

An Update to the Stochastically Perturbed Parameterization Scheme of HarmonEPS^①

ARISTOFANIS TSIRINGAKIS^{②,a}, INGER-LISE FROGNER,^b WIM DE ROOY,^a ULF ANDRAE,^c ALAN HALLY,^d SEBASTIÁN CONTRERAS OSORIO,^a SIBBO VAN DER VEEN,^a AND JAN BARKMEIJER^a

^a *Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands*

^b *Norwegian Meteorological Institute (Met Norway), Oslo, Norway*

^c *Swedish Meteorological and Hydrological Institute, Norrköping, Sweden*

^d *Irish Meteorological Service (Met Éireann), Dublin, Ireland*

(Manuscript received 22 September 2023, in final form 3 May 2024, accepted 15 May 2024)

ABSTRACT: High-resolution, limited-area forecasting is strongly affected by errors in the initial atmospheric state, lateral boundary conditions (LBCs), and physical parameterizations used by numerical weather prediction (NWP) models. These errors need to be accounted for through the introduction of uncertainty in an ensemble prediction system (EPS). One approach to account for model error is to use a stochastically perturbed parameterizations (SPPs) scheme. A first version of the SPP scheme of HARMONIE EPS (HarmonEPS) has been tested, with promising improvements in ensemble spread. However, it introduced systematic biases and deteriorated skill scores for some variables. Here, we investigate the performance of an updated version of the HarmonEPS SPP scheme, which includes (i) the use of uniform distributions, (ii) the correlation of stochastic patterns between key SPP parameters, and (iii) the introduction of four additional parameters, in the microphysics and mass-flux schemes. Two five-parameter SPP-based setups are compared against initial and LBC perturbations setups for five forecast periods: (i) 22–28 March 2019, (ii) 6–12 July 2020, (iii) 20–26 February 2021, (iv) 13–26 January 2021, and (v) 20 May–2 June 2021. We find that SPP-based experiments show better probabilistic metrics for near-surface and cloud-related variables than the non-SPP experiments. The SPP-based ensembles show increased spatial spread (as indicated by dFSS), while maintaining similar spatial skill (as indicated by eFSS) with the non-SPP experiment. In addition, the systematic bias in the ensemble members of the previous SPP iteration has been alleviated with the use of uniform distributions. Finally, the use of microphysical and mass-flux perturbations improves the ensemble scores for cloud-related variables, precipitation, and visibility.

SIGNIFICANCE STATEMENT: Ensemble prediction systems are essential for weather forecasting. Their skill depends on the representation of uncertainties arising from errors in the initial atmospheric state and model physics. In this study, we present an update to the stochastically perturbed parameterizations scheme, which represents uncertainty arising from model errors, of the limited-area HARMONIE ensemble prediction system. The update to the scheme includes the introduction of uniform distributions and correlated perturbations for the perturbed parameters, as well as the addition of four new perturbed parameters. Overall, this revision removes systematic biases in ensemble members, which were present in the previous iteration of the scheme, and shows increased ensemble spread (up to 70%) and reduced model errors (up to 5%) for near-surface and cloud-related variables.

KEYWORDS: Atmosphere; Uncertainty; Ensembles; Numerical weather prediction/forecasting; Model evaluation/performance

1. Introduction

Uncertainty is an intrinsic characteristic of the atmosphere due to its chaotic nature (Lorenz 1963). Accounting for this uncertainty is essential to forecast potential evolutions of the current atmospheric state in time, which is the aim of an ensemble prediction system (EPS). Uncertainty is generally

separated into two categories: atmospheric state uncertainty and uncertainty arising from the incomplete representation of physical properties and processes (Leutbecher et al. 2017).

Atmospheric state uncertainty is also known as initial state uncertainty. Initial uncertainty attempts to account for errors in the atmospheric observations (or their limited availability) and their assimilation in the generation of the analysis fields. Uncertainty can also arise from the incomplete representation of physical processes in the numerical weather prediction (NWP) models. This could be not only due to the lack of knowledge or limitations in model physics but also uncertainties in dynamics and dynamics–physics coupling. Although in theory the uncertainty sources should be separable, in practice, it is difficult to separate them as initial errors include physical model errors and vice versa since the analysis is generated by combining observations and model forecasts (Leutbecher and Palmer 2008).

^① Supplemental information related to this paper is available from the Journals Online website at <https://doi.org/10.1175/MWR-D-23-0212.s1>.

Tsiridakis's current affiliation: The European Centre for Medium-Range Weather Forecasts, Bonn, Germany.

Corresponding author: Aristofanis Tsiringakis, aristofanis.tsiringakis@ecmwf.int

Various methods have been developed over the years to represent initial state uncertainty. Most notable are the singular vectors (Buizza and Palmer 1995; Barkmeijer et al. 1998, 1999; Leutbecher and Lang 2014), ensemble Kalman filtering (Evensen 2003), breeding models (Toth and Kalnay 1993), and ensembles of data assimilation (EDA) (Buizza et al. 2008). For more details on the differences between these approaches, we refer to relevant review papers (Leutbecher and Palmer 2008; Buizza et al. 2008). In addition to the initial state perturbations, lateral boundary conditions (LBCs) perturbations are also important for a limited-area EPS (Romine et al. 2014; Frogner et al. 2019). Usually, the limited-area EPS is forced by a coarser-resolution EPS at the boundaries, and the coarser-resolution EPS is, sometimes, also used to generate initial uncertainty. When no coarser-resolution EPS is available, the scaled lagged average forecasting (SLAF) method (Hou et al. 2001) can be used to introduce initial state and LBCs uncertainty. EPSs that account only for initial and LBCs uncertainty are not able to adequately represent the actual forecast uncertainty (Palmer et al. 2005; Jankov et al. 2017; Leutbecher et al. 2017).

Increasing effort has been directed toward the representation of model uncertainty. Uncertainty associated with incomplete knowledge of land surface/soil parameters can be introduced through perturbations of the surface variables with substantial impact on near-surface meteorological conditions (Barthlott and Kalthoff 2011; Bouttier et al. 2016; Frogner et al. 2019). Model errors from parameterization schemes can be taken into account through the use of multi-physics schemes (Charron et al. 2010; Berner et al. 2012, 2015; McTaggart-Cowan et al. 2022) or the commonly used stochastically perturbed parameterization tendencies (SPPT) schemes (Palmer et al. 2009; Bouttier et al. 2012; Leutbecher et al. 2017).

The SPPT scheme, with all its variations (Sanchez et al. 2016; Christensen et al. 2017; Wastl et al. 2019a,b), is perturbing parameterized tendencies of atmospheric variables (horizontal wind components, specific humidity, and temperature). This approach has been very efficient in introducing model uncertainty and is thus exceptionally useful in improving the reliability and skill of an EPS (Palmer et al. 2009; Bouttier et al. 2012; Leutbecher et al. 2017; Wastl et al. 2019a). Yet, as noted by Leutbecher et al. (2017) and Ollinaho et al. (2017), the SPPT scheme does not respect energy and mass conservation laws and it has a smaller impact within the boundary layer, due to tapering functions commonly used to avoid numerical instabilities (Wastl et al. 2019a). In HARMONIE EPS (HarmonEPS), Frogner et al. (2022) have shown that the SPPT scheme has a small impact on the representation of model uncertainty, which is in contrast with the results for the AROME EPS (Bouttier et al. 2012), despite the fact that the two EPSs share the same dynamical core and similar physical parameterization schemes. Overall deficiencies in the SPPT scheme led to the development of the stochastically perturbed parameterizations (SPP) scheme (Ollinaho et al. 2017; Frogner et al. 2022).

Instead of perturbing parameterized tendencies, the SPP scheme is directly perturbing uncertain physical parameters in

the various parameterization schemes to account for model error. This alternative approach can increase the reliability and skill of an EPS (Ollinaho et al. 2017; Wastl et al. 2019a; Jankov et al. 2019; Lang et al. 2021; Frogner et al. 2022; McTaggart-Cowan et al. 2022), bringing it on par with the SPPT schemes or even surpassing them (Lang et al. 2021; Frogner et al. 2022). Parameter distributions for the SPP scheme are usually lognormal, to ensure that perturbed parameters do not receive negative or zero values (Palmer et al. 2009; Leutbecher et al. 2017; Ollinaho et al. 2017).

To match the ensemble scores achieved with the use of SPPT schemes, SPP schemes often need to employ large standard deviations for the underlying distributions from which the perturbed parameters are sampled (Wastl et al. 2019a; Lang et al. 2021). However, the combination of large standard deviation and the lognormal distributions can lead to skewed parameter distributions (Frogner et al. 2022). Moreover, due to nonlinear behavior, parameter values in the tail of the probability distribution functions (PDFs) can lead to extreme, unrealistic response for the perturbed physical processes. This can lead to systematic model biases in the SPP scheme and deteriorate its skill (Jankov et al. 2019; Frogner et al. 2022; McTaggart-Cowan et al. 2022).

To address these biases and improve the performance of the SPP scheme, it is essential to investigate different approaches for the representation of parameter distributions. One way to do this would be to look for ways to increase ensemble spread, without increasing the standard deviation of the underlying parameter distributions. In general, it is preferable that parameter perturbations are independent (Palmer et al. 2009; Wastl et al. 2019a; Leutbecher et al. 2017; Ollinaho et al. 2017) of each other. However, different studies have shown that correlation/anticorrelation between perturbations for parameters acting in a similar/opposite direction can increase ensemble spread and improve probabilistic performance of an ensemble (Jankov et al. 2019; Lang et al. 2021).

Another approach would be to use so-called stretched-uniform distributions, as seen in recent research (McTaggart-Cowan et al. 2022). The former have clearly defined edges, which eliminates the need of clipping. Moreover, in uniform distributions, all parameter values have equal probability of occurrence, which can increase the actual standard deviation of the parameter without the need for long distribution tails. Uniform distributions can be easily centered around the default value, to equally split probabilities between positive and negative perturbations, or even offset the distribution's center to account for nonsymmetry in the impact of the perturbation. This can help alleviate the systematic bias in ensemble members when using lognormal distributions (Frogner et al. 2022), as seen in the recent SPP literature (McTaggart-Cowan et al. 2022).

Finally, increasing the number of parameter perturbations to sample uncertainty from different physical parameterizations can also lead to increased ensemble spread (Jankov et al. 2019; Lang et al. 2021). Perturbations in microphysics (Thompson et al. 2021) and shallow-convection/mass-flux parameters (Fleury et al. 2022) can have a significant impact on the representation of clouds and precipitation, which are

physical processes that convection-permitting ensembles still struggle to improve (Frogner et al. 2019, 2022).

Therefore, to address the issues discussed above, we expanded the functionality of the scheme with the addition of (i) uniform distributions, (ii) correlated perturbations between selected parameters, (iii) four new SPP parameters, and (iv) changes in the standard deviation of lognormal distributions for some SPP parameters. In this manuscript, we compare three different SPP-based EPS suites (two new and one based on the previous SPP iteration) versus an EPS suite with only initial and LBC perturbations. In section 2, we describe the methodology followed and modifications implemented in the new iteration of the SPP scheme. The results are presented in section 3, while discussion and conclusions are provided in sections 4 and 5.

2. Methodology

a. Model description

The experiments were conducted using HarmonEPS (Frogner et al. 2019), which is the ensemble prediction system based on the HARMONIE–AROME limited-area model, at convection-permitting grid spacing (2.5 km) and 65 vertical levels. We use the CY43H.2.2 version of the HARMONIE–AROME code, which is based on the CY40 version described in Bengtsson et al. (2017), but includes updates to various physical parameterizations [e.g., the updates described in de Rooy et al. (2022)]. In HARMONIE–AROME, turbulence and convection are described in the coherent framework of eddy diffusivity mass flux (EDMF) (Siebesma et al. 2007). The eddy diffusive turbulent parameterization scheme HARMONIE with RACMO Turbulence (HARATU) is based on the prognostic TKE scheme of Lenderink and Holtslag (2004) with different length scale formulations for heat and momentum (Bengtsson et al. 2017). Shallow convection is parameterized as in de Rooy et al. (2022). Surface processes are parameterized using the surface externalized (SURFEX) model (Masson et al. 2013), version 8.1, while the microphysical process is parameterized following the ICE3 scheme (Lascaux et al. 2006). For a detailed description of the model dynamics and parameterization schemes of HARMONIE–AROME, we refer the reader to Bengtsson et al. (2017), de Rooy et al. (2022).

HarmonEPS is capable of variational data assimilation (3D-Var and 4D-Var) with 3-h cycling (Brousseau et al. 2011, 2012; Gustafsson et al. 2018). Uncertainty in the initial and lateral conditions can be introduced from a coarser regional or global model (in this case, we used the IFS). This is done by nesting HarmonEPS members into the coarser EPS's members, which provide the different initial and later boundary conditions for the HarmonEPS members. Alternatively, HarmonEPS can use the SLAF method of Hou et al. (2001), when boundary conditions are only available from deterministic IFS forecasts. Observation perturbations using EDA can also be applied to generate initial perturbations for the ensemble members of HarmonEPS (Frogner et al. 2019), while perturbations to surface parameters can also be applied

following Bouttier et al. (2016). In terms of model error representation, HarmonEPS can utilize a multiphysics ensemble (Frogner et al. 2016), a SPPT scheme (Bouttier et al. 2012; Frogner et al. 2019), and a SPP scheme (Frogner et al. 2022). The SPP scheme is discussed in detail in section 2b.

b. The HARMONIE–AROME SPP scheme

The SPP scheme can account for uncertainty arising from errors in model physics by perturbing uncertain parameters in the parameterization schemes of HARMONIE–AROME. The parameters that have been tested and perturbed (Frogner et al. 2022) can be seen in Table 1. From these 11 parameters, the most important in terms of ensemble spread were (i) the proportionality constant in the additional variance term in the cloud scheme (VSIQSAT) [see de Rooy et al. (2022)], (ii) the cloud thickness threshold for shallow/deep convection (CLDDPTHDP), which is shutting down moist convection above the threshold, (iii) the ice nuclei (ICENU), (iv) the top entrainment (RFAC_TWO_COEF), which is a proportionality constant in the top entrainment term, (v) the proportionality constant of the stable condition length scale (RZC_H), and (vi) the asymptotic free atmosphere length scale (RZL_INF) (Frogner et al. 2022; de Rooy et al. 2022).

The method used to create the stochastic perturbations to the SPP parameters can be separated into two parts: (i) the use of the stochastic pattern generator (SPG) (Tsyrlunikov and Gayfulin 2017) and (ii) the particular perturbation amplitude and distribution type applied to the individual parameters.

