**Impact of land surface stochastic physics in ALADIN-LAEF**

Yong Wang1 | Martin Belluš2 | Florian Weidle1 | Christoph Wittmann1 | Jian Tang3 | Florian Meier1 | Fan Xia4 | Endi Keresturi5

1 Department of Forecasting Models, Zentralanstalt für Meteorologie und Geodynamik, Vienna, Austria
2 NWP Division, Slovak Hydro-meteorological Institute, Bratislava, Slovakia
3 NMC, China Meteorological Administration, Beijing, China
4 Shangdong Meteorological Bureau, Jinan, China
5 NWP Section, Croatian Meteorological and Hydrological Service, Zagreb, Croatia

**Correspondence**
Yong Wang, Zentralanstalt für Meteorologie und Geodynamik, Hohe Warte 38, A-1190 Wien, Austria.
Email: yong.wang@zamg.ac.at

**Funding information**
International Cooperation

**Abstract**
To deal with the land surface physics uncertainties, a stochastic scheme based on stochastic perturbation of physics tendencies is implemented and tested. The impact of land surface physics uncertainties and their relative importance to land surface initial uncertainties are investigated in the regional ensemble forecasting system ALADIN-LAEF (Aire Limitée Adaptation Dynamique Développement InterNational – Limited Area Ensemble Forecasting). The land surface initial perturbation is generated by using an ensemble of land surface data assimilation; and the land surface physics uncertainties by applying the idea of stochastically perturbed parametrization tendencies (SPPT) scheme. Three experiments are conducted and compared with the reference ensemble over a 2-month period. The results show the introduction of land surface stochastic physics increases the ensemble spread, reduces the ensemble bias, and keeps neutral in deterministic forecast skill of the ensemble, its impact strongly depending on the quality of ensemble initial conditions. The ensemble land surface data assimilation has stronger positive impact on the ALADIN-LAEF than the land surface stochastic physics for screen-level temperature and humidity. There is not much impact on 10 m wind and precipitation. Best results are obtained when both the ensemble land surface data assimilation and land surface stochastic physics are used simultaneously; it gives a more reliable and statistically consistent forecast, which is contributed mainly by ensemble land surface data assimilation in the first forecast hours and largely by land surface stochastic physics in the later forecast hours.

**KEYWORDS**
ensemble prediction, land surface physics, numerical weather prediction, probabilistic forecast, stochastic physics, surface perturbation

**1 INTRODUCTION**
Errors in initial conditions could grow unpredictably with time and therefore limit the predictability due to the chaotic nature of the atmosphere (Lorenz, 1963; 1965). Model errors from physical parametrizations and dynamical core reduce the limit of the predictability even further (Buizza et al., 1999; Wandishin et al., 2001). An objective way to quantify model forecast uncertainty is ensemble forecasting. A skilful ensemble prediction system (EPS) should provide sharp and reliable
probabilistic forecasts (Sutton et al., 2006). However, forecasts of near-surface variables in many EPSs tend to be rather unreliable and under-dispersive (Hamill and Colucci, 1997; Mullen and Buizza, 2001; Sutton et al., 2006; Clark et al., 2009; Wang et al., 2012; Romine et al., 2014; Schwartz et al., 2014). Among many possible reasons why it is a particular problem near the surface, the lack or inadequate representation of errors in land surface physics and in land surface initial conditions (ICs) can be definitively counted as major contributors (Wang et al., 2010; Lavaysse et al., 2013; Tennant and Beare, 2014; Belluš et al., 2016; Bouttier et al., 2016). While perturbation of land surface ICs is very important, our focus here is the representation of land surface physics errors in an operational regional EPS. The land surface ICs perturbation in a regional EPS has been discussed in detail by Tennant and Beare (2014), Bouttier et al. (2016) and Belluš et al. (2016).

The representation of uncertainties due to model physics in the free atmosphere has been well studied using multi-physics, multi-model, perturbation of empirical parameters in physical parametrizations, and stochastic physics (Teixeira and Reynolds, 2008; Charron et al., 2010; Schwartz et al., 2010; Berner et al., 2011; 2012; Gebhardt et al., 2011; Hacker et al., 2011; Bengtsson et al., 2013; Christensen et al., 2015; Shutts, 2015; Kober and Craig, 2016; Christensen et al., 2017; Leutbecher et al., 2017; Ollinaho et al., 2017). One of the stochastic physics methods is the stochastically perturbed parametrization tendencies (SPPT) scheme, which addresses model uncertainty due to the physics parametrization schemes by perturbing the physics tendencies using multiplicative noise. It has been widely applied in many operational EPSs for dealing with the model uncertainties in the atmosphere (Buizza et al., 1999; Palmer et al., 2009; Charron et al., 2010; Bouttier et al., 2012; Romine et al., 2014; Sanchez et al., 2016). SPPT keeps the proper balance between the tendencies due to different parametrizations, and it is a simple and effective scheme in improving probabilistic skill in the EPSs (Leutbecher et al., 2017; Palmer, 2018).

There are few studies in the literature for quantifying land surface physics uncertainties in EPSs, especially in the operational regional EPSs. The land surface physics scheme includes a complex set of equations and parametrizations to represent mass and energy flux at the surface and below (Cloke et al., 2011). The quantification of the errors in the parametrization of the land surface physical processes is not straightforward, and the forecast uncertainties due to land surface parametrization are little studied. Du et al. (2007) suggested different land surface physics schemes in SREF (Short Range Ensemble Forecasting) at National Centers for Environmental Prediction (NCEP). A multi-scheme of land surface physics was applied by Hacker et al. (2011) and Berner et al. (2011) in the US Air Force Weather Agency joint mesoscale ensemble. Gebhardt et al. (2011) varied slightly some empirical parameters in their land physics scheme in COSMO-DE-EPS (COnsortium for Small scale Modelin DEutschland EPS). Orth et al. (2016) introduced land surface parameter perturbations into the coupled European Centre for Medium-range Weather Forecasts (ECMWF) forecast, showing improvements in forecast skill.

Lavaysse et al. (2013) applied a Markov chain process (Li et al., 2008; Charron et al., 2010) to randomly perturb some key land surface variables in the Canadian Regional Ensemble Prediction System. They found that the surface perturbation has especially positive impact on 2 m temperature and 10 m wind during daytime. Similarly Tennant and Beare (2014) perturbed sea-surface temperature (SST) with a first-order Markov process (Frankignoul, 1985) in the Met Office Global and Regional Ensemble Prediction System (MOGREPS: Bowler et al., 2008). It reduces the spread deficit of surface temperature. Bouttier et al. (2016) applied the same idea in a convective-scale ensemble Application of Research to Operations at MEsoscale (AROME-EPS), and showed that there is improvement on the forecast performance of 2 m temperature and 2 m humidity, but not on precipitation. In their method (Lavaysse et al., 2013; Tennant and Beare, 2014; Bouttier et al., 2016), the perturbations are generated for each member, starting date, and land surface climatological and prognostic variables. They are kept constant for the duration of the model integration, except for the prognostic variables. Their uncertainties can evolve freely in the model integration.

