Introducing Independent Patterns into the Stochastically Perturbed Parametrisation Tendencies (SPPT) scheme

H. M. Christensen,[a,b*] S.-J Lock[c], I. M. Moroz[d] and T. N. Palmer[a]

[a]*Atmospheric, Oceanic and Planetary Physics, University of Oxford, UK*

[b]*National Center for Atmospheric Research, Boulder, Colorado, USA*

[c]*European Centre for Medium -Range Weather Forecasts, Shinfield Park, UK*

[d]*Mathematical Institute, University of Oxford, UK*

*Correspondence to: H. M. Christensen, NCAR, P.O.Box 3000, Boulder, CO 80303, USA. E-mail: hannahc@ucar.edu

The Stochastically Perturbed Parametrisation Tendencies (SPPT) scheme is used at weather and climate forecasting centres worldwide to represent model uncertainty that arises from simplifications involved in the parametrisation process. It uses spatio-temporally correlated multiplicative noise to perturb the sum of the parametrised tendencies. However, SPPT does not distinguish between different parametrisation schemes, which do not necessarily have the same error characteristics. A generalisation to SPPT is proposed, whereby the tendency from each parametrisation scheme can be perturbed using an independent stochastic pattern. This acknowledges that the forecast errors arising from different parametrisations are not perfectly correlated. Two variations of this 'independent SPPT' (iSPPT) approach are tested in the Integrated Forecasting System (IFS). The first perturbs all parametrised tendencies independently while the second groups tendencies before perturbation. The iSPPT schemes lead to statistically significant improvements in forecast reliability in the tropics in medium range weather forecasts. This improvement can be attributed to a large, beneficial increase in ensemble spread in regions with significant convective activity. The iSPPT schemes also lead to improved forecast skill in the extra tropics for a set of cases in which the synoptic initial conditions were more likely to result in European 'forecast busts'. Longer 13-month simulations are also considered to indicate the effect of iSPPT on the mean climate of the IFS.

Prepared using qjrms4.cls

1. Introduction

Operational weather forecasting centres routinely produce probabilistic medium range weather forecasts. These consist of an ensemble forecast, in which the ensemble is designed to represent the uncertainty involved in making the forecast (Palmer 2001). This is of great benefit to the end user, who can assess the probabilities and risks associated with the forecast. They can make cost-effective decisions based on the information in the forecast and their acceptable level of risk (Murphy 1977; Zhu et al. 2002). In order to be useful, such probabilistic forecasts must be *reliable*: the ensemble should simulate all sources of uncertainty in the forecast, such that the verification behaves as if drawn from the forecast distribution (Anderson 1997; Wilks 2006).

There are two main sources of uncertainty in weather forecasts: *initial condition uncertainty* and *model uncertainty*. Initial condition uncertainty arises due to the limited availability of observational data combined with errors in the retrieval process. This results in errors in the initial conditions used for the forecast. This uncertainty is represented by perturbing the initial conditions of the different ensemble forecast members, for example by using a singular vector approach (see e.g. Leutbecher and Lang 2014) or by producing an ensemble of data assimilations (Isaksen et al. 2010).

The other key source of uncertainty in weather forecasts, model uncertainty, arises due to the simplifications and approximations used when formulating the model (Leutbecher and Palmer 2008). Stochastic parametrisation schemes are widely used to represent model uncertainty in weather forecasts. Whereas a deterministic parametrisation scheme represents the most likely effect of sub-grid scale processes on the resolved scale state, a stochastic parameterisation scheme represents the impact of a possible realisation of the sub-grid scale motion (Palmer et al. 2009). An ensemble of forecasts is produced, in which each perturbed ensemble member sees a different, but equally likely, stochastic forcing. Such schemes have been shown to significantly improve the reliability of medium-range and seasonal forecasts (Buizza et al. 1999; Weisheimer et al. 2014; Sanchez et al. 2016; Leutbecher et al. 2017).

Designing new stochastic schemes has been the target of much innovative research over the last decade. A focus has been on developing physically motivated approaches, whereby the sources of uncertainty in a particular parametrisation scheme are identified and addressed (Khouider et al. 2003; Plant and Craig 2008; Bengtsson et al. 2013; Christensen et al. 2015; Kober and Craig 2016). While many of these schemes show desirable features when tested in weather and climate models, they are yet to be implemented operationally. Instead, operational centres often take a holistic approach, whereby uncertainty from a number of parametrised processes are represented using a single scheme: the Stochastically Perturbed Parametrisation Tendencies (SPPT) scheme. SPPT was developed, and is used operationally at, the European Centre for Medium Range Weather Forecasts (ECMWF) in their Integrated Forecasting System (IFS) model (Buizza et al. 1999; Palmer et al. 2009). SPPT is one of the most extensively used stochastic parametrisation schemes, and is used at operational forecasting and research centres worldwide, including in the models of the U.K. Met Office (Sanchez et al. 2016) and Japan Meteorological Agency (Yonehara and Ujiie 2011), in the Application of Research to Operations at Mesoscale (AROME) model (Bouttier et al. 2012) and the Weather Research and Forecasting (WRF) model (Berner et al. 2015), and in the Community Earth System Model (Berner et al. 2016) and EC-Earth (Davini et al. 2016) climate models.

SPPT is a multiplicative noise scheme which perturbs the sum of the parametrised tendencies with spatio-temporally correlated multiplicative noise. By representing the uncertainty in all parametrised processes using a single stochastic scheme, SPPT ensures a consistent representation of model uncertainty in all sub-grid processes, and avoids the duplication of model error representation across multiple parametrisation schemes. It is also efficient and easy to implement, as it is used in conjunction with existing parametrisation schemes. It has a beneficial impact on the model, improving the reliability of medium-range and seasonal forecasts by reducing biases in the ensemble forecasts and improving forecast reliability (Buizza et al. 1999; Palmer et al. 2009; Weisheimer et al. 2014). SPPT also leads to an improvement of the modelled climate (Sanchez et al. 2016; Christensen et al. 2017; Watson et al. 2017), reducing biases in climate mean and variability.

A key principle behind the formulation of SPPT is to perturb the entire parametrised tendency together and not the tendencies associated with individual schemes. This is because the individual parametrised tendencies (e.g. from the cloud, radiation, convection and boundary layer schemes) are not independent of each other, but are in a state of partial balance (Beljaars et al. 2004). Perturbing the processes together maintains this balance.

Despite its widespread use, SPPT has several limitations, which are discussed in detail in Leutbecher et al. (2017). SPPT assumes the uncertainty in the parametrised tendencies is proportional to the total tendency, when it is likely that there are uncertainties that cannot be represented in this way. For example, this method cannot represent uncertainty in the vertical distribution of convective heating. SPPT does not perturb fluxes at the surface or top of atmosphere, introducing inconsistencies between the perturbed tendencies in a column and these fluxes. SPPT also imposes large spatio-temporal correlation scales when perturbing tendencies to represent the correlation of model uncertainties in space and time, but these correlation scales have not been measured, and are not tied to physical processes. Furthermore, in perturbing the sum of the physics tendencies, SPPT attributes the same error characteristics to all parametrisation schemes and assumes that the errors are perfectly correlated with each other. A recent coarse graining study by Shutts and Pallares (2014) measured the standard deviation of the error for each physics tendency as a function of the parametrised tendency. The study revealed that the different schemes do, indeed, have very different error characteristics, with the uncertainty in the cloud and convection tendencies being much larger than the radiation tendency. Additionally, the standard deviation of the cloud and convection tendencies were shown to be proportional to the square root of the parametrised tendency, unlike the representation in SPPT, which represents the standard deviation as proportional to the total parametrised tendency.

It is possible that maintaining the partial balance between schemes in SPPT is overly constrained, and representing the uncertainty in each individual parametrisation scheme separately could be justified. In the UK Met Office Unified Model, the SPPT-style scheme (the Stochastic Perturbation of Tendencies (SPT) scheme) goes some way towards recognising these differences in the error characteristics from different schemes: the multiplicative noise has a different standard deviation for different processes (Sanchez et al. 2016).

This paper proposes an alternative form of SPPT, which we call "independent SPPT" (iSPPT). In this paradigm, the uncertainty in each parametrisation tendency can be represented independently from the others. Whereas in SPPT the uncertainty is assumed to be proportional to the total net tendency, iSPPT represents the errors from the different parametrisation schemes as being uncorrelated, so that the uncertainty in the forecast is proportional to the tendencies from individual processes. iSPPT is an efficient representation of uncertainty from different parametrisation schemes. For example, it uses far fewer stochastic patterns than alternative stochastically perturbed parameter approaches (Bowler et al. 2008; Ollinaho et al. 2016). This reduces the computational burden of the approach and results in a scheme that is easier to tune and maintain as the deterministic model is updated. Through iSPPT, the perturbations to one parametrisation scheme can be 'turned off', allowing for an alternative representation of uncertainty in that scheme, for example, stochastic perturbation of uncertain parameters (Bowler et al. 2008; Ollinaho et al. 2016) or a stochastic scheme that targets one specific process, e.g. convection (e.g Plant and Craig 2008). By perturbing the tendencies from a single scheme, iSPPT also allows for sensitivity tests, whereby the growth of uncertainties associated with a particular atmospheric process can be explored.

We recognise that the different physics schemes in the IFS have been developed in tandem, and are called sequentially in the IFS to maintain a balance. For example, the cloud and convection schemes are closely linked, as the convection scheme represents the warming due to moist convection, while the cloud scheme calculates the cooling due to evaporation of detrained cloudy air. On the one hand, if the two schemes have been closely tuned to each other, potentially with compensating errors, decoupling the two schemes by using independent perturbations could reduce the forecast skill and introduce errors (and even numerical instability) into the forecasts. On the other hand, the close relationship means that the net tendency from the two schemes is smaller than each individual tendency, and SPPT attributes a correspondingly small level of uncertainty. iSPPT could be beneficial in this case, as it is able to represent the potentially large errors in each individual tendency.

As indicated by Shutts and Pallares (2014), it is unlikely that the uncertainties in different processes are precisely correlated as assumed under SPPT. A new coarse graining study is currently underway to test the assumptions made in the SPPT approach, including the correlation of errors arising from different parametrisation schemes in the IFS. Preliminary results indicate that the errors in the different parametrised processes are not perfectly correlated (H. Christensen, pers. comm.), providing physical motivation for the iSPPT scheme. In this paper we deliberately take the inverse approach to this coarse graining study and consider which of the SPPT and iSPPT schemes perform best when implemented in a weather or climate model.

In Section 2 we outline the operational weather forecasting model used in this study, the IFS, and the implementation of the operational SPPT and new iSPPT schemes. In Section 3 we describe the impact of the scheme in medium-range weather forecasts, focusing on the skill of the forecasts and the source of improved forecast reliability in the tropics. In Section 4 we consider the impact of the iSPPT schemes on a case study in which the synoptic conditions increase the likelihood of a 'forecast bust' occurring over Europe. In Section 5 we consider the impact of the iSPPT schemes on the mean model climate. Finally the results are discussed and conclusions are drawn in Section 6.

