

Stochastic Parametrization and Model Uncertainty

Palmer, T.N., R. Buizza, F. Doblas-Reyes,
T. Jung, M. Leutbecher, G.J. Shutts,
M. Steinheimer, A. Weisheimer

Research Department

October 8, 2009

*This paper has not been published and should be regarded as an Internal Report from ECMWF.
Permission to quote from it should be obtained from the ECMWF.*



European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:

<http://www.ecmwf.int/publications/>

Contact: library@ecmwf.int

©Copyright 2009

European Centre for Medium-Range Weather Forecasts
Shinfield Park, Reading, RG2 9AX, England

Literary and scientific copyrights belong to ECMWF and are reserved in all countries. This publication is not to be reprinted or translated in whole or in part without the written permission of the Director. Appropriate non-commercial use will normally be granted under the condition that reference is made to ECMWF.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error, omission and for loss or damage arising from its use.

Abstract

Stochastic parametrization provides a methodology for representing model uncertainty in ensemble forecasts, and also has the capability of reducing systematic error through the concept of nonlinear noise-induced rectification. The stochastically perturbed parametrization tendencies scheme and the stochastic backscatter scheme are described and their impact on medium-range forecast skill is discussed. The impact of these schemes on ensemble data assimilation and in seasonal forecasting is also considered. In all cases, the results are positive. Validation of the form of these stochastic parametrizations can be found by coarse-grain budgets of high resolution (e.g. cloud-resolving) models; some results are shown. Stochastic parametrization has been pioneered at ECMWF over the last decade, and now most operational centres use stochastic parametrization in their operational ensemble prediction systems - these are briefly discussed. The seamless prediction paradigm implies that serious consideration should now be given to the use of stochastic parametrization in next generation Earth System Models.

1 Introduction

Nowadays, most operational weather prediction centres run ensemble forecast systems. In doing so, they recognise that predicting the uncertainty in prognostic variables such as temperature, precipitation, wind speed and so on, is central for robust decision making across a range of weather forecast applications. Sources of forecast uncertainty include limitations on the accuracy and representativity of observations, on the methods by which these observations are assimilated into forecast models, and on the forecast models themselves.

This paper focuses on model uncertainty, i.e. the inherent uncertainties associated with computational representations of the underlying partial differential equations that govern atmospheric motion. The basis for stochastic parametrization ([Palmer, 1997, 2001](#)) is that whilst these partial differential equations may themselves be deterministic, at the computational level, the equations of motion for weather are not deterministic. For example, the bulk-formula parametrizations, largely based on the notion of ensembles of sub-grid processes in quasi-equilibrium with the grid scale flow, necessarily approximate sub-grid tendencies in a turbulent system like the atmosphere with its power-law energy spectrum. Hence we look for stochastic representations of the computational equations of motion. In an ensemble forecast, different realisations of these stochastic representations are used to generate the “model error” component of ensemble dispersion.

Sections 2 describe two complementary approaches to stochastic parametrization. The Stochastically Perturbed Parametrization Tendencies Scheme (SPPT) is based on the approach of [Buizza et al. \(1999\)](#). The Stochastic Backscatter Scheme is based on implementations in large-eddy simulation models ([Mason and Thomson, 1992](#)) and subsequent adaption for numerical weather prediction (NWP) by [Shutts \(2005\)](#). The two approaches address different aspects of the parametrization problem. SPPT is concerned with aspects of uncertainty in existing parametrization schemes. The origin of these uncertainties is partly associated with the fact that the sub-grid processes, whose effects are to be parametrized, may only have small ensemble size within a grid box. Moreover, SPPT can address additional errors that arise from the choice of the parametrization algorithms representing the physical processes and their interactions. This latter aspect includes for instance the choice of parameter settings in the parametrization algorithms. Hence SPPT generalises the output of existing sub-grid parametrizations as probability distributions. Backscatter, on the other hand, describes a physical process missing in conventional parametrization schemes. Because this process is inherently linked to underlying turbulent energy cascades in the atmosphere, the parametrized formulation of backscatter is necessarily stochastic. In this sense, backscatter can be thought of as describing aspects of structural uncertainty in conventional parametrization. The impact on medium-range forecast skill of the two schemes is documented and discussed.

Since observations are assimilated in a forecast model as part of the process of creating a set of initial conditions, model uncertainty is a component of initial condition uncertainty. At ECMWF, an ensemble data assimilation

system is being developed in order to estimate as explicitly as possible, a probability distribution of initial state. In Section 3, the impact of the stochastic parametrization schemes of Section 2 are studied in the context of this ensemble data-assimilation system.

Although devised principally for numerical weather prediction, the stochastic parametrization schemes described in Section 2 are relevant for longer timescale climate forecasts such as the seasonal to interannual timescales. There are two aspects relevant here. Firstly, as with numerical weather prediction, stochastic parametrization schemes can provide representations of model uncertainty on climate timescales. On seasonal and longer timescales, the conventional approach to the representation of model uncertainty is through the multi-model ensemble (e.g. [Palmer et al. 2004](#)). However, the multi-model ensemble concept is ad hoc: individual component models have not been developed to represent aspects of model uncertainty. As such it is to be hoped that ensembles which incorporate more rigorously defined stochastic parametrizations are capable of outperforming the multi-model ensemble. This issue is currently being studied in the ENSEMBLES project, and preliminary results are shown in Section 4. However, there is a second reason why stochastic parametrization might be especially relevant on longer climate timescales: the systematic bias of a climate model can be affected by nonlinear noise-induced rectification. Results from earlier cycles of the ECMWF model have already shown a positive impact of stochastic parametrization on model bias ([Jung et al., 2005](#); [Palmer et al., 2005](#); [Bernier et al., 2008, 2009a](#)). In Section 4 we discuss the impact of stochastic parametrization on systematic error from seasonal integrations.

How can one develop stochastic parametrizations rigorously? In Section 5 we describe an approach developed in [Shutts and Palmer \(2007\)](#) based on coarse-graining of a high resolution (potentially cloud-resolving) model. The basic approach is to define output from the high resolution model as “truth”, and to study the statistics of consequent error associated with a conventional parametrization based on coarse-grain grid-box averaged fields, where the size of these coarse-grain grid boxes are representative of typical numerical weather prediction or climate models. As discussed, a number of features of the stochastic schemes described in Section 2 can be validated using this approach.

Since the implementation of the [Buizza et al. \(1999\)](#) scheme, a number of other operational centres have begun implementing stochastic parametrization schemes in their forecast models. A brief description of these other stochastic parametrization schemes is given in Section 6.

Some concluding remarks are made in Section 7.

2 Stochastic Parametrizations

2.1 Stochastically Perturbed Parametrization Tendencies (SPPT)

In 1998, a representation of model uncertainty was introduced in the EPS by [Buizza et al. \(1999\)](#). The scheme perturbs the total parametrized tendency of physical processes with multiplicative noise. As the perturbations are applied to tendencies due to physical processes (as opposed to the dynamics), the term “stochastic physics” was coined. As this term is somewhat unspecific, it was decided to break with tradition and this class of model uncertainty representation is now referred to as *Stochastically Perturbed Parametrization Tendencies* (SPPT) in the following. The original version of SPPT by Buizza, Miller and Palmer, will be referred to as the BMP scheme. Since its introduction, the BMP scheme has not been changed. The BMP scheme uses random patterns that are piecewise constant in space and time. This was a convenient choice for the initial implementation. However, the discontinuities in the perturbations at the places and times where the random numbers change are somewhat unphysical. The revised pattern described here varies smoothly in space and time. In addition,

the distribution of perturbations is changed from a multivariate uniform distribution to a univariate Gaussian distribution. This latter change is designed to address the overprediction of heavy precipitation events.

2.1.1 The original SPPT scheme (BMP)

In the BMP scheme, multiplicative noise perturbs the net parametrized physics tendency ([Buizza et al., 1999](#)). The tendencies of the wind components u, v , temperature T and humidity q are perturbed. Let X denote the net parametrized physics tendency of any variable. For an unperturbed tendency X_c , the perturbed tendency X_p is computed as

$$X_p = (1 + r_X)X_c, \quad (1)$$

where r_X is a random number drawn from a uniform distribution in the range $[-0.5, 0.5]$. The perturbations are multivariate, i.e. different random numbers r_u, r_v, r_T, r_q are used for the four variables. In order to impose spatial correlations, the same random numbers are used in the whole column over boxes 10° by 10° in latitude and longitude. Temporal correlation is achieved by using the same random numbers over 6 consecutive model time steps, i.e. r is constant for 3 h and 4.5 h for T_L399 and T_L255 forecasts, respectively. If a critical humidity linked to saturation is exceeded due to the perturbations, the perturbations of temperature and humidity are not applied. The critical humidity is set to the saturation value for temperatures greater 250 K. For lower temperatures the critical value allows for some supersaturation in order to account for homogeneous nucleation ([Tompkins et al., 2007](#)).

2.1.2 The revised SPPT scheme

The revised SPPT scheme uses perturbations collinear to the unperturbed tendencies. For all variables $X \in \{u, v, T, q\}$, the perturbed tendency is obtained with the same random number r

$$X_p = (1 + r\mu)X_c. \quad (2)$$

The distribution from which r is drawn is close to a Gaussian distribution. The factor $\mu \in [0, 1]$ is used for reducing the perturbation amplitude close to the surface and in the stratosphere. The replacement of the multivariate distribution (r_u, r_v, r_T, r_q) of the BMP scheme by a univariate distribution is an attempt to introduce perturbations that are more consistent with the model physics. If the model state prefers to stay on a manifold, an attractor, the multivariate distribution of the BMP scheme is likely to frequently push the model state off its attractor. In contrast, the univariate perturbations of the revised scheme, will keep the perturbed state much closer to the model attractor as long as the perturbation amplitude and the curvature of the attractor are not too large. Initial testing of the revised scheme at T_L255 resolution confirmed that the introduction of univariate perturbations has a significant impact on the tail of the precipitation distribution.

The random numbers are described through a spectral pattern generator of the same type used in the kinetic energy backscatter scheme described by [Berner et al. \(2009b\)](#). This type of pattern generator has also been used by [Li et al. \(2008\)](#) to perturb parameters in order to represent model uncertainty. The spectral coefficients of r are described with auto-regressive processes of first order (AR(1)). The variance depends on the total wavenumber. A variance spectrum has been chosen that yields a spatial autocorrelation corresponding to the equivalent of a Gaussian on the sphere ([Weaver and Courtier, 2001](#)). Details are described in Appendix 8.1. Gaussian random numbers are used to force the AR(1) processes. Therefore, the pattern r in grid point space has also a Gaussian distribution as it is a linear combination of Gaussian random variables through the spectral transform. In order to limit perturbations to physically reasonable limits, the pattern is bounded to the range of ± 3 standard deviations.

Identifier	Description
NoTenPert	no tendency perturbations
SP _{BMP}	Stochastically Perturbed Parametrization Tendencies SPPT; original version cf. Buizza, Miller and Palmer (1999)
SP1 _M	revised SPPT scheme with moderate amplitude, $\sigma = 0.50$, clipped at $\pm 3\sigma$, $\tau = 6$ h, $L = 500$ km
SP1 _L	revised SPPT scheme with large amplitude, $\sigma = 0.75$, clipped at $\pm 2\sigma$, $\tau = 6$ h, $L = 500$ km
SP2	two-scale version of the revised SPPT scheme: scale 1: $\sigma_1 = 0.50$, $\tau_1 = 6$ h, $L_1 = 500$ km; scale 2: $\sigma_2 = 0.20$, $\tau_2 = 30$ d, $L_2 = 2500$ km; sum clipped at $\pm 3\sigma$, where $\sigma = (\sigma_1^2 + \sigma_2^2)^{1/2}$

Table 1: List of model uncertainty representations.

For reasons of numerical stability and physical realism, the perturbations have been tapered to zero in the lowermost atmosphere and in the stratosphere. In initial tests, tendencies were perturbed in the entire atmosphere. For standard deviations of 0.5, numerical instabilities were encountered. Further testing showed that the cause of the numerical instability are the perturbations in the lowermost part of the atmosphere. The reason is the delicate balance between model dynamics and vertical momentum transport which is established in the lowest model levels on timescales of the order of minutes. As a compromise between numerical stability and high probabilistic skill, the tendency perturbations were reduced towards zero close to the surface (factor μ in Eq. (2)). There are no tendency perturbations in the lowest ≈ 300 metres of the atmosphere. In a transition layer up to ≈ 1300 metres, the perturbations are smoothly ramped up to full amplitude. The reduction of the perturbations close to the surface removed the numerical instabilities and allowed to use perturbation amplitudes of 0.5 and larger.

