved (i.e., from <1 to >10 km) (Termonia et al. 2018). Deep convection is parameterized according to the Modular Multiscale Microphysics and Transport scheme (3MT), (Gerard et al. 2009). The turbulence scheme is based on K theory, using a prognostic turbulent kinetic energy (Duran et al. 2014). For radiation a broadband scheme with a single shortwave and a single longwave interval is used (Ritter and Geleyn 1992; Coiffier 2011). All experiments and operational configurations described

in this paper use HarmonEPS based on HARMONIE–AROME, if not otherwise explicitly stated.

Surface processes are modeled using SURFEX in both HarmonEPS configurations (Masson et al. 2013). SURFEX divides the surface into four main types, or tiles: nature, town, sea, and inland water. For each type there are a number of schemes to choose from depending on the application. For a description of how SURFEX is used in HARMONIE the reader is referred to Bengtsson et al. (2017).

In the standard setup the forecast model is run at 2.5-km horizontal grid spacing with 65 levels in the vertical. The upper-air data assimilation system in HarmonEPS is based on 3DVAR with a 3-h cycle capable of assimilating a wide range of conventional and nonconventional observations (Brousseau et al. 2011; Berre 2000; Randriamampianina 2006; Randriamampianina et al. 2011; Lindskog et al. 2012; Ridal and Dahlbom 2017; Valkonen et al. 2017). At the surface 2-m temperature and relative humidity as well as snow cover are assimilated using optimal interpolation (Giard and Bazile 2000).

3. Verification methodology

The verification of the different HarmonEPS configurations and experiments with different settings is done against point observations using a common software package developed for use by the HIRLAM and ALADIN consortia. For (near) surface parameters [2-m temperature (T2m), 2-m dewpoint (Td2m), 2-m relative humidity (RH2m), 10-m wind speed (S10m), accumulated precipitation(AccPcp)] and cloud cover, forecasts are verified against observations from SYNOP stations. For upper-air parameters, forecasts are verified against radiosonde observations. Observations were checked for quality using a gross error check to filter out unrealistic values. A further check of the observations was done against the ensemble forecasts themselves—the standard deviation of the forecasts aggregated over all stations was computed and observations that were more than six standard deviations away from the forecast values were removed. Previous experience has shown us that this removes only those observations with large representativeness errors that could overwhelm the verification statistics. Raw forecasts are horizontally interpolated to observation station locations using bilinear interpolation, and in the case of 2-m temperature, a correction is applied to account for height differences between the model and station elevations. This height correction applies the standard adiabatic lapse rate of $6.5 \, \text{K km}^{-1}$ to the elevation difference.

For different sets of experiments, a selection from the following objective scores, which are described in detail

by Wilks (2011), is used to show the relative performance of the different models and/or model configurations.

- The root-mean-square error (RMSE) of the ensemble mean of the forecast compared with observations.
- The ensemble spread is the standard deviation of the ensemble members around the ensemble mean. This reflects the uncertainty in the forecast that the ensemble is able to model. For a well calibrated ensemble, the ensemble spread should be equal to the RMSE.
- The continuous rank probability score (CRPS) of the ensemble. This measures the distance of a continuous distribution function constructed from the ensemble forecast to the observed value. For a single ensemble member the CRPS reduces to the mean absolute error of the forecast. It is therefore negatively oriented with a perfect score being zero.
- Rank histograms (sometimes referred to as Talagrand diagrams), which depict the distribution of observations into bins of ranked ensemble members. The shape of the rank histogram gives an indication of under (u shaped) or over (convex shaped) spread, or negative (weighted toward the right) or positive (weighted toward the left) bias. In this paper the count of observations in each bin is given as the normalized frequency whereby an ensemble with perfect spread would have a normalized frequency of 1.

Other metrics, such as relative operating characteristics (ROC), Brier (skill) score, reliability diagrams, and economic value were also used, but were found not to add any extra insight over the scores used herein.

The statistical significance of the differences between the scores for different models/model configurations was computed using a bootstrap approach with 10 000 replicates. Scores are computed independently at each lead time from the forecast/observation data pooled for each forecast start date. The mean score is then computed from these pooled scores using sampling with replacement. This means that the test is insensitive to spatial correlations (Bouttier et al. 2016). If the differences between the mean scores have the same sign for at least 95% of the replicates, the differences were considered to be significant at the 95% confidence level. While this information is not shown in the figures, the differences are significant unless stated otherwise in the text.

For the most part, observation errors are not taken into account. Our goal is to compare the relative performance of different ensemble models/model configurations rather than their absolute performance. However, the impact of taking observation errors into account is discussed in section 4b.

It should also be noted that due to the large computational expense of running ensemble experiments, it was not possible to verify each model configuration against the same set of observations for the same time period.

4. Accounting for uncertainties in HarmonEPS

a. Lateral boundaries

Several options exist in HarmonEPS for perturbing the lateral boundary conditions (LBCs). If HarmonEPS is nested in a coarser-resolution (possibly global) EPS, perturbed LBCs are naturally included. The simplest option for nesting is then to use the corresponding member from the coarser EPS in the limited-area ensemble. A way to control the spread induced from the boundaries is to pick members representing, for example, the maximum spread. In HarmonEPS the complete linkage clustering method is used (Molteni et al. 2001) targeting surface pressure, wind, and temperature at 850 and 925 hPa for forecast lengths at 24 and 36 h in the nesting model. Other clustering methods giving more evenly populated clusters have been investigated by Bouttier and Raynaud (2018). Nesting HarmonEPS in a coarser-resolution EPS is a natural extension of the way deterministic limited-area models are nested in a coarser-resolution deterministic model, but it is not the only option. If a coarser-resolution EPS is not readily available it is still possible to set up a limited-area EPS. One method for doing so is the scaled lagged average forecasting (SLAF) method (Ebisuzaki and Kalnay 1991; Hou et al. 2001), where the lateral boundary perturbations are computed as scaled differences between previous forecasts from a coarser-resolution deterministic model. In our case that is the ECMWF Integrated Forecasting System (IFS) deterministic forecast (IFS HRES), valid at the forecast time according to Eq. (1):

$$\mathrm{BC\_m} = \mathrm{IFS\_0} + \mathrm{K\_m} \times (\mathrm{IFS\_N} - \mathrm{IFS\_N\text{-}6}). \quad (1)$$

Here BC_m is the lateral boundary condition for member m, IFS_0 is the latest available forecast from the nesting model, IFS_N is a forecast from the nesting model with length N and IFS_N-6 is a 6-h shorter forecast, both valid at the same time as the analysis. The term K_m is a scaling factor. The perturbations that by construction include "errors of the day," are added and subtracted to the most recent coarser-resolution deterministic forecast (IFS_0), thus providing pairs of symmetric perturbations. The scaling is applied so as to ensure similar magnitude of all perturbations. There is also an option to use the total energy norm to define the perturbation magnitude (that is the K_ms) (Keller et al. 2008). The lag between the different forecasts used is currently 6 h. It was found
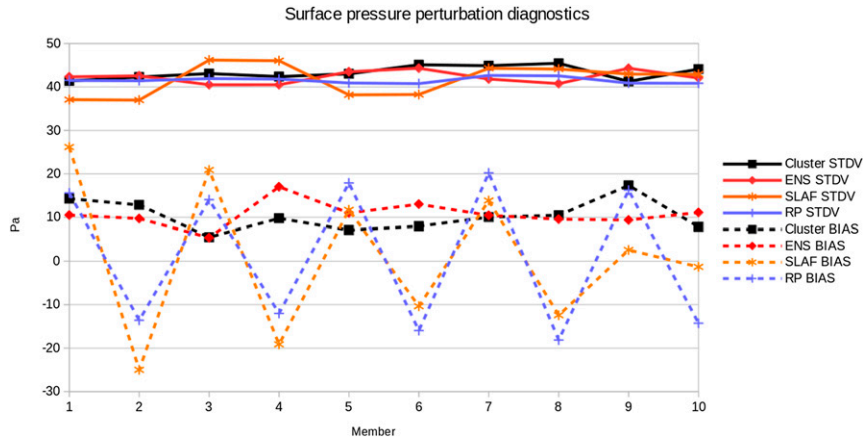
FIG. 2. Surface pressure perturbation diagnostics using different boundaries and perturbation strategies. Clustered IFS ENS in black, IFS ENS members 1–10 in red, SLAF perturbations in orange, and random perturbations (RP) in blue. Solid lines are standard deviation, and dashed lines are bias.

that larger differences introduced a clustering [e.g., mean sea level pressure (MSLP)] due to the increasing bias with forecast length in the ECMWF HRES forecasts used. One of the main drawbacks of the SLAF method is that the number of perturbations—and hence ensemble members—is limited by the length of the coarse-resolution deterministic forecast. In practice the ensemble size is limited to 10–12 perturbed members. The advantage of SLAF is that it offers an easy operational implementation with presumably already available deterministic data. Also, full model states of IFS ENS data are not archived for longer periods, and so experimentation is not possible using IFS ENS boundary data without the additional effort of archiving it before it is deleted from disk.

A somewhat similar approach to SLAF is random perturbations, following Magnusson et al. (2008). Here instead of using forecasts valid at the same time as the analysis to construct the perturbations, IFS_N and IFS_N-6 in Eq. (1) are replaced by forecasts that are valid at a randomly selected day within ±20 days from the analysis day at a randomly selected year. Random dates from the same year close to the analysis time are excluded, so that the random forecasts are always independent of the analysis.

The different methods are compared in Fig. 2 over the purple area in Fig. 1 for the period from 21 August 2017 to 20 September 2017. Shown are the standard deviation and bias for the boundary files at initial time as a function of ensemble member. The random perturbations are scaled using the total energy norm as described above, while SLAF uses fixed and tuned K_{ms}. By construction SLAF and random perturbations have pairwise symmetry that shows up as in the bias in Fig. 2.

Such symmetry is not seen for IFS ENS. Although the initial ENS perturbations by construction are symmetric for paired members, this is not the case for the 6-h forecast used here. This is due to IFS ENS perturbations having a small positive bias introduced by the SPPT scheme (Leutbecher et al. 2017). The average size of the perturbations is very similar for all four methods although it is clear that the energy norm used to construct the random perturbations gives the smallest variability between the members (blue, solid curve in Fig. 2).

In Fig. 3 we see the spread and RMSE for HarmonEPS driven by the first 10 members of IFS ENS, clustered IFS ENS, SLAF method and random perturbation method, for the same period and area as in Fig. 2. The SLAF and random perturbations are scaled so that they are of similar magnitude initially to the IFS ENS perturbations, which is then applied for all forecast lengths. Note that the initial conditions are also perturbed here, consistent with the LBC method used [see section 4c(1)]. The RMSE is very similar in all four cases whereas the spread develops differently. Using IFS ENS maintains the spread to RMSE ratio for MSLP in a better way throughout the forecast suggesting that it is a better choice, although the clustered version may give too large spread. For other variables like T2m the differences are smaller although we still see the largest spread for the clustered run at the end of the forecast length. The difference in initial spread is more related to how we construct our initial perturbations than to the evolution due to differences in the boundary forcing. With better maintenance of the spread to RMSE ratio with forecast time and less restriction on the number of members it is recommended
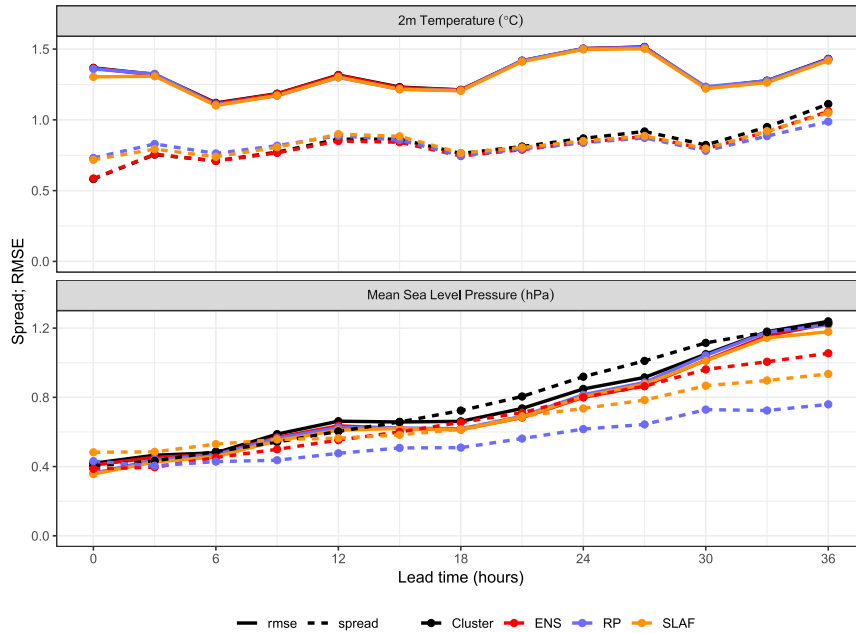
FIG. 3. Spread (dashed) and RMSE (solid) for (top) T2m at 846 stations and (bottom) MSLP at 633 stations. HarmonEPS nested in clustered IFS ENS in black, IFS ENS first 10 members in red, random perturbations in blue, and SLAF in orange.

to use IFS ENS over SLAF, possibly with a clustering option for IFS ENS. However, using SLAF does not degrade the performance much compared to nesting in IFS ENS.

b. Surface

Surface perturbations are applied to account for uncertainties in the turbulent fluxes emanating from interactions between the surface and the atmosphere. These uncertainties may come from both the specification of static physiographic fields and the analysis of prognostic surface parameters in the initial conditions. The method used to apply the surface perturbations is taken from Bouttier et al. (2016). For clarity, a brief explanation of the methodology follows, with key differences to Bouttier et al. (2016) highlighted.

The perturbations are applied to parameters in the SURFEX (see section 2) analysis after the surface data assimilation is completed and remain fixed throughout the forecast for static parameters. For prognostic parameters (i.e., soil temperature and soil moistures), the forecasts begin from the perturbed state and are then allowed to adjust dynamically to the model atmospheric forcing. For each ensemble member and parameter, an independent field of white noise is generated. A set of random seeds (one for each parameter) is generated for each ensemble member from a combination of the forecast analysis time and the ensemble member number. Using the forecast analysis time rather than system

time ensures reproducibility. A recursive Gaussian filter is applied to the white noise until a prescribed correlation length scale is reached. In experiments done for a 3-week period spanning July/August 2015 and a 3-week period spanning December 2015/January 2016 (not shown), it was found that a correlation length scale of ~150 km gave optimum results, compared to the ~400 km used by Bouttier et al. (2016). The spatially correlated random noise field is then clipped to the range ±2 and scaled depending on the parameter. A further clipping is applied after the scaling to ensure that the perturbed fields remain within realistic ranges.

The scaling of the random patterns is chosen, following Bouttier et al. (2016), such that the standard deviations of the perturbations are approximately equal to the precision with which the parameters are known. For sea surface temperature (SST), it was found that smaller perturbations than those used in Bouttier et al. (2016) were more realistic for the MEPS domain. This scaling is either additive or multiplicative depending on the parameter. Table 1 shows the standard deviation and type of scaling applied for each of the perturbed parameters. For the soil temperature and moisture, the uppermost two (of three) layers of the soil are perturbed and perturbations to the sea surface fluxes are made to simulate perturbations to the roughness length over the sea.

The impact of the surface perturbations is assessed for a 3-week period during the summer of 2017 for the MEPS domain (see Fig. 1) by comparing two experiments with

TABLE 1. The magnitude and type of perturbation applied to the surface parameters. For type, × means that the perturbations are multiplicative and + means that the perturbations are additive.

