Using Stochastically Perturbed Parameterizations to Represent Model Uncertainty.
Part I: Implementation and Parameter Sensitivity

RON MCTAGGART-COWAN, a LEO SEPAROVIC, a RABAH AIDER, a MARTIN CHARRON, a MICHEL DESGAGNÉ, a PIETER L. HOUTEKAMER, b DANAHÉ PAQUIN-RICARD, a PAUL A. VAILLANCOURT, b AND AYRTON ZADRA b

a Atmospheric Numerical Weather Prediction Research Section, Environment and Climate Change Canada, Dorval, Quebec, Canada
b Data Assimilation Research Section, Environment and Climate Change Canada, Dorval, Quebec, Canada

( Manuscript received 3 December 2021, in final form 23 June 2022)

ABSTRACT: Accurately representing model-based sources of uncertainty is essential for the development of reliable ensemble prediction systems for NWP applications. Uncertainties in discretizations, algorithmic approximations, and diabatic and unresolved processes combine to influence forecast skill in a flow-dependent way. An emerging approach designed to provide a process-level representation of these potential error sources, stochastically perturbed parameterizations (SPP), is introduced into the Canadian operational Global Ensemble Prediction System. This implementation extends the SPP technique beyond its typical application to free parameters in the physics suite by sampling uncertainty both within the dynamical core and at the formulation level using “error models” when multiple physical closures are available. Because SPP perturbs components within the model, internal consistency is ensured and conservation properties are not affected. The full SPP scheme is shown to increase ensemble spread to keep pace with error growth on a global scale. The sensitivity of the ensemble to each independently perturbed “element” is then assessed, with those responsible for the bulk of the response analyzed in more detail. Perturbations to surface exchange coefficients and the turbulent mixing length have a leading impact on near-surface statistics. Aloft, a tropically focused error model representing uncertainty in the advection scheme is found to initiate growing perturbations on the subtropical jet that lead to forecast improvements at higher latitudes. The results of Part I suggest that SPP has the potential to serve as a reliable representation of model uncertainty for ensemble NWP applications.

SIGNIFICANCE STATEMENT: Ensemble systems account for the negative impact that uncertainties in prediction models have on forecasts. Here, uncertain model parameters and algorithms are subjected to perturbations representing impact on forecast errors. By initiating error growth within the model calculations, the equally skillful members of the ensemble remain physically realistic and self-consistent, which is not guaranteed by other depictions of model error. This “stochastically perturbed parameterization” technique (SPP) comprises many small error sources, each analyzed in isolation. Each source is related to a limited set of processes, making it possible to determine how the individual perturbations affect the forecast. We conclude that SPP in the Canadian Global Ensemble Forecasting System produces realistic estimates of the impact of model uncertainties on forecast skill.

KEYWORDS: Ensembles; Numerical weather prediction/forecasting; Model errors; Stochastic models

1. Introduction

One of the primary goals of ensemble NWP systems is to provide information about the evolution of possible errors over the forecast period. The sources of these errors are the uncertainties inherent in the system, whose origins include incomplete observations, assimilation hypotheses, discretization approximations and deficiencies in the forecast model. Accurately representing these uncertainties is essential for generating a reliable depiction of guidance quality. Sources of uncertainty within the forecast model are the focus of this two-part study, which investigates the utility of a recently proposed form of stochastic model error representation in the context of an operational global ensemble prediction system.

Denotes content that is immediately available upon publication as open access.

Corresponding author: Ron McTaggart-Cowan, ron.mctaggart-cowan@ec.gc.ca

DOI: 10.1175/MWR-D-21-0315.1
For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Two broad classes of strategies to represent model uncertainties have been adopted in operational ensembles: multiformulation and stochastic (Berner et al. 2017). Multiformulation approaches employ different algorithms for different ensemble members to generate a diversity of possible solutions, and include multimodel ensembles (Krishnamurti et al. 1999, 2016; Hagedorn et al. 2005) and multiphysics ensembles (Houtekamer et al. 1996; Stensrud et al. 2000; Berner et al. 2011, 2015). Stochastic strategies depend on distinct perturbations to individual members that promote the divergence of the ensemble in sensitive regions. Stochastically perturbed parameterization tendencies (Buizza et al. 1999; Palmer et al. 2009; Christensen et al. 2015) are an example of such a strategy that has been widely adopted in operational systems around the world (Palmer et al. 2009; Charron et al. 2010; Sanchez et al. 2016; Leutbecher et al. 2017). Stochastic uncertainty representation can also be implemented at the level of individual physical processes including deep convection (Lin and Neelin 2000; Teixeira and Reynolds 2008; Plant and Craig 2008), shallow convection (Sakradzija et al. 2015) and boundary layer turbulence (Rasp et al. 2018;
Hirt et al. 2019; Clark et al. 2021). Such schemes depict only the sampling component of uncertainty (Plant et al. 2011); however, they benefit from the conceptual advantage of seeding ensemble diversity in close proximity to potential error sources and can improve ensemble predictions (Keane et al. 2014; Sakradzija et al. 2020; Wang et al. 2021).

This study focuses on a form of model uncertainty representation that has received increased recent attention: stochastically perturbed parameterizations (SPP). The distinguishing feature of this approach is that perturbations are made to uncertain internal model parameters rather than being applied in a post hoc manner to computed physics tendencies (Leutbecher et al. 2017). This incremental step toward process-level stochasticity allows the SPP scheme to depict uncertainty sources in a comprehensive way that ensures internal consistency and conservation (Lang et al. 2021). Unlike multif ormulation approaches, the SPP technique also preserves the exchangeability of members, an attribute that promotes Gaussianity and is a prerequisite for fair evaluation techniques (Ferro 2014). Building on the work of Li et al. (2008) and Bowler et al. (2008), Ollinaho et al. (2017) describe the development of an SPP scheme in an experimental version of the ECMWF global ensemble. They conclude that despite improvements in short-range and precipitation predictions, the SPP scheme is unable to produce forecasts that are as skillful as those of the operational ensemble. Jankov et al. (2017) suggest that the SPP scheme is unable to generate sufficient ensemble spread because its sampling of potential error sources is incomplete, a finding that is consistent with the demonstration by Jankov et al. (2019) that combining SPP with other stochastic methods currently yields optimal ensemble performance. However, Lang et al. (2021) note that care must be taken in selecting these techniques to retain the conceptual benefits that the SPP scheme affords.

Spurred by the roadmap for stochastic algorithm development proposed by Leutbecher et al. (2017), refinement of the SPP scheme in the ECMWF ensemble has resulted in a configuration that yields medium-range forecasts that are generally as skillful as those generated by the operational system (Lang et al. 2021). Despite this rapid progress, important questions remain about the effects of model resolution and cycling within the ensemble data assimilation system. Similarly, Kalina et al. (2021) show that the SPP technique yields physically realistic solutions that improve the reliability of a convection-permitting ensemble; however, the perturbations worsen a dry bias in the system so that the resulting forecasts lack the sharpness required for severe-weather forecasting. These results suggest that continued development of the SPP approach has the potential to lead to a scheme that is capable of serving as the primary representation of model uncertainty in an operational context.

This study documents the development of an extended SPP scheme in the Global Ensemble Prediction System (GEPS; Houtekamer et al. 2014). In addition to free-parameter perturbations (Ollinaho et al. 2017), uncertainties in the model’s dynamical core and in the formulation of physical closures are sampled in this implementation. The resulting scheme attempts to represent a broader range of potential error sources than did its predecessors; however, there is a commensurate increase in the computational cost and the complexity of the system. As a result, careful examination of the impact of each individual perturbation (referred to as SPP “elements” hereinafter) is required to ensure an optimal balance between these competing imperatives. This element-level investigation builds on the findings of studies such as Reynolds et al. (2011) and Frogner et al. (2022) to provide insight into the physics and dynamics of model error growth that affects guidance quality. At the same time, it expands the catalog of well-documented sensitivities that will inform the development of future model uncertainty schemes.

The GEPS is introduced in section 2, with the implementation of the SPP scheme described in section 3. Section 4 contains a summary of results from both the full SPP scheme and a series of sensitivity tests in which individual elements are activated. For some of the leading elements, further investigations of processes underlying the observed sensitivities are presented in section 5. The first part of the study concludes with a discussion in section 6, which sets the stage for comparison between results from the SPP scheme and those obtained with the existing operational model uncertainty representations in McTaggart-Cowan et al. [2022, hereinafter Part II (of this investigation)].

2. Data and methods

The Global Environmental Multiscale (GEM) model is employed for all simulations discussed in this study. Its dynamical core uses implicit two-time-level temporal discretization and semi-Lagrangian advection to solve the Euler equations (Côté et al. 1998; Yeh et al. 2002; Girard et al. 2014). The global domain is represented using a pair of overlapping limited-area Arakawa C grids (Arakawa 1988) in a so-called yin–yang configuration (Qaddouri and Lee 2011) with 0.35° grid spacing. The model employs 84 terrain-following log-hydrostatic-pressure-based vertical levels (Girard et al. 2014) with a top at 0.1 hPa.

