ORIGINAL ARTICLE

Constraining stochastic parametrisation schemes using high-resolution simulations

H. M. Christensen[1]

[1]University of Oxford

Correspondence
Hannah Christensen, Clarendon Laboratory, Parks Road, Oxford, OX1 3PU, UK.
Email: hannah.christensen@physics.ox.ac.uk

Funding information
European Research Council grant number 291406 and Natural Environment Research Council grant number NE/P018238/1

Stochastic parametrisations can be used in weather and climate

Abbreviations:

models to improve the representation of unpredictable unresolved processes. When compared to a deterministic model, a stochastic model represents 'model uncertainty', i.e., sources of error in the forecast due to the limitations of the forecast model. A technique is presented for systematically deriving new stochastic parametrisations or for constraining existing stochastic approaches. A high-resolution model simulation is coarse-grained to the desired forecast model resolution. This provides the initial conditions and forcing data needed to drive a Single Column Model (SCM). Comparing the SCM parametrised tendencies with the evolution of the high resolution model provides an estimate of the error in the SCM tendencies that a stochastic parametrisation seeks to represent. This approach is used to assess the physical basis of the widely used Stochastically Perturbed Parametrisation Tendencies (SPPT) scheme. Justification is found for the multiplicative nature of SPPT, and some evidence for the use of spatio-temporally correlated stochastic perturbations. Evidence that the stochastic perturbation should be positively skewed is found, indicating that occasional large-magnitude positive perturbations are physically realistic. However other key assumptions of SPPT are less well justified, including coherency of the stochastic perturbations with height, coherency of the perturbations for different physical parametrisation schemes, and coherency for different prognostic variables. Relaxing these SPPT assumptions allows for an error model that explains a larger fractional variance than traditional SPPT. In particular, it is suggested that independently perturbing the tendencies associated with different parametrisation schemes is justifiable, and would improve the realism of the SPPT approach.

Keywords stochastic parametrisation; coarse-graining; Stochastically Perturbed Parametrisation Tendencies (SPPT); model uncertainty; ensemble weather prediction; climate modelling

1 | INTRODUCTION

Weather and climate projections, spanning timescales from a few days to many decades, are routinely presented in a probabilistic manner (Houghton et al., 1990; Buizza, 2017). Such probabilistic predictions are required to support decision-making. This

enables preventative action to mitigate the impacts of extreme weather events or future climate change. However, overconfident predictions can lead users to make decisions that are costly, while the benefits of those decisions are never realised (Murphy, 1977). To be useful, a forecast must be reliable — it must accurately represent the likelihood of the forecast event — and so must represent all sources of uncertainty in the forecast (Stensrud et al., 2000).

A major source of error in both weather and climate prediction is the approximations made when developing the forecast model (Hawkins and Sutton, 2009). In particular, limited computer resources lead to the simplified representation of unresolved small-scale processes, such as convective clouds and turbulent transport, through parametrisation schemes. There is much debate as to the optimal representation of this *model uncertainty*, and several methods have been proposed (Stainforth et al., 2005; Bowler et al., 2008; Rougier et al., 2009; Kirtman et al., 2014; Ollinaho et al., 2017; Leutbecher et al., 2017). An attractive solution, initially proposed for use in weather forecasts, is the use of stochastic parametrisations. Here, atmospheric processes are represented as a combination of a predictable deterministic and an unpredictable stochastic component. Stochastic parametrisations have revolutionised probabilistic weather prediction (Palmer et al., 2009). They are now ubiquitous in operational forecasting centres worldwide (e.g. Sanchez et al., 2016), and have also been widely adopted by seasonal forecasting systems (e.g. Charron et al., 2010). They have been shown to outperform other representations of model uncertainty on weather and seasonal timescales (Christensen et al., 2015; Weisheimer et al., 2011). Recent work has considered the impact of stochastic parametrisations in climate models, where they have been found to substantially improve both mean state and variability (Wang et al., 2016; Christensen et al., 2017a; Berner et al., 2017; Davini et al., 2017; Strømmen et al., 2018). Note that stochastic parametrisation schemes include both probabilistic representations of a specific sub-grid process, and stochastic approaches to characterise uncertainty in otherwise deterministic models.

As summarised in Figure 1, the starting point for all stochastic parametrisation schemes should be identifying a source of uncertainty in a forecast model at a given resolution. Examples include the error in the representation of a specific process such as sub grid-scale turbulent mixing (Nie and Kuang, 2012; Gentine et al., 2013; Sušelj et al., 2014), convective entrainment (Romps and Kuang, 2010), variability in air-sea fluxes (Bessac et al., 2019), or the error in a collection of processes, such as uncertainty in the net parametrised physics tendency (Buizza et al., 1999). Having identified the model error that leads to uncertainty in the forecast, the characteristics of that model error must be predicted through theory, or assessed through measurements. This allows for a statistical representation of that error to be included into the forecast model.

Several schemes have been proposed whereby the characteristics of model error are predicted through theoretical understanding. For example, Plant and Craig (2008) propose a stochastic convection parametrisation based on the theory of Craig and Cohen (2006). Khouider et al. (2010) propose a stochastic multicloud model based on an understanding of tropical convection, with the transition rates between cloud types set by rules of thumb. Ollinaho et al. (2017) propose to stochastically vary 20 parameters in the European Centre for Medium-range Weather Forecasts (ECMWF) model, where the uncertain parameters were identified by experts and the optimal magnitude of the perturbations was tuned to maximise forecast skill. However, even with a physical foundation, these approaches tend to contain one or more parameters that must be estimated or tuned. To fully characterise model error and constrain these variables, we can augment such theoretical ideas with measurements. For example, Ollinaho et al. (2013) describe the use of a Bayesian parameter estimation framework to quantify the uncertainty in four parameters within the convection parametrisation scheme, which was used by Christensen et al. (2015) to develop a well-constrained stochastically perturbed parameter scheme. Dorrestijn et al. (2015) use observational data to estimate the transition probabilities between cloud types to underpin the stochastic multicloud approach of Khouider et al. (2010).

Recent years have seen a surge in the production of very high resolution atmospheric simulations. Continued increase in computational power has led to an increase in domain size and duration of simulations, with resolutions regularly reaching convection permitting, if not convection resolving, scales (Holloway et al., 2012; Satoh et al., 2014; Schalkwijk et al., 2015; Heinze et al., 2017; Satoh et al., 2017; Stevens et al., 2019). The availability of such datasets opens up the option of using high-resolution simulations as a proxy for the 'true atmosphere', and identifying the difference between a low resolution forecast

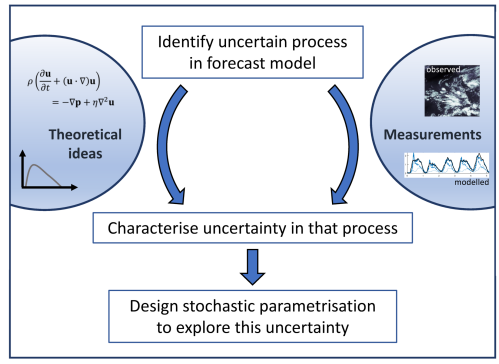

FIGURE 1 Schematic illustrating the process of developing stochastic parametrisation schemes.

model and a high-resolution simulation as the model error that a stochastic parametrisation seeks to represent (Shutts and Palmer, 2007; Shutts and Pallares, 2014). This allows for the derivation of data-driven stochastic representations of this sub-grid variability (Dorrestijn et al., 2013; Porta Mana and Zanna, 2014; Cooper and Zanna, 2015; Bessac et al., 2019). To date, this approach has been used to construct stochastic parametrisation schemes that are independent from (i.e. an alternative to) existing deterministic parametrisations.

In this study we revisit the use of very high resolution datasets to evaluate and constrain stochastic parametrisation schemes. Instead of deriving an independent stochastic scheme, we seek to characterise the error in existing deterministic parametrisations. We then assess whether the deterministic schemes could be augmented by a stochastic term to represent uncertainty in the parametrised tendencies. We follow the technique proposed in Christensen et al. (2018b), whereby existing high-resolution simulations are coarse-grained to provide the input for a low-resolution single column model (SCM), which provides the parametrised tendencies. As a case study, we use this coarse graining methodology to assess the widely used 'Stochastically Perturbed Parametrisation Tendencies' (SPPT) scheme. While the SPPT scheme was proposed based on theoretical ideas, several of the foundational assumptions of the scheme have never been assessed. This study will probe the validity of the underlying theory.

We present the SPPT scheme in Section 2. In section 3 we review the coarse-graining methodology presented in Christensen et al. (2018b), and highlight adaptations to the methodology necessary for this study. In section 4 we begin to assess the main assumptions underlying SPPT, and highlight successes of the scheme. In section 5 we demonstrate that other assumptions made in SPPT are not justifiable. In section 6 we discuss the results, and conclude in section 7 by recommending simple changes to the SPPT approach.

2 | CHOSEN APPLICATION: THE STOCHASTICALLY PERTURBED PARAMETRISATION TENDENCIES SCHEME

The stochastically perturbed parametrisation tendencies (SPPT) scheme is an attractive approach due to its ease of use and beneficial impact on ensemble forecast reliability (Palmer et al., 2009; Weisheimer et al., 2014). It runs in conjunction with operational physics parametrisation schemes, so can be easily adapted for use in different models. It is widely used in weather and seasonal forecasting centres worldwide, including at the European Centre for Medium-range Weather Forecasts (ECMWF) (Buizza et al., 1999; Palmer et al., 2009), the U.K. Met Office (Sanchez et al., 2016), the Japan Meteorological Agency (Yonehara

and Ujiie, 2011), in the Application of Research to Operations at Mesoscale (AROME) model (Bouttier et al., 2012), the Weather Research and Forecasting (WRF) model (Berner et al., 2015), and in the Community Earth System Model (Christensen et al., 2017a) and EC-Earth (Davini et al., 2017; Strømmen et al., 2019) climate models. While the essence of the scheme is the same across models, there are differences in precise implementation. The following discussion presents the details of the scheme as implemented at ECMWF (Palmer et al., 2009; Leutbecher et al., 2017).

The SPPT scheme addresses model uncertainty due to the parametrisation process. It does this by perturbing the sum of the parametrised physics tendencies using multiplicative noise:

$$\begin{aligned} \mathbf{T}_X &:= \frac{\partial X}{\partial t} \\ &= \mathbf{D}_X + (1+e)\sum_{i=1}^{I} \mathbf{P}_{i,X} \\ &= \mathbf{D}_X + (1+e)\mathbf{P}_X . \end{aligned} \tag{1}$$

where $\mathbf{T}_X$ is the total vector tendency in X, as a function of model level at a particular spatial grid point. $\mathbf{D}_X$ is the vector tendency from the dynamics, $\mathbf{P}_{i,X}$ is the vector tendency from the ith physics scheme, and e is a zero mean random perturbation. There are $I = 5$ key physics schemes in the IFS: radiation (RDTN); turbulence and orographic gravity wave drag (TGWD); non-orographic gravity wave drag (NOGW); convection (CONV); large-scale water processes (LSWP). The perturbation e is constant in the vertical though it is tapered in the boundary layer and stratosphere. The tapering in the boundary layer is to avoid exciting numerical instabilities in the model, while in the stratosphere, uncertainty in the parametrised tendencies is believed to be small due to the lack of uncertain moist processes.

The SPPT scheme perturbs the tendency for the four prognostic variables in the IFS: X = temperature (T), zonal and meridional wind speed (U and V respectively), and humidity (q). Each variable tendency is perturbed using the same random number field. The perturbation field is generated using a spectral pattern generator. The pattern at each time step is the sum of three independent random fields with horizontal correlation scales of 500, 1000 and 2000 km. These fields are evolved in time using a first order autoregressive (AR(1)) process with time scales of 6 hours, 3 days and 30 days respectively. The fields have standard deviations of 0.52, 0.18 and 0.06 respectively. It is expected that the smallest scale (500 km and 6 hours) will dominate at a 10 day lead time, while the larger scale perturbations are important for monthly and seasonal forecasts.

Underpinning SPPT are several theoretical statements (Buizza et al., 1999):

1. The larger the parametrised tendencies, the larger the potential random error. This is based on the concept that random model error arises due to unresolved organisation of sub-grid processes. A higher degree of sub-grid organisation will increase the mean sub-grid tendency, and also increase the variability.
2. Random error due to the parametrisation process will be coherent between the different physics parametrisation schemes. This is such that the balance between tendencies associated with different physics schemes is retained.
3. Random error due to the parametrisation process will be coherent between the parametrised tendencies for different prognostic variables (T, U, V, q). This ensures physical consistency in the model.
4. The random error is coherent across large spatial and temporal scales, due to the source of the random error being the lack of sub-grid organisation in the forecast model. Furthermore, the truncation of the model equations of motion is expected to introduce errors on both larger and smaller scales than the truncation scale, introducing correlations into the random error.

Some evidence supporting the first statement has been provided by past coarse graining studies. Shutts and Palmer (2007) defined an idealised cloud resolving model (CRM) simulation as truth. The high-resolution fields and their tendencies were

coarse grained to the resolution of a NWP model to study the sub-grid scale variability which a stochastic parametrisation seeks to represent. The 'true' convective heating on the coarse grid was calculated by averaging the convective heating on the fine grid. This was compared to the heating calculated from a convection parametrisation scheme on the coarse grid. The validity of the multiplicative noise in the SPPT scheme was analysed by studying histograms of the coarse grained 'true' heating conditioned on different ranges of the parametrised heating. The mean and standard deviation of the true heating were observed to increase as an approximately linear function of the parametrised heating, providing some support for the SPPT scheme. However, Shutts and Palmer (2007) only consider a single level in the atmosphere, and do not test the validity of the approach at other levels. Shutts and Palmer (2007) also focus on relatively large coarse-graining scales of 80–320 km, substantially coarser than the resolution of modern NWP models.

Despite some evidence in support of the first statement, coarse graining studies have indicated that the second underpinning statement may be less valid. A different coarse graining study by Shutts and Pallares (2014) estimated the standard deviation of the error for each physics tendency as a function of the parametrised tendency. The tendencies from the IFS at T1279 (16km) were defined to be "truth", and were compared to forecast tendencies from the IFS at T159 (130km). The study revealed that the different schemes have different error characteristics, with the uncertainty in the cloud and convection tendencies being much larger than the radiation tendency. Shutts and Pallares (2014) also found that the standard deviations of the cloud and convection tendencies were a non-linear function of the parametrised tendency. However, the T1279 simulation used as 'truth' is relatively low resolution, and includes parametrised convection. It is not clear how this would impact the analysis.