Parameter	Standard deviation	Type
Vegetation fraction	0.1	×
Leaf area index	0.1	×
Thermal coefficient of vegetation	0.1	×
Surface roughness length over land	0.2	×
Albedo	0.1	×
Sea surface temperature	0.25	+
Soil temperature	1.5	+
Soil moisture	0.1	×
Snow depth	0.5	×
Surface fluxes over sea	0.2	×

10 perturbed members. First, the reference experiment, REF, that includes perturbations to the boundary and initial conditions using the SLAF method and PertAna [see section 4c(1)]; and SFCPERT, which is the same as REF, but includes all of the surface perturbations described in Table 1. The verification is done for parameters that might be expected to be affected by surface influences, T2m, RH2m, S10m, and 12-h accumulated precipitation (AccPcp12h). The 12-h accumulated precipitation was chosen for a number of reasons. First, the largest number of reliable stations observing precipitation was available for 12-h accumulations. Second, longer accumulation times are likely to reduce the double penalty problem. Finally, observations were available at 0600 and 1800 UTC separating the precipitation into daytime and nighttime components.

The ensemble spread and RMSE are shown in Fig. 4. For T2m, S10m, and RH2m, the addition of surface perturbations has a statistically significant positive impact on the ensemble spread without any negative impact on the ensemble RMS errors. For AccPcp12h SFCPERT shows a small decrease in spread compared with REF that is accompanied by a decrease of similar magnitude in the RMSE. However, these differences are not significant at the 95% confidence level, so we do not investigate them further here.

The improvements in spread due to the surface perturbations are further confirmed by rank histograms shown in Fig. 5. For both REF and SFCPERT the rank histograms are U shaped for all parameters indicating that the ensembles are underdispersive. For AccPcp12h
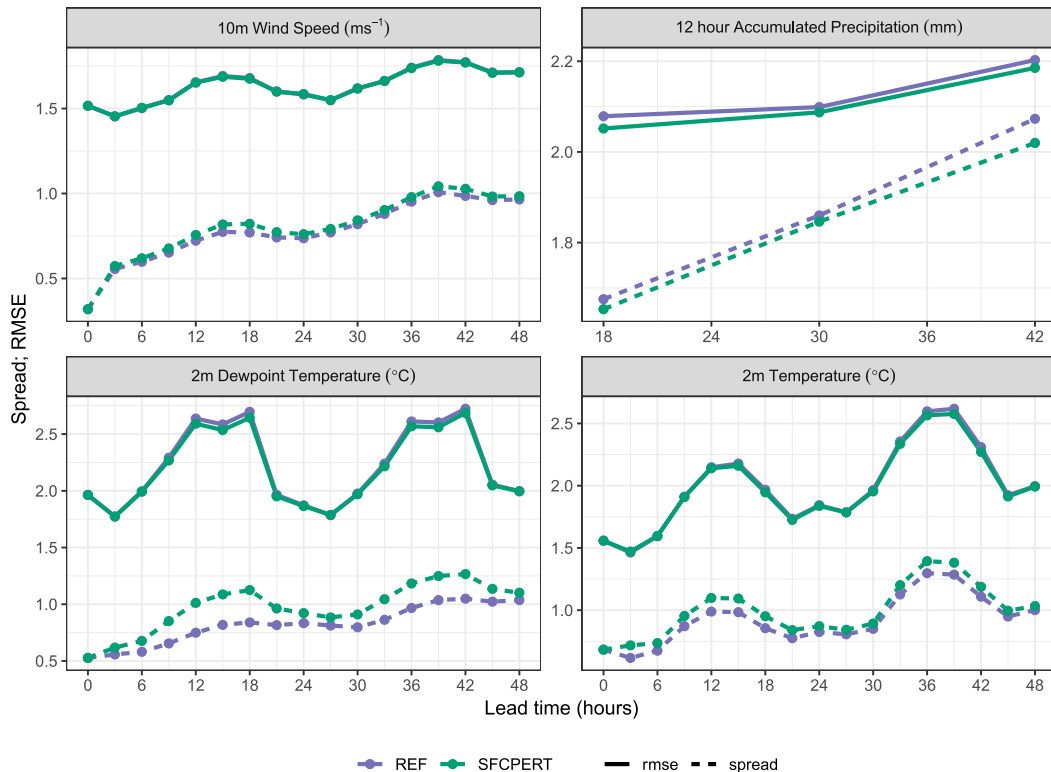

FIG. 4. RMS errors of the ensemble mean (solid line) and ensemble spread (dashed line) for the REF (purple) and SFCPERT (green) experiments for (top left) S10m at 743 stations, (top right) AccPcp12h at 557 stations, (bottom left) Td2m at 828 stations, and (bottom right) T2m at 791 stations. See text for a description of the experiments.
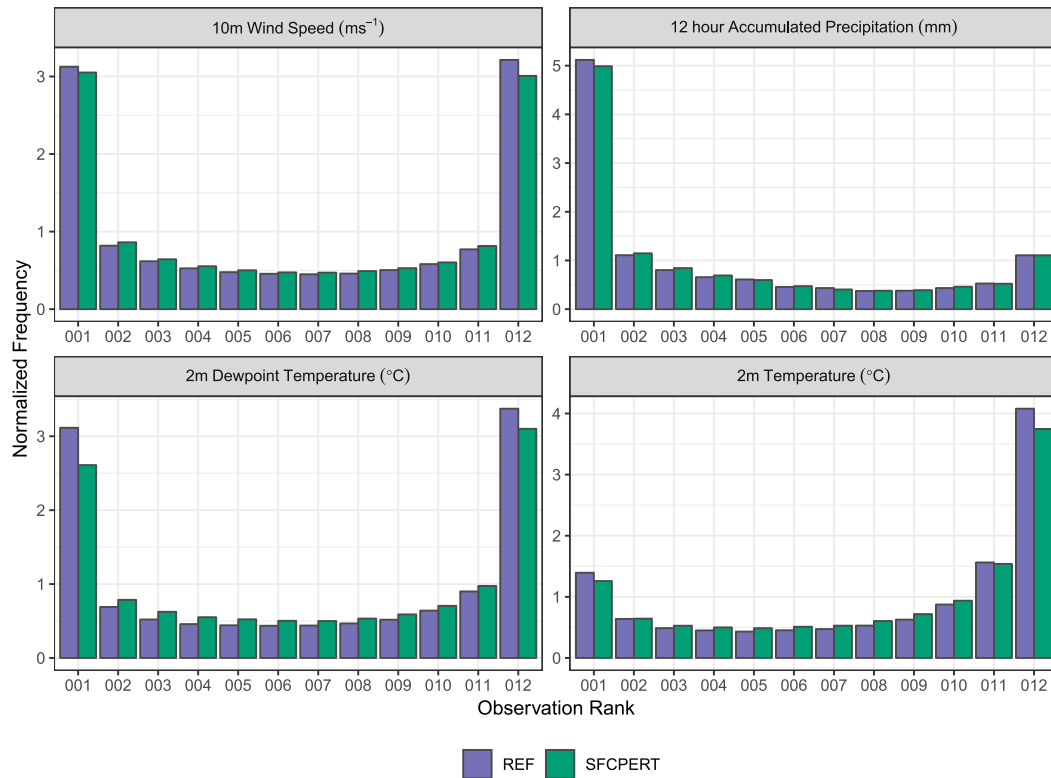
FIG. 5. Rank histograms for the REF (purple) and SFCPERT (green) experiments for (top left) S10m at 743 stations, (top right) AccPcp12h at 557 stations, (bottom left) Td2m at 773 stations, and (bottom right) T2m at 791 stations. See text for a description of the experiments.

there is an indication of a positive bias with the largest proportion of observations being ranked below all of the ensemble members. For T2m, there is an indication of a negative bias with the largest proportion of observations being ranked above all of the ensemble members. However, for all parameters the number of observations being ranked as outliers from the ensemble is smaller for SFCPERT than for REF.

Rank histograms are particularly sensitive to observation errors. Such errors can be taken into account by perturbing all of the ensemble members with an estimate of the observation error sampled from the error distribution for the observation in question (Hamill 2001). The standard deviation of the observation errors are estimated in the surface data assimilation for T2m and Td2m to be ~1 K. For S10m, a value of $1 \, \mathrm{m \, s^{-1}}$ is estimated, while for AccPcp12h a value of $0.2 + 0.2 \times$ AccPcp12h is used as in Bouttier et al. (2016). The rank histograms now tell a slightly different story (Fig. 6)— S10m forecasts are more evenly dispersed, while the negative bias in T2m is more obvious throughout the ensemble members. Td2m shows more a better dispersed ensemble than originally indicated though there remains a large number of observations that have values

smaller than the ensemble minimum. The rank histogram for AccPcp12h suggests that the positive bias is stronger than indicated before observation errors were taken into account. However, the fact that SFCPERT remains more dispersive than REF in Fig. 6 suggests that taking observational errors into account is not so important when assessing the relative performance of different model configurations in this case. Similar tests were done for all experiments verified in this paper and it was found that the conclusions about the relative performance of different model configurations were not affected. Since we do not have consistent and reliable estimates for observation errors for the different model domains and seasons used in this paper, it was decided not to include observation error estimates in the verification scores.

The biases shown in the rank histograms in Fig. 6 suggest that the bias is a significant component of the RMSE. It is not the role of the EPS to account for systematic biases in the numerical weather prediction (NWP) model and so it is expected that the spread–RMSE relationships shown in Fig. 4 would be improved if the ensembles were calibrated to at least make some correction for systematic biases.
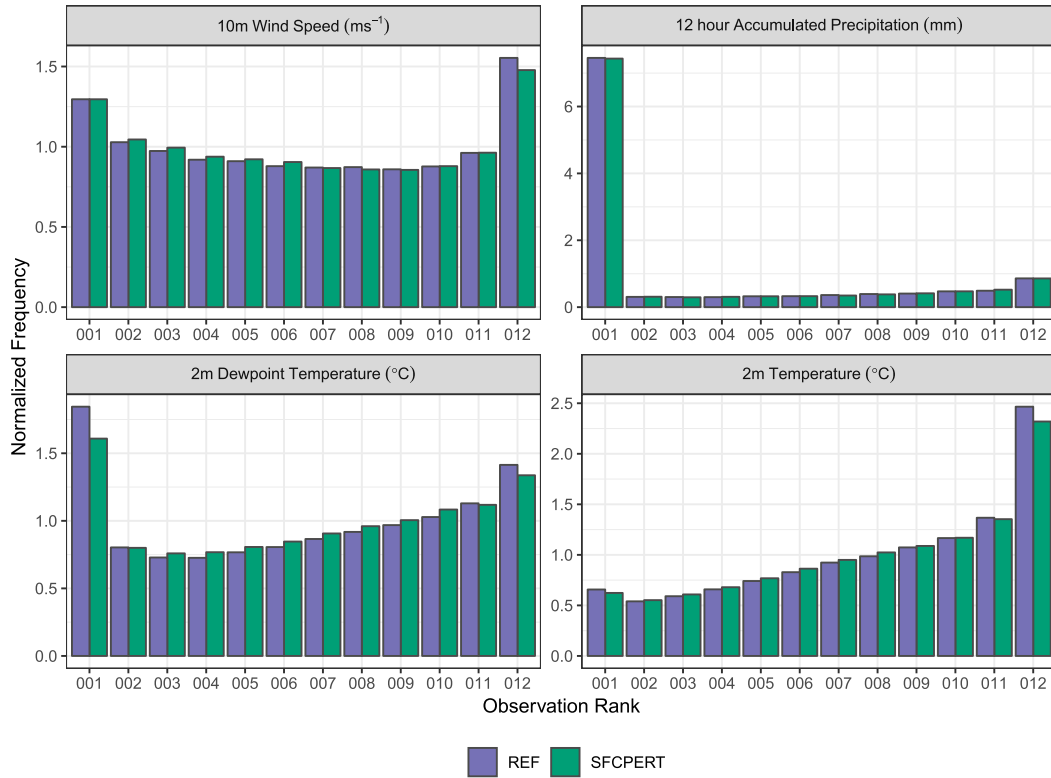
FIG. 6. As in Fig. 5, but with observation errors taken into account.

The surface perturbation scheme has been shown to result in a statistically significant increase in ensemble spread for near-surface parameters without having a statistically significant impact on the RMSE of the forecast. The representation of uncertainty for these parameters is therefore improved in HarmonEPS by the introduction of surface perturbations.

c. Initial conditions for upper air

There are several ways to construct initial states for an EPS. All of them try to sample the initial error, that is, the difference between the best initial state and the truth, which is always unknown. In this section the available generation techniques within the HarmonEPS system, at the moment of writing, are listed.

1) PERTURBATIONS FROM NESTING MODEL—PERTANA

The default perturbation strategy for upper-air fields at initial time is to add perturbations from the nesting model using the corresponding (interpolated) boundary file at initial time to the HarmonEPS control member. In the case of IFS ENS, SLAF, or random perturbations as boundary strategy, this is the natural extension of those methods to the analysis time. When using IFS ENS the perturbations are simply the difference between the

corresponding IFS ENS members and IFS ENS control. The perturbation can then be scaled. For SLAF and the random perturbations the differences between the two IFS HRES forecasts used for the LBC are used to generate the initial perturbation:

$$\text{IC\_m} = \text{A\_c} + \text{K\_m} \times (\text{IFS\_N} - \text{IFS\_N-6}), \quad (2)$$

where IC_m is the initial condition for member m, A_c is the HarmonEPS control analysis, K_m a scaling factor for member m, IFS_N is a HRES forecast with length N and IFS_N-6 is a 6-h shorter HRES forecast, both valid at the same time as the analysis. K_m is set so the members have a similar perturbation magnitude. In the experiments described in this paper the absolute value of K_m ranges from 1.6 for the smallest N to 0.86 for the largest N.

In the sections to follow, this default perturbation strategy is called "PertAna". The influence of PertAna has been tested in many HarmonEPS configurations and has been shown to be important in improving scores. An example is shown later in this paper in section 4c(3) (Fig. 8).

2) SURFACE ASSIMILATION FOR PERTURBED MEMBERS

In the standard setup HarmonEPS runs upper air assimilation only for the control member. However, the
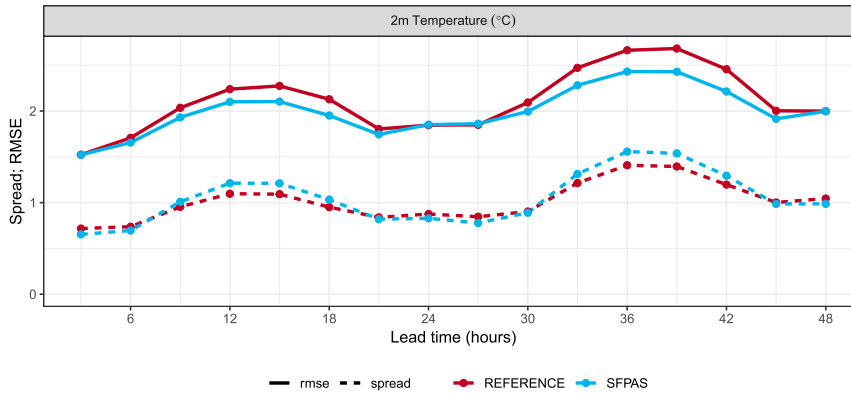
FIG. 7. Spread (dashed) and RMSE (solid) for T2m at 791 stations. REFERENCE shown in red and SFPAS in light blue.

corresponding surface assimilation is applied to each member separately but with identical observations. Each member has different LBCs and model error is represented in the background forecast through the surface perturbations. As the assimilation acts to keep the initial state closer to the observations one would possibly expect that this would hamper the evolution of the spread in the ensemble. This is not always the case as can be seen in Fig. 7. It shows a 3-week period in 2017 from 25 May to 15 June over the MEPS area, 10 + 1 members, where surface assimilation for the perturbed members has been switched on (SFPAS) and off (REFERENCE). The surface assimilation naturally gives a better RMSE. Each member is allowed to develop its own surface state around which perturbations are applied. This in turn gives higher spread compared to the nonassimilation case where perturbations are applied around the same state. Surface assimilation for perturbed members is recommended, and is done by default in HarmonEPS.