2. Extending SPPT to permit independent random patterns

2.1. *Standard SPPT*

The SPPT scheme, used operationally in the IFS model, represents uncertainties in the forecast that arise due to the simplified representations of parametrised atmospheric processes — the so-called 'physics' parametrisations. It does so by modifying the net tendencies from the physics parametrisations, $\mathbf{P}$, with multiplicative noise, r, to generate a perturbed physics tendency, $\hat{\mathbf{P}}$, such that:

$$\hat{\mathbf{P}} = (1 + \mu r)\,\mathbf{P} = (1 + \mu r)\sum_i \mathbf{p}_i, \qquad (1)$$

where $\mathbf{p}_i(z)$ is the vector of tendencies for the prognostic model variables (winds, temperature and humidity) that results from the i-th physics parametrisation; and $\mu \in [0, 1]$ is a tapering function that can reduce the noise to zero in parts of the model domain (explained more later).

As described in Palmer et al. (2009), the noise term r is a 2D random field in spectral space with prescribed standard deviation, temporal and spatial correlation lengths. The field uses random numbers from a Gaussian distribution with zero mean and unit variance. Each spectral coefficient is auto-correlated in space in the form of a Gaussian on the sphere (following Weaver and Courtier 2001); and the time correlations arise via a first-order auto-regressive process. The prescribed standard deviation defines the variance

j	σ_j	L_j (km)	τ_j (days)
1	0.52	500	0.25
2	0.18	1000	3
3	0.06	2000	30

Table 1. SPPT parameter values for the $j = 1:3$ random fields that comprise the 3-scale pattern used in the IFS: Standard deviation σ_j, horizontal correlation length L_j, time decorrelation scale τ_j.

of the resultant (total wavenumber) field, indicating the amplitude of the perturbations.

In the IFS, r is a three-scale pattern, composed from a linear combination of three independent random patterns that each describe different correlation scales (described in Shutts et al. 2011, and summarised here in Table 1). The three-scale pattern is dominated by the shortest correlation scales, which are associated with mesoscale processes. Uncertainties associated with longer-range weather features are accounted for by smaller amplitude contributions from the larger correlation scales.

The (3-scale) random pattern r is transformed from spectral to grid-point space using the spectral transform, and the resulting grid-point values are clipped to ensure $\hat{\mathbf{P}}$ retains the same sign as $\mathbf{P}$, such that $(1 + \mu r) \in [0, 2]$.

Each grid-point value from r is used according to (1) for all perturbations throughout the grid-column, thereby retaining the vertical structure that results from the physics parametrisations. The (optional) tapering function (μ) is applied in order to reduce the perturbations smoothly to zero in the boundary layer (to avoid exciting numerical instabilities), and in the stratosphere (where $\mathbf{P}$ is dominated by well-constrained clear-skies radiative transfer).

The scheme described above — hereafter, "standard SPPT" — takes no account of differences in the uncertainties associated with different physics parametrisation schemes: the uncertainty attributed to the physics schemes is simply proportional to the total tendency ($\mathbf{P}$). Thus, the variance of the resulting perturbed tendencies is given by

$$\hat{\sigma}^2 = \sigma^2 \mathbf{P}^2 = \sigma^2 \left(\sum_i \mathbf{p}_i \right)^2 , \qquad (2)$$

where σ is the standard deviation of the random pattern r. The variance of the perturbed tendencies can be seen to vanish as $\sum_i \mathbf{p}_i \rightarrow 0$. In other words, large uncertainty is attributed when the physics parametrisations yield a large net tendency. However, a small uncertainty is assumed when the net physics tendency is small, even if individual physics processes have large contributions.

2.2. *Independent SPPT (iSPPT)*

In this work, we propose extending standard SPPT to enable each physics process, $\mathbf{p}_i$, to be independently perturbed with multiplicative noise:

$$\hat{\mathbf{P}} = \sum_i (1 + \mu r_i)\, \mathbf{p}_i , \qquad (3)$$

where each random pattern, r_i, is generated using a different seed and evolves independently. The properties of the distributions that determine each r_i (standard deviation, spatial and temporal correlations) are individually specified, and so can differ from one random pattern to another. Under this approach, the variance of the total perturbed tendencies becomes

$$\hat{\sigma}^2 = \sum_i \left(\sigma_i^2 \mathbf{p}_i^2 \right) . \qquad (4)$$

Note that the same perturbation is applied to all state variables — the 'independence' refers to independent perturbation of tendencies from different parametrisation schemes. This is in contrast to the earliest version of SPPT described in Buizza et al. (1999), which independently perturbs the tendencies in different state variables.

This revised scheme — "iSPPT" — continues to attribute large uncertainty when the net physics tendencies are large; but in addition, large tendencies from different physics schemes which act in opposite directions can each be attributed large uncertainty. Indeed, the variance of the perturbed tendencies can only vanish when all physics parametrisations yield a small tendency: $\mathbf{p}_i \rightarrow 0, \forall i$.

At this stage, we introduce a naming convention for iSPPT. Numbers are used to indicate the indices for the pattern applied to tendencies, while the ordering indicates which of the 6 physics parametrisation schemes is perturbed with that pattern: 1st=radiation (RDTN), 2nd=turbulence, vertical mixing and orographic drag (TGWD), 3rd=convection (CONV), 4th=large scale water (cloud) processes (LSWP), 5th=non-orographic drag (NOGW), and 6th=methane oxidation (MOXI), i.e. "iSPPT [123456]" indicates that six independent patterns are applied, one to each of the six physics processes.

In the forecast model, during a model integration, the physics parametrisation schemes are called sequentially, and the tendencies that are output from one scheme become the input for the next. In the design of iSPPT, there is a choice over when to apply the perturbations to the tendencies from each physics scheme: at the end of all the calls to physics schemes (consistent with the application of the perturbations in standard SPPT) or as they are output from each physics scheme. The latter approach leads to perturbations to the outputs from an earlier scheme acting as inputs to the subsequent schemes, resulting in a non-linear attribution of model uncertainty. It then becomes difficult to avoid some duplication in attributing errors across schemes, and introduces inherent correlations between the perturbations applied to one scheme's outputs and the outputs of a later scheme. Our proposed approach is to apply the stochastic perturbations linearly — perturbing individual physics tendencies *after* the full sequence of physics schemes has been completed. In so doing, iSPPT[111111] remains equivalent to standard SPPT, which would not be the case with sequential perturbations from a single random pattern. Under this approach, it is possible to control the correlation between perturbations to different physics schemes explicitly by using the same random pattern for different processes. It also enables iSPPT to be used as a research tool, to understand the impact of perturbations applied to an individual scheme, while all other schemes are unperturbed.

By perturbing tendencies from each physics scheme independently, iSPPT [123456] samples a higher-dimensional phase-space than standard SPPT, which generates the greater variance expressed in (4). A concern with the iSPPT approach, however, is that the balances that result, by design, from the sequential calls to physics routines in the unperturbed model are no longer respected by the perturbed ensemble members. Inconsistencies may arise within the grid column by independently perturbing the tendencies from each physical parametrisation. For example, the convective tendency may be scaled down, reducing the vertical transport of water vapour, while the large scale water processes scheme may be amplified, indicating an increase in high-level cloud. In this study, one attempt to limit this problem has been explored, whereby independent perturbations are applied to the net tendencies from *groups* of physics processes: specifically, using only two independent patterns to perturb nominally dry (RDTN, TGWD, NOGW) and moist* (CONV, LSWP, MOXI) processes independently: "iSPPT [112212]". Under this configuration of iSPPT, it is recognised that the errors from certain processes (dry versus moist) are not perfectly correlated, while preserving some of the balances between the parametrisation tendencies.

Compared to standard SPPT, iSPPT introduces multiple additional parameters: each random pattern requires the specification of space and time correlation scales and a standard deviation (as summarised in Table 1 for standard SPPT). Ideally, the parameter values associated with the random pattern for each (set of) physics parametrisation(s) would be guided by our understanding of the uncertainties associated with the representation of that particular process. In the absence of such knowledge and for simplicity in this study, we apply the parameters of the 3-scale pattern from standard SPPT (Table 1) to each of the independent random patterns in the iSPPT experiments. The need to specify the standard deviation, spatial and temporal correlations for each random pattern separately could be considered a limitation of the iSPPT approach: testing iSPPT using the same pattern characteristics for each physical process

*i.e. those processes which produce a moisture tendency will indicate the impact of the independent seeds alone, without the complication of also specifying separate noise parameters.

The increased variance that arises due to iSPPT compared to standard SPPT will act to increase the ensemble spread (standard deviation) in the forecast. The question that arises is whether that increase in spread yields a more reliable forecast.

3. Impact on Medium-range Forecasts

A series of medium-range ensemble prediction experiments has been conducted with the IFS ensemble system (ENS) to consider the impact of the iSPPT scheme. The experiments are performed with model version CY43R1, using the cubic octahedral grid with truncation at wavenumber 399 (so-called, $T_{CO}399$) with 91 vertical levels, for a forecast range of 15 days. The atmospheric model is coupled to the dynamical ocean model (NEMO). Each ensemble forecast consists of one control and 20 perturbed members. The perturbed members arise due to the representation of initial state uncertainties via a combination of the ensemble of data assimilations (EDA) (Isaksen et al. 2010) and singular vectors (see e.g. Leutbecher and Lang 2014). Operationally, forecast uncertainties due to the model integrations are represented by the (standard) SPPT scheme and the Stochastic Kinetic Energy Backscatter (SKEB) scheme (Berner et al. 2009). SKEB has been shown to have only a small impact on ensemble forecasts in the IFS (Leutbecher et al. 2017), and so has been deactivated for these experiments; comparisons are made between the impact of using standard SPPT and iSPPT.

The SPPT approach is known to result in a systematic trend in globally averaged moisture in the model: SPPT leads to a systematic drying of the atmosphere. The simple moisture conservation fix described in Leutbecher et al. (2017) is used in all SPPT and iSPPT experiments discussed here to ensure moisture is globally conserved.

Results are derived from forecasts with 46 start dates every 8 days spanning January to December 2015. The impact of iSPPT is demonstrated with respect to an experiment only including perturbations to the initial state ("ICP only"), and with reference to "standard SPPT". We demonstrate two versions of the iSPPT scheme: iSPPT[123456], in which each parametrisation scheme sees an independent stochastic perturbation; and iSPPT[112212], where the net tendencies due to the moist processes are perturbed with one pattern, and the net tendencies due to the dry processes are perturbed with a different pattern. The experiments are summarised in Table 2.