Diabatic and unresolved processes are represented using a physical parameterization suite that recently underwent a major update (McTaggart-Cowan et al. 2019b). The SPP-based GEPS ensemble employs a single physics configuration that is summarized in Table 1. Although operational GEPS forecasts include coupling to the NEMO ocean model (Gurvan et al. 2017; Smith et al. 2018), the simulations here are atmosphere only, with sea surface temperatures fixed to their analyzed values as in the GEPS data assimilation cycle (Houtekamer et al. 2014).

The ensemble system employed in this study is a simplified form of the forecast component of the GEPS. It is initialized by subsampling analyses from the 256-member data assimilation ensemble over a 2-month boreal winter period (January–February 2020) using two different strategies. The system’s response to the full set of SPP elements is assessed in section 4 using a 20-member ensemble initialized at 36-h intervals. Forecasts for the 40 dates extend to 360-h lead times to duplicate the behavior of the operational system. To assess the isolated
Table 2. Summary of primary con-
a. The stochastic field generator

An important ingredient in stochastic representations of model error is the structure of the perturbations. Although white noise (without temporal or spatial correlations) could be considered as a possible method for representing the seeds of uncertainty in a chaotic system (Lorenz 1969), Zagar (2017) and Christensen (2020) document the prevalence of large-scale errors in NWP forecasts. As a result, temporally and spatially correlated perturbations are needed to inject variability on scales that accurately reflect the uncertainties in the system.

The GEM model supports the representation of stochastic processes by creating a set of independent pseudorandom patterns using first-order autoregressive [AR(1)] processes. These fields \( F(L, \phi, t) \), where \( L, \phi \), and \( t \) represent longitude, latitude, and time, respectively, can be used to seed the perturbations that represent uncertainty in any component of the model. The construction of \( F \) is described in detail in the appendix and differs from the ECMWF implementation (Ollinaho et al. 2017) in both formulation and configuration (the primary parameters for the stochastic field generator are identified in Table 2 for reference). An example of \( F \) produced by the stochastic field generator for the global wavenumber range 1–8 (Table 2) is shown in Fig. 1a. This wavenumber range corresponds to a spatial decorrelation length scale of \( \sim 1500 \) km, the distance at which autocorrelation falls to \( e^{-1/2} \). Although a comparison of Fig. 1a with Figs. 1b and 1c of Lang et al. (2021) confirms that the patterns in \( F \) fall within the 1000–2000-km range, the conversion between decorrelation length scale- and wavenumber-based definitions is nontrivial and will appear in a separate document currently in preparation.

The demands of SPP for a large number of independent perturbation fields prompted an optimization effort within the stochastic field generator. As a result, addition of the full set of perturbation fields incurs a negligible computational cost. However, the use of a coarse-resolution global grid in the inverse transform of the spherical harmonics reduces the variance of \( F \) and creates anisotropy near the pole (Fig. 1; details are provided in the appendix). The configuration of the stochastic field generator will be updated to rectify these issues as part of ongoing system development.

b. SPP implementation in GEM

The stochastic fields \( F \) described above are used to perturb individual elements within the model (Table 3). There