3) EDA

The ensemble of data assimilation (EDA) for the HARMONIE–AROME system was originally developed for the estimation of short-term forecasts (also called background) error covariance matrices needed for the variational data assimilation system. In the upper-air part of the data assimilation, the observations are perturbed in a similar way to what is described in Isaksen et al. (2007). For the surface data assimilation, the perturbation of 2-m temperature and 2-m humidity is done separately but using the same technique. The use of conventional observations [synop, ship, buoy, airplane (airep), radiosonde (temp), profilers, ...] as well as radiances (AMSU-A, AMSU-B/MHS, IASI) was implemented in the HARMONIE–AROME EDA system. HarmonEPS EDA is set up to run one 3DVAR

analysis with perturbed observations per member at the same resolution. The ensemble members then start directly from each EDA member. It is possible to inflate the EDA perturbations, but that has not been done in this study.

The importance of including PertAna in the EDA experiments is seen in Fig. 8, with CRPS for S10m and dewpoint at 850 hPa (Td850) as examples. The comparison with/without PertAna was done with 10 + 1 members and run over the MEPS area for 17 days in spring 2016. The score differences for S10m are statistically significant at the 95% level. However, for Td850 the sample is small due to the low number of available upper-air observations, and the score differences are significant for 12 h at the 95% level, for 0 h at the 90% level, and for 24 and 36 h at the 80% level. The spread for upper-air parameters is somewhat too large with both EDA and PertAna (not shown). The two methods for initial state perturbations should be tuned together to get the best balance, by, for example, reducing the amplitude of PertAna when introducing EDA.

A comparison between the reference HarmonEPS setup and different EDA approaches, all with 10 + 1 members, is now presented. The test period was 17 days in spring 2016 and the area was the MEPS domain seen in Fig. 1. The reference run (REF) was a HarmonEPS setup using SLAF (section 4a), PertAna [section 4c(1)], and the surface perturbation scheme (section 4b). In experiment EDA_surfobs each member runs its own analysis with perturbed observations both for the upper air and the surface. The surface perturbation scheme is switched off, otherwise it is as REF. Experiment EDA_surfpert is as the EDA_surfobs experiment, but instead of perturbing the observations in the surface analysis, we use the surface perturbation scheme, as in REF. No model uncertainty is included here, except through the surface perturbations. In Fig. 9 the response
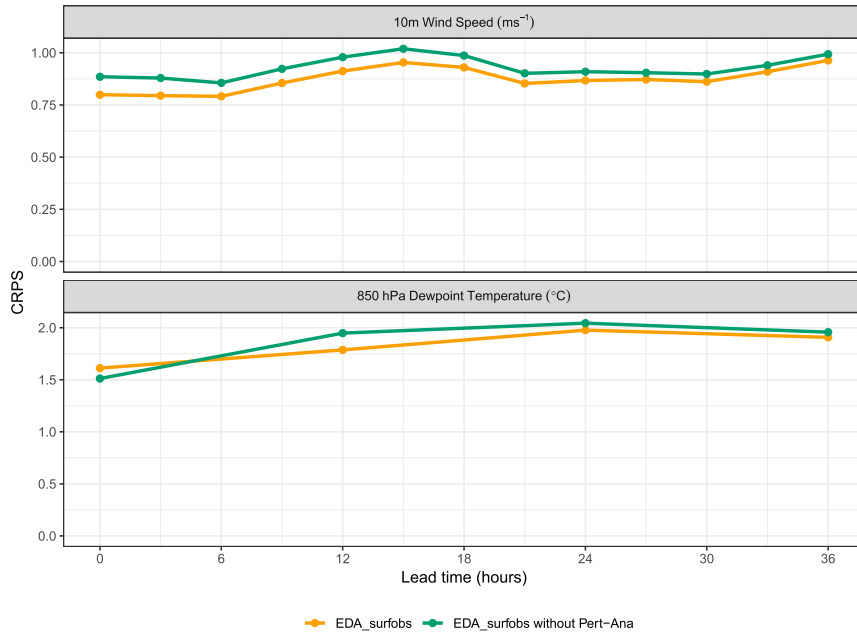
FIG. 8. CRPS for (top) S10m at 637 stations and (bottom) Td850 at 21 stations. Shown are EDA (orange) and the same as EDA, but without PertAna (green).

of the perturbations on T2m is shown for all three experiments, for one randomly chosen day and at initial time ($+0\,$h). While the perturbations are somewhat larger in amplitude for the two EDA experiments, the most striking difference is that EDA introduces more evenly distributed perturbations throughout the whole integration area. More perturbations are especially seen above the sea, probably due to perturbations of the radiances. The spatial scales are qualitatively the same, also for higher levels and parameters (not shown), hence EDA does not introduce finer-scale perturbations, at least not with the same amplitude. This is to be expected because

thinning of the observations is applied in the data assimilation process.

In Fig. 10 spread and RMSE are shown for REF, EDA_surfobs, and EDA_surfpert for S10m, T2m, S10m, Td850, and MSLP. The RMSE does not change much between the experiments, but there is a small tendency for EDA_surfobs to have larger RMSE than REF and EDA_surfpert. For MSLP both EDA experiments have larger RMSE than REF. It is clear that EDA_surfobs increases the initial spread, and gives a better spread to RMSE relationship for the near-surface parameters, but only for the first few hours. For the
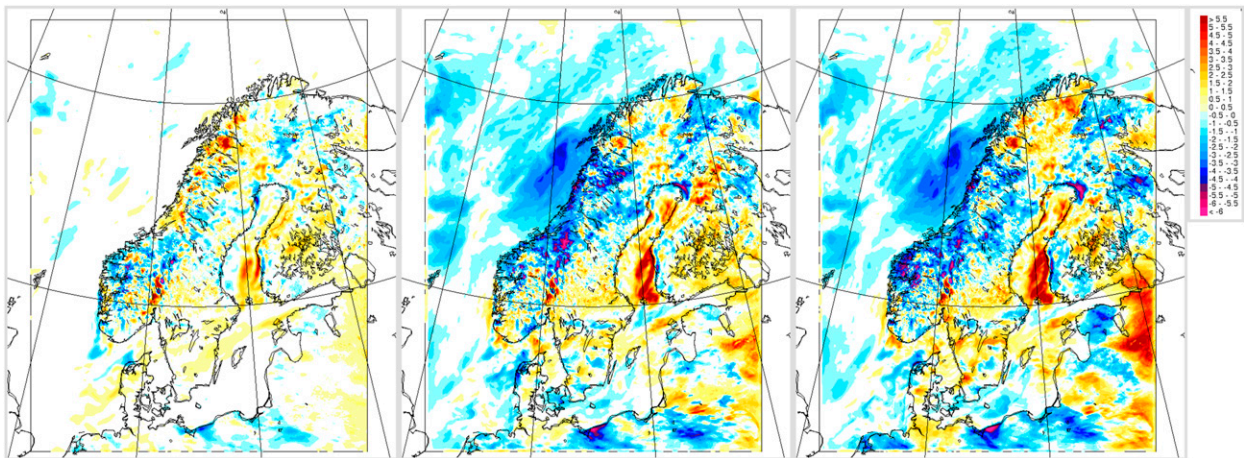

FIG. 9. Example of perturbation size for one randomly chosen day, for T2m for (left) REF, (center) EDA_surfobs, and (right) EDA_surfpert.
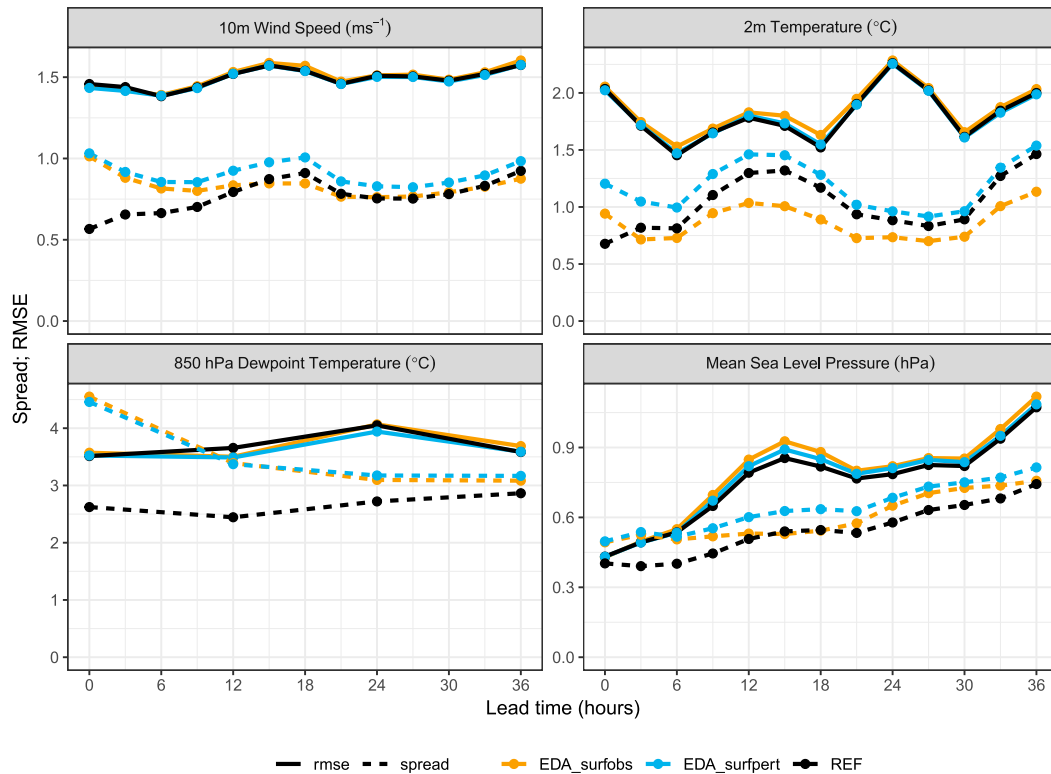
FIG. 10. Spread (dashed) and RMSE (solid), for REF (black), EDA_surfobs (orange), and EDA_surfpert (blue). (top left) S10m at 637 stations, (top right) T2m at 706 stations, (bottom left) Td850 at 21 stations, and (bottom right) MSLP at 491 stations.

upper-air parameters (here only Td850 is shown) the initial spread is too big, hence there is a need for tuning of EDA together with PertAna. EDA_surfpert has the best spread to RMSE relationship throughout the forecast range, due to increased spread, except that the spread is too large initially for MSLP and upper-air parameters. For CRPS (not shown) EDA_surfpert is better than or as good as REF for all parameters, both surface and upper air. Clearly EDA_surfobs does not verify as well as EDA_surfpert. There are some unwanted features seen in Fig. 10: somewhat increased RMSE for the two EDA experiments as well as a too large initial spread for some parameters. This has been investigated in another study (not shown), and reducing PertAna to 0.5 gave, as would be expected, less spread than with PertAna set to 1, but still significantly larger spread than without PertAna. There was no increase in the RMSE as was seen for, for example, MSLP in Fig. 10. This highlights the importance of testing and tuning the perturbation methods together.

The reason that the surface perturbation scheme performs better than when perturbing the observations in the surface analysis can be due to the limited number of observations perturbed at the surface. For EDA we only perturb the observations of T2m and RH2m. For the surface perturbation scheme, however, we perturb many more parameters (see section 4b, Table 1). Although they are kept constant throughout the forecast, except for the prognostic variables, which are freely evolving, they are different for different members. Another difference is that the surface perturbations are applied after the analysis, while obviously the perturbation of surface observations is done beforehand. Looking at the difference between member 1 and the control member for T2m (Fig. 9) it can be seen that the perturbation size and spatial scales are similar between EDA_pertobs and EDA_surfobs, so that cannot explain why an impact of the surface perturbation scheme is seen throughout the forecast range, and not when perturbing the surface observations. It is more interesting to look at a parameter with longer memory, like deep soil temperature (TG2). In Fig. 11 the standard deviation of the difference between member 1 and the control member for the whole forecast range is shown for TG2 for one random date (0000 UTC 1 June 2016). A much larger initial perturbation for EDA_surfpert is clearly seen. The perturbations also have larger scales initially (not shown). While the difference decreases
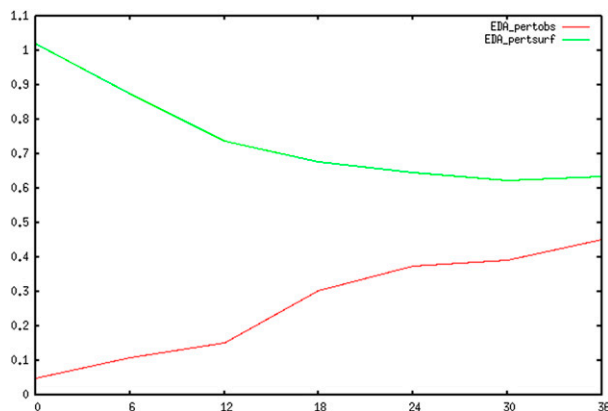
FIG. 11. Standard deviation of member 1 − control for TG2 for 0000 UTC 1 Jun 2016, with EDA_surfobs in red and EDA_surfpert in green.

with time for EDA_surfpert, it slightly increases for EDA_surfobs, but it is still larger at the end of the forecast range. It is likely that the EDA_surfpert perturbations are too big for the model to maintain, and that is the reason why it decreases. It could be that smaller perturbations would also grow, as for EDA_surfobs. Including EDA perturbations in the upper air in combination with the surface perturbations is clearly beneficial for most parameters. The combination of EDA_surfobs and EDA_surfpert where the observations are perturbed, including other perturbed parameters from the surface perturbation scheme, did not give any improvement in the scores over EDA_surfpert (not shown). This is probably due to the much larger initial perturbations coming from the surface perturbation scheme (see Fig. 11). A tuning of the initial size of the surface observation perturbations, together with perturbing more parameters like SST, may lead to better performance for EDA_surfobs. This will be looked into in a further study.

4) LETKF

Besides variational methods like 3DVAR (Anderson et al. 1998), 4DVAR (Courtier et al. 1994), and EDA (Isaksen et al. 2010), EnKF algorithms appear as an alternative to perform atmospheric analysis. EnKF algorithms are used operationally in some NWP centers like Deutscher Wetterdienst (DWD; Schraff et al. 2016), the Canadian Meteorological Centre (CMC; Houtekamer et al. 2005), or the National Centre for Environmental Prediction (NCEP; Pan et al. 2014). Within the family of EnKF algorithms, the local ensemble transform Kalman filter (LETKF) stands out mainly due to its high computational efficiency because it performs an analysis independently at each grid point, so it is highly parallelizable. In particular Schraff et al. (2016) describe the

implementation of LETKF to perform data assimilation for its high-resolution convective-permitting operational forecast. To have full details about the algorithm, the reader is referred to Hunt et al. (2007). ECMWF decided to code an EnKF system based on the IFS model to have an alternative data assimilation system that allowed comparisons to its 4DVAR-EDA operational system. Detailed information on the technical implementation of EnKF at ECMWF and its performance can be found in Hamrud et al. (2015) and Bonavita et al. (2015). IFS EnKF contains two EnKF algorithms, LETKF (Hunt et al. 2007) and EnSRF (Whitaker and Hamill 2002). From these, LETKF is used to perform analysis in model space and EnSRF in observation space in order to look at innovation statistics. The IFS EnKF code has been ported to the HARMONIE system and is available to perform high-resolution deterministic and probabilistic analyses and forecasts. For simplicity, from now on in this paper the acronym LETKF will be used to refer to the HARMONIE EnKF system.

