C-LAEF: Convection-permitting Limited-Area Ensemble Forecasting system

Clemens Wastl1 | Yong Wang1 | Aitor Atencia1 | Florian Weidle1 | Christoph Wittmann1 | Christoph Zingerle2 | Endi Keresturi3

Department of Forecasting Models, Zentralanstalt für Meteorologie und Geodynamik (ZAMG), Vienna, Austria
Customer Service of Tyrol and Vorarlberg, Zentralanstalt für Meteorologie und Geodynamik (ZAMG), Vienna, Austria
NWP Section, Meteorological and Hydrological Service, Zagreb, Croatia

Abstract

C-LAEF (Convection-permitting Limited-Area Ensemble Forecasting) has been developed at the Austrian national weather service ZAMG (Zentralanstalt für Meteorologie und Geodynamik) and has been running operationally at the European Centre for Medium Range Weather Forecasts (ECMWF) supercomputer since November 2019. It includes (a) an ensemble 3D variational blending technique to deal with atmospheric initial uncertainties, (b) an ensemble of land surface data assimilation to account for uncertainties in the initial land surface conditions, (c) a hybrid stochastic physics perturbation scheme to treat model uncertainties in the different physics parametrization schemes, and (d) a coupling with the global ensemble system IFS-ENS (Integrated Forecasting System-ENSemble) to consider uncertainties in the lateral boundary conditions. C-LAEF has a horizontal resolution of 2.5 km and consists of 16 perturbed members plus one unperturbed control run. It runs four times a day and provides probabilistic forecasts up to 60 hr on a domain covering the whole Alpine area. This article describes the C-LAEF system in detail and evaluates the relative contributions of the different perturbation techniques. The ensemble variational blending technique and the ensemble surface data assimilation provide additional spread in the first forecast hours, while the hybrid stochastically perturbed parametrization scheme improves the performance of C-LAEF during the whole forecast range. The performance of C-LAEF is evaluated extensively over one summer and one winter period and compared with its mesoscale counterpart ALADIN-LAEF (Aire Limitee Adaptation dynamique Developpement InterNational – Limited-Area Ensemble Forecasting) and the IFS-ENS. State-of-the-art probabilistic measures indicate that C-LAEF is able to outperform ALADIN-LAEF for all considered upper-air variables. At the surface, C-LAEF outperforms ALADIN-LAEF and IFS-ENS according to most conventional measures, particularly for wind and precipitation. C-LAEF benefits from...
the higher resolution and the explicit treatment of deep convection and can provide more accurate probabilistic information for weather warnings.

**KEYWORDS**
convection-permitting, ensemble system, error representation, verification

### 1 INTRODUCTION

The Austrian national weather service ZAMG (Zentralanstalt für Meteorologie und Geodynamik) started to work on limited-area ensemble prediction systems (EPSs) in 2008 by developing the ALADIN-LAEF (Aire Limitee Adaptation dynamique Developpement International – Limited-Area Ensemble Forecasting; Wang *et al.*, 2011) system. At that time, only a few national weather services in Europe had their own regional EPS, mainly the big “players” such as DWD (Deutscher Wetterdienst; COSMO-DE-EPS; Baldauf *et al.*, 2011), UK Met Office (MOGREPS; Bowler *et al.*, 2008) and Météo France (PEARO; Bouttier *et al.*, 2012). The development, implementation, and maintenance of a high-resolution ensemble system is very demanding on both human and computer resources, and many small- to medium-sized national weather services do not have the capacity to do this on their own (Wang *et al.*, 2018). Therefore, the development of the ALADIN-LAEF system has been done within the framework of the international cooperation of the ALADIN and RC LACE (Regional Cooperation for Limited Area modelling in Central Europe) consortia. ALADIN-LAEF was put into operation in 2011, received a major upgrade in 2013, and is still running operationally. The code is based on the hydrostatic version of the spectral limited-area model ALADIN (Termonia *et al.*, 2018). The operational setup of ALADIN-LAEF consists of 16 perturbed members (plus one unperturbed control run) with a horizontal resolution of 11 km and 45 vertical levels. It provides probabilistic forecasts up to 72 hr ahead twice a day (0000 and 1200 UTC). It has been upgraded to 5 km and 60 vertical levels in summer 2020.

In recent years, the horizontal resolution of many deterministic limited-area models (LAMs) has entered the convection-permitting scale (<3 km), and this is also the trend in ensemble prediction (Romine *et al.*, 2014). The main purpose of a convection-permitting EPS is to provide important and reliable information on predictability and on possible weather scenarios, with a special focus on local convective weather hazards (Schellander-Gorgas *et al.*, 2017). Such systems are becoming more widely used forecasting tools and are therefore operated by several national meteorological services in the meantime. PEARO (Bouttier *et al.*, 2012), COMSO-E (Klasa *et al.*, 2018), MOGREPS (Hagelin *et al.*, 2017), or HarmonEPS (Frogner *et al.*, 2019a) are only some examples of current operational convection-permitting EPSs in Europe. ZAMG has started to devote research to build up the Convection-permitting Limited-Area Ensemble Forecasting system (C-LAEF) in 2016, and it finally became operational in November 2019. C-LAEF is based on the non-hydrostatic spectral model AROME (Applications of Research to Operations at Mesoscale; Seity *et al.*, 2011) and is run on a 2.5-km grid covering large parts of Central Europe (Figure 1).

A complete ensemble system needs to address all potential sources of uncertainties in all parts of the forecasting system to provide useful probabilistic forecasts (Wastl *et al.*, 2019a). These include inaccuracies in the initial conditions, in the representation of surface conditions, and approximations in the model formulation and its physical parametrizations (model error). In the case of LAMs, nesting into a global model that provides lateral boundary conditions poses an additional source of uncertainty. There is a great variety of methods that have been developed over the years to best account for all these uncertainties in a convection-permitting EPS. Frogner *et al.* (2019a) implemented and tested different methods within HarmonEPS. They found that initial condition and LBC perturbations are very important to achieve enough spread in the first forecast hours, while surface perturbations and model error representations have a continuous impact on the scores during the whole forecasting range. Hagelin *et al.* (2017) evaluated different configurations for MOGREPS, while Klasa *et al.* (2018) focused on the effect of perturbations on precipitation forecasts with the COSMO-E system of Switzerland.

To account for all these kinds of uncertainties in C-LAEF, some innovative methodologies have been developed and tested at ZAMG in the past years:

1. An ensemble 3D variational blending technique (ensemble $J_k$; Keresturi *et al.*, 2019) to deal with atmospheric initial uncertainties. By blending large-scale perturbations from the driving IFS-ENS (Integrated Forecasting System—ENSEmble; ECMWF, 2018) directly into the three-dimensional variational data assimilation (3D-Var) of C-LAEF, the perturbation
FIGURE 1 The location of the C-LAEF domain in Europe (upper panel) and a zoom (lower panel) including the topography in m (colour), frontiers (black lines), and coastlines (gray lines). The red box in the lower panel shows the INCA domain for precipitation verification [Colour figure can be viewed at wileyonlinelibrary.com]

mismatch between the LAM and its global counterpart is reduced.

2. An ensemble of land surface data assimilation scheme where each member of the ensemble sees different stochastic surface observation perturbations (ESDA; Bellus et al., 2016).

3. A hybrid stochastic physics perturbation scheme (HSPP; Wastl et al., 2019b) to represent the model error. HSPP combines the advantages of a tendency (simple, effective) and a parameter perturbation approach (high physical consistency) by applying individually adapted stochastic perturbations directly to the different physics parameterization schemes.

The aim of the present article is to give an overview of the current operational configuration of C-LAEF with special focus on the different methodologies of error representation and their relative impact (Section 2). The performance of the full C-LAEF system is evaluated in Section 3 by an extensive observation-based verification over a summer and winter period. The global ensemble system (IFS-ENS) of the ECMWF (European Centre for Medium-Range Weather Forecasts; ECMWF, 2018) and the mesoscale counterpart of C-LAEF (ALADIN-LAEF; Wang et al., 2011) are taken as references. The article concludes with a discussion of the results and an outlook on planned upgrades of C-LAEF in Section 4.
2 | METHODOLOGY

2.1 | C-LAEF system configuration

The C-LAEF system has been developed at ZAMG and became fully operational in November 2019. It is running on the high-performance computer facilities at the ECMWF. The underlying limited-area model of C-LAEF is the convection-permitting model AROME (Seity et al., 2011), which is under active development within the ALADIN (Termonia et al., 2018), HIRLAM (High Resolution Limited Area Model; Bengtsen et al., 2017), and RC LACE (Wang et al., 2018) consortia. The AROME model has a non-hydrostatic dynamical kernel that has been initially developed for the ALADIN model (Bubnova et al., 1995). It solves the non-hydrostatic Eulerian equations in a mass-based vertical coordinate with semi-Lagrangian advection and semi-implicit time stepping (Bénard et al., 2010). The physics package used in AROME is mainly adopted from the research model Meso-NH (Mascart and Bougeault, 2011). It uses a one-moment bulk microphysical scheme (ICE 3; Pinty and Jabouille, 1998) and a 1D 1.5-order turbulence scheme (Cuxart et al., 2000). Shallow convection is described by a mass-flux type scheme with turbulence closure (Pergaud et al., 2009). A deep convection scheme is not used because it is assumed to be explicitly resolved by the dynamics. AROME uses a three-layer version of the surface scheme SURFEX (Surface Externalisé; Masson et al., 2013) with different subschemes for the tiles nature, town, sea, and lake. The radiation scheme of AROME is taken from the ECMWF IFS model (Fouquart and Bonnel, 1980; Mlawer et al., 1997). AROME has been used as the main operational NWP system at ZAMG since 2014.

