

IDŐJÁRÁS

*Quarterly Journal of the Hungarian Meteorological Service
Vol. 124, No. 3, July – September, 2020, pp. 401–418*

Forecast skill of regional ensemble system compared to the higher resolution deterministic model

Simona Tascu¹, Mirela Pietrisi^{3,1}, Christoph Wittmann², Florian Weidle²,
and Yong Wang²

¹ *Meteo Romania (National Meteorological Administration)
Bucharest, Romania*

² *Zentralanstalt für Meteorologie und Geodynamik (ZAMG)
Vienna, Austria*

³ *University of Bucharest, Faculty of Physics
P.O.BOX MG 11, Magurele, Bucharest, Romania*

**Corresponding author E-mail: mirela.niculae@meteoromania.ro*

(Manuscript received in final form July 8, 2019)

Abstract—The 11 km regional ensemble ALADIN-LAEF (Aire Limitée Adaption Dynamique Développement InterNational – Limited Area Ensemble Forecasting) is evaluated by comparison with the 5 km deterministic model ALARO (ALADIN and AROME combined model – Application of Research to Operations at Mesoscale), in order to investigate the advantages and disadvantages facing short-range ensemble and high-resolution forecasts. To make rational decisions about the benefits or challenges of both systems, the forecast skill was measured through probabilistic and deterministic approaches over a 2-month period from late spring to summer season of 2013. The verification uses observations from 1219 SYNOP stations and 1 km × 1 km precipitation analysis from INCA (*Integrated Nowcasting through Comprehensive Analysis*) nowcasting system. The evaluation was carried out for three essential meteorological variables: 2 m temperature, 10 m wind speed, and 6-hour cumulated precipitation. From the probabilistic point of view, the results show that ALADIN-LAEF outperforms ALARO-LAGGED ensemble system, being statistically more reliable. From the deterministic point of view, the high-resolution deterministic system simulates better the precipitation forecast structure with respect to the observations. Compared to the ensemble system, the deterministic system cannot provide guidance concerning the forecast uncertainties or probabilities, making the ensemble products a powerful tool for risk assessment and decision making.

Key-words: ensemble forecast, numerical models, time-lagged, ALADIN-LAEF, ALARO

1. Introduction

An accurate weather forecast is crucial in severe weather conditions that could lead to a large variety of damages including loss of lives. Furthermore, increasingly skilful numerical weather forecasts convey essential information for several domains of public interest, like health, agriculture, energy, tourism.

Despite the significant increase in forecast skill, uncertainties in the initial conditions and the model formulation itself affect the accuracy of the forecast. The inherent errors in the initial conditions could amplify in time, due to the complexity and chaotic character of the atmospheric system, leading sometimes to completely different solutions (*Lorenz, 1963*). These errors are due to the irregular distribution (relative sparseness) of the observation network, the measurement instruments errors, as well as to data assimilation. The benefit of using data assimilation algorithms is unquestionable in the numerical weather prediction field (*Law and Stuart, 2012*). Nonetheless, small uncertainties are coming from data assimilation algorithms due to mathematical assumptions (*Du, 2007*). The forecast uncertainty arisen from the imperfection of the model itself was quantified for the first time through stochastic perturbations of the physical tendencies (SPPT scheme) in the ECMWF model (*Buizza et al., 1999; Palmer, 2001*). Other approaches to simulate the model uncertainties are based on multi-physics schemes (*Murphy et al., 2004*) or on the use of Poor Man's Ensemble (*Ebert, 2001; Corazza et al., 2018*).

To quantify these uncertainties, a widely used technique is the ensemble prediction system (*Palmer, 2017*): a single deterministic forecast provides one single scenario and a simple decision strategy, while an ensemble system is able to provide probabilistic information and multiple scenarios. Several studies have been done on the comparison of global ensemble and deterministic systems (*Buizza, 2008; Palmer et al., 2005/06; Rodwell, 2005/06*). The results show that a coarser resolution ensemble with more members can outperform a higher resolution ensemble with less members.

Here we use the regional ensemble system ALADIN-LAEF (*Wang et al., 2011*) and the ALARO deterministic model (*Termonia et al., 2017*). The operational ALADIN-LAEF system has been developed at ZAMG (ZentralAnstalt für Meteorologie und Geodynamik, Vienna, Austria) within the framework of Regional Cooperation for Limited-Area modelling in Central Europe (RC-LACE, *Wang et al., 2018*). ALADIN-LAEF runs in operational mode on European Centre for Medium-Range Weather Forecasts's (ECMWF) supercomputer (*Wang et al., 2011*). The system is used in different downstream applications such as hydrology, transportation, energy, agriculture, and civil warnings (*Vokoun and Hanel, 2018; Wastl et al., 2018*). The limited area ALARO model development started in 2003 as a version of ALADIN system. It is able to cover a large variety of horizontal resolution ranges, up to the so-called convection permitting scales (*Termonia et al., 2017*).

This paper presents the benefit of using regional ensemble over deterministic systems – even if both the horizontal and vertical resolutions of the deterministic model are much higher. This issue is addressed by a verification study, from the probabilistic and deterministic points of view. The paper is structured as follows: Section 2 introduces the ALADIN-LAEF and ALARO models setup and the time-lagged ensemble; Section 3 describes the dataset and verification methodologies; Section 4 encapsulates the results from a 2-month period verification for surface parameters, and a case study is performed; the conclusions are drawn in Section 5.

2. Models setup and time-lagged ensemble method

2.1. ALADIN-LAEF – Limited Area Ensemble Forecasting

ALADIN-LAEF is a single model limited area ensemble system based on the ALARO model (*Wang et al.*, 2011, 2018) and runs operationally two times per day at 00 and 12 UTC, on a horizontal resolution at approximately 11 km, with 45 levels in the vertical and a forecast range up to 72-hour. The integration domain covers Europe and large parts of the North Atlantic, the Mediterranean Sea, and the Black Sea. ALADIN-LAEF consists of 16 perturbed members, using the first 16 members of 50 from ECMWF’s ensemble prediction system ENS (*Buizza et al.*, 2000) as coupling model (*Weidle et al.*, 2013).

To provide meaningful initial perturbations for the ensemble members, a breeding-blending cycle is applied for atmospheric fields (*Wang et al.*, 2011). This method combines large-scale perturbations from the driving ECMWF ENS members with ALADIN-LAEF breeding vectors that contain perturbations on scales that can be resolved by ALADIN-LAEF (*Toth and Kalnay*, 1993). To assure a smooth transition between large-scale and small-scale perturbations, a digital filter initialization (DFI) is used (*Wang et al.*, 2014). The DFI is applied on low truncations of both the ALADIN-breeding vectors and the fields from the driving model. The filtered breeding vectors are subtracted on the full resolution from the unfiltered ones, and this difference is then added to the filtered fields from ECMWF ENS. This method assures that the initial perturbations are consistent with both the driving ECMWF ENS member and ALADIN-LAEF itself. Surface perturbations are generated by running a surface assimilation scheme with randomly perturbed observations. An optimal interpolation (OI) assimilation of 2 m temperature and relative humidity observations is applied to perturb the uppermost surface fields in ALADIN-LAEF.