Here a 10-day evaluation (7 September 2018–16 September 2018, 0000 and 1200 UTC runs) of the probabilistic forecasting performance of LETKF is presented, taking as reference the method to sample initial uncertainties in HarmonEPS, PertAna [section 4c(1)]. In all the corresponding figures the LETKF experiment has the tag "LETKF" while the PertAna experiment is labeled "PERTANA." The main characteristics of the LETKF runs are: 10 ensemble members, $0.2 \log(Ps\,P^{-1})$ for vertical localization (where Ps is the surface pressure and P is the pressure), 200 km of horizontal localization, and both additive and multiplicative inflations. The experiments are run over the Iberian Peninsula (see Fig. 1, green domain) at 2.5-km horizontal resolution. The SLAF methodology is used to create the boundary conditions (section 4a).

In this experiment only conventional observations are assimilated (synop, ships, buoys, airep, and temp) as a first test. In particular T2m and Rh2m observations are assimilated in both LETKF and PertAna upper-air analyses. One week of spinup starting on 1 September 2018 has been done. The period of study is slightly unstable from a meteorological point of view, with Mediterranean convective situations typical of the end of summer starting to appear. The forecast range for the ensemble to be evaluated is 36 h, and the analysis cycling is 3 h. A clean comparison between LETKF and PertAna has been carried out. This means that the only difference in both systems is the way in which the initial perturbations of the upper air are constructed. Verification is done against conventional observations, that is, synop, ships, and buoys on the surface and temp in the vertical.

Figure 12 shows the spread and RMSE for surface parameters Td2m and T2m. In PertAna, like in LETKF,
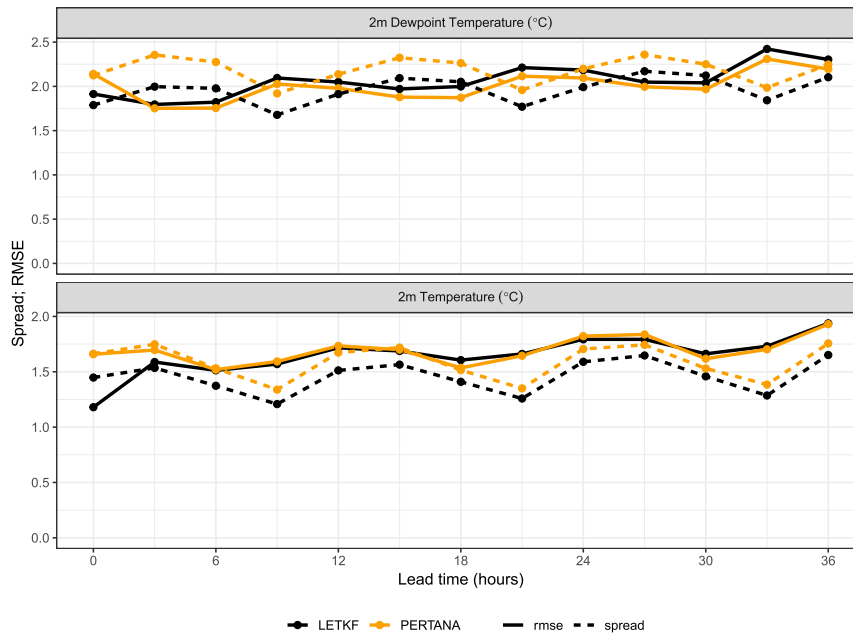
FIG. 12. Spread and RMSE of 36-h ensemble forecasts for surface parameters (top) Td2m at 169 stations and (bottom) T2m at 169 stations. Black lines are for LETKF and yellow lines for PertAna. Continuous lines are for RMSE and dashed for spread.

stochastic perturbations of surface fields after surface analysis (see section 4b) in each member are applied. It is worth noting that at analysis time the mean error in both parameters is lower for LETKF and that this error tends to increase more uniformly with lead time in the first hours than for PertAna. This behavior could be an indication of the more accurate construction of the initial states in the case of LETKF. At time range hour +3 LETKF also has lower error for T2m, with similar error to PertAna at time ranges thereafter. For Td PertAna is slightly better in the whole forecast range. Looking at spread, it is always larger in PertAna, perhaps somewhat too large. Nonetheless the stochastic perturbation of surface fields does a good job of introducing realistic spread, giving a quite balanced amount of spread and error for surface parameters, compared for example with upper-air parameters (Fig. 14).

An interesting feature of flow-dependent methods to estimate background error (like LETKF or EDA) is documented in several studies (e.g., Pu et al. 2013; Ha and Snyder 2014). The hypothesis is that the flow dependency in the background error term for LETKF (and EnKF in general) produces more realistic variances and so this results in more realistic analysis increments over complex terrains compared to variational methods where the background error is climatological and hence constant in time (not in space). Compared to other observations, surface observations have high spatial and temporal resolution, so better representation

of background errors in this case could result in an improvement of the analysis.

To better understand this hypothesis, a comparison of analysis increments of 3DVAR (in PertAna core) and LETKF is presented in Fig. 13, for temperature at model level 65 (between 10 and 15 m above the surface). The purpose of these figures is to show one of the clear differences between flow-dependent (here LETKF) and non-flow-dependent analysis methods (here 3DVAR), that is, the spatial structure of analysis increments. In the case of 3DVAR, due to the climatological definition of background error, the resulting analysis increments are quite uniform and smooth, and their spatial structure does not have a relationship with the coastlines or orographic features. On the other hand, LETKF analysis increments are clearly nonhomogeneous, neither uniform or smooth, and the spatial structure reflects to some extent the orographic features. In particular, in black circles the basins of two of the largest rivers in Spain, the Ebro and Guadalquivir Rivers, are shown. One could expect to some extent that weather patterns in those basins could be similar, leading to a spatial correlation of background errors and eventually resulting in similar analysis increments.

The spread and RMSE for several upper atmospheric parameters (or related) are plotted in Fig. 14: 700-hPa dewpoint temperature (Td700), temperature at 850 hPa (T850), 925-hPa wind speed (S925), and total cloud cover (TCC). These plots are a sample of what is
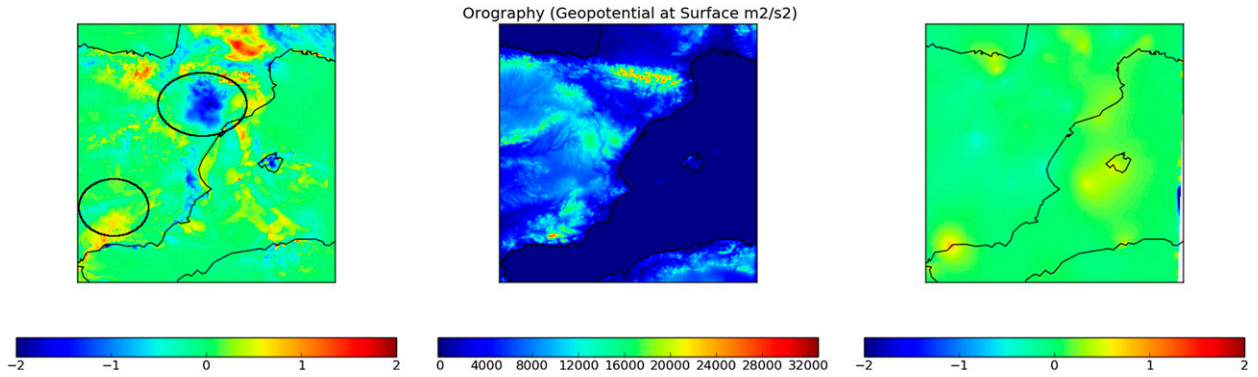
FIG. 13. Analysis increments of temperature (°C) at model level 65 at analysis cycle 1200 UTC 15 Apr 2017 for (left) LETKF and (right) 3DVAR. (center) Orography (geopotential at surface; $m^2 s^{-2}$) shown for reference. As an example of orographic features, black circles in the LETKF increment show the river Ebro and Guadalquivir basins.

happening in the vertical. Three basic results can be highlighted by looking at the vertical. First, LETKF seems to be clearly better than PertAna in terms of mean error for the humidity field Td700 and slightly better for TCC, at all forecast ranges. When looking at T850 and the S925 wind field the impact seems to be variable to neutral. This result would indicate that

LETKF could have a clear positive impact in humidity-related fields.

The second result has also been shown in Fig. 12, that is, the ability of LETKF to construct more accurate initial states. The measure of this more accurate construction is in terms of lower mean error and more realistic (increasing) evolution of this error in the first
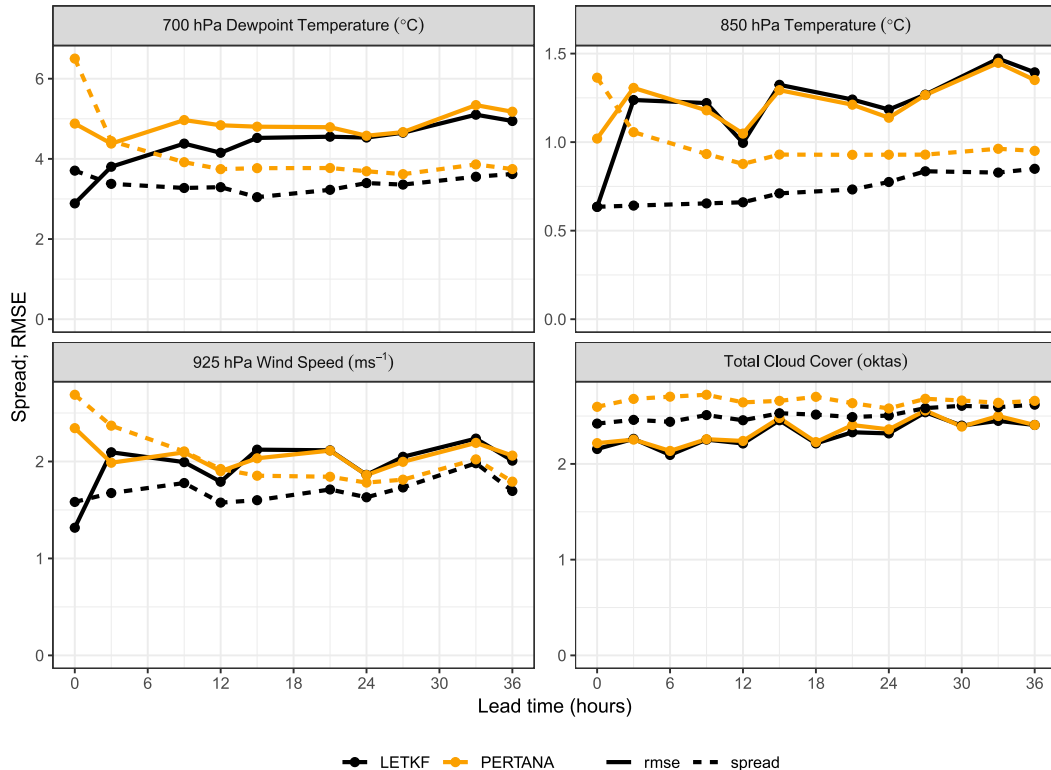

FIG. 14. Spread and RMSE of 36-h ensemble forecasts for upper-atmospheric parameters (and related) (top left) Td700 at 8 stations, (top right) T850 at 8 stations, (bottom left) S925 at 7 stations, and (bottom right) TCC at 151 stations. Black lines are for LETKF and yellow lines for PertAna. Continuous lines are for the RMSE of the ensemble mean and dashed for spread.

forecast hours, compared to PertAna. In particular the nondesired jumps in the 0–3-h range for Td700 and S925 from PertAna are totally eliminated in LETKF.

Finally, the third result is that although the amount of spread is still not realistic enough, PertAna is always able to produce more spread, which translates in general into better spread to RMSE relationships (except for TCC). Adding a flow-dependent perturbation coming from the boundary conditions to the initial state, as PertAna does [section 4c(1), Eq. (2)], seems to have a positive impact in the vertical, increasing the spread while maintaining the mean error. This result for PertAna is something to test in LETKF too. On the other hand, the fact that the spread in PertAna is decreasing with lead time is an indication that this boundary-dependent perturbation is too large at the initial time, and should be reduced.

In this simplified study it has been shown that LETKF produces more skillful probabilistic predictions than PertAna for upper-air humidity fields like Td700 and TCC, while having a neutral impact for T850 and wind at 925 hPa. It seems that the construction of the initial states via LETKF is more accurate than the PertAna approach, both in terms of mean analysis error at initial time and evolution of this error. For surface parameters the ability of the algorithms varies, showing LETKF overperformance in T2m and Td2m overperformance in PertAna. What seems to be clear is that probabilistic prediction in the surface is strongly influenced by stochastic perturbations of surface fields (section 4b), which creates quite balanced spread error relationships in both cases.

Although further investigation is needed, particularly testing the algorithm in an operational fully observing system, these results make LETKF a promising candidate as an initial states generator for probabilistic forecasting with HARMONIE.

5) BRAND

BRAND perturbations are based on the randomization of the climatological background error covariance matrix. This is an alternative scheme available in the HARMONIE–AROME forecasting system for the generation of the initial condition perturbations. A similar approach for perturbing initial conditions has been applied in Raynaud and Bouttier (2016). BRAND perturbations $\mathbf{x}_i$, $i = 1, \ldots, N_{\mathrm{ens}}$, are generated as standard Gaussian random numbers $N(0, 1)$ in the entire vector control space and are transformed to the physical model space through the square root of the climatological background error covariance:

$$\mathbf{x}_i = \sqrt{\mathbf{B}}\boldsymbol{\eta}_i, \quad \boldsymbol{\eta}_i \sim N(0,1). \qquad (3)$$

The model for the climatological background error covariance $\mathbf{B}$ is the same as the one used in variational data assimilation to form the analysis increments from observations (Berre 2000). The background error covariance is formulated in spectral space assuming homogeneity and isotropy of climatological statistics. A BRAND perturbation is generated as follows: A random vector of the size of the entire control vector space $\boldsymbol{\eta}_i$ is sampled from a standard Gaussian distribution, $\boldsymbol{\eta}_i \sim N(0,1)$. Then random spectral components corresponding to a particular 1D wavelength in spectral space are first transformed to impose vertical and horizontal correlation structures and then a separate per wavelength balance operator is applied. Finally an inverse 2D Fourier transform projects the perturbations to the gridpoint space. The obtained perturbation is relaxed toward a large-scale perturbation on the lateral boundaries. The formulation and the properties of the climatological background error covariance used in the HARMONIE–AROME system are extensively discussed in Bojarova and Gustafsson (2019).