C-LAEF comprises 16 perturbed members plus one unperturbed control. Boundary conditions for the perturbed members are taken every 3 hr from the first 16 out of a total of 51 members of the global ensemble system IFS-ENS by using a Davies relaxation scheme (Davies, 1976). In this scheme, the interior flow near the boundaries is relaxed to the external fully prescribed flow on a zone with an eight-point width along the physical domain border. The unperturbed control is coupled to the deterministic IFS-HRES (IFS-High RESolution; ECMWF, 2018) and serves as a backup for the deterministic AROME forecast at ZAMG. Coupling is not in real time but 6 hr lagged, which means, for example, that the 0000 UTC C-LAEF run is coupled to the 1800 UTC run of the IFS. Lag is introduced for operational reasons to guarantee a prompt initialization of the C-LAEF runs. Each C-LAEF member has its own 3D-Var using a wide range of conventional and nonconventional observations (Seity et al., 2011) in a −1.5 hr to +1.5 hr window. At the surface, screen level temperature and relative humidity are assimilated using an optimal interpolation method (Giard and Bazile, 2000). Snow cover is initialized by using the high-resolution (100 m) output of the ZAMG snow cover model SNOWGRID (Olefs et al., 2013).

The background error covariance matrices for C-LAEF have been created using a full C-LAEF ensemble over one summer and one winter month. The method of creation follows Berre (2000) and is formulated in spectral space and computes climatological background error statistics for vorticity, divergence, temperature, surface pressure, and specific humidity. Scale-dependent statistical regressions are used to estimate cross covariances.

C-LAEF is run on a 2.5-km horizontal grid with 90 vertical levels and a time step of 60 s. The integration domain covers an area of 1,500 × 1,080 km and is centred on Austria (Figure 1). Four C-LAEF runs a day are initialized at 0000, 0600, 1200, and 1800 UTC. The forecast range depends on the run and varies between +6 hr (0600 and 1800 UTC), +48 hr (1200 UTC), and +60 hr (0000 UTC). This is a compromise between the user requirements and the computational resources available at the ECMWF supercomputer facilities. The 0000 UTC run has the longest forecast range to meet the user requirements. The short intermediate runs (0600 and 1800 UTC) are not used for forecast purposes; they only serve to keep the assimilation cycle running.

The output frequency of C-LAEF forecasts is 1 hr, and the files are transferred from the ECMWF to ZAMG for postprocessing and visualization. Around 4 hrs after initialization time, the full C-LAEF data set is available at ZAMG. C-LAEF output is also the basis for the ensemble nowcasting and analysis system INCA (Integrated Nowcasting through Comprehensive Analysis; Haiden et al., 2011; Wang et al., 2017). To guarantee constant data availability for the users, a C-LAEF backup has been implemented. This backup consists of IFS-ENS forecast fields interpolated to the C-LAEF domain and resolution. It is running at the HPC at ZAMG and is used if a C-LAEF run is delayed for more than 1 hr.

2.2 | Perturbations

2.2.1 | Lateral boundary conditions

Lateral boundary coupling (LBC) is done by nesting each C-LAEF member within a different member of the global ensemble IFS-ENS. In this way, perturbed LBCs are naturally included since the different members of the driving EPS are perturbed individually. The initial condition perturbations in the IFS-ENS are created by a combination of the EDA (ensemble of data assimilation) and SV (singular vectors) techniques (ECMWF, 2018). Clustering or similar
methods to control the spread induced from the boundaries (e.g., Molteni et al., 2001) are not used in C-LAEF yet, but will be investigated in future. This means that C-LAEF member number 1 is at the moment always nested in IFS-ENS member 1 and so on.

2.2.2 Initial condition perturbations

It is important that the initial condition perturbations are consistent with the LBC perturbations. Using different and independent perturbations in the limited-area EPS (e.g., EDA) versus the perturbations in the host EPS (e.g., singular vectors) can lead to a conflict at the lateral boundaries generating spurious gravity waves that spread into the rest of the domain (Caron, 2013; Wang et al., 2014). Furthermore, typical initial perturbation methods applied in a LAM are usually not able to represent sufficient uncertainties in the large-scale flow because of the typically limited-size domain, less effective data assimilation, etc. (Gustafsson et al., 2018). This issue can be reduced by applying blending methods (e.g. Brožková et al., 2001) or by including global model information directly into a limited-area variational analysis (Guidard and Fischer, 2008). The latter study has motivated the development of an ensemble Jk method at ZAMG that is directly applied to the 3D-Var EDA of C-LAEF (Keresturi et al., 2019). The EDA in C-LAEF is created by running one 3D-Var analysis for each ensemble member with separately perturbed observations. The perturbations are Gaussian distributed with variances that are consistent with observational statistics for each instrument type (Bouittier et al., 2012). By applying the ensemble Jk method, the small-scale perturbations generated from 3D-Var EDA are directly blended with the large-scale perturbation generated from the driving IFS-ENS. The cost function $J$ of one ensemble member $i$ of the ensemble Jk method can be written as follows:

$$J_i(x_i) = \frac{1}{2}(x_i - x_{h,i})^T B^{-1} (x_i - x_{h,i}) + \frac{1}{2}(y_i - Hx_i)^T R^{-1} (y_i - Hx_i) + \frac{1}{2}(H(x_i - H_kx_{h,i}))^T V^{-1} (H(x_i - H_kx_{h,i}))$$

(1)

where $x$ is the C-LAEF state vector, $x_{h}$ the C-LAEF background, $y$ the observation and $H, H_1,$ and $H_k$ are nonlinear operators. $B, R,$ and $V$ are the error covariance matrices of the C-LAEF background, the observations, and the host model, respectively. Ensemble Jk method increases the skill and reliability of the C-LAEF ensemble in the first hours and reduces the perturbation mismatch between the LAM and its global counterpart (Keresturi et al., 2019). For the land surface data assimilation in C-LAEF, an EDA with observation perturbations is applied to the surface parameters 2 m temperature and 2 m humidity (ESDA, Bellus et al., 2016).

Inflation of ensemble Jk and ESDA perturbations is not applied in C-LAEF. The respective perturbed initial state is then the starting point for each ensemble member’s integration.

2.2.3 Model perturbations

The model error representation in the current operational C-LAEF ensemble is based on a hybrid system (HSPP; Wastl et al., 2019b) that combines a tendency perturbation approach with a parameter perturbation scheme. In HSPP, the partial tendencies of temperature, humidity, and wind of the physics parametrisation schemes of radiation, shallow convection, and microphysics are perturbed by applying a random noise created by a stochastic pattern generator (Wastl et al., 2019a). The horizontal and temporal structure of the stochastic pattern is adjusted individually to the uncertainty scale of the different physics parametrisations. This means that, for example, perturbations applied to the shallow convection scheme are much larger than for the microphysics scheme. The physics schemes are called sequentially in the C-LAEF system, and the tendencies are perturbed directly after the call of each scheme. This implies that the tendencies are passed through the different schemes together with the perturbations. In this way, a physically consistent relationship between the different parametrisation schemes is kept and the stability of the perturbed model is guaranteed (Wastl et al., 2019b). In the case of a possible supersaturation due to positive humidity or negative temperature perturbations, an iterative method of perturbation reduction is applied (Szűcs, 2016).

Uncertainty representation in the turbulence scheme is achieved by perturbing preselected parameters instead of tendencies. This method is based on the SPP scheme (Ollinaho et al., 2017) and has been chosen because tendency perturbations can induce model instabilities in the very sensible turbulence scheme. The selection of the parameters to be perturbed was proposed by physics experts who also provided a meaningful perturbation range. The six selected parameters address processes like dissipation, turbulent mixing, or wind-pressure correlation. The perturbation of the parameters is done by applying individually adapted stochastic fields following a Gaussian distribution with zero mean. The perturbed parameters $P_i'$ are obtained by multiplying the unperturbed parameters $(P_i)$ by the exponential of the
The advantage of such process-based perturbations is a much higher physical consistency, since uncertainties are treated directly at their origin (Leutbecher et al., 2017). In this way, the conservation laws of energy and moisture are respected, which is not the case for solely tendency perturbations methods such as SPPT (stochastically perturbed parametrization tendencies; Palmer et al., 2009).