SPG is based on a linear third-order stochastic model driven by white Gaussian noise, which is used to generate spatiotemporally correlated pseudorandom fields (Tsyrlunikov and Gayfulin 2017). The spatial and temporal perturbations are achieved by using Matérn class correlation functions (Handcock and Stein 1993; Guttorp and Gneiting 2006). This results in a 2D pseudorandom spatially and temporally correlated noise, which satisfies the proportionality of scales (Tsyrlunikov and Gayfulin 2017). The spatiotemporal pattern evolves in time using a spectrally based SPG numerical solver.

Four main parameters control the spatial SPG patterns: (i) the standard deviation (SDEV) of Gaussian noise field, (ii) the spatial length scale λ , (iii) the temporal decorrelation scale τ , and (iv) the proportionality of scale factor (SPGQ). These SPG parameters are the same for all SPP parameters as in Frogner et al. (2022). To ensure that each SPP parameter has a different SPG-generated pattern, a different random seeding per parameter is applied generating a unique SPG field (hereby denoted as Φ_j).

The SPG fields for each SPP parameter are then converted into lognormal fields. If we assume that X_j is the perturbed value of the j th parameter with default parameter value $\widehat{X}_j$, then the equation reads (Ollinaho et al. 2017)

$$X_j = \widehat{X}_j e^{(\mu_j + \Psi_j)}, \quad (1)$$

where Ψ_j is the pseudorandom standard Gaussian noise generated by the SPG (Φ_j) after the multiplication by CMPERT_j (the multiplicative noise factor for the j th parameter applied

TABLE 1. Parameters, descriptions, default values, recommended ranges, and multiplicative factor (CMPERT) used in the SPP scheme. Recommended ranges and CMPERT values are shown as in [Frogner et al. \(2022\)](#), while defaults values are those set in CY43H2.2 of HarmonEPS. CMPERT values have been adjusted for $SDEV = 1$, instead of $SDEV = 3$, as in [Frogner et al. \(2022\)](#). Parameter values for RZC_H and RZL_INF are based on [de Rooy et al. \(2022\)](#). Parameter values for RZALPHA, RZNUC, ZMFDY, and ZMBCLOSURE are newly introduced parameters here discussed in [section 2c\(3\)](#) and were not tested before ([Frogner et al. 2022](#)).

Parameter name	Description	Physics scheme	Default value (units)	Recommended range (units)	CMPERT (cy40.1.1)
VSIGQSAT	Saturation limit	Microphysics	0.02 (—)	0.0–0.6 (—)	1.2
CLDDPTHDP	Cloud thickness threshold for shallow/deep convection	Convection	4000 (—)	1000–8000 (m)	1.2
ICE_CLD_WGT	Cloud ice content impact on cloud thickness	Microphysics	1 (—)	0–2 (—)	1.2
ICENU	Ice nuclei	Microphysics	1 (—)	0.1–100 (—)	2.1
KGN_ACON	Kogan autoconversion speed	Microphysics	10 (—)	1–500 (—)	1.5
KGN_SBGR	Kogan subgrid scale (cloud fraction) sensitivity	Microphysics	0.5 (—)	0.01–1 (—)	0.6
RADGR	Graupel impact on radiation	Radiation	0.5 (—)	0–1 (—)	0.9
RADSN	Snow impact on radiation	Radiation	1 (—)	0–2 (—)	0.9
RFAC_TWO_COEF	Top entrainment efficiency	Turbulence	2 (—)	0.5–3 (—)	1.2
RZC_H	Proportionality constant in the stable length scale formulation	Turbulence	0.11 (—)	0.1–0.2 (—)	1.2
RZL_INF	Asymptotic free atmosphere length scale	Turbulence	40 (m)	20–200 (m)	1.8
RZALPHA	Parameter α of generalized γ distribution used in cloud droplet size distribution	Microphysics	3 (—)	—	—
RZNUC	Parameter ν of generalized γ distribution used in cloud droplet size distribution	Microphysics	3 (—)	—	—
ZMFDY	Dry updraft mass-flux parameter	Convection	1 (—)	0–2 (—)	—
ZMBCLOSURE	Mass-flux closure parameter	Convection	0.35 (—)	0–0.45 (—)	—

to Φ_j) and μ_j is the mean of the Ψ_j distribution, which can be set to either $\mu_j = -0.5(SDEV \text{ CMPERT}_j)^2$ (the mean of the lognormal distribution is equal to $\widehat{X}_j$) or $\mu_j = 0$ (the median of the lognormal distribution is equal to $\widehat{X}_j$). This effectively means that the amplitude of the total standard deviation of the j th parameter is $SDEV \text{ CMPERT}_j$.

c. SPP modifications

1) CORRELATED PERTURBATIONS

The concept of correlating parameter perturbations is based on the physical assumption that certain parameters are acting in a similar direction (e.g., increasing both parameters separately increases the turbulent mixing). For this study, we have identified two SPP parameters (RZC_H and RZL_INF) that modulate the mixing length scales in the boundary layer and free troposphere, mainly under neutral and stable conditions ([de Rooy et al. 2022](#)). From a physical process perspective, correlating their spatial perturbation patterns ensures similar perturbations in turbulent mixing within and above the boundary layer, and consequently, a similar influence on the strength of atmospheric inversions and stratus/stratocumulus dissipation ([de Rooy et al. 2022](#)).

The correlation between the perturbations is achieved by using the pseudorandom spatially and temporally correlated SPG field of Φ_j , instead of Φ_{j+1} for the j th + 1 parameter. This ensures that the resulting perturbations for j th + 1 parameter have the same spatial pattern with the j th parameter but different amplitude due to the different multiplication factor CMPERT_{j+1} .

2) UNIFORM DISTRIBUTIONS

The second modification to the SPP scheme is the addition of uniform distributions for the SPP parameters. The idea of the uniform distribution arises from the physical reasoning that the uncertainty in the reference value of a parameter does not always have a predefined probability distribution function (e.g., a normal or lognormal distribution); hence, all values within the uncertainty range are equally probable. This assumption may be true for highly uncertain physical parameters but may not hold for well-constrained parameters.

Uniform distributions are generated separately for each SPP parameter and are not affecting the generation of the SPG fields. To generate the uniform perturbations, the cumulative distribution function of the pseudorandom SPG field Φ_j of the j th parameter is calculated as

$$\text{CDF}_j = 0.5 \{ \text{erf}[(x - m_j)/(SDEV\sqrt{2})] + 1 \}, \quad (2)$$

where $\text{erf}[(x - m_j)/(SDEV\sqrt{2})]$ is the error function of the Gaussian distribution from the SPG field Φ_j with mean value m_j and SDEV. The perturbations for the j th SPP parameter are generated as

$$X_j = \widehat{X}_j [1 + \text{CMPERT}_j(\text{CDF}_j - \beta)], \quad (3)$$

where $\text{CDF}_j - \beta$ is the uniform SPG noise and β is a parameter used to adjust the center of the uniform distribution, with $\beta = 0.5$ leading to a uniform distribution centered around the default parameter value.

This approach ensures that the probability distribution of the perturbed parameters is close to uniform but retains the spatially and temporally correlated pattern of the SPG pseudorandom noise. For simplicity in the implementation, we make the assumption that m_j is equal to zero.

3) NEW MICROPHYSICS AND MASS-FLUX PARAMETERS

In terms of microphysics, we decided to add the parameters α and ν , which are controlling the shape of the generalized gamma distribution used to compute the cloud droplet size distribution (Contreras Osorio et al. 2022). Previous studies (Miles et al. 2000; Bari et al. 2015; Igel and van den Heever 2017; Wurtz et al. 2021; Thompson et al. 2021) have shown that the values of these parameters are highly uncertain. These parameters have a strong impact on cloud formation and precipitation and can also affect fog (Thompson et al. 2021; Barthlott et al. 2022; Boutle et al. 2022; Contreras Osorio et al. 2022). In HARMONIE-AROME, the parameters α and ν have two different values over the land and sea ($\alpha_{\text{sea}} = 3$, $\alpha_{\text{land}} = 1$; $\nu_{\text{sea}} = 1$, $\nu_{\text{land}} = 3$). Here, we decided to perturb α_{sea} and ν_{land} to display part of the sensitivities highlighted in Contreras Osorio et al. (2022), but in the future, we plan to include perturbations for all four parameters.

Apart from the uncertainty in several parameters of the shallow-convection scheme (de Rooy et al. 2022), there is also a more fundamental problem of using deterministic mass-flux schemes at high resolutions. These schemes assume convective quasi-equilibrium and consequently a one-to-one correspondence between the large-scale (grid point) forcing and the convective response. However, already at the current operational grid resolution of the HARMONIE-AROME model (2.5 km), the mass-flux response is just one realization from a broad probability density function of possible mass fluxes [see Plant and Craig (2008), Fig. 1]. In the future, with steadily increasing model resolution, the necessity to capture the stochastic nature of convection will only increase. One way to represent the variable convective response at high resolutions is the use of an intrinsically stochastic convection scheme [see, e.g., Dorrestijn (2016)]. However, another obvious option to capture the stochastic nature is to modulate the mass-flux response with SPP perturbations. Note that our shallow-convection scheme is based on a dual mass-flux approach (Neggers et al. 2009; de Rooy et al. 2022) in which we distinguish two updrafts: 1) a dry updraft describing all the thermals that do not convert into saturated updrafts in the cloud layer and 2) a moist updraft representing all updrafts that do reach the lifting condensation level (LCL) and continue their ascent in the cloud layer. Therefore, we modulate the mass flux of both updrafts separately. The dry updraft mass flux is perturbed by adding a control parameter (ZMFDRY, default value 1) to modulate the intensity of the dry updraft mass flux. Uncertainty in the moist updraft mass flux is represented by perturbations to the mass-flux closure parameter $c_b = 0.035$ (Brown et al. 2002; de Rooy et al. 2022). This is done by introducing a new perturbed parameter ZMBCLOSURE (reference value 10 c_b), which is then perturbed.

d. Experimental design

To test the impact of the SPP modifications, we conducted a series of experiments for different periods over the KNMI domain (Fig. 1): (i) 20–26 February 2021, (ii) 6–12 July 2020, and (iii) 22–28 March 2019; and over the MetCoOp domain (Fig. 1): (i) 13–26 January 2021 and (ii) 20 May–2 June 2021. These periods encompass a variety of weather conditions (e.g., cold and mild winter, convective weather during summer with thunderstorms) that are usually challenging for limited-area NWP models. Each domain is tested using a different version of HarmonEPS as we attempted to stay as close to the operational EPS versions used at the Royal Netherlands Meteorological Institute (KNMI) and MetCoOp (see differences in the appendix). Apart from the different domain and EPS setups, the MetCoOp domain is larger with 1069×949 grid cells versus 789×789 grid cells in the KNMI domain. Both domains used the same number of vertical levels (65).

Each EPS experiment consists of seven members (0–control, 1–6 perturbed members) running forecasts over the selected periods. Forecast lead times are set to 36 h starting each day at 0000 UTC (the KNMI domain) and 48 h starting at 1200 UTC (MetCoOp domain). In addition, 3-hourly forecasts are run every 3 h to keep data assimilation running continuously during the forecast period. For the KNMI domain, each member does its own assimilation of surface observations, but atmospheric assimilation is restricted to the control member. In the MetCoOp domain, all members do their own 3D-Var as EDA with observation perturbations is active. The control member does not use the SPP scheme. Data output is provided every 3 h. A summary of the HARMONIE-AROME configuration settings and the differences between each EPS version is provided in the appendix.

The experiments are split into two parts. In the first part, we evaluate the impact that the correlated perturbations and uniform distributions have on the performance of the SPP scheme. For this comparison, both KNMI and MetCoOp have selected an SPP setup consisting of five dominant SPP parameters (SPP_5P setup), which was tested using their respective model domains, HarmonEPS setups, and forecast periods. These SPP_5P experiments vary slightly between the domain as seen in Table 2 but are very similar; thus, we will refer to them as SPP_5P. These SPP_5P experiments are compared to reference HarmonEPS experiments (hereby CY43_REF), which do not include SPP perturbations and represent the current operational HarmonEPS setups used by KNMI and MetCoOp (valid until August 2022). Thus, the CY43_REF experiment for the KNMI domain includes perturbations in initial and LBC conditions using the SLAF approach, while over the MetCoOp domain, the CY43_REF experiment includes perturbations in initial and LBC conditions from IFS ensemble members, observation perturbations (EDA), and surface perturbations. Furthermore, to test the reduction in model biases reported in (Frogner et al. 2022), we ran two additional SPP_5P experiments (thereafter named SPP_5P_CY40S) for the two winter forecast periods but using the SPP setting of CY40h1.1 (i.e., old CMPERTs, not correlated



FIG. 1. Model domains used in HARMONIE–AROME for MetCoOp (orange domain; METCOOP25D) and KNMI (red domain; NETHERLANDS) experiments. For simplicity, we refer to them as MetCoOp and KNMI domains.

or uniform distributions) from [Frogner et al. \(2022\)](#) (see [Table A2](#)).

In addition, for the forecast periods over the KNMI domain, an additional SPP experiment (SPP_9P) has been conducted based on SPP_5P but with the addition of the four new SPP parameters [see [section 2c\(3\)](#)]. For a description of the experiments and the SPP settings used, please see [Table 2](#).

e. Evaluation metrics

The evaluation of the EPS experiments is conducted against point observations of 2-m temperature (T_{2m}), 2-m relative humidity (RH_{2m}), 2-m specific humidity (Q_{2m}), 10-m wind speed (S_{10m}), fractional total (CC_{tot}), and low cloud cover (CC_{low}), mean sea level pressure (PMSL,) and accumulated 3-hourly precipitation (AccPcp3h) using the HARP package (<https://github.com/harphub>). Observations are based on synoptic (SYNOP) stations across the modeled domains.

Here, we use the same evaluation metrics available in HARP as used in [Frogner et al. \(2022\)](#): (i) the root-mean-

square error (RMSE) of the ensemble mean, (ii) the bias of the ensemble mean, (iii) the ensemble spread (square root of the unbiased variance estimation of the ensemble members with respect to the ensemble mean), and (iv) the continuous ranked probability score (CRPS, negatively oriented). Testing for the significance in the verification metrics difference between each experiment has been done using bootstrapping methods, where we randomly sample forecast–observation pairs for the calculation of the verification metrics (as done in [Frogner et al. 2022](#)). This random sampling is repeated 20 times to compute the statistical significance of the results.

The ability of the ensemble to predict the occurrence of fog events is evaluated using the Brier score:

$$BS = \frac{1}{n} \sum_i \sum_j (p_{ij} - o_{ij})^2, \quad (4)$$

where n is the number of forecast–observation pairs, i is the verification point, j is the forecast number, and p_{ij} and o_{ij} are the modeled and observed probabilities for the j th forecast at

TABLE 2. SPP-based experiment settings using SPP_5P and SPP_9P for both domains. The format for each entry is name, default value, CMPERT, distribution.