Fig. 1. (a) Sample spatial pattern of \( F(\Lambda, \phi, t) \), for the perturbation range [−2.5, 2.5] as used by the continuous ml_emod SPP element (Table 3). (b) An example of the discrete perturbation pattern for the deeptrig SPP element is shown. Field values are color filled following the color bars below each panel.
**Table 3.** Description of parameters and algorithms perturbed with the SPP scheme, split into subsections by grouping of physical processes. The elements that are employed in the full SPP configuration (section 4a) are identified with boldface type. The “Applications” column describes the role of $\mathcal{F}$ in perturbation design. It can take the values of Factor ($\mathcal{F}$ is used as a multiplying factor), Value (the values of a parameter are given by $\mathcal{F}$), Closure (different closures are used depending on $\mathcal{F}$), or Error Model ($\mathcal{F}$ is used in an error model as described in section 3c). The “Range” column refers to the limits imposed on $\mathcal{F}$ ($\mathcal{F}_{\min}$–$\mathcal{F}_{\max}$), and the “Shape” column refers to the width of the field distribution $\gamma$ as defined in Table 2. Acronyms used in this table are as follows: BL indicates Bougeault and Lacarrere (1989), BLK indicates Blackadar (1962), BLK-BL indicates a blending of BL and BLK that is based on turbulence regime, KF indicates Kain and Fritsch (1993), KFU indicates Kain (2004), and TKE indicates turbulence kinetic energy–based trigger based on concepts introduced by Mapes (2000). The repeat count of acronyms in discrete range sets indicates the prevalence of the specified formulation across the grid.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Type</th>
<th>Application</th>
<th>Range</th>
<th>Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Surface and boundary layer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fh_mult</td>
<td>Turbulent exchange coefficient for scalar surface fluxes</td>
<td>Continuous</td>
<td>Factor</td>
<td>[0.5, 1.5]</td>
<td>1</td>
</tr>
<tr>
<td>fm_mult</td>
<td>Turbulent exchange coefficient for surface momentum fluxes</td>
<td>Continuous</td>
<td>Factor</td>
<td>[0.5, 1.5]</td>
<td>1</td>
</tr>
<tr>
<td>tkediff</td>
<td>Turbulent transport of turbulence kinetic energy</td>
<td>Continuous</td>
<td>Factor</td>
<td>[0, 2]</td>
<td>1</td>
</tr>
<tr>
<td>fnnreduc</td>
<td>Turbulent flux profile adjustment to represent mixing in boundary layer clouds</td>
<td>Continuous</td>
<td>Value</td>
<td>[0.5, 1.5]</td>
<td>1</td>
</tr>
<tr>
<td>ricmin</td>
<td>Critical Richardson no. for flow transition from laminar to turbulent</td>
<td>Continuous</td>
<td>Value</td>
<td>[0, 0.5]</td>
<td>1</td>
</tr>
<tr>
<td>ml_emod</td>
<td>Boundary layer mixing length scale</td>
<td>Continuous</td>
<td>Error Model</td>
<td>[−2.5, 2.5]</td>
<td>1</td>
</tr>
<tr>
<td>longmel</td>
<td>Formulation of the boundary layer mixing length scale</td>
<td>Discrete</td>
<td>Closure</td>
<td>[BL, BLK, BLK-BL]</td>
<td>1.39</td>
</tr>
<tr>
<td><strong>Moist convection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kfcrig4</td>
<td>Critical vertical motion for triggering deep convection over land</td>
<td>Continuous</td>
<td>Value</td>
<td>[0, 0.1] m s$^{-1}$</td>
<td>2</td>
</tr>
<tr>
<td>kfcrigwl</td>
<td>Threshold convective velocity scale for triggering convection under low-wind conditions over water</td>
<td>Continuous</td>
<td>Value</td>
<td>[0, 0.1] m s$^{-1}$</td>
<td>2</td>
</tr>
<tr>
<td>kfcrigwh</td>
<td>Threshold convective velocity scale for triggering convection under high-wind conditions over water</td>
<td>Continuous</td>
<td>Value</td>
<td>[−0.05, 0.05] m s$^{-1}$</td>
<td>2</td>
</tr>
<tr>
<td>mid_minemf</td>
<td>Triggering resolve mass-flux threshold for low-CAPE convection</td>
<td>Continuous</td>
<td>Value</td>
<td>[12, 110] × 10$^{-6}$ kg m s$^{-1}$</td>
<td>1</td>
</tr>
<tr>
<td>crad_mult</td>
<td>Deep convective updraft radius at cloud base</td>
<td>Continuous</td>
<td>Factor</td>
<td>[0.5, 1.5]</td>
<td>1</td>
</tr>
<tr>
<td>deeprate</td>
<td>Cloud–rain autoconversion rate in deep convective clouds</td>
<td>Continuous</td>
<td>Value</td>
<td>[0.005, 0.015] s$^{-1}$</td>
<td>1</td>
</tr>
<tr>
<td>dpdd_mult</td>
<td>Downdraft detrainment depth for deep convective clouds</td>
<td>Continuous</td>
<td>Factor</td>
<td>[0.5, 1.5]</td>
<td>1</td>
</tr>
<tr>
<td>deeptrig</td>
<td>Formulation of the trigger function for deep convection</td>
<td>Discrete</td>
<td>Closure</td>
<td>[KF, KF, TKE, KFU]</td>
<td>1</td>
</tr>
<tr>
<td>bkf_evaps</td>
<td>Fraction of detrained shallow-convective condensate evaporated within the scheme</td>
<td>Continuous</td>
<td>Value</td>
<td>[0, 1]</td>
<td>1</td>
</tr>
<tr>
<td><strong>Gridscale clouds and radiation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu0min</td>
<td>Threshold relative humidity for upper-level cloud development</td>
<td>Continuous</td>
<td>Value</td>
<td>[0.75, 0.95]</td>
<td>1</td>
</tr>
<tr>
<td>hu0max</td>
<td>Threshold relative humidity for lower-level cloud development</td>
<td>Continuous</td>
<td>Value</td>
<td>[0.81, 0.99]</td>
<td>1</td>
</tr>
<tr>
<td>cond_hcst</td>
<td>Cloud condensate threshold for autoconversion of cloud droplets to raindrops</td>
<td>Continuous</td>
<td>Value</td>
<td>[5, 15] × 10$^{-3}$ kg kg$^{-1}$</td>
<td>1</td>
</tr>
<tr>
<td>rew_mult</td>
<td>Effective cloud water droplet radius for radiative transfer</td>
<td>Continuous</td>
<td>Factor</td>
<td>[0.5, 1.5]</td>
<td>1</td>
</tr>
<tr>
<td>rei_mult</td>
<td>Effective cloud ice radius for radiative transfer</td>
<td>Continuous</td>
<td>Factor</td>
<td>[0.5, 1.5]</td>
<td>1</td>
</tr>
<tr>
<td>aero_mult</td>
<td>Aerosol concentration for radiative transfer</td>
<td>Continuous</td>
<td>Factor</td>
<td>[0.5, 1.5]</td>
<td>1</td>
</tr>
<tr>
<td>hetero_mult</td>
<td>Subgrid-scale cloud heterogeneity for radiation</td>
<td>Continuous</td>
<td>Factor</td>
<td>[0.5, 1.5]</td>
<td>1</td>
</tr>
<tr>
<td><strong>Momentum and dynamics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sgo_phic</td>
<td>Flow blocking by subgrid-scale orography</td>
<td>Continuous</td>
<td>Value</td>
<td>[0.07, 0.27]</td>
<td>1.39</td>
</tr>
<tr>
<td>rmscon</td>
<td>Velocity spectrum rms of nonorographic gravity waves at the launching level</td>
<td>Continuous</td>
<td>Value</td>
<td>[0.2, 1.8] m s$^{-1}$</td>
<td>1</td>
</tr>
<tr>
<td>adv_rhsint</td>
<td>Order of interpolation for departure-point value in semi-Lagrangian advection ($\tau = 24$ h)</td>
<td>Continuous</td>
<td>Error Model</td>
<td>[−0.4, 0.4] [Eq. (11)]</td>
<td>2</td>
</tr>
<tr>
<td>phycpl</td>
<td>Fraction of split-explicit dynamics–physics coupling</td>
<td>Continuous</td>
<td>Value</td>
<td>[0.6, 1]</td>
<td>2</td>
</tr>
</tbody>
</table>
FIG. 2. Evolution of global spread and error of (a),(d) 250-hPa zonal wind; (b),(e) 500-hPa geopotential height; and (c),(f) 850-hPa temperature in control and SPP-based ensembles, showing (left) the spread (dashed lines) and RMSE$_u$ [solid lines; Eq. (5)] for the control (black) and SPP-based (red) integrations and (right) the fractional contribution of the SPP scheme to ensemble spread computed using Eq. (2), expressed as a percentage ($100 \times C_f$). Also shown in (d)–(f) are color-shaded backgrounds for the forecast pentads.
are two forms of perturbations: continuous and discrete. For continuous perturbations (Fig. 1a), $F$ typically either takes on the values of a free parameter or serves as a multiplying factor. An exception to this basic application of continuous perturbations occurs in the form of SPP error models (section 3c). Discrete perturbations are designed to permit sampling of the uncertainty associated with algorithm formulation by allowing components of individual schemes to be employed in a stochastic framework. For example, the type of trigger function used for deep convection can be made to vary spatiotemporally in each ensemble member (Fig. 1b).
FIG. 4. Spread-reliability diagrams (Leutbecher et al. 2007) for 72-h forecasts of (left) 850-hPa temperature and (right) 250-hPa zonal winds in the control (black) and SPP-based (red) ensembles in (a), (d) the northern midlatitudes; (b), (e) the tropics; and (c), (f) the southern midlatitudes. Filled circles indicate means for 5-percentile bins, and filled squares indicate the overall means for the ensembles. The ideal 1:1 reference is shown with a dashed gray line. Note that the ranges of the tropical [(b)] and midlatitude [(a) and (c)] diagrams differ to reflect regional differences in spread and error. Both spread and error are normalized by the climatological standard deviation, computed using the ERA-Interim reanalysis (Dee et al. 2011) for the period 1989–2016.
Candidate elements to be perturbed by the SPP scheme
were identified by subject matter experts as those that were
both uncertain and likely to have a significant impact on
the model solution. The parameters for each $\mathcal{F}$ (Table 3)
were constructed on the basis of this input and adjusted using
spread-reliability evaluations (Leutbecher et al. 2007) in a
rapid test configuration (not shown). Although a uniform
sampling distribution ($\gamma = 1$) was prescribed for most SPP
elements to maximize the scheme’s impact, the stochastic
field generator yields distributions with a shape closer to
$\gamma = 1.45$ as described in the appendix. This implicit use of an
“informed prior” reduces the amplitude of SPP perturbations
but is more consistent with parameter uncertainty (Posselt
and Vukicevic 2010). Additional SPP elements are technically
simple to introduce, a feature that makes the scheme highly
flexible as new potential error sources are identified.

The use of 36-h relaxation $\tau$ for the majority of perturba-
tions (Table 3) is consistent with the long time scales em-
ployed in other SPP implementations (Lang et al. 2021) and
in the stochastic tendency perturbations used by the GEPS
until December 2021 (Houtekamer et al. 2009). The SPP
scheme’s impact depends on $\tau$ and other leading parameters
that define $\mathcal{F}$: a study is currently under way to quantify
these sensitivities.

c. SPP error models

If uncertainty in the value of a given parameter or the for-
mulation of a closure is flow dependent, then the associated
perturbations should adapt themselves to reflect this reality.
Such uncertainty awareness is possible when multiple esti-
mates ($N > 1$) of a single quantity $\chi$ are available: an “error
model” can be designed to scale perturbation amplitude by
the diversity in these estimates. Under the assumption that
the uncertainty of $\chi$ is reflected in the range of the $N$ com-
puted values ($\chi_{i-1}, \ldots, \chi_N$), the perturbed $\chi'$ samples around the
unperturbed configuration $\chi$ using the error model

$$\chi' = \chi + \mathcal{F} \sqrt{(\chi_i - \langle \chi \rangle)^2},$$

(3)

where angle brackets indicate the mean of the $N$ estimates
and $\mathcal{F}$ is centered on zero.

A concrete example of an error model as an SPP element
appears in the ml_emod configuration (sections 4b and 5b).
The turbulent mixing length $\ell$ in the boundary layer scheme is
computed using the Blackadar (1962) estimate in the con-
tr$\ell$ [in Eq. (3)], with perturbations constructed on the basis of the
difference between this estimate and one computed us-
ing the approach proposed by Bougeault and Lacarrère
(1989). With only one alternative estimate available (subscript
“alt”), Eq. (3) is simplified to

$$\chi' = \chi + \mathcal{F} |\chi_i - \chi_{\text{alt}}|,$$

(4)

a form that yields results similar to Eq. (3) when the range of
$\mathcal{F}$ is suitably rescaled (not shown).

Using either Eq. (3) or Eq. (4), the error model amplifies
perturbations under conditions in which differing closures yield
divergent estimates of $\chi$. The distributions of the perturbations
are centered on the unperturbed value, a prerequisite for the
development of an ensemble whose perturbed-member cli-
mates resemble that of the control. The SPP elements that em-
ploy error models are shown below to have significant impacts on
the forecast skill of the ensemble.