The combination of tendency and parameter perturbations in HSPP means that the tapering function is no longer needed. The tapering function reduces the perturbations in the upper- and lowermost parts of the atmosphere and has been implemented in many ensembles with tendency perturbations (e.g., IFS-ENS, PEARO) to avoid numeric instabilities. HSPP combines the strength of both schemes: the simplicity and effectiveness of tendency perturbations and the physical consistency of parameter perturbations.

At the moment, stochastic perturbations in the surface scheme during model integration are not applied in the operational C-LAEF setup. However, perturbations of the key parameters in SURFEX following Bouttier et al. (2016) are planned in the near future.

2.3 C-LAEF setup and verification methods

2.3.1 Test phase

To assess the impact of the different error representation methods in C-LAEF, several experiments are run for a 1-month test period in July 2016. These experiments include the following: (a) a C-LAEF experiment based on a pure downscaling of the IFS-ENS initial conditions without any additional error representation scheme (REF), (b) C-LAEF with initial condition perturbations (ensemble Jk blending method plus EDA and ESDA) and no model error representation scheme (Init), (c) C-LAEF with HSPP stochastic physics scheme and no initial condition error representation scheme (Stophy), and (d) C-LAEF with initial condition perturbations and HSPP (Combined). During this test phase, C-LAEF was running with a forecast range of 30 hr for the 0000 UTC run and 6 hr for the other runs (0600, 1200, and 1800 UTC) to keep the assimilation cycle running.

The verification of the experiments is done using the common software package “harp” (HIRLAM-ALADIN R Package, https://github/harphub) developed within the ALADIN and HIRLAM consortia. Surface verification of sea-level pressure, 2 m temperature, relative humidity, and 10 m wind is done against point observations at all SYNOP and TAWES (automated weather station network of ZAMG) observation sites (around 300) in Austria. For this purpose, the forecast values are bilinearly interpolated to the observation sites (Weidle et al., 2013). In the case of 2 m temperature, a height correction (6.5 K·km⁻¹) is applied to account for the elevation differences between model surface and the observation site. For precipitation, a grid-based verification is applied by using the INCA analysis system developed at ZAMG. It combines rain gauge measurements and radar data by a sophisticated algorithm on a 1-km grid (Haiden et al., 2011). The ability of the INCA radar–rain gauge merging algorithm to precisely analyze precipitation patterns has been tested in several studies in the past (Haiden et al., 2011; Kann et al., 2014; 2015a). The quality of such a grid-based precipitation analysis system strongly depends on the density and representativeness of the stations in use and the quality of the radar network (Kann et al., 2015b). However, a spatial verification is reasonable in the present study because in such complex terrain as the Alps, precipitation variability is high and a station network alone is insufficient for a representative verification. The precipitation fields of the considered ensembles are bilinearly interpolated to the INCA grid for the domain shown in Figure 1 (red box). The computation of probabilities in dependence of different thresholds for some probabilistic scores is done separately for the three EPSs, but if one of them does not reach the threshold at any pixel in the domain, all three ensembles are not used for this time. The upper-air weather variables of temperature, relative humidity, wind, and geopotential are verified against IFS-HRES analyses at the 500 and 850 hPa levels. It was decided to use IFS-HRES analyses instead of radiosoundings because only four locations with radiosoundings are available in Austria, which is far too few for a robust statistical verification.

2.3.2 Operational phase

To assess the performance and quality of the full C-LAEF system in different seasons, classical deterministic and probabilistic verification scores have been calculated for a summer (August 20, 2019–September 20, 2019) and a winter period (February 20, 2020–March 20, 2020). The first verification period in August/September 2019 has been chosen because at that time the preoperational C-LAEF mode had just started at the ECMWF HPC and it was the first time our forecasters received C-LAEF products. The second period in February/March 2020 has been chosen because of the need for a comparable period in winter.
The verification of C-LAEF is done against the reference systems ALADIN-LAEF (Wang et al., 2011), and the IFS-ENS from the ECMWF (ECMWF, 2018). Due to the fact that verification is performed against ECMWF analyses, the IFS-ENS is not included for the upper-air verification. Standard probabilistic scores (spread, root-mean-square error, Brier score, continuous ranked probability score, etc.) are calculated for the 0000 UTC run with a forecast range of 48 hr for the summer and 60 hr for the winter period (60 hr forecast range was not yet available in the summer period). A more detailed description of the probabilistic metrics used can be found in Wilks (2011). Statistical significance of the score differences between the three ensembles is determined by a moving-block bootstrap technique with a block size of 3 days and 10,000 re-samples. If the sign of the differences between the experiments is not contradicted by more than 10% of the sample, the score difference is deemed statistically significant. The significance information is not explicitly shown in the figures in order to keep the figures clear, but the statement “statistically significant” is given in the text if this criterion is met. Observation errors are not taken into account in the verification procedure, since the goal of this article is a comparison of the different ensembles rather than an assessment of their absolute performance.

3 RESULTS

3.1 Impact of the different error representation methodologies

Figure 2 shows the surface verification results for all three C-LAEF experiments during the test phase in July 2016. July 2016 was characterized by a generally strong convective activity in the target area with a lot of thunderstorms and high precipitation amounts. In the first half of the month, a very deep trough over the British Islands caused an extensive southwesterly flow in Central Europe resulting in a strong advection of warm and moist air masses towards the Alps. In the second part of July 2016, a very weak pressure gradient was established over Central Europe causing a lot of small-scale and stationary thunderstorms.

The gray line in Figure 2 refers to the “REF” experiment, the black line to “Init”, the blue line to “Stophy”, and the red one to “Combined”. To highlight the impact of the different methodologies, ensemble spread (dashed) and root-mean-square error of the ensemble mean (RMSE, solid) in panel (a) to (d) are given as average difference to REF, while the panel (e) and (f) show absolute numbers. For mean sea-level pressure (a), 2 m temperature (b) and relative humidity (c) and 10 m wind speed (d), the impact

![Figure 2](https://wileyonlinelibrary.com)
of initial condition perturbations on the ensemble spread is very high at initialization time and rapidly decreases with lead time. Initial condition perturbations in C-LAEF are very effective since they consist of a combination of three perturbation methods (ensemble Jk blending, EDA, and ESDA). Such characteristics with rapidly decreasing spread caused by initial condition perturbations in the first hours can also be found in other convection-permitting EPSs such as AROME-EPS (Bouttier et al., 2012) or HarmonEPS (Frogner et al., 2019a). However, an increased ensemble spread is desirable, since the C-LAEF ensemble suffers from a lack of spread during the whole forecasting range (panel e and f in Figure 2).

On the other hand, the impact of the HSPP scheme on the ensemble spread is rather small at the beginning and increases with lead time. The maximum gain of spread for the experiments with stochastic physics can be found in the afternoon and evening hours when convection is very active in summer (Wastl et al., 2019b). The magnitude of the additional spread due to model error representation is smaller for most surface variables than the impact of initial condition perturbations. These findings are confirmed by other studies investigating the impact of different perturbation methods in a convection-permitting EPS (e.g., Hagelin et al., 2017; Frogner et al., 2019b). The experiment with the highest ensemble spread for all lead times in Figure 2 is the combination of initial condition perturbations and stochastic physics (red line). The planned stochastic perturbations in the surface scheme SURFEX in C-LAEF should add some additional ensemble spread during the whole forecasting range (Bouttier et al., 2016).

The RMSE reveals much smaller differences between the experiments. Initial condition perturbations slightly increase RMSE of 2 m temperature and mean sea-level pressure in the first hours and decrease it later. However, this increase of RMSE at the beginning is much smaller than the gain in ensemble spread. For 2 m relative humidity and 10 m wind speed, initial condition perturbations slightly reduce the average RMSE of the ensemble mean. HSPP has generally a very small impact on the RMSE during the whole forecast range.

The effect of the different error representation methods on upper-air variables such as geopotential, humidity, or wind speed is similar to the surface (not shown): The impact of initial condition perturbations is very high in the first hours, while HSPP has a positive impact during the whole forecast range.

### 3.2 C-LAEF verification during the summer period

The considered period from August 20 to September 20, 2019 can be divided into two parts from a meteorological point of view. The first 2 weeks were characterized by a southwesterly flow over Austria advecting hot and moist air from the Mediterranean Sea and causing an unstable situation with several thunderstorms in the Alps. The second half of the period was dominated by a strong high-pressure system over Austria with only very few rain events.