Parameters	Experiment name		
	SPP_5P, MetCoOp	SPP_5P, KNMI	SPP_9P, KNMI
P1	VSIGQSAT, 0.02, 0.6, lognormal (mean)	VSIGQSAT, 0.02, 0.6, lognormal (mean)	VSIGQSAT, 0.02, 0.6, lognormal (mean)
P2	CLDDPTHDP, 4000, 0.6, lognormal (mean)	RFAC_TWOC, 2, 0.6, lognormal (mean)	RFAC_TWOC, 2, 0.6, lognormal (mean)
P3	ICE_CLD_WGT, 1, 1.2, uniform ($\beta = 0.5$)	ICE_CLD_WGT, 1, 1.5, uniform ($\beta = 0.5$)	ICE_CLD_WGT, 1, 1.5, uniform ($\beta = 0.5$)
P4	RZC_H, 0.11, 1.05, uniform ($\beta = 0.475$)	RZC_H, 0.11, 1.35, uniform ($\beta = 0.475$)	RZC_H, 0.11, 1.35, uniform ($\beta = 0.475$)
P5	RZL_INF, 40, 0.45, lognormal (mean)	RZL_INF, 40, 2.25, uniform ($\beta = 0.25$)	RZL_INF, 40, 2.25, uniform ($\beta = 0.25$)
P6	—	—	RZALPHA, 3, 1.5, uniform ($\beta = 0.57$)
P7	—	—	RZNUC, 3, 0.6, lognormal (mean)
P8	—	—	ZMFDRIY, 1, 0.8, lognormal (mean)
P9	—	—	ZMBCLOSURE, 0.35, 1.5, lognormal (mean)
Correlations	RZC_H + RZL_INF	RZC_H + RZL_INF	RZC_H + RZL_INF

location i , respectively. The Brier score is calculated for three threshold values for visibility (250, 500, and 1000 m), intended to separate the ability of the ensemble to capture dense, moderate, and light fog.

To complement the point-based model evaluation, we also conducted spatial evaluation on 3-hourly precipitation and cloud cover with the use of the fractions skill score (FSS) (Roberts and Lean 2008). Previous studies (Dey et al. 2014; Skok and Roberts 2016; Frogner et al. 2022) have shown that a spatial evaluation for precipitation and cloud cover can provide additional information on the spatial skill score and ensemble spread, especially for high-resolution simulations at convection-permitting resolutions. The spatial skill (eFSS) is a measure of the spatial difference between forecast and the observations. Here, we follow Dey et al. (2014) and compute eFSS scores for each member–observation pair and then average them to compute the mean eFSS. The dispersion of the FSS (dFSS) is derived by calculating the FSS between each member pair (e.g., member 0 vs member 1, member 2 vs member 3) and then averaging (Dey et al. 2014). The mean dFSS score gives a measure of the spatial agreement of the variable fields in the different ensemble members and hence is an indication of spatial ensemble spread (Dey et al. 2014). The eFSS score is a positive oriented score (larger is better), while the dFSS score is negatively oriented (lower score means more spatial ensemble spread).

Modeled cloud mask is evaluated against satellite-derived observations from the MSG-SEVIRI following the algorithms of Le Gleau (2019). The grid spacing of the retrieved cloud mask is 3 km at nadir and is regridded to match the model's domain and 2.5-km grid spacing using bilinear resampling from the pyresample python module (<https://pyresample.readthedocs.io/en/latest/>). MSG-SEVIRI retrievals provided cloud masks every 15 min (00, 15, 30, and 45 min); however, to match the modeled output times, we only used the retrieved cloud mask at 0000 UTC every 3 h (0000, 0300 UTC, etc.). Similar to Frogner et al. (2022), we decided to use a

cloud mask threshold value of 0.2, to ensure that grid cells with high-altitude clouds are excluded from the analysis. For the satellite scores, we used spatial scales of 2.5–192.5 km. Note that for the MetCoOp domain, the analysis of cloud cover with the eFSS score is limited to 75°N as the MSG-SEVIRI observations were not available above this latitude. The eFSS and dFSS scores for cloud cover are calculated for all forecast periods over the KNMI domain but only during the 20 May–2 June 2021 period for the MetCoOp domain due to the lack of model data availability for the spatial cloud mask during the winter period.

The evaluation of spatial precipitation skill is done only for the experiments conducted over the KNMI domain and only within the geographical borders of the Netherlands as the radar precipitation product is not available outside these geographical borders. The radar observations are a product of hourly precipitation estimates from radar, which are corrected using the rainfall gauge network of KNMI in the Netherlands. The methodological details of radar precipitation processing to generate the hourly precipitation products are thoroughly described in Overeem et al. (2009b,a, 2011). The hourly precipitation data are accumulated into 3-hourly values (starting from 0000 UTC every day) to match the maximum forecast lead time of the ensemble experiments. For the computation of the eFSS and dFSS scores, we evaluated the modeled 3-hourly accumulated precipitation fields (lead times 3–36 h) against observations for different precipitation thresholds (1st–99th percentiles of the observed precipitation) and spatial scales (2.5–92.5 km).

3. Results

a. Evaluation against point observations

We start with the evaluation of the improvements in the systematic bias of the ensemble members, before proceeding to the comparison of the CY43 SPP-based experiment with CY43_REF.

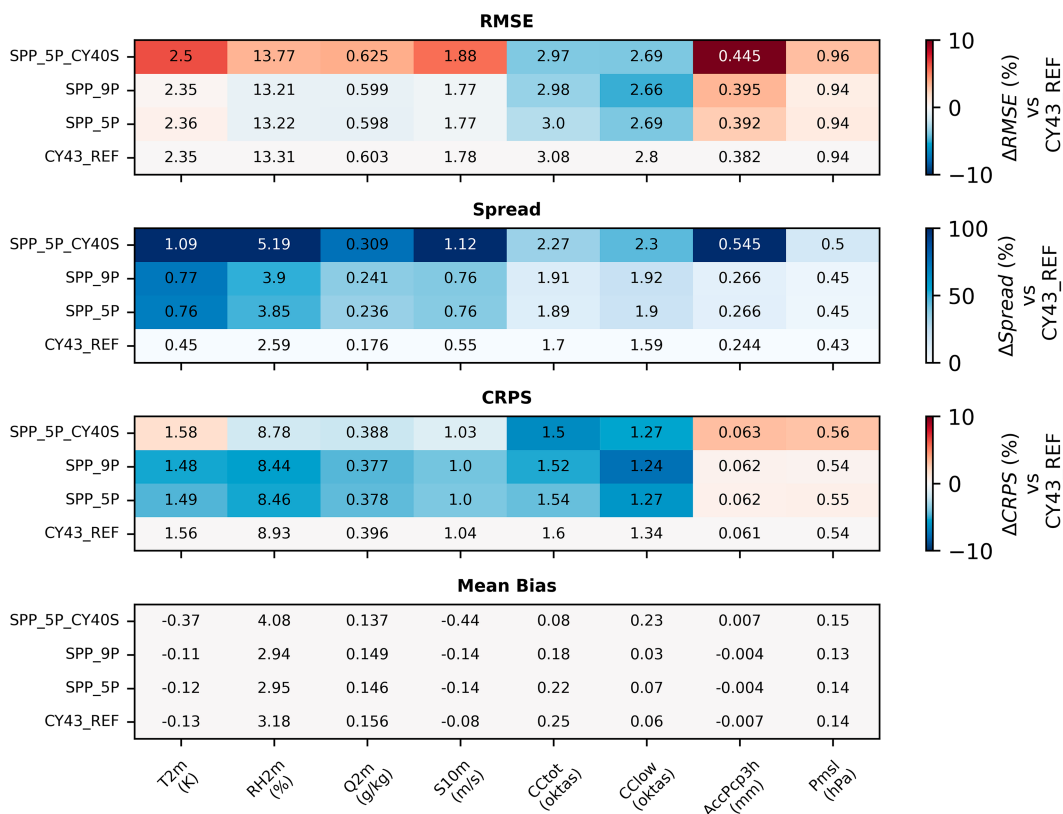


FIG. 2. Comparison of RMSE, spread, CRPS, and mean bias between the CY43_REF, SPP_5P, SPP_9P, and SPP_5P_CY40S for the KNMI domain during the 20–26 Feb 2021 period. Evaluation metrics are calculated against observations from point stations (>500 point stations included for T_{2m}) across the domain and are averaged over all forecast lead times (3–36 h) and forecast initialization days during the aforementioned forecast period. Actual metric values are given in the text. The differences (%) vs the CY43_REF experiment for RMSE, spread, and CRPS are presented using the color scale.

1) REDUCTION IN SYSTEMATIC BIASES

The reduction of systematic model bias is investigated for the winter periods over the KNMI and MetCoOp domains by comparing SPP_5P versus the SPP_5P_CY40S experiments.

The SPP_5P_CY40S experiments show a substantial increase in ensemble spread for the February and January 2021 periods over the CY43_REF experiments (Figs. 2 and 3). However, we do find a substantial deterioration of RMSE and mean ensemble bias in the SPP_5P_CY40S, leading to a substantial deterioration of the CRPS scores for T_{2m} compared to CY43_REF. Substantial changes in ensemble mean bias are seen mainly for all near-surface variables (except Q_{2m}) during the February period. These findings are in line with those from Frogner et al. (2022).

The change in the ensemble mean bias of SPP_5P_CY40S is systematic in nature and is driven by the SPP perturbations in SPP_5P_CY40S settings (Table A2), as the ensemble mean bias is not present in the CY43_REF experiment (Fig. 4). This is quite apparent looking at the bias of the perturbed ensemble members for T_{2m} and S_{10} in SPP_5P_CY40S versus those in CY43_REF (Fig. 4). In the SPP_5P_CY40S, the use of

lognormal distributions with large CMPERT values results in negative perturbations accounting for more than 70% of the total perturbations applied to some of the SPP parameters (see Fig. 5). More specifically for the RZC_H, the imbalance in negative versus positive perturbations results in lower-than-reference RZC_H values over most of the model domain. The disproportionality in the probability of negative and positive perturbations, and the fact that negative perturbations in RZC_H tend to have a larger impact than a similar magnitude positive perturbation, lead to weak vertical mixing during stable conditions, which results in the systematic decrease in T_{2m} and S_{10} in the perturbed members.

Systematic biases due to perturbations can lead to ensemble members with different mean model climate than that of the control member, which is not a desired trait for an EPS (Jankov et al. 2019; McTaggart-Cowan et al. 2022). Moreover, systematically biased perturbed members may limit the reliability of the ensemble.

Comparing the SPP_5P_CY40S experiment with the new SPP experiments (SPP_5P and SPP_9P) reveals that the differences in ensemble mean bias versus CY43_REF experiment decrease substantially. Systematic biases for T_{2m} and S_{10} have been decreased (Figs. 2–4), by reducing CMPERT

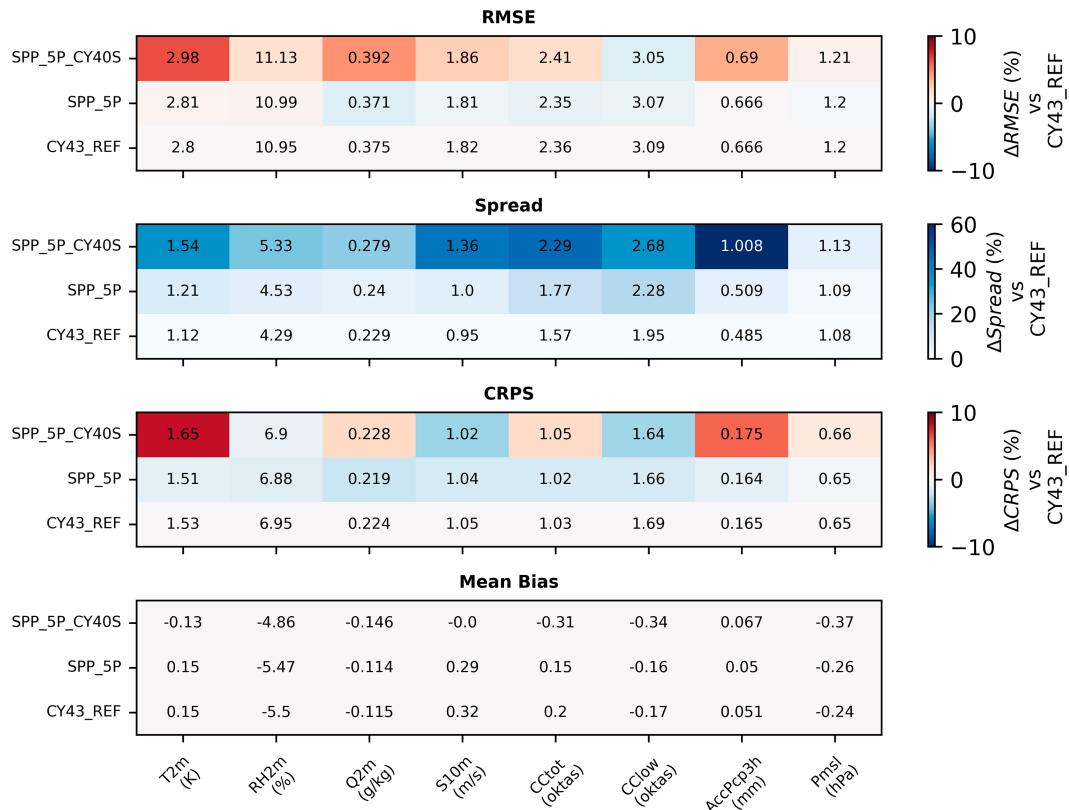


FIG. 3. Comparison of RMSE, spread, CRPS, and mean bias between the CY43_REF, SPP_5P, and SPP_5P_CY40S for the MetCoOp domain during the 13–26 Jan 2021 period. Evaluation metrics are calculated against observations from point stations across the domain and are averaged over all forecast lead times (3–48 h) and forecast initialization days during the aforementioned forecast period. Actual metric values are given in the text. The differences (%) vs the CY43_REF experiment for RMSE, spread, and CRPS are presented using the color scales.

values (VSIGOSAT), by changing to uniform distributions (for ICE_CLD_WGT, RZC_H, and RZL_INF), and by adjusting β for some SPP parameters to account for the nonsymmetric impact of the perturbations (Table 2). These changes result in a substantial reduction in ensemble spread compared to SPP_5P_CY40S but have mainly positive or neutral impact on RMSE and CRPS scores (Figs. 2 and 3). This is also confirmed by comparing the scorecards of SPP_5P_CY40S versus SPP_5P, which shows significant improvements for most of the evaluated variables, except CC_{tot} and C_{base} (Fig. S1 in the online supplemental material).