4. Ensemble predictions with SPP

The success of the SPP scheme as a representation of model
uncertainty will ultimately be determined by its ability to im-
prove ensemble forecast skill. This implies that the spread
generated by SPP should accurately reflect flow-dependent
forecast error. The impact of the full SPP scheme on GEPS
guidance is introduced in section 4a. The large number of
perturbed parameters, however, makes it difficult to identify the
sources of forecast improvements. This task is undertaken in
section 4b, where the sensitivity to individual SPP elements is
assessed.

a. Model uncertainty representation by SPP

A necessary condition for a well-balanced ensemble is that
error and spread growth rates be similar. Confirming this be-
havior would be an important milestone because SPP schemes have
been found to suffer from underdispersion (Ollinaho et al. 2017).
For this analysis, error is defined as the unbiased
RMSE estimator\footnote{Throughout this study the term “unbiased” refers to the stan-
dard statistical correction of an estimator to account for finite
ensemble size. The ensemble data themselves have not been
postprocessed for any of the analyses shown here.} of the ensemble mean (RMSE$_{\mu}$) for state variable $x$:

$$\text{RMSE}_{\mu}(x) = \sqrt{\frac{1}{M(M-1)} \sum_{j=1}^{M} (x_j - \mu_j(x))^2}$$

$$\text{RMSE}_x(x) = \sqrt{\frac{1}{M} \sum_{j=1}^{M} (x_j - \mu_j(x))^2}$$

$$= \sqrt{\text{RMSE}_x(x) - \frac{1}{M} \sigma^2(x)}.$$
where \( j \) indexes each of the \( M = 20 \) perturbed members, \( y \) is the analyzed value, \( \mu_e \) is the ensemble mean, \( \sigma_e^2 \) is the unbiased ensemble variance estimator, and overbars indicate spatio-temporal averages. In a perfect ensemble, RMSE\(_u\) (\( x \)) = \( \sigma_e \) (\( x \)) (Leutbecher and Palmer 2007).

The SPP scheme yields global spread growth that closely mirrors RMSE\(_u\) in the middle and upper troposphere (Figs. 2a,b). Temperatures at the top of the boundary layer remain underdispersed throughout the 15-day forecast as a result of insufficient spread in the initializations and a suppressed spread growth rate over the first 48 h of integration (Fig. 2c). The SPP scheme does not contribute to an increase in the error of the ensemble mean, a result that is consistent with the scheme’s objective to represent random errors within regions of enhanced uncertainty.

The SPP scheme’s fractional contribution \( C_f \) to ensemble spread \([f = \sigma_e(x)\] in Eq. (2)] reaches a maximum of ~18% in pentad-1 forecasts (Fig. 2, second column). A comparison of model uncertainty representations in section 3 of Part II shows that this is comparable to the contributions of schemes used in the operational GEPS. The peak \( C_f \) for upper-level winds (Fig. 2d) is delayed because few SPP elements directly affect this quantity; instead, spread appears to develop through balance adjustments to mass-field perturbations.

![Summary of sensitivities to SPP elements related to turbulent transport in the surface and boundary layer.](Fig. 5)

### Elements Related to Turbulent Transport in the Surface and Boundary Layer

**Fractional contributions to (a)–(c) ensemble spread and (d)–(f) fCRPS are plotted as in Fig. 3, except that the temporal abscissa is replaced by individual SPP elements (Table 3) such that each column in the panels represents pentad-1 sensitivity diagnosed from an independent test. The names of elements that are included in the full SPP configuration (section 4a) are shown in black, and those that were rejected are shown in gray. The range of the color bars has been reduced from the values used in Fig. 3 because of the limited amplitude of individual SPP element contributions.**

**Fig. 5.** Summary of sensitivities to SPP elements related to turbulent transport in the surface and boundary layer.
The decay of SPP’s relative contribution at longer lead times is the result of saturation as the ensemble approaches its climatological dispersion.

A more complete overview of the ensemble’s response to the full SPP scheme is shown in Fig. 3. The first column suggests that the spread increases noted in Fig. 2 are largest in the tropics (Fig. 3b), although the scheme also contributes to diversity in the northern (winter) midlatitudes and to near-surface spread in pentad-1 forecasts in the southern (summer) midlatitudes. The greater longevity of significant SPP contributions in the tropics (Fig. 3b) is a reflection of the slower approach to climatological spread in areas where moist baroclinic growth is not a primary driver of variability (Judt 2020). These process-level distinctions in the GEPS response imply that all sensitivities must be assessed in the tropical and midlatitude regions independently.

The impact of SPP on probabilistic forecast skill is evaluated through the scheme’s contribution to the fair continuous ranked probability score \( f_{\text{CRPS}} \) (Ferro et al. 2008; Fricker et al. 2013). This \( f_{\text{CRPS}} \) is computed using Eq. (2), with \( J \) representing \( f_{\text{CRPS}} \) defined following Leutbecher (2019) as follows:

\[
f_{\text{CRPS}}[(x_j, y)] = \frac{1}{M} \sum_{j=1}^{M} |x_j - y| - \frac{1}{2M(M-1)} \sum_{j=1}^{M} \sum_{k=1}^{M} |x_j - x_k|,
\]

where subscripts \( j \) and \( k \) refer to members. The fair variant differs from the standard score to correct for the limited ensemble size.

The full SPP scheme reduces the \( f_{\text{CRPS}} \) in all regions, indicative of forecast skill improvements (second column of Fig. 3). The largest impact of SPP is in the tropics, where significant

![Fig. 6](image-url)
forecast improvements are observed for all variables and lead times. In the midlatitudes, significant fCRPS reductions are generally restricted to lower-tropospheric temperatures and are most apparent in pentad 1.

Although these summary evaluations suggest that the SPP scheme improves ensemble behavior in the aggregate, a more refined assessment of the relationship between spread and error is possible using spread-reliability diagrams (Leutbecher et al. 2007). A well-balanced ensemble that reliably discriminates between high- and low-predictability conditions will have a distribution that falls along the diagonal of these plots; however, the limited GEPS ensemble size will tend to flatten this slope. An analysis of 72-h forecasts of 850-hPa temperature and 250-hPa zonal wind illustrates the impact of SPP on the ensemble from the spread-reliability perspective (Fig. 4). The control ensemble is underdispersed in all regions because it does not sufficiently account for uncertainties within the model. The slope of the distribution suggests that spread and error are well coupled in the midlatitudes (Figs. 4a,c); however, steepness in the tropics implies that large errors can develop in regions without increased spread (Fig. 4b). The SPP scheme shifts all distributions toward the diagonal, a result that is consistent across variables and lead times (not shown). This suggests that the SPP scheme implemented in GEPS improves the ensemble’s ability to depict the flow-dependent nature of spread and error growth.

b. Sensitivity to individual SPPs

The full SPP scheme comprises independent perturbations to numerous model components (Table 3), making it difficult to identify direct relationships between specific perturbations and the ensemble results described above. Such information, however, is extremely valuable to ensemble system designers, model developers and researchers interested in understanding how model uncertainties affect forecast error growth. Only candidate SPP elements that have a neutral or positive impact on GEPS performance from the spread or fCRPS perspective are considered for adoption in the full SPP configuration. Those with neutral impacts are included if they possess strong physical justification, often in the form of well-documented uncertainties, or if they are known to affect particular forms of high impact weather that are not well described by summary statistics.

In this section, sensitivity tests based on a minimized ensemble configuration (section 2; \(M = 10\)) are used to assess the impact of each SPP element in isolation. This framework reproduces the summary statistics of the full GEPS (Table 4), at a nearly 20-fold decrease in computational cost. Because of the potential for interactions between perturbations, the sensitivities identified by activating each SPP element in isolation will not sum to the total sensitivity shown in Fig. 3 (Posselt and Vukicevic 2010). However, this set of tests will make it possible to characterize the direct impact of each SPP element and to identify the most important perturbations so that they can be examined in more detail in section 5. The SPP elements are grouped by process as in Table 3 for this discussion to divide the large volume of information into physically relevant subsets.

1) SPP ELEMENT SENSITIVITY: SURFACE AND BOUNDARY LAYER

Perturbations to elements related to turbulent transport in the planetary boundary layer are highly effective at generating
spread and improving forecast skill (Fig. 5). The ensemble is particularly sensitive to the perturbation of surface exchange coefficients (fh_mult and fm_mult) and to the representation of the mixing length (ml_emod and longmel). Because of the large impact that these SPP elements have on the solution, they will be analyzed in more detail in sections 5a and 5b, respectively. In addition to their large amplitudes, these sensitivities are unique in their geographical scope, which extends through the midlatitudes in both hemispheres.

Other SPP elements related to the boundary layer have a limited impact on the ensemble metrics shown here; however, their intent is to sample uncertainties associated with specific conditions that are not well represented by summary statistics. The perturbations in ricmin sample uncertainty in the value of the critical Richardson number for turbulence transitions, a quantity that influences the model’s representation of freezing rain profiles (McTaggart-Cowan and Zadra 2015). Similarly, uncertainty in the transport of turbulence kinetic energy affects surface temperatures under very stable conditions but does not have a significant impact on hemispheric-scale statistics. Such elements are retained in the SPP configuration because their flow-dependent sensitivities may be particularly relevant to predictions of high impact weather events.