A multi-physics approach is implemented in ALADIN-LAEF to account for model uncertainties. Different model configurations with various parameterizations and tuning settings are applied for each perturbed ensemble member. Different settings are used for shallow and deep convection,

microphysics, radiation, and turbulence parameterizations. In addition, different approaches and tunings for screen-level diagnostics and gust diagnostics are applied in the multi-physics package of ALADIN-LAEF.

2.2. ALARO deterministic model

The ALARO model can be described as the further development of the spectral limited area model ALADIN, in order to be able to run the model using grid spacing around 5 km or below. These scales pose a particular challenge to model developers as convection is neither fully resolved nor can be sufficiently described with a classical parameterized convection scheme. Therefore, the main differences between ALARO and ALADIN concern the deep convection scheme in the physics package. For ALARO, a prognostic convection scheme called Modular Multiscale Microphysics and Transport scheme (3MT) was developed by *Gerard and Geleyn (2005)*, *Gerard (2007)*, *Gerard et al. (2009)*. Further, a prognostic microphysics scheme is included together with a statistical scheme describing the sedimentation of precipitating hydrometeors (*Geleyn et al., 2008*). Turbulence is parameterized using a so-called pseudo prognostic TKE (turbulent kinetic energy) scheme, details can be found in *Váňa et al. (2008)*. Surface processes are described using a two-layer version of the ISBA scheme (Interactions between Soil, Biosphere, and Atmosphere, *Noilhan and Planton, 1989*). At ZAMG, the ALARO model is used in operational mode since March 2011. It is run up to 72-hour lead time, four times per day, using approximately 5 km grid spacing and 60 levels in the vertical. While the three-dimensional (3D) initial state for the atmosphere and the lateral boundary conditions are provided by the high resolution analysis and forecast of the Integrated Forecasting System (IFS HRES) from ECMWF, the surface is initialized using an optimal interpolation method.

2.3. Time-lagged ensemble using ALARO deterministic model

Since ALARO deterministic model is integrated in operational mode at ZAMG, four times per day up to a lead time of 72-hour, it is possible to generate a time-lagged ensemble (ALARO-LAGGED) with no additional computational costs. The construction of this ensemble is based on a combination of several deterministic integrations (*Hoffmann and Kalnay, 1983*). In our case, the time-lagged ensemble covers a forecast range of 48-hour and consists of forecasts from consecutive ALARO model runs, covering the same target time period. For a given day, the ensemble with initial time at 12 UTC contains 5 members, as shown in *Table 1*.

Table 1. ALARO-AUSTRIA time-lagged ensemble

12 UTC(one day before):													
member 5 -	00	06	12	18	24	30	36	42	48	54	60	66	72
18 UTC(one day before):													
member 4 -		00	06	12	18	24	30	36	42	48	54	60	66
00 UTC(current day):													
member 3 -			00	06	12	18	24	30	36	42	48	54	60
06 UTC(current day):													
member 2 -				00	06	12	18	24	30	36	42	48	54
12 UTC(current day):													
member 1 -					00	06	12	18	24	30	36	42	48

3. Dataset and verification methodologies

3.1. Model dataset

The numerical forecast generated by the deterministic and ensemble systems is provided on regular latitude - longitude grids, having the resolution of $0.1^\circ \times 0.14^\circ$ for ALADIN-LAEF, while for ALARO the denser grid of $0.04^\circ \times 0.06^\circ$ is kept. The high-resolution grid for ALARO is used to preserve the mesoscale features of the forecasts. The verification domain covers most of Europe, a region between 2.55 to 31.8° E and 38.6 to 54.95° N (not shown).

3.2. Observation dataset

SYNOP data from 1219 stations are used for verification. Model data were interpolated to the observation sites through a bilinear interpolation.

INCA is an analysis and a nowcasting system (Haiden *et al.*, 2009) developed at ZAMG. The system is run on a 3D grid with a horizontal resolution of 1 km. Depending on the parameter, the analysis and nowcasting are performed on a two-dimensional (2D) grid (e.g., precipitation) or a 3D grid (e.g., temperature, wind). For the 3D grid, a vertical resolution around 100–200 m is used. The system is fed by surface station observations, remote sensing data (radars and satellites), numerical weather prediction model data, and high-resolution topographic data.

3.3. Verification methodologies

An important question arises from whether the high-resolution deterministic ALARO model could compete with the lower resolution ALADIN-LAEF ensemble. This question can be answered by assessing the two systems based on deterministic and probabilistic approaches.

Both systems were evaluated by applying the traditional verification scores (like BIAS and RMSE), as well as the following ones:

- spread - skill relationship, based on the scatter diagram of ensemble spread and root mean square error of the ensemble mean;
- continuous ranked probability skill score (CRPSS), a skill score which is based on CRPS value (*Hersbach, 2000*) and verifies the model performance related to a reference;
- percentage of outliers, a measure of reliability quantifying the relative number of observations which lie outside the predicted density function for a given parameter.

The pointwise verification is not suitable for the precipitation field, because these scores do not fully account for the unique characteristics due to its discontinuity in space and time (*Casati et al., 2004*).

The traditional metrics offer little diagnostics about the types of errors (displacement, location, and intensity), therefore an advanced verification technique is applied. One spatial verification approach used in this study is the object-based method SAL (structure-amplitude-location). This method is a spatial three-component feature-based quality measure which quantifies the precipitation forecast performance according to three aspects: structure of the precipitation field (S), amplitude (A), and location (L) of the predicted mass center of the precipitation field (*Wernli et al., 2008*).

4. Results

To evaluate the performance of the described ensemble systems, a 2-month verification was performed for the period ranging from April 23 to June 23, 2013. The forecast output frequency is 6-hour and the verification length is up to 48-hour lead time, which is the maximum forecast range of ALARO-LAGGED. Even though both runs (00 and 12 UTC) of ALADIN-LAEF are available, only the 12 UTC runs are considered for this study.

First, the probabilistic approach is applied to evaluate the surface parameters: 2 m temperature (T2m), 10 m wind speed (W10m) and 6-hour cumulated precipitation (PREC) against SYNOP data. The computed scores are CRPSS, percentage of outliers and spread - skill relationship, as well as BIAS and RMSE of ensemble mean and ensemble spread. To reduce the forecast bias

at each SYNOP station, the 2 m temperature forecasts are height corrected using a standard temperature gradient.