In general, the net physics tendency in the stratosphere is dominated by radiative forcing. Radiative tendencies are expected to be relatively accurate in the stratosphere and with errors that are predominantly large scale, i.e. with correlation lengths far larger than 500 km. Therefore, it was decided not to apply the tendency perturbations in the stratosphere in the revised scheme. The transition zone for the tapering has been placed in the layer from 50 to 100 hPa. Note, however, that this procedure will neglect perturbations to the gravity wave drag that may be locally significant in the stratosphere.

As in the BMP scheme, the perturbations of temperature and humidity are not applied if they lead to humidity values exceeding the critical (super)saturation value. In contrast to the BMP scheme, the perturbations to humidity and temperature are also not applied, if they lead to negative humidity.

2.1.3 Experimentation

Initial tests at T_L255 resolution gave a first insight into the behaviour of the revised SPPT scheme and were used to identify suitable configurations for testing at higher resolution. The experiments described here have a resolution of T_L399 up to day 10 and T_L255 from day 10 to 15 with 62 vertical levels. All experiments use the same set of singular vector-based initial perturbations (cf. [Leutbecher and Palmer 2008](#)) and have 50 perturbed forecasts. The experimentation is with model cycle 35R1. The different tendency perturbations are evaluated over a boreal winter period (November/December 2007, 20 cases) and a boreal summer period (July/August 2008, 20 cases).

Table 1 lists the configurations for model uncertainty representations that are considered here. There are two reference experiments: In Exp. NoTenPert, model tendencies are not perturbed. Experiment SP_{BMP} uses the BMP version of the stochastically perturbed parametrization tendencies.

Three different configurations of the revised SPPT scheme are considered. In Exps. SP1_M and SP1_L, the noise

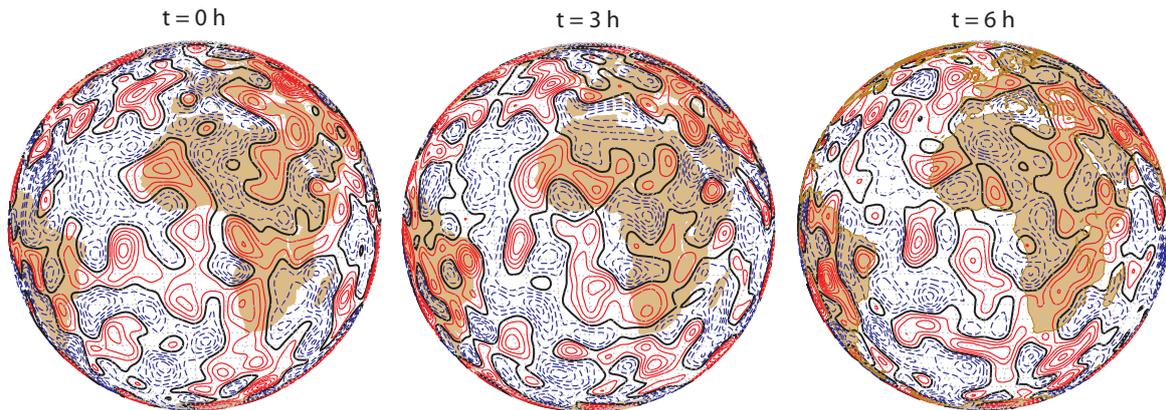


Figure 1: Example of the pattern r used in the revised scheme; contour interval 0.25; red (blue) contours correspond to positive(negative) values.

r has a horizontal correlation length scale of 500 km and a correlation time of 6 h. Thus the spatio-temporal coherence of the noise is roughly similar to the original BMP version. Figure 1 shows a particular realisation of the pattern for $SP1_M$ in three-hourly intervals. The standard deviation of the random pattern r has been set to 0.50 and 0.75 in $SP1_M$ and $SP1_L$, respectively. Coarse-graining diagnostic of Cloud Resolving Model output by [Shutts and Palmer \(2007\)](#) suggests that an uncertainty of this magnitude is not unrealistic (cf. Section 5).

In Exp. SP2, the pattern r is the sum of two independent patterns r_1 and r_2 that represent errors on different spatial and temporal scales. Both patterns r_1 and r_2 are defined by the same algorithm used for $SP1_M$ and $SP1_L$. Pattern r_1 represents fast evolving synoptic scale errors and has the same correlation characteristics and amplitude as the pattern in $SP1_M$. Pattern r_2 represents slower evolving, planetary scale errors; it has a correlation length scale of 2500 km and a correlation time of 30 d. The standard deviation of r_2 is set to 0.20. Therefore, the sum of r_1 and r_2 has a standard deviation of 0.54. Now, the different experiments will be compared in terms of upper air verification and precipitation.

2.1.4 Results: Upper air verification

Figure 2 shows the ensemble standard deviation and ensemble mean RMS error for 850 hPa temperature. All three experiments with the revised SPPT scheme have a larger ensemble spread than SP_{BMP} . In terms of increasing spread, the experiments are ordered SP_{BMP} , $SP1_M$, SP2, $SP1_L$. The increase in spread is small in the extra-tropics and quite substantial in the tropics. In all regions the agreement between ensemble RMS error and ensemble standard deviation is improved. The tropics also show a reduction of the ensemble mean RMS error with the revised SPPT.

The experiments with the revised SPPT scheme are slightly more skilful than SP_{BMP} in the extra-tropics and they are significantly more skilful in the tropics in terms of the Continuous Ranked Probability Skill Score (Figure 3). The order of experiments SP_{BMP} , $SP1_M$, SP2, $SP1_L$ in terms of probabilistic skill is consistent with the increased spread seen in Figure 2. The area under the Relative Operating Characteristic shows a more noticeable impact of the revised SPPT scheme in the extra-tropics than the Continuous Ranked Probability Score (Fig. 4).

Other variables ($Z500$, $v850$, $v200$, $u850$, $u200$, not shown) show a similar signal in terms of spread and probabilistic skill. Finally, we observe that the BMP scheme is also clearly beneficial, in particular in the tropics compared to Exp. NoTenPert.

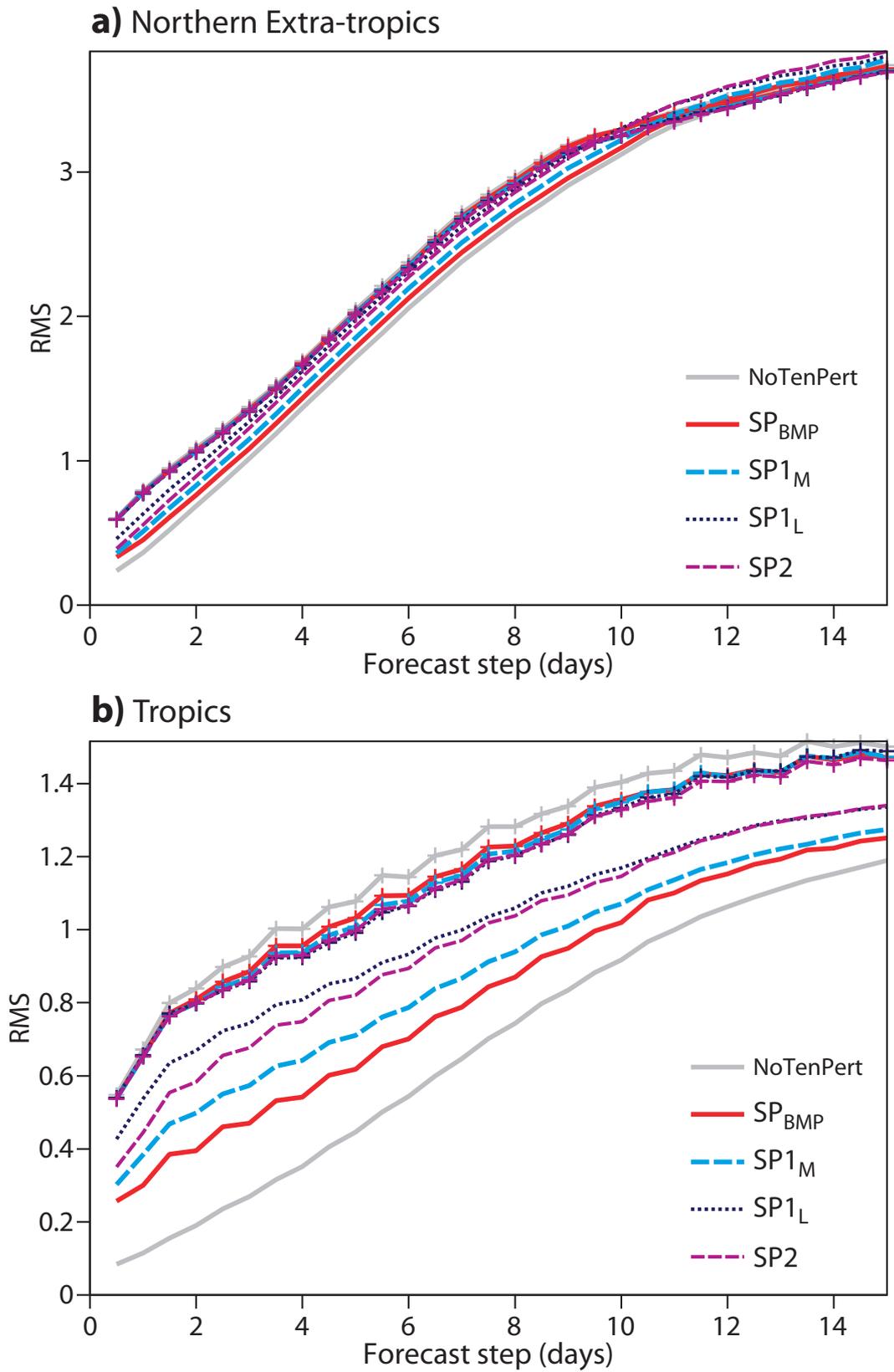


Figure 2: Ensemble standard deviation (plain curves) and ensemble mean RMS error (curves with symbols on) for 850 hPa temperature.