The second statement above also assumes that the errors from each physics parametrisation scheme are perfectly correlated — one random number field is used to perturb all schemes. This has not been assessed. It is unlikely that uncertainties in the different processes are precisely correlated, as modelled by SPPT. An alternative is to use independent random fields for each physics tendency. This "independent SPPT" (iSPPT) was considered by Christensen et al. (2017b), and resulted in a significant improvement in the reliability of ensemble forecasts, particularly in regions with significant convective activity.

Regarding the third statement, the original implementation of SPPT perturbed different prognostic variable tendencies with different random numbers (Buizza et al., 1999). The move to using a single pattern to perturb all prognostic variable tendencies was proposed later to ensure physical consistency (Palmer et al., 2009), for example, accounting for the relationship between temperature and humidity tendencies for thermodynamic processes. Moving to a single pattern was found to benefit (reduce) the frequency of strong precipitation events in the forecast (Palmer et al., 2009). Nevertheless, the validity of the third statement has not been tested.

The validity of the fourth statement has also not been tested. It is well known that stochasticity must be applied on large spatial and temporal scales to noticeably impact the forecast, as grid-scale noise is readily dissipated by the model equations. However, it is not clear whether there is a physical origin for these large spatio-temporal correlation scales. The chosen scales in SPPT have been simply tuned to give the best results.

3 | THE COARSE-GRAINING FRAMEWORK

In this study, we use a high-resolution convection permitting atmospheric simulation to address the theoretical ideas underpinning SPPT. We consider the multiplicative nature of SPPT, the coherency of the uncertainty arising from different physics schemes, the coherency of the uncertainty in different prognostic tendencies, and the existence of large spatio-temporal correlation scales in the optimal perturbation.

The high resolution simulation used here was one of several simulations produced by the UK 'Cascade' project, funded by the Natural Environment Research Council. The chosen simulation was produced using the UK Met Office Unified Model (MetUM) at 4 km resolution, covering the Indo-Pacific Warm Pool region, 20°S–20°N, 42–177°E (Holloway et al., 2012). The

model is semi-Lagrangian and non-hydrostatic, and uses Smagorinsky sub-grid mixing in the horizontal and vertical dimensions. At 4 km resolution, the model is 'convection permitting'. The closure of the convection scheme is adapted such that almost all rainfall is generated explicitly, with the convection scheme only active in weakly unstable cases[1]. The model has 70 terrain following levels in the vertical with a model top at 40 km, and uses a time step of 30s. The lateral boundary conditions were provided by relaxing the simulation to a 12 km parametrised run through a nudged rim of 8 grid points. The simulation begins on 6 April 2009, and spans 10 days. The start date was selected to study an active Madden-Julian Oscillation (MJO) event. The data is stored at full resolution in space, and once an hour in time, and is available on request from the NERC Centre for Environmental Data Analysis (CEDA). For further details of the simulation, see Holloway et al. (2012).

The high-resolution simulation realistically simulates tropical meteorology, as reported on in Holloway et al. (2012, 2013, 2015). At 4 km, the simulation does not fully resolve convection. However, the simulation was more realistic than a similar 1.5 km simulation (C. Holloway, pers. comm.) and showed substantial improvements over simulations with parametrised convection, including simulating a realistic rainfall distribution (Holloway et al., 2012), vertical heating structure (Holloway et al., 2015), relationship between precipitation rate and tropospheric humidity (Holloway et al., 2012, 2013), and realistic generation of eddy available potential energy (Holloway et al., 2013). The ability to simulate these basic physical relationships allows for a realistic MJO, including simulating a degree of convective organisation, MJO strength, and propagation speed that match observations (Holloway et al., 2013). The simulation captures dynamical features, including a realistic representation of horizontal and vertical wind speeds, though ascent is more spatially confined in observations than in the simulation (Holloway et al., 2013). The first day of the Cascade simulation showed a very strong spin-up (Holloway et al., 2012). We therefore discard the first day, and focus our analysis on the remaining nine days.

This study will treat the high-resolution Cascade simulation as the 'truth' which we would like a low resolution forecast model to be able to mimic. We first coarse-grain the high resolution simulation to the resolution of the forecast model. Ideally, a low resolution forecast model would be able to predict these coarse-grained fields. The difference between the low resolution forecast model and this coarse-grained 'truth' is defined to be the model error in the low resolution forecast model. It is this error that a stochastic parametrisation seeks to represent. By characterising the statistics of the model error, we can design a stochastic parametrisation scheme to represent this model error in forecasts, following Figure 1.

3.1 | The IFS SCM

The low-resolution forecasting model used in this study is the ECMWF Integrated Forecasting System (IFS), model version CY40R1, at $\mathsf{T}_\mathsf{L}639$ resolution (approximately 30 km grid box) with 91 vertical levels and a timestep of 15 minutes. This is a typical resolution used in a global ensemble prediction system, and is a substantially finer resolution than considered in previous coarse-graining studies. To combine the coarse-graining procedure with the low-resolution forecast model, we adapt the methodology described in Christensen et al. (2018b). Instead of considering forecasts made with the global IFS, the IFS Single Column Model (SCM) is used to integrate forward the equations of motion in each coarse-scale grid column. The CY40R1 IFS SCM has been released through the OpenIFS project.

The IFS SCM represents a single column taken from the global IFS model. The code base is the same as for the global IFS model, and includes the atmospheric physics parametrisations and the land surface scheme. The IFS SCM requires initial conditions for the atmospheric column and boundary conditions describing the sea surface temperature (SST), orography, vegetation, and surface fluxes. The IFS SCM does not contain the dynamics subroutines of the global IFS model. Instead it relies on the specification of external dynamical forcing fields including the vertical velocity, geostrophic winds, and advective tendencies of T, U, V, and q. The SCM contains a simplified set of dynamical equations which combine these specified forcings to estimate the total dynamical tendency for each prognostic variable (T, U, V, q). Given a full set of input fields, the SCM

[1] Less than 0.1% of precipitation in the Cascade simulation is due to parametrised convection.

predicts the future evolution of the atmospheric column. This includes a decomposition of the change in the prognostic variables into a component from each physical parametrisation scheme and a component from the dynamics (i.e. advection and diffusion).

3.2 | Coarse-graining the Cascade dataset

The coarse-graining methodology is detailed in Christensen et al. (2018b). An overview is provided here for ease of reference.

All SCM input fields will be taken from the coarse-grained Cascade dataset. The IFS $\mathsf{T}_\mathsf{L}639$ reduced gaussian grid is used to define the latitude and longitude coordinates that make up the coarse-scale SCM grid. The fields from Cascade are coarsened onto the $\mathsf{T}_\mathsf{L}639$ grid using local area averaging. This allows for high-resolution grid boxes to contribute a fractional component to several coarse-resolution grid boxes:

$$\overline{\psi}_{n,k} = \sum_f W_{n,f}\,\psi_{f,k} \tag{2}$$

where ψ_f denotes the field on the fine grid and $\overline{\psi}_n$ denotes the field on the coarsened grid. The coarse (fine) grid box is identified by the index n (f). $W_{n,f}$ indicates the fraction of fine grid box f within coarse grid box n, and the vertical level of the field is indicated by index k. Note that this is one choice of coarse-graining procedure, with alternatives including Gaussian (Bolton and Zanna, 2019) or spectral filters (Shutts and Pallares, 2014).

The fine- and coarse-resolution datasets are defined on model levels, and interpolation must also be performed in the vertical. We choose to perform vertical interpolation second, after first averaging horizontally across each model level. The first field to be coarsened in this way is the surface pressure. The low-resolution surface pressure field is combined with the ECMWF hybrid height coefficients, A_k and B_k, to define the pressure on the ECMWF hybrid model levels. The coarse-grained Cascade dataset is then interpolated, logarithmically in pressure, from the Cascade model levels to the ECMWF model levels. The Cascade model top is at 40 km altitude. Above this level, the fields are padded using ECMWF operational analysis data. The Cascade and analysis datasets are smoothly blended over five levels. Finally, a 9-point gaussian smoother is applied to all initial condition fields after coarse graining. This removes small scale features present in the Cascade simulation that are unresolved on the low resolution grid, and which therefore appear as grid-point noise.

The IFS SCM assumes all dynamical forcing fields are instantaneous. The advected tendencies of the prognostic variables (T, U, V, q) are calculated along IFS model levels from the coarsened fields:

$$\mathrm{adv}(\psi)|_{n,k} = -\overline{\mathbf{u}}_{n,k} \cdot \overline{\nabla}_k(\overline{\psi_{n,k}}) \tag{3}$$

for variable ψ. A centred finite difference scheme is used to estimate the vector gradient in ψ before the dot product is taken with the coarse-grained vector wind field, $\overline{\mathbf{u}}_{n,k}$. The required geostrophic wind forcing and vertical velocity forcing are also evaluated using the coarse-grained fields: see Christensen et al. (2018b) for more details.

The constant boundary fields required by the SCM are taken from the ECMWF archive at $\mathsf{T}_\mathsf{L}639$ resolution, ensuring the SCM has the same boundary conditions as the global model. Interactive land surface processes are turned off in the SCM, and replaced with time varying latent and sensible heat fluxes from the Cascade simulation.

The coarse-grained Cascade data sets required to drive the IFS SCM are archived at the Natural Environmental Research Council Centre for Environmental Data Analysis (Christensen et al., 2018a).

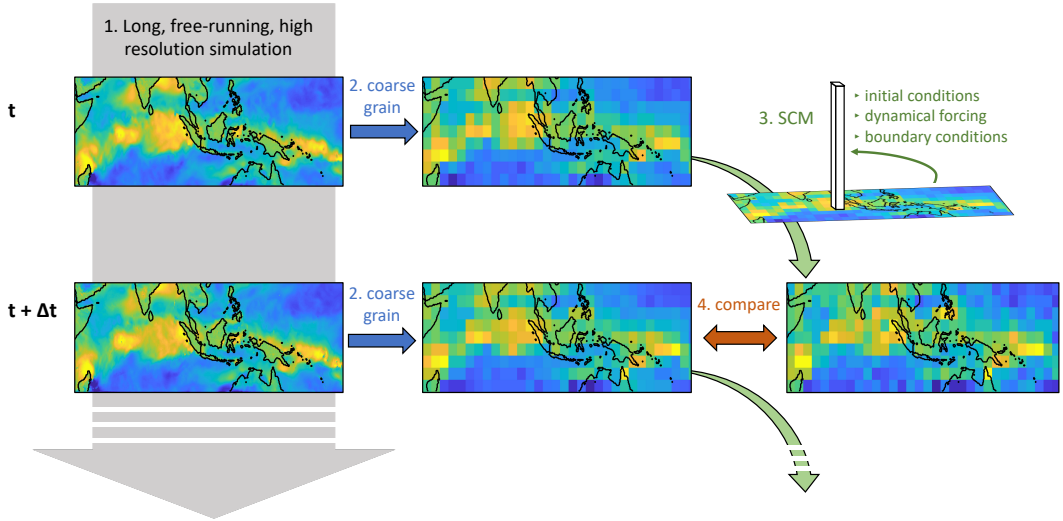

FIGURE 2 Schematic summarising the experimental procedure. 1. An existing high-resolution simulation is selected. 2. At each archived timestep, the high-resolution simulation is coarse grained to the resolution of the forecast model of interest. 3. The coarse-grained fields provide the initial conditions, boundary conditions and dynamical forcing for a Single Column Model (SCM). The SCM is used to step forward the low-resolution fields from t to $t + \Delta t$, independently for each coarse-grained grid box. 4. The resultant low-resolution forecast is compared to the coarse-grained high-resolution simulation at $t + \Delta t$. This procedure is repeated to span the entire length of the high-resolution simulation, as indicated by the dashed arrows.

3.3 | Experimental details

Figure 2 summarises the experimental procedure. The Cascade simulation is coarse-grained to provide the initial conditions, boundary conditions, and dynamical forcings required by the IFS SCM. An SCM integration is initialised every hour for each grid box of the coarse-grained Cascade simulation, starting at 00 UTC on 7 April 2009. The IFS shows a spin-up period over the first few timesteps, with anomalously high rain rates and associated moisture and temperature tendencies (Christensen et al., 2018b), indicating that the IFS parameterisation schemes rapidly equilibrate to a dryer mean-state compared to the MetUM physics in the Cascade simulation. It has not been assessed whether this is due to a bias in the IFS physics, the MetUM physics, or both. However, since SPPT acts over the entire forecast duration, this study is primarily concerned with errors in the evolution of the weather patterns which occur after this rapid adjustment. The statistics of these errors are relevant to the bulk of the model simulation and so can be used to inform SPPT. For this reason, each SCM simulation was run for two hours (eight SCM timesteps), the first hour of each SCM simulation is discarded and the second hour considered for analysis. Note that focusing on the first hour of each simulation could provide useful information on systematic model biases associated with the model spin-up: such analysis is left for a future study. It is necessary to consider the cumulative error over four timesteps because the Cascade fields are saved once an hour. The SCM is not nudged to the observed Cascade fields, but instead evolves freely. The lowest 60 IFS model levels are considered, excluding those above the Cascade model top. Note that model level 1 corresponds to the IFS model top, while model level 91 is closest to the ground. For a conversion between model levels and characteristic pressure levels, see Table S1 in the online supporting information.

The high resolution Cascade simulation is nudged towards a lower resolution simulation over a rim of points 32km wide. We therefore discard a rim of two coarse-grained points before analysis. The remaining domain spans 138 latitudinal by 476 longitudinal coarse-grained points, i.e. in excess of 65,000 independent SCM simulations per time step.

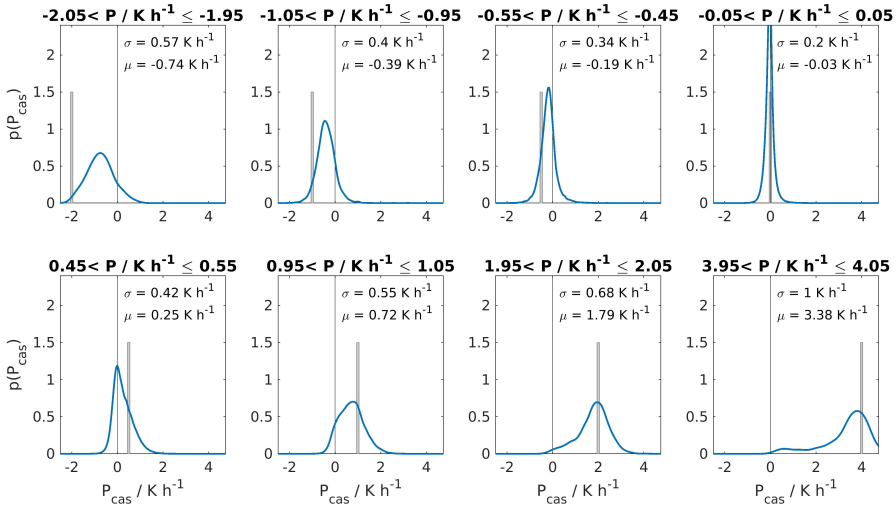

FIGURE 3 Testing the multiplicative noise hypothesis. Results are shown for the T tendency at model level 77 (approximately 850 hPa). Each subplot shows the distribution of $\mathbf{P}_{cas}$ conditioned on the physics tendency predicted by the SCM indicated by the title of each panel. The mean and standard deviation of the conditional $\mathbf{P}_{cas}$ distribution are shown in each panel. The grey rectangle shows the $\mathbf{P}$ distribution for each panel, though note that the height of this distribution is truncated to 1.5 for clarity: it should extend to 10. The number of data points used to estimate the pdf are 131, 4123, 29438, 1114817, 87759, 27608, 4020, and 162 for each panel respectively.