Ideally, the surface perturbations should be time dependent, because the corresponding errors may evolve over time (Bouttier et al., 2016) and interact between the parameters. An attractive technique for physics perturbation is the SPPT scheme. Recently, MacLeod et al. (2016) applied the SPPT approach and a perturbed parameter approach to the land surface tendencies in the ECMWF seasonal forecast, finding an improved representation of the 2003 European heatwave.

In this article a Stochastic perturbation of land surface PHYSics tendencies (SPHY) scheme is implemented in the regional ensemble system ALADIN-LAEF (Aire Limitée Adaptation dynamique Développement InterNational – Limited Area Ensemble Forecasting: Wang et al. 2010; 2011). The idea of SPHY is rather the same as the SPPT of ECMWF (Palmer et al., 2009), but applied in the land surface physics scheme Interactions Soil–Biosphere–Atmosphere (ISBA: Noilhan and Planton, 1989; Giard and Bazile, 2000) in ALADIN-LAEF. The purpose of this article is threefold: (a) the impact of SPHY on the performance of ALADIN-LAEF is studied with the focus on the screen-level weather parameters; (b) the performance of SPHY is compared against the land surface initial condition perturbation scheme Ensemble of land Surface Data Assimilation (ESDA: Belluš et al., 2016) of ALADIN-LAEF aiming at better understanding of the relative importance of land surface physics perturbation; (c) the forecast impact of combined land surface initial
condition (ESDA) and physics (SPHY) perturbations on ALADIN-LAEF performance is evaluated. All experiments are conducted with ALADIN-LAEF, and a two-month late spring to summer trial in 2011 is chosen for the statistical evaluation and comparison.

In the following, section 2 gives an overview of ALADIN-LAEF configuration. Section 3 describes the Stochastic perturbation of land surface PHYsics tendencies scheme (SPHY). Section 4 presents the design of experiments and defines the verification metrics. Section 5 discusses in detail the results from a two-months summer period, and a summary and conclusions follow in section 6.

2 ALADIN-LAEF

The operational regional ensemble ALADIN-LAEF has been developed in the framework of Regional Cooperation for Limited Area modelling Central Europe (RC LACE: Wang et al., 2018). The model core ALADIN is a hydrostatic spectral limited-area model based on hybrid vertical coordinate, semi-implicit semi-Lagrangian two-time-level advection scheme, fourth-order implicit linear horizontal diffusion, digital filter initialisation, and set of parametrizations of unresolved physics processes (Wang et al., 2006). Within ALADIN-LAEF, ALADIN is integrated over a domain of 600×500 horizontal grid points with 11 km grid spacing, and 45 levels in the vertical from surface to 10 hPa. The domain includes most of Europe and large parts of the North Atlantic, the Mediterranean Sea and the Black Sea (Figure 1). ALADIN-LAEF has 16 perturbed members, and runs operationally twice per day at 0000 and 1200 UTC with a 72 h forecast range.

FIGURE 1 Domain and model topography of ALADIN-LAEF. Verification of ensemble experiments is done over the inner limited-area domain in red

The land surface physics scheme in ALADIN-LAEF is based on the Interactions Soil–Biosphere–Atmosphere (ISBA) scheme described by Noilhan and Planton (1989) and Giard and Bazile (2000). It follows the force-restore method (Bélair et al., 2003), which is a method to approximate the surface energy budget that accounts for storage of heat in the top layer of soil. In this method the ground is conceptually split into two layers: a relatively thin layer near the top with uniform temperature, and a deep soil layer also of uniform but different temperature. The thickness of the top layer is carefully chosen based on the region where a periodic temperature surface forcing has appreciable effect. The net result is that flux from the deep soil layer tends to restore the top layer, opposing any radiative forcing from the atmosphere. This method is an alternative to a multilayer soil model, which is computationally more expensive (Bhumralkar, 1975; Blackadar, 1976; Deardorff, 1978).

In ALADIN-LAEF a two-layer version (surface layer and deep soil layer) of ISBA is in operation where overall 10 prognostic variables are used to describe the state of the soil and the exchange with the atmosphere, for example, surface temperature, deep soil temperature, surface reservoir water content, deep reservoir water content, interception reservoir water content, and the water content of the snow cover. Physical characteristics include vegetation fraction, land-use type (sea, ice, low vegetation or desert, high vegetation), minimum surface resistance, clay and sand percentages, active soil depth, leaf-area index, and thermal roughness length.

The perturbed initial atmospheric conditions are generated by a blending cycle (Wang et al., 2014), which combines large-scale perturbations from the first 16 initial ECMWF-EPS perturbed members with small-scale perturbations from ALADIN-LAEF breeding vectors. The perturbed lateral boundary conditions of ALADIN-LAEF are provided by the corresponding ECMWF-EPS members and are updated every 6 h (Weidle et al., 2013). A multi-physics approach is implemented in ALADIN-LAEF to account for model uncertainties in the atmosphere. Model configurations with different combinations of parametrizations and tuning settings are applied for each perturbed ensemble member (Wang et al., 2011). The perturbed initial land surface conditions, such as soil moisture and surface temperature, are obtained through an ensemble of land surface data assimilation (Beluš et al., 2016).

3 PERTURBATIONS

This section introduces the land surface physics perturbation scheme: Stochastic perturbation of land surface PHYsics tendencies (SPHY).

The SPHY is an adaption of ECMWF’s SPPT scheme (Palmer et al., 2009) implemented on ISBA land surface scheme. SPPT in the atmosphere is designed to represent
the uncertainty arising from the parametrization, which is attempting to represent unresolved subgrid processes. Because there are unresolved subgrid-scale processes in the surface (e.g. heterogeneous flow), the vision of SPPT as a representation of uncertainty from unresolved subgrid processes is thus still valid in the land surface. Further, because its formulation is general (i.e. not tied to a particular process), SPPT can also be used to represent errors in resolved processes, and one can use SPPT as a tool to model their statistical effect on an ensemble (Bouttier et al., 2012). In the implementation of SPHY a multiplicative noise is added to the net tendencies of surface prognostic variables. The net tendency is defined as the amount of changes in prognostic variables contributed by physical processes during one time step. For a given surface variable the tendency \( X_s \) is replaced by the formulation

\[
X'_s = X_s \left( 1 + r(\lambda, \varphi, t) \right),
\]

with the perturbed tendency \( X'_s \) and a random pattern \( r \), where \( \lambda, \varphi, t \) are longitude, latitude and time respectively. The random pattern \( r \) is the crucial part in the stochastic representation of model uncertainties in SPHY. Buizza et al. (1999) proposed the random pattern by drawing random numbers from a uniform distribution covering the range \([-0.5, 0.5]\). In order to include spatial correlation, the same random numbers are then assigned for an area on the model domain with a given size (e.g. \(1^\circ \times 1^\circ\)). Temporal correlation is included by keeping the random numbers constant for a given number of model time steps. Palmer et al. (2009) revised the random pattern generation in a more sophisticated way. They used a spectral random generator, i.e. the random pattern \( r \) is generated in spectral space \( (r_{mn} \in \mathbb{C}) \) and transformed back to grid-point space to be applied in the tendency equations:

\[
r = \sum_{mn} \hat{r}_{mn} Y_{mn}.
\]