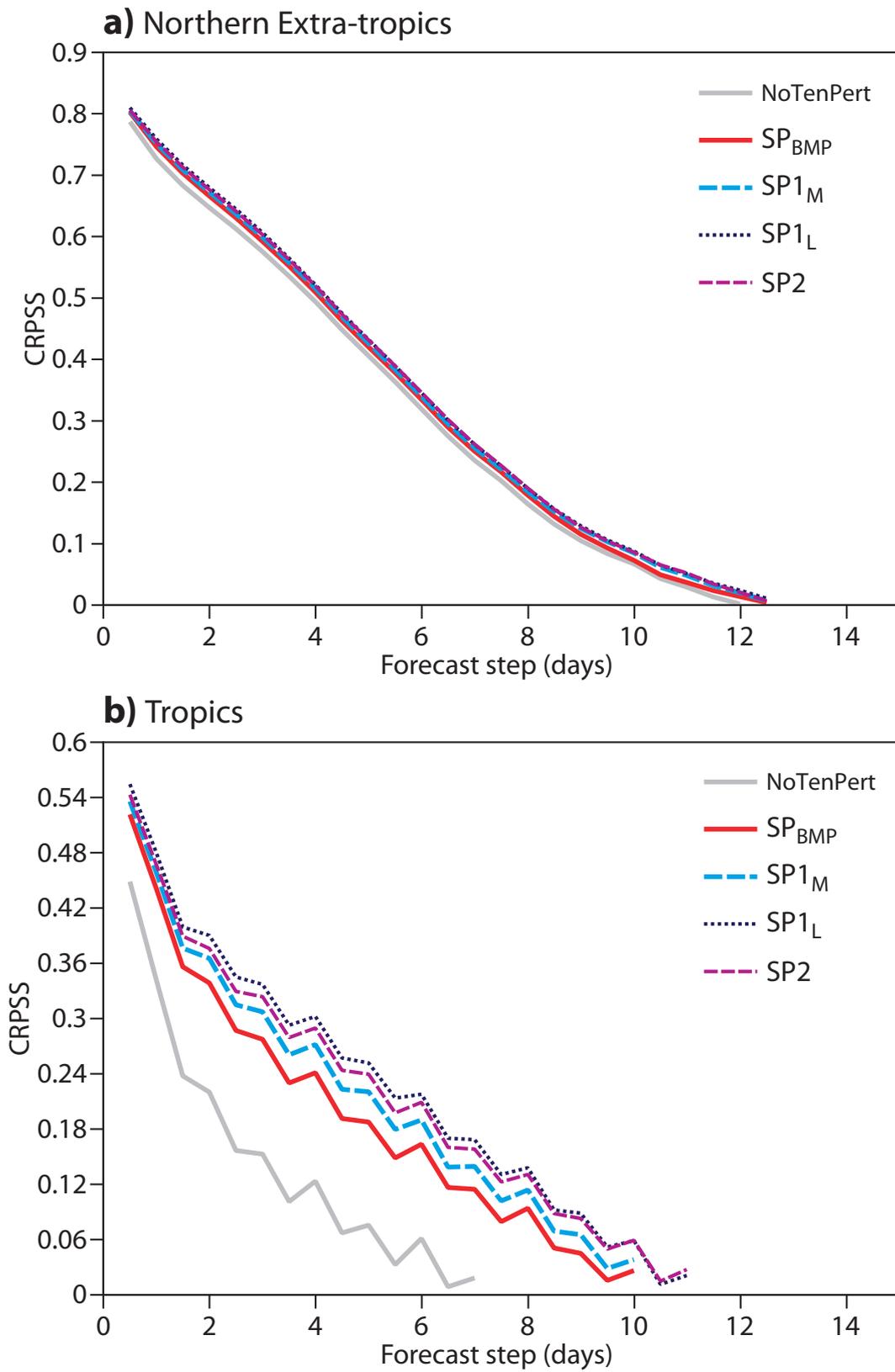


Figure 3: Continuous Ranked Probability Skill Score for 850 hPa temperature.

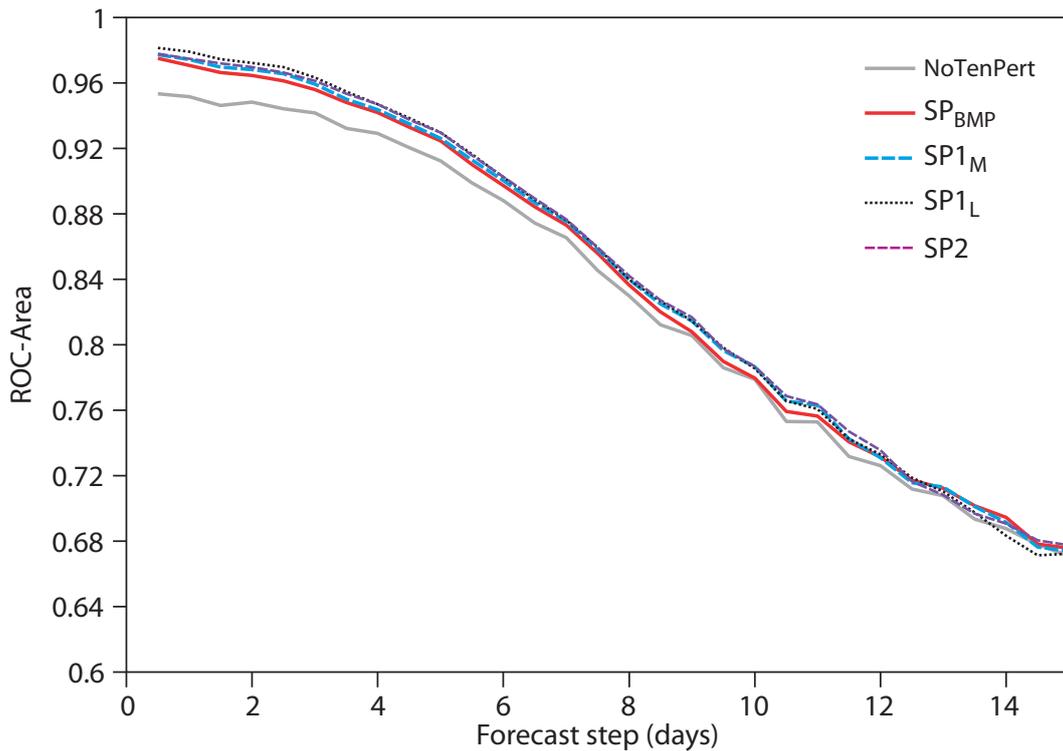


Figure 4: Area under the Relative Operating Characteristic for 850 hPa temperature and events where the anomaly is below minus one climatological standard deviation. Area: Northern Extra-tropics.

2.1.5 Results: Precipitation

The impact of the revised SPPT on precipitation has been evaluated in two ways. Firstly, the model climatology of forecasts with tendency perturbations is compared with the model climatology of forecasts from NoTenPert. Secondly, the ensemble forecasts are verified against 24-hour SYNOP precipitation data.

Proxy climatological distributions of the model precipitation have been determined on the grid-scale for 6-hour accumulations. The diagnostic is performed on the reduced Gaussian grid of the T_L399 model (N200) and for all forecast steps from 6 to 240 h. The frequency of events ranging from 0.1 mm / 6 h to 1000 mm / 6 h has been determined on the reduced Gaussian grid and then integrated over different regions. The averaging involves all 40 start dates and all 50 members of the ensemble. Experiment NoTenPert serves as a reference. Instantaneous rain rates from unperturbed T_L799 and T_L1279 model runs have been compared with 1D-Var rain rates derived from SSMI radiances up to rain rates equivalent to 40mm/6h (Bechtold, personal communication). This comparison shows a good agreement between model and data up to the highest rain rates.

Figure 5 shows the ratio of the frequencies of precipitation events between forecasts with tendency perturbations and forecasts without tendency perturbations. For the moderate thresholds up to 10 mm/6 h, the forecasts with tendency perturbations have nearly the same climatology as the forecasts without tendency perturbations. However, for the heavier precipitation events, there is a significant discrepancy. Experiment SP_{BMP} has about 50% more events of 40 mm/6 h in the extra-tropics than NoTenPert. In the tropics, the impact of BMP is more drastic. There are 3 times as many events of 40 mm/6 h in SP_{BMP} than in NoTenPert. The tail of the precipitation distribution is less affected by the revised SPPT scheme in particular $SP1_M$ and $SP2$. For events exceeding 10 mm/6 h, the ratio between perturbed and unperturbed forecasts is significantly closer to one in all

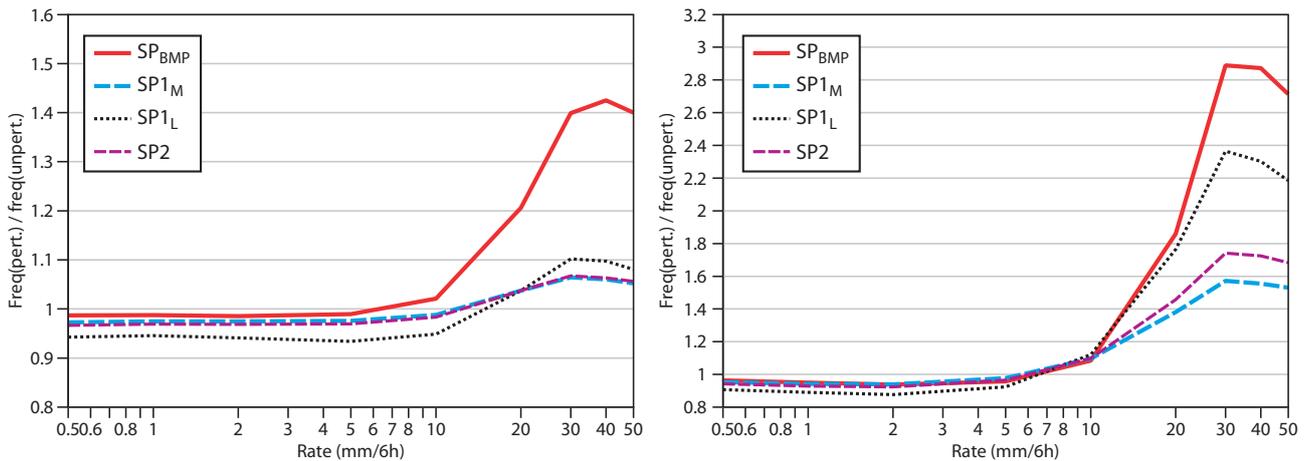


Figure 5: Precipitation frequency ratios between forecasts using tendency perturbations and forecasts without tendency perturbations. Northern Extra-tropics (left), Tropics (right).

regions. The revision of SPPT has also a small impact on the frequency of the moderate precipitation events in the extra-tropics. The ratio is slightly smaller than one with the revised scheme whereas it is closer to one with the BMP scheme. The reduction of the precipitation frequency for the moderate thresholds is somewhat larger for $SP1_L$.¹

The impact of the revision of SPPT has been evaluated with 24-hour accumulations of precipitation from SYNOP reports. Verification of this kind quantifies exactly the skill of ECMWF products such as the Meteogram. However, it is unclear whether such verification should be used to optimise the EPS. This is because precipitation exhibits variability on scales not resolved by the model (although the use of a long accumulation period of 24 hours will ameliorate this representativeness issue to some extent).

The bias of the predicted precipitation normalised with a climatological mean has been determined from station data. Experiment $SP1_M$ has a lower bias than SP_{BMP} in the Northern Extra-tropics and the Tropics at all lead times (not shown). The reduction is statistically significant at all lead times. The reduced bias is consistent with the overall reduction of the precipitable water noted in the climate runs that will be discussed in Section 4.1. Compared to NoTenPert, SP_{BMP} reduces (increases) the bias in the extra-tropics (tropics).

The probabilistic score has been evaluated with the Brier score. Experiment $SP1_M$ is more skilful than SP_{BMP} for precipitation events with thresholds between 1 and 10 mm per day in the extra-tropics and for events of 1 and 5 mm per day in the tropics (5 mm/d shown in Fig. 6, other thresholds not shown). The impact on the prediction of events of 20mm per day is neutral in the extra-tropics and negative in the tropics. Although the negative impact for 20 mm/d is statistically significant in the tropics this is not of serious concern as the more extreme events are also likely to be those affected most by the representativeness issue mentioned before. The $SP2$ is more skilful than $SP1_M$ and $SP1_L$ is even more skilful than $SP2$ for both tropics and extra-tropics and all thresholds. Last but not least, NoTenPert is statistically significantly worse than SP_{BMP} for all thresholds and both tropics and extra-tropics.

2.1.6 Discussion

Currently, configuration $SP1_M$ is being tested for operational implementation in cycle 35r3 in E-suite mode. The results indicate that further improvements in skill are feasible by either increasing the perturbation amplitude

¹see also discussion on drying in Section 4.1.

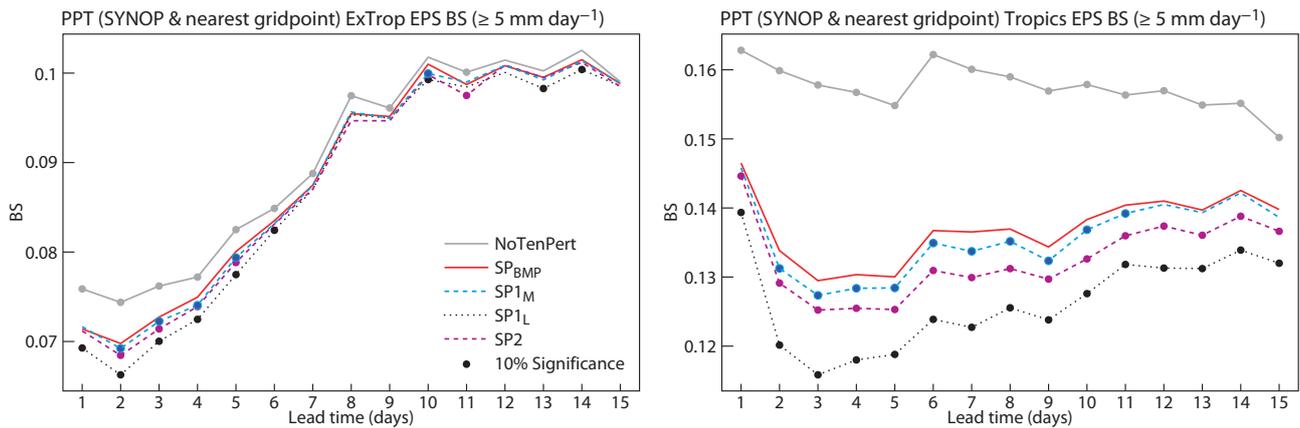


Figure 6: Brier score for 24-hour precipitation accumulations for events of 5 mm/d^{-1} . Left: Northern Extra-tropics (30°N – 90°N), right: Tropics (30°S – 30°N). The lead times at which the score differences (between an experiment and SP_{BMP}) are statistically significantly different from zero at the 10% level are marked with a dot. Verification against SYNOP data, joint sample of 40 cases.