BRAND ensemble members $\{\mathbf{X}_i, i = 1, \ldots, N_{\mathrm{ens}}\}$ can be generated in two different modes, the deterministic mode and the EPS mode. In the deterministic mode each scaled perturbation $\{\mathbf{x}_i\}$ is added to the same first guess $\mathbf{F}_c$ of the control:

$$\mathbf{X}_i = \mathbf{F}_c + \alpha\mathbf{x}_i, \quad i = 1, \ldots, N_{\mathrm{ens}}. \qquad (4)$$

In the EPS mode the ith perturbation $\{\mathbf{x}_i\}$ is added to the first guess $\mathbf{F}_i$ or to the analysis $\mathbf{A}_i$ of the ith ensemble member depending on the configuration. Under configuration "before DA" the ith scaled perturbation $\mathbf{x}_i$ is added to its own first guess $\mathbf{F}_i$ before the data assimilation is done:

$$\mathbf{X}_i = \mathbf{F}_i + \beta\mathbf{x}_i, \quad i = 1, \ldots, N_{\mathrm{ens}}. \qquad (5)$$

Under the configuration "after DA", the ith scaled perturbation $\mathbf{x}_i$ is added to its own analysis:

$$\mathbf{X}_i = \mathbf{A}_i + \gamma\mathbf{x}_i, \quad i = 1, \ldots, N_{\mathrm{ens}}. \qquad (6)$$

Here α, β, and γ are tunable scalar parameters that control the amplitude of the perturbations. Then a nonlinear forecast model is applied to propagate the ensemble forward in time. Note that the BRAND ensemble in the deterministic mode is centered around the deterministic control at initial time, while the BRAND ensemble in EPS mode is not. The spread of the BRAND ensemble in EPS mode is constrained by assimilating the same observations for all ensemble members. There is a possibility to control how strongly the ensemble members are drawn to observations. Although $\mathbf{x}_i$ perturbations are drawn from the climatological background error statistics, the BRAND
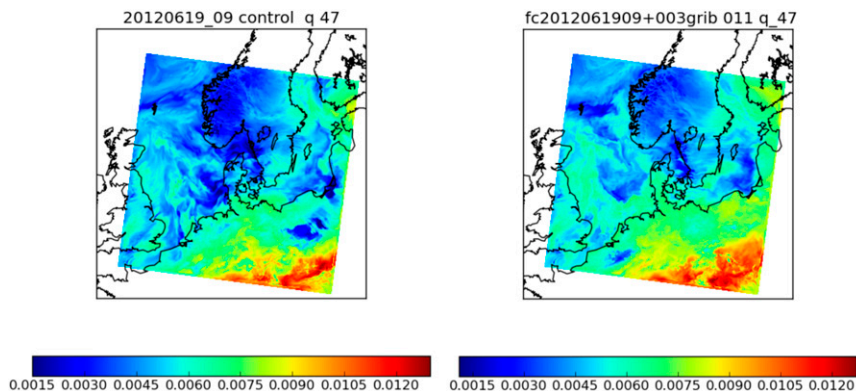
FIG. 15. (left) Control +3-h forecast and (right) the BRAND ensemble member for the specific humidity model field at approximately 850 hPa valid at 1200 UTC 19 Jun 2012.

ensemble in the EPS mode reflects flow dependency and has larger spread in the areas where the evolution is sensitive to the initial conditions. In addition the ensemble spread is reduced in the areas of dense observation coverage because the same observations are assimilated for all ensemble members.

Figure 15 shows the control member and a randomly chosen ensemble member, in this case number 11 from a 20-member BRAND ensemble in EPS mode (configuration "after DA"), for the specific humidity model field at approximately 850 hPa (model level 47 in the HARMONIE–AROME configuration). The fields are +3-h forecasts from a HARMONIE–AROME 2.5-km configuration valid at 1200 UTC 19 June 2012. In Fig. 16 the mean and standard deviation computed from the BRAND ensemble are shown for the same field. One can clearly see a much smoother structure of the mean field (Fig. 16, left) in comparison to the control field (Fig. 15, left), and similarities in structures between the control field and the ensemble member, even if the fields have obvious differences (Fig. 15, left and right). The standard deviation is an inhomogeneous and anisotropic field with a large amplitude in the areas of dynamically active areas. One may also notice a smaller amplitude of the standard deviation in the areas over land where a dense observational network is available. This is an attractive feature of the BRAND initial condition perturbations that makes them sensitive both to the dynamically unstable areas and the density and quality of the observing network.

At this stage it is not clear how EDA, LETKF, and BRAND compare. This will be addressed in a further study.

d. Representation of model uncertainty

Forecast skill and predictability are also affected by model errors. Model error can arise from unresolved

processes at the subgrid scale that need to be parameterized, from simplifications in the process description, from incomplete knowledge of the process itself, or from uncertain parameters, whether the parameters represent a physical quantity or not. Here we present a number of different approaches that are available in HarmonEPS: multiphysics, multimodel, and stochastically perturbed parameterization tendencies (SPPT).

1) MULTIPHYSICS AND MULTIMODEL

When using multiphysics (MP) to account for model uncertainty different parameterization schemes for turbulence, microphysics and radiation available within HarmonEPS are utilized. Each ensemble member in HarmonEPS then has a unique combination of physics parameterization schemes. In the multimodel setup, different ensemble members use different models, often in connection with other representations of uncertainty as the number of available models is limited. In HarmonEPS there is an option to build a multimodel ensemble from HARMONIE–AROME and HARMONIE–ALARO with differences both in physics and dynamics (see section 2). The hypothesis is that for multiphysics/multimodel different configurations and approximations performed in each parameterization/model, developed by different scientists/meteorological centers, contain a valid measure to describe the uncertainty and the errors of the parameterization/model itself. The multimodel approach has been demonstrated to be highly skillful (Beck et al. 2016; Frogner et al. 2016; Smet et al. 2012; Iversen et al. 2011; García-Moya et al. 2011). The reader is referred to section 5 to have a look at how a multimodel verifies.

A 3-week MP experiment was run for a period in summer 2015 (20 July 2015–10 August 2015) and compared to the basic setup of HarmonEPS with the same number of members (8 + 1), the same boundary
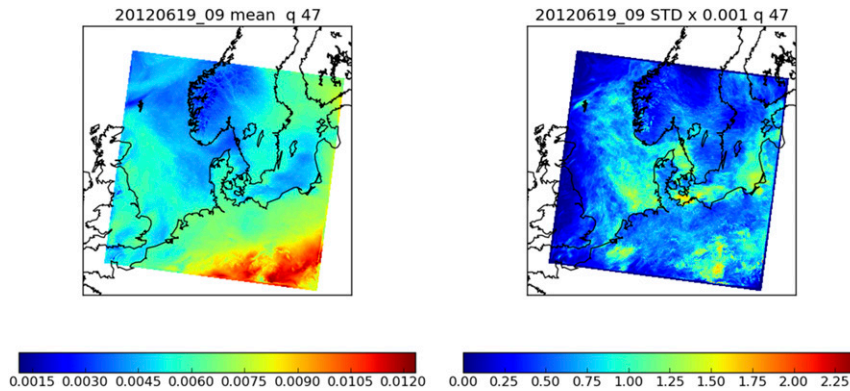
FIG. 16. The (left) mean and (right) standard deviation of the +3-h forecasts of the specific humidity model field at approximately 850 hPa computed from the 20 members of the BRAND ensemble in EPS mode (configuration "after DA"). The fields are valid at 1200 UTC 19 Jun 2012.

conditions and initial conditions and run over the MEPS domain (see Fig. 1). There is a tendency for the RMSE to be lower and the spread to be higher for HarmonEPS that includes multiphysics, but the differences are small and mostly not significant (not shown).

When applying multiphysics it is probable that different members of the ensemble will have different characteristics, or biases, as different parameterizations for the same processes have different biases. As expected this is the case for HarmonEPS with multiphysics (Fig. 17). One clear outlier (member 3) and two members with different evolutions with forecast length (members 1 and 8) can be seen. For other parameters other members are outliers. In the reference experiment the members have the same characteristics. In the same way, for a multimodel setup, the members from the different models can cluster so that the members from the same models have the same characteristics, but different from the members of the other model(s).

In MP and multimodel the improved scores can be due to different biases in the members/models, as seen in Fig. 17. The ensemble spread should ideally not come from different biases, but rather from nonsystematic and random forecast uncertainty. In Frogner et al. (2016), calibration was used on multimodel HarmonEPS, and even after removal of the biases the multimodel still resulted in better scores, indicating that a MP/multimodel approach goes beyond the effects of error cancellation. It is plausible that MP/multimodel is able to treat some uncertainties that a single parameterization/model is not able to, even when including methods for describing model uncertainty, simply because a single parameterization/model might not have the possibility to span all possible physical developments. The inconsistency in the ensemble arising from MP/multimodel, as seen in Fig. 17, can be both a challenge and an advantage: an ensemble where some members are, for example, always predicting more clouds than other members, can be challenging for the users, and the members will not have the same probability. On the other hand, some MP/multimodel members can be better than others in certain situations, giving indications of high impact weather that would otherwise have been missed by a single-physics/model ensemble. An important part of model uncertainty is the uncertainty in the dynamics (Bowler et al. 2008a), which is automatically taken into account when different models are used in a multimodel approach. The main drawback with multimodel/MP is the need to install and maintain several parameterizations/models.

2) SPPT

HarmonEPS has the possibility to account for model errors by the use of the stochastic perturbation of parameterizations tendencies scheme. It was adapted from the ECMWF implementation in the context of AROME-EPS (Bouttier et al. 2012). SPPT was first implemented operationally to represent parameterization uncertainties in the ECMWF's EPS in 1998 (Buizza et al. 1999) and later it was implemented in several global EPSs, for instance by Environment and Climate Change Canada and by the Japan Meteorological Agency, with quite successful performance in all of them (Charron et al. 2010; Separovic et al. 2016). The main practical motivation at that time for ECMWF and other centers was to increase the EPS spread, especially in the medium range, in order to have more reliable EPSs. SPPT also proved to be able to increase the skill by reducing the RMSE of the ensemble mean that has been explained through the concept of nonlinear noise-induced rectification (Palmer et al. 2009).

The basic SPPT is based on perturbing the output of the net physic tendencies with 2D random multiplicative noise in a different way for each ensemble member (see
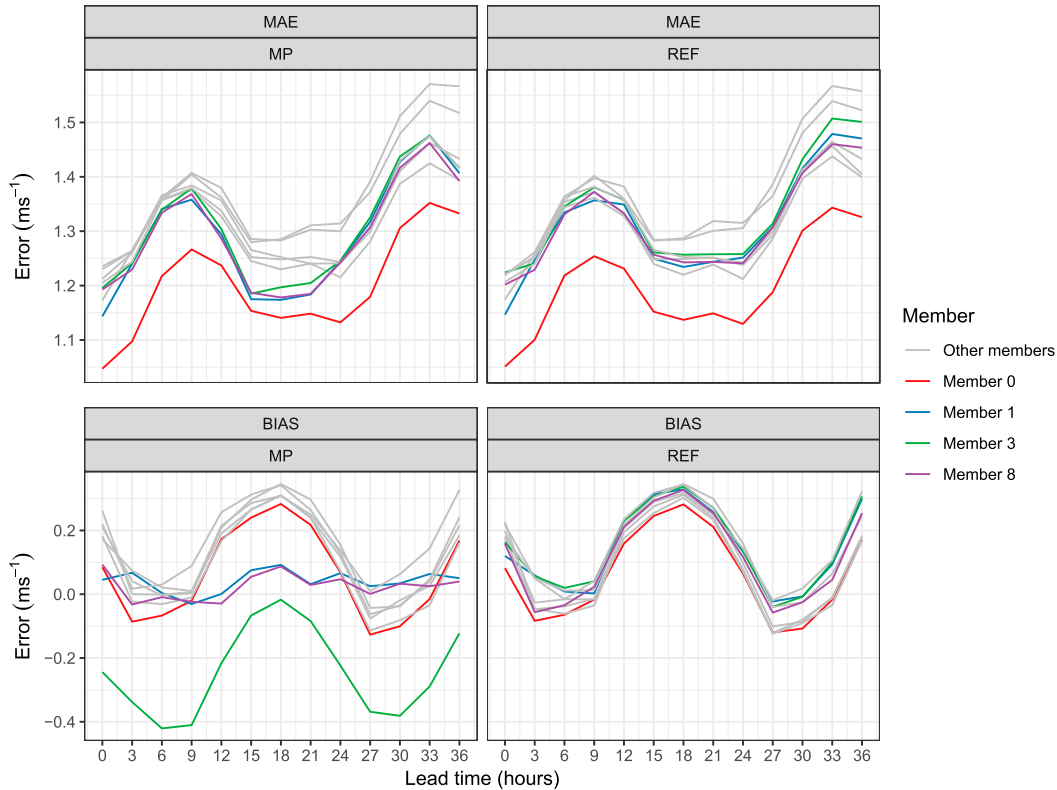
FIG. 17. (top) MAE and (bottom) bias for all members of the (left) MP experiment and (right) REF experiment, for S10m at 781 stations. Members 0, 1, 3, and 8, which are discussed in the text, are highlighted in red, blue, green, and purple, respectively.

details of SPPT in Palmer et al. 2009 and Leutbecher et al. 2017). An optional tapering in the range [0, 1] to avoid perturbations in the planetary boundary layer (PBL) and in the stratosphere can be applied.

A key feature of SPPT and other stochastic parameterizations such as SKEB (Berner et al. 2009; Shutts 2005; Bouttier et al. 2012) is that the perturbations are sampled from a pattern generator with spatial and temporal correlations as for instance shown in Fig. 18. The latter correlations "reinforce," in addition to the dynamics, the relationship of the subgrid realizations of close grid points in space and time. For example, warming/cooling a region, through locally increasing/decreasing temperature tendencies could foster/inhibit convection developments. It could be argued that pattern generators in stochastic parameterizations partially help in an ensemble context to alleviate a structural deficiency of NWP models that are built on the parameterization assumption that a spectral energy gap exists in the scale truncation between grid and subgrid processes, which is not observed in the atmospheric energy spectrum (Palmer 1997).

The three main parameters to be set in SPPT are as follows: σ is the standard deviation of the pattern

generator, L is the horizontal length scale, and τ is the time scale of the decorrelations. Unfortunately with the current SPPT adaptation to LAM-EPS (Bouttier et al. 2012) the spectral pattern generator does not correspond to what is intended by setting σ and L (M. Szűcs 2017, personal communication) and it generates a quite distinct pattern with different spatial correlations than the ones expected. This problem has motivated the development within HarmonEPS of the pattern generator on a bi-Fourier plane (work in progress) instead of the current projection to the plane from the quasi-Gaussian pattern on the sphere used in ECMWF's EPS SPPT, and to implement another pattern generator called the stochastic pattern generator (SPG; Tsyrulnikov and Gayfulin 2017).

A test with SPPT was run for three weeks in spring 2017 (0000 UTC 26 May 2017–15 June 2017) over the MEPS domain and with 10 + 1 members, with two slightly different settings, both with the same horizontal length scale (see Fig. 18) and same time scale (8 h). Experiment SPPT_0.2 had a standard deviation of 0.2 and experiment SPPT_0.33 a standard deviation of 0.33. SPPT_0.2 corresponds to the example pattern to the left in Fig. 18 and SPPT_0.33 to the pattern on the right.
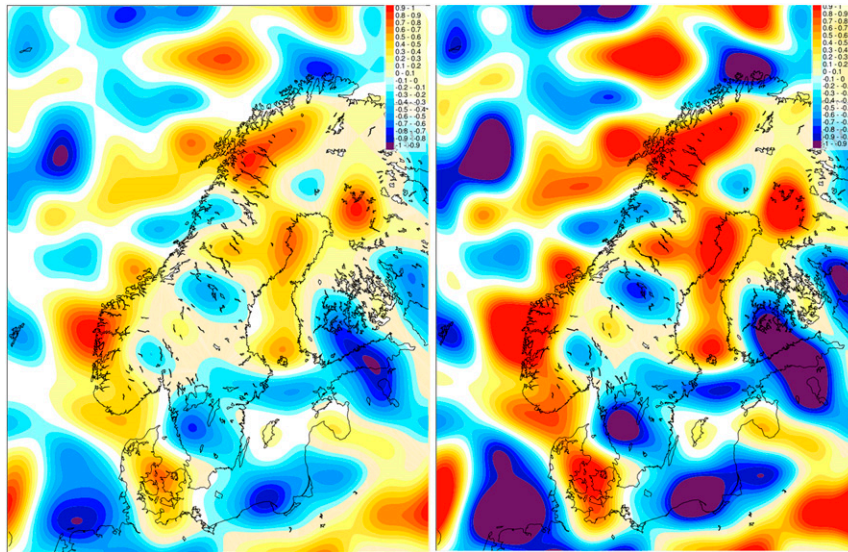
FIG. 18. Example of random patterns used in SPPT in HarmonEPS, with a standard deviation of (left) 0.2 and (right) 0.33.