Second, the deterministic approach is applied to verify the 6-hour cumulated precipitation against INCA analysis field (which represent a combination of radar and rain gauge data). The gridded analysis data are used for the computation of SAL score. The comparison of ALADIN-LAEF's ensemble mean and median and ALARO's most recent run was carried out.

4.1. Comparison of ALADIN-LAEF with ALARO-LAGGED

This section compares, from the probabilistic point of view, the skill of ALADIN-LAEF and ALARO-LAGGED against SYNOP data. *Fig. 1* shows the metrics (BIAS and RMSE of the ensemble mean and spread) of both systems for the surface parameters.

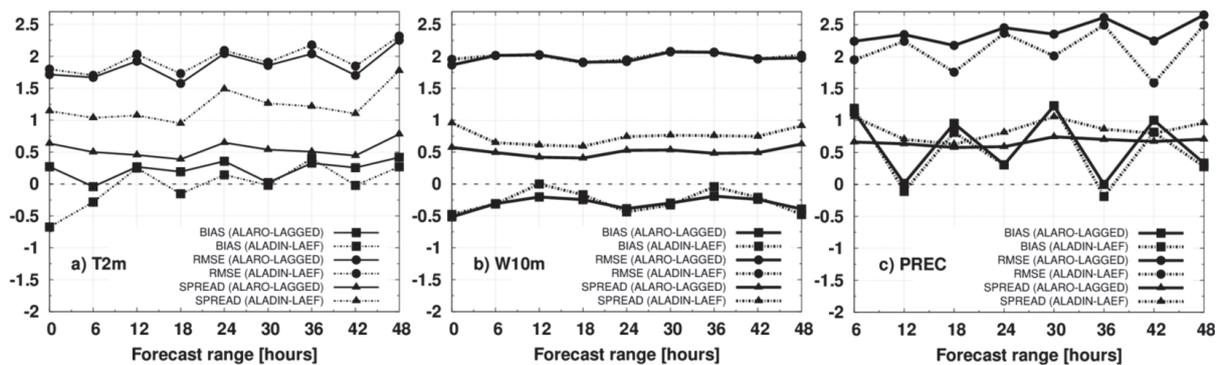


Fig. 1. BIAS and RMSE of ensemble mean and ensemble spread for ALARO-LAGGED (solid) and ALADIN-LAEF (dashed): a) T2m (°C); b) W10m (m/s); and c) PREC (mm).

For T2m, ALARO-LAGGED performs better than the limited area ensemble system in the first lead times (00 and 06 forecast ranges), for which ALADIN-LAEF has a strong negative BIAS (*Fig. 1a*). After 12-hour lead time, ALADIN-LAEF shows slightly less or equal BIAS than ALARO-LAGGED, except for 18 and 42 forecast ranges (morning hours) when the BIAS is considerably smaller. *Fig. 1b* shows that the W10m BIAS of both systems have comparable negative values for all lead times. The exception is made for 12- and 36-hour lead time when ALADIN-LAEF has almost no BIAS. For PREC, *Fig. 1c* shows a similar BIAS for both systems. ALARO-LAGGED system presents only positive values, better values for 12 and 36h (BIAS close to 0), while ALADIN-LAEF presents small negative values for these forecast ranges.

Likewise, *Fig. 1* presents RMSE and spread as a function of lead time for surface parameters. In terms of RMSE, the forecast errors are similar for both

systems. The ALADIN-LAEF system has a larger spread, which means that it is more reliable than ALARO-LAGGED, but the discrepancy between the ensemble spread and RMSE leads to the fact, that both systems are not enough statistically reliable (Buizza *et al.*, 2005). It is expected that ALARO-LAGGED will have a small spread, considering the lagged ensemble members are partially correlated. They are obtained using the same model, high-resolution forecasts with different ages.

The spread - skill relationship is presented in *Fig. 2*. The scatter diagram shows that ALARO-LAGGED simulates too little uncertainty, meaning the spread is underestimated, having only points which are not uniformly distributed. *Fig. 2c* (PREC) indicates a good correlation between spread and skill, especially for ALADIN-LAEF the relation is slightly better. Thereby, the ALADIN-LAEF system is more reliable than ALARO-LAGGED. Statistical reliability can be underlined by the usage of percentage of outliers, quantifying the number of cases where the analysis is outside of the predicted density function. As it can be seen from *Fig. 3*, ALARO-LAGGED has more outliers than ALADIN-LAEF.

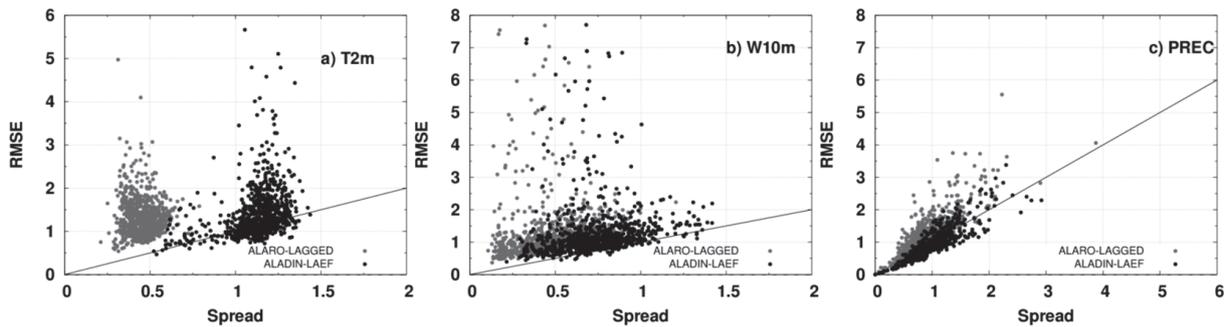


Fig. 2. RMSE-Spread relation for ALARO-LAGGED (grey) and ALADIN-LAEF (black): a) T2m (°C); b) W10m (m/s); and c) PREC (mm).

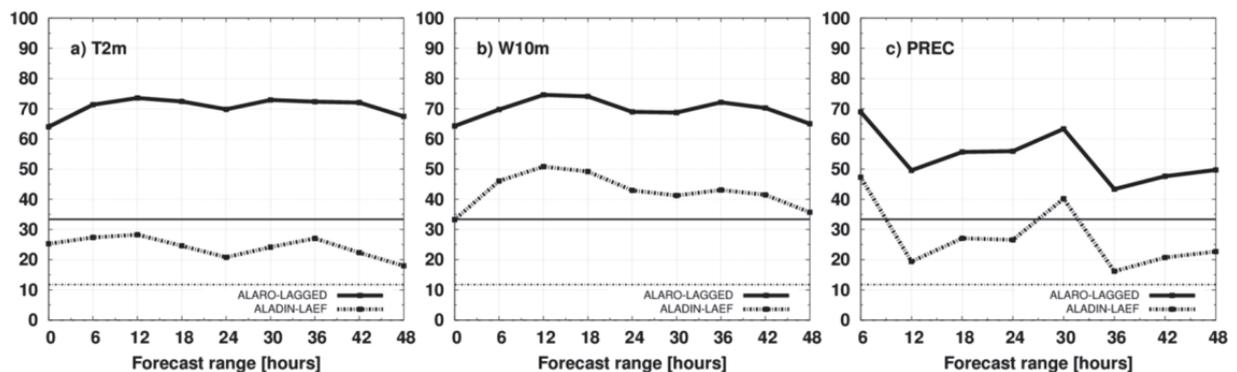


Fig. 3. Percentage of outliers (%) for ALARO-LAGGED (solid) and ALADIN-LAEF (dashed): a) T2m; b) W10m; and c) PREC.