RMSE of the ensemble mean (solid) and ensemble spread (dashed) of six selected surface parameters for this summer period are given in Figure 3. For mean sea-level pressure (a), the scores of the three ensembles are quite close without significant differences. Ensemble spread is continuously increasing with the lead time, which is not surprising since the forecast uncertainties are increasing with time as well. The peak of spread at initialization time for ALADIN-LAEF (blue) and C-LAEF (red) is caused by the surface observation perturbation schemes. Overall, ALADIN-LAEF shows the highest ensemble spread but also the highest RMSE. IFS-ENS and C-LAEF are behaving very similarly. RMSE has a clear diurnal cycle with highest values in the afternoon hours when predictability is generally lower in summer because of convection processes (Wastl et al., 2019b).

The differences between the three considered ensembles are much stronger when looking at the surface parameters 2 m temperature (b) and 2 m relative humidity (c). For both parameters, ALADIN-LAEF offers the highest ensemble spread but also the highest RMSE, followed by C-LAEF and IFS-ENS. In the case of C-LAEF, the RMSE is quite high during the night, especially for relative humidity. A detailed investigation of this feature indicated that C-LAEF suffers from a warm and dry bias during the night, especially in the Alpine valleys in the western part of the domain (not shown). The formation of cold air pools and thus the decoupling of the surface boundary layer during the night in the valleys is not well represented in C-LAEF, which results in a warm and dry bias. A significant improvement could be reached by replacing the digital terrain elevation data SRTM (Jarvis et al., 2008) with GMTED (Danielson and Gesch, 2010) and applying a Gaussian filter. This change had been made before the full operationalization of C-LAEF in November 2019, and therefore the problem is much smaller in the winter period in Section 3.3. The main reason for the bad 2 m scores in C-LAEF is the use of the CANOPY surface boundary layer scheme (Masson and Seity, 2009). This scheme causes problems with 2 m diagnostics for high-resolution models in areas with steep terrain (Teixeira et al., 2016). C-LAEF test runs without CANOPY scheme have shown promising results in the Alpine valleys, but a slight score degradation in the flat areas. A further investigation of this problem is currently ongoing at ZAMG.
C-LAEF shows the best performance of all three considered ensemble systems for the $u$ and $v$ components of the 10 m wind in Figure 3(d), (e). The ensemble spread is significantly higher and RMSE lower than for ALADIN-LAEF and IFS-ENS throughout the whole forecast range, indicating increased ensemble reliability. In particular, the small-scale wind features (not shown) in the complex topography of the Alps are much better represented in the higher-resolved (2.5 km) convection-permitting C-LAEF than in ALADIN-LAEF (11 km) and IFS-ENS (18 km). The explicit treatment of convection in C-LAEF has also a positive impact on the hourly accumulated precipitation scores, as can be seen in panel (f) of Figure 3. Both RMSE and spread of precipitation reveal a strong diurnal cycle with the highest values in the afternoon when small-scale convective precipitation and thunderstorms are dominating in this season. While RMSE is very similar for all three ensembles, ensemble spread is significantly higher in C-LAEF than in ALADIN-LAEF and IFS-ENS. Furthermore, in the case of C-LAEF, spread and RMSE are very close throughout the whole forecast range, which is desirable in a consistent ensemble system.

The findings presented here are comparable to those of other studies on convection-permitting EPSs. Frogner et al. (2019a) for example, found that different configurations of the HarmonEPS are able to produce higher spread with fewer members than IFS-ENS, and mostly better RMSE for precipitation. Klasa et al. (2018) investigated the performance of the COSMO-E ensemble of Switzerland in comparison with the global IFS-ENS for three contrasting precipitation events. They concluded that COSMO-E is able to outperform IFS-ENS in all cases in terms of precipitation pattern but the precipitation amounts (RMSE) are very similar in both ensembles.

Calculating ensemble spread and RMSE separately for the first (wet and unstable) and second part (dry and stable) of the summer period reveals strong differences in the forecast performances (not shown). During the first part, ensemble spread and RMSE values are generally high and the differences between the three ensemble systems are strong, with C-LAEF showing the best scores. In the second part, predictability is generally much higher and the differences between the ensemble systems are rather small.

Many convection-permitting EPSs are suffering from too little spread compared to RMSE (e.g. Chen et al., 2018). Also, C-LAEF is underdispersive for many surface parameters, as can be clearly seen in the rank histogram in Figure 4. The rank histogram divides observations into
bins of ranked ensemble members. A clear U-shape is present for most shown variables, which means that the observations are often outside the spread of the ensembles or, in other words, that the spread of the ensembles is too small. The numbers in Figure 4 are given as normalized frequency, whereby the perfect score in each of the 18 bins would be 1. For mean sea-level pressure (a) and temperature (b), there is an indication of a positive bias for all three considered ensemble systems with the largest proportion of observations ranked below all ensemble members. In the case of 10 m wind speed (c), no bias is obvious in the rank histogram. The most uniform distribution among the ranks for this parameter is present in C-LAEF, indicating highest reliability, while IFS-ENS has the strongest U-shape. For precipitation (d), the picture is quite different between the models. While ALADIN-LAEF and especially C-LAEF tend to underestimate the precipitation amounts, no bias but a clear U-shape is obvious for IFS-ENS. Observation errors are not taken into account here, but it can be assumed that the forecast ranks would be more evenly dispersed in this case (Frogner et al., 2019a).

Another possibility of assessing the quality of an ensemble is the continuous ranked probability score (CRPS), which is shown for selected parameters in Figure 5. This negatively oriented score measures the skill of the ensemble mean forecast as well as the ability of the perturbations to capture the deviations around it (Bowler et al., 2008). For the parameters mean sea-level pressure (a) 2 m temperature, (b) wind (d and e), and precipitation (f), C-LAEF is able to outperform ALADIN-LAEF and IFS-ENS for almost all lead times. Especially in the first 6 hrs, the additional skill of C-LAEF compared with the other ensembles is very high and statistically significant. In this period, C-LAEF benefits from ensemble Jk and more current observations since it uses a 6-hourly assimilation cycle, while ALADIN-LAEF, for example, only has two runs per day.

Another probabilistic score that is especially useful for verifying dichotomous variables such as precipitation is the Brier score. Figure 6 shows the Brier score for a threshold of 0.1 mm·hr\(^{-1}\) (a) and 5.0 mm·hr\(^{-1}\) (b) in this summer period. The score is negatively oriented, which means that smaller values of the score indicate better forecasts. For precipitation events with more than 0.1 mm·hr\(^{-1}\) in Figure 6(a), C-LAEF shows the best scores, followed by ALADIN-LAEF and IFS-ENS. For heavier precipitation events with more than 5 mm·hr\(^{-1}\), the situation is not so clear and the three ensembles perform quite similarly. Precipitation events in summer are generally dominated by convection and therefore a strong diurnal cycle is present, as can be seen in panel (c) of Figure 6. The highest observed hourly precipitation intensity averaged over the whole domain and summer period occurs between 1400 and 1700 UTC (about 0.2 mm·hr\(^{-1}\)), while the lowest values of precipitation intensity can be found in the morning hours between 0500 and 1000 UTC.
FIGURE 5  CRPS as a function of lead time in hr for (a) mean sea-level pressure (hPa), (b) 2 m temperature (K), (c) 2 m relative humidity (%), (d) \( u \) and (e) \( v \) component of 10 m wind (m s\(^{-1}\)), and (f) 1-hourly accumulated precipitation (mm) in the summer period [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 6  Brier score as a function of lead time in hr for 1-hourly accumulated precipitation in the summer period for the thresholds (a) 0.1 mm hr\(^{-1}\) and (b) 5 mm hr\(^{-1}\). Panel (c) shows the diurnal cycle of the observed precipitation intensity (mm hr\(^{-1}\)) for the summer period [Colour figure can be viewed at wileyonlinelibrary.com]
One of the main applications of convection-permitting EPSs at the meteorological services is the prediction of strong convection and thunderstorms. To compare the performance of C-LAEF with the coarser resolved ensembles of ALADIN-LAEF and IFS-ENS for such events, the area under the Relative Operating Characteristic curve (ROC_auc; Wilks, 2011) is shown in Figure 7 for four different precipitation thresholds. This score is a measure of the potential usefulness of a forecast and measures the ability of the forecast to discriminate between two alternative outcomes defined by a threshold. The ROC_auc score can reach values between 0 and 1, with 0.5 indicating no skill and 1 indicating a perfect forecast. Figure 7 shows that for small precipitation thresholds (0.1 and 0.5 mm hr⁻¹) the skill of the three ensembles is very similar, but for stronger precipitation events (1.0 and 5.0 mm hr⁻¹) C-LAEF is able to outperform the other two ensembles for most lead times.