2) EVALUATION OF SPP SETUPS

The SPP_5P and SPP_9P experiments show improved verification metrics compared to CY43_REF, but these improvements can vary drastically for the different forecast periods.

For the experiments conducted over the KNMI domain, SPP perturbations substantially increase ($>10\%$) the ensemble spread (Figs. 2, 6, and 7) for most variables, except for 3-hourly precipitation and mean sea level pressure (only about 1%–2% increase). The largest increase occurs in the February 2021 forecast period, with ensemble spread for T_{2m} increasing by 70%, while for the rest of the variables, the

increase in ensemble spread ranges between 5% and 50% compared to CY43_REF (Fig. 2). An increase in ensemble spread during the July 2020 and March 2019 periods is smaller (around 5%–15%). The inclusion of four new perturbations in the mass-flux and microphysics parameters (SPP_9P setup, Table 2) results in an additional increase of about 5% in ensemble spread for all variables (except 3-hourly precipitation and PMSL) versus the SPP_5P experiment for the July 2020 forecast period and about 1%–2% increase during the February 2021 and March 2019 periods.

For the experiments in the MetCoOp domain, the increase in ensemble spread in the SPP_5P experiment is smaller (around 5%–10%). The stark difference in the ensemble spread increase between the two domains can be attributed mainly to the use of EDA and surface perturbations in CY43_REF and SPP_5P experiments in the MetCoOp domain. This leads to generally larger ensemble spread compared to the SLAF approach used in the KNMI domain and also highlights the fact that in the presence of an already reasonably dispersive ensemble, SPP perturbations will have a smaller impact. Another reason for the differences could be the lower amplitude of the perturbations for RZC_H used in the MetCoOp domain (Table 2), since this parameter has

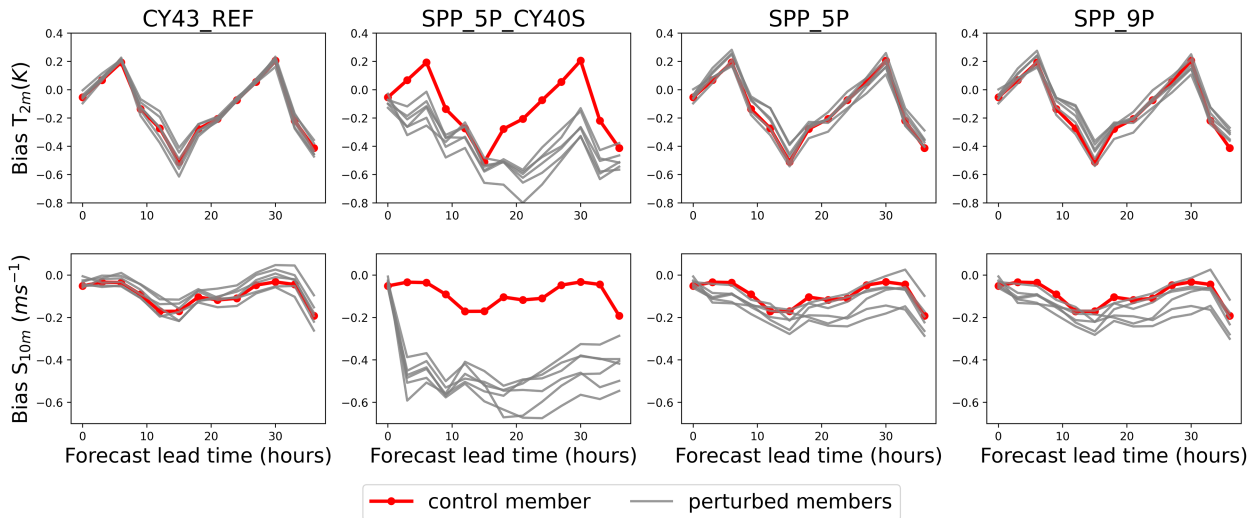


FIG. 4. Comparison of ensemble members bias in (top) T_{2m} and (bottom) S_{10m} between CY43_REF, SPP_5P_CY40S, SPP_5P, and SPP_9P experiments for the control member (red) and the perturbed members (gray). Evaluation metrics are calculated against observations from point stations during the 20–26 Feb 2021 period.

been shown to increase ensemble spread substantially in [Frogner et al. \(2022\)](#).

The difference in ensemble spread between SPP_5P and CY43_REF experiments for near-surface variables (T_{2m} , RH_{2m} , Q_{2m} , and S_{10m}) varies substantially between the different forecast periods over the KNMI domain. These differences can be attributed to the different impact of the SPP perturbations during these periods. For instance, perturbations to

RZC_H show the strongest impact on ensemble spread for near-surface variables during winter ([Frogner et al. 2022](#)), as these perturbations have a much larger impact on the near-surface profiles during stable stratification ([de Rooy et al. 2022](#)). In addition, the perturbations to mass-flux parameters (in SPP_9P) show a stronger impact during the summer period, probably due to the larger impact of the mass-flux perturbations in a summer period when there is substantial convective activity.

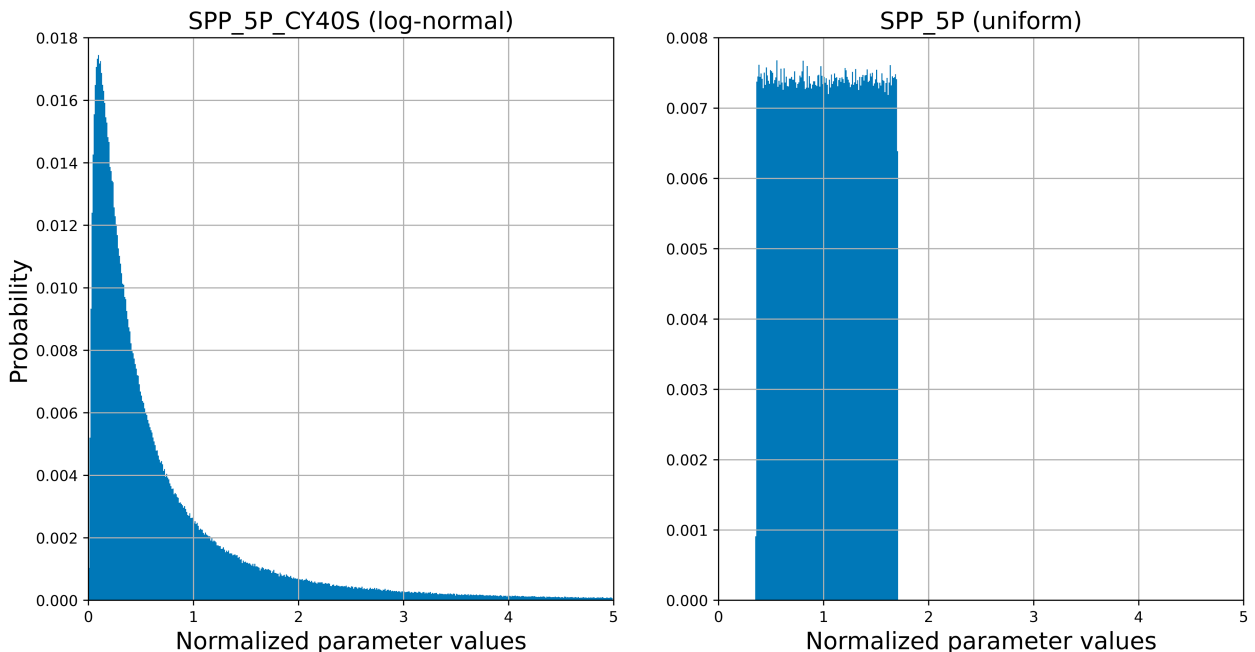


FIG. 5. Comparison of idealized lognormal and uniform distribution of the perturbed values of RZC_H (normalized by the reference value) generated from the SPP settings for the parameter in the SPP_5P_CY40S and SPP_5P experiment setups.

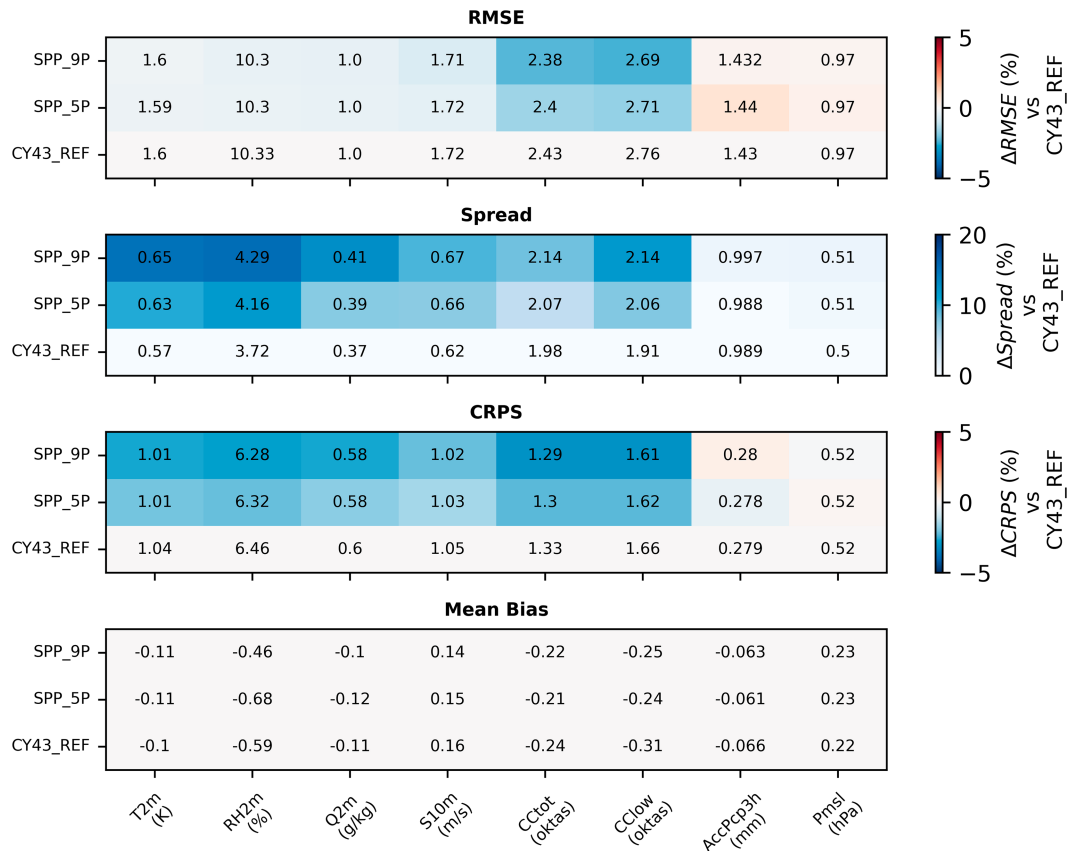


FIG. 6. As in Fig. 2, but for the 6–12 Jul 2020 period and without the SPP_5P_CY40S experiment.

Changes in RMSE between the three setups (CY43_REF, SPP_5P, and SPP_9P) are rather neutral ($<1\%$) for near-surface variables, but RMSE is reduced by around 2%–10% for cloud-related variables depending on the forecast period and domain evaluated (Figs. 2, 3, and 6–8). A small increase in RMSE is visible for 3-hourly precipitation for the winter and summer periods, though not for the spring period. The inclusion of the four new perturbations (SPP_9P) results in small decreases in RMSE (about 1%) compared to the SPP_5P experiment, mainly for CC_{tot} and CC_{low} and mainly for the winter and summer periods. The changes in RMSE and spread lead to an increase in the spread-to-RMSE ratio for the SPP_5P and SPP_9P experiments for all forecast periods. Considering that the spread-to-RMSE ratios for the CY43_REF experiment are substantially smaller than the optimal value of 1, an increase in spread-to-RMSE ratios indicates that the SPP-based ensembles are less underdispersive.

The CRPS is also reduced for most variables when SPP perturbations are active (Figs. 2, 3, and 6–8). Cloud-related variables show the largest reduction in the CRPS (3%–5%) over the KNMI domain, with a smaller reduction over the MetCoOp domain (1%). A substantial reduction is also visible for T_{2m} , RH_{2m} , Q_{2m} , and S_{10m} , but 3-hourly accumulated precipitation and PMSL show mixed results with

slight improvement in some forecast periods (summer period KNMI domain and winter period MetCoOp domain) and deterioration for the rest. The differences in ensemble spread and CRPS are statistically significant for most forecast lead times across all forecast periods (Figs. S2–S4).

During the 20–26 February 2021 period, there was significant fog over the Netherlands and northern Europe. Thus, we decided to also evaluate the ensembles using the Brier score [Eq. (4)] for visibility thresholds of 250, 500, and 1000 m. We found that the SPP perturbations reduce the Brier score for all thresholds (except in the SPP_5P_CY40S experiment where the BS increases), especially for visibility < 250 m, i.e., dense fog (Fig. 9). There is a similar decrease in the Brier score reliability for all thresholds (see Fig. S5), indicating a better match between forecast and observed probabilities for the selected visibility thresholds. On the contrary, in the SPP_5P_CY40S experiment, Brier score reliability increases. This could be due to an increase in forecast probabilities for visibility below the specified thresholds, due to substantially colder 2-m temperatures (as seen in Fig. 4). In Fig. 12 of Frogner et al. (2022), a similar SPP setup as SPP_5P_CY40S experiment resulted in a substantial increase in the forecast probability for fog over the MetCoOp domain. The addition of perturbations to the cloud droplet distribution parameters (RZALPHA and RZNUC) in the SPP_9P setup improves the

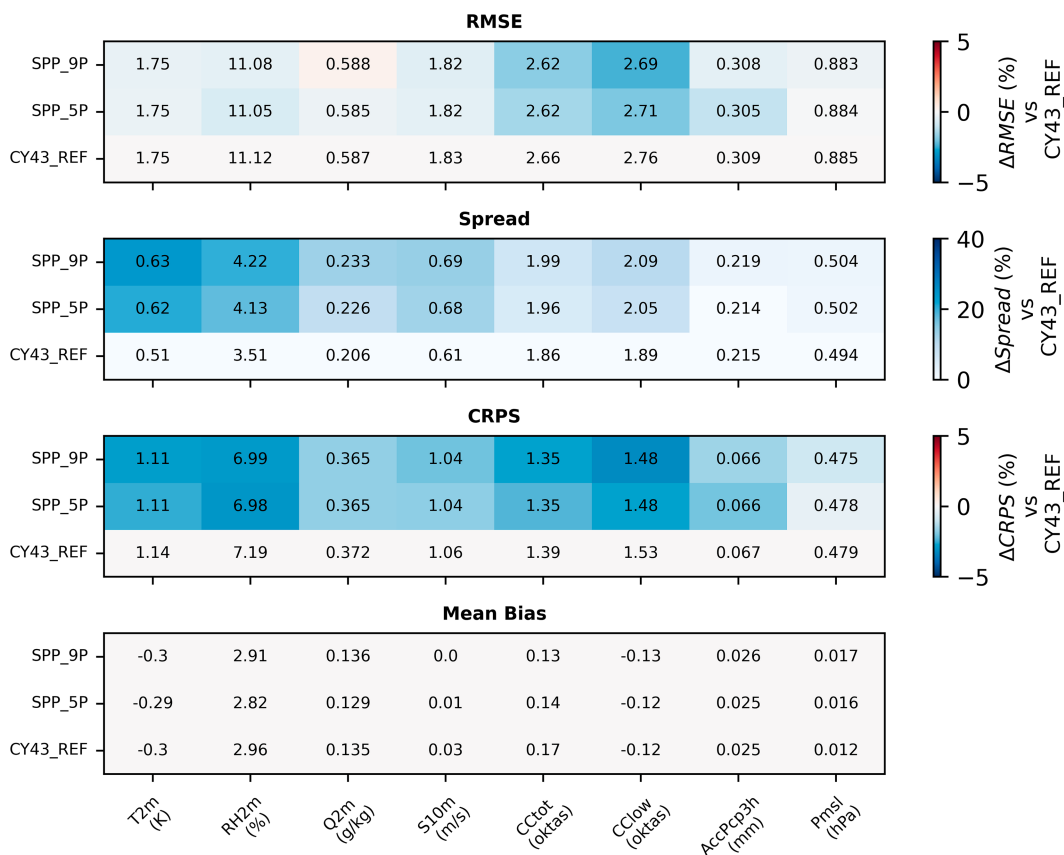


FIG. 7. As in Fig. 2, but for the 22–28 Mar 2019 period and without the SPP_5P_CY40S experiment.