The formulation of the spectral random generator in SPHY follows Palmer et al. (2009) and was modified by Bouttier et al. (2012) to work with the Bi-Fourier functions used in ALADIN model. The evolution of the spectral coefficients is described through an autoregressive model of order 1 (AR1):

\[
\hat{r}_{mn}(t + \Delta t) = \phi \hat{r}_{mn}(t) + \sigma_n \eta_{mn}(t).
\]

For a correlation time-scale \( \tau \), \( \phi = \exp \left( -\frac{\Delta t}{\tau} \right) \) defines the temporal correlation and real and imaginary part of the random numbers, \( \eta_{mn} \in \mathbb{C} \) are drawn from a Gaussian distribution with zero mean and unit variance. The variance factor \( \sigma_n \) controls the amplitude of the random numbers or the size of the distribution from which we draw them. As described in Palmer et al. (2009), the definition of \( \sigma_n \) includes the so-called correlation length-scale \( L \), which can be used to control the horizontal structure of the pattern. Finally, the amplitude, horizontal shape and evolution in time of the random pattern \( r \) can be controlled by the factors \( L, \sigma \) and \( \tau \). In the SPHY implementation, we applied the standard deviation of normal distribution \( \sigma = 0.25 \), spatial correlation length-scale \( L = 500 \) km and time correlation \( \tau = 2 \) h. These values are chosen based on several SPHY experiments with ALADIN-LAEF. For more details about the stochastic settings please refer to Belluš (2014).

During each model integration, the randomly generated pattern \( r \) is applied on the net tendency of surface physical variables, which are summarized in Table 1.

We intentionally do not perturb the deep soil prognostic variables (e.g. deep soil temperature), because such variables are naturally changing very slowly in time and their perturbation could be contradictory due to their slow response. On the other side, we found the perturbation of (skin) surface prognostic variables very important for generating enough spread for screen-level variables in the regional EPS.

To demonstrate the impact of the stochastic perturbation, we made some analysis of the perturbed land surface fields themselves. Figure 2 shows a case-study for comparison between the ensemble mean (left) and ensemble spread (right) for surface temperature (upper) and surface water (bottom). Ensemble mean and spread were computed using 16 perturbed members. It presents a positive between the spread of surface temperature and the spread of surface water content. We also compared the perturbations of surface ice, surface snow and surface temperature after 18 h of integration for one selected EPS member (not shown). There is no clear response between ices, snow with the surface temperature; it might be due to the time of year of the experiment.

## 4 EXPERIMENTAL SET-UP

### 4.1 Experiments

To investigate the impact of the land surface stochastic physics in ALADIN-LAEF, four experiments were conducted. Table 2 gives the description of the ALADIN-LAEF

| Variable | Description | Unit |
|----------|-------------|------|
| \( T_s \) | Surface temperature | (K) |
| \( W_s \) | Surface reservoir water content | (kg m\(^{-2}\)) |
| \( W_i \) | Interception reservoir volumetric water contents | (kg m\(^{-2}\)) |
| \( S_n \) | Snow reservoir snow water equivalent | (kg m\(^{-2}\)) |
| \( R_n \) | Snow density | (kg m\(^{-2}\)) |
| \( A_n \) | Snow albedo |
| \( W_{si} \) | Surface soil ice reservoir | (kg m\(^{-2}\)) |
**FIGURE 2** (a,c) Ensemble mean and (b,d) ensemble spread for (a,b) surface temperature (°C) and (c,d) surface water content (kg/m²). Ensemble mean and spread were computed using 16 perturbed SPHY members, starting at 1200 UTC 15 May 2011 for +24 h forecast.

**TABLE 2** ALADIN-LAEF set-up for experiments regarding the land surface initial perturbation, and land surface physics perturbation

| Experiment | Land surface initial perturbation | Land surface physics perturbation |
|------------|----------------------------------|----------------------------------|
| REF0       | No                               | No                               |
| SPHY       | No                               | Stochastics perturbation of land surface physics tendencies |
| ESDA       | Ensemble of land surface data assimilation | No |
| COMB       | Ensemble of land surface data assimilation | Stochastics perturbation of land surface physics tendencies |

set-up for the experiments regarding the land surface initial perturbation, and land surface physics perturbation. REF0 is the reference experiment in which there is no land surface stochastic physics and no ensemble of land surface data assimilation; SPHY is the experiment with introduction of land surface physics uncertainties, where the land surface physics in ALADIN-LAEF is perturbed by using stochastic perturbation of land surface physics tendencies; ESDA is the ensemble with land surface initial uncertainties, where the land surface initial conditions in ALADIN-LAEF are perturbed by using an ensemble of land surface data assimilation; and COMB is the combination of SPHY and ESDA, where both the land surface physics and initial conditions in
ALADIN-LAEF are perturbed by using the same perturbation techniques as in SPHY and ESDA.

The ensemble of land surface data assimilation scheme in ESDA is described in detail in Belluš et al. (2016). It is based on the ALADIN optimal interpolation technique Code d’Analyse Nécessaire à ARPEGE pour ses Rejets et son Initialisation (CANARI: Mahfouf, 1991, Bouttier et al., 1993). The CANARI land surface assimilation provides the initial conditions of four prognostic variables in ISBA: surface temperature $T_s$, deep soil temperature $T_p$ and soil moisture content ($W_S$, $W_p$) in both layers, while the other prognostic variables, such as snow reservoir $S_n$, snow density $\rho_s$, snow albedo $A_n$ and water content $W_r$ of the interception layer, are kept unchanged from the first guess. It is noted that the perturbed variables in ESDA are different from those in SPHY. For a consistent comparison between ESDA and SPHY, it would make more sense to perturb the same variables in both experiments. In ESDA, the increments of the variables $T_s$, $T_p$, $W_S$ and $W_p$ are not analysed directly, but rather are linked to the analysed 2 m temperature and 2 m relative humidity increments, which are transformed to soil temperature $\Delta T_S$, $\Delta T_P$ and soil moisture $\Delta W_S$, $\Delta W_P$ increments using the empirically calibrated linear relations from Giard and Bazile (2000). The observation error statistics and the background error statistics are kept unchanged in time and space. CANARI uses an isotropic and with distance exponentially decaying spatial background error correlation function (Giard and Bazile, 2000). In ESDA all the observations are randomly perturbed by Gaussian distributed uncorrelated random numbers with zero mean normed by the assumed observation errors (Storto and Randriamampianina, 2010; Horanyi et al., 2011) using a random number generator based on the code “Zufall” (Petersen, 1994; Burns and Pryor, 1998). The CANARI was run in a 12 h assimilation cycle, where the short range ALADIN-LAEF forecasts were used as the perturbed model backgrounds for the ensemble of assimilation.