Another score to assess the quantitative precipitation performance of an ensemble is the reliability diagram, which plots the observed frequency against the forecast probability. We have calculated the reliability diagram for four different thresholds (0.1, 0.5, 1 and 5 mm hr⁻¹) and found a general tendency towards overforecasting (too high probabilities) for all three ensembles. The benefit of C-LAEF is especially pronounced for the stronger precipitation events with more than 5.0 mm hr⁻¹ (not shown).

To assess the performance of the different ensembles at higher levels in the atmosphere, verification scores have been calculated for upper-air variables at 850 and 500 hPa. Figure 8 shows the CRPS for the parameters temperature (a), relative humidity (b), wind speed (c), and geopotential (d) at 500 hPa. IFS-ENS is not included in the upper-air verification, because ECMWF analyses are used as a reference and therefore a comparison with the other two ensemble systems would not be fair. Figure 8 shows that C-LAEF is able to outperform the ALADIN-LAEF ensemble system for the considered parameters at almost all lead times. Only for temperature at +18 and +42 hr and relative humidity at +18 hr of lead time does ALADIN-LAEF have a slightly lower CRPS. The CRPS values generally increase with lead time, which indicates that the uncertainties increase and predictability decreases with time. This is not completely the case for the geopotential where CRPS is higher at initialization time than in the subsequent hours. This feature originates in the ensemble data assimilation, which induces perturbations of geopotential at initialization time. However, especially in the case of ALADIN-LAEF, the perturbations seem to be too strong because RMSE is extremely high at that time (not shown). A diurnal cycle is not visible in Figure 8, which is not surprising since the 500 hPa level is at a height of about 5,500 m and thus far away from the strong diurnal variability of surface processes.

**Figure 7** Area under the ROC curve (ROC_auc) as a function of lead time in hr for 1-hourly accumulated precipitation in the summer period for the thresholds of (a) 0.1 mm hr⁻¹, (b) 0.5 mm hr⁻¹, (c) 1 mm hr⁻¹, and (d) 5 mm hr⁻¹. [Colour figure can be viewed at wileyonlinelibrary.com]
The RMSE of the ensemble mean and ensemble spread at 850 hPa are shown for C-LAEF and ALADIN-LAEF in Figure 9. ALADIN-LAEF exhibits a significantly higher spread for temperature (a), especially from a lead time of +10 hr onwards, but on the other hand also the RMSE is higher. This is not the case for relative humidity (b), where C-LAEF shows better performance with a higher spread and smaller RMSE values for all lead times. Also for wind speed (c) the ensemble spread is significantly higher for C-LAEF. The influence of perturbing initial conditions on the verification scores can be clearly seen in Figure 9 and is similar to other studies (e.g., Bouttier et al., 2012). The RMSE of the ensemble mean and ensemble spread are for most parameters higher at initialization time than at +6 hr. However, these perturbations are necessary to cope with the uncertainties in the observations and data assimilation procedure. ALADIN-LAEF and C-LAEF are underdispersive at both considered levels, as can be seen in that the RMSE is higher than the ensemble spread. When comparing the scores at 500 and 850 hPa it becomes obvious that the predictability at 500 hPa is generally higher, which is related to an overall higher stability at upper levels. In other words, the influence of small-scale surface processes and mainly the impact of the complex topography on the flow is much smaller at 500 hPa than at 850 hPa.

### 3.3 C-LAEF verification during the winter period

The selected winter period from February 20, 2020 to March 20, 2020 was characterized by a strong westerly flow and several low-pressure systems and associated fronts passing Central Europe. Therefore, precipitation occurred predominantly in the western and northern parts of the domain, while the south (in the lee of the Alps) and the east were rather dry. Because of the high temperatures, snow cover was present only at the higher altitudes (>1,000 m) of the Alps.

The C-LAEF system has undergone two major changes between the considered summer and winter period: a change in the model orography from unfiltered SRTM to filtered GMTED (due to bad 2 m scores of temperature and humidity) and an extension of the forecast range from 48 to 60 hr for the 0000 UTC run. Several deterministic and probabilistic verification scores have been calculated for surface and upper-air variables for this winter period. Figure 10 shows the RMSE of the ensemble mean (solid) and ensemble spread (dashed) for mean sea-level pressure (a), 2 m temperature (b) and humidity (c), u (d), and v (e) components of 10 m wind and precipitation (f). For mean sea-level pressure, ALADIN-LAEF shows the highest RMSE, especially at initialization time and
**FIGURE 9** RMSE of the ensemble mean (solid lines) and ensemble spread (dashed lines) at 850 hPa as a function of lead time in hr for (a) temperature (K), (b) relative humidity (%), (c) wind (m⋅s$^{-1}$), and (d) geopotential (m$^2$⋅s$^{-2}$) in the summer period. [Colour figure can be viewed at wileyonlinelibrary.com]

**FIGURE 10** RMSE of the ensemble mean (solid lines) and ensemble spread (dashed lines) as a function of lead time in hr for (a) mean sea-level pressure (hPa), (b) 2 m temperature (K), (c) 2 m relative humidity (%), (d) $u$ and (e) $v$ component of 10 m wind (m⋅s$^{-1}$), and (f) 1-hourly accumulated precipitation (mm) in the winter period. [Colour figure can be viewed at wileyonlinelibrary.com]
from 30 hr onwards, while RMSE for C-LAEF and IFS-ENS are quite similar. The ensemble spread is highest for ALADIN-LAEF especially at the beginning, while IFS-ENS has the lowest spread over the complete forecast range. The situation is quite similar for the 2 m temperature, where ALADIN-LAEF shows the highest spread, followed by C-LAEF and the IFS-ENS, while RMSE is highest for IFS-ENS. All three ensembles generally suffer from a too small diurnal cycle of 2 m temperature with a positive bias during the night and a negative one during the day (not shown). This feature is especially pronounced in the mountainous areas, as a verification of stations at different altitudes has shown. In the case of 2 m relative humidity (c), ALADIN-LAEF has very high RMSE values during the day caused by a positive humidity bias. The warm and dry bias of the summer period in C-LAEF is still present in winter, but is of much smaller extent. For 10 m wind and precipitation, C-LAEF outperforms the other two ensembles with significantly higher spread and lower RMSE values for most lead times. The fact that higher resolved regional EPSs are able to outperform their global counterparts especially for parameters with high spatial and temporal variability such as wind or precipitation is also the outcome of similar studies (Duc et al., 2013; Hagelin et al., 2017; Klasa et al., 2018; Wang et al., 2020).

However, all three ensembles in the present study generally suffer from too little spread compared with the RMSE at the surface. This underdispersive situation becomes obvious in the normalized rank histograms, which are all of U-shape (not shown). CRPS, which is a good indicator of the overall performance of an ensemble, shows that C-LAEF performs best for wind and precipitation while for 2 m temperature and mean sea-level pressure the scores of the three ensembles are similar. The Brier score is shown in Figure 11 to evaluate the probabilistic performance of the ensembles for different precipitation thresholds. For all precipitation events in Figure 11 (a), the Brier score of C-LAEF is significantly lower than that of ALADIN-LAEF and IFS-ENS throughout the whole forecast range. For stronger precipitation events with more than 5 mm·hr⁻¹ (b), the performance of the three ensemble systems is very similar. However, the Brier score in Figure 11 (b) is generally very small, due to only a very small number of events that exceeded the threshold of 5 mm·hr⁻¹ in the winter period. The forecast quality of precipitation shows a clear diurnal cycle with the lowest predictability in the evening hours. This is a quite surprising result since precipitation in winter is predominantly stratiform and the diurnal distribution of the observed precipitation intensity (Figure 11(c)) is much smoother than in summer (Figure 6(c)). Therefore, the forecast quality should not suffer from a diurnal cycle. The different stability of the boundary layer in the mountainous regions between day and night in the models could be a possible reason for this. However, this needs a further investigation that was beyond the scope of this article.