3.1. Forecast reliability and skill

The reliability of ensemble forecasts can be evaluated by comparing the spread (standard deviation) of the ensemble to the root mean square error (RMSE) between the ensemble mean and the verification as a function of time. The operational ECMWF analysis is taken as the verification. For a reliable forecast, i.e. a probabilistic forecast that is well calibrated and that correctly represents all sources of uncertainty in the forecast, the average ensemble spread should be a good predictor of the RMSE (Leutbecher 2010). We start by evaluating the RMSE and average ensemble spread as a function of time for each forecast experiment. Figure 1 shows the results for forecasts of temperature at 850 hPa (T850) averaged over each standard ECMWF region: the northern extra-tropics (NET) is defined as north of $20°N$, the southern extra-tropics (SET) is defined as south of $20°S$, and the tropics is defined as $20°S - 20°N$.

From the first row of Figure 1, it is apparent that the iSPPT approach has only a small effect on the RMSE, and a small effect on the ensemble spread in the extra tropics where the operational SPPT ensembles are well calibrated (the ensemble spread and RMSE are similar as a function of time). Both iSPPT schemes have a large positive impact on the ensemble spread in the tropics, where the operational SPPT ensemble forecasts are under-dispersive.

Experiment name	RDTN	TGWD	CONV	LSWP	NOGW	MOXI
ICP only	0	0	0	0	0	0
Standard SPPT	1	1	1	1	1	1
iSPPT [123456]	1	2	3	4	5	6
iSPPT [112212]	1	1	2	2	1	2

Table 2. Details of the experiments used to investigate the impact of independent SPPT. The numbers indicate the random pattern used to stochastically perturb each parametrisation. For the experiment with initial condition perturbations only (ICP only), the zeroes indicate an absence of stochastic perturbations to the physics schemes.

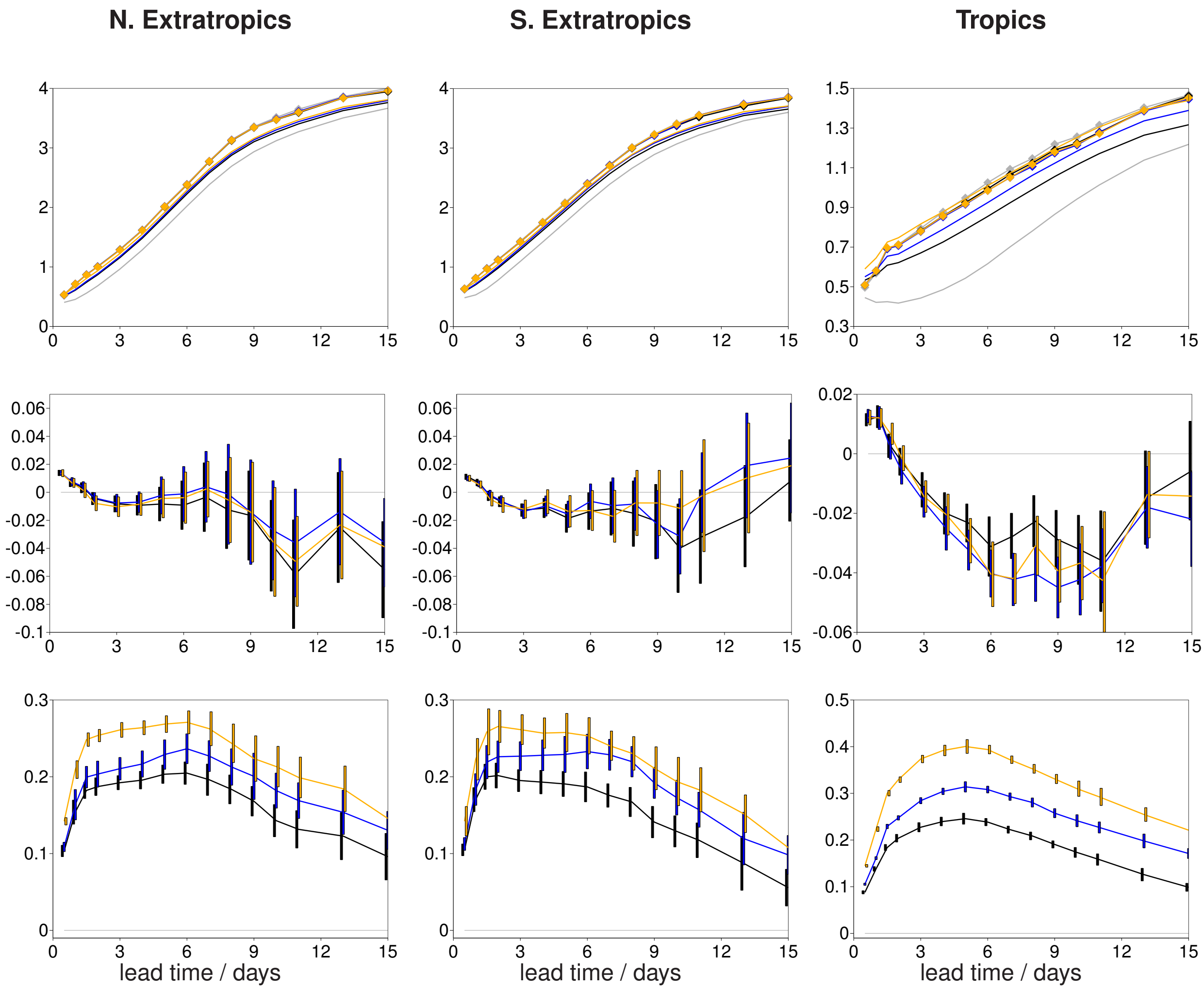

Figure 1. RMS ensemble spread and error as a function of lead time for forecasts of temperature at 850 hPa (K) for the experiments in Table 2. First row: RMS error (lines with crosses) and spread (lines with no symbols). Second row: Ensemble error relative to an experiment with initial condition perturbations only (ICP only). Third row: ensemble spread relative to ICP only. Bars indicate the significance of the differences at the 95% confidence interval. First column: northern extra-tropics. Second column: southern extra-tropics. Third column: tropics. ICP only — grey; standard SPPT — black; iSPPT [112212] — blue; iSPPT [123456] — yellow. Note that in the second and third rows, the grey ICP only line corresponds to the zero line.

The second row shows the difference between the RMSE for each experiment and the RMSE for an experiment with initial condition perturbations (ICP) only. A difference of less than zero indicates a reduced (i.e. improved) RMSE in the SPPT or iSPPT experiment. The error bars show the significance at the 95% confidence interval in each diagnostic calculated across different start dates, and so indicate the uncertainty in the diagnostic due to the sample size. Note that the significance of the difference between SPPT and each iSPPT scheme, as stated in the following discussion, cannot be directly inferred from Figures 1 and 2 due to the dependency of the forecast skill on the forecast start date. It was therefore calculated separately by considering how the difference of spread, error or skill score between the schemes varies across all start dates. There is no significant difference in RMSE in the NET, while in the

SET there is a small but significant increase in RMSE for iSPPT [112212] compared to SPPT. In the tropics there is a small but significant improvement in ensemble mean error for iSPPT [112212] from a lead time of two to ten days, and for iSPPT [123456] from a lead time of five to seven days.

The third row indicates the difference between ensemble spread for each experiment and ensemble spread for the experiment with ICP only. Both SPPT and the iSPPT schemes result in a significant increase of spread globally and at all lead times. The iSPPT [112212] scheme increases spread compared to standard SPPT, with the iSPPT [123456] scheme leading to further increases in spread compared to iSPPT [112212]. The relative impact of the SPPT schemes is largest in the tropics, where the ensemble spread is low for the ICP only experiment (note the different vertical axis). The largest impact is observed in the tropics for the iSPPT [123456] scheme, though both iSPPT schemes tested led to a large, significant increase in ensemble spread over standard SPPT at all lead times in the tropics.

The Continuous Ranked Probability Score (CRPS) can be used to evaluate the skill of the forecasts. It is a proper score, and is therefore sensitive to both reliability and resolution (Bröcker 2009).

$$\mathrm{CRPS} = \frac{1}{N}\sum_{n=1}^{N}\int_{x=-\infty}^{x=\infty}\left(F_n^f(x) - F_n^o(x)\right)^2, \quad (5)$$

where $F_n^f(x)$ is the forecast probability cumulative distribution function (cdf) for the nth forecast-verification case, and $F_n^o(x)$ is the observed probability cdf for the nth forecast-verification case. The CRPS is averaged over N forecast-verification pairs. The standard CRPS is adjusted, following Ferro (2014), to compensate for the finite ensemble size to ensure that the score is 'fair'. The smaller the CRPS, the more skilful the forecast.

Figure 2 shows the CRPS evaluated for each experiment in turn, minus the CRPS evaluated for the experiment with ICP only. Negative values indicate the CRPS has improved compared to that from the experiment with no stochastic physics. The first row shows the CRPS for forecasts of T850. For the NET and SET, the schemes have a largely neutral impact compared to standard SPPT. However in the tropics both iSPPT schemes lead to a significant improvement in CRPS at all lead times greater than one day compared to SPPT, due to both the increase in ensemble spread and the reduction in ensemble mean error.

Figure 2 also shows the CRPS evaluated for forecasts of (second row) the temperature at 200 hPa (T200), and (third row) the zonal wind at 200 hPa (U200). These variables were chosen as representative of all variables examined, and for brevity the ensemble mean error and spread are not shown for these variables.

In the NET and SET, the iSPPT schemes lead to a significant improvement in CRPS for T200 compared to standard SPPT due to significantly increased ensemble spread and, in the case of iSPPT [123456] forecasts, significantly reduced ensemble mean error. The CRPS for forecasts of U200 shows no significant change in these regions.

The third column of Figure 2 shows the CRPS evaluated in the tropics. Both iSPPT schemes lead to a significant improvement in CRPS compared to standard SPPT. The iSPPT [112212] scheme significantly improves T200 forecasts at lead times greater than two days, while the iSPPT [123456] scheme improves forecasts between three and ten days, though short range forecasts out to two days are degraded due to increased ensemble mean error. Both iSPPT schemes significantly reduce the CRPS for U200 at all lead times due to increased ensemble spread.

3.2. *Source of improved tropical reliability*

The two iSPPT schemes tested lead to a significant improvement in forecast reliability in the tropics, while having a broadly neutral impact in the extra-tropics where forecasts were already well calibrated. What is the cause of the improved spread in the tropics when the iSPPT scheme is activated?