Brier score for visibility compared to the SPP_5P experiment (viz., a reduction in BS reliability). This supports our understanding that perturbations to the microphysics and mass-flux parameters can have a substantial impact on clouds and fog, as seen in previous studies (Thompson et al. 2021; Contreras Osorio et al. 2022; Fleury et al. 2022).

b. Spatial evaluation

Cloud mask and precipitation often have a highly localized nature, which may lead to an insufficient evaluation through SYNOP observations alone. Thus, we decided to apply the FSS, to evaluate the ensemble versus spatial observations (eFSS score) (Roberts and Lean 2008; Frogner et al. 2022) and to get a measure of the spatial spread of the ensemble (dFSS score) (Dey et al. 2014).

1) CLOUD MASK

The mean eFSS and dFSS scores are calculated by averaging individual scores for each member–observation pair, as in Dey et al. (2014) to allow us to directly compare eFSS and dFSS. Similar to Frogner et al. (2022), we compute the 25th and 75th percentiles of eFSS and dFSS scores. Additionally, we used bootstrapping with resampling (1000 times) to obtain the 95th confidence intervals of the mean eFSS and dFSS.

As expected, both eFSS and dFSS increase with increasing spatial scales, as the displacement error for clouds is minimized when averaging over larger spatial scales (Figs. 10 and 11). Over the KNMI domain, eFSS for the CY43_REF, SPP_5P, and SPP_9P experiments is very similar, with the mean eFSS for CY43_REF being slightly higher at larger spatial scales for the February 2021 and March 2019 periods (Fig. 10). Yet, given the large range of the 75th–25th percentiles and the 95% of the confidence intervals for the mean eFSS, these differences are probably not significant. On the contrary, the eFSS for the SPP_5P over the MetCoOp domain deteriorates compared to the CY43_REF setup and across all spatial scales (Fig. 11).

The SPP_5P and SPP_9P experiments exhibit lower dFSS compared to CY43_REF across all spatial scales over the KNMI domain (Fig. 10), which indicates lower agreement in cloud fraction between the ensemble members and hence more spatial ensemble spread (Dey et al. 2014). These differences in mean dFSS scores are substantial, and there is a clear downward shift in the 75th–25th percentiles of the dFSS in the SPP_5P and SPP_9P setups. Over the MetCoOp domain, the differences in dFSS between CY43_REF and SPP_5P increase and the 75th–25th percentile range are not overlapping (Fig. 11), which indicates an even larger improvement in spatial ensemble spread for SPP_5P. The addition of the four

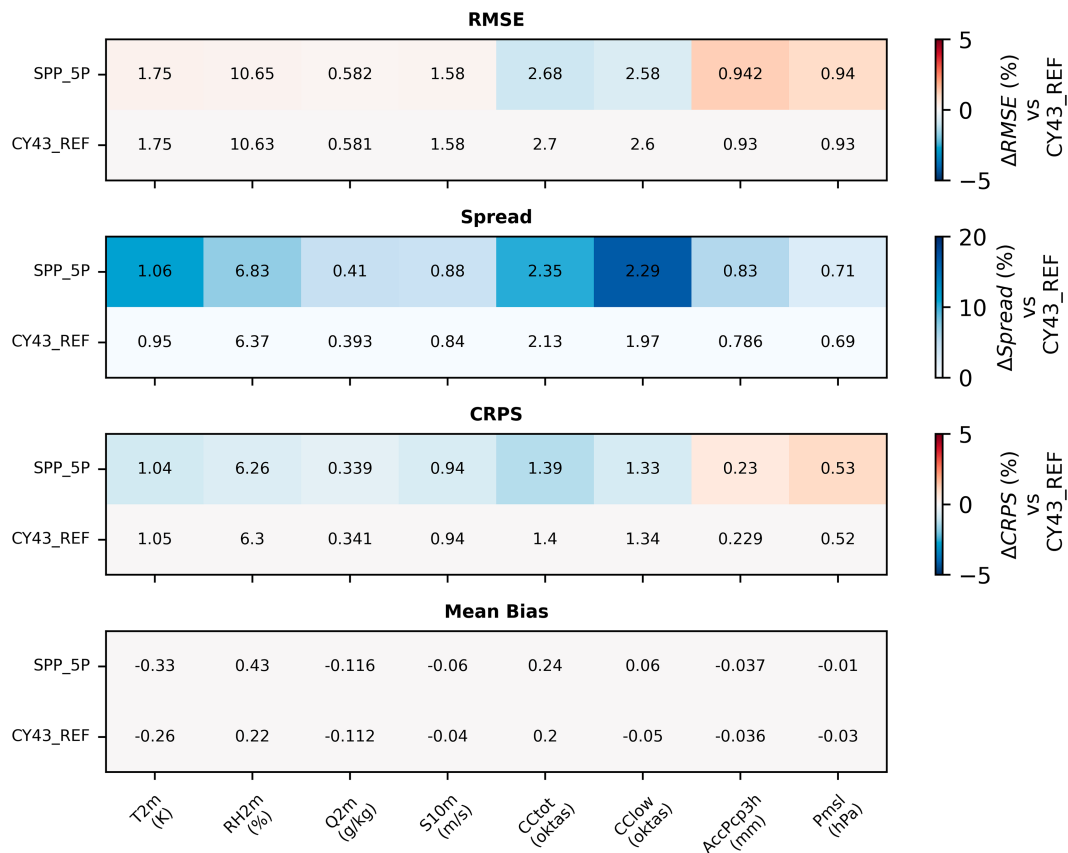


FIG. 8. As in Fig. 3, but for the 20 May–2 Jun 2021 period and without the SPP_5P_CY40S experiment.

new perturbations in SPP_9P leads to lower dFSS compared to SPP_5P (more spread) for the 6–12 July 2020 period, which is in line with our findings for total cloud cover in Fig. 6. This is likely to be caused by the mass-flux perturbations, which are more active during the summer period, leading to additional ensemble spread for cloud fraction. Overall, these results indicate that the new SPP-based experiments should be less underdispersive compared to the CY43_REF experiments, with only a small deterioration in forecast skill.

2) PRECIPITATION

The calculation of mean eFSS and dFSS for 3-hourly precipitation is only done for days (in the periods 20–26 February 2021, 6–12 July 2020, and 22–28 March 2019) and 3-hourly precipitation fields, when the 99th percentile of the 3-hourly observed precipitation over the Netherlands exceeds 1 mm.

The mean eFSS for CY43_REF decreases for larger precipitation percentiles and smaller spatial scales (Fig. 12). The ensemble is generally not able to achieve eFSS > 0.5 (a measure of the usefulness of the ensemble forecast) above the 50th precipitation percentile. This indicates that the ensemble is not able to reasonably predict intense localized precipitation (e.g., convective precipitation). The mean eFSS is improved if we calculate it using 36-hourly accumulated precipitation (not shown), but given the importance of accurate precipitation timing, we

consider that the evaluation of 3-h accumulated precipitation is appropriate. Comparing the SPP_5P and SPP_9P experiments with the CY43_REF reveals that the eFSS slightly decreases (less than 5%) for most percentiles and spatial scales but with a few exceptions. SPP_5P and SPP_9P experiments have similar or slightly better eFSS compared to CY43_REF for the 1st–10th percentiles and also for the 95th–99th percentiles at scales < 25 km (only for SPP_9P). The small improvement in mean eFSS for 95th–99th percentiles in SPP_9P may be related to the perturbations of the mass-flux SPP parameters, which can result in suppressed mass-flux that helps to build up organized convection and intense precipitation.

In terms of spatial ensemble spread, the mean dFSS for the CY43_REF experiment follows a similar pattern as the mean eFSS (Fig. 13), which indicates that the ensemble members differ less at lower precipitation percentiles and larger spatial scales. There is an increase in spatial ensemble spread (lower dFSS) for the SPP_5P and SPP_9P experiments compared to CY43_REF for all spatial scales. The reduction in dFSS reaches up to 10% for the 75th and 90th precipitation percentiles in the SPP_9P experiment. A decrease in dFSS scores between the SPP_5P and SPP_9P experiments indicates that the perturbations in mass flux and microphysical SPP parameters in SPP_9P also increase the spatial precipitation differences between the ensemble members.

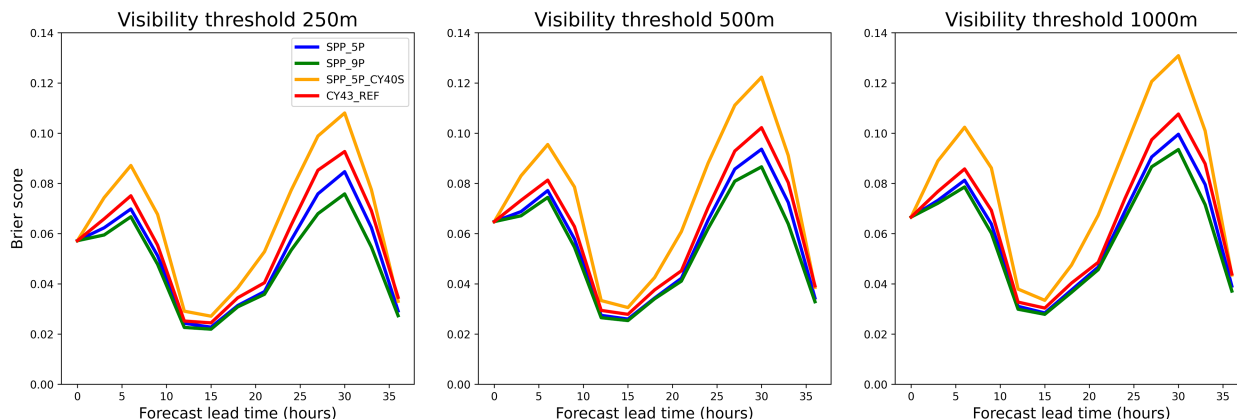


FIG. 9. Brier score for visibility thresholds of 250, 500, and 1000 m for the CY43_REF (red), SPP_5P (blue), SPP_9P (green), and SPP_5P_CY40S (orange) setups for the 20–26 Feb 2021 period. Evaluation is done against point observations across the domain.

Overall, the inclusion of SPP perturbations can decrease dFSS for 3-hourly accumulated precipitation, while only showing a small reduction in the eFSS. We also tested applying different thresholds [2 mm (3 h)^{-1} or 4 mm (3 h)^{-1}] when filtering the 3-hourly precipitation fields for the calculation of mean eFSS and dFSS and found similar results (see Figs. S6–S9).

4. Discussion

The use of SPP schemes has been shown to improve the performance of an EPS for near-surface and cloud-related variables (Wastl et al. 2019b; Lang et al. 2021; Frogner et al. 2022; McTaggart-Cowan et al. 2022). However, the previous version of the SPP scheme of HarmonEPS introduced systematic biases and increased RMSE and CRPS scores for T_{2m} , RH_{2m} , and CC_{low} (Frogner et al. 2022). Changes in ensemble mean bias with the use of a SPP scheme were also seen in other implementations of SPP (Jankov et al. 2019; McTaggart-Cowan et al. 2022). Systematic differences in mean bias between control forecast and perturbed members are not a desirable trait of an ensemble, as it can lead to worse RMSE and CRPS scores and reduce the reliability of the ensemble. We address systematic biases by introducing an option for uniform distributions in our SPP scheme. One could also consider to reduce the standard deviations for the perturbed parameters, while maintaining the original lognormal distributions. This may result in a simpler and easier way to maintain SPP scheme, due to the reduction in testing required to optimize the different distributions and perturbation magnitudes for all SPP parameters. However, this will also limit the flexibility of adjusting individual parameter perturbations and distributions to correct for additional systematic biases that may arise, when default values of SPP parameters change.

Similar to previous research (Jankov et al. 2019; Wastl et al. 2019b; Frogner et al. 2022; Lang et al. 2021; McTaggart-Cowan et al. 2022), our findings show a significant improvement in the CRPS and ensemble spread with the use of the SPP scheme. The exact magnitude of these improvements varies depending on the forecast period and model domain evaluated.

This may depend on the underlying uncertainty in the large-scale synoptic forcing, as reported by studies looking at differences in the impact of perturbations during strong and weak synoptic forcing Hally et al. (2014), Keil et al. (2019), and/or the different impact of the SPP perturbations during these periods. Similar to Frogner et al. (2022), we find larger improvements during the winter period, which may be associated with the substantial impact of the perturbations in the turbulent-mixing parameters (RZC_H and RZC_INF) under stable conditions.