2) SPP ELEMENT SENSITIVITY: MOIST CONVECTION

Despite the known limitations of convection parameterizations (Molinari and Dudek 1992; Arakawa 2004) and their impacts on forecast skill (Rodwell et al. 2013; Lillo and Parsons 2017), sensitivity to SPP elements related to moist convection is relatively small, particularly in the midlatitudes over the period of interest [Fig. 6; (Reynolds et al. 2011)]. Perturbation

Fig. 8. Summary of sensitivities to SPP elements related to gridscale clouds, cloud–radiation interactions, and radiative transfer, plotted as in Fig. 5.
of the cloud updraft radius (crad_mult) yields significant improvements in the tropics, likely because this element samples uncertainties in the entrainment and detrainment rates that affect cloud depth and transport processes (de Rooy et al. 2013). Perturbing the formulation of the trigger function for deep convection (deeptrig) introduces significant variability (Suhas and Zhang 2014); however, the worsening of a warm bias at 850 hPa (not shown) increases the fCRPS and precludes the use of this element in the full SPP configuration. Despite this problem, the limited sensitivity to perturbations within the existing trigger (kfcstrig, ktcstrigw and kfcstrigh) suggests that sampling formulation uncertainty has the potential to contribute to forecast skill improvements in the future (Li et al. 2008).

The other convective element not adopted in the proposed SPP configuration (bkf_evaps) also has a primarily positive impact on the fCRPS. Allowing a variable fraction of detrained shallow convective condensate to be evaporated within the scheme, as opposed to being handled by gridscale condensation, favorably cools the lower troposphere but worsens the 500-hPa tropical height bias to increase the fCRPS (Figs. 6e and 7). This outcome highlights the impact of both a tropical surface pressure bias (negative 1000-hPa height bias in Fig. 7b) and an indirect perturbation response, here in the form of reduced deep convective heating in the upper troposphere (Fig. 7a). These sensitivities also illustrate a problem with perturbations to quantities whose limiting values are employed in the control configuration. Assuming a linear response, centering the ensemble mean on the control member requires symmetric perturbations. In this case, all detrained condensate in the control member is passed to the gridscale scheme, such that perturbations can only act to reduce the transfer. This creates a conceptual problem for such SPP elements, which in this case manifests as an increased bias and fCRPS degradation.

3) SPP ELEMENT SENSITIVITY: GRIDSCALE CLOUDS AND RADIATION

Perturbations to parameters within the gridscale condensation scheme and affecting cloud–radiation interactions generally have relatively little impact on the summary metrics shown in Fig. 8, although sampling uncertainty in the relative humidity threshold required for cloud formation in the lower troposphere [hu0max; (Li et al. 2008)] yields modest fCRPS improvements. Other SPP elements in this group are adopted for the full SPP configuration because of known uncertainties. For example, the concept of effective ice particle radius is ill-defined for current microphysical schemes given the range of cloud–radiation interactions possible for different crystal habits (McFarquhar and Heymsfield 1998). Sampling this uncertainty with the rei_mult element is physically justified and has a direct effect on high-cloud albedo estimates (Fig. 9a). Despite minimal sensitivity in the summary statistics (Fig. 8), this change in reflection can have a large impact on the local surface energy budget (±20%; Fig. 9b) and is known to affect both tropospheric stability and tropical cyclogenesis (Caron et al. 2012). Perturbation of the highly parameterized subgrid-scale cloud variability estimate [hetero_mult; (Oreopoulos and Barker 1999)] induces systematic upper-level cooling that increases the fCRPS, rendering this element unsuitable for the full SPP configuration. Further efforts will be required to determine the optimal approach for sampling this uncertainty.

4) SPP ELEMENT SENSITIVITY: MOMENTUM AND DYNAMICS

Sensitivities to SPP elements related to momentum transport and dry dynamics are shown in Fig. 10. Those affecting parameterized drag processes (sgo_phic and rmscon) have
minimal impacts on the summary statistics; however, perturbing the velocity spectrum of nonorographic gravity waves promotes diversity in the wave-breaking layers of the upper stratosphere. Evaluations using satellite data show that the additional spread is beneficial for the assimilation component of the GEPS (these improvements will be documented in a forthcoming paper).

The two SPP elements that perturb quantities outside the physical-parameterizations suite are adv_rhsint and phycpl (Fig. 10). The former uses an error model to estimate uncertainty within the advection scheme and will be discussed in more detail in section 5c. The latter shows that the choice of dynamics-physics coupling strategy has a significant impact on forecast quality (Gross et al. 2018). Activating the phycpl SPP element replaces the split-explicit treatment of the model components with the implicit handling of a portion of the total physics tendencies. This decreases ensemble spread because the implicit tendency treatment damps variance over a broad range of wavenumbers (Fig. 11). Although arguably occurring at scales that are unpredictable by the end of pentad-1 forecasts, this spread reduction increases the fCRPS through the second term on the rhs of Eq. (6). This degradation precludes the adoption of the phycpl element in the full SPP configuration. Further investigation will be required to determine whether this sensitivity is indicative of true uncertainty in the system that should be sampled as a potential source of model error, or whether the current approach is simply optimal for the adopted model configuration.

5. Leading SPP parameters

The preceding analysis of individual SPP element sensitivities reveals a broad range of responses to the stochastic

![Fig. 10. Summary of sensitivities to SPP elements related to momentum transport and dry dynamics, plotted as in Fig. 5.](image)
perturbations. Even if the level of uncertainty in two parameters is of the same order, perturbations may have significantly different impacts on the solution as a result of differing and possibly flow-dependent sensitivities within the model. In this section, the SPP elements that were found to have the largest impacts on the GEPS are analyzed in more detail to provide physical explanations for the behavior of the SPP-based ensemble.

**a. Surface exchange coefficient perturbations**

Perturbations to the surface turbulent exchange coefficients for scalars (fh_mult) and momentum (fm_mult) have large impacts on ensemble spread and fCRPS at a global scale (Fig. 5). Surface fluxes \( F_x \) are computed using the bulk flux formula

\[
F_x = C_x(x_a - x_s),
\]

with \( x \) as temperature \( T \), specific humidity \( q \), or wind components \( \mathbf{V} = (u, v) \) at the lowest model level (subscript “a”) and the surface (subscript “s”), modulated by exchange coefficients \( C_x \). The interfacial flux \( B_x \) that serves as the lower boundary condition for the implicit solution of boundary layer equations is perturbed to yield

\[
B'_x = \mathcal{F} F_x = -\mathcal{F} C_x(x_a - x_s) = -C'_x(x_a - x_s),
\]

with \( \mathcal{F} = [0.5, 1, 5] \), where the final equality shows that the fh_mult and fm_mult SPP elements represent exchange coefficient perturbations for the boundary layer scheme \( (C'_s = \mathcal{F} C_s) \). They thereby directly affect heat and momentum transfers within the atmospheric column rather than adjusting the surface fluxes themselves.

---

**FIG. 11.** Contribution by the phycpl SPP element to (a) the mean global kinetic energy spectrum for 250-hPa zonal wind and (b) the mean global 850-hPa temperature variance spectrum after 120 h of integration. This contribution is computed using Eq. (2) and is expressed as a percentage \((100 \times C_f)\) such that negative values are associated with a decrease in spectral variance. The 95% confidence interval for the sensitivity is shown in light gray shading, with intervals that do not reach 0 (horizontal gray dashed line) indicative of significant departures based on a 1000-member bootstrap.

**FIG. 12.** Sensitivity of (a) screen-level temperature spread \((^\circ C)\) and (b) fCRPS \((^\circ C)\) to the fh_mult SPP element, computed as differences with respect to the control ensemble averaged over pentad 1. Differences that are determined to be statistically significant for a false discovery rate (Wilks 2016) of 5% based on a 1000-member bootstrap are dark-shaded following the bottom color bar on each panel, and those that fail to meet this level of significance are light-shaded following the top color bar.
The multiplicative scaling used in Eq. (8) means that the direct impacts of these perturbations are maximized in regions where turbulent surface exchanges are typically large (Fig. 12; results shown only for fh_mult because of the large sensitivity shown in Fig. 5). Differences therefore peak along the western boundary currents (Munk 1950) and cold outbreak regions in the Northern Hemisphere winter (Fletcher et al. 2016; Smith and Sheridan 2020), and over landmasses in the tropics and Southern Hemisphere (summer). The enhanced spread in continental regions is accompanied by a significant decrease in the fCRPS, indicating that the broad perturbations range reflects underlying uncertainty in the forecasts.