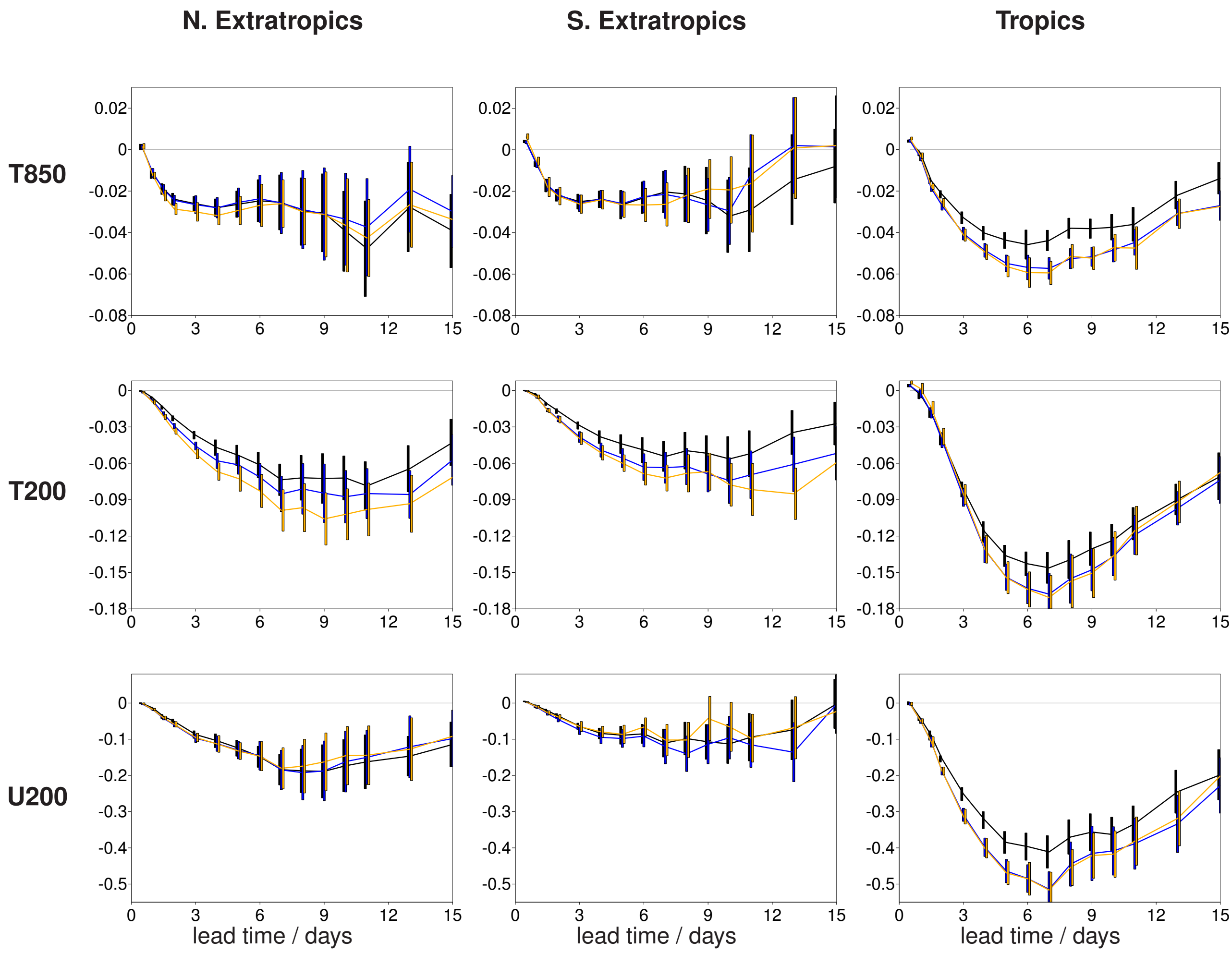

Figure 2. Continuous Ranked Probability Score (CRPS) as a function of lead time for the experiments in Table 2. First column: northern extra tropics; Second column: southern extra-tropics; Third column: tropics. First row: T850 (K); Second row: T200 (K); Third row: U200 (m/s). All panels show the CRPS relative to an experiment with initial condition perturbations only (ICP only). As CRPS is negatively orientated, differences less than zero indicate an improvement. ICP only — grey; standard SPPT — black; iSPPT [112212] — blue; iSPPT [123456] — yellow. Note that the grey ICP only line corresponds to the zero line.

From our knowledge of the IFS physics parametrisation schemes, we hypothesise that it could be decoupling the convection scheme from other parametrisation schemes that would lead to the observed large increase in spread. It is known that a large positive tendency from the convection scheme is often accompanied by a cooling tendency from both large scale water processes and radiation, which gives a small net tendency even if the individual tendencies are large. In addition, in Christensen et al. (2015), a version of SPPT was tested in which the perturbation to the convection scheme was switched off, while the other schemes were perturbed using a single pattern as in SPPT: a large increase in spread in the tropics was observed in forecasts using this approach. If decoupling convection from the other physics schemes is indeed the main source of the improvement, the large impact in the tropics and small impact in extra-tropics would be expected, as convection tendencies are larger in the tropics.

To investigate this hypothesis, we consider areas in the tropics where there is little convection, and compare these to areas where convection is the dominant process. The regions of interest are defined following Christensen et al. (2015). IFS model data from the Year of Tropical Convection (YOTC) project is used, which archived high resolution forecast and analysis data for the dates May 2008 — April 2010. The ratio between the magnitude of the 24 hour convection temperature tendency at 850 hPa and the sum of the magnitudes of all parametrised tendencies is calculated, for forecasts initialised every five days. If this ratio is large, it indicates convection is the dominant process. Since the

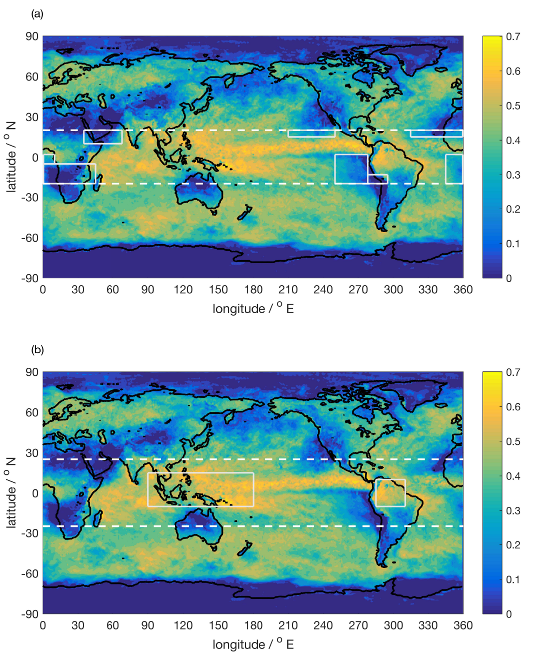

Figure 3. Convection diagnostic (colour) derived from the IFS tendencies calculated as part of the YOTC project (see text for details). (a) Regions where the diagnostic is close to zero (bounded by grey boxes), indicating there is little convection. (b) Regions where the diagnostic is large (bounded by grey box), indicating convection is the dominant process. The dashed lines indicate the tropics.

forecast skill of the IFS is dependent on latitude, both the regions with significant and little convection are defined in the tropics. The regions are approximately the same size, and cover similarly sized areas of both land and ocean. The regions are indicated by the grey boxes in figure 3.

Figure 4 shows the RMSE and spread as a function of time for (a–c) regions with significant convection and (d–f) regions with little convection. The operational SPPT ensembles (grey lines) are under-dispersive at all times, for all variables considered, in both regions. The under-dispersion is greater in regions with significant convection, while regions with little convection are better calibrated. The iSPPT approach increases the ensemble spread in both regions, however, the increase in spread is considerably greater in regions with significant convection. The spread of the ensembles closely matches the RMSE, though is over-dispersive for T850 for the iSPPT [123456] scheme in these regions. For other variables, the impact of the two iSPPT approaches is similar.

To probe further into the mechanism by which iSPPT increases the ensemble spread, a series of six experiments was carried out. In each experiment, just one of the physics schemes was perturbed with an independent random number field to the other schemes (i.e., [122222], [121111], ...). These experiments comprise forecasts from 23 start dates every 16 days spanning the same period. These 'individually independent' experiments will demonstrate further whether it is decoupling one particular scheme from the others that results in the large increase in spread, or if it is important that all schemes are treated independently.

Figure 5 shows the ensemble spread as a function of time for a subset of the 'individually independent' experiments. Perturbing TGWD, NOGW or MOXI independently from the other schemes leads to very little impact on ensemble spread, so they are not shown for clarity. The small impact is likely because either the schemes have very small tendencies (MOXI), or because the schemes act mainly in the boundary layer (TGWD) or stratosphere (NOGW), away from the verified variables and where the stochastic perturbations are tapered. Independently perturbing the convection tendency, iCONV [112111], has the largest impact on ensemble spread. The increase in spread is equal to or greater than independently perturbing all physics schemes. The next most influential scheme is radiation, though the iRDTN[122222] experiment shows smaller ensemble spread than iSPPT [123456]. For variables at 850 hPa, iLSWP has a large impact, though the impact is smaller at higher altitudes. For all schemes, the impact is larger in regions with significant convection than regions with little convection.

4. Case study: forecast busts over Europe

In Section 3, the iSPPT schemes were shown to significantly increase the spread of ensemble forecasts in the tropics, with the largest impact in regions with significant convection. The impact of the schemes in the extra tropics was shown to be small. Here, the forecast spread is already well calibrated, and forecasts are generally skilful. However,

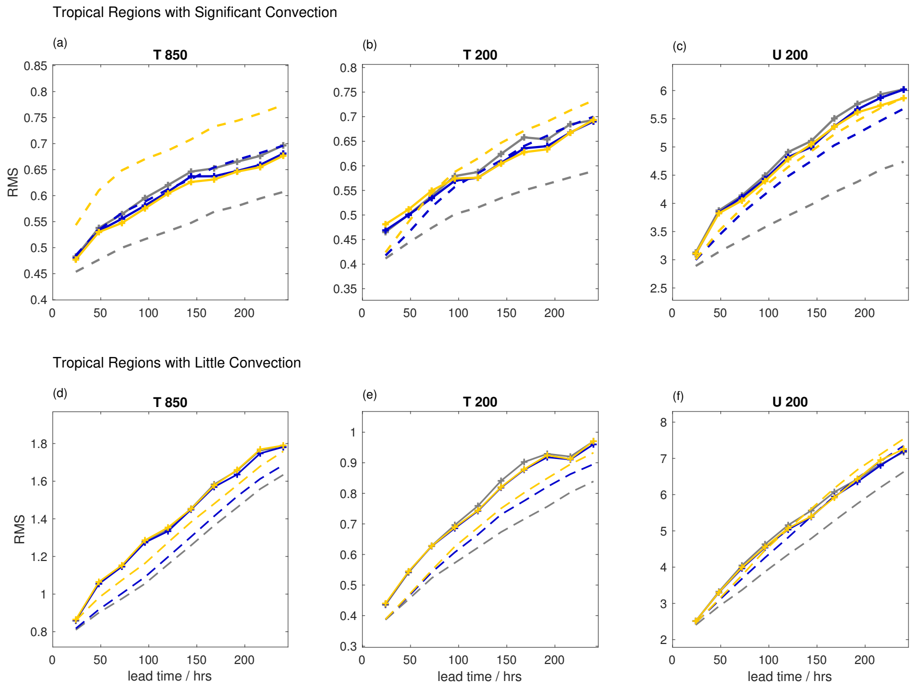

Figure 4. RMS spread (dashed line) and error (solid line with crosses) as a function of time for the experiments in Table 2. The three columns show results for T850, T200 and U200 respectively. The top row shows results for *tropical regions with significant convective activity*, the bottom row shows results for *tropical regions with little convective activity*. standard SPPT — grey; iSPPT [112212] — blue; iSPPT [123456] — yellow.