In all four experiments, ALADIN-LAEF couples with ECMWF EPS, which provides the atmospheric initial and lateral perturbations. For clean comparison, there are no other perturbations employed in the ALADIN-LAEF experiments for this study, such as blending of initial atmospheric perturbations and multi-physics in the atmospheric model. All ensemble experiments run once per day with 16 perturbed members, which started at 1200 UTC and integrated up to 54 h forecasts. A 2-month period (15 May–15 July 2011) covering late spring and early summer is chosen for the investigation. It is expected that the weather during this period was more sensitive to the land surface physics due to more pronounced evapotranspiration and less large-scale forcing but more convective activity.

### 4.2 Verification

The experiments were evaluated using some standard verification measures which assess the quality of probabilistic forecasts of scalar quantities. Those are the bias, ensemble spread, root-mean-square error (RMSE) of ensemble mean, spread-error skill and continuous ranked probability score (CRPS). One could find detailed discussions on the scores in Hamill and Colucci (1997), Jolliffe and Stephenson (2003), Stanski et al. (1989) and Wilks (2006).

- **Bias.** It is defined as the ensemble systematic error in the ensemble mean.
- **RMSE of ensemble mean.** It is the squared error of the ensemble mean, which describes the deterministic accuracy of the ensemble and gives a measure of the overall ensemble performance.
- **Ensemble spread.** It is the standard deviation of the ensemble members with respect to their mean. It measures the dispersion of the ensemble forecast.
- **Spread-Error skill.** It gives an evaluation of the statistical reliability of the ensemble. It is the relationship between spread and the ensemble RMSE. For a statistically consistent ensemble, the magnitude of ensemble spread should correspond to the magnitude of the RMSE of ensemble mean. A large difference between the RMSE of the ensemble mean and the ensemble spread is an indication of statistical inconsistency (Buizza et al., 2005).
- **Continuous Ranked Probability Score (CRPS).** It is an overall measure of the skill of a probabilistic forecast, measuring the skill of the ensemble mean forecast as well as the ability of the perturbations to capture the deviations around it (Bowler and Mylne, 2009). CRPS is the generalized form of the discrete ranked probability score, simulating the mean over all possible thresholds. As noted by Hersbach (2000), CRPS is analogous to an integrated form of the Brier score, which can be decomposed into reliability, resolution, and uncertainty. The CRPS orientates negatively, so smaller values are better; and it rewards concentration of probability around the step function located at the observed value. A perfect CRPS score is zero, as with the Brier score (Wilks, 2006).

The verification is performed at the observation station sites within a limited area of interest over Central Europe as shown in Figure 1, in which there are more than 1,200 observation stations in the verification domain and all of them were used for verification (not shown). The statistical scores are computed by using ensemble forecast against in situ observation. Ensemble forecast values are interpolated to the observation site for smoothly varying fields, such as 2 m temperature, 2 m humidity and 10 m wind speed. For the forecast variable with strong spatial gradients, such as precipitation, the observation is matched to the nearest grid point.
The performance of ensemble is rather difficult to score objectively in some sense, as it requires very large independent samples and observations for an objective verification (Candille and Talagrand, 2008; Bouttier et al., 2016). It is then very important to apply the statistical significance test in the ensemble verification. We use a bootstrap method (Wilks, 2006) for statistical significance test to score the differences between the experiments. The 95% and 5% confidence intervals are used for the ensemble experiments to decide whether an average score difference between the ensembles is statistically significant or not.

5 | RESULTS

In this study, we focus our investigation on the screen-level variables, such as 2 m temperature (T2m), 2 m relative humidity (RH2m), 10 m wind (W10m) and precipitation (PREC), as these variables are of most interest in an operational regional ensemble system.

5.1 | Impact of the stochastic perturbation of land surface physics tendencies

In this section the impact of the stochastic land surface physics perturbation is evaluated by comparing the experiment SPHY with the reference experiment REF0. Figure 2 shows the ensemble bias of all the four experiments for T2m, W10m, RH2m and PREC, averaged over the verification domain and period. The introduction of land surface stochastic physics SPHY improves the T2m bias during the night, which reduces the diurnal cycle of the T2m bias. For other variables W10m, RH2m and PREC, the overall impact of SPHY is neutral; there is little difference in their ensemble bias score. Both experiments REF0 and SPHY show a strong cold bias in T2m and moist bias in RH2m especially at daytime. This could largely be due to the surface coupling of ALADIN with ECMWF, namely, the different surface parametrization schemes used in the ALADIN and ECMWF models. This inconsistency, in particular in the soil moisture and soil temperature, introduces strong bias in the 2 m temperature and 2 m humidity (Wang et al., 2012).

The RMSE of the ensemble mean and the ensemble spread of T2m, RH2m, W10m and PREC for REF0 and SPHY are shown in Figure 3. The ensemble RMSE score of SPHY is very similar to REF0 for all those screen-level variables; it means that SPHY has little impact on the forecast quality in a deterministic sense. SPHY improves the ensemble spread of T2m and RH2m statistically significantly, while it is rather neutral for W10m and PREC. Both experiments are very under-dispersive, as the ensemble spread is much smaller than the RMSE of the ensemble mean.

Figure 4 presents the RMSE of the four screen-level variables as a function of the forecast range. The positive impact of SPHY is obvious for T2m, especially during the night hours, while for the other variables SPHY does not improve the probability skill.

In summary, SPHY keeps similar quality on the forecast in the deterministic sense, but it has higher skill in the spread of T2m and RH2m, smaller bias, and therefore a better CRPS skill. Such a result is expected, because the stochastic physics perturbation should keep the same deterministic quality of the model but increase the ensemble diversity in a probabilistic sense. Further, the perturbation on the tendency in SPHY is mainly focused on those tendencies related to soil temperature and moisture, which have direct impact on the T2m and RH2m. The short-range forecasts of W10m and PREC are influenced largely by physics and dynamics in the upper air and in the lower boundary layer, therefore SPHY could not have much impact on them.

It could be concluded that the surface stochastic physics would increase the ensemble spread, but not improve the RMSE of the ensemble mean, which is the overall deterministic quality score of an ensemble. It seems that the surface stochastic physics alone would have little higher skill in predicting the surface weather variables in the short range. A better surface initial analysis and its perturbation for the ensemble might be necessary to improve the short-range forecast of screen-level weather variables. The impact of the land surface initial perturbations on ALADIN-LAEF comparing to SPHY is investigated in the next section.

5.2 | Importance of land surface initial perturbations and stochastic physics perturbations

In this section, the performance of SPHY is compared against the land surface initial condition perturbation scheme ESDA (Belluš et al., 2016) of ALADIN-LAEF aiming at a better understanding of the relative importance of land surface physics perturbation.