In the North Atlantic Ocean, northerly flow in the wake of oceanic cyclones (Tilinina et al. 2018) and high winds associated with the Greenland tip jet (Doyle and Shapiro 1999) draw Arctic air across the extensive ice edge to create an environment conducive to frequent large-flux events (Papritz and Spengler 2017; Geerts et al. 2022). Despite symmetric perturbations to $B'_x$, the fh_mult ensemble predicts colder screen-level temperatures in this region (Figs. 13a,b). This cooling occurs despite locally increased surface heat fluxes (Figs. 13c,d) as a nonlinear response to this SPP element. For $\mathcal{F} \lesssim 0.8$, reduced $B'_x$ values severely limit atmospheric heating and prevent the erosion of the Arctic inversion. The persistence of the very stable boundary layer regime (Abraham and Monahan 2019) yields near-surface temperatures that remain colder than in the control (Fig. 14a). The increased air-sea temperature difference subsequently enhances the unperturbed heat fluxes across the interface [Eq. (7) and Fig. 14b]. Because this process is active under specific conditions favored in the North Atlantic, its influence on globally averaged fluxes remains small ($\lesssim 0.2 \text{ W m}^{-2}$). An experiment in which $\mathcal{F}$ is restricted to the linear response range in Fig. 14 ($0.8 \lesssim \mathcal{F} \lesssim 1.2$) successfully eliminates the temperature bias noted in Fig. 13; however, this ensemble suffers from a significant reduction in forecast skill improvement relative to the full fh_mult configuration (not shown). It therefore seems reasonable to accept that such changes in the mean state represent a true nonlinear response to a potential error source within the model rather than a failure of the SPP strategy in this case.

b. Uncertainties in the boundary layer mixing length

Both representations of uncertainty in the boundary layer mixing-length estimate shown in Fig. 5 (ml_emod and longmel) significantly influence ensemble behavior. This quantity modulates turbulent mixing-induced tendencies, which are computed via the $K$-theory diffusion analog as follows:

$$\frac{\partial \xi}{\partial t} = \frac{\partial}{\partial z} \left[ -K_x \frac{\partial \xi}{\partial z} - \Gamma \right],$$

(9)

for $x \in \{T, q, u, v\}$, where $z$ is height and $\Gamma$ depicts nongradient transport. In the adopted 1.5-order closure, the turbulent diffusion coefficient $K_x$ is estimated as follows:

$$K_x \propto \lambda \sqrt{E},$$

(10)

where $E$ is the turbulence kinetic energy and $\lambda$ is the mixing length, a height scale associated with Prandtl’s 1925 theory (Bradshaw 1974). Because of uncertainty in the estimate of $\lambda$,
three distinct formulations are available within the model's parameterization suite (Table 3).

The longmel SPP element makes direct use of the multiple $k$ estimates to sample uncertainty by adopting different closures through discrete perturbations (section 3b). Despite this introduction of spatiotemporal algorithmic variability, the longmel SPP element has limited impact on ensemble spread within the boundary layer (Fig. 15a) and leads to a bias-

![Graphs and diagrams](image)

**Fig. 14.** Oceanic (a) screen-level temperature and (b) sensible heat flux responses to exchange coefficient perturbations in the fh_mult ensemble averaged over the North Atlantic region (Figs. 13b,d). Pairwise differences are computed at 6-h intervals between the fh_mult and control ensemble members and plotted using dots that are color coded with the value of the quantity in the fh_mult ensemble as shown on the color bars. The sensible heat flux shown in (b) is computed from Eq. (7) before perturbations are applied. A kernel density estimate of the scatterplot is shown in solid black lines for reference. The best-fit linear model for perturbations over the range [0.8–1.2] (bounds highlighted with vertical green dashed lines) is shown with the dashed red line, and the average temperature departure for 10 perturbation percentile bins is shown with the blue line and blue points at the bin centroids. The perturbation origin ($\mathcal{F} = 1$ and zero response) is indicated with dashed gray lines. The response time of these surface layer properties (min) is short relative to the perturbation time scale ($\tau = 36$ h). As a result, these relationships are robust throughout the 5-day integration, with the perturbation-induced divergence of solutions leading only to increased dispersion of the points (not shown).

![Graphs and diagrams](image)

**Fig. 15.** Sensitivity of (a),(c) 850-hPa temperature spread ('C) and (b),(d) fCRPS ('C) for the (left) longmel and (right) ml_emod ensembles. Plotting follows the conventions employed in Fig. 12, except that regions in which surface pressure is less than 850 hPa are masked with light-green shading.
induced deterioration in the fCRPS at higher latitudes (Fig. 15b). Improved fCRPS values in the stratocumulus regions also originate from bias changes (not shown), suggesting that the Bougeault and Lacarrère (1989) $\lambda$ introduced in this configuration may yield locally improved basic states. Overall, however, the longmel SPP element leads to significant deteriorations that preclude its adoption in the GEPS configuration (Fig. 5).

FIG. 16. Mixing-length properties (m) at model hybrid level ~0.92 (~500 m above the surface) averaged over pentad 1. Shown are (a) the mean mixing length in the control ensemble, (b) the mean difference between the ml_emod and control ensemble (computed as ml_emod minus control), and (c) the difference in ensemble standard deviation [computed as for (b)]. Values are color coded according to the color bars at the bottom of each panel, with dark shading in (b) and (c) indicative of difference that are statistically significant for a false discovery rate of 5% based on a 1000-member bootstrap. The North Pacific domain used in Fig. 17 is outlined in black in (b) and (c).

FIG. 17. Properties of mixing-length estimates at hybrid level ~0.92 (~500 m above the surface) over the North Pacific basin (outlined in black in Figs. 16b,c) for a control simulation initialized at 0000 UTC 19 Jan 2020 with perturbed mixing-length diagnostics obtained by piggyback calculations (Grabowski 2014): (a) the cumulative frequency distributions of the Blackadar (1962) (black), Bougeault and Lacarrère (1989) (blue), and perturbed (red) mixing lengths, with the unphysical region $\lambda < 0$ gray shaded, and (b) the relationship between the two estimates, with green dots for individual points and a kernel density estimate in solid black lines (powers of 2 used for contour intervals; “e$^{-0X}$” indicates multiplication by 10 raised to the $-X$ power). The 1:1 line appears in gray dashes for reference. The regions in which large absolute values of $\lambda$ could lead to a negative perturbed mixing length for the $[-2.5, 2.5]$ range used in the ml_emod ensemble are computed via Eq. (4) and are shown with gray shading, and the equivalent for $[-0.5, 0.5]$ is shown with a thin red dashed line. These plots are computed from data at 6-hourly intervals that are spatially subsampled at 10-gridpoint intervals to limit data volume and promote independence.
Table 5. Globally averaged pentad-1 fractional contribution [Eq. (2), in percentage as $100 \times C^i_j$] to the ensemble spread and fCRPS by the ml_emod SPP element and a sensitivity test using the “height restricted” form of the error model described in the text. Differences between the SPP-based experiments and the control that are statistically significant at the 99% level according to a 1000-member bootstrap test are shown in boldface type.

<table>
<thead>
<tr>
<th>State variable</th>
<th>Spread contribution (%)</th>
<th>fCRPS contribution (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ml_emod</td>
<td>Height-restricted</td>
</tr>
<tr>
<td>250-hPa wind</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>500-hPa height</td>
<td>4.5</td>
<td>4.5</td>
</tr>
<tr>
<td>850-hPa temperature</td>
<td>4.4</td>
<td>4.5</td>
</tr>
<tr>
<td>850-hPa humidity</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Screen-level temperature</td>
<td>5.3</td>
<td>5.2</td>
</tr>
<tr>
<td>Sea level pressure</td>
<td>6.6</td>
<td>6.7</td>
</tr>
</tbody>
</table>

The mixing-length error model employed by the ml_emod SPP element proves to be more successful in sampling uncertainty in $\lambda$ without introducing systematic errors (Figs. 15c,d). The Blackadar (1962) and Bougeault and Lacarrère (1989) estimates are used in Eq. (4) as $\chi_i$ and $X_{alt}$, respectively, for $\mathcal{F} = [-2.5, 2.5]$. This symmetric range is intended to center $\lambda$ perturbations around the control; however, $\lambda$ values are larger in ml_emod than in the control ensemble, particularly in regions with increased $\lambda$ spread (Fig. 16). Results from a representative case study show that the perturbation-broadened distribution of $\lambda$ (shallower slope of the cumulative distribution in Fig. 17a) includes negative values. These are the result of large differences between $\lambda$ estimates, which occur primarily under stable conditions in which $\lambda \approx 200$ m (Fig. 17b). The application of a standard limit of $\lambda \geq 1$ m prevents numerical errors and yields near-zero diffusion coefficients consistent with a very stable environment [Eq. (10)]. As a result, there is physical symmetry in the turbulent-transport response to the perturbations despite truncation of the small-$\lambda$ end of the distribution. This outcome combines with the positive impact of the ml_emod SPP element (Fig. 5) to justify its inclusion in the GEPS configuration. However, care will have to be taken in the design of future error models to mitigate such behavior in cases in which imposing a lower bound is not appropriate.