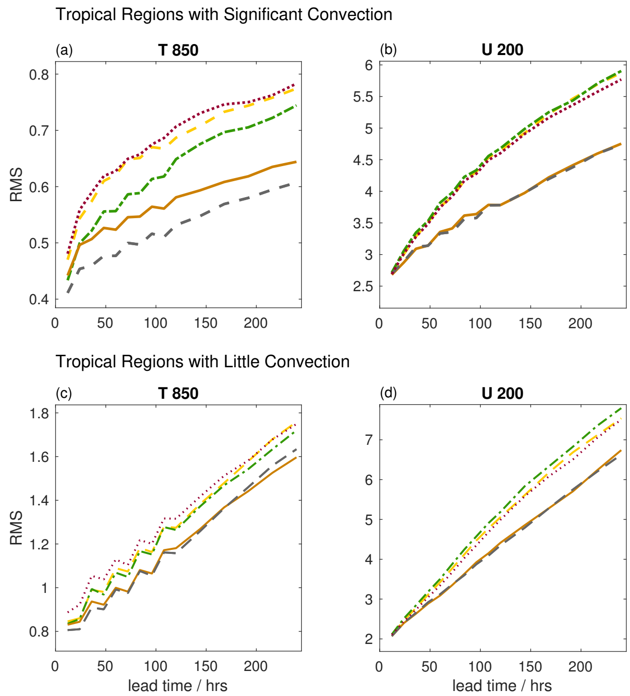

Figure 5. RMS spread as a function of time for the 'individually independent' experiments. The two columns show results for T850 and U200 respectively. The top row shows results for *tropical regions with significant convective activity*, the bottom row shows results for *tropical regions with little convective activity*. standard SPPT — grey dashed; iRDTN[122222] — green dash-dot; iCONV[112111] — magenta dotted; iLSWP[111211] — orange solid; iSPPT [123456] — yellow dashed.

there are occasions when the skill of forecasts in the extra tropics is considerably poorer than usual. For example, on average there are between five and ten European 'forecast busts' a year in the operational ECMWF forecasts, when the day-six error in forecasts of the geopotential height at 500 hPa (Z500) over Europe exceeds 60m, and the anomaly correlation coefficient drops below 40% (Rodwell et al. 2013). Ensemble forecasts of these events are systematically under-dispersive (Rodwell et al. 2013, 2015).

In Rodwell et al. (2013), the authors identify that European forecast busts are more likely to occur in the presence of a specific anomaly in the initial conditions over North America, namely a trough over the Rocky mountains combined with a region of high Convectively Available Potential Energy (CAPE) over Eastern North America. By considering this initial condition anomaly, the large forecast error over Europe can be largely attributed to errors in the representation of mesoscale convective systems (MCSs) that form in regions of high CAPE over Eastern North America (Grazzini and Isaksen 2002; Rodwell et al. 2013). Since iSPPT increases the spread of ensemble forecasts in regions with significant convection, it is of interest to investigate whether including iSPPT can reduce the likelihood or severity of extra-tropical forecast busts over

Europe through better representation of uncertainty in the initiation and development of the associated MCS.

We test this hypothesis by producing six-day forecasts from ten start dates which fulfil the 'trough/CAPE composite' criterion described in Rodwell et al. (2013), as well as from a further ten start dates which do not. The ensemble forecast setup is the same as in Section 3: a 21-member ensemble forecast at $T_{CO}399$ is performed for each start date for both the standard SPPT and iSPPT [123456] forecast models. However, unlike the experiments in Section 3, no initial condition perturbations are applied to the ensemble members, to enable a clearer diagnosis of the impact of iSPPT versus SPPT. The start dates are selected from across all seasons as in Section 3, but here we select from among the dates used in Rodwell et al. (2015), between May 2014 and April 2015. We require all 20 start dates to be separated by more than six days, such that the forecast periods do not overlap[†]. Hereafter, we refer to these sets of dates as 'bust' and 'non-bust' dates. However, note that these dates are selected based on the presence of an initial condition anomaly that is more (or less) likely to produce a forecast bust, and not on the presence of a bust in the operational forecast at day-six over Europe. We evaluate the forecasts over a North Atlantic/European region defined as $30-70°\text{N}, 30°\text{W}-30°\text{E}$.

Figure 6 shows the CRPS for forecasts of 500 hPa geopotential height (Z500) as a function of lead time for bust and non-bust dates. The CRPS is averaged over the North Atlantic/European region defined above. Each panel shows the CRPS for iSPPT [123456] with respect to standard SPPT. While there is no significant difference in CRPS for the non-bust dates, there is a small but significant reduction (improvement) in CRPS for the bust dates for lead times of five days or more.

[†]The dates selected are as follows, dd.mm.yyyy, all initialised at 00UTC. 'Bust' dates which fulfil the trough/CAPE criterion: 10.05.2014, 29.06.2014, 22.07.2014, 21.08.2014, 03.09.2014, 19.09.2014, 02.10.2014, 23.11.2014, 08.04.2015, 25.04.2015. 'Non-bust' dates which do not fulfil the trough/CAPE criterion: 01.05.2014, 04.06.2014, 10.07.2014, 06.08.2014, 27.08.2014, 11.09.2014, 26.09.2014, 28.10.2014, 28.02.2015, 17.04.2015.

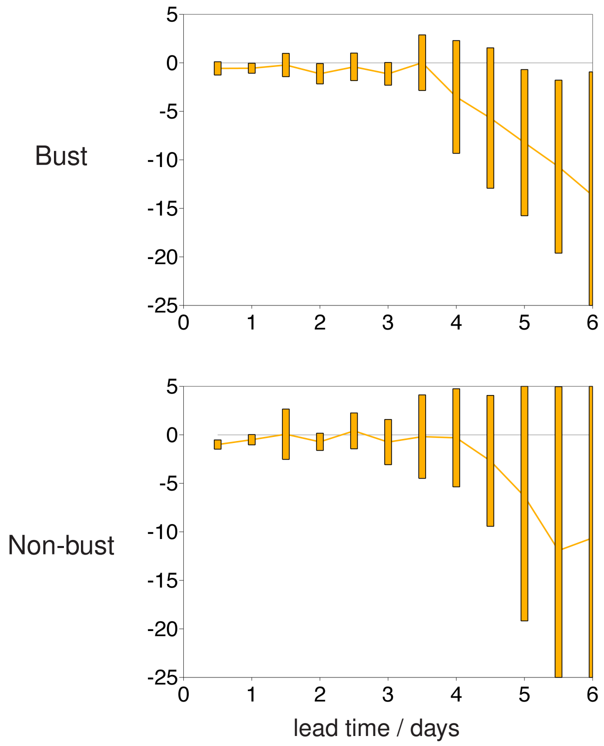

Figure 6. Continuous Ranked Probability Score (CRPS) for Z500 (dm) as a function of lead time for the forecast bust experiments. Top: CRPS for initial conditions which fulfilled the 'trough/CAPE' bust criterion. Bottom: CRPS for initial conditions which did not fulfil the 'trough/CAPE' bust criterion. The CRPS is shown for the iSPPT [123456] experiment relative to the experiment with standard SPPT, and is averaged over the North Atlantic/European region $30-70°\text{N}, 30°\text{W}-30°\text{E}$. Bars indicate the significance of the differences at the 95% confidence interval. As CRPS is negatively orientated, differences less than zero indicate an improvement.

To understand the improvement in CRPS observed for the bust dates at longer lead times, Figure 7 shows the average Z500 ensemble spread and RMSE at day six for the bust and non-bust dates. Forecasts using standard SPPT and iSPPT [123456] are compared. For both SPPT and iSPPT, the bust dates show an enhanced RMSE over the North Atlantic compared to the non-bust dates: this region of high RMSE explains the presence of a European 'forecast bust' for some such cases. In contrast, the ensemble spread is similar between the bust and non-bust dates. Figure 7 shows that the primary impact of iSPPT [123456] is to reduce the RMSE for the bust dates in the region of the largest error, while the RMSE shows little change for the non-bust dates. Consistent with our earlier analyses in the extra-tropics, there is a small increase in spread from iSPPT[123456] compared to standard SPPT, which applies to both sets of dates.

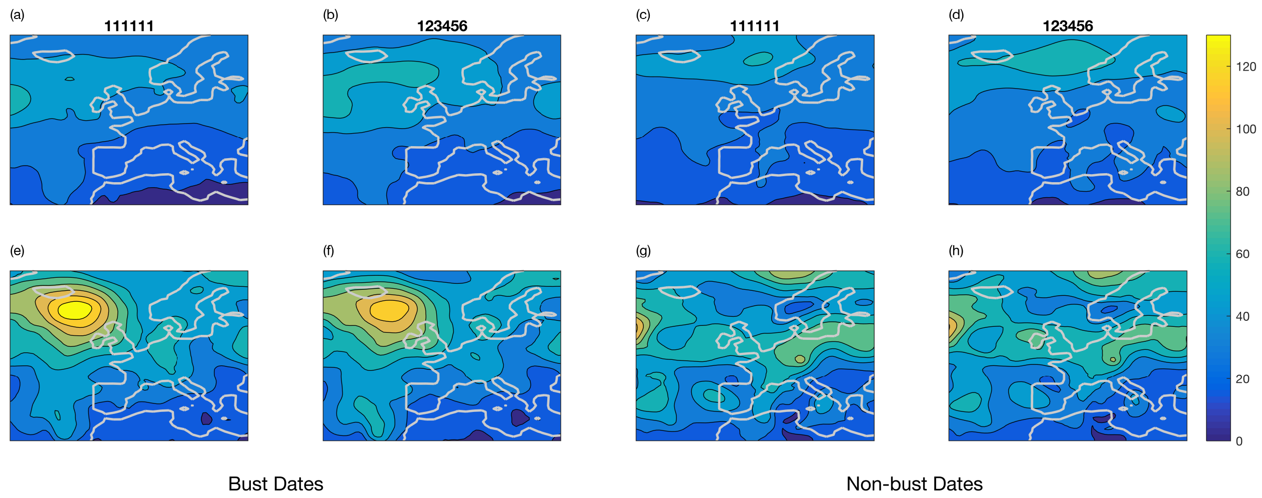

Figure 7. Mean ensemble spread (top row) and root mean square error (bottom row) for 144 hr forecasts of geopotential height at 500 hPa (Z500; m) for dates selected to target sensitivity to forecast busts. (a, b, e, f) Initial conditions were selected which fulfilled the "Trough/CAPE" criterion indicating a forecast bust is likely. (c, d, g, h) Initial conditions were selected which did not fulfil the "Trough/CAPE" criterion, indicating a bust is unlikely. Results are shown for standard SPPT (a, c, e, g) and iSPPT [123456] (b, d, f, h). The colour bar corresponds to all panels. The results are shown over the North Atlantic/European region, with the grey lines indicating the coastline.