($SP1_L$) or using the two-scale version (SP2) or even, conceivably, a two-scale version with increased amplitude for the fast synoptic-scale pattern. Work is in progress to improve the treatment of supersaturation to avoid a dry humidity bias.

2.2 Spectral Stochastic Backscatter (SPBS)

2.2.1 Background

At the time of development of BMP, Paul Mason at the Met Office suggested the possibility of perturbing the dynamical tendencies of the forecast model in manner analogous to the stochastic backscatter technique used in Large Eddy Simulation (LES) (*Mason and Thomson, 1992*). A simple spectral implementation of kinetic energy backscatter was tested in the Met Office Unified model and shown to be suitable for use in ensemble prediction systems (*Evans et al., 1998*). The scheme computed a streamfunction forcing pattern that was expressed as a truncated sum of spherical harmonics with each amplitude coefficient evolving according to a first-order autoregressive process (AR1) in time. The Met Office study demonstrated the feasibility of generating an impact on forecasts by day 8 using forcing perturbations that corresponded to flow accelerations of a few ms^{-1} per day. Cloud-resolving model simulations of deep convection carried out at the Met Office had suggested an upscale energy scale leading to mesoscale flow accelerations of this order (*Vallis et al., 1997*).

In 2003 work began on a stochastic kinetic energy backscatter scheme that it was hoped would supplement BMP. The dissipation rate calculation, which is fundamental to LES, was reinterpreted for the purposes of its use in NWP and deemed to be composed of contributions relating to numerical energy loss due to explicit and implicit diffusion; energy loss from the gravity wave drag parametrization (including flow blocking), and convective kinetic energy input (*Shutts, 2005*). Notably however, the contribution to the total dissipation rate from vertical mixing was omitted on the basis that the notional eddy size (e.g. of boundary layer eddies) is much too small, and the number of turbulent eddies per gridbox too large, for statistical fluctuation in the near-gridscale Reynolds' stresses. In contrast, mesoscale convective systems, which are neither explicitly parametrized nor sufficiently well resolved, can locally inject large flow perturbations into regions with large convective available potential energy. Likewise, mesoscale mountain ranges that are poorly resolved in forecast models may be associated with large, unpredictable wave drag or eddy stresses that impact on the large-scale, balanced

dynamics.

By perturbing streamfunction (or equivalently vorticity), the forecast model responds with less noise than if the wind components were independently perturbed (in fact options exist in the present code to compute temperature perturbations consistent with the linear or non-linear balance equations). Unlike stochastic backscatter's use in LES, the streamfunction forcing perturbations in the IFS are horizontally non-divergent and provide substantial energy input across the whole spectrum of wavenumbers. As will be seen later, generating sufficient impact on ensemble spread requires considerable input of energy in the sub-synoptic scales of motion. In this respect our approach to backscatter differs from *Frederiksen and Davies (1997)* who consider upscale energy transfer in two-dimensional flow on the sphere. Their work emphasises the injection of energy near the truncation scale of a forecast model whereas stochastic backscatter includes upscale energy transfer from unbalanced motions associated with convection and gravity waves.

The first stochastic backscatter scheme tested at ECMWF was called CASBS (Cellular Automaton Stochastic Backscatter) and used a Cellular Automaton (CA) pattern generator to define the form of the streamfunction forcing field. As in LES backscatter, the amplitude of the forcing function was modulated by the square-root of a dissipation rate field. The choice of a CA was originally motivated by the idea of generating patterns that would resemble the organization of convective cloud clusters in which the scale and evolution of the patterns could be linked to the local large-scale flow through their corresponding rulesets (*Palmer, 2001*). Whilst this approach is still under investigation using probabilistic CAs, a more pragmatic approach was to revert to the original spectrally-based scheme that had been tested at the Met Office.

Using a triangularly-truncated spherical harmonic expansion it is possible to ensure that the streamfunction forcing field is spatially-isotropic on the sphere and has a power spectrum of a chosen form. The auto-correlation time scale of the AR1 process for the spectral coefficients can also be made a function of wavenumber and calibrated using the procedure outlined in Section 5. The resulting scheme, called Spectral Stochastic Backscatter Scheme (SPBS) is described fully in *Berner et al. (2009b)* and has been the subject of considerable development in the last year. A major new extension not present in the Berner et al formulation is the introduction of vertical phase correlations for which autocorrelation scale is a function of horizontal spectral wavenumber.

2.2.2 The pattern generator and net streamfunction forcing

SPBS is conveniently formulated in terms of a streamfunction forcing function (F_ψ) prescribed on each model level and given by:

$$F_\psi = \left(\frac{b_R D_{tot}}{B_{tot}} \right)^{1/2} F_{\psi*} \quad (3)$$

where $F_{\psi*}$ is a three-dimensional random pattern field constructed to have specific statistical properties and which, if used by itself as a streamfunction forcing, would imply an ensemble-mean energy input rate of B_{tot} . $D_{tot}(x, y, z, t)$ is the total dissipation rate and b_R is a backscatter ratio so that the factor $(b_R D_{tot}/B_{tot})^{1/2}$ locally rescales the energy input rate to be $b_R D_{tot}$. Since F_ψ only affects the rotational wind, it is less likely to destroy the dynamical balance between pressure and wind fields than an arbitrary wind forcing. The spectral power distribution pattern field is assumed to be of the form of a power law in spherical harmonic degree n with the exponent inferred from a coarse-graining procedure (see Section 5). In all of the forecasts presented here that use SPBS, the auto-correlation time scale is set 25000 s (i.e. ≈ 7 hours). Further mathematical details of SPBS can be found in Appendix 8.2.

2.2.3 The dissipation rate calculation

The total dissipation rate D_{tot} is computed on all model levels and is made up of three components:

- numerical dissipation rate (D_{num}) as an estimate of the combined contributions from explicit biharmonic diffusion and interpolation error in the semi-Lagrangian advection scheme
- the combined effect of kinetic energy dissipation due to orographic Gravity Wave Drag (GWD) and flow blocking (D_{GWD})
- the rate of kinetic energy export from sub-gridscale deep convection into the resolved flow (D_{con}).

The numerical kinetic energy dissipation rate is given by the expression

$$D_{num} = \alpha_{num} K |\nabla \zeta|^2 \quad (4)$$

where ζ is the relative vorticity, K is the bi-harmonic diffusion coefficient and α_{num} is a factor to augment the dissipation rate to include the effects of semi-Lagrangian interpolation error (currently set to 3).

The gravity wave/mountain drag contribution to the energy dissipation rate (D_{GWD}) is obtained simply as the vector product of the parametrized wind vector tendency with the wind itself.

The convective contribution is not a dissipation rate as such but a term that represents the rate of transformation of convectively-generated kinetic energy to quasi-balanced, resolved flow. Specifically, D_{con} is given by:

$$D_{con} = \alpha_{con} \left(\sin(\phi) + \frac{\zeta}{2\Omega} \right)^2 M_d \rho \bar{w}^2 \quad (5)$$

where α_{con} is an area fraction factor, ϕ is latitude, Ω is the Earth's rotation rate, M_d is the convective mass detrainment rate and \bar{w} is a vertically-averaged convective updraught velocity. Excluding the bracketed factor, this term is a parametrization for the rate of detrainment of convective cloud kinetic energy to the resolved scales. The bracketed factor is a normalized absolute vorticity equivalent to $(\zeta + f)/2\Omega$. It approximately expresses the fraction of the kinetic energy released that is captured in balanced motion and will be referred to as the 'inertial stability factor' (*Shutts and Gray, 1994*). Deep convective mass transfer creates distinctive potential vorticity (PV) features in an environment with background PV. Near the equator, the background PV is very weak and so most of the convective energy release goes into unbalanced modes whereas in middle latitudes mesoscale convective systems generate intense PV anomalies that influence the development of baroclinic wave systems.

The streamfunction forcing F_ψ given by eq. (3) is evolved in time for each EPS member using the AR1 process defined in eq. (21) (see appendix) and the different random number sets $r_{m,n}^j$ ensure different model error forcing patterns.

2.2.4 Impacts in the IFS

Our results focus on ensemble forecasts with start dates in July and August so that the effect on tropical cyclone frequency could be monitored. Early experience had shown that SPBS in conjunction with SP_{BMP} led to an excessively high frequency of tropical cyclones and tropical depressions. In this section, attention will be focused on the temperature at 850 hPa (T850) and for reference the corresponding results for SP_{1M} are shown.

Figure 7 shows the error of the ensemble-mean and spread (left column) together with the associated Continuous Rank Probability Skill Score (CRPSS; right column) for T_L255 ensemble forecasts made with (i) no stochastic tendency perturbations; (ii) revised SPPT ($SP1_M$) and (iii) SPBS alone with $b_R = 0.085$. It can be seen that the impact of SPBS is comparable with $SP1_M$ except in the first 3 days where SPBS produces less additional spread compared to $SP1_M$. This is probably due to the fact that SPBS does not directly force temperature whereas SPPT does. Both SPBS and $SP1_M$ substantially increase ensemble spread (particularly in the tropics) and this is reflected in the CRPSS which are increased over the forecasts without stochastic perturbations. The error in the ensemble-mean is reduced with the biggest reduction being in the tropics with SPBS.

Figure 8 shows the corresponding results in T_L399 EPS forecasts using a somewhat higher backscatter ratio of 0.1. Even with this higher value of b_R the amount of spread using backscatter is less than that due to $SP1_M$ and points to a model resolution dependence of SPBS. The improvement in CRPSS is greatest in the tropics where the spread deficiency is worst and least in the southern hemisphere extra-tropics.

2.3 SPPT and SPBS combined

In this section the impact of backscatter is examined when it is added to the revised SPPT scheme $SP1_M$ since it is envisaged that the revised SPPT scheme is likely to become operational first given its clear improvement over the current scheme SP_{BMP} . As in Section 2.2, Figure 9 shows T850 spread/error and CRPSS for T_L255 EPS forecasts (for the same 20 start-dates) using $SP1_M$ and $SP1_M+SPBS$ with $b_R = 0.07$. It is evident that both spread and skill are improved with the addition of backscatter although without more experimentation one cannot be sure this blend of the two schemes is optimal. For the most part there is still insufficient spread to match the error of the ensemble-mean and it is not clear that increasing both tendency variance and backscatter ratio would increase skill.

For the corresponding EPS forecasts ($T_L399 \rightarrow T_L255$ at day 10) using $b_R = 0.1$ the additional spread generated by SPBS is noticeably smaller (Figure 10) and so the impact on the CRPSS is reduced relative to the T_L255 EPS forecasts. The reason for this is currently under investigation and it maybe that a larger b_R would suffice to increase the spread without introducing undesirable effects (e.g. excessive tropical cyclone frequency). One problem that has yet to be addressed for EPS forecasts is that the reduction in horizontal resolution at day 10 means that the chosen b_R becomes inappropriate for the T_L255 phase of forecast. Whilst not that noticeable in these T850 results, this leads to unreasonable spread growth after day 10 and a deterioration in many skill scores.