For comparison we use a reference HarmonEPS run (REF), which is identical to SPPT experiments, except SPPT is not activated. In Fig. 19 we see the spread and RMSE for S10m and Td2m. We see a slight increase in the spread when SPPT is activated, giving a somewhat better spread to RMSE relationship when SPPT is applied, but overall the impact of the current SPPT implementation in HarmonEPS is small (also for parameters not shown). Some suggestions for further improvements of SPPT in HarmonEPS are discussed in section 6.

5. Operational and preoperational implementations of HarmonEPS

Several institutes run implementations of HarmonEPS. Presently all systems run with 2.5-km horizontal grid spacing and 65 vertical levels. At the time of writing four systems are operational and two have preoperational status. The different domains for these six systems are displayed in Fig. 1, and their basic characteristics are in Table 2.

The first system to become operational in November 2016 was the MetCoOp Ensemble Prediction system (MEPS). MEPS is jointly operated by the national meteorological institutes of Finland, Norway, and Sweden within the MetCoOp cooperation. The control member runs a 3-hourly assimilation cycle as described in section 2 and the nine perturbed members run a 6-hourly surface assimilation [section 4c(2)]. Perturbations are generated using the SLAF technique (section 4a), PertAna [section 4c(1)], and surface perturbations (section 4b). [Note that snow depth, leaf area index (LAI),

vegetation, and the vegetation thermal inertia coefficient are not perturbed in MEPS.] The ensemble runs up to +54 h four times a day for the purple domain in Fig. 1. The different members are distributed over three different supercomputers as a way to share resources and achieve good redundancy.

Continuously updated mesoscale EPS (COMEPS) from the Danish Meteorological Institute (DMI) has been operational since June 2017. COMEPS is best described as a multi-HarmonEPS system. To satisfy as many users as possible, while at the same time limit the use of computer resources, half of the members are run on a big domain that includes all of Scandinavia and the Baltic and North Seas, while the other half are run on a smaller domain (see Fig. 1). Subsequently, the members on the big domain are interpolated to the small domain and added to the small domain ensemble. The spatial resolution is the same for the two domains (2.5 km and 65 vertical levels). New members, currently one control and two perturbed, are run for both domains every hour and added to an ensemble that contains not only the latest members, but also lagged members from the previous five runs. The perturbations are configured as if every perturbed member were only updated every 6 h, but new control runs, including assimilation of the latest observations, are run every hour. With new observations every hour and variation in some of the observation types (radar and satellite data) the control runs comprise a simple ensemble data assimilation system that samples observational uncertainty. The control runs use standard 3-hourly data assimilation cycling, but run in separate independent cycles in order
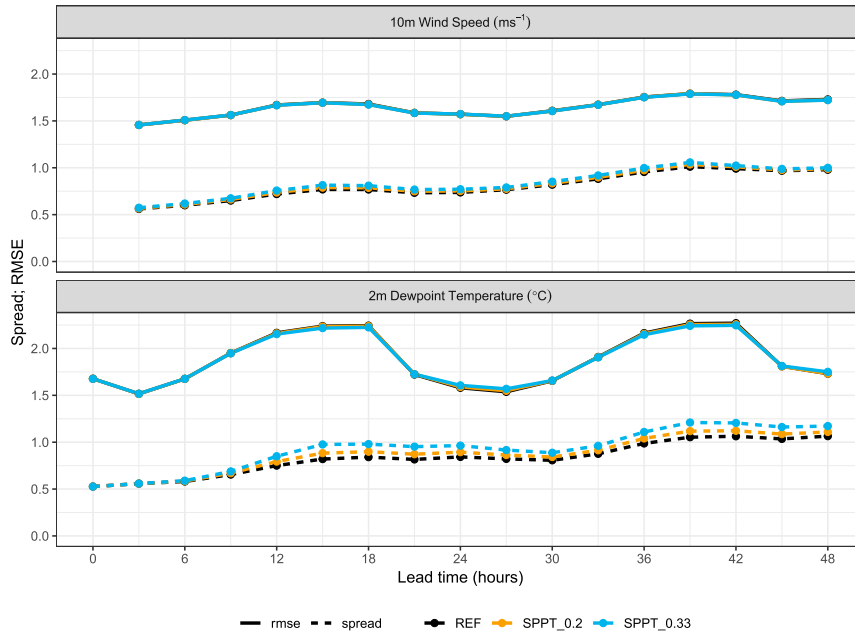
FIG. 19. Spread (dashed) and RMSE (solid) for (top) S10m at 744 stations and (bottom) Td2m at 825 stations for REF (black), SPPT_0.2 (orange), and SPPT_0.33 (blue).

to get new control runs every hour. This cycling strategy is also believed to reduce spinup problems for moist variables seen in 1-hourly assimilation cycling. Surface data are assimilated using pseudo 1-hourly assimilation cycling where first-guess data are taken from the latest cycle. Only one control run is included in the lagged ensemble; the other control runs are short forecasts used only for data assimilation cycling purposes. In addition to including observational uncertainty the hourly updates of the ensemble distributes the computational load throughout the day instead of imposing a massive peak in computational load every six hours. Initial and LBC perturbations include both SLAF, PertAna, and random field perturbations [sections 4a and 4c(1), as SLAF alone will not allow for enough members], and in addition, random surface perturbations as described in section 4b. Model perturbations include alternative turbulence and mass-flux schemes, use of a condensation threshold function, subgrid-scale orography, and microphysics modifications [section 4d(1)].

The Irish Regional Ensemble Prediction System (IREPS) is a configuration of the HarmonEPS that became operational at the Irish Meteorological Service, Met Éireann, in October 2018. IREPS is similar to the MEPS implementation of HarmonEPS being constructed of 10 perturbed ensemble members and 1 control member. The perturbed members are generated using the SLAF technique (section 4a) as well as having

perturbations applied to certain surface parameters following the methodology described in section 4b. Two cycles a day, at 0000 and 1200 UTC, are run out to a forecast length of +36 h and cover the domain in black shown in Fig. 1.

AEMET-γSREPS is the ensemble system running operationally at AEMET with more than 3000 probabilistic products available through a web page in the forecaster's offices at AEMET, Spain. It is a multiboundary and multimodel configuration. It consists of 20 members coming from crossing five different boundaries with four distinct nonhydrostatic convection-permitting LAM-NWP models. Its multiboundaries multimodel design is quite similar to its ancestor AEMET-SREPS, which had been operational from 2006 to 2014 (García-Moya et al. 2011). Since 26 April 2016 AEMET-γSREPS is being integrated at 0000 and 1200 UTC cycles up to 36 h and extended to 48 h in 2018 over the Iberian Peninsula (green domain in Fig. 1). Since 13 November 2018 it is also run operationally over the Canary Islands at 0000 UTC out to 48 h. And since 1 December 2018 it is integrated at 0000 UTC out to 48 h on a domain around Livingston Island (Antarctica), but only with 16 members and during the Antarctic Campaign (from 1 December to 31 March) in order to support Spanish Antarctic research activities. The multiboundary approach deals with the initial and lateral boundary uncertainties taking the boundary conditions from five Meteorological Centers that execute global NWP models

TABLE 2. Summary of operational implementations of HarmonEPS described in section 5.

	Meteorological institute(s)	Members	Initial and boundary perturbations	Assimilation	Observations	Model perturbations	Output
MEPS	Finland, Norway, and Sweden	1 + 9	SLAF, surface perturbations, PertAna	3 h 3DVAR and surface for control and 6 h surface for members	Conventional, AMSU-A, AMSU-B, MHS, RADAR (dBZ), ASCAT, GNSS, IASI		+54 h four times a day
COMEPS	Denmark	1 + 24 (small domain)	SLAF, random field perturbations, surface perturbations, PertAna, observations	3 h 3DVAR and 1 h surface for control	Conventional, AMSU-A, AMSU-B, MHS, AMV, MODE-S, ATMS, RO, radar reflectivity	Turbulence, mass-flux, condensation, subgrid-scale orography, microphysics	+51 h four times a day
IREPS	Ireland	1 + 10	SLAF, surface perturbations	3 h 3DVAR and surface for control. 6 h surface for members	Conventional, ASCAT		+36 h two times a day
RMI-EPS	Belgium	2 + 20	IFS ENS	6 h 3DVAR for both control members and 6 h surface for all members	Conventional	Multimodel approach: AROME and ALARO models/physics	+36 h two times a day
AEMET-γSREPS	Spain	20	Multiboundaries approach: ECMWF, NCEP, Météo-France, CMC and JMA	No	No	Multimodel approach: AROME, ALARO, WRF and NMMB models/physics	+48 h two times a day
KEPS	The Netherlands	1 + 10	SLAF	Both 3 h 3DVAR and 3 h surface for all members	Conventional, ASCAT, mode-S		+48 h eight times a day

(center, NWP): ECMWF, IFS; MétéoFrance, ARPÈGE; Japanese JMA, GSM; NOAA NCEP, GFS; and Canadian CMC, GEM. The multimodel technique addresses the model errors and uncertainties executing four different NWP models (center or consortium, NWP): HIRLAM, HARMONIE–AROME (Bengtsson et al. 2017); HIRLAM–ALADIN, HARMONIE–ALARO (Termonia et al. 2018); NOAA NCAR, WRF-ARW (Skamarock et al. 2008); and NOAA NCEP, NMMB (Janjić and Gall 2012). Due to its relatively small area (565×469 grid points) for the three domains, it could be stated that synoptic- and meso-α-scale uncertainties are taken into account from global NWP models through boundary conditions meanwhile meso-β-scale uncertainties are tackled mainly with the multimodel approach. Future plans of AEMET-γSREPS are in the short term to extend the Iberian Peninsula domain to 789×637 grid points, to include LETKF assimilation [see LETKF section 4c(4)] and in the longer term, a fifth convection-permitting NWP model in order to have 25 members.

A prototype convection-permitting EPS is under development at the Royal Meteorological Institute of Belgium (RMI), called RMI-EPS. A combination of the HarmonEPS system with RMI preprocessing and postprocessing scripts is used. Since September 2017, the system runs twice a day on the ECMWF's HPC infrastructure preoperationally (i.e., the system runs as if it were operational, but without a guarantee of timely delivery). Currently RMI-EPS consists of 22 ensemble members. There are two control members, one using the HARMONIE–AROME configuration, and the other using the HARMONIE–ALARO configuration, both described in section 2. Each control member has 10 corresponding perturbed members. Initial perturbations and boundary conditions for these members are taken from IFS ENS. Both control members have a 3DVAR upper-air data assimilation cycle. Each member also has its own surface assimilation cycle, as in the standard HarmonEPS setup [section 4c(2)]. The RMI-EPS system has two main cycles a day (0000 and 1200 UTC) with a forecast range of +36 h and covers the red domain in Fig. 1. Additionally, there are two 6-h data assimilation cycles (0600 and 1800 UTC). Some more details of the system, together with results for several thunderstorm events that occurred over Belgium in August 2015, can be found in Smet (2017).

The Royal Netherlands Meteorological Institute (KNMI) runs a preoperational HarmonEPS implementation, KEPS, since January 2018 for the dark green domain in Fig. 1. This system has 1 control member and 10 perturbed members using the SLAF method for the boundary conditions (section 4a). Upper-air analysis is with 3DVAR for all members that all use the same observations, and every member has also surface data assimilation [section 4c(2)]. Cycling frequency is every 3 h, with forecast lengths of 48 h for each cycle. A very recent update of the KEPS configuration is the use of PertAna for computing different analysis perturbations for all the members except the control. Upper-air analysis is also now computed only for the control run. This new configuration is not indicated in Table 2, which relates to the results of November 2018 shown in Fig. 20.

In Fig. 20 scores from the (pre)operational HarmonEPS implementations are shown for 12-h accumulated precipitation for November 2018, and compared with IFS ENS. HarmonEPS in different configurations is able to produce higher spread with less members than IFS ENS, and mostly better or comparable RMSE. A more indepth investigation of the added value of one of the operational systems, MEPS, over IFS ENS can be found in Frogner et al. (2019). This paper also investigates the added value of EPS over deterministic forecasts.

6. Outlook

As described in this paper, HarmonEPS includes a range of different choices for perturbations to different parts of the system, some of which can be combined. As seen in section 5, different operational institutes have chosen differently between these available perturbation strategies, and all perform well compared with IFS ENS. There is a trade-off between providing flexibility and the possibility to choose between different perturbation strategies and focusing human and computational resources on developing and maintaining the "best" or "correct" perturbation strategies. At present, it cannot be declared what perturbation strategies are best or correct, and therefore what perturbation strategies should be skipped, as they all show some advantages. On the other hand, it is desirable to take HarmonEPS in a direction of perturbations that represent errors close to their source. However, the perturbations that are thought to be theoretically most correct, might not be the ones giving the best scores, and in developing operational systems there might be conflicts between pragmatic views of obtaining the best scores and what theoretically is seen as the physically correct perturbations. It is not known if the perturbations so far introduced are best suited for a convection-permitting ensemble, as many of them are similar to those being used in coarser-resolution EPSs. Further work is necessary to look into the distinctiveness of convection-permitting ensembles, including utilizing diagnostics and verification metrics suitable for small-scale phenomena.