Figure 6 shows the CRPS evaluated for the experiments of SPHY and ESDA, minus the CRPS evaluated for the reference experiment REF0; negative values mean that the CRPS of the given experiment has improved in comparison to the reference one without the surface ensemble data assimilation and without surface stochastic physics.

SPHY has positive impact on T2m in the night hours, and neutral impact on RH2m compared to the reference REF0, while ESDA significantly improves the T2m and RH2m forecast for all lead times; it is more skillful than SPHY. ESDA has larger positive impact on the forecast during the day than in the night. It is obvious that ESDA starts with a lower CRPS score at the lead time zero. This result indicates that the statistically more reliable performance of ESDA might be largely due to the surface data assimilation and the observation perturbation. It provides better surface initial conditions for
FIGURE 3  Ensemble bias as function of lead time (h) for forecasts of experiments REF0 (yellow lines), SPHY (blue lines), ESDA (green lines) and COMB (red lines) for (a) T2m (K), (b) W10m (m/s), (c) RH2m (%), and (d) PREC (mm/12 h), averaged over the verification domain (Figure 1) and over the verification period. Bars show the 95% and 5% confidence intervals.

ALADIN-LAEF than the experiments of SPHY and REF0, in which the surface initial conditions come from the downsampling of ECMWF ensemble.

For W10m, ESDA leads to a slightly positive impact, while SPHY is largely neutral to the reference. It could be explained by the fact that we perturb and assimilate only the surface temperature and soil moisture content by using the increment of the 2 m temperature and relative humidity in ESDA (Belluš et al., 2016); hence ESDA has more influence on T2m and RH2m, but not much on W10m.

For PREC, both ESDA and SPHY cannot outperform the reference REF0 at most lead times, and ESDA is even worse than SPHY. As an example, we give the comparison of the spread-skill relationship valid at 18, 30, 42 and 54 h in Figure 7. It shows the difficulty of improving the ensemble precipitation forecast with perturbed land surface ICs or surface stochastic physics alone. It would be very interesting to investigate the impact of combined use of ESDA and SPHY.

5.3 Impact of combining the land surface initial and physical perturbations

We evaluate the performance of the combination of land surface initial condition and physics perturbations (experiment COMB) by comparing with the experiments ESDA and SPHY. The ensemble bias of all those experiments can be
found in Figure 3. In general, ESDA performs clearly better than SPHY, especially for T2m and RH2m. It is because the surface ensemble data assimilation improves the ICs of ESDA, which provides more accurate and reasonable ICs and perturbations than SPHY. The combination experiment COMB has the smallest bias for all the variables among the three experiments. It is even statistically significant for W10m and PREC. The bias of COMB is similar to that of ESDA but still slightly better. It seems that COMB integrates the advantage of ESDA and SPHY; this result confirms the conclusion of subsections 5.1 and 5.2.

The comparison of RMSE of the ensemble mean and spread for T2m, RH2m, W10m and PREC for the three experiments is shown in Figure 4. For all the three experiments, their RMSE is much larger than the spread, which implies that the ensemble system is under-dispersive and statistically inconsistent. For T2m and RH2m, the combination COMB has the similar ensemble RMSE as ESDA, but slightly larger spread. COMB is the most skilful option for ALADIN-LAEF. It reduces the RMSE of ensemble for T2m and RH2m because of the introduction of surface ensemble data assimilation; and increases the spread due to the application of land surface stochastic physics in ALADIN-LAEF. For W10m and PREC, COMB has neutral or slightly positive impact.

Figure 5 compares the CRPS of those three experiments for T2m, RH2m, W10m and PREC. The clear superiority of COMB to ESDA and SPHY can be observed for all the
FIGURE 5  Same as Figure 3, but for continuous ranked probability score (CRPS)

variables and over the whole forecast range. The COMB gives a more reliable and statistically consistent forecast than the other two experiments. The beneficial effect of COMB is attributed to the smaller RMSE of the ensemble, less ensemble bias contributed by ESDA; and slightly larger spread contributed by SPHY. It can be concluded that it is beneficial if an ensemble system perturbs both the land surface ICs and the land surface physics. To further illustrate the results, Figure 8 gives a comparison of T2m statistical skill scores (ensemble bias, RMSE, spread and CRPS) at the forecast hour +24 h for all the three experiments for the whole 62-day verification period. It is clear that COMB performs the best at the daily basis for the whole period, which confirms the aforementioned results again.

COMB is the experiment of SPHY on top of ESDA. To better understand the impact of stochastic perturbation of land surface physics tendencies, we compared the CRPS difference between SPHY and REF0, and the difference between COMB and ESDA. It is shown in Figure 9. Those are the impact of the SPHY with different IC perturbations. The performance of SPHY strongly depends on the quality of ensemble ICs. While SPHY with ICs downscaled from ECMWF EPS has small impact on T2m, and is neutral on RH2m, W10m and PREC, SPHY with ensemble surface data assimilation has positive impact on all the screen-level variables, including W10m and PREC.

To quantify the relative contributions of each experiment to the improvement in the scores, we compute the score
FIGURE 6  Impact of experiments SPHY (blue lines) and ESDA (green lines) on (a) T2m, (b) W10m, (c) RH2m and (d) PREC, relative to the reference experiment REF0 (grey zero line). It is the CRPS score averaged over the verification domain (Figure 1) and over the verification period. Shaded areas show the 95% and 5% confidence intervals for each experiment scaled as: 100% * (SPHY – REF) / (COMB – REF), so if the score is equal to the REF, then the scaled score would be 0%. If the improvement is almost the same as COMB then the scaled score would be near 100%. Figure 10 shows the relative CRPS contribution of SPHY (blue line) and ESDA (green line) to the improvement made by COMB, using REF0 the downscaling of ECMWF EPS as the reference. The relative contribution of SPHY using ESDA as reference (SPHY* in red line) is also shown in Figure 10. The CRPS difference between COMB and REF0 as the total improvement for all the four weather variables in grey bars are also shown in Figure 10. As expected, the contribution to the improvement for temperature and moisture is mainly coming from the ESDA, except for temperature during the night. For 10 m wind, the contribution of SPHY is not neglectable in the later forecast hours; for precipitation, neither SPHY nor ESDA could contribute positively. However, it is very interesting if we consider the relative contribution of SPHY starting with ESDA. Its relative contribution for temperature is positive in general, and it is stronger in the night; for moisture is about 10% and clearly more positive than the SPHY with downscaled ECMWF ICs. For wind, its impact is clearly stronger than the SPHY on top of REF0. The impact becomes larger than those from ESDA with the forecast time. For precipitation, it gives the most contribution on the improvement by COMB. One has to notice that the impressive performance
FIGURE 7  Spread-skill score of experiments SPHY (blue dots), REF0 (yellow dots) and ESDA (green dots) for PREC, valid at the forecast hours (a) 18 h, (b) 30 h, (c) 42 h, and (d) 54 h

of SPHY at 100% on precipitation forecast at hour 30 h is mainly due to the very small total improvement by COMB, as shown in Figure 10.