The correlation of the SPG patterns between the SPP parameters was limited to RZC_H and RZC_INF, based on the relation between these parameters in de Rooy et al. (2022). In preliminary tests conducted to investigate the impact of the correlation between RZC_H and RZC_INF, we found that their correlations increased ensemble spread by 2%–5%, without deteriorating RMSE and CRPS. Hence, we decided to correlate their SPG noise in the SPP_5P and SPP_9P experiments. We also considered correlating perturbations between the new mass-flux and microphysical parameters. It is likely that the mass-flux parameters should be anticorrelated to increase their impact at the initiation of deep convection. For instance, a positive perturbation of dry updraft mass-flux transport will increase the convective inhibition (CIN), which hampers the triggering of moist updrafts or resolved convection. On the other hand, a negative perturbation of the dry updraft transport, and therewith the CIN, will increase the moist updraft mass flux and the subgrid parameterized moist convection. The option to correlate or anticorrelate any SPP parameter with another has now been implemented but was not yet available when the experiments were conducted. Additional testing will be needed to find the appropriate correlation pairs between the different SPP parameters. This can be a computationally expensive endeavor and thus may best be done with the use of a single-column (1D version of the HARMONIE-AROME model) to identify the impacts of the correlated perturbations on the relevant physical process. Such an analysis in a 1D model can be done with the use of a “process diagram” as seen in Sterk et al. (2013), Tsiringakis et al. (2020).

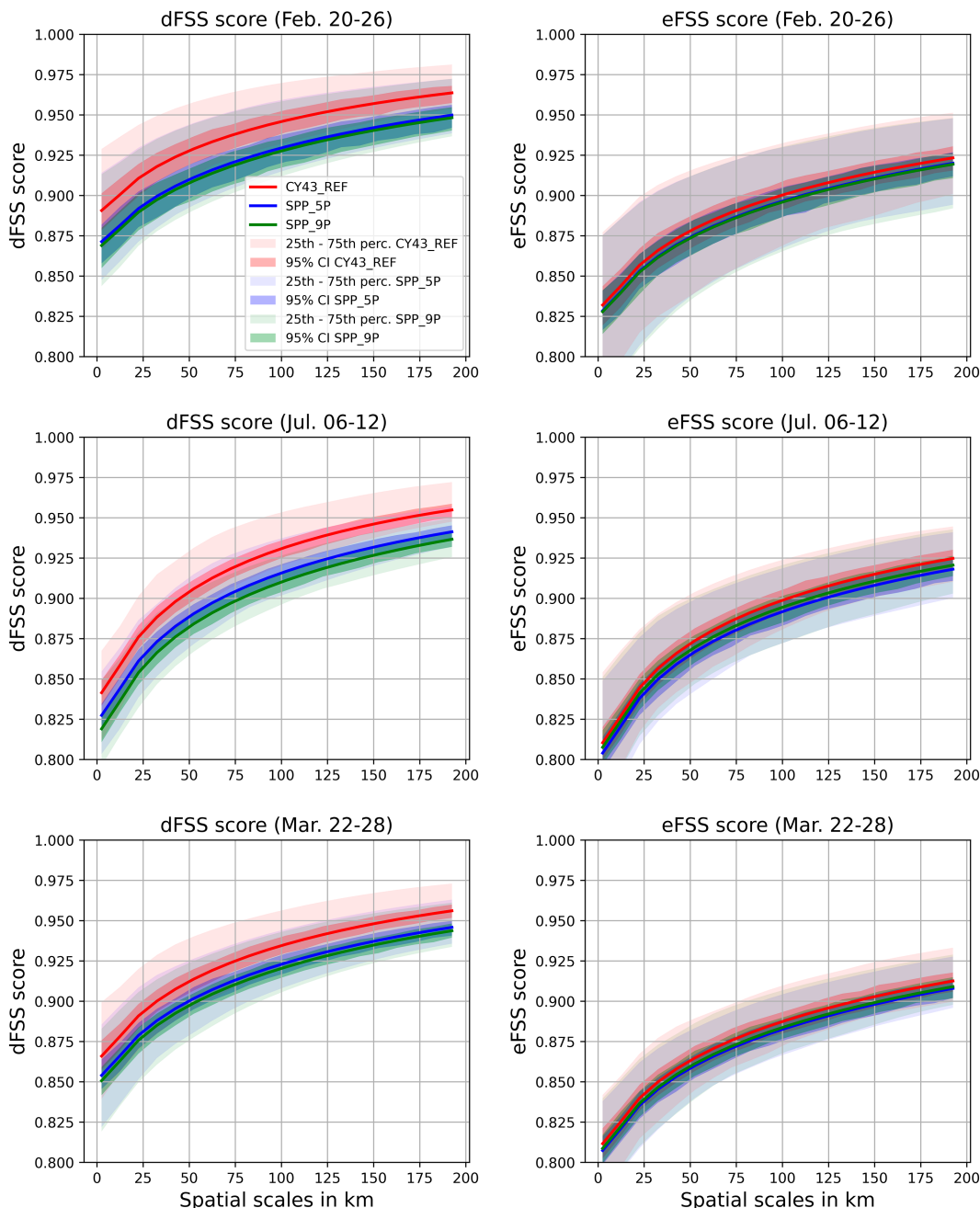


FIG. 10. Evolution of mean (right) eFSS and (left) dFSS for 3-hourly cloud cover over the KNMI domain across different spatial scales for CY43_REF (red), SPP_5P (blue), and SPP_9P (green) in (top) the 20–26 Feb 2021, (middle) 6–12 Jul 2020, and (bottom) 22–28 Mar 2019 periods. The scores are an average of all ensemble members—observation (eFSS) and member–member (dFSS) pairs over all 3-hourly output time steps. The 75th and 25th percentiles from the original sample of eFSS and dFSS (light shadow area) are also plotted. The 95% confidence interval for the mean eFSS and dFSS scores is calculated using bootstrapping (with resampling). Smaller dFSS values indicate more spatial ensemble spread.

Although the evaluation indicates that SPP-based ensembles perform better than non-SPP ensembles, special attention needs to be given to the reliability/sharpness of the ensemble for localized forecasts. In areas where the amplitude

of the perturbations reaches close to the maximum (or minimum) of the distribution, the perturbations for certain parameters may substantially alter the modeled physical processes locally. This can result in situations where ensemble members

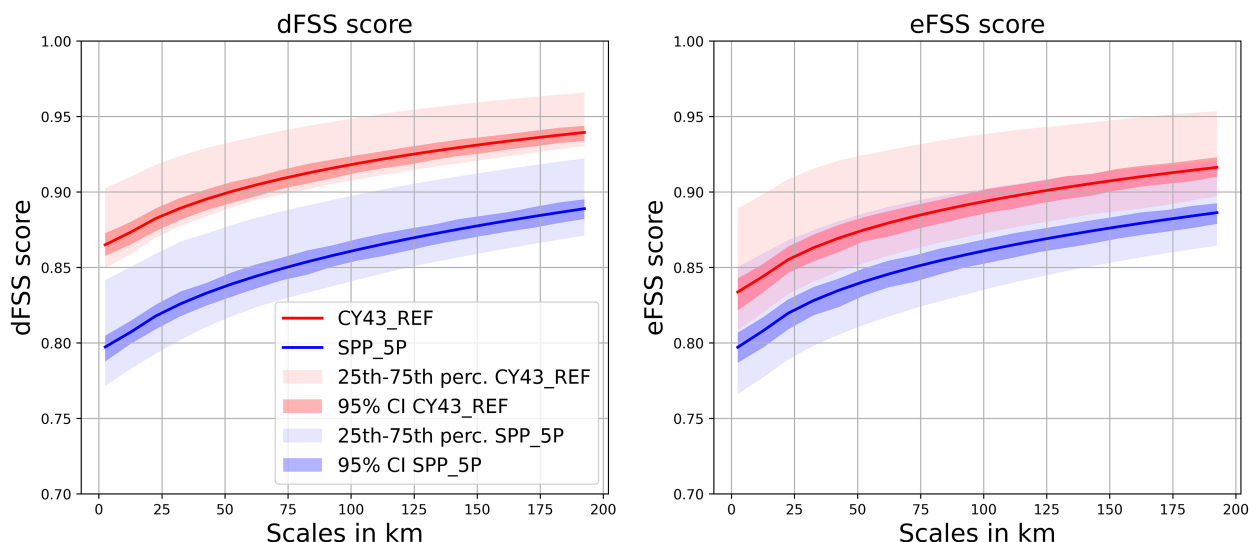


FIG. 11. Evolution of mean (right) eFSS and (left) dFSS for 3-hourly cloud cover over the MetCoOp domain across different spatial scales for CY43_REF (red) and SPP_5P (blue) for the 20 May–2 Jun 2021 period. The scores are an average of all ensemble members–observation (eFSS) and member–member (dFSS) pairs over all 3-hourly output time steps. The 75th and 25th percentiles from the original sample of eFSS and dFSS (light shadow area) are also plotted. The 95% confidence interval for the mean eFSS and dFSS scores is calculated using bootstrapping (with resampling). Smaller dFSS values indicate the higher spatial ensemble spread.

might have locally unrealistically high/low turbulent mixing in the boundary layer, which can erode nocturnal inversions and lead to unrealistic boundary layer heights during nighttime (not shown). This tends to be a local effect (with scales of few grid cells) and can lead to ensemble members being intermittently highly inaccurate when looking at forecasts for individual locations (point or grid cells), even if domain-wide probabilistic metrics are improved. Such issues can be alleviated by investigating closely the impact of parameter perturbations

on processes such as turbulent mixing, surface energy balance, and fluxes and adjusting the perturbation magnitudes and distribution types accordingly.

In section 3a(2), we showed that SPP_5P and SPP_9P experiments had lower ensemble mean RMSE, CRPS, and larger ensemble spread for total and low cloud cover compared to the CY43_REF experiment, when evaluated against in situ observations. Yet, during the spatial evaluation of cloud fraction, we found no improvement in eFSS. This may be due to

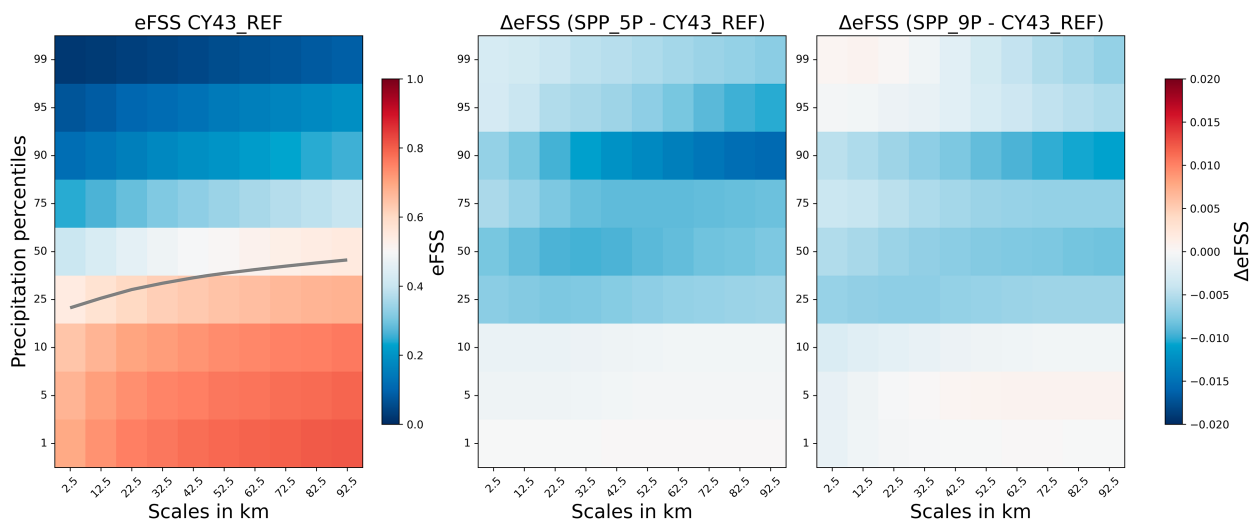


FIG. 12. The mean eFSS from all ensemble members–observations pairs for different 3-hourly precipitation percentiles and spatial scales for (left) CY43_REF and the differences of eFSS scores $\Delta eFSS$ between (middle) SPP_5P–CY43_REF and (right) SPP_9P–CY43_REF over the Netherlands. The scores are averaged over all 3-h accumulated precipitation outputs (3–36-h forecast lead times) and forecast initialization days across the three periods (20–26 Feb 2021, 6–12 Jul 2020, and 22–28 Mar 2019) for which the 99th observed precipitation percentile is larger than 1 mm (3 h)^{−1}. The line shows the contour corresponding to the eFSS = 0.5 threshold.

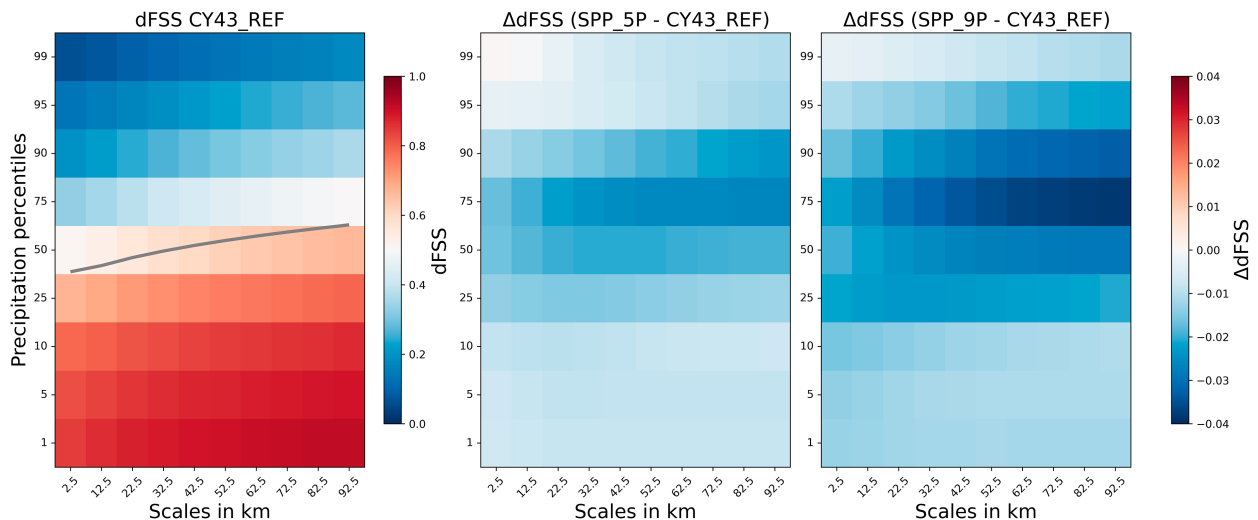


FIG. 13. The mean dFSS for all member–member pairs for different 3-hourly precipitation percentiles and spatial scales for (left) CY43_REF and the differences of eFSS scores Δ dFSS between (middle) SPP_5P–CY43_REF and (right) SPP_9P–CY43_REF over the Netherlands (the radar extent is limited within the geographical boundaries of the country). The scores are averaged over all 3-h accumulated precipitation outputs (3–36-h forecast lead times) and forecast initialization days across the three periods (20–26 Feb 2021, 6–12 Jul 2020, and 22–28 Mar 2019) for which the 99th observed precipitation percentile is larger than 1 mm (3 h)^{-1} . The line shows the contour corresponding to the dFSS = 0.5 threshold. Smaller dFSS values indicate more spatial ensemble spread.

the difference in the metrics used when evaluating against in situ (cloud cover units are in oktas) or satellite observations (cloud fraction is evaluated for threshold values > 0.2). Our findings suggest it may be appropriate to evaluate cloud cover and precipitation with the use of spatially based scores as done here and suggested by the previous literature (Roberts and Lean 2008; Crocker and Mittermaier 2013; Dey et al. 2014).