The use of multiple $\lambda$ estimates has the disadvantage of incurring additional computational cost, with model run time increased by 28%. However, restricting the Bougeault and Lacarrère (1989) calculations to lower levels ($\leq 2500$ m AGL) reduces the added cost to just 2% while yielding qualitatively identical results (Table 5). Although this configuration was identified too late in the development cycle for implementation in the 2021 GEPS upgrade, it suggests that SPP error models may benefit from having multiple “cheap” inputs rather than a very limited number of computationally intensive estimates.

2 Limiting the range of $\mathcal{F}$ to $[-0.5, 0.5]$ to reduce the potential for negative values of perturbed $\lambda$ (the red dashed line in Fig. 17b) eliminates the mean change in $\lambda$; however, it also renders the error model ineffective at generating ensemble diversity and is therefore rejected as a candidate configuration (not shown).

c. Advection error model

Although the large majority of SPP elements focus on components of the physical parameterization suite, uncertainties that exist within the dynamical core can have significant impacts on ensemble statistics (Fig. 10). The net effect of these potential error sources is represented in this implementation by the adv_rhsint SPP element, which uses an error model to sample uncertainty in the advection scheme. The semi-Lagrangian formulation begins by determining the departure point of the trajectory that arrives at each grid point over a time step. The advected quantities must then be interpolated to the departure point to give the upstream value: cubic polynomials are typically used for this operation in GEM (Fig. 18). The adv_rhsint error model samples uncertainties in these calculations by using cubic and linear interpolation estimates as $\chi$ and $X_{alt}$ in Eq. (4) with $\mathcal{F} = [-0.4, 0.4]$.

This error model generates perturbations where local gradients are highly spatially variable, such that the first- and third-order interpolation estimates are likely to differ (Fig. 18a). In such regions, the effects of discretization combine with uncertainties associated with the interpolants themselves to yield a potentially important error source. Conversely, the range of values computed using the different polynomials will be relatively small where the advected quantities vary smoothly in space (Fig. 18b). Unlike the mixing length estimates described above, this error model incurs no additional computational cost because the linearly interpolated value is precomputed within the advection scheme.

Although the adv_rhsint SPP element reduces fCRPS in all regions (Fig. 10), its impact on ensemble spread is primarily tropical (Fig. 19a) because a latitudinal tapering is applied to $\mathcal{F}$:

$$\mathcal{F}' = \begin{cases} \cos^2 \left( \frac{\phi}{\phi_o} \right) & \text{if } |\phi| < \phi_o \\ 0 & \text{otherwise} \end{cases}$$

(11)

where $\mathcal{F}'$ is used in Eq. (4) and $\phi_o = 45^\circ$ in adv_rhsint. This heuristic is consistent with enhanced dynamical core sensitivity in the tropics (Subramanian et al. 2019) and prevents the direct injection of perturbations along the midlatitude waveguides (Hoskins et al. 1985), in which both spread and error
grow rapidly if it is not applied (Figs. 19c,d). This suggests that although adv_rhsint can effectively identify regions with enhanced uncertainty in adiabatic dynamics, it induces perturbations that exceed the associated errors. Refinement of SPP within the dynamical core will require the identification of additional algorithmic uncertainties, such that smaller-amplitude perturbations can be introduced in closer connection to error sources.

Despite the restriction of adv_rhsint perturbations to tropical regions [Eq. (11)], 500-hPa height forecasts improve at high latitudes in the North Pacific basin (Fig. 19b). This sensitivity originates primarily within the East Asian subtropical jet (Yang et al. 2002; Luo and Zhang 2015) in the form of enhanced ensemble spread (Figs. 19a and 20a). The perturbations appear to promote diversity in the initiation of Rossby wave packets (Rothlisberger et al. 2018), which propagate downstream at rates that fall between the expected phase speed ($c_p \approx 10 \text{ m s}^{-1}$) and group velocity ($c_g \approx 20 \text{ m s}^{-1}$; Fragkoulidis and Wirth 2020). The associated fCRPS improvements (Fig. 20b) suggest that the perturbations are effectively identifying uncertainty in this tropical–extratropical interaction and the resulting Rossby wave evolution (Wirth et al. 2018). Although perturbations deeper within the tropics are sometimes associated with Rossby wave initiation (Stan et al. 2017), an experiment in which $f_{\text{off}} = 30$ [Eq. (11)] shows that this indirect mode of communication is relatively inefficient (Figs. 20c,d). The ability of adv_rhsint to represent the remote impact of tropical uncertainties, thereby reducing the fCRPS in all domains (Fig. 10), makes it an important component of the SPP configuration for the GEPS.

6. Discussion

The SPP scheme implemented in GEM employs spatiotemporally correlated random fields as the basis for perturbations to free parameters, physical closures and the dynamical core. These perturbations represent inherent uncertainties in the model in an internally consistent way, yielding a diversity of solutions under conditions in which the associated errors adversely affect practical predictability.

The impacts of 27 SPP elements have been evaluated for potential inclusion in the GEPS. Together, the proposed configuration generates sufficient ensemble spread without increasing RMSE$_u$. Individually, they trigger a broad range of responses in terms of contributions to ensemble spread and forecast skill. The spatial distributions of these sensitivities are closely related to the processes represented by the SPP elements themselves.

The perturbation-process connection is a conceptual strength of SPP; however, it introduces a high level of complexity into the system. The large number of degrees of freedom associated with the SPP technique makes identifying the optimal ensemble configuration a computationally intensive task. The extensive set of perturbations discussed here represents a small fraction of the potential error sources in the model. As more SPP elements are introduced, the amplitudes of existing perturbations will need to be revised to represent better the underlying uncertainty. Additional efforts are also needed to characterize...
sensitivity to different spatiotemporal scales in the perturbation patterns (Lin and Neelin 2000). Preliminary results suggest that the ensemble is highly sensitive to shortening these scales toward those associated with the processes themselves (Jankov et al. 2019; Kalina et al. 2021); however, Bessac et al. (2019) show that spatial error correlation scales can approach those employed here. The results of these analyses will be presented in a future study.

The formulation of SPP elements introduces an additional dimension to the problem of accurate model error representation. Parameter ranges and multiplicative scalings are simple descriptions of uncertainty that compose the majority of implemented SPP elements (Table 3). However, nonlinear responses can lead to systematic changes in perturbed members. Reducing perturbation ranges can mitigate associated biases; however, this also limits the SPP element’s overall effectiveness. The error models designed for select SPP elements appears to be able to represent uncertainty effectively when multiple estimates of a quantity are available. Further efforts are needed to assess the applicability of error models to other elements and to determine whether sampling from a larger range of computationally inexpensive estimates provides a more robust representation of uncertainty [Eq. (3)] than the limited approach currently adopted [Eq. (4)].

As the number of SPP elements increases, so too does the range of possible sensitivities. While some of these might be well represented by traditional summary statistics, others will not. Perturbations to the critical Richardson number and to the rms of gravity wave spectrum have negligible impacts on tropospheric fCRPS, but they, respectively, affect high impact freezing rain forecasts and assimilation-relevant stratospheric wind spread. Future assessments of the utility of SPP elements will therefore need to account for the full range of potentially useful sensitivities within the ensemble.

Extensions to the proposed SPP technique may lead to further improvements to the system. The atmosphere-only configuration of the GEPS used here permitted only a rudimentary representation of uncertainties at the surface interface through perturbed exchange coefficients in the boundary layer scheme. While this provides useful guidance about the magnitude of surface flux uncertainties, work within the fully coupled system is needed to make consistent perturbations across model components. Within the atmosphere itself, the current use of 2D perturbing fields implies perfect vertical uncertainty correlation, a simplification that is unlikely to be valid for processes such as boundary layer turbulence that are marked by sharp vertical transitions. The efficiency of the field generator makes it possible to create 3D perturbations; however, the impact of this generalization has not yet been assessed.

Although the specific SPP elements described here are unique to GEM, the SPP technique itself is applicable to any model. Moreover, the scheme’s ability to generate a physically realistic range of ensemble member solutions suggests that it may represent a viable alternative to existing estimates of model uncertainty. Part II of this investigation will therefore focus on an intercomparison between differing model uncertainty representations (multiphysics, stochastic tendency...
perturbation and SPP), an assessment that led to the autumn-2021 implementation of the SPP scheme in the operational GEPS.

Acknowledgments. The authors thank Dr. Carolyn Reynolds for productive discussions and suggestions on an earlier draft of this paper and Dr. Justin McLay for his advice on the ensemble sensitivity formalism. We also acknowledge the contribution of Dr. Xingxiu Deng in the form of help with GEPS configuration. Doctor Hannah Christensen and two anonymous reviewers provided excellent suggestions that improved the paper, as did Dr. Tommaso Benacchio as the handling editor.

Data availability statement. Given the large volume of model outputs analyzed in this study, transfer to an independent repository is impractical; however, all data used in this study will be made freely available upon request. The full ERA Interim reanalysis dataset is archived by Copernicus Climate Services (https://climate.copernicus.eu/). The GEM model, including the implementation of SPP described here, is available online (https://github.com/ECCC-ASTD-MRD/gem).