4 | ASSESSING SPPT: MULTIPLICATIVE NOISE

4.1 | Testing the multiplicative noise hypothesis

The experimental procedure outlined above allows the estimation of all the key terms in the SPPT equation 1. The total tendency, $\mathbf{T}$ for each variable is defined as the change in a prognostic variable between two consecutive coarse-grained Cascade fields, $(t, t + 1\text{hr})$. This is the 'target' which the forecast model should be able to predict. The SCM integration produces a tendency from each parametrised physics scheme, $\mathbf{P}_i$ over the same one-hour window, having been initialised at $t - 1\text{hr}$ and the first hour discarded. The SCM combines the provided forcing files from Cascade to produce a dynamics tendency, $\mathbf{D}$. Each of these tendencies are available for each prognostic variable, (T, U, V, q), as a function of model level, across the Cascade domain.

Firstly it is assessed whether multiplicative noise is a suitable model for the uncertainty in the IFS parametrisation schemes. To do so, an observed "Cascade physics tendency" is constructed as $\mathbf{P}_{cas} = \mathbf{T} - \mathbf{D}$ [2]. Treating each prognostic variable independently, the data are sorted by the predicted SCM physics tendency, and grouped into bins of equal width. Figure 3 shows the distributions of $\mathbf{P}_{cas}$ conditioned on $\mathbf{P}$ predicted by the SCM for T tendencies at 850 hPa. The chosen range in $\mathbf{P}$ is indicated by the figure panel titles and by the grey rectangle in each figure. Qualitatively the SCM parametrisation schemes are performing well: the average $\mathbf{P}_{cas}$ is well predicted by the SCM for positive tendencies, though there is a bias in the negative tendencies, where the mean $\mathbf{P}_{cas}$ has a smaller magnitude than $\mathbf{P}$. Figure 3 also shows that the uncertainty in the true tendency increases as the tendency increases, as modelled by SPPT. However, the distributions are not Gaussian, and there is a non-negligible probability that the 'true' tendency has the opposite sign to the predicted tendency. This is not allowed by SPPT, as the distribution of e in equation 1 is truncated at -1. The fourth panel shows the distribution of $\mathbf{P}_{cas}$ when the predicted tendency is zero. The standard

[2] Note that SPPT assumes there is no error in the dynamics tendency. This assumption is not questioned in this paper.

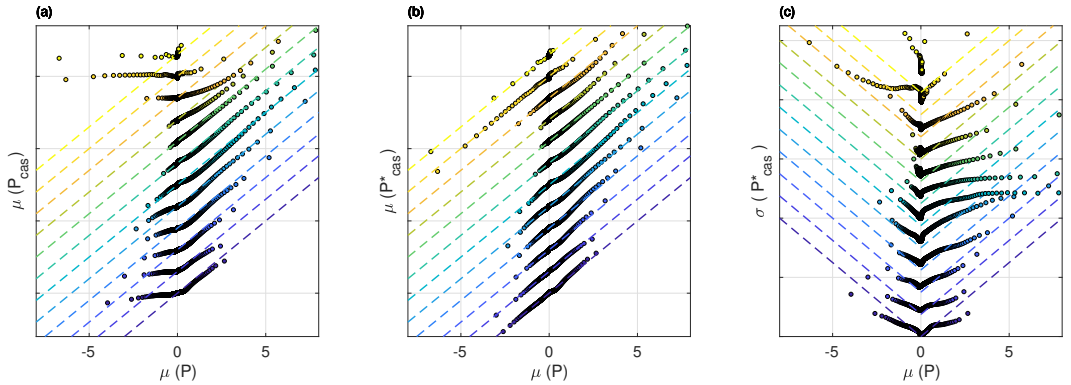

FIGURE 4 Testing the multiplicative noise hypothesis. Results are shown for the T tendency. (a) The mean of $\mathbf{P}_{cas}$ and (b) the mean of the bias corrected $\mathbf{P}^*{}_{cas}$, both conditioned on $\mathbf{P}$. The dashed lines indicate the one-to-one line for each level group. (c) The standard deviation of the corrected $\mathbf{P}^*{}_{cas}$ conditioned on $\mathbf{P}$. The dashed lines have a gradient of 0.5. In all panels the colours indicate the model level group, from dark blue (levels 87–91 / 980–1000 hPa) through to yellow (levels 32–36 / 60–80 hPa). Each level group is plotted with a vertical offset for clarity. The bin edges are defined based on the percentiles of the predicted tendency distribution for each level group, and so are equally populated.

deviation is substantial. Multiplicative noise is not able to represent this observed model uncertainty for small tendencies.

Figure 3 tests the multiplicative noise hypothesis at one height in the atmosphere. Figure 4 summarises this analysis for all model levels for T (the equivalent figures for q, U, and V are included in the supplementary material, figures S1, S2 and S3). In the vertical, the levels are grouped into 12 groups of five levels, and the data from each group of five levels are binned into 100 equally populated bins. Panel (a) shows the mean of $\mathbf{P}_{cas}$ conditioned on $\mathbf{P}$. If the SCM physics parametrisations are able to predict the 'observed' Cascade physics tendency accurately, the scattered points in Figure 4(a) should lie on the one-to-one dashed line. The positive temperature tendencies are well calibrated across almost the whole vertical domain, while the bias in negative tendencies highlighted in Figure 3 is confined to the lower troposphere and stratosphere.

Stochastic parametrisation schemes are designed to represent random model error. Therefore the systematic biases identified in Figure 4(a) will be modelled separately. Motivated by figure 4(a) and adopting the simplest functional form, the systematic difference between the average Cascade physics tendency, $\mu(\mathbf{P}_{cas})$ and the SCM physics tendency $\mathbf{P}$ is modelled as

$$\mu(\mathbf{P}_{cas}) = a_1\mathbf{P} + m, \qquad \mathbf{P} > 0 \tag{4}$$

$$\mu(\mathbf{P}_{cas}) = a_2\mathbf{P} + m, \qquad \mathbf{P} < 0 \tag{5}$$

Where a linear functional form has been assumed. The gradient is calculated separately for positive and negative tendencies, but a common intercept ensures a continuous function. The fit parameters are calculated separately for each model level and variable, and are available graphically in the supplementary material, Figure S4. The bias, $\mathbf{b}(\mathbf{P})$, is defined such that $\mu(\mathbf{P}_{cas}) = \mathbf{P} - \mathbf{b}(\mathbf{P})$. The SPPT equation 1 is rewritten to include this representation of both sources of error:

$$\mathbf{T} = \frac{\partial X}{\partial t} = \mathbf{D} + (1 + e)\mathbf{P} - \mathbf{b}(\mathbf{P}) \tag{6}$$

This error model is used for the rest of the paper.

To probe the statistics of the random model error, e, a corrected cascade tendency is defined $\mathbf{P}^*{}_{cas} = \mathbf{P}_{cas} + \mathbf{b}(\mathbf{P})$. Figure 4(b) shows that this simple linear approach is able to remove the majority of the systematic bias: the tendencies are now well calibrated.

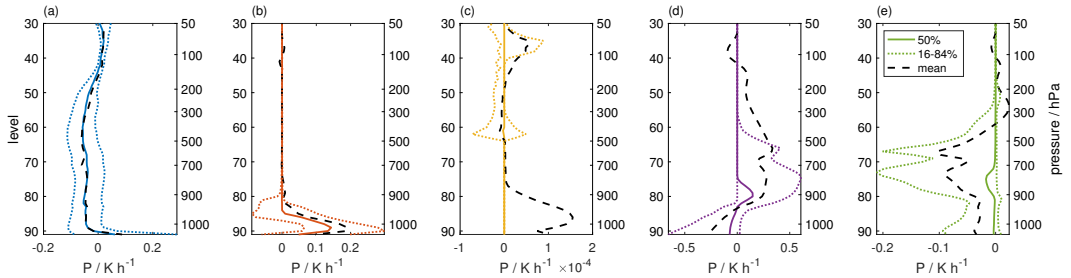

FIGURE 5 The distribution of tendencies from each parametrisation scheme as a function of model level. (a) Radiation. (b) Turbulence and orographic gravity wave drag. (c) Non-orographic gravity wave drag. (d) Convection. (e) Large-scale water processes. In each panel the coloured dotted lines indicate the 16th and 84th percentiles of the distribution as a function of model level; the coloured solid line indicates the median; the black dashed line indicates the mean. The mean often differs substantially from the median, indicating a strongly non-normal distribution. The right hand vertical axis shows characteristic pressure levels for reference.

Figure 4 (c) shows the standard deviation of $\mathbf{P}^*_{cas}$ conditioned on $\mathbf{P}$. If multiplicative noise is a good representation of model error in the IFS, the standard deviation should be proportional to the magnitude of the mean tendency, such that the scattered points lie on straight lines through the origin. The gradient of these lines indicates the optimal standard deviation of the stochastic perturbation. Straight lines are plotted in figure 4 (c) to guide the eye. The figure indicates there is some justification for the multiplicative noise hypothesis. The standard deviation is an approximately linear function of the mean SCM tendency for negative tendencies across the troposphere. The relationship is also approximately linear for small positive tendencies up to 750 hPa (the lowest four level blocks). In the mid troposphere, the standard deviation does increase with $\mathbf{P}$, but the relationship is markedly nonlinear. The largest warming tendencies do not show a correspondingly large standard deviation. In the upper troposphere and lower stratosphere, the linear relationship returns.

To attribute these observed characteristics to the behaviour of different parametrisation schemes, the levels at which each parametrisation is active must be characterised. Figure 5 summarises the probability distribution of T tendencies from each physics scheme as a function of model level. The mean and median of the distributions, and the 16th and 84th percentiles (equal to $\pm 1\sigma$ for a Normal distribution) are shown. Convection is the dominant parametrisation scheme that warms the mid troposphere, such that the well calibrated positive tendencies at those levels can be attributed to a well calibrated convection scheme. Similarly, the non-linear relationship between standard deviation and mean tendency at these levels is indicative of the statistics of model uncertainty in the convection parametrisation scheme. The turbulence and gravity wave drag scheme is the primary scheme warming the atmosphere at the lowest model levels. The positive tendencies at these levels appears well calibrated. The uncertainty in the tendencies is a linear function of the mean tendency, though the gradient changes between small tendencies and large tendencies. The TGWD tendency contains contributions from a number of processes including turbulence, mass flux in the boundary layer, and orographic gravity wave drag. It is possible that these different contributions have different uncertainty characteristics.

4.2 | The statistical characteristics of the optimal perturbation

Section 4.1 indicates that multiplicative noise is a reasonable first-order approximation to the uncertainty in the IFS physics parametrisation schemes, providing support for the use of SPPT. To inform the properties of the stochastic perturbation to be used in SPPT, including magnitude and spatio-temporal correlation scales, the optimal multiplicative perturbation is calculated for every grid point and time step, i.e. the perturbation, e, which best maps the forecast tendency onto the 'true' tendency. A

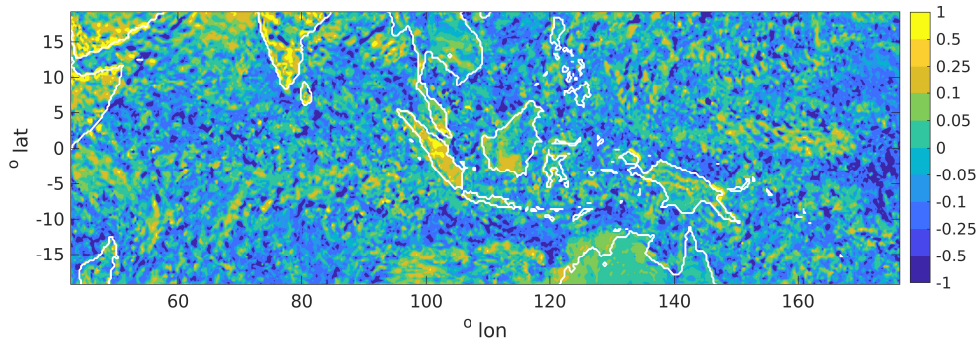

FIGURE 6 A snapshot of the optimal multiplicative perturbation required to map the SCM forecast onto the observed evolution of Cascade. The SCM forecasts were initialised at 00UTC, 7 April 2009.

forecast model cannot predict this optimal perturbation, but the statistics of this perturbation can be included in the forecast model. Each ensemble member in an ensemble forecast will experience a different realisation of this stochastic perturbation. The ensemble will thereby encompass the 'true' optimal perturbation in the ensemble, accounting for model uncertainty, and producing a reliable forecast.

To calculate the optimal multiplicative perturbation, we return to the SPPT equation 6 and rearrange such that the random error in the SCM parametrisation scheme, i.e. the difference between $\mathbf{P}*_{cas}$ and $\mathbf{P}$, is represented as a function of the SCM net physics tendency $\mathbf{P}$:

$$\mathbf{T} - \mathbf{D} - \mathbf{P} + \mathbf{b}(\mathbf{P}) = e\mathbf{P} \tag{7}$$

This is an over-constrained equation for the optimal instantaneous multiplicative perturbation, e. Equation 7 is solved simultaneously for all prognostic variable tendency vectors. The different prognostic variable tendencies have different units and therefore substantially different magnitudes (e.g. compare Figure 4 with Figures S1, S2 and S3). To remove this dependency on unit and to ensure each variable tendency is weighted equally, each tendency is first divided through by a scale factor $s_X = \sigma(\mathbf{T}_X)$ for variable X. In the IFS, SPPT is tapered in the boundary layer and upper stratosphere for stability reasons. These levels are therefore excluded from the analysis. Since the physics tendencies are smaller in the upper troposphere and lower stratosphere (Figure 5)[3] it was found that these higher levels dominated the procedure for fitting e. To focus on levels where the tendencies have an appreciable magnitude these levels are also excluded. Only levels between 45 and 87 (inclusive) are used to calculate e. The sensitivity of e to the choice of level is shown in the supplementary material, Figure S5. Equation 7 is solved by minimising the mean squared residual[4] The solution is the optimal e as a function of horizontal position and time.