The updated version of the HarmonEPS SPP scheme offers many optimization options with respect to the temporal and spatial correlation time scales and length scales, and also in relation to the sharpness of the generated noise features (sections 2b and 2c). Currently, these parameters are used according to Frogner et al. (2022), but it would be possible to optimize them further. Optimization of CMPERT values for the rest of the SPP parameters can be done based on ongoing SPP tuning experiments and their results. Currently, an operational SPP-based EPS, following the MetCoOp domain SPP_5P setup, is already used in the operational ensemble prediction system by MetCoOp (MEPS). At a later stage, CMPERT and SPG parameter values could be optimized and calibrated with the use of machine-learning. However, such an endeavor would require a large amount of data, generated via the operational use of the SPP scheme to create a large enough dataset needed for a machine-learning based approach to optimize the SPP scheme for the different meteorological conditions throughout the year.

5. Conclusions

We have investigated the impact of an updated version of the stochastically perturbed parameterizations (SPPs) scheme within HarmonEPS. We have introduced a series of modifications

in the SPP scheme: (i) the option to use uniform distributions for SPP parameters, (ii) the option to correlate perturbations between SPP parameters, (iii) the introduction of four new microphysical and mass-flux parameters, and (iv) small changes in the standard deviation of lognormal distributions for some SPP parameters. Two new SPP-based experimental setups with five (SPP_5P) and nine (SPP_9) SPP parameters are compared to an EPS setup with only initial and LBC perturbations (CY43_REF) and an SPP setup with five SPP parameters (SPP_5P_CY40S) based on the first version of the SPP scheme (Frogner et al. 2022). This comparison was conducted for two different model domains, totaling five different forecast periods, to evaluate the impact of these modifications under different weather conditions.

The use of the updated SPP scheme results in a substantial improvement in probabilistic score metrics (ensemble spread and CRPS) for 2-m temperature, 2-m relative humidity, 2-m specific humidity, 10-m wind speed, total cloud cover, and low cloud cover compared to CY43_REF. Mixed results are shown for 3-hourly precipitation and mean sea level pressure. RMSE for cloud cover improves in the SPP-based experiments, while no deterioration in RMSE occurs for the near-surface variables. This results in increased spread/RMSE ratios, indicating a better calibrated and less underdispersive ensemble.

The updates on the first version of the SPP scheme presented in Frogner et al. (2022) have resulted in the reduction of systematic biases for 2-m temperature, 2-m relative humidity, 10-wind speed, and cloud cover and a decrease in the CRPS. This can be attributed mainly to the use of uniform distributions for selected SPP parameters, which reduce systematic biases, and to correlating perturbations between the two turbulent-

mixing parameters. The addition of the four new parameters results in an increase in ensemble spread and a decrease in the CRPS scores for most of the investigated variables during the summer period and an improvement in the Brier score for visibility during the winter period compared to the SPP_5P experiments.

The spatial verification for cloud fraction and 3-hourly precipitation in the SPP_5P, SPP_9P, and CY43_REF experiments, with the use of eFSS (measure of spatial skill) and dFSS (measure of spatial spread) metrics, reveals not only very small reductions in eFSS with the use of the SPP scheme but also a substantial increase in spatial ensemble spread (lower dFSS). These results generally agree with the finding from the point-based evaluation of the experiments. However, the eFSS for cloud fraction suggests that there is no improvement in ensemble skill when using the SPP scheme, which is contrary to the evaluation using point-based observations. This highlights the need to evaluate ensembles with spatial skill and spread metrics, in addition to the standard evaluation against SYNOP observations.

Overall, the use of the updated SPP scheme leads to an improvement in the ensemble spread versus the non-SPP-based EPS, without sacrificing deterministic scores. The flexibility that the new SPP scheme offers allows us to maintain a substantial improvement over the non-SPP-based EPS, while using narrower and more realistic parameter distributions compared to the previous iteration of the scheme. This will allow us to introduce, expand, and balance the SPP scheme better in the future, especially when there are changes in the model physics in the HARMONIE–AROME model.

As all SPP configurations need to be recalibrated when model physics are updated, the use of the URANIE software platform (Blanchard et al. 2019) is planned in order to minimize the need for long drawn-out manual sensitivity tests. It is also in our plans to investigate the use of SPP during the generation of the EDA analysis, as well as expanding the use of SPP to perturb the surface/soil parameters. As is the case with other EPSs from limited-area models, a move to using single precision will allow resources to be reinvested in more ensemble members and an expansion of our SPP scheme to include a much larger suite of perturbed parameters.

Acknowledgments. We want to acknowledge the financial support from KNMI's multiannual RP3 project. We want to thank ECMWF for providing access to HPC facilities used to conduct the experiments and for the utilization of their operational analysis data. We also want to thank the three anonymous reviewers for their valuable comments and suggestions. This study was conducted as part of the High Resolution Limited Area Model (HIRLAM) consortium (<http://hirlam.org/>; last accessed on 7 July 2023).

Data availability statement. This study uses the HARMONIE–AROME, version CY43H.2.2. The HARMONIE–AROME model system is part of the ALADIN–HIRLAM system. The model code is not open-access. Based on the ALADIN–HIRLAM collaboration agreements, members institutes of the consortia are allowed to share the model code following nonanonymous requests in their home country and only for noncommercial research. Access to the HARMONIE–AROME can be provided after contacting any core-system staff from the member institutes of HIRLAM consortium (information in <https://hirlam.github.io/HarmonieSystemDocumentation/dev/System/GitDeveloperDocumentation>) and signing a standardized ALADIN–HIRLAM licensing agreement. We provide access to modified model code routines and example configuration files used in this study in the following Zenodo repository Tsiringakis et al. (2024). Detailed documentation of various HARMONIE–AROME EPS configurations and settings is available here (<https://hirlam.github.io/HarmonieSystemDocumentation/dev/EPS/>). Radar precipitation data are available in the KNMI data platform (<https://dataplatfom.knmi.nl/dataset/rad-nl21-rac-mfbs-01h-netcdf4-2-0>) under Creative Commons Attribution 4.0 license. Satellite-derived cloud mask from MSG–SEVIRI is available in the EUMETSAT NWC SAF website (<https://www.nwcsaf.org/>) after user registration. Model and observational data used for model initialization (and verification against SYNOP stations) were obtained from the Meteorological Archival and Retrieval System (MARS) of ECMWF. Retrieval of the data is done automatically through the HARMONIE–AROME model code (for information, see <https://hirlam.github.io/HarmonieSystemDocumentation/>). For access information to MARS archive data, we refer to the following link: <https://www.ecmwf.int/en/forecasts/accessing-forecasts>. The verification package HARP is freely available in the following repository: <https://github.com/harphub>. Example scripts used for ensemble verification, postprocessing, and plotting, along with a small subset of model output and postprocessed observational data, are provided in the same Zenodo repository Tsiringakis et al. (2024). The remaining model output data (i.e., those not included in the repository) are available upon request to the author.

APPENDIX

Differences in HarmonEPS Suites and Experiment Settings between KNMI and MetCoOp

Table A1 shows differences in the experimental HarmonEPS suites between KNMI and MetCoOp. Table A2 shows differences in SPP settings for SPP_5P_CY40S experiments between KNMI and MetCoOp.

TABLE A1. General settings for the HarmonEPS experimental setups for the different periods and domains.

Test periods	13–26 Jan 2021, 20 May–2 Jun 2021	20–26 Feb 2021, 6–12 Jul 2020, 22–28 Mar 2019
Physics suite	HARMONIE–AROME (CY-43h.2.2)	HARMONIE–AROME (CY-43h.2.2)
Domain name	METCOOP25D	NETHERLANDS
Ensemble members	7 (6 + control)	7 (6 + control)
Forecast time	48 h starting at 1200 UTC each day, 3-h forecasts at (0300, 0600 UTC, ..., etc.)	36 h starting at 0000 UTC each day, 3-h forecasts at (0300, 0600 UTC, ..., etc.)
Data assimilation	3D-Var (all members) CANARI surface assimilation (all members)	3D-Var (control only) CANARI surface assimilation (all members)
Initial conditions and LBCs	ENS 3 hourly (source: ECMWF EPS members)	SLAF 1-hourly (source: ECMWF deterministic forecast)
Initial and LBCs perturbations	ENS initial conditions + EDA (observation perturbations)	SLAF (source: IFS deterministic forecast)
Surface perturbations	Yes	No
SPG settings	$\lambda = 200$ km, $\tau = 12$ h, SPGQ = 0.5	$\lambda = 200$ km, $\tau = 12$ h, SPGQ = 0.5

TABLE A2. SPP-based experiment settings for SPP_5P_CY40S for both domains. The format for each entry is name, default value, CMPERT, distribution.

Parameters	Experiment name	
	SPP_5P_CY40S, MetCoOp	SPP_5P_CY40S, KNMI
P1	VSIGQSAT, 0.02, 1.2, lognormal (mean)	VSIGQSAT, 0.02, 1.2, lognormal (mean)
P2	CLDDPTHDP, 4000, 1.2, lognormal (mean)	RFAC_TWOC, 2, 1.2, lognormal (mean)
P3	ICE_CLD_WGT, 1, 1.2, lognormal (mean)	ICE_CLD_WGT, 1, 1.2, lognormal (mean)
P4	RZC_H, 0.11, 1.2, lognormal (mean)	RZC_H, 0.11, 1.2, lognormal (mean)
P5	RZL_INF, 40, 1.8, lognormal	RZL_INF, 40, 1.8, lognormal (mean)

REFERENCES

- Bari, D., T. Bergot, and M. El Khilfi, 2015: Numerical study of a coastal fog event over Casablanca, Morocco. *Quart. J. Roy. Meteor. Soc.*, **141**, 1894–1905, <https://doi.org/10.1002/qj.2494>.
- Barkmeijer, J., F. Bouttier, and M. Van Gijzen, 1998: Singular vectors and estimates of the analysis-error covariance metric. *Quart. J. Roy. Meteor. Soc.*, **124**, 1695–1713, <https://doi.org/10.1002/qj.49712454916>.
- , R. Buizza, and T. N. Palmer, 1999: 3D-Var hessian singular vectors and their potential use in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2333–2351, <https://doi.org/10.1002/qj.49712555818>.
- Barthlott, C., and N. Kalthoff, 2011: A numerical sensitivity study on the impact of soil moisture on convection-related parameters and convective precipitation over complex terrain. *J. Atmos. Sci.*, **68**, 2971–2987, <https://doi.org/10.1175/JAS-D-11-027.1>.
- , A. Zarbo, T. Matsunobu, and C. Keil, 2022: Importance of aerosols and shape of the cloud droplet size distribution for convective clouds and precipitation. *Atmos. Chem. Phys.*, **22**, 2153–2172, <https://doi.org/10.5194/acp-22-2153-2022>.
- Bengtsson, L., and Coauthors, 2017: The HARMONIE-AROME model configuration in the ALADIN-HIRLAM NWP system. *Mon. Wea. Rev.*, **145**, 1919–1935, <https://doi.org/10.1175/MWR-D-16-0417.1>.
- Berner, J., T. Jung, and T. N. Palmer, 2012: Systematic model error: The impact of increased horizontal resolution versus improved stochastic and deterministic parameterizations. *J. Climate*, **25**, 4946–4962, <https://doi.org/10.1175/JCLI-D-11-00297.1>.
- , K. R. Fossell, S.-Y. Ha, J. P. Hacker, and C. Snyder, 2015: Increasing the skill of probabilistic forecasts: Understanding performance improvements from model-error representations. *Mon. Wea. Rev.*, **143**, 1295–1320, <https://doi.org/10.1175/MWR-D-14-00091.1>.
- Blanchard, J., G. Damblin, J. Martinez, G. Arnaud, and F. Gaudier, 2019: The Uranie platform: An open-source software for optimisation, meta-modelling and uncertainty analysis. *EPJ Nucl. Sci. Technol.*, **5**, 4, <https://doi.org/10.1051/epjn/2018050>.
- Boutle, I., and Coauthors, 2022: Demistify: A large-eddy simulation (LES) and single-column model (SCM) intercomparison of radiation fog. *Atmos. Chem. Phys.*, **22**, 319–333, <https://doi.org/10.5194/acp-22-319-2022>.
- Bouttier, F., B. Vie, O. Nuissier, and L. Raynaud, 2012: Impact of stochastic physics in a convection-permitting ensemble. *Mon. Wea. Rev.*, **140**, 3706–3721, <https://doi.org/10.1175/MWR-D-12-00031.1>.
- , L. Raynaud, O. Nuissier, and B. Ménétrier, 2016: Sensitivity of the AROME ensemble to initial and surface perturbations during HyMeX. *Quart. J. Roy. Meteor. Soc.*, **142**, 390–403, <https://doi.org/10.1002/qj.2622>.
- Brousseau, P., L. Berre, F. Bouttier, and G. Desroziers, 2011: Background-error covariances for a convective-scale data-assimilation system: AROME–France 3D-Var. *Quart. J. Roy. Meteor. Soc.*, **137**, 409–422, <https://doi.org/10.1002/qj.750>.
- , —, —, and —, 2012: Flow-dependent background-error covariances for a convective-scale data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **138**, 310–322, <https://doi.org/10.1002/qj.920>.
- Brown, A. R., and Coauthors, 2002: Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land. *Quart.*