Results revealed that ALARO-LAGGED has a similar RMSE (forecast accuracy), a smaller spread, and more outliers when compared to the ALADIN-LAEF. Therefore, having less dispersion, it is not able to cover many possible atmospheric situations. Considering all these findings, it can be concluded that ALADIN-LAEF is statistically more reliable.

4.2. Skill score of ALADIN-LAEF and ALARO-LAGGED (reference ALARO)

The computation of the skill score CRPSS uses as reference system the ALARO deterministic model. Fig. 4 shows the quantitative skill of both ensembles. For all verified parameters, CRPSS has positive values throughout the forecast ranges. Even though both systems present positive values of the score, for ALADIN-LAEF they are significantly higher than ALARO-LAGGED. Based on the CRPSS results, ALADIN-LAEF ensemble is more skilful than ALARO-LAGGED. It is worth to underline that ALARO-LAGGED has an advantage since the most recent run from ALARO-LAGGED is used as the reference deterministic model.

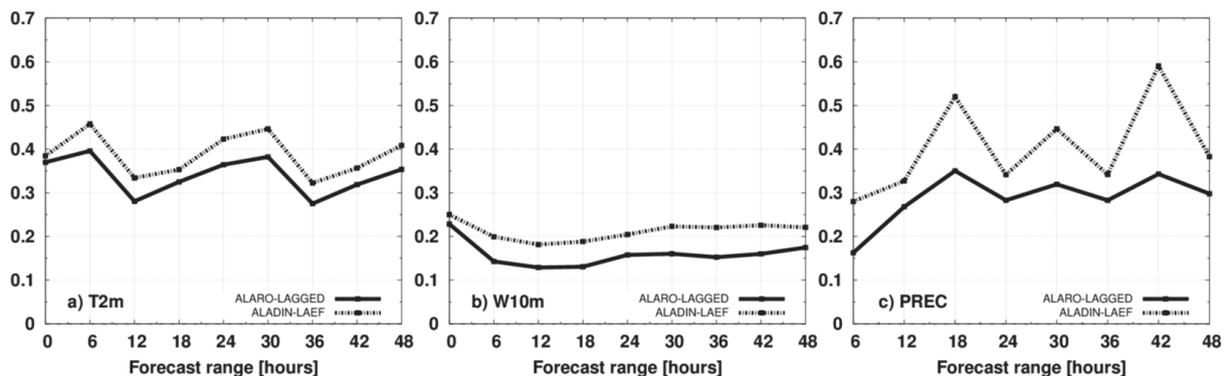


Fig. 4. Comparison of CRPSS for ALARO-LAGGED (solid) and ALADIN-LAEF (dashed): a) T2m; b) W10m; and c) PREC.

4.3. Deterministic comparison of ALADIN-LAEF and ALARO systems

To perform the comparison of both systems in a deterministic sense, a single forecast from each system is needed. The solutions provided by ALARO deterministic model and ALADIN-LAEF (mean and median of the ensemble) are evaluated in this section. The SAL component scores have been computed to evaluate the precipitation forecasts of ALADIN-LAEF (median and mean) and ALARO. Verification was performed using several domains in Austria that are characterized by different topographic conditions (flatlands, alpine region, and hilly region).

In order to get more details, the evaluation was done for different observed precipitation thresholds, i.e., the SAL components are calculated separately for events with observed areal precipitation means greater than 0.1 mm, 1 mm, 2 mm, 3 mm, 5 mm, and 10 mm. Taking into account the chosen size of the domains, area mean values greater than 5 mm or even 10 mm represent rare extreme cases and are therefore not taken into account for computing mean S, A, or L values over the given verification period. *Table 2* shows the mean values for S, A, and L over all verification domains for different lead times.

Table 2. The S, A, L values: mean over all domains for ALARO (ALARO), ALADIN-LAEF's ensemble mean (LMEAN), and ALADIN-LAEF's ensemble median (LMEDI)

	06	12	18	24	30	36	42	48
S-ALARO	0.04	-0.22	-0.29	0.23	0.04	-0.23	-0.27	0.31
S-LMEAN	1.10	0.72	0.54	1.24	1.15	0.76	0.74	1.28
S-LMEDI	0.91	0.55	0.29	1.07	0.98	0.42	0.26	1.05
A-ALARO	-0.53	-0.49	-0.30	-0.03	-0.25	-0.55	-0.52	0.05
A-LMEAN	0.17	-0.33	-0.31	0.45	0.20	-0.36	-0.36	0.54
A-LMEDI	-0.47	-0.59	-0.54	0.17	-0.24	-0.49	-0.47	0.31
L-ALARO	0.22	0.21	0.21	0.20	0.23	0.23	0.24	0.24
L-LMEAN	0.22	0.20	0.20	0.27	0.23	0.23	0.23	0.29
L-LMEDI	0.20	0.19	0.22	0.26	0.21	0.24	0.25	0.29

Fig. 5 shows the mean structure components for the 2-month period as a function of lead time, for a representative region in the southern part of Austria. The number of events used to build the mean values is indicated by the thin black bars in the same figure. It can be seen that:

- ALARO yields the best precipitation forecasts in terms of structure compared to the ensemble mean and median of ALADIN-LAEF, i.e., the size and shape of the precipitation simulated by ALARO corresponds better to the observed object characteristics (S values closer to 0). This result can be expected, as ALARO is run on a significantly higher horizontal resolution than ALADIN-LAEF.
- Comparing the ensemble mean and median of ALADIN-LAEF, the median yields slightly more structured forecasts, i.e., smaller values for S than the mean. This fact also meets expectations as the mean usually yields a smoother field.
- A diurnal cycle is visible for all systems, showing that the difference in terms of the observed objects gets usually bigger during night, leading to even negative S values. This diurnal behavior could be explained on the assumption that the transition from primary convection to more organized

one is not sufficiently simulated. This is mainly caused by a wrong or too early timing of triggering in the models, conducting to an early decrease of convection objects, leading to smaller objects compared to observations during the evening and early night. In reality, convection usually starts later during the day and lasts longer, as it is often more organized than in the models (*Wittmann et al., 2010*). This behavior is also visible in the A component as described later.