These results indicate that the improved skill (CRPS) in Z500 forecasts for this region beyond day-five can be directly traced to the reduction in the RMSE of the ensemble mean over the region. While we note that the sample of ten start dates used for each experiment is small, and that a larger study would be required to ensure a robust result, the fact that Figure 7 shows that the largest RMSE reduction is where the errors are largest indicates that iSPPT is beneficial in these bust conditions.

5. Impact on Model Climate

Finally, we consider the impact of the iSPPT scheme on the model climate. A set of 13-month model integrations have been performed, consistent with the earlier medium-range experiments. The long integrations are performed with the atmosphere-only model on the linear reduced Gaussian grid with truncation at wavenumber 255 (T_L255) and 137 vertical levels. Sea surface temperatures (SST) are prescribed from observations. Each experiment comprises integrations from four different start dates (00 UTC, 1st August 2000; 06 UTC, 1st August 2001; 12 UTC, 1st August 2002; 18 UTC, 1st August 2003) to sample initial state uncertainty ("ICP only"). In addition, experiments with stochastic perturbations activated from standard SPPT and iSPPT are considered. Annual means are constructed from the final 12 months of each integration (i.e. the first month's data are discarded). Comparisons between experiments and with respect to observations have been considered.

Figure 8 shows the distribution of mean precipitation for the ICP only, standard SPPT, and iSPPT [123456] experiments. The iSPPT [112212] experiment gave results between standard SPPT and iSPPT [123456], so is not shown for brevity. SPPT has a small impact on the average precipitation, slightly increasing the bias with respect to GPCP when compared to the control integration. The impact of iSPPT is to exacerbate the impact of SPPT on model climate. While the pattern of bias is very similar, the magnitude is approximately 10% higher. The model precipitation field can be decomposed into convective and large scale precipitation. The iSPPT [123456] integration has less convective precipitation compared to standard SPPT, but more large-scale precipitation (not shown). The iSPPT run led to an increase in mean surface latent heat flux into the atmosphere in the Indian Ocean and West Pacific (not shown), which could result in the observed increase in large scale precipitation through enhanced moisture flux. Interestingly, standard SPPT also results in a systematic reduction in convectively available potential energy (CAPE) across the tropics, reducing the globally averaged CAPE

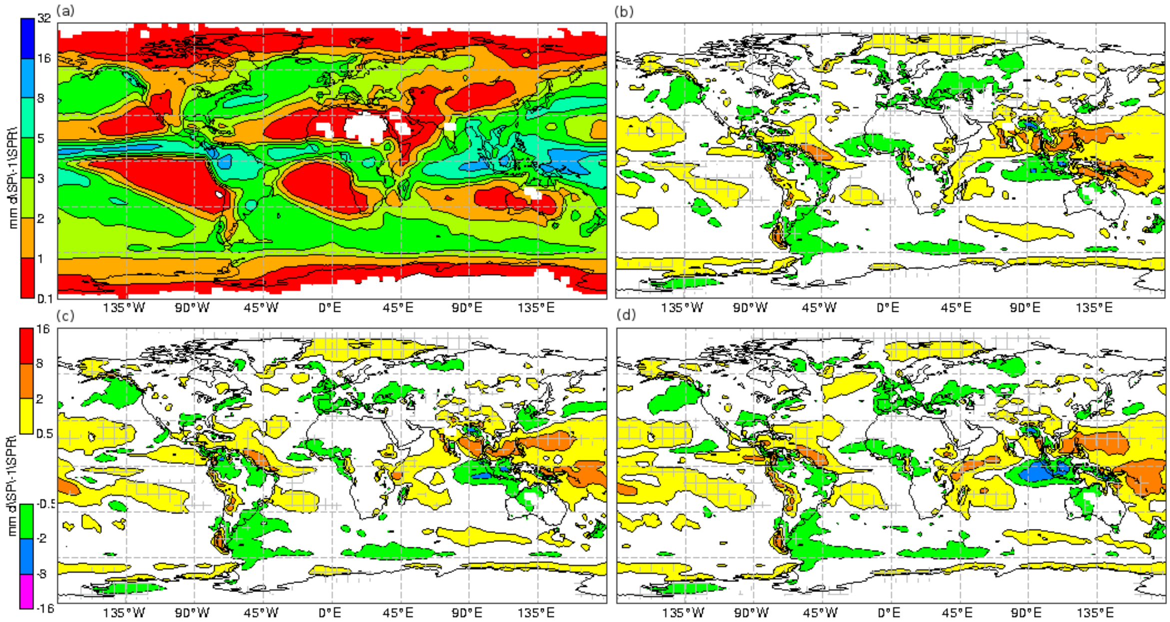

Figure 8. Climatological distribution of yearly—averaged precipitation from 13-month climate integrations with prescribed SST. (a) GPCP2.2 observational dataset. (b) Bias in ICP only model integration: [000000].(c) Bias in model integration with standard SPPT. (d) Bias in model integration with iSPPT: [123456]. The legend in (c) also corresponds to (b) and (d).

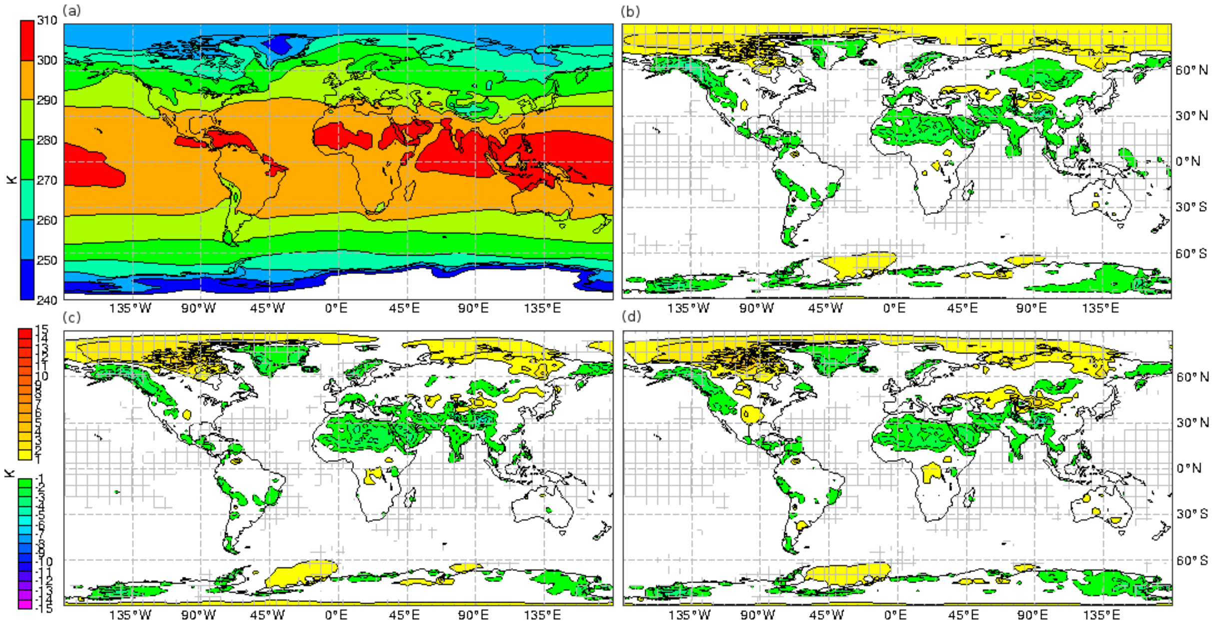

Figure 9. Climatological distribution of yearly—averaged temperature at 2m (T2m) from 13-month climate integrations with prescribed SST. (a) ERA-Interim reanalysis dataset. (b) Bias in ICP only model integration: [000000]. (c) Bias in model integration with standard SPPT. (d) Bias in model integration with iSPPT: [123456]. The legend in (c) also corresponds to (b) and (d).

from 200 J/kg to 176 J/kg, while the iSPPT [123456] integration shows a further reduction to 163 J/kg (not shown). This reduction in CAPE is not understood, though it could explain the reduction in convective precipitation.

Figure 9 shows the distribution of mean two-metre temperature (T2m) for the ICP only, standard SPPT and [123456] experiments. As for precipitation, the distribution of errors in the mean T2m are similar for the control and stochastic experiments. Compared to ICP only, the standard SPPT integration has slightly reduced the warm bias over the Arctic, though the cold bias over northern Africa and parts of Eurasia has not improved. The [123456] integration has similarly improved the warm Arctic bias compared to ICP only, and has slightly reduced the cold bias over parts of Eurasia. This has led to a reduction in the global mean bias compared to standard SPPT and ICP only.

Table 3 summarises the global RMS of the spatial distribution of annual mean biases for a number of climate variables. The results are mixed, with no one model performing clearly better than the others. Only for precipitation (PPT) and top of atmosphere short-wave radiation (TSW) does the ICP only integration show the lowest RMSE, for all other variables one of the stochastic models performs the best. The two-pattern [112212] integration usually shows RMSE between the standard SPPT and the fully independent [123456] integration. Overall, the impact on modelled mean climate is modest, with only small changes observed for the different iSPPT schemes tested, though no adverse impact of the schemes is observed.

6. Discussion and Conclusions

6.1. Medium range weather forecasts

The new iSPPT schemes result in a significant increase of spread for all variables at all lead times. In the extra-tropics, the ensemble forecasts are well calibrated. While there is a statistically significant increase in ensemble spread in these regions, it tends to be small, and does not result in a statistically significant change in CRPS (Figure 2). The only variable for which this is not the case is T200, where the iSPPT [112212] and iSPPT [123456] schemes result in a significantly improved CRPS for a range of lead times in the northern and southern extra tropics.

In the tropics, forecasts made with standard SPPT are significantly under-dispersive. Both iSPPT [112212] and iSPPT [123456] generate significantly greater spread than when SPPT is used, resulting in a statistically significant improvement to the CRPS for all variables tested at a range of lead times. The iSPPT schemes produce a flow-dependent estimate of forecast uncertainty, with a larger impact on SPPT forecasts that were more under-dispersive (in the tropics) than those that were well calibrated (in the extra tropics).