- J. Roy. Meteor. Soc.*, **128**, 1075–1093, <https://doi.org/10.1256/003590002320373210>.
- Buizza, R., and T. N. Palmer, 1995: The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, **52**, 1434–1456, [https://doi.org/10.1175/1520-0469\(1995\)052<1434:TSVSOT>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<1434:TSVSOT>2.0.CO;2).
- , M. Leutbecher, and L. Isaksen, 2008: Potential use of an ensemble of analyses in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **134**, 2051–2066, <https://doi.org/10.1002/qj.346>.
- Charron, M., G. Pellerin, L. Spacek, P. L. Houtekamer, N. Gagnon, H. L. Mitchell, and L. Michelin, 2010: Toward random sampling of model error in the Canadian ensemble prediction system. *Mon. Wea. Rev.*, **138**, 1877–1901, <https://doi.org/10.1175/2009MWR3187.1>.
- Christensen, H. M., S.-J. Lock, I. M. Moroz, and T. N. Palmer, 2017: Introducing independent patterns into the Stochastically Perturbed Parametrization Tendencies (SPPT) scheme. *Quart. J. Roy. Meteor. Soc.*, **143**, 2168–2181, <https://doi.org/10.1002/qj.3075>.
- Contreras Osorio, S., D. Martín Pérez, K.-I. Ivarsson, K. P. Nielsen, W. C. de Rooy, E. Gleeson, and E. McAufield, 2022: Impact of the microphysics in HARMONIE-AROME on fog. *Atmosphere*, **13**, 2127, <https://doi.org/10.3390/atmos13122127>.
- Crocker, R., and M. Mittermaier, 2013: Exploratory use of a satellite cloud mask to verify NWP models. *Meteor. Appl.*, **20**, 197–205, <https://doi.org/10.1002/met.1384>.
- de Rooy, W. C., and Coauthors, 2022: Model development in practice: A comprehensive update to the boundary layer schemes in HARMONIE-AROME cycle 40. *Geosci. Model Dev.*, **15**, 1513–1543, <https://doi.org/10.5194/gmd-15-1513-2022>.
- Dey, S. R. A., G. Leoncini, N. M. Roberts, R. S. Plant, and S. Migliorini, 2014: A spatial view of ensemble spread in convection permitting ensembles. *Mon. Wea. Rev.*, **142**, 4091–4107, <https://doi.org/10.1175/MWR-D-14-00172.1>.
- Dorrestijn, J., 2016: Stochastic convection parameterization. Ph.D. thesis, Delft University of Technology, 149 pp., <https://doi.org/10.4233/uuid:d80246c5-41dc-451d-9beb-c293c445a8f3>.
- Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, <https://doi.org/10.1007/s10236-003-0036-9>.
- Fleury, A., F. Bouttier, and F. Couvreur, 2022: Process-oriented stochastic perturbations applied to the parametrization of turbulence and shallow convection for ensemble prediction. *Quart. J. Roy. Meteor. Soc.*, **148**, 981–1000, <https://doi.org/10.1002/qj.4242>.
- Frogner, I.-L., T. Nipen, A. Singleton, J. B. Bremnes, and O. Vignes, 2016: Ensemble prediction with different spatial resolutions for the 2014 Sochi Winter Olympic Games: The effects of calibration and multimodel approaches. *Wea. Forecasting*, **31**, 1833–1851, <https://doi.org/10.1175/WAF-D-16-0048.1>.
- , and Coauthors, 2019: HarmonEPS—The HARMONIE ensemble prediction system. *Wea. Forecasting*, **34**, 1909–1937, <https://doi.org/10.1175/WAF-D-19-0030.1>.
- , U. Andrae, P. Ollinaho, A. Hally, K. Hamalainen, J. Kauhanen, K.-I. Ivarsson, and D. Yazgi, 2022: Model uncertainty representation in a convection-permitting ensemble—SPP and SPPT in HarmonEPS. *Mon. Wea. Rev.*, **150**, 775–795, <https://doi.org/10.1175/MWR-D-21-0099.1>.
- Gustafsson, N., and Coauthors, 2018: Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Quart. J. Roy. Meteor. Soc.*, **144**, 1218–1256, <https://doi.org/10.1002/qj.3179>.
- Guttorp, P., and T. Gneiting, 2006: Studies in the history of probability and statistics XLIX on the Matérn correlation family. *Biometrika*, **93**, 989–995, <https://doi.org/10.1093/biomet/93.4.989>.
- Hally, A., E. Richard, and V. Ducrocq, 2014: An ensemble study of HyMeX IOP6 and IOP7a: Sensitivity to physical and initial and boundary condition uncertainties. *Nat. Hazards Earth Syst. Sci.*, **14**, 1071–1084, <https://doi.org/10.5194/nhess-14-1071-2014>.
- Handcock, M. S., and M. L. Stein, 1993: A Bayesian analysis of kriging. *Technometrics*, **35**, 403–410, <https://doi.org/10.1080/00401706.1993.10485354>.
- Hou, D., E. Kalnay, and K. K. Droegemeier, 2001: Objective verification of the SAMEX '98 ensemble forecasts. *Mon. Wea. Rev.*, **129**, 73–91, [https://doi.org/10.1175/1520-0493\(2001\)129<0073:OVOTSE>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0073:OVOTSE>2.0.CO;2).
- Igel, A. L., and S. C. van den Heever, 2017: The importance of the shape of cloud droplet size distributions in shallow cumulus clouds. Part I: Bin microphysics simulations. *J. Atmos. Sci.*, **74**, 249–258, <https://doi.org/10.1175/JAS-D-15-0382.1>.
- Jankov, I., and Coauthors, 2017: A performance comparison between multiphysics and stochastic approaches within a North American RAP ensemble. *Mon. Wea. Rev.*, **145**, 1161–1179, <https://doi.org/10.1175/MWR-D-16-0160.1>.
- , J. Beck, J. Wolff, M. Harrold, J. B. Olson, T. Smirnova, C. Alexander, and J. Berner, 2019: Stochastically perturbed parameterizations in an HRRR-based ensemble. *Mon. Wea. Rev.*, **147**, 153–173, <https://doi.org/10.1175/MWR-D-18-0092.1>.
- Keil, C., F. Baur, K. Bachmann, S. Rasp, L. Schneider, and C. Barthlott, 2019: Relative contribution of soil moisture, boundary-layer and microphysical perturbations on convective predictability in different weather regimes. *Quart. J. Roy. Meteor. Soc.*, **145**, 3102–3115, <https://doi.org/10.1002/qj.3607>.
- Lang, S. T. K., S.-J. Lock, M. Leutbecher, P. Bechtold, and R. M. Forbes, 2021: Revision of the stochastically perturbed parameterisations model uncertainty scheme in the integrated forecasting system. *Quart. J. Roy. Meteor. Soc.*, **147**, 1364–1381, <https://doi.org/10.1002/qj.3978>.
- Lascaux, F., E. Richard, and J.-P. Pinty, 2006: Numerical simulations of three different map IOPs and the associated microphysical processes. *Quart. J. Roy. Meteor. Soc.*, **132**, 1907–1926, <https://doi.org/10.1256/qj.05.197>.
- Le Gleau, H., 2019: Algorithm theoretical basis document for the cloud product processors of the NWC/GEO. NWC SAF Tech. Rep., 125 pp., https://www.nwcsaf.org/Downloads/GEO/2018/Documents/Scientific_Docs/NWC-CDOP2-GEO-MFL-SCI-ATBD-Cloud_v2.1.pdf.
- Lenderink, G., and A. A. M. Holtslag, 2004: An updated length-scale formulation for turbulent mixing in clear and cloudy boundary layers. *Quart. J. Roy. Meteor. Soc.*, **130**, 3405–3427, <https://doi.org/10.1256/qj.03.117>.
- Leutbecher, M., and T. N. Palmer, 2008: Ensemble forecasting. *J. Comput. Phys.*, **227**, 3515–3539, <https://doi.org/10.1016/j.jcp.2007.02.014>.
- , and S. T. K. Lang, 2014: On the reliability of ensemble variance in subspaces defined by singular vectors. *Quart. J. Roy. Meteor. Soc.*, **140**, 1453–1466, <https://doi.org/10.1002/qj.2229>.
- , and Coauthors, 2017: Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Quart. J. Roy. Meteor. Soc.*, **143**, 2315–2339, <https://doi.org/10.1002/qj.3094>.
- Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141, [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2).

- Masson, V., and Coauthors, 2013: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes. *Geosci. Model Dev.*, **6**, 929–960, <https://doi.org/10.5194/gmd-6-929-2013>.
- McTaggart-Cowan, R., and Coauthors, 2022: Using stochastically perturbed parameterizations to represent model uncertainty. Part I: Implementation and parameter sensitivity. *Mon. Wea. Rev.*, **150**, 2829–2858, <https://doi.org/10.1175/MWR-D-21-0315.1>.
- Miles, N. L., J. Verlinde, and E. E. Clothiaux, 2000: Cloud droplet size distributions in low-level stratiform clouds. *J. Atmos. Sci.*, **57**, 295–311, [https://doi.org/10.1175/1520-0469\(2000\)057<0295:CDSIL>2.0.CO;2](https://doi.org/10.1175/1520-0469(2000)057<0295:CDSIL>2.0.CO;2).
- Neggers, R. A. J., M. Köhler, and A. C. M. Beljaars, 2009: A dual mass flux framework for boundary layer convection. Part I: Transport. *J. Atmos. Sci.*, **66**, 1465–1487, <https://doi.org/10.1175/2008JAS2635.1>.
- Ollinaho, P., and Coauthors, 2017: Towards process-level representation of model uncertainties: Stochastically perturbed parametrizations in the ECMWF ensemble. *Quart. J. Roy. Meteor. Soc.*, **143**, 408–422, <https://doi.org/10.1002/qj.2931>.
- Overeem, A., T. A. Buishand, and I. Holleman, 2009a: Extreme rainfall analysis and estimation of depth-duration-frequency curves using weather radar. *Water Resour. Res.*, **45**, W10424, <https://doi.org/10.1029/2009WR007869>.
- , I. Holleman, and A. Buishand, 2009b: Derivation of a 10-year radar-based climatology of rainfall. *J. Appl. Meteor. Climatol.*, **48**, 1448–1463, <https://doi.org/10.1175/2009JAMC1954.1>.
- , H. Leijnse, and R. Uijlenhoet, 2011: Measuring urban rainfall using microwave links from commercial cellular communication networks. *Water Resour. Res.*, **47**, W12505, <https://doi.org/10.1029/2010WR010350>.
- Palmer, T. N., G. J. Shutts, R. Hagedorn, F. J. Doblas-Reyes, T. Jung, and M. Leutbecher, 2005: Representing model uncertainty in weather and climate prediction. *Annu. Rev. Earth Planet. Sci.*, **33**, 163–193, <https://doi.org/10.1146/annurev.earth.33.092203.122552>.
- , R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. ECMWF Tech. Memo., 598, 44 pp., <https://doi.org/10.21957/ps8gbwbdv>.
- Plant, R. S., and G. C. Craig, 2008: A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci.*, **65**, 87–105, <https://doi.org/10.1175/2007JAS2263.1>.
- Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, <https://doi.org/10.1175/2007MWR2123.1>.
- Romine, G. S., C. S. Schwartz, J. Berner, K. R. Fossell, C. Snyder, J. L. Anderson, and M. L. Weisman, 2014: Representing forecast error in a convection-permitting ensemble system. *Mon. Wea. Rev.*, **142**, 4519–4541, <https://doi.org/10.1175/MWR-D-14-00100.1>.
- Sanchez, C., K. D. Williams, and M. Collins, 2016: Improved stochastic physics schemes for global weather and climate models. *Quart. J. Roy. Meteor. Soc.*, **142**, 147–159, <https://doi.org/10.1002/qj.2640>.
- Siebesma, A. P., P. M. M. Soares, and J. Teixeira, 2007: A combined eddy-diffusivity mass-flux approach for the convective boundary layer. *J. Atmos. Sci.*, **64**, 1230–1248, <https://doi.org/10.1175/JAS3888.1>.
- Skok, G., and N. Roberts, 2016: Analysis of fractions skill score properties for random precipitation fields and ECMWF forecasts. *Quart. J. Roy. Meteor. Soc.*, **142**, 2599–2610, <https://doi.org/10.1002/qj.2849>.
- Sterk, H. A. M., G. J. Steeneveld, and A. A. M. Holtslag, 2013: The role of snow-surface coupling, radiation, and turbulent mixing in modeling a stable boundary layer over arctic sea ice. *J. Geophys. Res. Atmos.*, **118**, 1199–1217, <https://doi.org/10.1002/jgrd.50158>.
- Thompson, G., J. Berner, M. Frediani, J. A. Otkin, and S. M. Griffin, 2021: A stochastic parameter perturbation method to represent uncertainty in a microphysics scheme. *Mon. Wea. Rev.*, **149**, 1481–1497, <https://doi.org/10.1175/MWR-D-20-0077.1>.
- Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330, [https://doi.org/10.1175/1520-0477\(1993\)074<2317:EFANTG>2.0.CO;2](https://doi.org/10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2).
- Tsiringakis, A., A. A. M. Holtslag, S. Grimmond, and G. J. Steeneveld, 2020: Surface and atmospheric driven variability of the single-layer urban canopy model under clear-sky conditions over London. *J. Geophys. Res. Atmos.*, **125**, e2019JD032167, <https://doi.org/10.1029/2019JD032167>.
- , I.-L. Frogner, W. de Rooy, U. Andrae, A. Hally, S. Contreras Osorio, S. van der Veen, and J. Barkmeijer, 2024: HarmonEPS modified routines, post-processing scripts and example data used in Tsiringakis, A., Frogner, I.L., de Rooy, W., Andrae, U., Hally, A., Contreras Osorio, S., van der Veen, S. and Barkmeijer, J., an update to the stochastically perturbed parametrizations scheme of HarmonEPS. Monthly Weather Review. Zenodo, <https://doi.org/10.5281/zenodo.11104411>.
- Tsyrlunikov, M., and D. Gayfulin, 2017: A limited-area spatio-temporal stochastic pattern generator for simulation of uncertainties in ensemble applications. *Meteor. Z.*, **26**, 549–566, <https://doi.org/10.1127/metz/2017/0815>.
- Wastl, C., Y. Wang, A. Atencia, and C. Wittmann, 2019a: A hybrid stochastically perturbed parametrization scheme in a convection-permitting ensemble. *Mon. Wea. Rev.*, **147**, 2217–2230, <https://doi.org/10.1175/MWR-D-18-0415.1>.
- , —, —, and —, 2019b: Independent perturbations for physics parametrization tendencies in a convection-permitting ensemble (PSPPT). *Geosci. Model Dev.*, **12**, 261–273, <https://doi.org/10.5194/gmd-12-261-2019>.
- Wurtz, J., D. Bouniol, B. Vié, and C. Lac, 2021: Evaluation of the AROME model's ability to represent ice crystal icing using in situ observations from the HAIC 2015 field campaign. *Quart. J. Roy. Meteor. Soc.*, **147**, 2796–2817, <https://doi.org/10.1002/qj.4100>.