- *Table 2* shows the S values averaged over all regions. The results differ in terms of absolute values of the S component for each domain (not shown), although the differences among the systems (ALARO vs. ALADIN-LAEF) and the characteristic of the diurnal cycle are consistent over all domains.
- A comparison of the results for different area mean thresholds seems to reveal that structure scores tend to get better, i.e., closer to 0, when concentrating on cases with higher precipitation rates (not shown). Also, the diurnal cycle is less pronounced. This could be correlated with the fact that strong precipitation events are often related to some large scale forcing which is less dependent on the general diurnal cycle of convection.

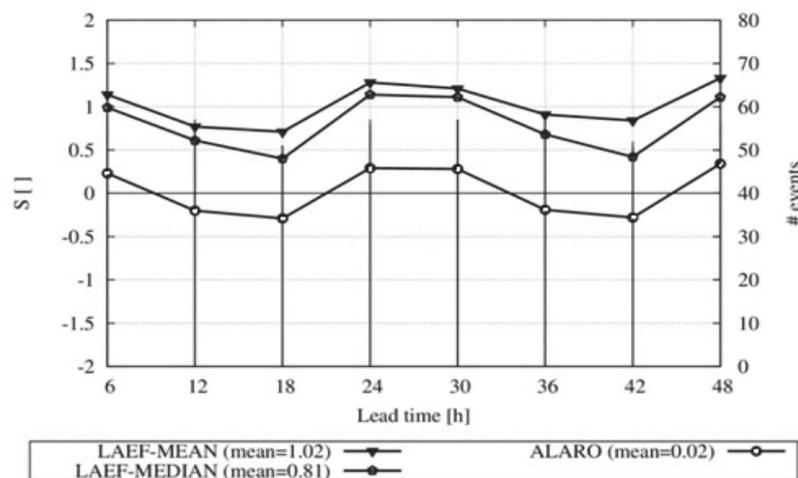


Fig. 5. S component of: ALARO (ALARO), ALADIN-LAEF's ensemble mean (LAEF-MEAN) and ALADIN-LAEF's ensemble median (LAEF-MEDIAN) for the southern part of Austria, for area mean precipitation greater than 0 mm.

The interpretation of the A component (evaluating the area mean precipitation for a given domain) is more difficult, as the results are less clear than for the structure component. *Fig. 6* and *7* show the A component for a domain in the central part of Austria.

- The diurnal cycle already mentioned above is also visible in the amplitude component: during (early) day, precipitation activity is overestimated, while there is an underestimation during evening and night. Again, this can possibly be explained by a general shift of the (convective) precipitation diurnal cycle, with the convection starting too early in the model (*Fig. 6*).
- For higher precipitation thresholds, underestimation gets larger (especially for ALADIN-LAEF), and there is a decrease in overestimation, therefore the diurnal cycle is less pronounced (*Fig. 7*).

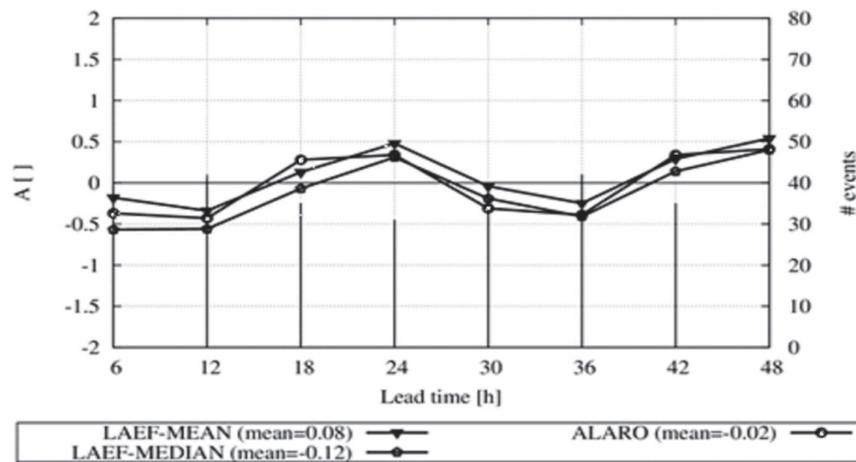


Fig. 6. A component of ALARO (ALARO), ALADIN-LAEF's ensemble mean (LAEF-MEAN), and ALADIN-LAEF's ensemble median (LAEF-MEDIAN) for the central part of Austria, for area mean precipitation greater than 0.1 mm.

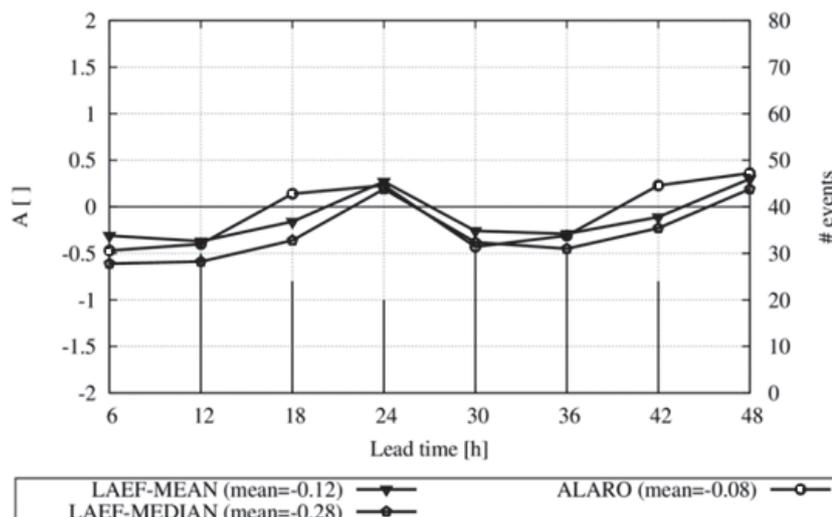


Fig. 7. A component of: ALARO (ALARO), ALADIN-LAEF's ensemble mean (LAEF-MEAN) and ALADIN-LAEF's ensemble median (LAEF-MEDIAN) for the central part of Austria, for area mean precipitation greater than 0.5 mm.

The evaluation of the mean L component scores does not reveal a clear view, as the results do not differ significantly among the model systems. As it was already stated in other studies (*Wittmann et al.*, 2010), the L score seems to be more suitable to be used in case studies. Mean values over an extended verification period do not reveal a clear picture.

4.4. Case study

In addition to the statistical scores computed for the 2-month period, a comprehensive case study is carried out for a major flood event that took place in Central Europe from May 31 to June 3, 2013, affecting the Danube, Elbe, and Moldova river catchments.