Figure 6 shows a map of the instantaneous optimal multiplicative perturbation, e, for forecasts initialised at 00UTC on April 7th 2009. While there is substantial small-scale variability in e, this is embedded within larger-scale correlated structures. The day-night boundary (at approx. 90 E) is visible in the increased errors over land in the night regions.

The distribution of e is summarised in Figure 7 in terms of the mean, standard deviation, skewness and kurtosis as a function of local time of day. The dashed lines in Figure 7 show the statistics calculated using each day separately to indicate the variability in each statistic. The statistics aggregated over the entire domain (black) are compared to those aggregated over only land regions (green) and those aggregated over only ocean regions (blue). Over land regions, the mean perturbation shows a marked diurnal

[3] Note that the radiation tendencies are only small because shortwave and longwave tendencies approximately balance in the stratosphere for the daily average.

[4] This is achieved using the MatLab backslash operator, which in this case employs a QR decomposition to find the solution.

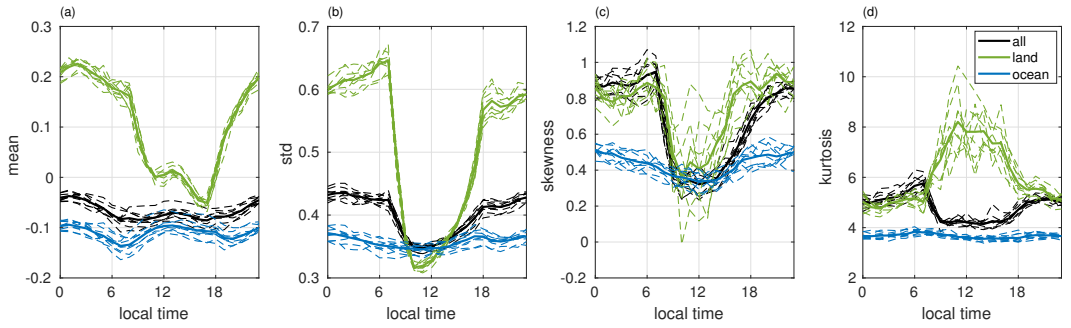

FIGURE 7 Summary statistics of the distribution of e as a function of local time of day. The statistics are shown calculated over the entire domain (black) as well as over only land points (green) and over only ocean points (blue). The solid line indicates the average statistics over the entire dataset, while the dashed lines indicate the statistic calculated for each of the nine days separately.

cycle, with perturbations positive on average at night, and zero or slightly negative during the day. The mean perturbation over ocean points is negative, indicating a slight systematic reduction in the magnitude of the physics tendencies of order 10%.

The standard deviation over land also shows a marked diurnal cycle, with larger magnitude perturbations at night and smaller magnitude perturbations during the day. In contrast, perturbations over ocean points do not show a strong diurnal cycle. The standard deviation of the perturbation over the ocean is 0.36. Overall, the standard deviation aggregated over the whole domain is 0.40, in contrast to the value of 0.55 used operationally in the IFS.

It is known that forecast models struggle to represent the diurnal cycle in convection over land and that many do not predict a peak in convective activity at night as observed (Guichard et al., 2004; Love et al., 2011; Couvreux et al., 2015). The statistics of error in the convection parametrisation scheme will therefore vary with time of day over land. The probability distribution of tendencies from each physics scheme (Figure 5) was calculated separately for land and ocean points during night and day hours (not shown). While the distribution of parametrised tendencies over ocean does not vary with time of day, at night the parametrised physics tendencies over land have smaller magnitude, reflecting the lack of convection at night in the forecast model. A multiplicative perturbation is not able to adequately represent the error in this small tendency.

The calculated e show positive skewness and positive kurtosis over both land and ocean regions. This is in contrast to the perturbations used in the IFS, which are normally distributed. The positive skewness indicates that large negative e, which could change the sign of the parametrised tendency, are less common than large positive e. The kurtosis greater than three indicates the distribution of e has fat tails. Both positive skewness and excess kurtosis indicate the need for large positive perturbations more often than would be modelled by a Normal distribution. Large positive perturbations could be due to the lack of quasi-equilibrium in the boundary layer (Bechtold et al., 2014).

The SPPT scheme used operationally in the IFS uses a sum over three spectral patterns with specified spatial and temporal correlation scales. To inform the optimal length and time scales used in SPPT, e was similarly modelled as a sum over a number of independent patterns. The autocorrelation of e was estimated separately in longitude, latitude, and time. It was found that representing the spatial and temporal autocorrelation functions a sum over N AR(1) processes represented the error processes well. For details of the fitting process, see the Appendix. For each dimension, N was chosen to give the best representation of the estimated correlation function: the domain is large enough in longitude to fit $N = 3$ AR(1) processes, whereas the latitudinal and temporal extents did not permit fitting a third pattern such that $N = 2$. Figure 8 shows the fitted processes for each spatio-temporal direction using data aggregated from over the whole domain.

The autocorrelation functions were also estimated separately for land and ocean regions as for Figure 7. The fitted e over

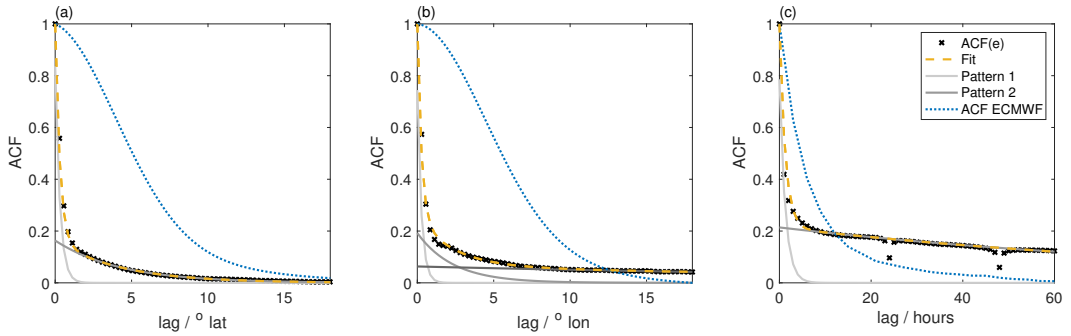

FIGURE 8 The autocorrelation of e in the (a) latitudinal, (b) longitudinal and (c) temporal dimensions (black crosses). The autocorrelation of e is modelled as the sum over N AR(1) processes, where $N = 2$ in the latitudinal and temporal directions and $N = 3$ in the longitudinal direction. The individual AR(1) processes are shown as the grey solid lines, whereas the yellow dashed line is the sum over these N processes. The goodness-of-fit is indicated by the match between the yellow dashed line and the black crosses. The autocorrelation of e as used operationally at ECMWF is shown by the blue dotted line.

ocean has longer spatial correlation scales than over land, whereas e over land points has a longer decorrelation timescale than over ocean (see supplementary material, Figures S6 and S7). The shorter spatial scales over land could be due to the presence of local topographic features or variability in the land surface. However, persistent errors in land surface parameters such as soil moisture could also give rise to extended decorrelation timescales when compared to those measured over ocean.

The parameters fitted in each dimension were combined to produce a spatial and temporal correlation for each scale pattern. Since the longitudinal autocorrelation indicates the existence of a third scale, it is assumed that three patterns exist in each spatio-temporal direction, but the restricted domain size prevents the third pattern from being identified in the latitude and temporal dimensions. The variance ratio for pattern three is therefore taken from the longitudinal fit, and the latitudinal and temporal variance ratios for patterns one and two rescaled to account for the unmeasured third scale. Finally, the relative magnitudes and decorrelation scales of the first two patterns are estimated by averaging the three estimates from each spatio-temporal dimension. For simplicity, this ignores the spatial anisotropy of e.

These data are shown in Table 1, and compared to the values used operationally in the IFS. The first fitted pattern has spatial scale of 32 km and a temporal scale of 1 h. These represent the spatial and temporal resolution of the coarse-scale data, such that this first field corresponds to white noise in time and space. The magnitude of this pattern is smaller than that assigned to the first pattern in operational SPPT. The second pattern accounts for a relatively larger fraction of the variance than in the operational settings, and shows substantial spatial and temporal correlation scales. The fitted spatial scale of 370 km is similar to that used in the first operational SPPT pattern, while the fitted temporal correlation scale of 4.3 d is similar to the second operational SPPT pattern. The third pattern, with expected spatial scales on the order of thousands of km and temporal scales of several weeks, is too large to be constrained by the Cascade dataset, though the third pattern in the longitudinal direction indicates the presence of structures in e with scales of order 8,000 km.

5 | BEYOND SPPT

Section 4 presented evidence for the multiplicative noise hypothesis, providing some justification for SPPT. Within this framework, the statistical characteristics of the optimal perturbation were estimated. While the leading order pattern is white in time and space, there is evidence for large scale correlations in the second and third patterns, providing some justification for the use of correlated noise in stochastic parametrisation schemes. However it is evident that SPPT is not a perfect representation of

	Operational SPPT			Fitted SPPT		
$\mu(e)$	0.0			-0.07		
$\sigma(e)$	0.55			0.40		
σ_j	0.52,	0.18,	0.06	0.35,	0.17,	0.10
L_j (km)	500,	1000,	2000	32,	370,	–
τ_j	6 h,	3 d,	30 d	1.2 h,	4.3 d,	–

TABLE 1 SPPT parameter values for the random fields $j = 1, 2, 3$ that comprise the 3-scale pattern used in the IFS: Standard deviation σ_j, horizontal correlation length L_j, time decorrelation scale τ_j. The spatial and temporal scale of the third pattern cannot be accurately estimated due to the limited size of the domain and length of dataset.

uncertainty in IFS. At some vertical levels, the uncertainty in the parametrised tendency is not a simple linear function of the mean tendency. Furthermore, section 4 did not assess all the assumptions made by SPPT, including the coherency of uncertainty in the vertical, between different prognostic variables, and between different parametrisation schemes. These assumptions are tested in the following section.

5.1 | Vertical coherence of perturbations

We first assess whether a single e is a good representation of model error as a function of model level. In other words, is the vector error in the tendency proportional to the vector tendency. The validity of this assumption was discussed by Leutbecher et al. (2017), who point out that, for example, uncertainty in the shape of a tendency profile cannot be represented using a constant perturbation as a function of height. We probe this question by considering the characteristics of the optimal perturbation fitted separately to each model level, e_z.

Figure 9 summarises the distribution of the optimal e_z as a function of model level in terms of its deciles. It is evident that the characteristics of e_z vary as a function of height in the atmosphere. The standard deviation of e_z is smaller lower in the atmosphere, between levels 85 and 91. The fitted e_z distributions are roughly symmetrical between levels 50 and 91, though for the highest levels in the atmosphere the mean perturbation is positive. Separating the e_z data into (b) land and (c) ocean regions reveals differences. For example, over ocean, e_z for the lowest levels have negative mean and median, which is not the case over land. Over land, the distribution of e_z for levels in the free troposphere is positively skewed, whereas the distribution over ocean is roughly symmetric. Separating the e_z data into (d) day and (e) night reveals smaller uncertainties at night, and higher uncertainties during the day.

Figure 10 shows the correlation between perturbations fitted to different levels. If the correlation matrix shows 'square' features, this indicates a block of levels that are mutually highly correlated. In general, the perturbations fitted to different levels are weakly correlated. A notable exception comes in the lowest model levels (85-91), where perturbations show high mutual correlations. Consideration of Figure 5 shows that it is the turbulence and orographic gravity wave drag scheme that is active over these levels. Calculating the correlations separately for land and ocean regions shows this enhanced correlation is present for all grid points (not shown). Other regions also show time of day-dependent enhanced correlations. At night (panel c), the optimal perturbations show substantial correlations between levels 43 and 82. Considering diurnal variations in tendencies indicates that it is the radiation scheme which is active over these levels, with a net cooling tendency at night (see supplementary material figure S8). During the day (panel b), enhanced correlations are found between levels 48 and 61. Figure S8 shows several schemes are active over those levels, but note that the large scale water processes scheme shows a warming (drying) tendency confined to that vertical block. During the day enhanced correlations are also found above level 42. This could be due to enhanced radiative

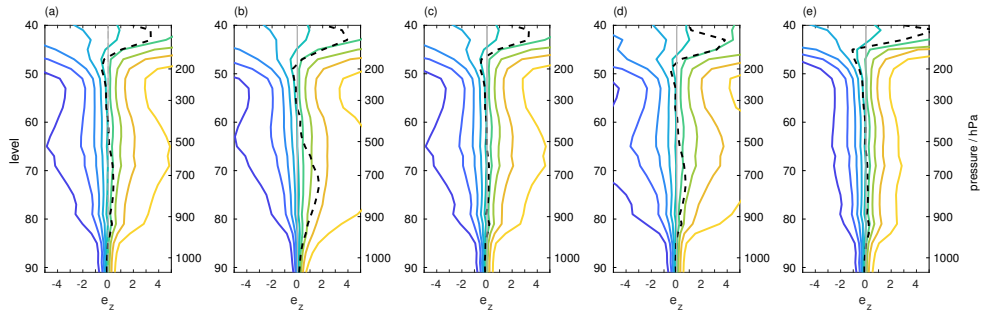

FIGURE 9 Distributions of optimal e_z as a function of height. The pdf is summarised by showing each decile as a function of model level from dark blue (10th percentile) to yellow (90th percentile). The black dashed line indicates the mean of the distribution. (a) all data, (b) e_z over land points, (c) e_z over ocean points, (d) e_z during the day: 6 am to 6 pm local time and (e) e_z during the night: 6 pm to 6 am local time. The right hand vertical axis shows characteristic pressure levels for reference.