It is logical that if we apply the stochastic physics on biased ICs (downscaled ECMWF EPS), it is hard to have better results. ESDA plays a double role in this experiment: (a) it perturbs the initial state of the surface prognostic fields; and (b) at the same time it reduces (systematic) model bias due to the different surface treatment in ECMWF IFS and ALADIN.

The ESDA involves the observations, but it uses also already well balanced ALADIN first guess. SPPT then perturbs the surface prognostic fields during the integration, which in case of formerly applied ESDA happens around the less biased model state. This can explain why the COMB gives better results.

6  SUMMARY AND CONCLUSIONS

To account for uncertainty in the land surface physics, a stochastic scheme based on stochastic perturbation of physics tendencies is implemented and tested. Its impact on the ALADIN-LAEF forecasts, its relative importance to the land surface initial perturbation, and the impact of the combination of land surface initial perturbation and stochastic physics perturbation on ALADIN-LAEF performance are evaluated. Three experiments are conducted and compared with the
reference ensemble without land surface physics perturbation and land surface initial perturbation.

The three experiments are evaluated by using several standard ensemble and probabilistic scores on the forecast performance of four screen-level variables T2m, RH2m, W10m and PREC over a 2-month summer period in 2011.

As expected, the introduction of land surface stochastic physics increases the ensemble spread, reduces the ensemble bias, and keeps neutral in deterministic forecast skill of the ensemble. SPHY has a better spread-skill relationship in comparison to the reference, and improves slightly the overall skill of the ALADIN-LAEF probabilistic forecast measured by CRPS. The results show that SPHY has small but evident positive impact on T2m and RH2m. It could be related to the SPHY implementation, in which the SPHY perturbation is mostly on those tendencies related to soil temperature and moisture, which has strong impact on the T2m and RH2m. SPHY does not have significant but rather neutral impact on W10m and PREC, which could be explained as the short-range forecast of W10m and PREC is influenced largely by physics and dynamics in the atmosphere.

The performance of ALADIN-LAEF strongly depends on the quality of land surface ICs used for the model integration. The ensemble land surface data assimilation ESDA has clearly stronger positive impact on the ALADIN-LAEF than the land surface stochastic physics SPHY. ESDA significantly improves the overall skill of the ALADIN-LAEF probabilistic forecast.
FIGURE 9  CRPS difference between SPHY and REF0 (blue lines) and between COMB and ESDA (red lines) on (a) T2m, (b) W10m, (c) RH2m, and (d) PREC. The CRPS score is averaged over the verification domain (Figure 1) and over the verification period. Shaded areas show the 95% and 5% confidence intervals.

Improves the probabilistic skill of T2m and RH2m forecasts at all lead times, it has larger positive impact on the forecast during the daytime than at night. It is very likely due to the land surface data assimilation in ESDA, which provides more accurate initial surface ICs than those in the experiment SPHY. The ESDA impact on W10m is positive compared to SPHY, but not as evident as for T2m and RH2m. It might be related to the fact that we perturb and assimilate only the surface temperature and soil moisture content in ESDA (Belluš et al., 2016).

For PREC, it remains a challenge to obtain an improvement with perturbations in surface ICs and surface stochastic physics alone. ESDA reduces the spread, keeps the ensemble RMSE almost unchanged, resulting in a slight deterioration of the CRPS score for precipitation, and thus making ALADIN-LAEF less reliable. It could make the forecast statistically not so reliable because of the spread-skill relationship.

When ALADIN-LAEF simulates the effect of both surface initial and surface physics related uncertainties on the forecast error (experiment COMB), its forecast performance is the best among the three experiments. COMB has the smallest ensemble bias for all the variables, which is even statistically significant for W10m and PREC. COMB reduces clearly the ensemble RMSE for T2m and RH2m because of the introduction of surface ensemble data assimilation, and increases the spread due to the application of land surface stochastic physics in ALADIN-LAEF. For W10m and PREC, COMB
The relative contribution to the improvement by COMB in the scores, computed by CRPS difference of experiment to the reference in percentage, e.g. 100% * (SPHY – REF)/(COMB – REF). The relative contribution of SPHY to COMB with REF0 as reference is in blue line; the relative contribution of ESDA to COMB with REF0 as reference is in green line; and the relative contribution of SPHY to COMB with ESDA as reference is in red line (SPHY*). The grey bars are the CRPS differences between COMB and REF0, showing the total improvement of COMB. (a) T2m, (b) W10m, (c) RH2m, and (d) PREC.

has a neutral or slightly positive impact. The superiority of COMB to ESDA and SPHY can be observed for all the variables and over the whole forecast range. The COMB gives a more reliable and statistically consistent forecast than the other two experiments. The overall results demonstrate that it would be beneficial if an ensemble system takes both the land surface ICs and the land surface physics perturbations into account.

A detailed evaluation of the relative contribution of SPHY and ESDA to the improvement by COMB shows that the ESDA contributes mainly in the first forecast hours, and the SPHY improves the ensemble forecast performance largely in the later forecast hours; this positive contribution is also true for the precipitation due to the better-balanced ICs.

Although the introduction of IC and physics perturbation improves the ALADIN-LAEF performance, the ensemble forecasts are still quite under-dispersive. This indicates that the ensemble is not fully capturing all uncertainties. The research on the perturbation of the surface analysis and physics will be continued. It is planned to investigate the impact of land surface stochastic physics in a convection-permitting ensemble. The land surface SPPT will still be one of the focuses in the future work, because of its simplicity and effectiveness, although it is somehow ad hoc.
and not very physically based (Palmer, 2018). The spatial variability and the impact of forecast error of the surface prognostic variable and its tendency should be investigated. We will try to find the optimal stochastic representation on the individual surface prognostic variable, and the physical consistency in perturbations among the prognostic variables will be checked. A new spectral pattern generator which was developed originally for usage of limited-area model (Tsyrulnikov and Gayfulin, 2017) will be tested. Furthermore, parameter-based land surface stochastic physics, similar to the ECMWF Stochastically Perturbed Parametrization scheme (SPP) will be implemented and evaluated in a convection-permitting EPS. Some encouraging results with the perturbed parameter approach for representing the land surface uncertainty have been shown by MacLeod et al. (2016) in the seasonal prediction, which suggests that the perturbed parameter approach may provide a path to improved weather and climate forecasts.

ACKNOWLEDGEMENTS

We gratefully acknowledge all the LACE colleagues who have contributed to the development of ALADIN-LAEF. ECMWF has provided the computer facilities for running ALADIN-LAEF. The work has been partly funded by ÖAD (Austrian Agency for international Cooperation in Education & Research, Project CN 14/2013).