The large amount of precipitation was linked to the subsequent passage of three cyclones formed over Southeastern Europe, traveling around the quasi-stationary upper-level low. A detailed description of the synoptic situation is given by *Grams et al.* (2014). The effect of this precipitation event on the runoff and level of various rivers in Central Europe was huge, since the soil was already saturated or close to saturation before the start of the event. From climatological point of view, May 2013 was one of the three wettest months of May in the last 150 years in the Upper Danube basin in Austria (*Blöschl et al.*, 2013).

To evaluate the performance of the systems described in the previous sections, the investigation of this case will focus on Austria and the surrounding regions, using several rectangular domains. The domain “RR Zentrum” (11.00 to 14.00 E longitude, 47.10 to 48.40 N latitude) is the most affected one, with peak precipitation values up to 300 mm in 72 hours (*Fig. 8*). The 72-hour cumulated precipitation map (May 31, 2013, 00 UTC – June 3, 2013, 00 UTC) is provided by INCA analysis. The highest precipitation rates are identified in the interval from June 1, 2013, 18 UTC to June 2, 2013, 06 UTC. Thereby, the analysed interval is divided in two periods of 6-hour forecast range, as shown in *Fig. 9* (a: June 1, 2013, 18 UTC – June 2, 2013, 00 UTC and b: June 2, 2013, 00 UTC – June 2, 2013, 06 UTC).

Focus will be on the capabilities of the ALADIN-LAEF and ALARO-LAGGED ensembles to simulate the quantitative precipitation forecasts on the short time range. To assess the uncertainty of the analyzed ensembles, a so-called “spaghetti” plot is represented in *Fig. 10*. It shows the area mean 6-hour cumulated precipitation for forecast ranges up to 48-hour, for both ensembles and the INCA analysis, covering the “RR-Zentrum” domain. The forecasts are initialized at different times: May 31, 2013, 12 UTC (*Fig. 10a*) and June 1, 2013, 12 UTC (*Fig. 10b*).

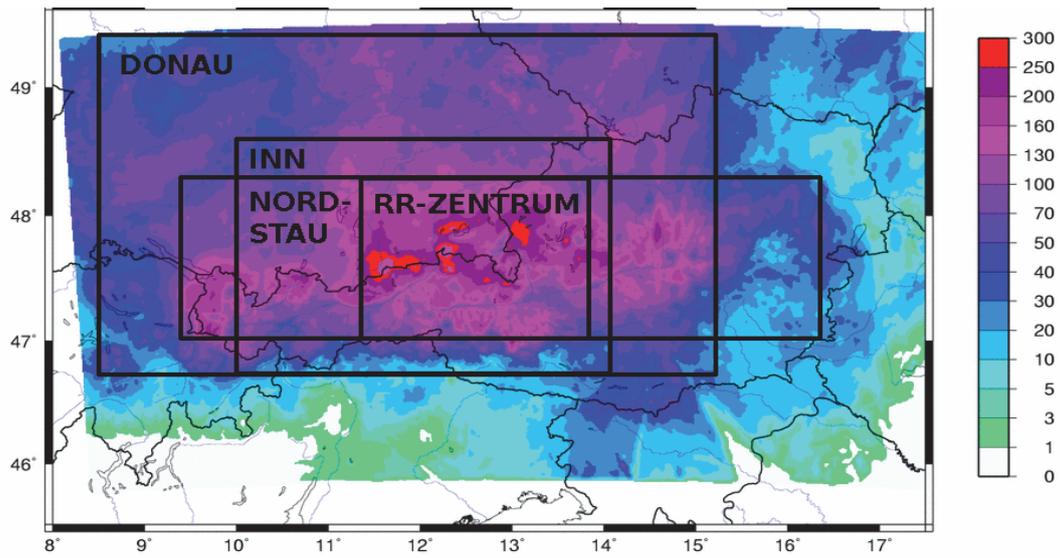


Fig. 8. INCA 72-hour cumulated precipitation [mm] in the period of May 31, 2013, 00 UTC – June 3, 2013, 00 UTC.

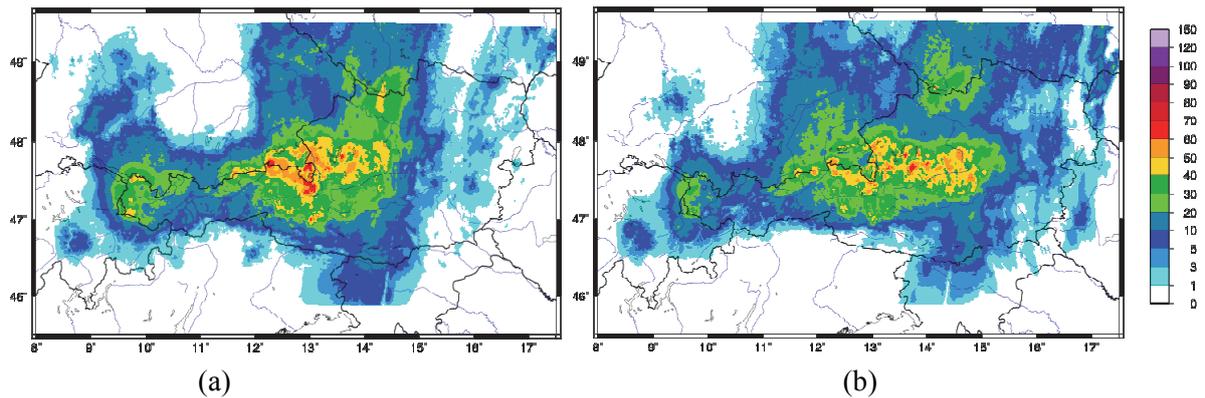


Fig. 9. INCA 6-hour cumulated precipitation [mm] in the periods of June 1, 2013, 18 UTC – 02.06.2013, 00 UTC (a) and 02.06.2013, 00 UTC – 02.06.2013, 06 UTC (b).