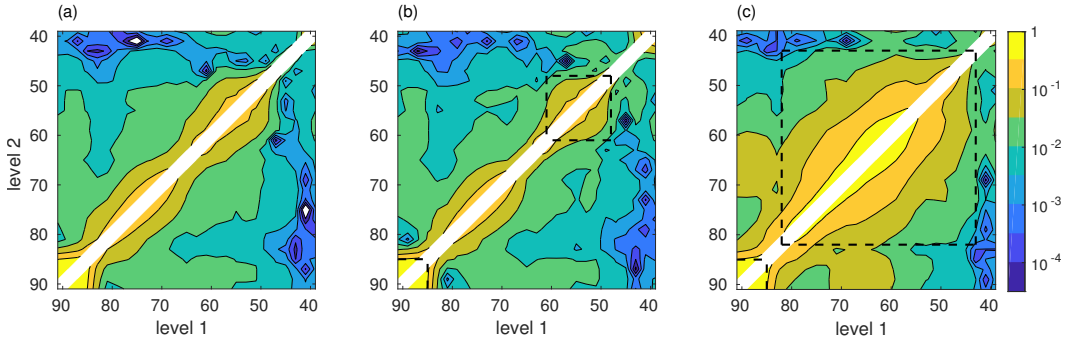

FIGURE 10 Correlation between the optimal e_z fitted to different model levels. The correlation of e_z with itself at a given level is one, and so has been masked out. (a) all data, (b) e_z during the day: 6 am to 6 pm local time and (c) e_z during the night: 6 pm to 6 am local time. Square features in the correlation indicate a block of levels with high inter-correlations. Dashed black lines highlight such features.

tendencies at those levels, present during the day but not at night (Figure S8).

Comparing the vertical regions of enhanced correlation in Figure 10 to the distribution of e_z as a function of height in Figure 9 reveals that vertical regions with enhanced correlations coincide with vertical regions over which the statistical properties of e_z are reasonably constant in height. It seems justifiable to conclude that a single multiplicative perturbation could be appropriate over those vertical regions.

5.2 | Coherency of optimal perturbations between different prognostic tendencies

SPPT assumes that the random error due to the parametrisation process is coherent between the parametrised tendencies for different prognostic variables. Most parametrisation schemes produce tendencies for more than one prognostic variable, the exception being radiation which only impacts T. In general the tendencies in different variables are physically related, for example thermodynamic process impact both T and q, or convective processes transport momentum, heat, and moisture aloft. Perturbing the different prognostic tendencies with the same random number is designed to improve the physical consistency of

	T			q			U			V		
$\mu(e_X)$	-0.060			-0.017			-0.37			-0.52		
$\sigma(e_X)$	0.70			0.65			1.7			1.9		
σ_j	0.66,	0.17,	0.13	0.60,	0.22,	0.10	1.6,	0.47,	0.18	1.8,	0.54,	0.18
L_j (km)	39,	400,	–	33,	430,	–	28,	270	–	26,	290,	–
τ_j	0.6 h,	3.5 d,	–	1.2 h,	4.3 d,	–	1.2 h,	3.8 d,	–	1.2 h,	4.2 d,	–

TABLE 2 As for table 1 except treating each prognostic variable ($X = T, q, U, V$) independently. Parameter values are shown for the random fields $j = 1, 2, 3$ that comprise the 3-scale pattern similar to that used in the IFS: Standard deviation σ_j, horizontal correlation length L_j, time decorrelation scale τ_j. The spatial and temporal scale of the third pattern cannot be accurately estimated due to the limited size of the domain and length of dataset.

SPPT, representing an instantaneous amplification or reduction of the strength of the parametrised processes.

To test this assumption, we take the opposite position and consider a generalised 'variable SPPT' (vSPPT) in which each prognostic variable, $X = \{T, q, U, V\}$, is perturbed with an independent pattern. The optimal perturbation, e_X, is fitted independently for each prognostic variable. If the statistical characteristics of the patterns are similar for each prognostic variable, and if there is a high degree of correlation between the e_X fitted to different variable tendencies, then we consider that evidence that a single perturbation should be used for all prognostic tendencies as proposed in SPPT.

The moments and spatio-temporal correlations are computed separately for each e_X following the methodology used in section 4. These statistics are summarised in Table 2, showing clear differences to the statistics of e shown in Table 1. The standard deviations of the e_X are higher than the standard deviation of e. There is a clear grouping into two pairs: e_U and e_V have a substantially higher standard deviation but smaller spatial decorrelation scales than e_T and e_q. The wind perturbations e_U and e_V also have a substantial negative mean, whereas the means for e_T and e_q are close to zero.

Aside from similarities to e_q in terms of mean, standard deviation, and spatial decorrelation scales, the e_T pattern shows certain characteristics not shared by any other e_X. The temperature perturbations e_T decorrelate more rapidly in time than e or the other e_X. The temperature perturbation e_T also shows a different spatial correlation structure, with a *negative* correlation at a lag of 1.5°. This is not captured by the AR(1) model (not shown), though at larger spatial lags the AR(1) model is a good representation of the correlation structure of e_T. Finally, the standard deviation of e_T is higher over ocean than over land in contrast to e and the other tendencies (see supplementary online material Figure S9).

Fitting a separate e_X to each prognostic tendency is expected to improve the error characterisation compared to fitting a single pattern because of the additional degrees of freedom in the fitting procedure. To quantify the improvement, the measured error between the SCM and Cascade,

$$\mathbf{d}_X = \mathbf{T}_X - \mathbf{D}_X - \mathbf{P}_X + \mathbf{b}(\mathbf{P}_X), \tag{8}$$

and the modelled errors,

$$\mathbf{d}_X^{\mathrm{SPPT}} = e\mathbf{P}_X,$$
$$\mathbf{d}_X^{\mathrm{vSPPT}} = e_X\mathbf{P}_X, \tag{9}$$
$$\tag{10}$$

are calculated for each variable, X. The mean square difference (MSD) between measured and modelled error is calculated for

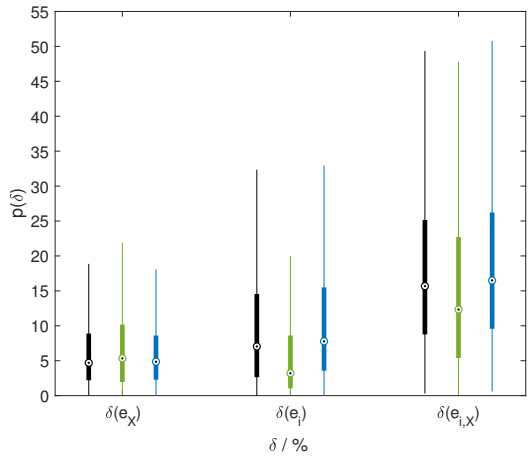

FIGURE 11 Distribution in the change of the mean square difference between the actual error, $T - D - \Sigma P$ and the modelled error, $e\Sigma P$ for different modifications to SPPT: when independent e_X are used for different variables; when independent e_i are used for different parametrisation schemes; when independent $e_{i,X}$ are used for different parametrisation schemes and variables. δ is the percentage improvement in the representation of the error on using the modified SPPT approach. The data are shown as box plots, where the median δs are shown as a circle, and the boxes indicate the 25th and 75th percentiles. Black: data aggregated from across whole domain. Green: only land points. Blue: only ocean points.

each model and averaged over all variables, as a function of time and spatial position. The percentage improvement,

$$\delta = 100 \cdot \frac{MSD^{\text{SPPT}} - MSD^{\text{vSPPT}}}{MSD^{\text{SPPT}}}, \tag{11}$$

is evaluated. Figure 11 summarises the distribution of δ over time and spatial positions using a box and whisker diagram. The median improvement is 5%, with the whiskers extending to 20%. The median fractional variance explained by the approach (i.e., the ratio of modelled variance, $\sigma^2(e\mathbf{P}_X)$, to the variance of the measured error, $\sigma^2(\mathbf{d}_X)$ increases by a factor of 2.7 on moving from SPPT to vSPPT (see Supplementary Figure S10).

Figure 12 shows the correlation between e_X fitted for different X as a function of local time of day. Most variable pairs show a modest correlation of order 0.1. However, there are noticeable correlations of 0.3 to 0.45 between e_T and e_q, peaking in the mid morning and mid afternoon. This high correlation between e_T and e_q is expected given the physical relationship between T and q tendencies associated with thermodynamic processes. Consideration of the diurnal cycle in parametrised processes (supplementary online material Figure S8) indicates that convective activity peaks in mid morning, while large scale water processes peaks in mid afternoon, explaining the peak in correlation at those times.

Given the high correlations between e_T and e_q, and the similarity in standard deviation and spatial correlations between these variable perturbations, perturbing T and q with the same pattern as in SPPT seems physically justified. It is interesting to note that there is not a high correlation between e_U and e_V, despite the intimate relationship between U and V. It is possible that consideration of wind magnitude and direction, or divergent and rotational flow, would indicate correlated errors in these two variables, not evident when considering U and V.

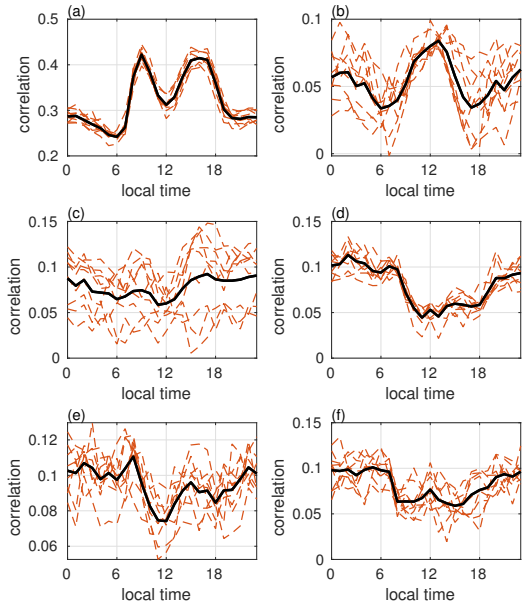

FIGURE 12 Correlation between e_X fitted to different variable tendencies, as a function of local time of day. (a) $r(e_T, e_q)$; (b) $r(e_T, e_U)$; (c) $r(e_T, e_V)$; (d) $r(e_q, e_U)$; (e) $r(e_q, e_V)$; (f) $r(e_U, e_V)$. The solid black line indicates the average statistics over the entire dataset, while the dashed lines indicate the the statistic calculated for each of the nine days separately.

5.3 | Coherency of perturbations between different parametrisation schemes: iSPPT

The final assumption of SPPT to consider is that the random error is coherent between the different physics parametrisation schemes, such that the balance between tendencies from different physics schemes is retained. However this may not accurately reflect variations in model uncertainty (Leutbecher et al., 2017). As in section 5.2, we test this by taking the opposite stance and consider a generalisation to SPPT where each physics parametrisation scheme is perturbed using an independent pattern:

$$\mathbf{T} = \mathbf{D} + \sum_{i=1}^{I} (1 + e_i)\mathbf{P}_i - b(\mathbf{P}). \tag{12}$$

In this 'independent SPPT' (iSPPT) approach, proposed by Christensen et al. (2017b), the spatio-temporal characteristics of the independent patterns e_i can be specified by the user, allowing for the representation of different model error characteristics associated with each physical parametrisation scheme. We can use the coarse-graining framework to assess whether the statistical characteristics of the e_i are indeed different, or if using a single e as in SPPT is sufficient to represent model uncertainty in the IFS.

The e_i are estimated by solving the over-constrained vector equation:

$$\mathbf{T} - \mathbf{D} - \mathbf{P} + \mathbf{b}(\mathbf{P}) = \underline{\underline{\mathbf{P}}}\mathbf{e} \tag{13}$$

at every spatial location and time step, where the matrix $\underline{\underline{\mathbf{P}}}$ consists of I columns each containing parametrised tendency i as a function of height. The $I \times 1$ vector $\mathbf{e}$ contains the e_i optimal perturbations. It was found that fitting an independent perturbation for non-orographic gravity wave drag (NOGW) led to instabilities in the fitting procedure, because that scheme

	RDTN			TGWD			CONV			LSWP (LSWP in CONV regions)		
$\mu(e_X)$	-0.023			-0.043			-0.16			-0.34 (0.18)		
$\sigma(e_X)$	1.4			0.66			4.0			14 (2.7)		
σ_j	1.4,	0.32,	0.24	0.59,	0.28,	0.11	3.9,	0.87,	0.15,	14 (2.6),	3.2 (0.59),	1.8 (0.34)
L_j (km)	38,	570,	–	27,	330,	–	16,	240,	–	33,	370,	–
τ_j	0.79 h,	4.2 d	–	1.4 h,	5.0 d	–	0.72 h,	5.6 d,	–	0.86 h,	6.5 d	–

TABLE 3 As for table 2 except treating each parametrisation scheme (RDTN, TGWD, CONV, LSWP) independently. The bracketed numbers in the LSWP column indicate the pattern parameters if analysis is only carried out in regions where the convection parametrisation has triggered. Parameter values are shown for the random fields $j = 1, 2, 3$ that comprise the 3-scale pattern similar to that used in the IFS: Standard deviation σ_j, horizontal correlation length L_j, time decorrelation scale τ_j. The spatial and temporal scale of the third pattern cannot be accurately estimated due to the limited size of the domain and length of dataset.

produces tendencies that are three orders of magnitude smaller than the other parametrisation schemes (Figure 5). The NOGW tendencies were therefore excluded from $\underline{\underline{\mathbf{P}}}$ and instead moved to the left hand side of equation 13, and $\mathbf{e}$ calculated for the remaining four schemes.

Figure 13 shows an instantaneous snapshot of the optimal e_i for each of the physical parametrisation schemes considered: radiation (RDTN), turbulence and gravity wave drag (TGWD), convection (CONV), and large scale water processes (LSWP). A number of interesting results are apparent. Firstly, different parametrisation schemes show different error characteristics: for example, the optimal perturbation to the radiation scheme appears to have significantly smaller scales than the other schemes. Secondly, the characteristics of e_{TGWD} are remarkably similar to the characteristics of the e fitted to the total net tendency. Only the TGWD and CONV schemes produce tendencies in all four prognostic variables. Errors in the U and V tendencies therefore must be accounted for by perturbing the TGWD tendency in both the SPPT and iSPPT framework, especially since CONV does not trigger everywhere, explaining the similarities between e and e_{TGWD}. Thirdly, in regions where the convection scheme did not trigger (white in panel three), the perturbation in both the RDTN and LSWP schemes are of larger magnitude than in regions where convection did trigger. This could indicate that the convection parametrisation scheme has uncertainties that are not well represented by (independent) SPPT, for example, errors in triggering of convective events. This would justify stochastically perturbing physical processes within the convection scheme, such as triggering, instead of relying on the multiplicative approach of SPPT. In regions where convection did not trigger, the analysis approach taken here could incorrectly attribute errors in the convection parametrisation scheme to other parametrisation schemes, because of the multiplicative nature of SPPT.