Figure 12 shows the area under the ROC curve (ROC_auc) for the four different precipitation thresholds of 0.1 mm·hr⁻¹ (a), 0.5 mm·hr⁻¹ (b), 1 mm·hr⁻¹ (c), and 5 mm·hr⁻¹ (d). For the low precipitation thresholds
in panel (a) and (b), the ROC_auc is very similar for all three EPSs, while with increasing precipitation intensity C-LAEF is able to outperform the other two coarser-resolved ensemble systems. Especially for the highest threshold of 5 mm hr\(^{-1}\), the ALADIN-LAEF and IFS-ENS are not able to provide a skillful forecast. The superior performance of C-LAEF for strong precipitation scores is also confirmed by calculating the reliability diagrams (not shown). Some selected scores from the upper-air verification are shown in Figures 13 and 14. Figure 13 shows the CRPS of temperature (a), relative humidity (b), wind speed (c), and geopotential (d) at 500 hPa. C-LAEF shows a smaller CRPS and thus provides a better forecast than ALADIN-LAEF for all considered variables and for all lead times, except for temperature at +30 hr and beyond. Statistically significant differences of the ensembles can be found mainly for relative humidity, wind speed, and geopotential and especially for the higher lead times. The very high CRPS of geopotential in ALADIN-LAEF at initialization time is also present in the winter period (d). The CRPS values at 500 hPa in the winter period are generally a bit higher than in the summer period (Figure 8), which is related to a higher baroclinicity at these latitudes in winter.

The RMSE of the ensemble mean (solid) and ensemble spread (dashed) for the 850 hPa level is shown in Figure 14. The ensemble spread is higher for the C-LAEF system in the case of relative humidity (b) and wind speed (c), while it is quite similar in both ensembles for geopotential (d). RMSE is significantly higher for ALADIN-LAEF in the case of relative humidity (b) and geopotential (d).

# 4 CONCLUSIONS AND OUTLOOK

This study investigated the performance of the recently implemented convection-permitting ensemble system C-LAEF of the Austrian weather service ZAMG. C-LAEF has a horizontal resolution of 2.5 km and runs four times a day on the high-performance computing facilities of the ECMWF. It is driven by the global ensemble system of the ECMWF (IFS-ENS) with a coupling frequency of 3 hr. The forecast range lies between 6 hr (0600 and 1800 UTC run), 48 hr (1200 UTC), and 60 hr (0000 UTC) and is a compromise between the available computer resources and the scientific/operational requirements. The final operational setup of C-LAEF has been defined in close cooperation between model developers and forecasters at ZAMG.

The main outstanding scientific features of C-LAEF are:

1. The ensemble \( J_k \) method, which blends large-scale perturbations from the driving IFS-ENS directly into
the 3D-Var of C-LAEF and thus reduces the perturbation mismatch between the LAM and its global counterpart (Keresturi et al., 2019).

2. ESDA—an ensemble of land surface data assimilation scheme where each member of the ensemble sees different stochastic surface observation perturbations that represent the initial condition surface uncertainties (Bellus et al., 2016).

3. HSPP—a hybrid stochastic physics perturbation scheme that is a combination of a tendency and a
parameter perturbation approach and treats model uncertainties in the physics parametrization schemes of C-LAEF (Wastl et al., 2019b).

All three methodologies have been developed to address the key problems of many limited-area ensemble prediction systems—an insufficient representation of initial condition uncertainties and a lack of ensemble spread during the whole forecast range. Ensemble Jk and ESDA are able to increase the ensemble spread of C-LAEF at initialization time and in the first hours, while HSPP improves the ensemble spread throughout the whole forecast range, as an extensive verification during a test month (July 2016) has shown. Nevertheless, the C-LAEF ensemble is still underdispersive for many parameters, as can be seen in the results presented here. This is a typical characteristic of most limited-area ensemble systems, and therefore the research on improved methodologies for perturbations or other uncertainty representations is a hot topic in the NWP community (Dey et al., 2016; Frogner et al., 2019b; Wang et al., 2020).

A focus of this article lies in an overall assessment of C-LAEF in comparison with the previously operational ensemble system at ZAMG - ALADIN-LAEF (Wang et al., 2011) and the global benchmark in ensemble prediction—the IFS-ENS (ECMWF, 2018). An extensive verification over a summer and a winter period has shown that C-LAEF is able to outperform the other ensemble systems for several upper-air and surface parameters. The difference is mainly obvious in an increased accuracy of the ensemble mean and ensemble probabilities, as well as an increased ensemble reliability in C-LAEF. However, a significant score degradation for C-LAEF could be found for 2 m temperature and humidity during the night in summer. A detailed investigation showed that C-LAEF suffers from a significant warm and dry bias during the night and morning hours at stations in deep Alpine valleys, especially in clear sky conditions. The development of a temperature inversion due to descending of cold air to the valley floor and a resulting decoupling of the lower levels is not well represented in C-LAEF. Literature research has shown that this is not a problem of C-LAEF alone, but that many high-resolution models also suffer from a similar underestimation of the diurnal temperature cycle in narrow valleys (e.g., Pagès and Miró, 2010). Surprisingly, the scores of IFS-ENS and ALADIN-LAEF are much better in such situations. However, it has to be considered that narrow Alpine valleys are not properly resolved in the coarse 11/18 km grids of ALADIN-LAEF and IFS-ENS. A direct comparison is therefore not really fair, even though a simple temperature reduction to the real station height is applied. To reduce this problem in C-LAEF, we have implemented and tested different approaches. The best result could be achieved by applying a Gaussian filter to the subgrid orography of C-LAEF and replacing the digital terrain elevation data SRTM (Jarvis et al., 2008) with GMTED (Danielson and Gesch, 2010). Doing this results in a loss of some subgrid orographic information, but the verification has shown that the temperature/humidity problem could be reduced and that no significant impact on the other scores occurred. The change in the model orography was done before the final operationalization of C-LAEF in November 2019. The verification of the winter period has shown that the problem is significantly reduced. However, especially in the context of higher resolutions in the near future, this is not a satisfying methodology, and therefore further investigations are necessary. One possible solution addresses the surface boundary layer scheme CANOPY (Masson and Seity, 2009), which was initially introduced to the AROME model by Météo France several years ago to improve the 2 m diagnostics. It is a high-vertical-resolution subgrid 1D column model that adds six levels between the lowermost model level and the surface. However, with increasing resolution of AROME and connected steeper slopes, CANOPY has produced some score degradation of 2 m temperature. Therefore, it has been switched off at Météo France when going from 2.5 km to 1.2 km horizontal resolution in their operational AROME model. The effect of switching off CANOPY in C-LAEF is currently under investigation for a longer period. First tests have shown some promising results in the Alps, but also a slight score degradation in the flat areas of the domain.

The biggest improvement of C-LAEF compared with ALADIN-LAEF and IFS-ENS could be found for the surface variables wind and precipitation, which is in accordance with similar studies (Hagelin et al., 2017; Chen et al., 2018; Klasa et al., 2018). In this context, C-LAEF benefits from the higher resolution and the explicit treatment of convection. Especially for convective situations in summer, C-LAEF is able to provide important probabilistic information on the occurrence of small-scale thunderstorms with high precipitation amounts or dangerous wind gusts, which is very helpful in issuing warnings (forecasters’ feedback). Six months after the final operational implementation of C-LAEF, the acceptance and usage of the new ensemble system among the operational forecasters at ZAMG is already very high, as a survey has just shown.

In the near future, with a new HPC at the ECMWF it is planned to extend the forecast range of the intermediate runs and to introduce a 3-hourly assimilation cycle for at least some members. The possibility of independent assimilation cycles within the ensemble members is not new and has already been successfully used, for example, by the COMEPS ensemble of the Danish Meteorological Institute (Frogner et al., 2019a). Furthermore,
an increase of the coupling frequency to 1 hr is envisaged for an update of the C-LAEF suite at the new ECMWF HPC in 2021. Future plans for C-LAEF include also a closer consideration of uncertainties in all the aspects of the model, for example, an improved perturbation scheme for surface parameters to better account for uncertainties in the interaction between surface and atmosphere. Ideas in this context are the perturbation of climatological physiographic fields (e.g., topography, vegetation types, aerosol climatology, etc.) or the perturbation of parameters in the surface scheme SURFEX such as soil temperature, soil moisture, etc. Bouttier et al. (2016) or Frogner et al. (2019a) have shown that this has a positive impact on the scores during the whole forecasting range. The extension of the stochastic physics parameter perturbation scheme to other parametrizations such as shallow convection, microphysics, or radiation to describe errors closer to the source and to increase the physical consistency of the ensemble is also planned. Another interesting point worth considering would be a kind of flow-dependent pattern generator that creates noise only in areas with high uncertainties.

ACKNOWLEDGMENTS

We want to thank all colleagues from ALADIN/LACE/HIRLAM consortia who have contributed to the development of the C-LAEF ensemble system in the past years. We gratefully acknowledge the ECMWF for the provisioning of the computer facilities and the technical help to implement the C-LAEF suite at the ECMWF supercomputer. We are also grateful to the two anonymous reviewers for their valuable comments.

ORCID

Clemens Wastl https://orcid.org/0000-0003-2210-5201

REFERENCES

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M. and Reinhartd, T. (2011) Operational convective-scale numerical weather prediction with the COSMO model: description and sensitivities. Monthly Weather Review, 139, 3887–3905. https://doi.org/10.1175/MWR-D-10-05013.1.