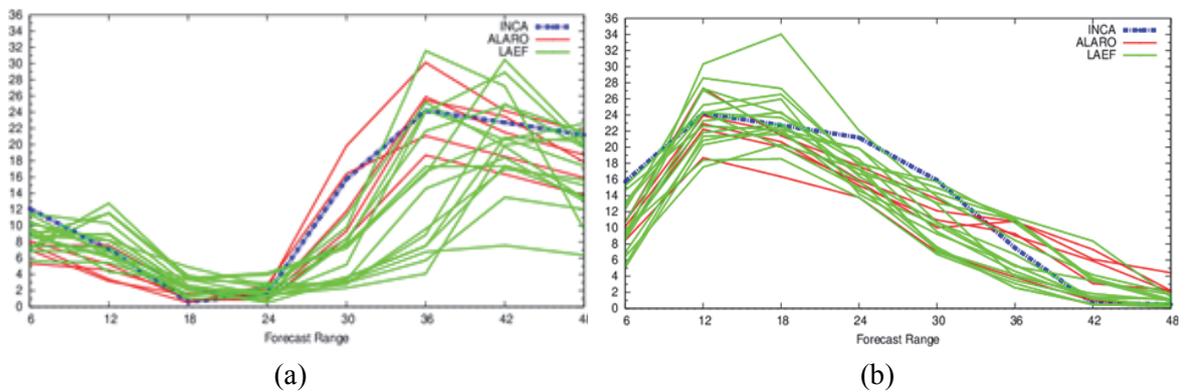


Fig. 10. 6-hour cumulated precipitation [mm] for RR-Zentrum starting from May 31, 2013, 12 UTC (a) and June 1, 2013, 12 UTC (b): INCA (blue), ALADIN-LAEF members (LAEF, green), and ALARO-LAGGED members (ALARO, red).

The members of ALADIN-LAEF and ALARO-LAGGED are rather close to INCA in the first 24 hours, showing only small spread (*Fig. 10a*). After 24-hour lead time, when the precipitation rates start to increase (up to 24 mm in 6-hour at lead time 36-hour), the ALADIN-LAEF shows larger spread than ALARO-LAGGED. Also, the mean precipitation rates are underestimated by many ALADIN-LAEF members with respect to the INCA analysis. These results are underlined by the verification scores obtained for the 2-month verification period, shown in subsection 4.1.

In *Fig. 10b*, it can be noticed that both systems have a smaller spread at the beginning of the interval, the maximum precipitation amount (at 12-hour lead time) is now better captured by more members of ALADIN-LAEF. Yet, ALARO-LAGGED underestimates the precipitation rate occurring from 12 to 30-hour lead time and decreases the precipitation too early. For higher lead times (36–48-hour), ALARO-LAGGED slightly overestimates, while ALADIN-LAEF stays closer to the observed values.

5. Conclusions

The forecast skill of the 11 km regional ensemble ALADIN-LAEF, compared to the 5 km deterministic ALARO model, has been investigated, in order to understand the advantages and shortcomings of both systems. ALADIN-LAEF's members are generated by using the first 16 members (from 50) of ECMWF ENS, with different initial conditions perturbations. To quantify the uncertainties in the forecasting system, the breeding-blending and multi-physics approaches are applied. The time-lagged ensemble comprises 5 successive forecasts of ALARO deterministic model from different runs. The forecast quality of the systems was assessed through probabilistic and deterministic measures over a 2-month period, at the beginning and during the convective season of 2013, for the main surface parameters (2 m temperature, 10 m wind speed, and 6-hour cumulated precipitation). A special focus on mesoscale convective systems that generated heavy precipitation over Central Europe was also evaluated.

The main conclusions of this study are:

- (1) In general, the probabilistic verification reveals that ALADIN-LAEF, compared to ALARO-LAGGED, shows more spread, less outliers, and it is more skilful in terms of CRPSS. Hence, the low-resolution ensemble system is able to cover many possible scenarios through the spread, and it is statistically more reliable.
- (2) Regarding the deterministic verification (conducted for 6-hour cumulated precipitation forecasts), in terms of SAL, the results reveal the advantages of the high-resolution deterministic model ALARO compared to the lower-

resolution regional ensemble ALADIN-LAEF. The ALARO precipitation forecast structure is better simulated compared to the observed objects. The structure and amplitude components show similar behavior of the diurnal cycle for both systems. The premature trigger of the convection masks sometimes the difficulty to simulate the diurnal cycle, mainly because the convective precipitation starts too early in the model.

- (3) The high-resolution deterministic model is able to better reproduce the mesoscale convective systems and to solve the complexity of topography. However, the deterministic system is not able to provide information about forecast uncertainties or probabilities.
- (4) The regional ensemble system provides advantages compared to the deterministic one. Thus, the ensemble products represent a powerful tool in risk assessment and decision making.

Acknowledgments: The results are obtained in the frame of the RC-LACE project. We would like to thank our colleagues from the INCA department for providing the observation datasets.

References

- Blöschl, G., Nester, T., Komma, J., Parajka J., and Perdigao, R., 2013: The June 2013 flood in the Upper Danube Basin, and comparisons with the 2002, 1954 and 1899 floods. HESS 17, 5197–5212. <https://doi.org/10.5194/hess-17-5197-2013>*
- Buizza, R., Miller, M., and Palmer, T.N., 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system, Quart. J. Roy. Meteorol. Soc. 125, 2887–2908. <https://doi.org/10.1002/qj.49712556006>*
- Buizza, R., Barkmeijer, J., Palmer T.N., and Richardson, D.S., 2000: Current status and future developments of the ECMWF Ensemble Prediction System. Meteorol. Appl. 7, 163–175. <https://doi.org/10.1017/S1350482700001456>*
- Buizza, R., Houtekamer, P.L., Toth, Z., Pellerin, G., Wei, M., and Zhu, Y., 2005: A comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems, Mon. Weather Rev. 133, 1076–1097. <https://doi.org/10.1175/MWR2905.1>*
- Buizza, R., 2008: Comparison of a 51-Member Low-Resolution (TL399L62) Ensemble with a 6-Member High-Resolution (TL799L91) Lagged-Forecast Ensemble. Mon. Weather Rev. 136, 3343–3362. <https://doi.org/10.1175/2008MWR2430.1>*
- Casati, B., Ross, G. and Stephenson, D. B., 2004: A new intensity-scale approach for the verification of spatial precipitation forecasts. Met. Apps. 11, 141–154. <https://doi.org/10.1017/S1350482704001239>*
- Corazza, M., Sacchetti, D., Antonelli, M., and Drofa, O., 2018: The ARPAL operational high resolution Poor Man’s Ensemble, description and validation. Atmos. Res. 203, 1–15. <https://doi.org/10.1016/j.atmosres.2017.11.031>*
- Du, J., 2007: Uncertainty and ensemble forecast. Sci. Technol. Inf. Lecture Ser. Available online at <https://www.nws.noaa.gov/ost/climate/STIP/uncertainty.htm>*
- Ebert, E., 2001: Ability of a Poor Man's Ensemble to Predict the Probability and Distribution of Precipitation. Mon. Weather Rev. 129, 2461–2480. [https://doi.org/10.1175/1520-0493\(2001\)129<2461:AOAPMS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2461:AOAPMS>2.0.CO;2)*
- Geleyn, J.-F., Catry, B., Bouteloup, Y., and Brožková, R., 2008: A statistical approach for sedimentation inside a microphysical precipitation scheme. Tellus A 60, 649–662. <https://doi.org/10.1111/j.1600-0870.2008.00323.x>*