ORCID

Yong Wang https://orcid.org/0000-0002-4246-6525
Endi Keresturi https://orcid.org/0000-0002-9626-1947

REFERENCES

Bélar, S., Crevier, L.P., Mailhot, J., Bilodeau, B. and Delage, Y. (2003) Operational implementation of the ISBA land surface scheme in the Canadian regional weather forecast model. Part I: Warm season results. Journal of Hydrometeorology, 4, 352–370.
Belluš, M. (2014) Combination of IC and model uncertainties for the surface prognostic variables in ALADIN-LAEF system. RC LACE internal report, 19. http://www.rclace.eu/File/Predictability/reports/Report_ZAMG_2015_06_mbell.pdf.
Belluš, M., Wang, Y. and Meier, F. (2016) Perturbing surface initial conditions in a regional ensemble prediction system. Monthly Weather Review, 144, 3377–3390.
Bengtsson, L., Steinheimer, M., Bechtold, P. and Geleyn, J.F. (2013) A stochastic parametrization for deep convection using cellular automata. Quarterly Journal of the Royal Meteorological Society, 139, 1533–1543.
Berner, J., Ha, S.Y., Hacker, J.P., Fournier, A. and Snyder, C. (2011) Model uncertainty in a mesoscale ensemble prediction system: stochastic versus multiphysics representations. Monthly Weather Review, 139, 1972–1995.
Berner, J., Jung, T. and Palmer, T.N. (2012) Systematic model error: the impact of increased horizontal resolution versus improved stochastic and deterministic parameterizations. Journal of Climate, 25, 4946–4962.
Bhumralkar, C.M. (1975) Numerical experiments on the computation of ground surface temperature in an atmospheric general circulation model. Journal of Applied Meteorology, 14, 67–100.
Blackadar, A.K. (1976) Modeling the nocturnal boundary layer. In: Proceedings of Third Symposium on Atmospheric Turbulence, Diffusion and Air Quality. Boston, MA: American Meteorological Society, pp. 46–49.
Bouttier, F., Mahfouf, J.-F. and Noilhan, J. (1993) Sequential assimilation of soil moisture from atmospheric low-level parameters. Parts I and II. Journal of Applied Meteorology, 32, 1335–1364.
Bouttier, F., Vié, B., Nuissier, O. and Raynaud, L. (2012) Impact of stochastic physics in a convection-permitting ensemble. Monthly Weather Review, 140, 3706–3721.
Bouttier, F., Raynaud, L., Nuissier, O. and Ménétrier, B. (2016) Sensitivity of the AROME ensemble to initial and surface perturbations during HyMeX. Quarterly Journal of the Royal Meteorological Society, 142, 390–403.
Bowler, N.E. and Mylene, K.R. (2009) Ensemble transform Kalman filter perturbation for a regional ensemble prediction system. Quarterly Journal of the Royal Meteorological Society, 135, 757–766.
Bowler, N.E., Arribas, A., Mylene, 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, 703–722.
Buizza, R., Milleer, M. and Palmer, T.N. (1999) Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. Quarterly Journal of the Royal Meteorological Society, 125, 2887–2908.
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. Monthly Weather Review, 133, 1076–1097.
Burns, P.J. and Pryor, D.V. (1998) Surface radiative transport at large scale via Monte Carlo. Annual Review of Heat Transfer, 9, 79–158.
Bæcile, G. and Talagrand, O. (2008) Impact of observational error on the validation of ensemble prediction systems. Quarterly Journal of the Royal Meteorological Society, 134, 959–971.
Charrier, M., Pellerin, G., Spacek, L., Houtekamer, P.L., Gagnon, N., Mitchell, H.L. and Michelin, L. (2010) Toward random sampling of model error in the Canadian ensemble prediction system. Monthly Weather Review, 138, 1877–1901.
Christensen, H.M., Moroz, M.I. and Palmer, T.N. (2015) Stochastic and perturbed parameter representations of model uncertainty in convection parameterization. Journal of the Atmospheric Sciences, 72, 2525–2544.
Christensen, H.M., Lock, S.J., Moroz, I.M. and Palmer, T.N. (2017) Introducing independent patterns into the Stochastically Perturbed Parametrization Tendencies (SPPT) scheme. Quarterly Journal of the Royal Meteorological Society, 143, 2168–2181.
Clark, A.J., Gallus, W.A., Jr., Xue, M. and Kong, F. (2009) A comparison of precipitation forecast skill between small convection-allowing and large convection-parameterizing ensembles. Weather and Forecasting, 24, 1121–1140.
Cloeke, H., Weisheimer, A. and Pappenberger, F. (2011) Representing uncertainties in land surface hydrology: fully coupled simulations with the ECMWF land surface scheme. ECMWF Workshop on Model Uncertainty, 20–24. (June 2011).
Deardorff, J.W. (1978) Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation. *Journal of Geophysical Research*, 83(C4), 1889–1903.

Du J, Gayno G, Mitchell KE, Toth Z, Di Mego G. (2007) Sensitivity study of T2m and precipitation forecasts to soil moisture initial condition using NCEP WRF ensemble system. 18th Conference on Numerical Weather Prediction and 22nd Conference on Weather Analysis and Forecasting, Park City, UT, Paper 11B.3. American Meteorological Society.

Frankignoul, C. (1985) Sea surface temperature anomalies, planetary waves, and air–sea feedback in the middle latitudes. *Reviews of Geophysics*, 23(4), 357–390.

Gebhardt, C., Theis, S.E., Paulat, M. and Ben Bouallègue, Z. (2011) Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. *Atmospheric Research*, 100, 168–177.

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.

Hacker, J.P., Ha, S.Y., Snyder, C., Berner, J., Eckel, F.A., Kuchera, E., Pocernich, M., Rugg, S., Schramm, J. and Wang, X. (2011) The US Air Force Weather Agency’s mesoscale ensemble: scientific description and performance results. *Tellus*, 63A, 625–641.

Hamill, T.M. and Colucci, S.J. (1997) Verification ofEta–RSM short-range ensemble forecasts. *Monthly Weather Review*, 125, 1312–1327.

Hersbach, H. (2000) Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15, 559–570.

Horanyi, A., Mile, M. and Szticz, M. (2011) Latest developments around the ALADIN operational short-range ensemble prediction system in Hungary. *Tellus*, 63, 642–651.

Jolliffe, I.T. and Stephenson, D.B. (2003) *Forecast Verification: A practitioner’s guide in atmospheric science*. Hoboken, NJ: Wiley.

Kober, K. and Craig, G.C. (2016) Physically based stochastic perturbations (PSP) in the boundary layer to represent uncertainty in convective initiation. *Journal of the Atmospheric Sciences*, 73, 2893–2911.

Lavaysse, C., Carrera, M., Béclair, S., Gagnon, N., Frenette, R., Charron, M. and Yau, M.K. (2013) Impact of surface parameter uncertainties within the Canadian regional ensemble prediction system. *Monthly Weather Review*, 141, 1506–1526.