Bellus, M., Wang, Y. and Meier, F. (2016) Perturbing surface initial conditions in a regional ensemble prediction system. Monthly Weather Review, 144, 3377–3390. https://doi.org/10.1175/MWR-D-16-0038.1.

Bénard, P., Vivoda, J., Mašek, J., Smolíková, P., Yessad, K., Smith, C., Brožková, R. and Geleyn, J.F. (2010) Dynamical kernel of the Aladin-NH spectral limited-area model: revised formulation and sensitivity experiments. Quarterly Journal of the Royal Meteorological Society, 136, 155–169. https://doi.org/10.1002/qj.522.

Bengtsson, L., Andrae, U., Aspelien, T., Batrak, Y., Calvo, J., De Rooy, W., Gleeson, E., Hansen-Sass, B., Homleid, M., Hortal, M., Ivarsson, K., Lenderink, G., Niemelä, S., Nielsen, K.P., Onvlee, J., Rontu, L., Samuelsson, P., Muñoz, D.S., Subias, A., Tijm, S., Toll, V., Yang, X. and Kolczew, M.O. (2017) The HARMONIE–AROME model configuration in the ALADIN–HIRLAM NWP system. Monthly Weather Review, 145, 1919–1935. https://doi.org/10.1175/MWR-D-16-0417.1.

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

Bouttier, F., Raynaud, L., Nuissier, O. and Ménétretier, B. (2016) Sensitivity of the AROME ensemble to initial and surface perturbations during HyMeX. Quarterly Journal of the Royal Meteorological Society, 142, 390–403. https://doi.org/10.1002/qj.2622.

Bouttier, F., Vie, B., Nuissier, O. and Raynaud, L. (2012) Impact of stochastic physics in a convection-permitting ensemble. Monthly Weather Review, 140, 3706–3721. https://doi.org/10.1175/MWR-D-12-00031.1.

Bowler, N.E., Arribas, A., Mylete, K.R., Robertson, K.B. and Beare, S.E. (2008) The MOGREPS short-range ensemble prediction system. Quarterly Journal of the Royal Meteorological Society, 134(632), 703–722. https://doi.org/10.1002/qj.234.

Brožková, R., Klaricić, D., Iivatek-Sahdan, S., Geleyn, J.F., Cassé, V., Široka, M., Radnötí, G., Janoušek, M., Stadbacher, K. and Seidl, H. (2001) DF1 blending: an alternative tool for preparation of the initial conditions for LAM. In: Ritchie, H. (Ed.) Research Activities in Atmospheric and Oceanic Modelling, (Vol. 31, pp. 1.7–1.8). Geneva, Switzerland: WMO CAS/JSC WGNE Report.

Bubnova, R., Hello, G., Benard, P. and Geleyn, J.F. (1995) Integration of all the fully elastic equations cast in the hydrostatic pressure terrain-folllowing coordinate in the framework of the ARPEGE/ALADIN NWP system. Monthly Weather Review, 123, 515–535. https://doi.org/10.1175/1520-0493(1995)123<0515:IOTFEE>2.0.CO;2.

Caron, J.F. (2013) Mismatching perturbations at the lateral boundaries in limited area ensemble forecasting: a case study. Monthly Weather Review, 141, 356–374. https://doi.org/10.1175/MWR-D-12-00051.1.

Chen, X., Yuan, H. and Xue, M. (2018) Spatial spread-skill relationship in terms of agreement scales for precipitation forecasts in a convection-allowing ensemble. Quarterly Journal of the Royal Meteorological Society, 144, 85–98. https://doi.org/10.1002/qj.3186.

Cuxart, J., Bouguault, P. and Redelsperger, J.L. (2000) A turbulence scheme allowing for mesoscale and large-eddy simulations. Quarterly Journal of the Royal Meteorological Society, 126, 1–30. https://doi.org/10.1002/qj.49712656202.

Danielson, J. and Gesch, D. B. (2010) Global multi-resolution terrain elevation data 2010 (GDEM2). OFR No. 2011-1073.

Davies, H. (1976) A lateral boundary formulation for large-eddy simulations in limited area models. In: Ritchie, H. (Ed.) Research Activities in Atmospheric and Oceanic Modelling. (Vol. 1, pp. 1.7–1.8). Geneva, Switzerland: WMO CAS/JSC WGNE Report.

Dey, S.R., Plant, R.S., Roberts, N.M. and Migliorini, S. (2016) Assessing spatial precipitation uncertainties in a convective-scale ensemble. Quarterly Journal of the Royal Meteorological Society, 142, 2935–2948. https://doi.org/10.1002/qj.2893.

Duc, L., Saito, K. and Seko, H. (2013) Spatial-temporal fractions verification for high-resolution ensemble forecasts. Tellus A, 65, 1–23. https://doi.org/10.3402/tellusa.v65i0.18171.
ECMWF (2018) IFS Documentation. Available at: https://www.ecmwf.int/en/publications/ifs-documentation [Accessed May 23, 2020].

Fouquart, Y. and Bonnel, B. (1980) Computations of solar heating of the earth's atmosphere: a new parameterization. Beiträge zur Physik der Atmosphäre, 35, 35–62. https://doi.org/10.1029/JD093iD09p11063.

Frogner, I.-L., Andrea, U., Bojarova, J., Callado, A., Escriba, P., Feddersen, H., Hallé, A., Kauhanen, J., Randriamaniamina, R., Singleton, A., Smet, G., Van der Veen, S. and Vignes, O. (2019a) Harmonie—The HARMONIE ensemble prediction system. Weather and Forecasting, 34, 1909–1937. https://doi.org/10.1175/WAF-D-19-0030.1.

Frogner, I.-L., Singleton, A., Koltzow, M. and Andrea, U. (2019b) Convection-permitting ensembles: challenges related to their design and use. Quarterly Journal of the Royal Meteorological Society, 145, 90–106. https://doi.org/10.1002/qj.3525.

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

Guidard, V. and Fischer, C. (2008) Introducing the coupling information in a limited area variational assimilation. Quarterly Journal of the Royal Meteorological Society, 134, 723–736. https://doi.org/10.1002/qj.215.

Gustafsson, N., Janjić, T., Schraff, C., Leuenberger, D., Weissman, M., Reich, H., et al. (2018) Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. Quarterly Journal of the Royal Meteorological Society, 144, 1218–1256. https://doi.org/10.1002/qj.3179.

Hagelin, S., Son, J., Swinbank, R., McCabe, A., Roberts, N. and Tennant, W. (2017) The Met Office convective-scale ensemble, MOGREPS-UK. Quarterly Journal of the Royal Meteorological Society, 143, 2846–2861. https://doi.org/10.1002/qj.3135.

Haiden, T., Kann, A., Wittmann, C., Pistotnik, G., Bica, B. and Gruber, C. (2011) The integrated Nowcasting through comprehensive analysis (INCA) system and its validation over the eastern alpine region. Weather and Forecasting, 26(2), 166–183. https://doi.org/10.1175/2010WAF2222451.1.

Jarvis, A., Reuter, H.I., Nelson, A. and Guevara, E. (2008) Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database. Available at: http://srtm.csi.cgiar.org [Accessed May 23, 2020].

Kann, A., Kršmanc, R., Habrovský, R., Šajn Slak, A., Bujnák, R., Schmid, F., Tarjáni, V., Wang, Y., Wastl, C., Bica, B. and Meirold-Mautner, I. (2014) High-resolution nowcasting and its application in road maintenance: experiences from the INCA system and its validation over the eastern alpine region. Weather and Forecasting, 30, 1077–1089. https://doi.org/10.1175/WAF-D-15-0001.1.

Kann, A., Wittmann, C., Bica, B. and Wastl, C. (2015b) Evaluation of high-resolution precipitation analyses using a dense station network. Hydrology and Earth System Sciences, 19, 1547–1559. https://doi.org/10.5194/hess-19-1547-2015.

Kerestuci, E., Wang, Y., Meier, F., Weidle, F., Wittmann, C. and Atencia, A. (2019) Improving initial condition perturbations in a convection-permitting ensemble prediction system. Quarterly Journal of the Royal Meteorological Society, 145, 993–1012. https://doi.org/10.1002/qj.3473.

Klasa, C., Arpagaus, M., Walser, A. and Wernli, H. (2018) An evaluation of the convection-permitting ensemble COSMO-E for three contrasting precipitation events in Switzerland. Quarterly Journal of the Royal Meteorological Society, 144, 744–764. https://doi.org/10.1002/qj.3245.