Table 3 shows the moments and spatio-temporal correlations calculated for each e_i. The perturbations fitted to the TGWD scheme show similar statistics to the SPPT e perturbation, including similar standard deviation, spatial and temporal correlations, and relative magnitudes of the three patterns. The e_i fitted to other parametrisation schemes show substantially different statistics. The radiation perturbations have a larger standard deviation than e. The first pattern explains substantially more variance, explaining the shorter correlation scales observed for e_{RDTN} in Figure 13 than for e. Inspection of the full spatial decorrelation structure shows AR(2) behaviour similar to that observed for e_T: it is likely that the AR(2) behaviour in e_T can be traced to uncertainty in the radiation scheme. The perturbations e_{CONV} and e_{LSWP} also show larger standard deviations and a larger weighting on the first pattern than e, though they also show larger temporal correlation scales. The standard deviation of e_{CONV} is larger than the SPPT e perturbation, in particular over land regions, and shows diurnal structure with a peak in standard deviation at night (Supplementary figure S11). This is likely linked to errors in simulating the diurnal cycle of convection, as previously discussed. The standard deviation of e_{LSWP} is very large. This can be traced to regions where the convection scheme has not triggered.

The correlation between the perturbations fitted to different schemes as a function of local time of day is shown in Figure 14.

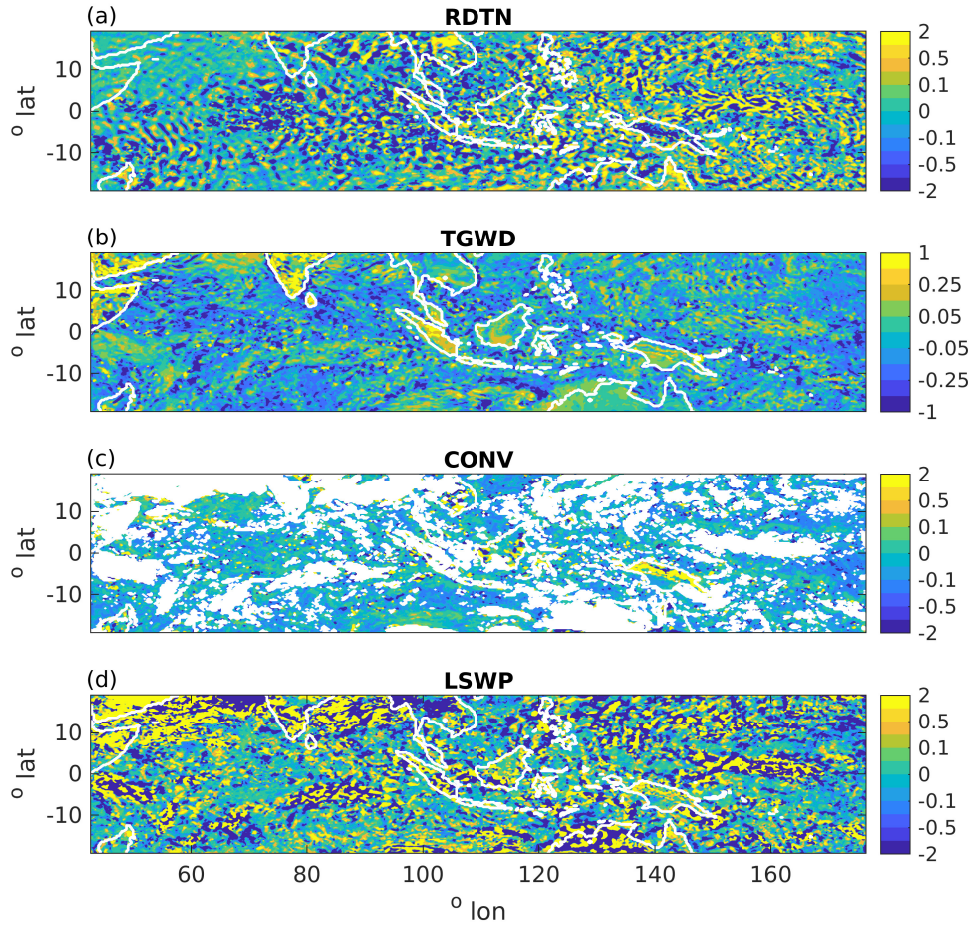

FIGURE 13 A snapshot of the optimal multiplicative perturbations, e_i, as defined in the iSPPT framework. The optimal perturbation to (a) the radiation scheme; (b) the turbulence and gravity wave drag scheme; (c) the convection scheme; (d) the Large-scale water processes scheme. The SCM forecasts were initialised at 00UTC, 7 April 2009. Note that the colour bar saturates in panels (a), (c) and (d).

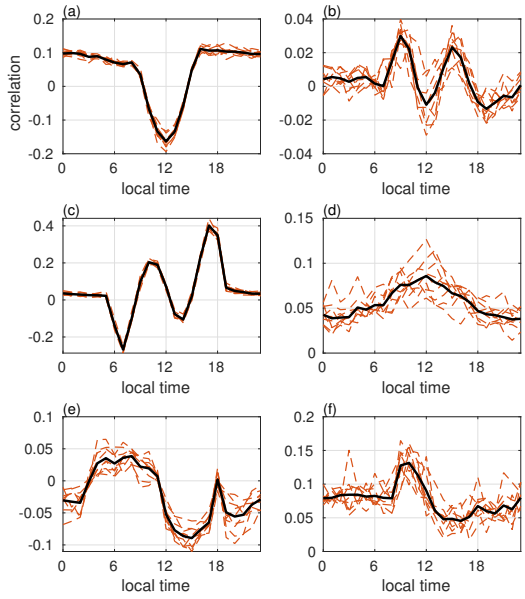

FIGURE 14 Correlation between e_i fitted to different parametrisation schemes, as a function of local time of day, as defined in the iSPPT framework. (a) $r(e_{\mathrm{RDTN}}, e_{\mathrm{TGWD}})$; (b) $r(e_{\mathrm{RDTN}}, e_{\mathrm{CONV}})$; (c) $r(e_{\mathrm{RDTN}}, e_{\mathrm{LSWP}})$; (d) $r(e_{\mathrm{TGWD}}, e_{\mathrm{CONV}})$; (e) $r(e_{\mathrm{TGWD}}, e_{\mathrm{LSWP}})$; (f) $r(e_{\mathrm{CONV}}, e_{\mathrm{LSWP}})$. The solid black line indicates the average statistics over the entire dataset, while the dashed lines indicate the the statistic calculated for each of the nine days separately.

Correlation are generally small, but show a clear diurnal structure. The correlations between the perturbations fitted to different schemes are likely due to physical relationships between the schemes. However, the fitted correlations between different schemes should be treated with caution, because of the aforementioned difficulties in attribution of error to a particular scheme in regions where the convection scheme did not trigger.

As for vSPPT, fitting a separate e_i to each parametrisation scheme is likely to improve the error characterisation because of the additional degrees of freedom in the fitting procedure. Figure 11 summarises the improvement in fit due to iSPPT over SPPT in terms of the distribution of δ. The median improvement is 7%, with the whiskers extending to over 30%. If the vSPPT and iSPPT approaches are combined (ivSPPT), and a separate $e_{i,X}$ fitted for each parametrisation scheme for each variable, the median improvement increases to 15%, with tails extending to 50%. Similarly, the median fractional variance explained by the approach increases by a factor of 3.4 (5.0) on moving from SPPT to iSPPT (ivSPPT) (see Supplementary Figure S10). For both iSPPT and ivSPPT, the improvement is larger over ocean points than over land points. The remaining residual between measured and modelled error cannot easily be represented through a multiplicative approach.

6 | DISCUSSION

Coarse-graining is an attractive technique for characterising model error. A high resolution simulation, which resolves the processes of interest, is taken as a benchmark. The simulation is coarsened to the resolution of the forecast model, before its evolution is compared to forecasts made with the low resolution model. The difference between the coarse-grained high resolution simulation and the forecast is considered the 'error' in the forecast. There are many benefits of using a high-resolution simulation for coarse-graining studies, such as excellent spatio-temporal coverage, and the availability of all model fields including those

poorly constrained by observations (e.g. vertical velocity) (Christensen et al., 2018b). The coarse-graining approach is especially useful if the high resolution simulation resolves a process of interest that is unresolved at the forecast resolution (e.g. convection), or if it makes use of a more sophisticated parametrisation scheme (e.g. 2-moment microphysics). However, it is important to note that the high-resolution simulation is only a proxy for the truth. The high-resolution simulation must be thoroughly validated against observational data before coarse-graining. This ensures any biases in the high-resolution simulation are understood, to avoid conflation with the forecast 'errors' assessed in the coarse-graining analysis.

This paper has proposed a new technique within the coarse-graining framework for assessing systematic and random model error. Instead of requiring the user to simultaneously produce a pair of low- and high-resolution simulations, it makes use of existing high-resolution simulations. This sets a relatively low bar for carrying out such coarse-graining studies, and makes use of the wealth of high-resolution simulations available (Heinze et al., 2017; Satoh et al., 2014, 2017; Schalkwijk et al., 2015). The coarse-grained dataset is coupled to a forecast model through the use of a single column model (SCM) framework. The initial conditions and forcing fields (including the dynamics tendencies) required by the SCM are provided by the coarse-grained dataset. The SCM is run independently in each coarse-grained grid box and provides the physics tendencies as a function of model level. This allows for a 3D characterisation of error in the parametrised physics tendencies.

The technique can be used to study the development of systematic biases between the high resolution simulation and the low resolution forecast model, developing over just a few time steps. Evaluating bias development over short time periods ensures that any biases identified are unlikely to be due to a remote source (Rodwell et al., 2016). This contamination is a problem when using standard validation techniques, e.g. the evaluation of parametrisation schemes in long climate simulations, where errors from remote sources can compensate or be conflated with errors arising in the region of interest (Williams et al., 2013). The presence of compensating errors is particularly problematic for the parametrisation development process, where it can be hard to assess the benefits of a new scheme if that scheme no longer compensates for other errors in the model. Localisation of errors in space and time enables identification of the sources of those errors.

In this study, the focus was on using the coarse-graining approach to characterise and understand random errors in a low resolution forecast model: the ECMWF Integrated Forecasting System (IFS) at resolution $\mathsf{T}_\mathsf{L}639$. To illustrate the approach, the widely used Stochastically Perturbed Parametrisation Tendencies (SPPT) scheme (Buizza et al., 1999; Palmer et al., 2009) was taken as an example of a stochastic parametrisation. Despite its widespread use and beneficial impacts in weather and seasonal forecasts, there is limited evidence supporting the theoretical foundations of the scheme. Through characterising the statistics of random model error in low resolution forecasts, this study seeks to indicate whether the theoretical foundations of SPPT are sound.

We find evidence that uncertainty in the parametrised tendencies increases with the magnitude of the parametrised tendency. Multiplicative noise is therefore a good first-order model for uncertainty in the tendencies, providing support for the use of SPPT. To inform the properties of the stochastic perturbation to be used in SPPT, we calculate an optimal multiplicative perturbation as a function of space and time, and assess its statistical properties. The standard deviation of the optimal perturbation is approximately 30% smaller than that used operationally at ECMWF. However, the optimal perturbation also has positive skewness and positive excess kurtosis. This results in a distribution with fatter tails than the Normal distribution used at ECMWF, and would increase the frequency of large-magnitude perturbations. The positive skewness also indicates a reduced likelihood of large negative perturbations which change the sign of the parametrised tendency. At ECMWF, the SPPT perturbations are truncated at plus or minus one (Leutbecher et al., 2017)[5] to ensure SPPT does not invert the sign of the tendency. Using a skewed distribution would reduce the need for this truncation. Overall, it seems the estimated characteristics of SPPT are not very different to those used operationally. It is reassuring to find that the 'top down' approach of tuning the SPPT scheme to produce reliable forecasts is able to find similar parameter values to those found in this 'bottom up' coarse-graining approach.

Importantly, the optimal multiplicative perturbation was found to be correlated in space and time. Spatio-temporally

[5] Note that this truncation reduces the actual standard deviation of the perturbation by around 6% from that stated.

correlated noise has long been recognised as necessary for a skilful stochastic parametrisation scheme (Buizza et al., 1999), and while this can be motivated by theoretical considerations, no coarse-graining studies have presented evidence that this is physically justified. This coarse-graining study provides evidence in support of correlated noise. The decorrelation scales were estimated in space and time and compared to those used in SPPT. The measured leading order pattern is white in both space and time, unlike that used in SPPT. This leads to a measured decorrelation that is initially more rapid than that used operationally. However, large scale spatio-temporal correlations appear in the second and third order patterns. At moderate temporal lags over 12 hours and large spatial lags over 15 degrees, the estimated e showed higher correlations than those used in the operational scheme.

Each SCM simulation is produced independently from its space-time neighbours. Any spatio-temporal correlations in the optimal perturbation must be due to correlated errors in SCM behaviour under different meteorological or boundary conditions. For example, wind shear may introduce convective organisation (e.g. Rotunno et al., 1988) which is not well represented by parametrised models (Liu and Moncrieff, 2001), leading to correlated errors in convective tendencies. The use of spatio-temporally correlated noise in stochastic parametrisations allows for the representation of such errors. A secondary motivation for spatio-temporally correlated noise is to couple unresolved small-scale processes directly with larger scales (Palmer, 2019). This facilitates up-scale energy transfer. It also avoids imposing a hard truncation scale which is inconsistent with the scaling symmetries of the Navier-Stokes equations (Palmer, 2019). Evaluating missing up-scale energy transfer is not possible within this SCM framework, so it is possible that larger correlation scales than those measured could be warranted.

The estimated correlation scales could be used in SPPT. However, it is not clear to what extent the statistics would change for regions other than the Tropical Pacific, and for other time periods. Furthermore, the study is limited by the spatio-temporal domain of the high-resolution simulation, which restricts the ability to estimate correlations on the largest space and time scales. It is possible that the longest space- and time-scales used in operational SPPT actually represent errors due to the coupled ocean-atmosphere system. These errors cannot be assessed in this framework. In any case, such errors would be better represented by including stochasticity into the ocean model (Juricke et al., 2017).

This study also indicates several limitations of SPPT. The optimum perturbation shows substantial variation in its distribution and correlation characteristics between land and sea points. Over land the perturbation shows a marked diurnal cycle, with a substantially higher standard deviation at night than during the day (Figure 7). This reveals a limitation of the SPPT approach. It is known that models struggle to represent the diurnal cycle of convection over land, and predict a peak of precipitation during the morning or early afternoon instead of during the evening as observed (Love et al., 2011; Bechtold et al., 2014). This results in too-small parametrised tendencies at night, such that a multiplicative perturbation must be very large to represent uncertainty in these tendencies. It is possible that an additional, state independent uncertainty representation could be appropriate here, as well as further systematic improvements to parametrisation schemes. Other approaches could also be used to improve the representation of the diurnal cycle in uncertainty. For example, Lock et al. (2019) highlight a recent development to SPPT which accounts for low uncertainty in clear sky radiative tendencies. This leads to differing representations of uncertainty as a function of time of day, due to diurnal variations in the radiative tendencies.