The iSPPT [112212] and iSPPT [123456] schemes have a comparatively small impact on the forecast RMSE, showing no significant change in any region compared to the standard SPPT experiments at most lead times for most variables. Furthermore, no cases of numerical instability were encountered in any of the experiments performed.

The impact of iSPPT in tropical regions with significant convection is considerably greater than in tropical regions with little convection (Figure 4). This is beneficial for forecast skill, as it is regions where there is significant convection that are under-dispersive when using the standard SPPT scheme. For example, Rodwell et al. (2016) show that ensemble forecasts made within the data assimilation cycle are systematically under-dispersive in convective regions over South America: the short lead time of these forecasts helps us attribute this lack of spread to deficiencies in representing uncertainty due to convection. Both the iSPPT [112212] and [123456] schemes improve the calibration of forecasts in convecting regions, while maintaining the good calibration of forecasts in regions with little convection.

The large impact of iSPPT in convecting regions indicates that convection, together with its interactions with other physics schemes, is a key process by which iSPPT impacts the ensemble, and it is this aspect of the scheme in particular which likely produces flow-dependent estimates of uncertainty (Christensen et al. 2015). Equations (2) and (4) highlight that the forecast uncertainty represented by iSPPT will be greater than SPPT for regions where the model tendencies act in opposite directions, i.e., where the individual tendencies are large but the net tendency is small. In tropical regions with significant convection, this is indeed the case for the IFS. The convection scheme parametrises the effect of convective latent heating on the atmosphere. This scheme interacts directly with the large scale water processes (clouds) scheme: water detrained from the convective plume acts as a source of water for clouds in the large scale water processes scheme, which then calculates the effect of evaporative cooling

Variable	ICP only [000000]	standard SPPT [111111]	iSPPT [112212]	iSPPT [123456]
10W (m/s)	1.31	1.28	**1.27**	**1.27**
2D (K)	1.11	1.05	1.07	**1.04**
2T (K)	0.842	0.836	**0.825**	0.830
PPT (mm/d)	**0.881**	0.909	0.937	0.982
TCC (%)	11.8	11.5	**11.1**	11.6
TCWV (kg/m^2)	2.74	**2.68**	2.70	2.73
TCLW (g/m^2)	35.9	34.6	33.6	**33.1**
TLW (W/m^2)	5.05	**4.65**	4.78	5.63
TSW (W/m^2)	**8.74**	8.93	9.18	9.62

Table 3. Global root mean square of the annual mean biases from 13-month climate integrations with prescribed SST. 10W: 10m wind, compared to SSMI; 2D: 2m dewpoint, compared to ERA Interim; 2T: 2m temperature, compared to ERA Interim; PPT: total precipitation, compared to GPCP2.2; TCC: total cloud cover, compared to MODIS; TCWV: Total Column Water Vapour, compared to SSMI; TCLW: Total Column Liquid Water, compared to SSMI; TLW: Top of atmosphere long wave radiation, compared to CERES-EBAF; TSW: Top of atmosphere short wave radiation, compared to CERES-EBAF.

on the atmosphere (ECMWF 2015). This interaction means that a warming due to convection tends to be associated with a cooling from the cloud scheme. The opposing nature of these tendencies results in the significant increase in ensemble spread associated with iSPPT in these regions. The individually independent SPPT experiments also suggest that decoupling clouds and convection from each other results in a large increase in spread for T850, as both the iCONV and iLSWP experiments showed increases in spread compared to standard SPPT (Figure 5).

The convection scheme also directly interacts with the radiation parametrisation scheme through its impact on cloud fraction: both short- and long-wave radiative transfer are sensitive to the cloud fraction predicted by the cloud scheme. In particular, low level cloud is often associated with cooling from the radiation scheme (Morcrette 2012), which opposes the warming from convection. Furthermore, it is known that in the tropics convective warming is statistically in balance with radiative cooling, with convective motions reacting to destabilisation of the atmosphere due to radiative cooling (Manabe and Wetherald 1967; Tompkins and Craig 1998a,b)[etc.]. The interaction between radiation and convection will contribute to the observed increase in spread for the iSPPT forecasts in regions with significant convection. The iRDTN experiment showed that decoupling radiation from the other parametrisation schemes also results in a large increase in spread (Figure 5), supporting this hypothesis. Results from the iSPPT [112212] experiment demonstrate that independently perturbing moist processes (dominated by convection) and dry processes (dominated by radiation) can give rise to much of the increased spread from iSPPT [123456] with respect to standard SPPT.

The observed improvement in ensemble forecast reliability and skill is a good indication that the proposed iSPPT approach improves the representation of model uncertainty in the IFS. In particular, the observed impact in regions with significant convective activity suggests iSPPT improves the representation of model uncertainty related to convective processes. It seems that requiring the maintenance of the balance between parametrisation schemes at each time step may be more than is necessary. Enabling independent perturbation of the component parametrised tendencies can generate additional spread and increase forecast reliability.

In particular, it seems that the error in the convective tendency is not perfectly correlated with the error in either the radiative or cloud tendencies, despite the physical relationship between the tendencies themselves. Decoupling convection from radiation, as in iSPPT [112212], or convection from both radiation and clouds, as in iSPPT [123456], and perturbing these tendencies separately, models this lack of correlation between the errors in the different schemes. These results motivate the development of stochastic parametrisation schemes that directly target uncertainty in convection, including how these uncertainties can influence the large-scale flow (e.g Shutts 2015).

6.2. *'Forecast bust' case study*

It has been documented that the IFS is systematically under-dispersive in regions with significant convection in the extra-tropics as well as in the tropics. For example, Rodwell et al. (2013) analysed cases where a 'forecast bust' was observed over Europe and demonstrated that the bust could be traced back to the occurrence of mesoscale convective systems (MCSs) over North America a few days previously. In particular, Rodwell et al. (2013) demonstrated that the IFS ensemble forecast showed a lack of spread after an MCS event, which propagated downstream and made these 'forecast busts' over the UK and Europe more likely.

We considered the impact of iSPPT on ensemble forecasts for a set of initial condition dates likely to develop a forecast bust over Europe, compared to a set of dates where a bust was unlikely. The standard SPPT scheme was compared to the iSPPT [123456] scheme. On average at day-six, the RMSE in Z500 was higher for the bust dates than for the non-bust dates over a North Atlantic/European region. The iSPPT scheme led to a small but significant reduction in CRPS in this region, while the CRPS in other regions (not shown) and for the non-bust dates, was unchanged. The improvement in CRPS due to iSPPT is associated with a reduction in the RMSE of the ensemble mean for the region for the bust dates, which was not observed for the non-bust dates. Maps of the region highlight a clear reduction of the largest RMSE, which would be associated with the 'forecast busts'.

Given the large impact of iSPPT in convecting regions, as discussed in Section 6.1, it seems possible that iSPPT has indeed improved the representation of uncertainty in the development of MCSs over North America. By sampling this uncertainty near the start of the forecast, the forecast error at day six is reduced. Given these results, it is possible that iSPPT would be beneficial in the ECMWF EDA system, by improving the representation of initial condition uncertainty in MCS regions. We emphasise that these results are a first indication of the potential impact of iSPPT in the extra-tropics. A larger sample of dates is required to make firm statements, together with a more complete analysis tracing the observed impacts of iSPPT back to their source. This further analysis is left for a future study.

6.3. *Model climate*

The impact on model climate was evaluated using four years of forecast data. Because of this limited dataset, only the impact on the mean climate could be evaluated. This could explain the small impact observed by the iSPPT schemes. Previous work indicates that SPPT is able to have a large beneficial improvement on the modelled climate, but the largest impact is on improving modes of climate variability such as North Atlantic weather regimes (Dawson and Palmer 2015), the El Nino-Southern Oscillation (Christensen et al. 2017), or the Indian monsoon (Strømmen et al. 2017). In a future study, we will evaluate the impact of the iSPPT scheme on longer coupled integrations in the climate model EC-Earth (which uses the IFS as its atmospheric component). This will allow us to consider the impact of iSPPT on climate variability as well as the mean state.

Nevertheless, these 13-month experiments are valuable as they demonstrate that the iSPPT approach does not adversely affect the mean climate of the IFS, and therefore that the average balance between the different parametrisation schemes, has not been adversely affected. The primary effect of the scheme is in improving the reliability of weather forecasts. The scheme could therefore also be useful in probabilistic seasonal forecasting, where it is imperative to produce probabilistic forecasts, but where longer-time scale climate biases can also affect the accuracy of the forecast.

6.4. Concluding comments

The proposed iSPPT approach for representing model uncertainty in the IFS leads to large and statistically significant improvements in forecast reliability in the tropics, while maintaining the calibration of forecasts in the extra topics. In particular, the scheme improves the representation of uncertainty associated with convective processes, having a large positive impact in regions where there is significant convection.

Operational ensemble forecasts in the extra-tropics are very well calibrated, e.g. for T850, while forecasts in the tropics remain somewhat under dispersive (Haiden et al. 2015). It is possible that the iSPPT [123456] scheme tested here at $T_{CO}255$ will result in over dispersive forecasts in the tropics at operational resolution, though the iSPPT [112212] scheme may improve the forecast reliability. However, we have not performed any tuning for the results presented here. When implemented at a higher resolution, the imposed iSPPT standard deviations could be reduced to improve the calibration of the forecasts. The demonstrated flow-dependent impact of the iSPPT scheme leads us to believe that the scheme could improve the reliability of tropical forecasts at operational resolution, while maintaining the calibration of forecasts in the extra-tropics.

It is likely that the 'true' errors in the parametrisation schemes are neither perfectly correlated as in SPPT, nor perfectly uncorrelated as in iSPPT [123456]. A further interesting line of enquiry would be to introduce correlations between the noise patterns used for different parameters. Instead of using two independent patterns in iSPPT [112212], perturbation patterns for the wet processes could be partially correlated with each other, while perturbations for the dry processes could also be partially correlated. This potentially more realistic representation of the errors in the forecast model could further improve the forecast reliability.

A key benefit of the iSPPT approach is the ability to specify different noise characteristics for different parametrisation schemes, according to our understanding of the model uncertainty stemming from each parametrisation. For example, in standard SPPT, the perturbations are tapered in the boundary layer and the stratosphere following arguments specific to individual parametrisation schemes (numerical instabilities associated with the boundary layer scheme and well constrained radiative tendencies in the stratosphere). Using the iSPPT approach, tapering can be directly targeted at the relevant schemes. We have not yet explored the potential of this flexible approach, in part due to the large parameter space that would need to be explored to tune the scheme. However, a coarse-graining study is currently underway that will explicitly measure the characteristics of the optimum perturbation to each parametrisation scheme, which will guide future implementation of iSPPT (Christensen et al, in prep).