Leutbecher, M., Lack, S.J., Ollinaho, P., Simon, T.K., Balsamo, G., Bechtold, P., Bonavita, M., Christensen, H.M., Diamantakis, M., Dutra, E., English, S., Fisher, M., Forbes, R.M., Goddard, J., Haiden, T., Hogan, R.J., Juricke, S., Lawrence, H., MacLeod, D., Magnusson, L., Malardel, S., Massart, S., Sandu, I., Smolkarkiewicz, P.K., Subramanian, A., Vitart, F., Wedi, N. and Weisheimer, A. (2017) Stochastic representations of model uncertainties at ECMWF: state of the art and future vision. Quarterly Journal of the Royal Meteorological Society, 143(707), 2315–2339. https://doi.org/10.1002/qj.3094.

Mascart, P.J. and Bougeault, P. (2011) The Meso-NH atmospheric simulation system: Scientific documentation. Technical Report. Meteo France.

Masson, V. and Seity, Y. (2009) Including atmospheric layers in vegetation and urban offline surface schemes. Journal of Applied Meteorology and Climatology, 48, 1377–1397. https://doi.org/10.1175/2009JAMC1866.1.

Masson, V., Le Moigne, P., Martin, E., Paroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P., Brun, E., Calvet, J.C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essaouini, K., Glibelin, A.L., Giordani, H., Habets, F., Jidane, M., Kerdraon, G., Kourzeneva, E., Lafayse, M., Lafont, S., Lebeaupin Brossier, C., Lemonsu, A., Mahfouf, J.F., Marquin-auld, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F., Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V. and Voldoire, A. (2013) The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes. Geoscientific Model Development, 6, 929–960. https://doi.org/10.5194/gmd-6-929-2013.

Mlawer, E.J., Taubman, S.J., Brown, P.D., Iacono, M.J. and Clough, S.A. (1997) Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. Journal of Geophysical Research: Atmospheres, 102(D14), 663–682. https://doi.org/10.1029/97JD00237.

Molteni, F., Marsigli, C., Montani, A., Nerozzi, F. and Paccagnella, T. (2001) A strategy for high-resolution ensemble prediction. I: definition of representative members and global-model experiments. Quarterly Journal of the Royal Meteorological Society, 127, 2069–2094. https://doi.org/10.1002/qj.49712757612.

Olefs, M., Schöner, W., Suklitsch, M., Wittmann, C., Niedermoser, B., Neururer, A. and Wurzer, A. (2013) SNOWGRID – A New Operational Snow Cover Model in Austria, Proceedings of the International Snow Science Workshop Grenoble Chamonix Mont-Blanc.

Ollinaho, P., Lock, S.J., Leutbecher, M., Bechtold, P., Beljaars, A., Bozza, A., Forbes, R.M., Haiden, T., Hogan, R.J. and Sandu, I. (2017) Towards process-level representation of model uncertainties: stochastically perturbed parametrisations in the ECMWF ensemble. Quarterly Journal of the Royal Meteorological Society, 143, 408–422. https://doi.org/10.1002/qj.2931.
study from the Pyrenees. *Meteorological Applications*, 17, 53–63. https://doi.org/10.1002/met.160.

Palmer, T. N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G. J., Steinheimer, M. and Weisheimer, A. (2009) Stochastic parametrization and model uncertainty. Reading: ECMWF. ECMWF Technical Memorandum, Vol. 598.

Pergaud, J., Masson, V. and Malardel, S. (2009) A parameterization of dry thermals and shallow cumuli for mesoscale numerical weather prediction. *Boundary-Layer Meteorology*, 132, 83–106. https://doi.org/10.1007/s10546-009-9388-0.

Pinty, J.P. and Jabouille, P. (1998) A mixed-phase cloud parameterization for use in mesoscale non-hydrostatic model: simulations of a squall line and of orographic precipitations. *Proc. Conf. of Cloud Phys.*, 1999, 217–220. https://doi.org/10.1256/qj.02.50.

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

Schellander-Gorgas, T., Wang, Y., Meier, F., Weidle, F., Wittmann, C. and Kann, A. (2017) On the forecast skill of a convection-permitting ensemble. *Geoscientific Model Development*, 10, 35–56. https://doi.org/10.5194/gmd-10-35-2017.

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C. and Masson, V. (2011) The AROME-France convective-scale operational model. *Monthly Weather Review*, 139, 976–991. https://doi.org/10.1175/2010MWR3425.1.

Szucs, M. (2016) SPPT in AROME and ALARO. Presentation at HIRLAM WW on EPS and Predictability 2016-2, Helsinki, Finland.

Termonia, P., Fischer, C., Bazile, E., Bouyssel, F., Brožková, R., Bénard, P., Bochenek, B., Degrauwe, D., Derková, M., El Khatib, R., Hamdi, R., Mašek, J., Pottier, P., Pristov, N., Seity, Y., Smolíková, P., Spaniel, O., Tudor, M., Wang, Y., Wittmann, C. and Joly, A. (2018) The ALADIN system and its canonical model configurations AROME CY41T1 and ALARO CY40T1. *Geoscientific Model Development*, 11, 257–281. https://doi.org/10.5194/gmd-11-257-2018.

Teixeira, M.A.C., Kirshbaum, D.J., Olafsson, H., Sheridan, P.F. and Stiperski, I. (2016) *The atmosphere over mountainous regions*. Frontiers in Earth Science. ISBN 9782889450169. Lausanne, Switzerland: Frontiers Media SA, 162. https://doi.org/10.3389/978-2-88945-016-9.

Wang, L., Shen, X., Liu, J. and Wang, B. (2020) Model uncertainty representation for a convection-allowing ensemble prediction system based on CNOP-P. *Advances in Atmospheric Sciences*, 37, 817–831. https://doi.org/10.1007/s00376-020-9262-2.

Wang, Y., Bellus, M., Ehrlich, A., Mile, M., Pristov, N., Smolíková, P., Spaniel, O., Trojakova, A., Brozкова, R., Cedilnik, J., Klarić, D., Kovacic, T., Masek, J., Meier, F., Szintai, B., Tascu, S., Vivoda, J., Wastl, C. and Wittmann, C. (2018) 27 years of regional cooperation for limited area Modelling in Central Europe (RC LACE). *Bulletin of the American Meteorological Society*, 99, 1415–1432. https://doi.org/10.1175/BAMS-D-16-0321.1.

Wang, Y., Belluš, M., Geleyn, J.F., Ma, X., Tian, W. and Weidle, F. (2014) A new method for generating initial condition perturbations in a regional ensemble prediction system: blending. *Monthly Weather Review*, 142, 2043–2059. https://doi.org/10.1175/MWR-D-12-00354.1.

Wang, Y., Bellus, M., Wittmann, C., Steinheimer, M., Weidle, F., Kann, A., Ivatek-Sahdan, S., Tian, W., Ma, X., Tascu, S. and Bazile, E. (2011) The central European limited-area ensemble forecasting system: ALADIN-LAEF. *Quarterly Journal of the Royal Meteorological Society*, 137, 483–502. https://doi.org/10.1002/qj.751.

Wang, Y., Meirold-Mautner, I., Kann, A., Slak, A., Simon, A., Vivoda, J., Bica, B., Böcskör, E., Brezková, L., Dantinger, J., Giszterowicz, M., Heizler, G., Iwanski, R., Jachs, S., Bernard, T., Kršmanc, R., Merše, J., Micheletti, S., Schmid, F. and Vadislavsky, E. (2017) Integrating nowcasting with crisis management and risk prevention in a transnational and interdisciplinary framework. *Meteorologische Zeitschrift*, 26, 459–473. https://doi.org/10.1127/metz/2017/0843.

Wastl, C., Wang, Y., Atencia, A. and Wittmann, C. (2019a) Independent perturbations for physics parameterization tendencies in a convection permitting ensemble. *Geoscientific Model Development*, 12, 261–273. https://doi.org/10.5194/gmd-12-261-2019.

Wastl, C., Wang, Y., Atencia, A. and Wittmann, C. (2019b) A hybrid stochastically perturbed parametrization scheme in a convection permitting ensemble. *Monthly Weather Review*, 147, 2217–2230. https://doi.org/10.1175/MWR-D-18-0415.1.

Weidle, F., Wang, Y., Tian, W. and Wang, T. (2013) Validation of strategies using clustering analysis of ECMWF-EPS for initial perturbations in a limited area model ensemble prediction system. *Atmosphere-Ocean*, 51, 284–295. https://doi.org/10.1080/07055900.2013.802217.

Wilks, D. (2011) *Statistical Methods in the Atmospheric Sciences*, 3rd edition. (Vol. 100, p. 704). Cambridge, MA: Academic Press.

**How to cite this article:** Wastl C, Wang Y, Atencia A, et al. C-LAEF: Convection-permitting Limited-Area Ensemble Forecasting system. *QJR Meteorol Soc*. 2021;147:1431–1451. https://doi.org/10.1002/qj.3986