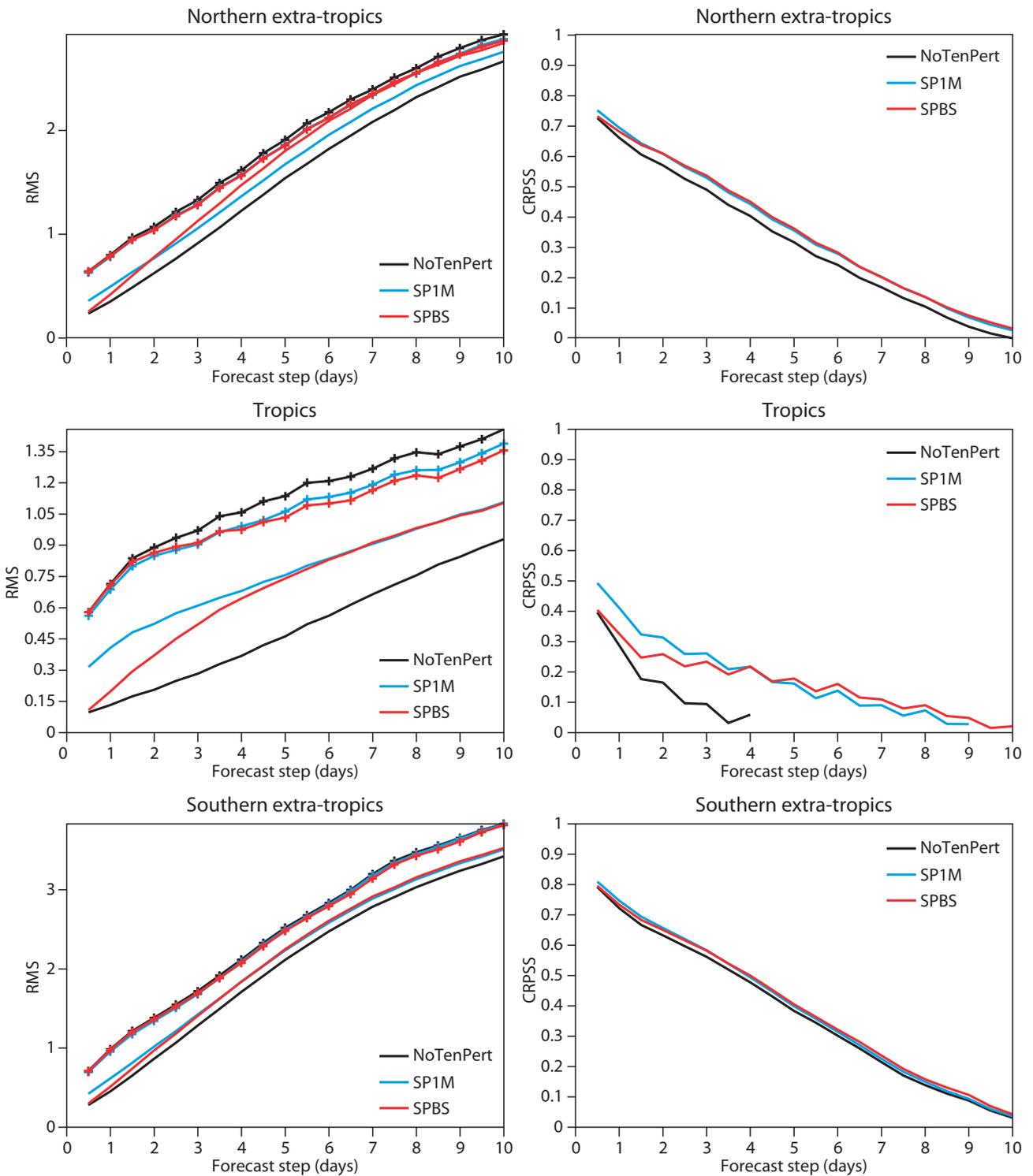


Figure 7: T_{L255} 850 hPa temperature ensemble spread/error and skill scores for NoTenPert, $SP1_M$ and SPBS (BSR=0.085) using 20 start dates running every other day from July 24 2008 to August 31 2008. Left column shows rms error of the ensemble-mean (lines with cross) and rms spread about the ensemble-mean. The right column show the Continuous Rank Probability Skill Scores. The top row is for the northern hemisphere extra-tropics; the middle row is for the tropics and the bottom row is for the southern hemisphere extra-tropics.

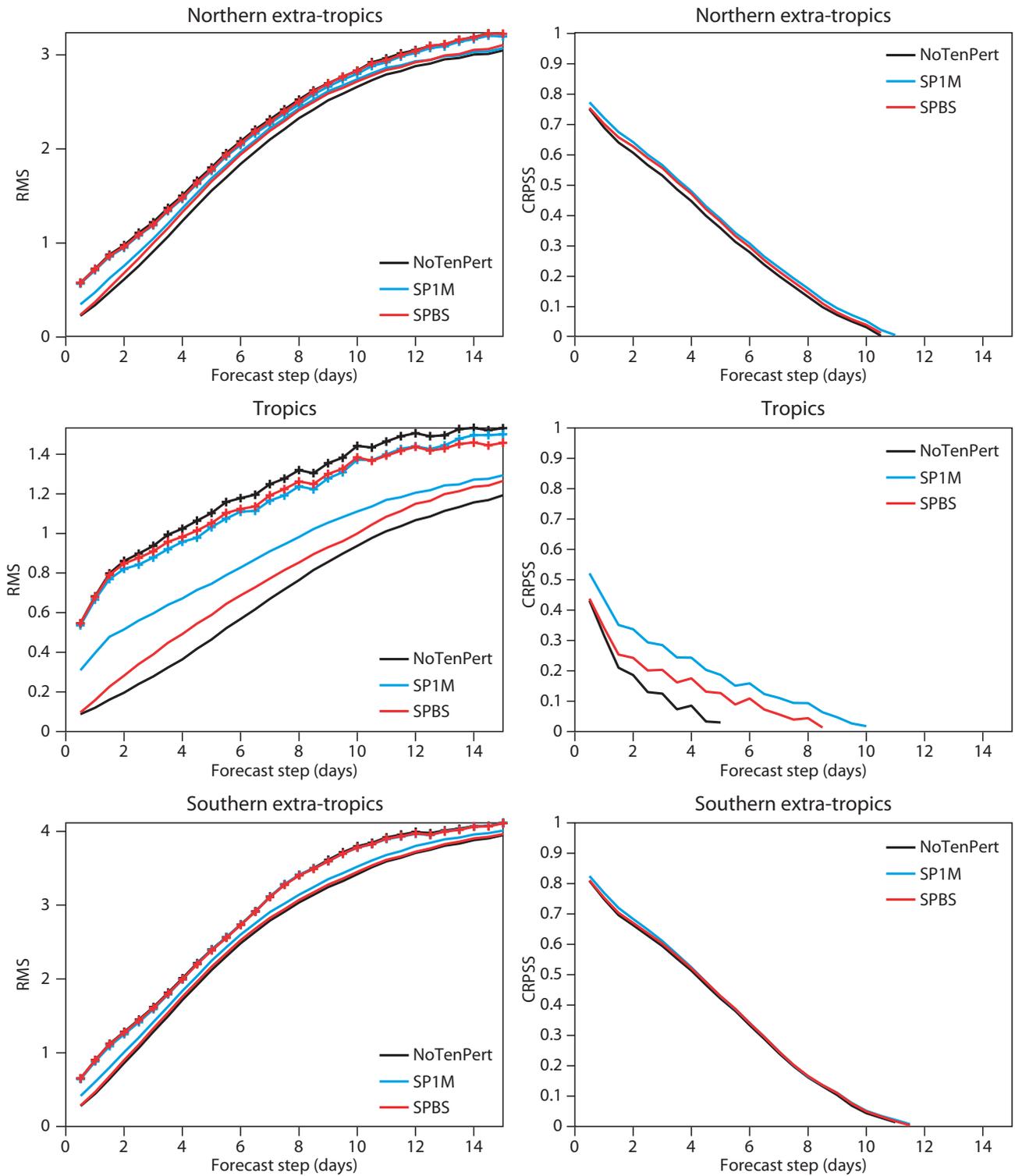


Figure 8: EPS ($T_{L399} \rightarrow T_{L255}$) 850 hPa temperature ensemble spread/error and skill scores for NoTenPert, SP1M and SPBS(BSR=0.1) 20 start dates running every other day from July 24 2008 to August 31 2008. Left column shows rms error of the ensemble-mean (lines with cross) and rms spread about the ensemble-mean. The right column show the Continuous Rank Probability Skill Scores. The top row is for the northern hemisphere extra-tropics; the middle row is for the tropics and the bottom row is for the southern hemisphere extra-tropics.

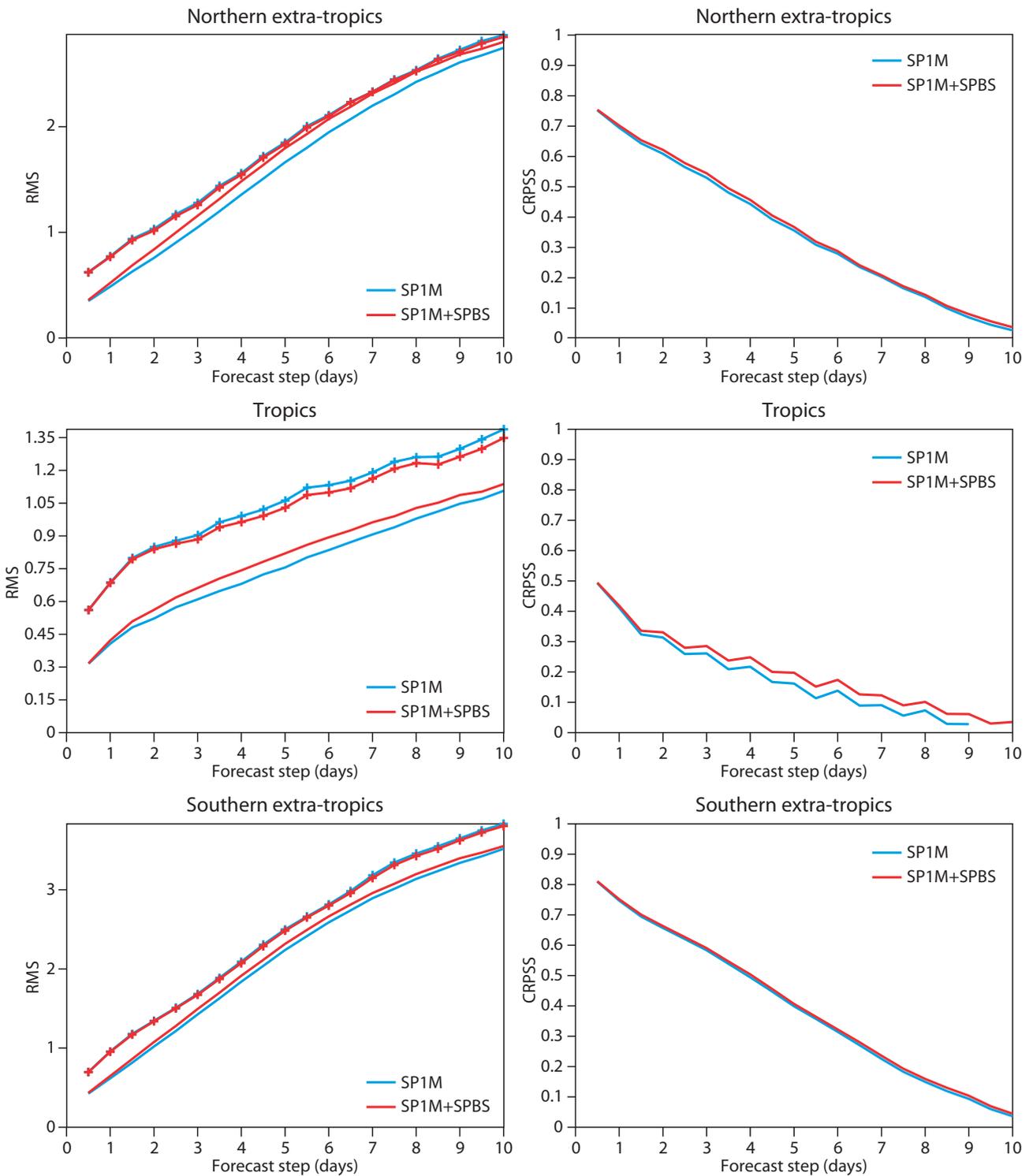


Figure 9: T_{L255} 850 hPa temperature ensemble spread/error and skill scores for $SP1_M$, $SP1_M+SPBS(BSR=0.07)$ using 20 start dates running every other day from July 24 2008 to August 31 2008. Left column shows rms error of the ensemble-mean (lines with cross) and rms spread about the ensemble-mean. The right column show the Continuous Rank Probability Skill Scores. The top row is for the northern hemisphere extra-tropics; the middle row is for the tropics and the bottom row is for the southern hemisphere extra-tropics.