The iSPPT approach provides the possibility to explore the model response to perturbations to individual physics processes, which will help inform future developments of model uncertainty representation. This is beyond the scope of this manuscript but will form the basis of future efforts.

Extending SPPT to enable independent perturbation of tendencies from different parametrisation schemes implies additional computational cost compared to standard SPPT. The additional costs arise from multiple sources: the additional storage and computations required for evolving multiple random patterns; and the requirement to separately store tendencies from different (sets of) parametrisation schemes. A crude estimate of the additional cost has been made using results from the 13-month model climate integrations, which indicate only a small relative increase (3%) in runtime when comparing iSPPT[123456] to standard SPPT[‡]. This moderate increase in computational cost is smaller than, or comparable to, alternative stochastic approaches for representing model uncertainty (e.g. Ollinaho et al. 2016), and so is affordable

[‡]The relative increase in computational cost depends on many factors, including forecast resolution, forecast range, I/O demand; and will display considerable variability between forecast runs. The 3% estimate is presented as guidance.

for implementation in operational forecasting centres. Furthermore, Leutbecher et al. (2017) describe a number of options to reduce the computational cost of random pattern generation, which would be beneficial when using iSPPT.

Overall, given the comparable performance of iSPPT[112212] and iSPPT[123456] for both short range and climate forecasts at T399, together with the smaller 'gap' between ensemble mean error and spread at operational resolution in the tropics, and the reduced computational cost of iSPPT[112212], we feel that using independent patterns for groups of parametrisation schemes as in iSPPT[112212] may be sufficient to realise the benefits of iSPPT in an operational forecast setting.

Since iSPPT is a generalisation of SPPT, any planned combination of SPPT with other approaches, such as with the Stochastically Perturbed Parametrisations scheme described in Ollinaho et al. (2016) and in Leutbecher et al. (2017), would also be possible with iSPPT. The impact of the iSPPT schemes has been observed to be similar across different model resolutions (previous experiments, not shown here) and across several IFS model cycles. For example, very similar results were obtained when the iSPPT scheme was evaluated in the older IFS cycle CY37R3 (Arnold 2013), which differs from CY42R1 in a number of key ways. Most notably for our discussion, CY37R3 closes the deep convection scheme using a CAPE formulation (Bechtold et al. 2008), whereas more recent model cycles use the 'PCAPE' formulation outlined in Bechtold et al. (2014). The robustness of the results gives us some confidence that the results presented here will be generally applicable to other models which use the SPPT parametrisation scheme, and that the iSPPT approach could also be used in those systems to improve the reliability of ensemble forecasts.

Acknowledgements

The research of H.M.C. and T.N.P. was supported by European Research Council grant number 291406. The authors would like to express their particular thanks to Martin Leutbecher, Mark Rodwell and Peter Bechtold (ECMWF) for their extensive input into this work through discussions and access to diagnostic tools; and to Richard Forbes, Erland Källén and Nils Wedi (ECMWF) for their helpful advice and comments on this manuscript. Thanks also to Alfons Callado Pallarès (AEMet) for his early input into the iSPPT approach, and to three anonymous reviewers for their constructive comments and suggestions which have improved this manuscript.

References

J. L. Anderson. The impact of dynamical constraints on the selection of initial conditions for ensemble predictions: Low-order perfect model results. *Mon. Weather Rev.*, 125(11):2969–2983, 1997.

H. M. Arnold. *Stochastic Parametrisation and Model Uncertainty*. PhD thesis, University of Oxford, 2013.

P. Bechtold, M. Köhler, T. Jung, F. Doblas-Reyes, M. Leutbecher, M. J. Rodwell, F. Vitart, and G. Balsamo. Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales. *Q. J. Roy. Meteor. Soc.*, 134(634):1337–1351, 2008.

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.

A. C. M. Beljaars, P. Bechtold, M. Köhler, J.-J. Morcrette, A. M. Tompkins, P. Viterbo, and N. Wedi. The numerics of physical parameterization. In *Proceedings, Seminar on Recent Developments in Numerical Methods for Atmospheric and Ocean Modelling*, pages 113–134, Shinfield Park, Reading, 2004. ECMWF.

L. Bengtsson, M. Steinheimer, P. Bechtold, and J.-F. Geleyn. A stochastic parametrization for deep convection using cellular automata. *Q. J. Roy. Meteor. Soc.*, 139(675):1533–1543, 2013.

J. Berner, G. J. Shutts, M. Leutbecher, and T. N. Palmer. A spectral stochastic kinetic energy backscatter scheme and its impact on flow dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.*, 66(3):603–626, 2009.

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, H. M. Christensen, and D. Coleman. Stochastic parameterizations in the community atmosphere model. *J. Climate*, 2016. Submitted.

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.

J. Bröcker. Reliability, sufficiency, and the decomposition of proper scores. *Q. J. Roy. Meteor. Soc.*, 135(643):1512–1519, 2009.

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.

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, 2017.

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 climate simulations. *GMD*, 2016.

A. Dawson and T. N. Palmer. Simulating weather regimes: impact of model resolution and stochastic parametrisation. *Clim. Dynam.*, 44 (7):2177–2193, 2015.

ECMWF. *IFS documentation CY41r1*. ECMWF, Shinfield Park, Reading, RG2 9AX, U.K., 2015. https://software.ecmwf.int/wiki/display/IFS/CY41R1+Official+IFS+Documentation.

C. A. T. Ferro. Fair scores for ensemble forecasts. *Q. J. Roy. Meteor. Soc.*, 140(683):1917–1923, 2014.

F. Grazzini and L. Isaksen. North american increments. Technical report, European Centre for Medium-Range Weather Forecasts, Shinfield park, Reading, 2002.

T. Haiden, M. Janousek, P. Bauer, J. Bidlot, M. Dahoui, L. Ferranti, F. Prates, Richardson D. S., and F. Vitart. Evaluation of ECMWF forecasts, including 2014-2015 upgrades. Technical Report 765, European Centre for Medium-Range Weather Forecasts, Shinfield park, Reading, 2015.

L. Isaksen, M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher, and L. Raynaud. Ensemble of data assimilations at ECMWF. Technical Report 636, European Centre for Medium-Range Weather Forecasts, Shinfield park, Reading, 2010.

B. Khouider, A. J. Majda, and M. A. Katsoulakis. Coarse-grained stochastic models for tropical convection and climate. *P. Natl. Acad. Sci. U.S.A.*, 100(21):11941–11946, 2003.

K. Kober and G. C. Craig. Physically-based stochastic perturbations (PSP) in the boundary layer to represent uncertainty in convective initiation. *J. Atmos. Sci.*, 2016.

M. Leutbecher. Diagnosis of ensemble forecasting systems. In *Seminar on Diagnosis of Forecasting and Data Assimilation Systems, 7 - 10 September 2009*, pages 235–266, Shinfield Park, Reading, 2010. ECMWF.

M. Leutbecher and S. T. K. Lang. On the reliability of ensemble variance in subspaces defined by singular vectors. *Quarterly Journal of the Royal Meteorological Society*, 140(682):1453–1466, 2014. doi: 10.1002/qj.2229. URL http://dx.doi.org/10.1002/qj.2229.

M. Leutbecher and T. N. Palmer. Ensemble forecasting. *J. Comput. Phys.*, 227(7):3515–3539, 2008.

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*, 2017. submitted.

S. Manabe and R. T. Wetherald. Thermal equilibrium of the atmosphere with a given distribution of relative humidity. *J. Atmos. Sci.*, 24(3): 2411–259, 1967.

J.-J. Morcrette. Radiation and cloud radiative properties in the European Centre for Medium Range Weather Forecasts forecasting system. *J. Geophys. Res.-Atmos.*, 96(D5):9121–9132, 2012.

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.

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.*, 2016.

T. N Palmer. A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrisation in weather and climate prediction models. *Q. J. Roy. Meteor. Soc.*, 127(572): 279–304, 2001.

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.

M. J. Rodwell, L. Magnusson, P. Bechtold, M. Bonavita, C. Cardinali, M. Diamantakis, P. Earnshaw, A. Garcia-Mendez, L. Isaksen, E. Källén, D. Klocke, P. Lopez, T. McNally, A. Persson, F. Prates, and N. Wedi. Characteristics of occasional poor medium-range weather forecasts for europe. *B. Am. Meteorol. Soc.*, 94(9):1393–1405, 2013.

M. J. Rodwell, L. Ferranti, T. Haiden, L. Magnusson, J. Bidlot, N. Bormann, M. Dahoui, G. De Chiara, S. Duffy, R. Forbes, E. Hólm, B. Ingleby, M. Janousek, S. T. K. Lang, K. Mogensen, F. Prates, F. Rabier, D. S. Richardson, I. Tsonevsky, and M. Vitart, F. andYamaguchi. New developments in the diagnosis and verification of high-impact weather forecasts. Technical Report 759, European Centre for Medium-Range Weather Forecasts, Shinfield park, Reading, 2015.

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.

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.

G. Shutts, M. Leutbecher, A. Weisheimer, T. Stockdale, L. Isaksen, and M. Bonavita. Representing model uncertainty: stochastic parametrizations at ECMWF. *ECMWF Newsletter*, 129:19–24, 2011. [Available online at: http://www.ecmwf.int/publications/newsletters/pdf/129.pdf].

G. J. Shutts. A stochastic convective backscatter scheme for use in ensemble prediction systems. *Q. J. Roy. Meteor. Soc.*, 141(692): 2602–2616, 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.

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*, 2017.

A. M. Tompkins and G. C. Craig. Radiativeconvective equilibrium in a three-dimensional cloud-ensemble model. *Q. J. Roy. Meteor. Soc.*, 124(550):2073–2097, 1998a.

A. M. Tompkins and G. C. Craig. Time-scales of adjustment to radiative-convective equilibrium in the tropical atmosphere. *Q. J. Roy. Meteor. Soc.*, 124(552):2693–2713, 1998b.

P. A. G. Watson, J. Berner, S. Corti, P. Davini, J. von Hardenberg, C. Sanchez, A. Weisheimer, and T. Palmer. The impact of stochastic physics on tropical rainfall variability on daily time scales in global weather and climate models. *J. Geophys. Res.*, 2017. In review.

A. Weaver and P. Courtier. Correlation modelling on the sphere using a generalized diffusion equation. *Q. J. Roy. Meteor. Soc.*, 127(575):1815–1846, 2001. doi: 10.1002/qj.49712757518. URL http://dx.doi.org/10.1002/qj.49712757518.

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.

D. S. Wilks. *Statistical Methods in the Atmospheric Sciences*, volume 91 of *International Geophysics Series*. Elsevier, second edition, 2006.

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.

Y. Zhu, Z. Toth, R. Wobus, D. Richardson, and K. Mylne. The economic value of ensemble-based weather forecasts. *B. Am. Meteorol. Soc.*, 83(1):73–83, 2002.