Leutbecher, M., Lock, S.-J., Ollinaho, P., Lang, S.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., Smolarkiewicz, 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, 2315–2339.

Li, X., Charron, M., Spacek, L. and Candille, G. (2008) A regional ensemble prediction system based on moist targeted singular vectors and stochastic parameter perturbation. *Monthly Weather Review*, 136, 443–462.

Lorenz, E.N. (1963) Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20, 130–141.

Lorenz, E.N. (1965) A study of the predictability of a 28-variable atmospheric model. *Tellus*, 17, 321–333.

MacLeod, D.A., Cloke, H.L., Pappenberger, F. and Weisheimer, A. (2016) Improved seasonal prediction of the hot summer of 2003 over Europe through better representation of uncertainty in the land surface. *Quarterly Journal of the Royal Meteorological Society*, 142, 79–90.

Mahfouf, J.-F. (1991) Analysis of soil moisture from near-surface parameters: a feasibility study. *Journal of Applied Meteorology*, 30, 1534–1547.

Mullen, S.L. and Buizza, R. (2001) Quantitative precipitation forecast over the United States by the ECMWF ensemble prediction system. *Monthly Weather Review*, 129, 638–663.

Noilhan, J. and Planton, S. (1989) A simple parameterization of land surface processes for meteorological models. *Monthly Weather Review*, 117, 536–549.

Ollinaho, P., Lock, S.-J., Leutbecher, M., Bechtold, P., Beljaars, A., Bozzo, A., Forbes, R.M., Haiden, T., Hogan, R.J. and Sandu, I. (2017) Towards process-level representation of model uncertainties: stochastically perturbed parametrizations in the ECMWF ensemble. *Quarterly Journal of the Royal Meteorological Society*, 143, 408–422.

Orth, R., Dutra, E. and Pappenberger, F. (2016) Improving weather predictability by including land surface model parameter uncertainty. *Monthly Weather Review*, 146, 1551–1569.

Palmer, T.N. (2018) The ECMWF ensemble prediction system: looking back (more than) 25 years and projecting forward 25 years. *Quarterly Journal of the Royal Meteorological Society*, 144, 1–13.

Palmer TN, Buizza R, Doblas-Reyes F, Jung T, Leutbecher M, Shutts GJ, Steinheimer M, Weisheimer A. (2009) Stochastic parametrization and model uncertainty. Technical Report, ECMWF Technical Memo. 598, 42 pp. Available at: http://www.ecmwf.int/publications/.

Petersen, W.F. (1994) Lagged Fibonacci series random number generator for the NEC SX-3. *International Journal of High Speed Computing*, 6, 387–398.

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.

Sanchez, C., Williams, K.D. and Collins, M. (2016) Improved stochastic physics schemes for global weather and climate models. *Quarterly Journal of the Royal Meteorological Society*, 142, 147–159.

Schwartz, C.S., Kain, J.S., Weiss, S.J., Xue, M., Bright, D.R., Kong, F., Thomas, K.W., Levit, J.J., Coniglio, M.C. and Wandishin, M.S. (2010) Toward improved convection-allowing ensembles: model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Weather and Forecasting*, 25(1), 263–280.

Schwartz, C.S., Romine, G.S., Smith, K.R. and Weisman, M.L. (2014) Characterizing and optimizing precipitation forecasts from a convection-permitting ensemble initialized by a mesoscale ensemble Kalman filter. *Weather and Forecasting*, 29, 1295–1318.

Shutts, G.J. (2015) A stochastic convective backscatter scheme for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society*, 141, 2602–2616.

Stanski HR, Wilson LJ, Burrows WR. (1989) Survey of common verification methods in meteorology. WMO/WWW Technical Report 8, 114 pp.

Storto, A. and Randriamampianina, R. (2010) Ensemble variational assimilation for the representation of background error covariances.
in a high-latitude regional model. *Journal of Geophysical Research*, 115, D17204. https://doi.org/10.1029/2009JD013111.

Sutton, C., Hamill, T.M. and Warner, T.T. (2006) Will perturbing soil moisture improve warm-season ensemble forecasts? A proof of concept. *Monthly Weather Review*, 134, 3174–3189.

Teixeira, J. and Reynolds, C.A. (2008) Stochastic nature of physical parameterizations in ensemble prediction: a stochastic convection approach. *Monthly Weather Review*, 136, 483–495.

Tennant, W. and Beare, S. (2014) New schemes to perturb sea-surface temperature and soil moisture content in MOGREPS. *Quarterly Journal of the Royal Meteorological Society*, 140, 1150–1160.

Tsyrulnikov, M. and Gayfulin, D. (2017) A limited-area spatio-temporal stochastic pattern generator for simulation of uncertainties in ensemble applications. *Meteorologische Zeitschrift*, 26, 549–566.

Wandishin, M.S., Mullen, S.L., Stensrud, D.J. and Brooks, H.E. (2001) Evaluation of a short-range multimodel ensemble system. *Monthly Weather Review*, 129, 729–747.

Wang, Y., Haiden, T. and Kann, A. (2006) The operational limited area modelling system at ZAMG: ALADIN-AUSTRIA. *Österreichische Beiträge zu Meteorologie und Geophysik*, 37, 1–33.

Wang, Y., Kann, A., Belluš, M., Pailleux, J. and Wittmann, C. (2010) A strategy for perturbing surface initial conditions in LAMEPS. *Atmospheric Science Letters*, 11, 108–113.

Wang, Y., Belluš, M., Wittmann, C., Steinheimer, M., Weidle, F., Kann, A., Ivatek-Šahdan, S., Tian, W., Ma, X., Tascu, S. and Bazile, E. (2011) The Central European limited-area ensemble forecasting system: ALADIN-LAEF. *Quarterly Journal of the Royal Meteorological Society*, 137, 483–502.

Wang, Y., Tascu, S., Weidle, F. and Schmeisser, K. (2012) Evaluation of the added value of regional ensemble forecasts on global ensemble forecasts. *Weather and Forecasting*, 27, 972–987.

Wang, Y., Belluš, M., Geleyn, J.F., Tian, W., Ma, X. and Weidle, F. (2014) A new method for generating initial perturbations in regional ensemble prediction system: blending. *Monthly Weather Review*, 142, 2043–2059.

Wang, Y., Belluš, M., Ehrlich, A., Mile, M., Pristov, N., Smolíková, P., Španiel, O., Trojáková, A., Brožková, R., Cedilnik, J. and Klarić, D. (2018) 27 years of regional cooperation for limited area modelling in Central Europe (RC LACE). *Bulletin of the American Meteorological Society*, 99, 1415–1432.

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.

Wilks, D.S. (2006) *Statistical Methods in the Atmospheric Sciences*, 2nd edition. London: Academic Press.

---

**How to cite this article**: Wang Y, Belluš M, Weidle F, *et al*. Impact of land surface stochastic physics in ALADIN-LAEF. *Q J R Meteorol Soc*. 2019;145:3333–3350. https://doi.org/10.1002/qj.3623