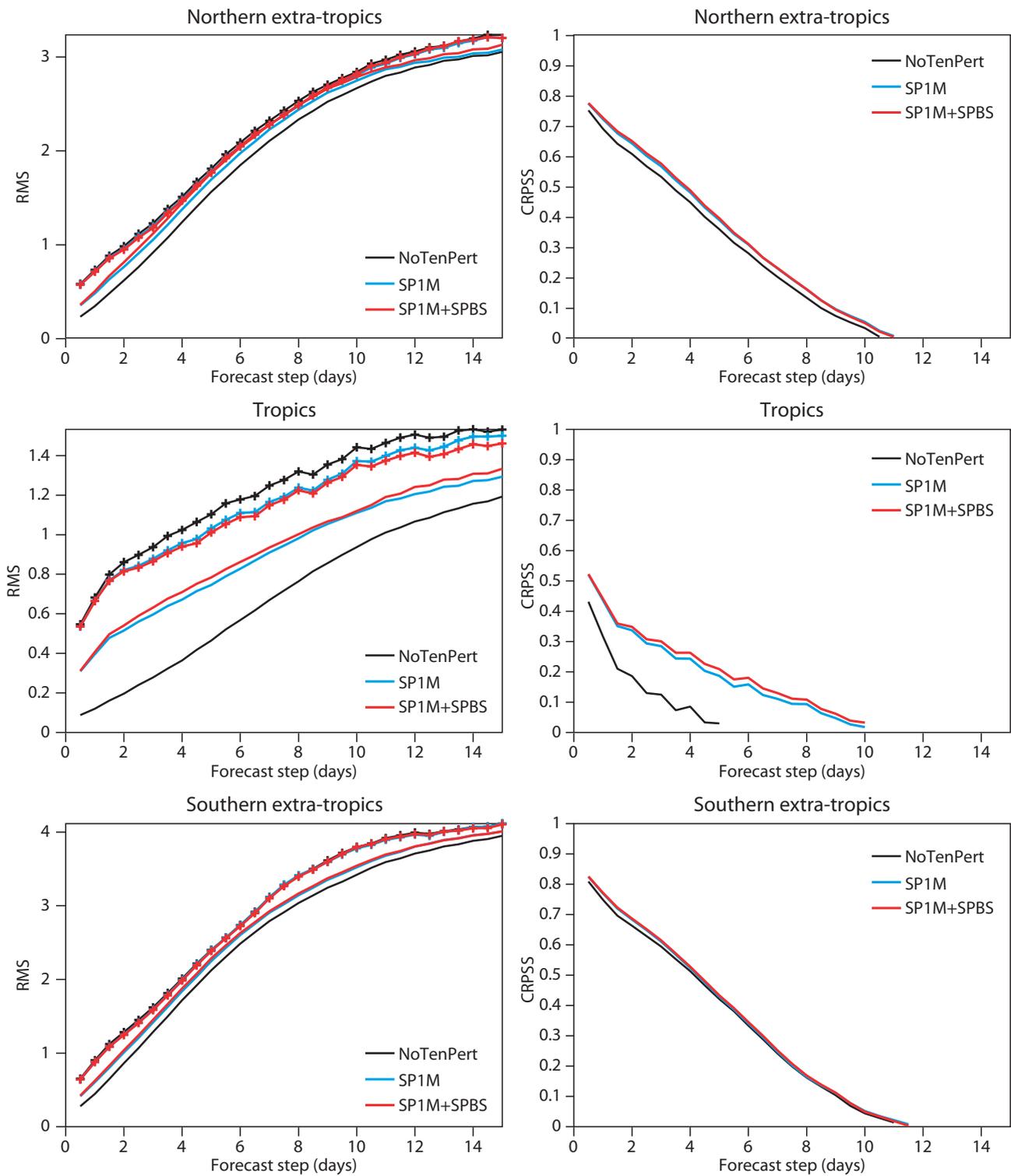


Figure 10: EPS ($T_{L399} \rightarrow T_{L255}$) 850 hPa temperature ensemble spread/error and skill scores for $SP1_M$, $SP1_M+SPBS(BR=0.1)$ using 20 start dates running every other day from July 24 2008 to August 31 2008. Left column shows rms error of the ensemble-mean (lines with cross) and rms spread about the ensemble-mean. The right column show the Continuous Rank Probability Skill Scores. The top row is for the northern hemisphere extra-tropics; the middle row is for the tropics and the bottom row is for the southern hemisphere extra-tropics.

3 Use of stochastic parametrization schemes in ensemble data assimilation

One of the key aspects that ensemble prediction systems need to simulate to provide accurate probabilistic predictions is the effect on forecast error of initial uncertainties. At ECMWF these uncertainties have been simulated with singular vectors (SVs), perturbations characterized by the fastest growth, measured using a total energy norm (*Buizza and Palmer, 1995; Palmer et al., 1998*), over a finite time interval. In the current system different sets of singular vectors are used to better sample the initial uncertainties. Initial-time SVs growing into the first 48 hours of the forecast range, which represent uncertainties growing during the forecast time, are mixed with evolved SVs computed to grow during the 48 hours leading to the analysis time, which represent uncertainties that have been growing during the current and past data-assimilation cycles. The initial-time and evolved SVs are combined and scaled to have an amplitude comparable to the analysis error estimate provided by the ECMWF data assimilation system (*Barkmeijer et al., 1999*).

Buizza et al. (2008) showed that perturbations defined by an ensemble data assimilation (EDA) system provide a better representation of the uncertainties growing during the data assimilation cycle than the set of the evolved singular vectors currently used in the operational EPS. In their work, ensembles of 10 perturbed and 1 unperturbed analyses were generated by randomly perturbing the observations in each single assimilation cycle at $T_L399L91$ resolution, with a 12-hour 4D-Var with $T_L159L91$ inner loops, and by including a version of the stochastic backscatter scheme (*Shutts, 2005; Berner et al., 2009b*). For each observation (apart from cloud-track winds) perturbations were defined by randomly sampling a Gaussian distribution with zero mean and standard deviation defined by the observation error standard deviation. For cloud-track wind observations, perturbations were horizontally correlated (*Bormann et al., 2003*). Sea-surface temperature fields were also perturbed with correlated patterns as in the current operational ECMWF seasonal ensemble forecasting system (*Vialard et al., 2005*). *Buizza et al. (2008)*'s results indicated that if used alone to define EPS initial uncertainties, EDA-based perturbations lead to an under-dispersive and less skilful ensemble than the one based on initial-time SVs only. By contrast, combining the EDA and the initial-time SVs improves the ensemble spread, reduces the ensemble-mean error, and provides more skilful probabilistic forecasts than the current operational system based on initial-time and evolved SVs.

Work is in progress to assess whether the initial-time spread of the EDA-perturbations can be increased by using the more recent versions of the stochastic parametrization schemes discussed above (see *Isaksen et al. 2007* for earlier results). Ensembles of analyses have been generated without stochastic schemes EDA(NOST), with a version of the SPBS scheme with vertically-correlated perturbations EDA(SPBS_{VC}), with a version of the revised SPPT scheme EDA(SP1_{M45}) (generated with the revised scheme with a 0.45 standard deviation instead of 0.5), and with a combination of the two schemes EDA(SP1_{M45}+SPBS_{VC}).

Figure 11 shows the impact of the stochastic schemes on the 10-day average standard deviation of the EDA analyses over an area covering the tropical and northern Pacific Ocean for kinetic energy at 700 hPa (KEN700)—one of the variables for which the stochastic schemes have the largest impact. Figure 12 gives a more complete view of the increase in spread that the two stochastic scheme induce when used alone or in conjunction, measured in terms of geopotential height at 500 (Z500), temperature and kinetic energy at 850 and 700 hPa over both the Northern Hemisphere and the tropics.

Results indicate that over Northern Hemisphere the revised SPPT scheme (SP1_{M45}) increases the average ensemble spread by 20-to-40% and the backscatter scheme SPBS_{VC} by 5-to-20%, with both schemes having the largest impact on the ensemble spread in terms of kinetic energy. Over the tropics (between 30 degrees South and 30 degrees North), the revised SPPT scheme (SP1_{M45}) has a larger impact than over the Northern Hemisphere, with the ensemble spread increasing by 30-to-70%. Over the tropics, the backscatter scheme SPBS_{VC} increases the spread only by 2-to-7%. Over both areas, the apparent lack of impact of SPBS on T850 ensemble spread is partly due to the absence of direct temperature forcing in the scheme and partly due to the choice

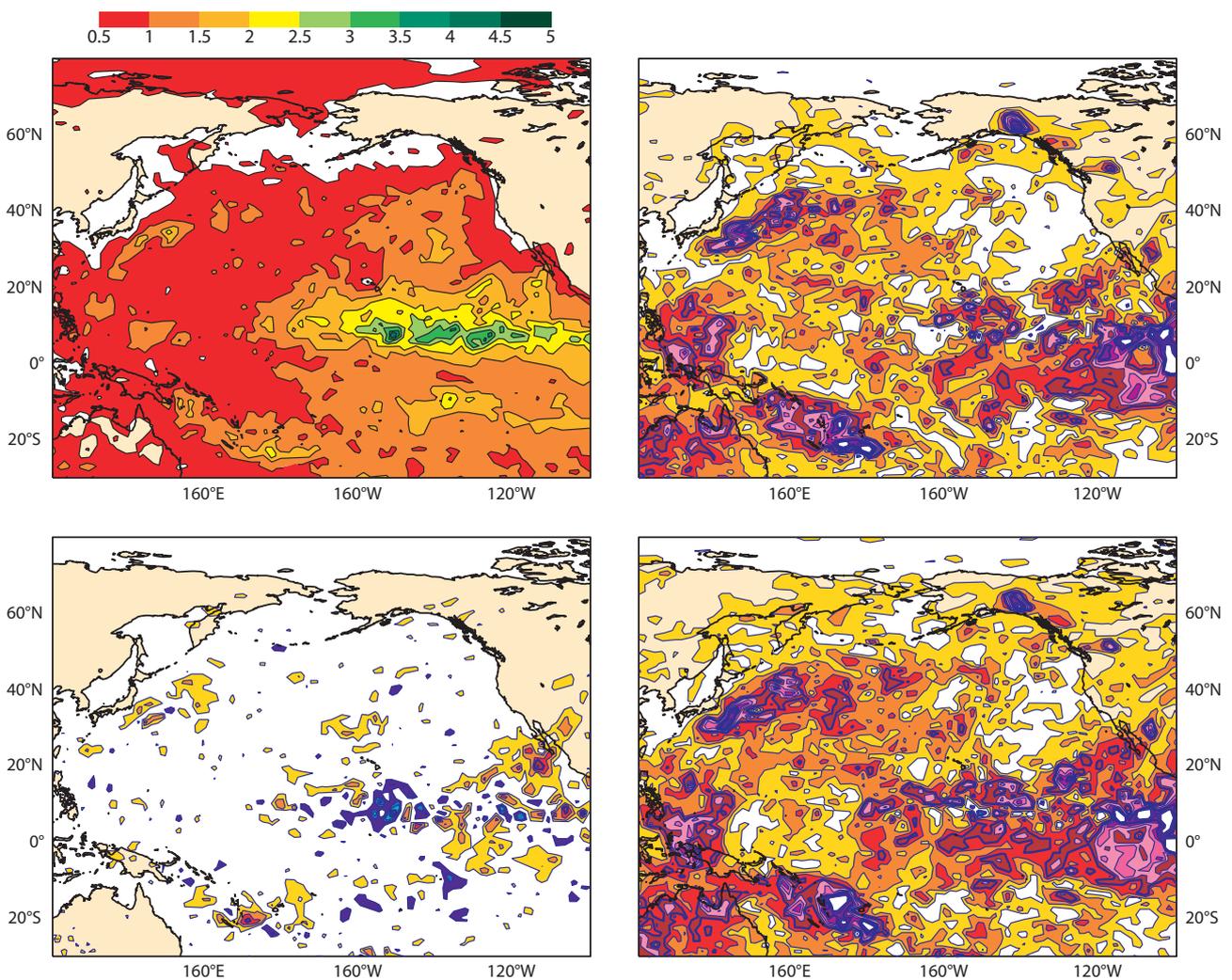


Figure 11: Top-left panel: 10-day average (from 2008-10-10 to 20) std of the EDA measured in terms of kinetic energy per unit mass at 700 hPa generated without stochastic schemes. Top-right: difference between the std of the EDA generated with the revised SPPT scheme $SP1_{M45}$ and without stochastic schemes. Bottom-left: difference between the std of the EDA generated with the stochastic backscatter scheme $SPBS_{VC}$ and without stochastic schemes. Bottom-right: difference between the std of the EDA generated with $SP1_{M45}+SPBS_{VC}$ and without stochastic schemes. Contour interval: $0.5 \text{ m}^2/\text{s}^2$ for full field, and $0.25 \text{ m}^2/\text{s}^2$ for differences.

of backscatter ratio. For T_L399 forecasts, a backscatter ratio of twice the value used here would have been a more appropriate choice. The experiments will be rerun using the latest implementations of the schemes, which include a standard deviation of 0.5 for SPPT and a backscatter ratio of 0.1 for SPBS. Overall, the revised SPPT scheme ($SP1_{M45}$) has a larger impact, and the combined use of the two schemes leads to a stronger impact on the ensemble spread.