It still remains a goal to gradually move HarmonEPS in the direction of a system with physically consistent perturbation strategies. One such example is the work
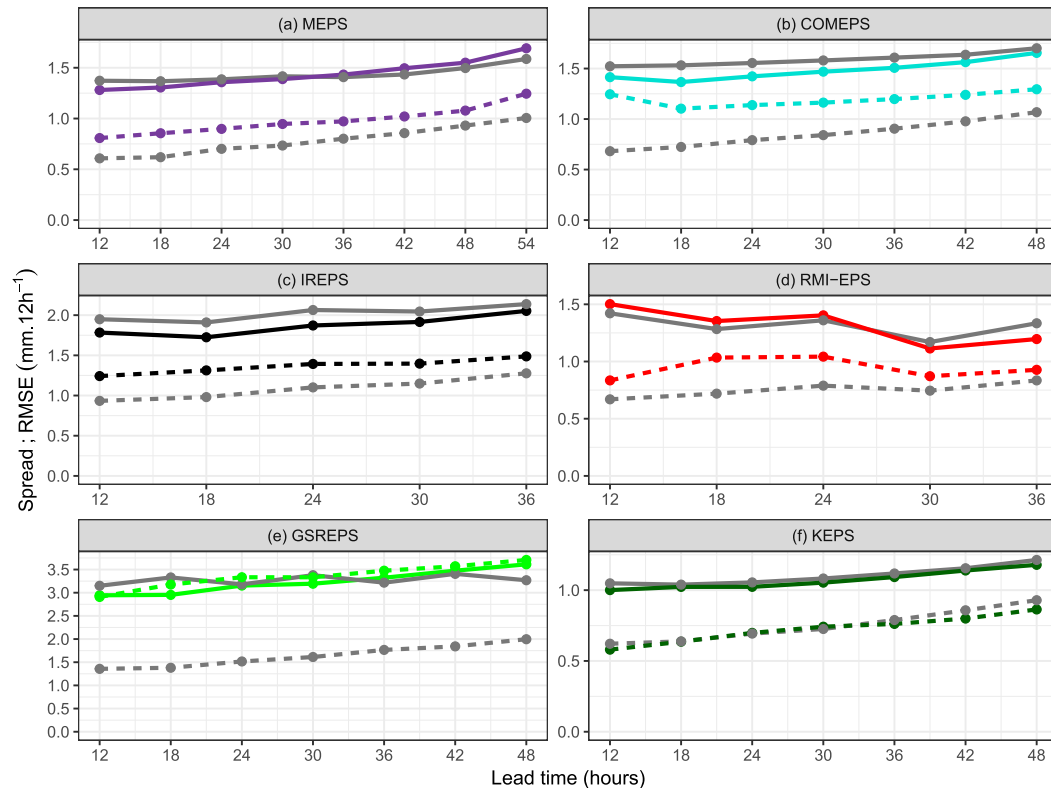
FIG. 20. Spread (dashed) and RMSE (solid) for the different HarmonEPS systems compared with IFS ENS (gray) for 12-h accumulated precipitation for the combined daily runs in November 2018. Note the different scales in the plots. (a) MEPS for 643 stations in the MetCoOp domain (purple area in Fig. 1), (b) COMEPS for 660 stations over the small COMEPS domain (the smaller cyan domain in Fig. 1), (c) IREPS for 226 stations in the IREPS domain (black area in Fig. 1), (d) RMI-EPS for 8 stations in Belgium, (e) AEMET-γSREPS for 145 stations over the IBERIA_2.5 domain (green domain in Fig. 1), and (f) KEPS for 191 stations in a 300 × 300 gridpoint area around the Netherlands, which is smaller than and contained in the 800 × 800 gridpoint area represented by the dark green model domain in Fig. 1.

on uncertain parameters in the parameterizations, and how to represent the uncertainty at the sources of the individual physical processes. There are several ways of perturbing uncertain parameters in the parameterizations, with varying complexity. The simplest is to assign to each member of the ensemble parameters that are fixed during the integration, sometimes referred to as fixed parameter perturbations. Like for MP, this can lead to different members having different biases. A somewhat more stochastic approach is random perturbed parameters (RPP) where each member has a different value of one or a few parameters fixed during the integration, but where the parameters are randomly chosen from a prescribed distribution for each member and cycle. This ensures statistical indistinguishability of the members. Both approaches are described in Marsigli et al. (2014b). A scheme developed for the UKMO ensemble systems (Bowler et al. 2008b; Baker et al. 2014) called random parameters (RP) introduced stochastic parameters that vary discontinuously in time. Another

way is to randomly and gradually change the parameter during the forecast, depending on space and time. ECMWF are working on the stochastically perturbed parameterizations scheme (Ollinaho et al. 2017), a scheme where perturbations evolve in time and space according to the same pattern generator as is explained above for SPPT. SPP samples a lognormal distribution for most parameters with independent distributions for each parameter and variable, making sure the perturbations are uncorrelated. SPP has an advantage over SPPT in that it represents the errors close to their source, it respects local budgets of moisture, momentum, and energy, and can also represent uncertainty beyond a simple amplitude error (Leutbecher et al. 2017). However, SPP is more complex to develop, improve, and maintain. The SPP approaches for perturbing uncertain parameters in the parameterizations are at the time of writing being developed and tested in HarmonEPS. RPP is just a special case of SPP and will be tested as well and compared with SPP. SPP in

HarmonEPS is implemented with the same framework and same characteristics as in IFS, but obviously the parameterization schemes and parameters are different. Currently 14 parameters are implemented in HarmonEPS SPP, and work is ongoing to implement and test more parameters. The work to identify sensitive and uncertain parameters from the parameterizations of microphysics, cloud processes, convection and radiation is done in close cooperation with HARMONIE–AROME physics experts. Perturbations to the dynamics will also be included. There are many different sources of model error, hence it is not presently clear if SPP will be sufficient. Neither is it not known whether one approach to model error description is better than another. For some time to come, it might be beneficial to combine SPP and SPPT to cover a greater part of the uncertainties. The benefit of combining SPP and SPPT was shown in Jankov et al. (2017) in the Rapid Refresh ensemble system based on the Weather Research and Forecasting Model.

Work will continue to improve SPPT in HarmonEPS. Below is a summary of a number of foreseeable developments that are planned to extend SPPT capabilities in HarmonEPS, which could improve its performance, with the drawback, on the other hand, that they are going to significantly increase the number of parameters to be experimentally tuned:

- Combining several spatiotemporal-scale patterns as is done with three scales at the ECMWF (Palmer et al. 2009; Leutbecher et al. 2017).
- Using a 3D pattern generator instead of the current 2D.
- Perturbing independently each parameterization (Arnold 2013; Christensen et al. 2017), as well as partial SPPT (Wastl et al. 2017), and as it was suggested by Shutts and Callado Pallarès (2014), diagnose coarse-grained comparisons between tendencies from ECMWF/IFS integrations with different horizontal resolutions that show distinct uncertainties for each parameterization. Furthermore it is planned to try a combined pattern with a common fraction for all parameterizations added to another independent fraction for each one with the idea to ensure at least some physical consistency.
- Perturb independently each variable, as some coarse-graining results suggest, as well as different uncertainties in different variables.
- Better adjusting the PBL and upper-atmosphere SPPT tapering for LAM-EPS, or even not apply it at all.

Another example of moving toward a system with representation of errors close to their source was discussed in section 4c(3) with the introduction of perturbed observations (EDA). Until now initial condition uncertainty in HarmonEPS has been coming solely from the coarser-resolution nesting model (the method of PertAna), and while that cannot be claimed to be unphysical, it introduces some noise in the first few hours of the forecast. The introduction of uncertainty in the observations in the HarmonEPS EDA is accounting for known uncertainties in the observations and representativeness issues, and it is an approach that does not introduce noisy patterns (as in e.g., PertAna). Including observation perturbations also resulted in a need to reduce the perturbations coming from the nesting model so as not to increase the RMSE. In this test, model uncertainty was not accounted for in the EDA or in the forecasts. This could be important and will be tested in the future. With the setup used here where the EDA and ensemble generation are integrated into one system, the model uncertainty will be consistent between assimilation and forecast. A tuning of the size of the EDA perturbations together with the PertAna perturbations, and also the surface perturbations and model error perturbations that are not part of this exercise, is recommended before being used operationally. Although EDA costs more as each member has to run its own analysis, it is not more time consuming in a standard operational cycling. Without EDA the members still have to wait for the control analysis to finish as they use the control analysis with perturbations added.

It was seen in section 4b that the surface perturbations are effective in increasing the spread of the undispersed ensemble. However, it was also seen in section 4c(3) when comparing the size of the perturbations for variables with long memory (Fig. 11) that the size of the surface perturbations is very large and is actually decreasing with lead time. This can indicate that the perturbations are too large for the model to maintain. Also, further improvements could be made. In the experiments discussed herein, the perturbation fields all had the same spatial scale, regardless of parameter. It may be more realistic to perturb different parameters at different spatial scales depending on the parameter. Furthermore, uncertainties in vegetation fraction and leaf area index may depend on both vegetation type and season and so different perturbations could be applied dependent on those factors. Work is ongoing to investigate these issues and to explore perturbing other surface parameters, such as soil ice content in the winter and sea ice concentration/extent. It remains to be seen if the more realistic perturbations will give verification scores that are as satisfactory as the current scheme.

In the next few years HIRLAM EPS work will focus on improving and including more sources of uncertainty in all aspects of the model, and will strive to move in the direction of describing the errors close to the source and to design perturbation strategies that are suitable for the convection-permitting scales. This includes getting SPP

operational, refining the surface perturbation scheme, further understanding and developing perturbations for the initial conditions, and uncovering how best to create an ensemble that also fits with the needs of data assimilation. The fact that there are several different operational implementations of HarmonEPS is an advantage in the development process. The different institutes with their different needs and weather forecasting challenges help us to build a system that performs well in different climates in Europe, which can also lead to important lessons learned for the future with a changing climate.

APPENDIX

List of Acronyms

3DVAR	Three-dimensional variational data assimilation
3MT	Modular Multiscale Microphysics and Transport scheme
4DVAR	Four-dimensional variational data assimilation
ALADIN	Aire Limitée Adaptation Dynamique Développement International
Alaro	ALADIN–AROME
AROME	Applications of Research to Operations at Mesoscale
ARPÉGE	Action de Recherche Petite Echelle Grande Echelle
BRAND	B matrix randomization
COMEPS	Continuously Updated Mesoscale Ensemble Prediction System
DA	Data assimilation
ECMWF	European Centre for Medium-Range Weather Forecasts
EDA	Ensemble data assimilation
EnKF	Ensemble Kalman filtering
EnSRF	Ensemble square root filter
EPS	Ensemble prediction system
GLAMEPS	Grand Limited Area Modeling Ensemble Prediction System
HarmonEPS	HARMONIE Ensemble Prediction System
HARMONIE	HIRLAM–ALADIN Research on Mesoscale Operational NWP in Euromed
HIRLAM	The international research program High Resolution Limited Area Model
HPC	High-performance computing
IFS	Integrated Forecasting System
IFS HRES	Integrated Forecasting System—High Resolution (deterministic)
IFS ENS	Integrated Forecasting system—Ensemble
IREPS	Irish Regional Ensemble Prediction System
KEPS	The Netherlands Meteorological Institute ensemble prediction system
LAM	Limited-area model
LBC	Lateral boundary conditions
LETKF	Local ensemble transform Kalman filter
MEPS	MetCoOp Ensemble Prediction System
MetCoOp	Meterological Cooperation on Operational Numeric Weather Prediction between the Finnish Meteorological Institute, MET Norway, and Swedish Meteorological and Hydrological Institute
MP	Multiphysics
NWP	Numerical weather prediction
PertAna	Method for generation of initial condition perturbations
RMI-EPS	Belgian Meteorological Institute ensemble prediction system
RP	Random parameters
RPP	Random perturbed parameters
SKEB	Stochastic kinetic energy backscatter
SLAF	Scaled lagged average forecasting
SPG	Stochastic pattern generator
SPP	Stochastically perturbed parameterizations scheme
SPPT	Stochastically perturbed parameterization tendencies
SURFEX	Land and ocean surface model
TKE	Total kinetic energy
γSREPS	Spanish Meteorological Institute Short Range Ensemble Prediction System

REFERENCES

Andersson, E., and Coauthors, 1998: The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part III: Experimental results. *Quart. J. Roy. Meteor. Soc.*, **124**, 1831–1860, https://doi.org/10.1002/qj.49712455004.

Arnold, H. M., 2013: Stochastic parametrisation and model uncertainty. Ph.D. thesis, University of Oxford, Oxford, United Kingdom, 238 pp.

Aspelien, T., T. Iversen, J. B. Bremnes, and I.-L. Frogner, 2011: Short-range probabilistic forecasts from the Norwegian limited-area EPS: Long-term validation and a polar low study. *Tellus*, **63A**, 564–584, https://doi.org/10.1111/j.1600-0870.2010.00502.x.

Baker, L., A. Rudd, S. Migliorini, and R. Bannister, 2014: Representation of model error in a convective-scale ensemble prediction system. *Nonlinear Processes Geophys.*, **21**, 19–39, https://doi.org/10.5194/npg-21-19-2014.

Beck, J., F. Bouttier, L. Wiegand, C. Gebhardt, C. Eagle, and N. Roberts, 2016: Development and verification of two convection-allowing multi-model ensembles over Western Europe. *Quart. J. Roy. Meteor. Soc.*, **142**, 2808–2826, https://doi.org/10.1002/qj.2870.

Bénard, P., J. Vivoda, J. Mašek, P. Smolíková, K. Yessad, C. Smith, R. Brožková, and J.-F. Geleyn, 2010: Dynamical kernel of the Aladin–NH spectral limited-area model: Revised formulation and sensitivity experiments. *Quart. J. Roy. Meteor. Soc.*, **136**, 155–169, https://doi.org/10.1002/qj.522.

Bengtsson, L., and Coauthors, 2017: The HARMONIE–AROME model configuration in the ALADIN–HIRLAM NWP system. *Mon. Wea. Rev.*, **145**, 1919–1935, https://doi.org/10.1175/MWR-D-16-0417.1.

Berner, J., G. J. Shutts, M. Leutbecher, and T. N. Palmer, 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, https://doi.org/10.1175/2008JAS2677.1.

Berre, L., 2000: Estimation of synoptic and mesoscale forecast error covariances in a limited-area model. *Mon. Wea. Rev.*, **128**, 644–667, https://doi.org/10.1175/1520-0493(2000)128<0644:EOSAMF>2.0.CO;2.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436, https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.

Bojarova, J., and N. Gustafsson, 2019: Relevance of climatological background error statistics for mesoscale data assimilation. *Tellus*, **71A**, 1–22, https://doi.org/10.1080/16000870.2019.1615168.

Bonavita, M., M. Hamrud, and L. Isaksen, 2015: EnKF and hybrid gain ensemble data assimilation. Part II: EnKF and hybrid gain results. *Mon. Wea. Rev.*, **143**, 4865–4882, https://doi.org/10.1175/MWR-D-15-0071.1.

Bouttier, F., and L. Raynaud, 2018: Clustering and selection of boundary conditions for limited-area ensemble prediction. *Quart. J. Roy. Meteor. Soc.*, **144**, 2381–2391, https://doi.org/10.1002/qj.3304.

——, B. Vie, O. Nuissier, and L. Raynaud, 2012: Impact of stochastic physics in a convection-permitting ensemble. *Mon. Wea. Rev.*, **140**, 3706–3721, https://doi.org/10.1175/MWR-D-12-00031.1.

——, L. Raynaud, O. Nuissier, and B. Ménétrier, 2016: Sensitivity of the AROME ensemble to initial and surface perturbations during HyMeX. *Quart. J. Roy. Meteor. Soc.*, **142**, 390–403, https://doi.org/10.1002/qj.2622.

Bowler, N. E., A. Arribas, and K. R. Mylne, 2008a: The benefits of multianalysis and poor man's ensembles. *Mon. Wea. Rev.*, **136**, 4113–4129, https://doi.org/10.1175/2008MWR2381.1.

——, ——, ——, K. B. Robertson, and S. E. Beare, 2008b: The MOGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **134**, 703–722, https://doi.org/10.1002/qj.234.

Brousseau, P., L. Berre, F. Bouttier, and G. Desroziers, 2011: Background-error covariances for a convective-scale data-assimilation system: AROME–France 3D-Var. *Quart. J. Roy. Meteor. Soc.*, **137**, 409–422, https://doi.org/10.1002/qj.750.

Buizza, R., and T. Palmer, 1995: The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, **52**, 1434–1456, https://doi.org/10.1175/1520-0469(1995)052<1434:TSVSOT>2.0.CO;2.

——, M. Miller, and T. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908, https://doi.org/10.1002/qj.49712556006.

——, M. Leutbecher, and L. Isaksen, 2008: Potential use of an ensemble of analyses in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **134**, 2051–2066, https://doi.org/10.1002/qj.346.

Charron, M., G. Pellerin, L. Spacek, P. L. Houtekamer, N. Gagnon, H. L. Mitchell, and L. Michelin, 2010: Toward random sampling of model error in the Canadian Ensemble Prediction System. *Mon. Wea. Rev.*, **138**, 1877–1901, https://doi.org/10.1175/2009MWR3187.1.

Christensen, H. M., S. Lock, I. M. Moroz, and T. N. Palmer, 2017: Introducing independent patterns into the Stochastically Perturbed Parametrization Tendencies (SPPT) scheme. *Quart. J. Roy. Meteor. Soc.*, **143**, 2168–2181, https://doi.org/10.1002/qj.3075.

Clark, A. J., and Coauthors, 2011: Probabilistic precipitation forecast skill as a function of ensemble size and spatial scale in a convection-allowing ensemble. *Mon. Wea. Rev.*, **139**, 1410–1418, https://doi.org/10.1175/2010MWR3624.1.