We see further evidence that uncertainty in the IFS is not perfectly multiplicative in Figure 4. Between levels 66 and 52 (555–240 hPa), uncertainty in positive temperature tendencies is markedly non-linear, increasing at a slower rate than linear. By considering the levels at which different parametrisation schemes are active, these features can be attributed to uncertainty in the convection parametrisation. A non-linear relationship between uncertainty in convection and the magnitude of convective tendencies was also highlighted by Shutts and Pallares (2014). Several alternative approaches have been proposed to represent uncertainty in convection parametrisation. Of particular note is the Plant-Craig scheme (Plant and Craig, 2008). The underlying theory was proposed with tropical convection in mind (Craig and Cohen, 2006), but the generality of the theory for convection over land has recently been demonstrated (Rasp et al., 2018). The Plant-Craig approach represents the uncertainty in the convective mass flux as proportional to the square root of its mean. Many convection parametrisation schemes begin by

estimating the mass-flux as a function of stability, convectively available potential energy or moisture budgets, before this is used as an input to an entraining parcel model. While in this study, the uncertainty in the *output* of the convection parametrisation is considered, as opposed to the uncertainty in what is effectively an *input*, there seems to be a consistency between our results and the Plant-Craig approach.

In section 5, the coarse-graining analysis was used to assess three further assumptions made in SPPT. Firstly, we consider the assumption that the SPPT perturbation is constant in the vertical, such that the whole vector tendency is scaled up or down. To assess the justification for this, a separate multiplicative perturbation was fitted at each vertical level, and the correlation calculated between perturbations at different levels. In general, correlations were weak, with the strongest correlations found between levels affected by the same parametrisation schemes. This was particularly evident for the boundary layer scheme, and for night time radiative tendencies. Over those levels, the statistics of the optimal perturbation were also found to be approximately constant. The use of a constant perturbation in height ensures consistency is maintained for schemes which represent transport processes: the whole tendency is scaled up or down to ensure conservation of mass and tracers. It seems that using a constant perturbation for each parametrisation scheme would be sufficient, instead of for the whole vertical column.

The second assumption considered is the use of a single perturbation for all prognostic variable tendencies. As for vertical coherence, this assumption was tested by relaxing the assumption. The statistics of the optimal perturbation fitted to each variable tendency were then considered. The optimal perturbation fitted to q and T was found to be correlated throughout the day. For moist processes, changes in q are associated with a change in T, explaining the correlation in errors for these tendencies. We would not expect this correlation to be perfect as dry processes can also change T. We would also expect the zonal and meridional wind tendencies to be related. However, the correlation between errors in these tendencies is small. It is possible that expressing the wind as stream function and velocity potential would be informative. The statistical properties of the optimal perturbation fitted to U and V tendencies are similar, and the perturbations fitted to T and q are also similar. This motivates the development of stochastic parametrisations separately for thermodynamic and dynamic processes (e.g. Holm, 2015).

The final SPPT assumption considered in this study is that the error is coherent between all physical parametrisation schemes. This was assessed by fitting a separate perturbation to each parametrisation scheme. The correlation between different schemes was generally weak, though the radiation perturbation showed robust diurnally-varying correlations with both the cloud and boundary layer schemes, due to physical relationships between these processes. The perturbations fitted to separate schemes were found to have markedly different statistical characteristics, including magnitude of perturbation and correlation scales. Together with the generally weak correlations between perturbations fitted to different schemes, and the assessment of vertical coherence of perturbations, we find support for the 'independent SPPT' (iSPPT) approach proposed by Christensen et al. (2017b). It is known that the iSPPT approach improves forecast reliability in the IFS, and that it has its largest beneficial impact in the tropics (Christensen et al., 2017b). This study considers a tropical domain, so it is again reassuring that the 'bottom up' approach of estimating instantaneous error statistics reaches the same conclusions as the 'top down' approach of assessing medium-range forecast reliability. However it is possible that if an extra-tropical domain were used for coarse-graining analysis, the SPPT approach would appear more favourable compared to iSPPT.

Finally, even allowing for independent perturbations to each scheme and to each variable, it is not possible to fully account for uncertainty in the parametrised tendencies using a multiplicative approach. Allowing for these generalisations results in an improvement of the fit between modelled and measured error by up to 50% over SPPT, but the average improvement is only 15%. This indicates that there are model errors in the IFS which cannot be represented using multiplicative noise. This motivates the continued development of new stochastic parametrisations to better characterise model error.

While the Cascade simulation has been thoroughly validated against observations, it is possible that the results presented here are sensitive to errors in the truth simulation, or to other details of the truth model. For example, at 4km resolution, the Cascade simulation does not fully resolve convective motions. While assessing this sensitivity is outside the scope of this study, future work will evaluate the sensitivity of the results to the truth model. Other details of the experimentation will also be

considered, including a comparison between different forecast models, domains, and meteorological conditions.

7 | CONCLUSIONS AND RECOMMENDATIONS

We conclude by using the results of this study to suggest some recommendations for stochastic parametrisation. In order of priority:

1. There is evidence that multiplicative noise is a reasonable first-order representation of uncertainty in the IFS parametrised tendencies, providing some support for the use of SPPT. However, the evidence also suggests that the convection scheme in particular could benefit from an alternative approach, such that the uncertainty in the convection tendencies increase at a rate slower than linear.
2. There is some evidence of a physical basis for the spatio-temporal correlations used in stochastic parametrisation schemes such as SPPT, which are therefore not only necessary for pragmatic reasons.
3. The standard deviation of perturbations used in SPPT should be reduced, but the random perturbations should also be drawn from a skewed distribution. This will reduce the need for truncating the distribution to avoid negative perturbations.
4. The iSPPT approach (Christensen et al., 2017b) seems to account for many of the results shown, including.
   a. The correlation between perturbations at different vertical levels is limited to within parametrisations.
   b. A low correlation is found between perturbations fitted to different parametrisation schemes.
   c. Perturbations fitted to different schemes show very different noise characteristics. These different model error characteristics can be specified within the iSPPT approach.
   d. The correlations between perturbations applied to different variables are due to the physical relationship between these variables, as represented in the parametrisation schemes.

   This approach also enables multiplicative noise to be easily replaced by alternative model uncertainty representations for specific schemes.
5. There is some evidence that uncertainty in the thermodynamic (T, q) and dynamic (U, V) tendencies should be treated differently, as they exhibit different error characteristics. The optimal method for achieving this is left for future research.

8 | ACKNOWLEDGEMENTS

The research of H.M.C. was supported by European Research Council grant number 291406 and Natural Environment Research Council grant number NE/P018238/1. The author would like to express her particular thanks to Andrew Dawson (ECMWF) for his extensive input into writing software used in this work. The author would also like to thank Chris Holloway (University of Reading) for providing the Cascade data used here, and advising on its use. Thanks also to Tim Palmer (University of Oxford), Judith Berner (NCAR), Martin Leutbecher and Sarah-Jane Lock (both ECMWF) for helpful advice and input into this work. Thanks to Filip Vana (ECMWF) for support with using the IFS SCM. The author is grateful to the ECMWF OpenIFS project (https://www.ecmwf.int/en/research/projects/openifs) for providing access to the IFS SCM. The coarse-grained data used and produced in this study are archived at the Centre for Environmental Data Analysis (http://catalogue.ceda.ac.uk/uuid/bf4fb57ac7f9461db27dab77c8c97cf2).

A | ESTIMATING THE SPATIAL AND TEMPORAL CORRELATION COEFFICIENTS

The optimal perturbation, $e(\phi, \lambda, t)$ is modelled as a sum over N AR1 processes, separately for each spatial dimension (longitude, ϕ, and latitude, λ) and in time (t). For illustration, consider the time decomposition:

$$e(t) = \sum_{i=1}^{N} X_i(t), \tag{14}$$

$$X_i(t) = \phi_i X_i(t-1) + \sigma_i (1-\phi_i^2)^{\frac{1}{2}} \xi \tag{15}$$

where ϕ_i and σ_i are the lag-1 autocorrelation and standard deviation of the ith scale respectively, and ξ is white noise, $\xi \sim \mathcal{N}(0, 1)$. The X_i are ordered such that the first scale decorrelates the fastest, and the Nth scale decorrelates the slowest. Since the X_i are uncorrelated, the variance and autocorrelation of e can be written:

$$\sigma_e^2 = \sum_{i=1}^{N} \sigma_i^2 \tag{16}$$

$$\rho_e = \frac{\sum_{i=1}^{N} \sigma_i^2 \phi_i^\tau}{\sum_{i=1}^{N} \sigma_i^2} \tag{17}$$

To select the optimal number of scales, the log of the autocorrelation of e is plotted, revealing a number, N, of straight-line sections. For large τ, we assume $\phi_i << \phi_N, i \neq N$, and approximate the autocorrelation as

$$\rho_e = \frac{\sigma_N^2 \phi_N^\tau}{\sum_{i=1}^{N} \sigma_i^2}, \tag{18}$$

In this way, the variance ratio, $\frac{\sigma_N^2}{\sum_{i=1}^{N} \sigma_i^2}$, and autocorrelation, ϕ_N, of the largest scale can be estimated from the graph of the log of the autocorrelation function at large τ. The modelled ρ_N is subtracted from ρ_e, and the method repeated for each of the next slowest scales in turn.

REFERENCES

P. Bechtold, N. Semane, P. Lopez, J.-P. Chaboureau, A. Beljaars, and N. Bormann. Representing equilibrium and nonequilibrium convection in large-scale models. *J. Atmos. Sci.*, 71(2):734–753, 2014.

J. Berner, K. R. Fossell, S.-Y. Ha, J. P. Hacker, and C. Snyder. Increasing the skill of probabilistic forecasts: Understanding performance improvements from model-error representations. *Mon. Weather Rev.*, 143:1295–1320, 2015.

J. Berner, U. Achatz, L. Batte, A. De La Cámara, H. Christensen, M Colangeli, D. R. Coleman, D. Crommelin, S. Dolaptchiev, C. L. E. Franzke, P. Friederichs, P. Imkeller, H. Järvinen, S. Juricke, V. Kitsios, F. Lott, V. Lucarini, S. Mahajan, T. N. Palmer, C. Penland, J.-S. Von Storch, M. Sakradzija, M. Weniger, A. Weisheimer, P. D. Williams, and Y.-I. Yano. Stochastic parameterization: Towards a new view of weather and climate models. *B. Am. Meteorol. Soc.*, 98:565–588, 2017.

J. Bessac, A. Monahan, H. M. Christensen, and N. Weitzel. Stochastic parameterization of subgrid-scale velocity enhancement of sea surface fluxes. *Mon. Weather Rev.*, 147:1447–1469, 2019.

Thomas Bolton and Laure Zanna. Applications of Deep Learning to Ocean Data Inference and Subgrid Parameterization. *JAMES*, 11 (1):376–399, 2019. doi: 10.1029/2018MS001472.

F. Bouttier, B. Vié, O Nuissier, and L. Raynaud. Impact of stochastic physics in a convection-permitting ensemble. *Mon. Weather Rev.*, 140(11):3706–3721, 2012.

N. E. Bowler, A. Arribas, K. R. Mylne, K. B. Robertson, and S. E. Beare. The MOGREPS short-range ensemble prediction system. *Q. J. Roy. Meteor. Soc.*, 134(632):703–722, 2008.

R. Buizza. 25 years of ensemble forecasting at ecmwf. *ECMWF Newsletter*, 153:20–31, 2017.

R. Buizza, M. Miller, and T. N. Palmer. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Q. J. Roy. Meteor. Soc.*, 125(560):2887–2908, 1999.

M. Charron, G. Pellerin, L. Spacek, P. Houtekamer, N. Gagnon, H. Mitchell, and L. Michelin. Toward random sampling of model error in the canadian ensemble prediction system. *Mon. Weather Rev.*, 138:1877–1901, 2010.

H. M. Christensen, I. M. Moroz, and T. N. Palmer. Stochastic and perturbed parameter representations of model uncertainty in convection parameterization. *J. Atmos. Sci.*, 72(6):2525–2544, 2015.

H. M. Christensen, J. Berner, D. Coleman, and T. N. Palmer. Stochastic parametrisation and the El Niño-Southern Oscillation. *J. Climate*, 30(1):17–38, 2017a.

H. M. Christensen, S.-J. Lock, I. M. Moroz, and T. N. Palmer. Introducing independent patterns into the stochastically perturbed parametrization tendencies (SPPT) scheme. *Q. J. Roy. Meteor. Soc.*, 143(706):2168–2181, 2017b.

H. M. Christensen, A. Dawson, and C. Holloway. Forcing files for the Integrated Forecasting System (IFS) Single Column Model (SCM) over Indian Ocean/Tropical Pacific derived from a 10-day high resolution simulation. Dataset record, Centre for Environmental Data Analysis, 2018a. http://catalogue.ceda.ac.uk/uuid/bf4fb57ac7f9461db27dab77c8c97cf2.

H. M. Christensen, A. Dawson, and C. E. Holloway. Forcing single column models using high-resolution model simulations. *JAMES*, 10, 2018b. doi: https://doi.org/10.1029/2017MS001189.

F. C. Cooper and L. Zanna. Optimisation of an idealised ocean model, stochastic parameterisation of sub-grid eddies. *Ocean Model.*, 88:38–53, 2015.

F. Couvreux, R. Roehrig, C. Rio, M. P. Lefebvre, M. Caian, T. Komori, S. Derbyshire, F. Guichard, F. Favot, F. D'Andrea, P. Bechtold, and P. Gentine. Representation of daytime moist convection over the semi-arid Tropics by parametrizations used in climate and meteorological models. *Q. J. Roy. Meteor. Soc.*, 141(691):2220–2236, 2015. doi: 10.1002/qj.2517.

G. C. Craig and B. G. Cohen. Fluctuations in an equilibrium convective ensemble. part i: Theoretical formulation. *J. Atmos. Sci.*, 63(8): 1996–2004, 2006.

P. Davini, J. von Hardenberg, S. Corti, H. M. Christensen, S. Juricke, A. Subramanian, P. A. G. Watson, A. Weisheimer, and T. N. Palmer. Climate sphinx: evaluating the impact of resolution and stochastic physics parameterisations in the ec-earth global climate model. *GMD*, 10(3):1383–1402, 2017.

J. Dorrestijn, D. T. Crommelin, J. A. Biello, and S. J. Böing. A data-driven multi-cloud model for stochastic parametrization of deep convection. *Phil. Trans. R. Soc. A*, 371(1991), 2013.

J. Dorrestijn, D. T. Crommelin, A. P. Siebesma, H. J. J. Jonker, and C. Jakob. Stochastic parameterization of convective area fractions with a multicloud model inferred from observational data. *J. Atmos. Sci.*, 72:854–869, 2015.