In data-assimilation, an ensemble of analyses can be used to compute flow-dependent background error statistics to improve the assimilation of observations (more specifically, the plan is to run the latest high-resolution data-assimilation cycle with background error statistics computed using a lower-resolution ensemble of analyses run during the previous data assimilation cycle). Work is in progress to assess the impact of the stochastic schemes on the spread of the EDA analyses, and thus on the background statistics.

In ensemble prediction, a very recent set of experiments has indicated that the use of the revised SPPT scheme in

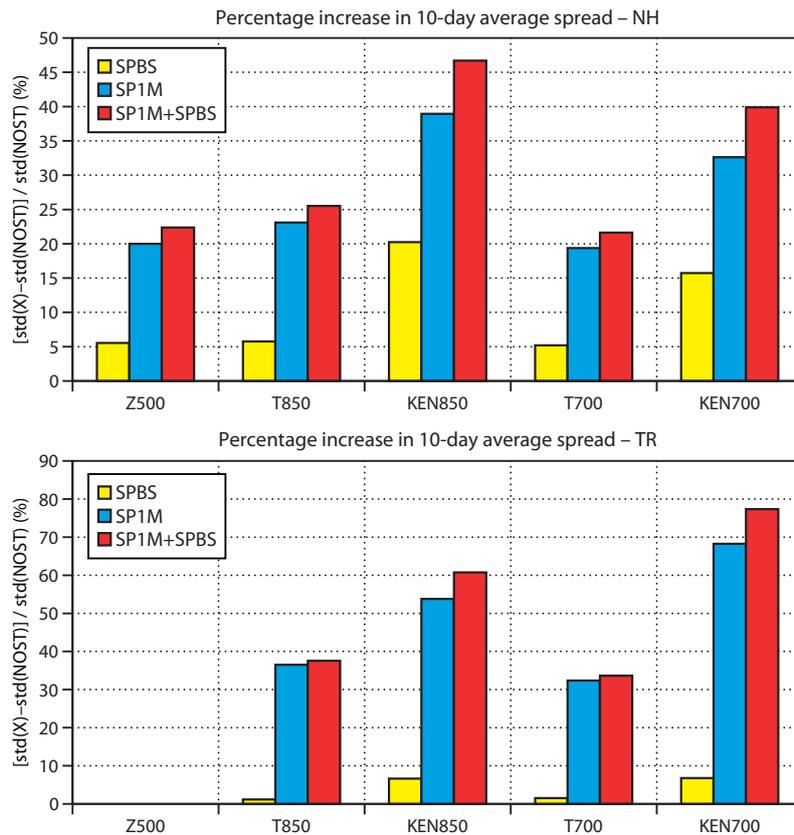


Figure 12: Top panel: increase in the 10-day average (from 2008-10-10 to 20) ensemble spread induced by the stochastic backscatter scheme $SPBS_{VC}$ (yellow bars), the revised SPPT scheme $SP1_{M45}$ (blue bars) and the two schemes used together (red bars), with ensemble spread measured in terms of 500 hPa geopotential height, and temperature and kinetic energy per unit mass at 850 and 700 hPa over the Northern Hemisphere. Bottom panel: as top panel but for the tropics.

the EDA generation and in the ensemble forecasts leads to higher skill. The ensembles of analyses EDA(NOST), EDA($SP1_{M45}$), EDA($SPBS_{VC}$) and EDA($SP1_{M45}+SPBS_{VC}$) have been used either alone, or with initial-time SVs, to generate the EPS initial perturbations. Results based on 18 cases have indicated that replacing the evolved SVs with EDA($SP1_{M45}$) has a positive impact on the ensemble scores, especially over the tropics, thus confirming the conclusions of *Buizza et al. (2008)*. Figure 13 shows the impact of using the EDA($SP1_{M45}$)-based initial perturbations and $SP1_{M45}$ (in ensemble forecasts) on the quality of probabilistic 850 hPa temperature forecasts over the Northern Hemisphere and the tropics. Both changes improve the EPS performance, especially over the tropics where they have a large, positive impact on the ensemble spread (not shown).

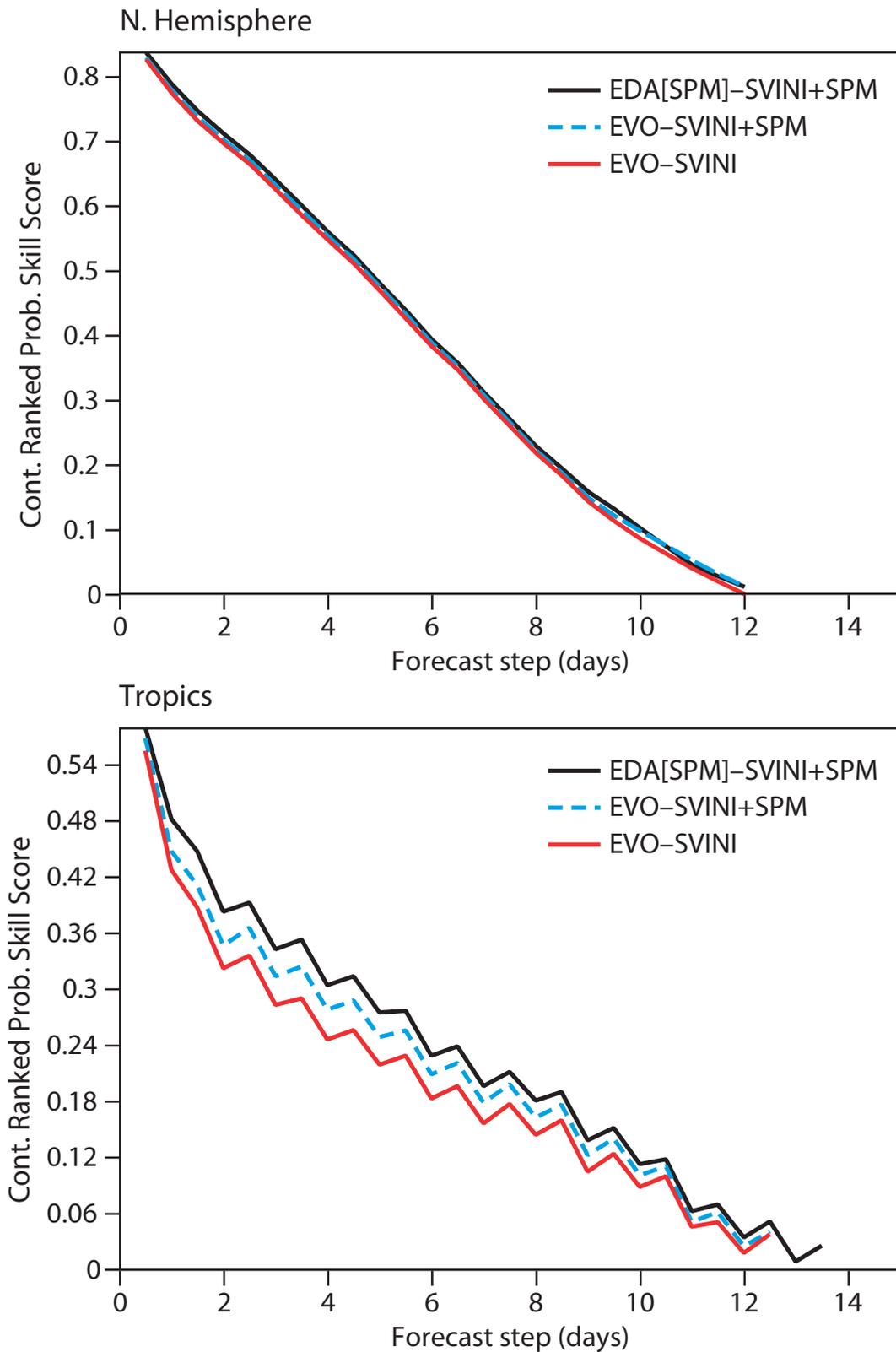


Figure 13: Top panel: 18-case average Continuous Rank Probability Skill Score (CRPSS) of the probabilistic prediction of 850 hPa temperature over NH for the operational ensemble with the original SPPT scheme (solid red), the operational ensemble with the revised SPPT scheme $SP1_{M45}$ (dashed blue) and of an ensemble with EDA-based initial perturbations generated with $SP1_{M45}$ and initial-time singular vectors (solid black). Bottom panel: as top panel but for the tropics.

4 Impact on seasonal integrations

As part of the routine assessment of changes in the physics of the IFS, long uncoupled (atmosphere-only) climate integrations are usually carried out at a relatively low horizontal resolution. Diagnosis of these long integrations provides a good idea as to whether the proposed changes have neutral, positive or detrimental impacts on certain aspects of the model's climate (*Rodwell and Jung, 2008*). Given that the implementation of stochastic parametrization schemes constitutes a substantial modification of the physics of the IFS, it has been decided to carry out a detailed investigation to understand the impact that the various stochastic parametrization schemes have on the model's climate. In fact, in previous studies it has been found that stochastic parametrization generally improve the climate of the model, especially in terms of the extratropical atmospheric circulation (*Jung et al., 2005; Palmer et al., 2005; Berner et al., 2009a*). Results from these latest experiments are discussed in Section 4.1.

Since the operational introduction of System 3 in 2007, seasonal forecast and hindcast ensembles with the coupled IFS-HOPE system have been using the original version of SPPT (SP_{BMP}) for representing model uncertainty. However, the ensembles of seasonal integrations are, in general, underdispersive, i.e., the ensemble standard deviation around the ensemble mean is substantially smaller than the RMSE of the ensemble mean. The lack of spread becomes apparent from the first month of the integrations and remains relatively constant over all lead times.

An earlier version of the stochastic backscatter scheme based on a cellular automaton pattern generator (CASBS) had been applied to a test set of seasonal hindcasts using the IFS atmospheric cycle CY29R2. The results showed a reduction in some of the systematic errors, for example for tropical rainfall and SSTs in the ENSO region, and significant improvements in probabilistic skill scores for a number of variables (*Berner et al., 2008*).

In the ENSEMBLES project, the CASBS simulations were compared with two alternative ways to address model uncertainty on the seasonal time range, namely with a multi-model ensemble and a perturbed physical parameter ensemble. It was found in *Doblas-Reyes et al. (2009)* that the multi-model ensemble performed best on lead times shorter than five months.

In Section 4.2 we discuss recent progress that has been made in coupled IFS-HOPE seasonal integrations with the atmospheric cycle CY35R2 using the revised SPPT and SPBS schemes. A total of 3 different experiments were performed: $SP2$ using the two-scale version of the revised SPPT scheme with the parameter settings according to Table 1; $SPBS$ with a backscatter ratio $b_R = 0.05$; and the combination of $SP2 + SPBS$. The control experiment $Ctrl$ was done using the original version of SPPT, SP_{BMP} .