Coiffier, J., 2011: *Fundamentals of Numerical Weather Prediction.* Cambridge University Press, 368 pp.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367–1388, https://doi.org/10.1002/qj.49712051912.

Done, J., C. A. Davis, and M. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the weather research and forecasting (WRF) model. *Atmos. Sci. Lett.*, **5**, 110–117, https://doi.org/10.1002/asl.72.

Du, J., and M. Tracton, 2001: Implementation of a real-time short range ensemble forecasting system at NCEP: An update. Preprints, *Ninth Conf. on Mesoscale Processes,* Fort Lauderdale, FL, Amer. Meteor. Soc., P4.9, https://ams.confex.com/ams/pdfpapers/23074.pdf.

Duran, I. B., J.-F. Geleyn, and F. Vana, 2014: A compact model for the stability dependency of TKE production–destruction–conversion terms valid for the whole range of Richardson numbers. *J. Atmos. Sci.*, **71**, 3004–3026, https://doi.org/10.1175/JAS-D-13-0203.1.

Ebisuzaki, W., and E. Kalnay, 1991: Ensemble experiments with a new lagged average forecasting scheme. *WMO Research Activities in Atmospheric and Oceanic Modeling*, Vol. 15, WMO, 308 pp.

Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, https://doi.org/10.1007/s10236-003-0036-9.

Frogner, I.-L., and T. Iversen, 2002: High-resolution limited-area ensemble predictions based on low resolution targeted singular vectors. *Quart. J. Roy. Meteor. Soc.*, **128**, 1321–1341, https://doi.org/10.1256/003590002320373319.

——, H. Haakenstad, and T. Iversen, 2006: Limited-area ensemble predictions at the Norwegian Meteorological Institute. *Quart. J. Roy. Meteor. Soc.*, **132**, 2785–2808, https://doi.org/10.1256/qj.04.178.

Frogner, I., T. Nipen, A. Singleton, J. B. Bremnes, and O. Vignes, 2016: Ensemble prediction with different spatial resolutions

for the 2014 Sochi Winter Olympic Games: The effects of calibration and multimodel approaches. *Wea. Forecasting*, **31**, 1833–1851, https://doi.org/10.1175/WAF-D-16-0048.1.

Frogner, I.-L., A. T. Singleton, M. Ø. Køltzow, and U. Andrae, 2019: Convection-permitting ensembles: Challenges related to their design and use. *Quart. J. Roy. Meteor. Soc.*, **145**, 90–106, https://doi.org/10.1002/qj.3525.

García-Moya, J.-A., A. Callado, P. Escribà, C. Santos, D. Santos-Muñoz, and J. Simarro, 2011: Predictability of short-range forecasting: A multimodel approach. *Tellus*, **63A**, 550–563, https://doi.org/10.1111/j.1600-0870.2010.00506.x.

Gebhardt, C., S. E. Theis, P. Krahe, and V. Renner, 2008: Experimental ensemble forecasts of precipitation based on a convection-resolving model. *Atmos. Sci. Lett.*, **9**, 67–72, https://doi.org/10.1002/asl.177.

Gerard, L., J.-M. Piriou, R. Brožková, J.-F. Geleyn, and D. Banciu, 2009: Cloud and precipitation parameterization in a meso-gamma-scale operational weather prediction model. *Mon. Wea. Rev.*, **137**, 3960–3977, https://doi.org/10.1175/2009MWR2750.1.

Giard, D., and E. Bazile, 2000: Implementation of a new assimilation scheme for soil and surface variables in a global NWP model. *Mon. Wea. Rev.*, **128**, 997–1015, https://doi.org/10.1175/1520-0493(2000)128<0997:IOANAS>2.0.CO;2.

Ha, S., and C. Snyder, 2014: Influence of surface observations in mesoscale data assimilation using ensemble Kalman filter. *Mon. Wea. Rev.*, **142**, 1489–1508, https://doi.org/10.1175/MWR-D-13-00108.1.

Hacker, J. P., and Coauthors, 2011: The U.S. Air Force Weather Agency's mesoscale ensemble: Scientific description and performance results. *Tellus*, **63A**, 625–641, https://doi.org/10.1111/j.1600-0870.2010.00497.x.

Hagelin, S., J. Son, R. Swinbank, A. McCabe, N. Roberts, and W. Tennant, 2017: The Met Office convective-scale ensemble, MOGREPS-UK. *Quart. J. Roy. Meteor. Soc.*, **143**, 2846–2861, https://doi.org/10.1002/qj.3135.

Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560, https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2.

Hamrud, M., M. Bonavita, and L. Isaksen, 2015: EnKF and hybrid gain ensemble data assimilation. Part I: EnKF implementation. *Mon. Wea. Rev.*, **143**, 4847–4864, https://doi.org/10.1175/MWR-D-14-00333.1.

Hou, D., E. Kalnay, and K. K. Droegemeier, 2001: Objective verification of the SAMEX '98 ensemble forecasts. *Mon. Wea. Rev.*, **129**, 73–91, https://doi.org/10.1175/1520-0493(2001)129<0073:OVOTSE>2.0.CO;2.

Houtekamer, P. L., L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225–1242, https://doi.org/10.1175/1520-0493(1996)124<1225:ASSATE>2.0.CO;2.

——, H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133**, 604–620, https://doi.org/10.1175/MWR-2864.1.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation of spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, **230**, 112–126, https://doi.org/10.1016/j.physd.2006.11.008.

Isaksen, L., M. Fisher, and J. Berner, 2007: Use of analysis ensembles in estimating flow-dependent background error variance. *Proc. ECMWF Workshop on Flow-dependent Aspects of Data Assimilation*, Reading, United Kingdom, ECMWF, 65–86, https://www.ecmwf.int/sites/default/files/elibrary/2007/10127-use-analysis-ensembles-estimating-flow-dependent-background-error-variance.pdf.

——, M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher, and L. Raynaud, 2010: Ensemble of data assimilations at ECMWF. Research Department, Tech. Memo. 636, 48 pp., https://www.ecmwf.int/en/elibrary/10125-ensemble-data-assimilations-ecmwf.

Iversen, T., A. Deckmyn, C. Santos, K. Sattler, J. B. Bremnes, H. Feddersen, and I.-L. Frogner, 2011: Evaluation of 'GLAMEPS'—A proposed multimodel EPS for short range forecasting. *Tellus*, **63A**, 513–530, https://doi.org/10.1111/j.1600-0870.2010.00507.x.

Janjić, Z. I., and R. Gall, 2012: Scientific documentation of the NCEP Nonhydrostatic Multiscale Model on the B Grid (NMMB). Part 1: Dynamics. NCAR/TN-489+STR, 75 pp., https://doi.org/10.5065/D6WH2MZX.

Jankov, I., and Coauthors, 2017: A performance comparison between multiphysics and stochastic approaches within a North American RAP ensemble. *Mon. Wea. Rev.*, **145**, 1161–1179, https://doi.org/10.1175/MWR-D-16-0160.1.

Keller, J. D., L. Kornblueh, A. Hense, and A. Rhodin, 2008: Towards a GME ensemble forecasting system: Ensemble initialization using the breeding technique. *Meteor. Z.*, **17**, 707–718, https://doi.org/10.1127/0941-2948/2008/0333.

Kiktev, D., and Coauthors, 2017: FROST-2014: The Sochi Winter Olympics International Project. *Bull. Amer. Meteor. Soc.*, **98**, 1908–1929, https://doi.org/10.1175/BAMS-D-15-00307.1.

Klasa, C., M. Arpagaus, A. Walser, and H. Wernli, 2018: An evaluation of the convection-permitting ensemble COSMO-E for three contrasting precipitation events in Switzerland. *Quart. J. Roy. Meteor. Soc.*, **144**, 744–764, https://doi.org/10.1002/qj.3245.

Leutbecher, M., and Coauthors, 2017: Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Quart. J. Roy. Meteor. Soc.*, **143**, 2315–2339, https://doi.org/10.1002/qj.3094.

Lindskog M., M. Dahlbom, S. Thorsteinsson, P. Dahlgren, R. Randriamampianina, J. Bojarova, 2012: ATOVS processing and usage in the HARMONIE reference system. *HIRLAM Newsl.*, **59**, 33–43.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141, https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

Magnusson, L., E. Källén, and J. Nycander, 2008: Initial state perturbations in ensemble forecasting. *Nonlinear Processes Geophys.*, **15**, 751–759, https://doi.org/10.5194/npg-15-751-2008.

Marsigli, C., F. Boccanera, A. Montani, and T. Paccagnella, 2005: The COSMO-LEPS mesoscale ensemble system: Validation of the methodology and verification. *Nonlinear Processes Geophys.*, **12**, 527–536, https://doi.org/10.5194/npg-12-527-2005.

——, A. Montani, and T. Paccagnella, 2014a: Provision of boundary conditions for a convection-permitting ensemble: Comparison of two different approaches. *Nonlinear Processes Geophys.*, **21**, 393–403, https://doi.org/10.5194/npg-21-393-2014.

——, ——, and ——, 2014b: Perturbation of initial and boundary conditions for a limited-area ensemble: Multi-model versus single-model approach. *Quart. J. Roy. Meteor. Soc.*, **140**, 197–208, https://doi.org/10.1002/qj.2128.

Masson, V., and Coauthors, 2013: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes. *Geosci. Model Dev.*, **6**, 929–960, https://doi.org/10.5194/gmd-6-929-2013.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF Ensemble Prediction System: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119, https://doi.org/10.1002/qj.49712252905.

——, ——, C. Marsigli, A. Montani, F. Nerozzi, and T. Paccagnella, 2001: A strategy for high-resolution ensemble prediction. I: Definition of representative members and global-model experiments. *Quart. J. Roy. Meteor. Soc.*, **127**, 2069–2094, https://doi.org/10.1002/qj.49712757612.

Ollinaho, P., and Coauthors, 2017: Towards process-level representation of model uncertainties: Stochastically perturbed parametrizations in the ECMWF ensemble. *Quart. J. Roy. Meteor. Soc.*, **143**, 408–422, https://doi.org/10.1002/qj.2931.

Palmer, T. N., 1997: On parametrising scales that are only somewhat smaller than the smallest resolved scale, with application to convection and orography. *Proc. ECMWF Workshop on New Insights and Approaches to Convective Parametrisation*, Reading, United Kingdom, ECMWF, 328–337.

——, R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. ECMWF Tech. Memo 598, 42 pp.

Pan, Y., K. Zhu, M. Xue, X. Wang, M. Hu, S. G. Benjamin, S. S. Weygandt, and J. S. Whitaker, 2014: A GSI-based coupled EnSRF–En3DVar hybrid data assimilation system for the operational Rapid Refresh model: Tests at a reduced resolution. *Mon. Wea. Rev.*, **142**, 3756–3780, https://doi.org/10.1175/MWR-D-13-00242.1.

Pu, Z., H. Zhang, and J. Anderson, 2013: Ensemble Kalman filter assimilation of near-surface observations over complex terrain: Comparison with 3DVAR for short-range forecasts. *Tellus*, **65A**, 19620, https://doi.org/10.3402/tellusa.v65i0.19620.

Randriamampianina, R., 2006: Impact of high resolution satellite observations in the ALADIN/HU model. *Idojaras*, **110**, 329–347.

——, T. Iversen, and A. Storto, 2011: Exploring the assimilation of IASI radiances in forecasting polar lows. *Quart. J. Roy. Meteor. Soc.*, **137**, 1700–1715, https://doi.org/10.1002/qj.838.

Raynaud, L., and F. Bouttier, 2016: Comparison of initial perturbation methods for ensemble prediction at convective scale. *Quart. J. Roy. Meteor. Soc.*, **142**, 854–866, https://doi.org/10.1002/qj.2686.

Ridal, M., and M. Dahlbom, 2017: Assimilation of multinational radar reflectivity data in a mesoscale model: A proof of concept. *J. Appl. Meteor. Climatol.*, **56**, 1739–1751, https://doi.org/10.1175/JAMC-D-16-0247.1.

Ritter, B., and J.-F. Geleyn, 1992: A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations. *Mon. Wea. Rev.*, **120**, 303–325, https://doi.org/10.1175/1520-0493(1992)120<0303:ACRSFN>2.0.CO;2.

Romine, G. S., C. S. Schwartz, J. Berner, K. R. Fossell, C. Snyder, J. L. Anderson, and M. L. Weisman, 2014: Representing forecast error in a convection-permitting ensemble system. *Mon. Wea. Rev.*, **142**, 4519–4541, https://doi.org/10.1175/MWR-D-14-00100.1.

Schraff, C., H. Reich, A. Rhodin, A. Schomburg, K. Stephan, A. Periáñez, and R. Potthast, 2016: Kilometre-scale ensemble data assimilation for the COSMO model (KENDA). *Quart.*

J. Roy. Meteor. Soc., **142**, 1453–1472, https://doi.org/10.1002/qj.2748.

Schwartz, C. S., G. S. Romine, K. R. Fossell, R. A. Sobash, and M. L. Weisman, 2017: Toward 1-km ensemble forecasts over large domains. *Mon. Wea. Rev.*, **145**, 2943–2969, https://doi.org/10.1175/MWR-D-16-0410.1.

Separovic, L., M. Charron, A. Erfani, N. Gagnon, A. Zadra, and P. Vaillancourt, 2016: Model error representation in the Canadian ensemble prediction systems. *ECMWF/WWRP Workshop on Model Uncertainty,* ECMWF, Reading, United Kingdom, 24 pp., https://www.ecmwf.int/en/elibrary/16742-model-error-representation-canadian-ensemble-prediction-systems.

Shutts, G., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079–3102, https://doi.org/10.1256/qj.04.106.

——, and A. Callado Pallarès, 2014: Assessing parametrization uncertainty associated with horizontal resolution in numerical weather prediction models. *Philos. Trans. Roy. Soc.*, **372**, https://doi.org/10.1098/rsta.2013.0284.

Skamarock, W., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., https://doi.org/10.5065/D68S4MVH.

Smet, G., 2017: RMI-EPS: A prototype convection-permitting EPS for Belgium. *ALADIN-HIRLAM Newsl.*, **8**, 73–79.

——, P. Termonia, and A. Deckmyn, 2012: Added economic value of limited area multi-EPS weather forecasting applications. *Tellus*, **64A**, 18901, https://doi.org/10.3402/tellusa.v64i0.18901.

Termonia, P., and Coauthors, 2018: The ALADIN system and its canonical model configurations AROME CY41T1 and ALARO CY40T1. *Geosci. Model Dev.*, **11**, 257–281, https://doi.org/10.5194/gmd-11-257-2018.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330, https://doi.org/10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2.

Tsyrulnikov, M., and D. Gayfulin, 2017: A limited-area spatio-temporal stochastic pattern generator for simulation of uncertainties in ensemble applications. *Meteor. Z.*, **26**, 549–566, https://doi.org/10.1127/metz/2017/0815.

Valkonen, T., H. Schyberg, and J. Figa-Saldaña, 2017: Assimilating advanced scatterometer winds in a high-resolution limited area model over northern Europe. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **10**, 2394–2405, https://doi.org/10.1109/JSTARS.2016.2602889.

Wang, Y., and Coauthors, 2011: The central European limited-area ensemble forecasting system: ALADIN-LAEF. *Quart. J. Roy. Meteor. Soc.*, **137**, 483–502, https://doi.org/10.1002/qj.751.

Wastl, C., Y. Wang, and C. Wittmann, 2017: A new stochastic scheme perturbing partial model tendencies. *EMS Annual Meeting*, Dublin, Ireland, European Meteorological Society, Vol. 14, EMS2017-207.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences.* 3rd ed. International Geophysics Series, Vol. **100**, Academic Press, 704 pp.