APPENDIX

The Stochastic Field Generator

All stochastic representations of model uncertainty in GEM use a common stochastic field generator. The factory described here flexibly and efficiently produces a large number of independent fields \( \mathcal{F}(\Lambda, \phi, t) \) that each have their own temporal and spatial correlation properties (\( \Lambda, \phi \) and \( t \) refer to the
longitude, latitude and time dimensions, respectively). Individual instances of \( F \) are used to control the evolution of each of the SPP elements described in section 3 of the main text.

a. Double Markov chains for smooth temporal evolution

Time series governed by a pair of first-order autoregressive (Markov) processes are used as spectral coefficients of an expansion on spherical harmonics. A 2D random function on the sphere, correlated in space and time, with a PDF symmetric around the mean \( \mu \) is defined as follows:

\[
f(\Lambda, \phi, t) = \mu + \sum_{l=L_{\text{min}}}^{L_{\text{max}}} \sum_{m=-l}^{l} a_{l,m}(t) Y_{l,m}(\Lambda, \phi), \tag{A1}
\]

where the parameters \( L_{\text{min}} \) and \( L_{\text{max}} \) define the spectral range of the random function: their inverses can be interpreted in terms of decorrelation length scales. The \( Y_{l,m} \) functions are spherical harmonics (\( l \) is the total horizontal wavenumber, and \( m \) is the zonal wavenumber), which are normalized by

\[
\int_0^{2\pi} d\Lambda \int_{-1}^{1} d(\sin \phi) Y_{l,m}(\Lambda, \phi) Y_{l,m}^*(\Lambda, \phi) = \delta_{l,l'} \delta_{m,m'}, \tag{A2}
\]

where \( \delta \) represents the Kronecker delta function.

The spectral coefficients in Eq. (A1) are defined as follows:

\[
a_{l,m}(t + \Delta t) = e^{-\Delta t \gamma} a_{l,m}(t) + \frac{(1 - e^{-2 \Delta t \gamma})}{\sqrt{(1 + e^{-2 \Delta t \gamma})}} b_{l,m}(t + \Delta t), \tag{A3}
\]

where \( \gamma \) is the single-chain decorrelation time scale, \( \Delta t \) is the time step of the numerical model, and

\[
b_{l,m}(t + \Delta t) = e^{-\Delta t \gamma} b_{l,m}(t) + \sqrt{\frac{4 \pi \sigma^2 S(t)(1 - e^{-2 \Delta t \gamma})}{2l + 1}} R_{l,m}(t), \tag{A4}
\]

is a second autoregressive process. The complex \( R_{l,m} \)s are uncorrelated random processes with mean zero and variance \( \langle |R_{l,m}|^2 \rangle \) of unity, where the angle brackets denote a statistical mean. The global mean of the random function \( f \) is \( \mu \) when \( L_{\text{min}} > 0 \), and its prescribed variance is \( \sigma^2 \). For the real random fields used here,

\[
a_{l,m} = (-1)^m a_{l,-m}^* \quad \text{and} \quad b_{l,m} = (-1)^m b_{l,-m}^* \tag{A5}
\]

must apply, where the asterisk denotes the complex conjugate.

The function \( S(l) \) in Eq. (A4) is proportional to the prescribed mean power spectrum of \( f \), which satisfies the following condition:

\[
\sum_{l=L_{\text{min}}}^{L_{\text{max}}} S(l) = 1. \tag{A6}
\]

The statistical properties of \( f \) are homogeneous and isotropic, implying that \( S(l) \) is independent of \( m \). In this implementation, the \( R_{l,m} \) functions are Gaussian processes. The choice \( S(l) = \text{constant} \) therefore generates a white-noise signal in space between \( L_{\text{min}} \) and \( L_{\text{max}} \).

The use of two Markovian processes in the construction of the spectral coefficients [Eqs. (A3) and (A4)] renders the temporal evolution of \( f \) smooth in the first derivative. However, correlation between \( a_{l,m} \) and \( b_{l,m} \) extends the decorrelation time scale to \( \tau = 2.146 \tau \) (\( \Delta t \ll \tau \)) and leads to a reddening of the power spectrum in the double Markov chain formulation.

To keep the perturbing field within specified bounds and to permit flexibility in defining the shape of the fields PDF, \( f \) is stretched as follows:

\[
F = \left( \frac{F_{\text{max}} - F_{\text{min}}}{2} \right) \text{erf} \left[ \frac{f - \mu}{\gamma \sigma \sqrt{2}} \right] + \frac{F_{\text{min}} + F_{\text{max}}}{2}, \tag{A7}
\]

where the error function is used to ensure that \( F \) is uniform and insensitive to \( \sigma \) if \( f \) is normally distributed and \( \gamma = 1 \). The latter adjusts the shape of the PDF of \( F \), which is uniformly distributed when \( \gamma = 1 \) and increasingly peaked at \( \mu \) as \( \gamma \) rises (Fig. A1).

---

**Table A1.** Global variance of stochastic fields \( F \) and \( F \) for configurations using different intermediate grids. These representative values are computed as averages of the variances of the eight independent SPP elements with ranges of [0.5, 1.5] and \( \gamma = 1 \) (Table 3) for the forecast initialized at 0000 UTC 1 Jan 2020. The variance of \( F \) is computed as the average of variances on the yin and yang native model domains.

<table>
<thead>
<tr>
<th>Intermediate grid</th>
<th>Variance of ( F )</th>
<th>Variance of ( F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 × 8</td>
<td>0.077</td>
<td>0.045</td>
</tr>
<tr>
<td>64 × 32</td>
<td>0.076</td>
<td>0.075</td>
</tr>
<tr>
<td>384 × 192</td>
<td>0.076</td>
<td>0.079</td>
</tr>
</tbody>
</table>
b. Stochastic fields on the native model grid

The stochastic field $\mathcal{F}$ is computed directly from results of the inverse Fourier transform of the spectral coefficients $f$ using Eq. (A7). It is therefore defined on a global domain that serves as the target for the transformation. The values of $\mathcal{F}$ on the native yin–yang model grid are computed as follows:

$$\mathcal{F} = \mathcal{L}(\mathcal{F}),$$ (A8)

where $\mathcal{L}$ is the bilinear spatial interpolation operator.

The stochastic pattern generator used to produce $\mathcal{F}$ for all SPP elements in this study was configured to use a low-resolution (global 16 × 8 points) latitude–longitude intermediate grid for $\mathcal{F}$. Although the resolution of this grid is sufficient to represent the low-wavenumber range of the spherical harmonics (Table 2), the application of $\mathcal{L}$ leads to significant smoothing because of the low-resolution interpolation source. As shown in Table A1, the variance of $\mathcal{F}$ is reduced by almost 40% with respect to both $\mathcal{F}$ and the equivalent value computed using a higher resolution (384 × 192) intermediate grid.

The application of $\mathcal{L}$ acts as a filter with a poor response function, damping variance near the wavenumber-8 cutoff as shown in Fig. A2. Interpolation from the 16 × 8 intermediate grid also leads to the generation of persistently

![Stochastic Field Variance Spectrum](image)

**Fig. A2.** Variance spectra of $\mathcal{F}$ for 16 × 8 (red), 64 × 32 (blue), and 384 × 192 (black) intermediate global grids. Spectra are computed after cubic interpolation of the native yin–yang outputs to an 800 × 600 Gaussian grid for the same eight independent SPP elements used for the variance calculations shown in Table A1. To increase the robustness of the spectral analysis, spectra are computed and averaged from the 6-, 60-, and 120-h lead times from the 0000 UTC 1 Jan 2020 initialization of the $M = 10$ ensemble, such that a total of 240 individual spectra contribute to each of the results shown here. The shaded area around each line represents the width of the 95% confidence interval for the mean.

![Global 16x8](image)

**Global 16x8**

![Global 384x192](image)

**Global 384x192**

**Fig. A3.** Representative stochastic fields $\mathcal{F}$ generated using (a) 16 × 8 and (b) 384 × 192 intermediate grids, plotted as in Fig. 1a but for the range [0.5, 1.5] consistent with the SPP elements used in this analysis.
Note that $R_m(0)$ takes a different random complex value for each spectral element. Note also that $R_m^R(0) = 0$, which comes from the requirement that $a_{lm}(0)$ and $b_{lm}(0)$ be real [Eq. (A5)]. The normalization factor $\beta_m$ is derived from the requirement that the random variable $R_m$ have unit variance.

To ensure the smooth evolution of perturbations, Markov chains persist within the data assimilation cycle. For all simulations discussed here, spectral coefficients from the appropriate trial integration are therefore re-ingested by the model instead of being initialized at the beginning of the integration.

**REFERENCES**