Pierre Gentine, Alan K. Betts, Benjamin R. Lintner, Kirsten L. Findell, Chiel C. van Heerwaarden, Alexandra Tzella, and Fabio D'Andrea. A Probabilistic Bulk Model of Coupled Mixed Layer and Convection. Part I: Clear-Sky Case. *J. Atmos. Sci.*, 70(6): 1543–1556, 2013. doi: 10.1175/jas-d-12-0145.1.

Francoise Guichard, J. C. Petch, J. L. Redelsperger, P. Bechtold, J. P. Chaboureau, S. Cheinet, W. Grabowski, H. Grenier, C. G. Jones, M. Köhler, J. M. Piriou, R. Tailleux, and M. Tomasini. Modelling the diurnal cycle of deep precipitating convection over land with cloud-resolving models and single-column models. *Q. J. Roy. Meteor. Soc.*, 130 C(604):3139–3172, 2004. doi: 10.1256/qj.03.145.

E. Hawkins and R. Sutton. Decadal predictability of the atlantic ocean in a coupled gcm: Forecast skill and optimal perturbations using linear inverse modeling. *Journal of Climate*, 22(14):3960–3978, 2009. doi: 10.1175/2009JCLI2720.1.

R. Heinze, A. Dipankar, C. C. Henken, C. Moseley, O. Sourdeval, S. Trömel, X. Xie, P. Adamidis, F. Ament, H. Baars, C. Barthlott, A. Behrendt, U. Blahak, S. Bley, S. Brdar, M. Brueck, S. Crewell, H. Deneke, P. Di Girolamo, R. Evaristo, J. Fischer, C. Frank, P. Friederichs, T. Göcke, K. Gorges, L. Hande, M. Hanke, A. Hansen, H.-C. Hege, C. Hoose, T. Jahns, N. Kalthoff, D. Klocke, S. Kneifel, P. Knippertz, A. Kuhn, T. van Laar, A. Macke, V. Maurer, B. Mayer, C. I. Meyer, S. K. Muppa, R. A. J. Neggers, E. Orlandi, F. Pantillon, B. Pospichal, N. Röber, L. Scheck, A. Seifert, P. Seifert, F. Senf, P. Siligam, C. Simmer, S. Steinke, B. Stevens, K. Wapler, M. Weniger, V. Wulfmeyer, G. Zängl, D. Zhang, and J. Quaas. Large-eddy simulations over Germany using ICON: a comprehensive evaluation. *Q. J. Roy. Meteor. Soc.*, 143(702):69–100, 2017.

C. E. Holloway, S. J. Woolnough, and G. M. S. Lister. Precipitation distributions for explicit versus parametrized convection in a large-domain high-resolution tropical case study. *Q. J. Roy. Meteor. Soc.*, 138(668):1692–1708, 2012.

C. E. Holloway, S. J. Woolnough, and G. M. S. Lister. The effects of explicit versus parameterized convection on the MJO in a large-domain high-resolution tropical case study. part I: Characterization of large-scale organization and propagation. *J. Atmos. Sci.*, 70 (5):1342–1369, 2013.

Christopher E. Holloway, Steven J. Woolnough, and Grenville M. S. Lister. The Effects of Explicit versus Parameterized Convection on the MJO in a Large-Domain High-Resolution Tropical Case Study. Part II: Processes Leading to Differences in MJO Development. *J. Atmos. Sci.*, 72(7):2719–2743, 2015. doi: 10.1175/JAS-D-14-0308.1.

Darryl D. Holm. Variational principles for stochastic fluid dynamics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2176), 2015. doi: 10.1098/rspa.2014.0963.

J. T. Houghton, G. J. Jenkins, and J. J. Ephraums, editors. *Report prepared for Intergovernmental Panel on Climate Change by Working Group I*. Cambridge University Press, 1990.

Stephan Juricke, Tim N. Palmer, and Laure Zanna. Stochastic subgrid-scale ocean mixing: Impacts on low-frequency variability. *J. Climate*, 30(13):4997–5019, 2017. doi: 10.1175/JCLI-D-16-0539.1.

B. Khouider, J. Biello, and A. J. Majda. A stochastic multicloud model for tropical convection. *Commun. Math. Sci.*, 8(1):187–216, 2010.

B.P. Kirtman, D. Min, J. M. Infanti, J. L. Kinter, D. A. Paolino, Q. Zhang, H. van den Dool, S. Saha, M. P. Mendez, E. Becker, P. Peng, P. Tripp, J. Huang, D. G. DeWitt, M. K. Tippett, A. G. Barnston, S. Li, A. Rosati, S. D. Schubert, M. Rienecker, M. Suarez, Z. E. Li, J. Marshak, Y. Lim, J. Tribbia, K. Pegion, W. J. Merryfield, B. Denis, and E. F. Wood. The north american multimodel ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *B. Am. Meteorol. Soc.*, 95:585–601, 2014.

M. Leutbecher, S.-J. Lock, P. Ollinaho, S. T. K. Lang, G. Balsamo, P. Bechtold, M. Bonavita, H. M. Christensen, M. Diamantakis, E. Dutra, S. English, M. Fisher, R. M. Forbes, J. Goddard, T. Haiden, R. J. Hogan, S. Juricke, H. Lawrence, D. MacLeod, L. Magnusson, S. Malardel, S. Massart, I. Sandu, P. K. Smolarkiewicz, A. Subramanian, F. Vitart, N. Wedi, and A. Weisheimer. Stochastic representations of model uncertainties at ecmwf: State of the art and future vision. *Quarterly Journal of the Royal Meteorological Society*, 143(707):2315–2339, 2017.

Changhai Liu and Mitchell W. Moncrieff. Cumulus Ensembles in Shear: Implications for Parameterization. *J. Atmos. Sci.*, 58(18): 2832–2842, 2001. doi: 10.1175/1520-0469(2001)058<2832:ceisif>2.0.co;2.

S.-J. Lock, S. T. K. Lang, M. Leutbecher, R. J. Hogan, and F. Vitart. Treatment of model uncertainty from radiation by the stochastically perturbed parametrization tendencies (SPPT) scheme and associated revisions in the ECMWF ensembles. *Q. J. Roy. Meteor. Soc.*, 2019.

B. S. Love, A. J. Matthews, and G. M. S. Lister. The diurnal cycle of precipitation over the Maritime Continent in a high-resolution atmospheric model. *Q. J. Roy. Meteor. Soc.*, 137(657):934–947, 2011. doi: 10.1002/qj.809.

A. H. Murphy. The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situatiuon. *Mon. Weather Rev.*, 105(7):803–816, 1977.

Ji Nie and Zhiming Kuang. Responses of Shallow Cumulus Convection to Large-Scale Temperature and Moisture Perturbations: A Comparison of Large-Eddy Simulations and a Convective Parameterization Based on Stochastically Entraining Parcels. *J. Atmos. Sci.*, 69(6):1936–1956, 2012. doi: 10.1175/jas-d-11-0279.1.

P. Ollinaho, P. Bechtold, M. Leutbecher, M. Laine, A. Solonen, H. Haario, and H. Järvinen. Parameter variations in prediction skill optimization at ecmwf. *Nonlinear Proc. Geoph.*, 20(6):1001–1010, 2013.

P. Ollinaho, S.-J. Lock, M. Leutbecher, P. Bechtold, A. Beljaars, A. Bozzo, R. M. Forbes, T. Haiden, R. J. Hogan, and I. Sandu. Towards process-level representation of model uncertainties: Stochastically perturbed parametrisations in the ECMWF ensemble. *Q. J. Roy. Meteor. Soc.*, 143(702):408–422, 2017.

T. N. Palmer. Stochastic weather and climate models. *Nature Reviews Physics*, 1(7):463–471, 2019.

T. N. Palmer, R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer. Stochastic parametrization and model uncertainty. Tech. Mem. 598, European Centre for Medium-Range Weather Forecasts, Shinfield park, Reading, 2009. URL `http://www.ecmwf.int/en/elibrary/technical-memoranda`.

R. S. Plant and G. C. Craig. A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci.*, 65(1): 87–104, 2008.

Pier Gian Luca Porta Mana and Laure Zanna. Toward a stochastic parameterization of ocean mesoscale eddies. *Ocean Modelling*, 79: 1–20, 2014. doi: 10.1016/j.ocemod.2014.04.002.

Stephan Rasp, Tobias Selz, and George C. Craig. Variability and Clustering of Midlatitude Summertime Convection: Testing the Craig and Cohen Theory in a Convection-Permitting Ensemble with Stochastic Boundary Layer Perturbations. *J. Atmos. Sci.*, 75(2): 691–706, 2018. ISSN 0022-4928. doi: 10.1175/JAS-D-17-0258.1.

M. J. Rodwell, S. T. K. Lang, B. Ingleby, N. Bormann, E. Hólm, F. Rabier, D. S. Richardson, and M. Yamaguchi. Reliability in ensemble data assimilation. *Q. J. Roy. Meteor. Soc.*, 142(694):443–454, 2016.

David M. Romps and Zhiming Kuang. Nature versus Nurture in Shallow Convection. *J. Atmos. Sci.*, 67(5):1655–1666, 2010. doi: 10.1175/2009jas3307.1.

Richard Rotunno, Joseph B. Klemp, and Morris L. Weisman. A Theory for Strong, Long-Lived Squall Lines. *J. Atmos. Sci.*, 45(3): 463–485, 1988. doi: 10.1175/1520-0469(1988)045<0463:atfsll>2.0.co;2.

J. Rougier, D. M. H. Sexton, J. M. Murphy, and D. Stainforth. Analyzing the climate sensitivity of the HadSM3 climate model using ensembles from different but related experiments. *J. Climate*, 22:3540–3557, 2009.

C. Sanchez, K. D. Williams, and M. Collins. Improved stochastic physics schemes for global weather and climate models. *Q. J. Roy. Meteor. Soc.*, 142(694):147–159, 2016.

M. Satoh, H. Tomita, H. Yashiro, H. Miura, C. Kodama, T. Seiki, A. T. Noda, Y. Yamada, D. Goto, M. Sawada, T. Miyoshi, Y. Niwa, M. Hara, T. Ohno, S. Iga, T. Arakawa, T. Inoue, and H. Kubokawa. The non-hydrostatic icosahedral atmospheric model: description and development. *Prog. Earth Planet. Sc.*, 1(18):1–32, 2014.

Masaki Satoh, Hirofumi Tomita, Hisashi Yashiro, Yoshiyuki Kajikawa, Yoshiaki Miyamoto, Tsuyoshi Yamaura, Tomoki Miyakawa, Masuo Nakano, Chihiro Kodama, Akira T. Noda, Tomoe Nasuno, Yohei Yamada, and Yoshiki Fukutomi. Outcomes and challenges of global high-resolution non-hydrostatic atmospheric simulations using the K computer. *Progress in Earth and Planetary Science*, 4(13):1–24, 2017. doi: 10.1186/s40645-017-0127-8.

J. Schalkwijk, H. J. J. Jonker, A. P. Siebesma, and E. Van Meijgaard. Weather forecasting using GPU-based large-eddy simulations. *B. Am. Meteorol. Soc.*, 96(5):715–723, 2015.

G. J. Shutts and A. C. Pallares. Assessing parametrization uncertainty associated with horizontal resolution in numerical weather prediction models. *Phil. Trans. R. Soc. A*, 372(2018), 2014.

G. J. Shutts and T. N. Palmer. Convective forcing fluctuations in a cloud-resolving model: Relevance to the stochastic parameterization problem. *J. Climate*, 20(2):187–202, 2007.

D. A Stainforth, T. Aina, C. Christensen, M. Collins, N. Faull, D. J. Frame, J. A. Kettleborough, S. Knight, A. Martin, J. M. Murphy, C. Piani, D. Sexton, L. A. Smith, R. A Spicer, A. J. Thorpe, and M. R Allen. Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, 433(7024):403–406, 2005.

D. J. Stensrud, J.-W. Bao, and T. T. Warner. Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Weather Rev.*, 128(7):2077–2107, 2000.

Bjorn Stevens, Masaki Satoh, Ludovic Auger, Joachim Biercamp, Christopher S Bretherton, Xi Chen, Peter Düben, Falko Judt, Marat Khairoutdinov, Daniel Klocke, Chihiro Kodama, Luis Kornblueh, Shian-jiann Lin, Philipp Neumann, William M Putman, Niklas Röber, Ryosuke Shibuya, Benoit Vanniere, Pier Luigi Vidale, Nils Wedi, and Linjiong Zhou. Open Access DYAMOND : the DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains. *Progress in Earth and Planetary Science*, 2019.

K. Strømmen, H. M. Christensen, J. Berner, and T. N. Palmer. The impact of stochastic parametrisations on the representation of the asian summer monsoon. *J. Climate*, 50(5–6):2269–2282, 2018.

K. Strømmen, H. M. Christensen, D. MacLeod, S. Juricke, and T. N. Palmer. Introducing the probabilistic earth-system model: Examining the impact of stochasticity in ec-earth v3.2. *Geosci. Model Dev.*, 2019. In Press.

K. Sušelj, T. F. Hogan, and J. Teixeira. Implementation of a stochastic eddy-diffusivity/mass-flux parameterization into the navy global environmental model. *Weather Forecast.*, 29:1374–1390, 2014.

Yong Wang, Guang J. Zhang, and George C. Craig. Stochastic convective parameterization improving the simulation of tropical precipitation variability in the NCAR CAM5. *Geophys. Res. Let.*, 43(12):6612–6619, 2016. doi: 10.1002/2016GL069818.

A. Weisheimer, T. N. Palmer, and F. J. Doblas-Reyes. Assessment of representations of model uncertainty. *Geophys. Res. Let.*, 38, 2011.

A. Weisheimer, S. Corti, T. N. Palmer, and F. Vitart. Addressing model error through atmospheric stochastic physical parametrizations: impact on the coupled ECMWF seasonal forecasting system. *Phil. Trans. R. Soc. A*, 372, 2014.

Keith D. Williams, A. Bodas-Salcedo, M. Déqué, S. Fermepin, B. Medeiros, M. Watanabe, C. Jakob, S. A. Klein, C. A. Senior, and D. L. Williamson. The transpose-AMIP II experiment and its application to the understanding of southern ocean cloud biases in climate models. *J. Climate*, 26(10):3258–3274, 2013. doi: 10.1175/JCLI-D-12-00429.1.

H. Yonehara and M Ujiie. A stochastic physics scheme for model uncertainties in the JMA one-week ensemble prediction system. Technical Report 41, CAS/JSC WGNE Research Activities in Atmospheric and Oceanic Modelling, 2011.