- Gerard, L. and Geleyn, J.-F., 2005: Evolution of a subgrid deep convection parametrization in a limited-area model with increasing resolution. *Quart. J. Roy. Meteorol. Soc.* 131, 2293–2312. <https://doi.org/10.1256/qj.04.72>
- Gerard, L., 2007: An integrated package for subgrid convection, clouds and precipitation compatible with the meso-gamma scales. *Quart. J. Roy. Meteorol. Soc.* 133, 711–730. <https://doi.org/10.1002/qj.58>
- Gerard, L., Piriou, J.-M., Brožková, R., Geleyn, J.-F., Banciu, D., 2009: Cloud and precipitation parameterization in a meso-gamma scale operational weather prediction model. *Mon. Weather Rev.* 137, 3960–3977. <https://doi.org/10.1175/2009MWR2750.1>
- Grams, C.M., Binder, H., Pfahl, N.P.S., and Wernli, H., 2014: Atmospheric processes triggering the Central European floods in June 2013. *Nat. Hazards Earth Syst. Sci. Discuss.* 2, 427–458. <https://doi.org/10.5194/nhessd-2-427-2014>
- Haiden, T., Kann, A., Pistotnik, G., Stadlbacher, K., Wittmann, C., 2009: Integrated Nowcasting through Comprehensive Analysis (INCA) system description., *ZAMG Rep.*, Vienna, Austria. http://www.zamg.ac.at/fix/INCA_system.pdf
- Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction system, *Weather Forecast*, 15, 559–570. [https://doi.org/10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2)
- Hoffmann, R.N. and Kalnay, E., 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting, *Tellus*, 35A, 100–118. <https://doi.org/10.1111/j.1600-0870.1983.tb00189.x>
- Law, K. J. and Stuart, A. M., 2012: Evaluating Data Assimilation Algorithms. *Mon. Wea. Rev.* 140, 3757–3782. <https://doi.org/10.1175/MWR-D-11-00257.1>
- Lorenz, E., 1963: Deterministic non-periods flows. *J. Atmos. Sci.* 20, 130–141. [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2)
- Murphy, J., Sexton, D.M.H., Barnett, D.N., Jones, G.S., Webb, M.J., Collins, M., Stainforth, D.A., 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations, *Nature* 430, 768–772. <https://doi.org/10.1038/nature02771>
- Noilhan, J. and Planton, S., 1989: A Simple Parameterization of Land Surface Processes for Meteorological Models. *Mon. Weather Rev.* 117, 536–549. [https://doi.org/10.1175/1520-0493\(1989\)117<0536:ASPOLS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<0536:ASPOLS>2.0.CO;2)
- Palmer, T., 2001: A nonlinear dynamical perspective on model error: A proposal for nonlocal stochastic-dynamic parametrisation in weather and climate prediction models. *Quart. J. Roy. Meteorol. Soc.* 127, 279–304. <https://doi.org/10.1002/qj.49712757202>
- Palmer, T., Buizza, R., Hagedorn, R., Lawrence, A., Leutbecher, M., and Smith, L., 2005/06: Ensemble prediction: A pedagogical perspective, *ECMWF Newsletter* 106.
- Palmer, T., 2017: The primacy of doubt: Evolution of numerical weather prediction from determinism to probability, *J. Adv. Model. Earth Syst.* 9, 730–734. <https://doi.org/10.1002/2017MS000999>
- Rodwell, M., 2005/06: Comparing and combining deterministic and ensemble forecasts: How to predict rainfall occurrence better. *ECMWF Newsletter* 106.
- 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., Smoliková, P., Španiel, O., Tudor, M., Wang, Y., Wittmann, C., Joly, A., 2017: 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 Kalnay, E., 1993: Ensemble forecasting at NMC: the generation of perturbation. *Bull. Amer. Meteor. Soc.* 74, 2317–2330. [https://doi.org/10.1175/1520-0477\(1993\)074<2317:EFANTG>2.0.CO;2](https://doi.org/10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2)
- Váňa, F., Bénard, P., Geleyn, J.-F., Simon, A., and Seity, Y., 2008: Semi-Lagrangian advection scheme with controlled damping: An alternative to nonlinear horizontal diffusion in a numerical weather prediction model. *Quart. J. Roy. Meteor. Soc.* 134, 523–537. <https://doi.org/10.1002/qj.220>
- Vokoun, M. and Hanel, M., 2018: Comparing ALADIN-CZ and ALADIN-LAEF Precipitation Forecasts for Hydrological Modelling in the Czech Republic, *Advances in Meteorology*, 2018, 14 pages. <https://doi.org/10.1155/2018/5368438>

- Wang, Y., Belluš, M., Wittmann, C., Steinheimer, M., Weidle, F., Kann, A., Ivatek-Šahdan, S., Tian, W., Ma, X., Taşcu, S., Bazile, E., 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>
- Wang, Y., Bellus, M., Geleyn, J.-F., Ma, X., Tian, W. and Weidle, F., 2014: A new method for generating initial condition perturbations in a regional ensemble prediction system: blending. *Mon. Weather Rev.* 142, 2043–2059. <https://doi.org/10.1175/MWR-D-12-00354.1>
- Wang, Y., Belluš, M., Ehrlich, A., Mile, M., Pristov, N., Smoliková, P., Španiel, O., Trojáková, A., Brožková, R., Cedilnik, J., Klarić, D., Kovačić, T., Mašek, J., Meier, F., Szintai, B., Taşcu, S., Vivoda, J., Wastl, C., Wittmann, C2018: 27 years of Regional Cooperation for Limited Area Modelling in Central Europe (RC LACE). *Bull. Amer. Meteorol. Soc.* 99, 1415–1432. <https://doi.org/10.1175/BAMS-D-16-0321.1>
- Wastl C., Simon, A., Wang, Y., Kulmer, M., Baár, P., Bölöni, G., Dantinger, J., Ehrlich, A., Fischer, A., Frank, A., Heizler, Z., Kann, A., Stadlbacher, K., Szintai, B., Szűcs, M., Wittmann, C2018: A seamless probabilistic forecasting system for decision making in Civil Protection, *Meteorol. Zeit.* 27, 417–430. <https://doi.org/10.1127/metz/2018/902>
- 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, *Atmosph.-Ocean* 51, 284–295. <https://doi.org/10.1080/07055900.2013.802217>
- Wernli, H., Paulat, M., Hagen, M., and Frei, C., 2008: SAL - a novel quality measure for the verification of quantitative precipitation forecasts. *Mon. Weather Rev.* 136, 4470–4487. <https://doi.org/10.1175/2008MWR2415.1>
- Wittmann, C., Haiden, T., and Kann, A., 2010: Evaluating multi-scale precipitation forecasts using high resolution analysis. *Adv. Sci. Res.* 4, 89–98. <https://doi.org/10.5194/asr-4-89-2010>