4.1 Systematic error in uncoupled integrations

In order to study the impact of the different stochastic parametrization schemes on the climate of the ECMWF model, a large set of 13-months long integrations, started on 15 November of each of the years 1990–2005, was carried out for the different model formulations summarized in Tab. 1. All integrations are based on cycle 35R1 with a horizontal resolution of T_L159 and 91 levels in the vertical. Observed SST and sea ice fields were prescribed as lower boundary conditions. Here we analyse the impact of stochastic parametrization schemes on the first season of the integrations only (i.e. DJF).

The impact on the wintertime atmospheric circulation over the Northern Hemisphere can be inferred from Fig. 14. Albeit much reduced compared to IFS cycles used before 2008 (*Jung et al., 2009*), there are still notable systematic Z500 error for NoTenPert (cycle 35R1) over the Northern Hemisphere, especially in the northeastern North Atlantic and the North Pacific (Fig. 14a). Generally, mean Z500 differences between the

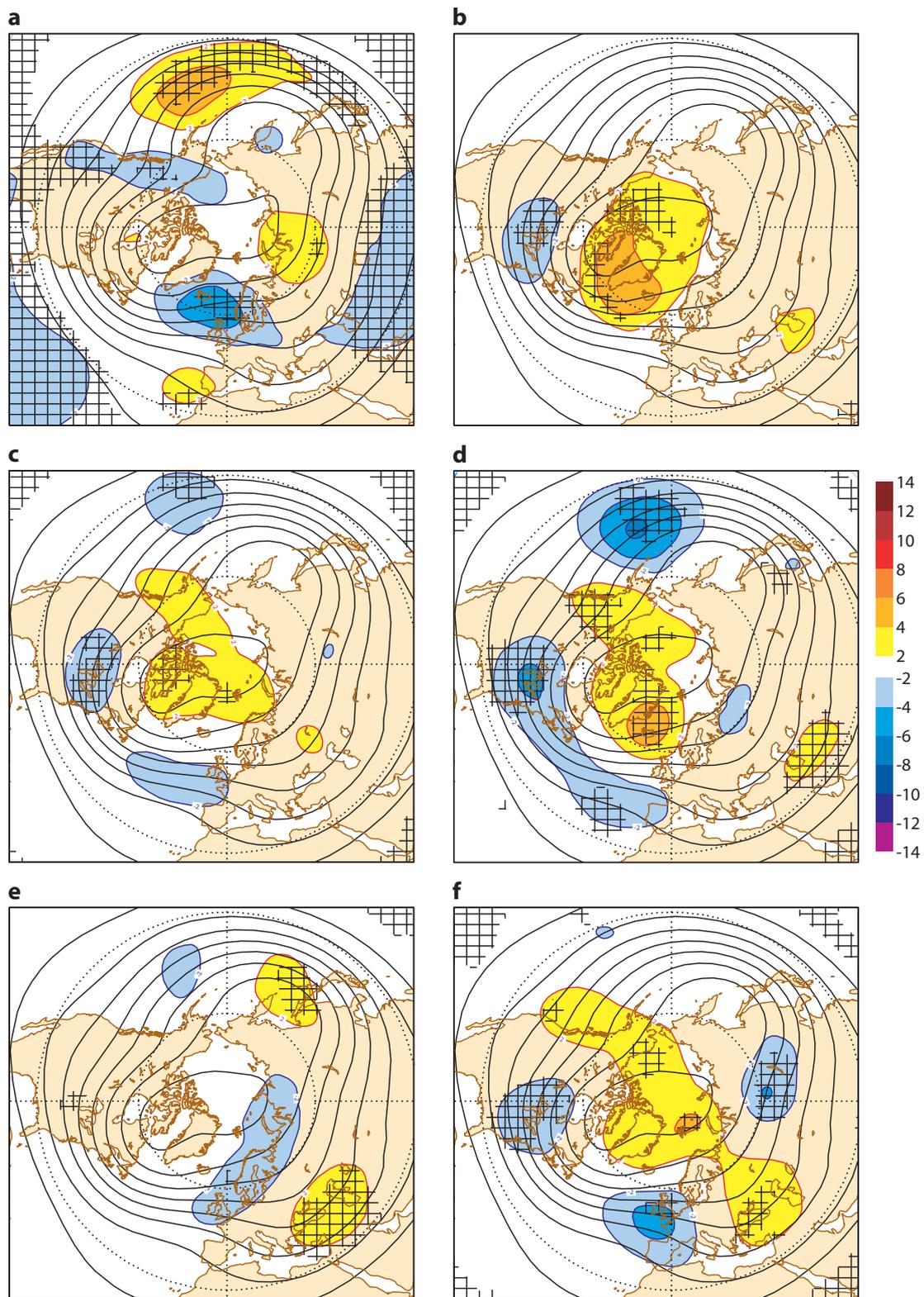


Figure 14: (a) Systematic Z500 error (dam) for NoTenPert. Also shown are mean Z500 differences from NoTenPert for (b) SP_{BMP}, (c) SP_{I_M}, (d) SP_{I_L}, (e) SP₂ and (f) SP_{I_M}+SP_{BS}. Results are shown for winters (December–February) of the period 1990–2005. Statistically significant differences (at the 95% confidence level) are hatched.

sensitivity experiments with stochastic parametrizations and NoTenPert show that the impact of the various stochastic parametrization schemes is only slightly smaller than the magnitude of systematic Z500 error for NoTenPert (Fig. 14b–f). The first thing to notice is that none of the schemes leads to a significant deterioration of the mean horizontal circulation. The strongest, and largely beneficial impact on the mean Z500 field is found for the revised SPPT scheme with large perturbations (SP1_L, Fig. 14d). The smallest impact is found for the revised two-scale SPPT scheme, SP2 (Fig. 14e). A comparison between SP1_M and SP1_M+SPBS suggests that the revised SPPT scheme has a larger impact onto the model climate than the stochastic backscatter component (similar results are found for other parameters, regions and seasons). During boreal summer the impact of the stochastic parametrization schemes on the mean horizontal circulation over the Northern Hemisphere is generally smaller (not shown).

The impact of the stochastic parametrization schemes on synoptic activity in the Northern Hemisphere winter can be summarized as follows: synoptic activity in the subtropical and polar regions (mid-latitude storm tracks) is increased (reduced) compared to NoTenPert. These changes are generally positive for they tend to oppose systematic errors apparent in the integrations with NoTenPert.

In the tropics, the largest impact of the stochastic parametrization schemes can be found during boreal summer. In particular the different versions of the revised SPPT scheme (SP1_M, SP1_L and SP2) lead to a drying of the tropical atmosphere. A comparison with total precipitable water (TPW) climatologies over the oceans from SSM/I data suggests that this drying is largely beneficial. At least some of the drying, however, might actually be artificial due to the treatment of supersaturation; in fact, preliminary tests with a more realistic supersaturation treatment suggest that the drying is significantly reduced, especially for very strong forcing (SP1_L).

The Indian Summer Monsoon, the African Summer Monsoon and precipitation over the Maritime Continent tend to weaken with the introduction of the revised SPPT schemes (the stronger the forcing the larger the weakening). At least for the Indian Summer Monsoon and precipitation over the Maritime Continent, this weakening is beneficial. The sole influence of the SPBS scheme is generally smaller than that from the revised SPPT schemes.

In terms of tropical variability (OLR and rainfall) it is found that the use of the revised SPPT schemes leads to increased (reduced) levels of synoptic (low-frequency intraseasonal) variability, especially during boreal winter; the impact of the SPBS scheme is relatively small. None of the schemes cures the problems seen in the NoTenPert integration (and all previous model cycles) in simulating a realistic Madden-Julian Oscillation (MJO). There is some evidence, however, that the revised SPPT schemes increase quasi-periodic variability with a period of about 30 days (the observed ‘peak’ is found in the 40–60 day range), especially during boreal summer.

In summary, it can be concluded that the impact of the various stochastic parametrization schemes on the climate of the (uncoupled) atmospheric model in the extra-tropics is comparable to the errors present in the experiment without tendency perturbations. In the tropics, however, the impact is significantly smaller than the systematic errors present in the control integration.

4.2 Coupled integrations

The experimental set-up of the seasonal hindcasts consists of 7-month long integrations with the coupled IFS-CY35R2-HOPE system in T_L159L62 resolution. The hindcasts were started twice a year on the 1st of May and November and thus cover, with different lead times, all 4 seasons. The hindcasts were run over the 18-year period 1991-2008, although some of the results shown will be for the slightly shorter period 1991-2005 because the ENSEMBLES multi-model ensemble is not available after 2005. The seasonal hindcast ensemble are based

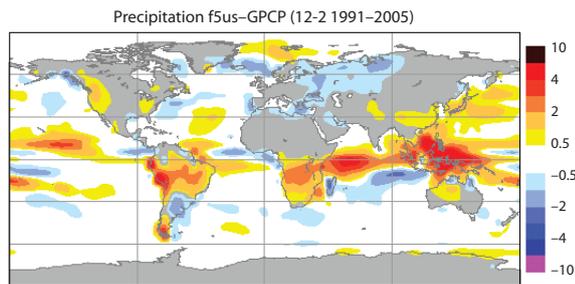


Figure 15: Systematic error of precipitation in DJF (*Ctrl* minus GPCP).

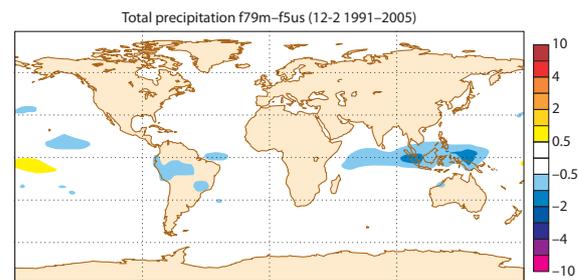


Figure 16: Precipitation difference *SP2 + SPBS* minus *Ctrl* in DJF

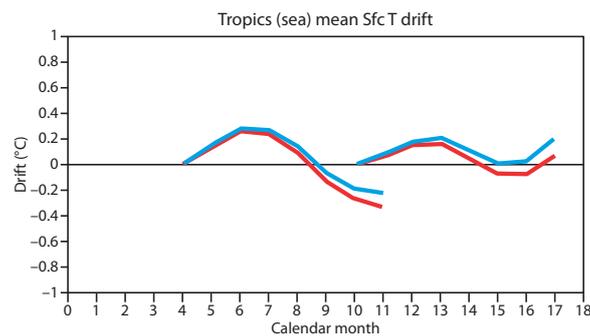


Figure 17: Tropical SST drift in coupled seasonal hindcasts estimated for May and Nov start dates over the period 1991-2005. Red: *Ctrl* and blue: *SP2 + SPBS*

on 9 ensemble members that were created by initialising the hindcasts from 5 different ocean analyses with added SST perturbations and atmospheric singular vectors.

4.2.1 Systematic errors

For the two start dates of seasonal hindcasts in May and November, an assessment of the systematic errors in the atmosphere for the DJF and JJA seasons with lead times 2-4 months has been performed. The error was estimated using all available hindcasts and ensemble members.

In general, relatively little impact of the tested stochastic parametrization schemes on systematic errors was found. The most impressive result is a significant reduction of excessive rainfall in the tropical belt. Fig. 15 shows the systematic error of precipitation in DJF of the *Ctrl* simulation compared to GPCP indicating too much rainfall over parts of South America, South Africa, the tropical Indian ocean and the warm pool area. In the *SP2 + SPBS* seasonal hindcasts, some of these errors were significantly reduced, especially over the Indian Ocean and the Maritime continent (Fig. 16). A similar improvement was found in JJA over the ITCZ in the Eastern Pacific and the warm pool areas (not shown). In agreement with the uncoupled runs, the excessive Indian monsoon precipitation has also been reduced.

The stochastic parametrization schemes have little impact on the model SST drift. As an example Fig. 17 displays the evolution of the SST drift over the tropics for the May and November start dates. Compared with the *Ctrl* run, the stochastic parametrization scheme *SP2 + SPBS* tends to warm the tropical ocean by approx. 0.2 degrees after 7 months. For the May start dates this corresponds to an improvement, whereas for the November starts, the warming leads to a small positive bias.

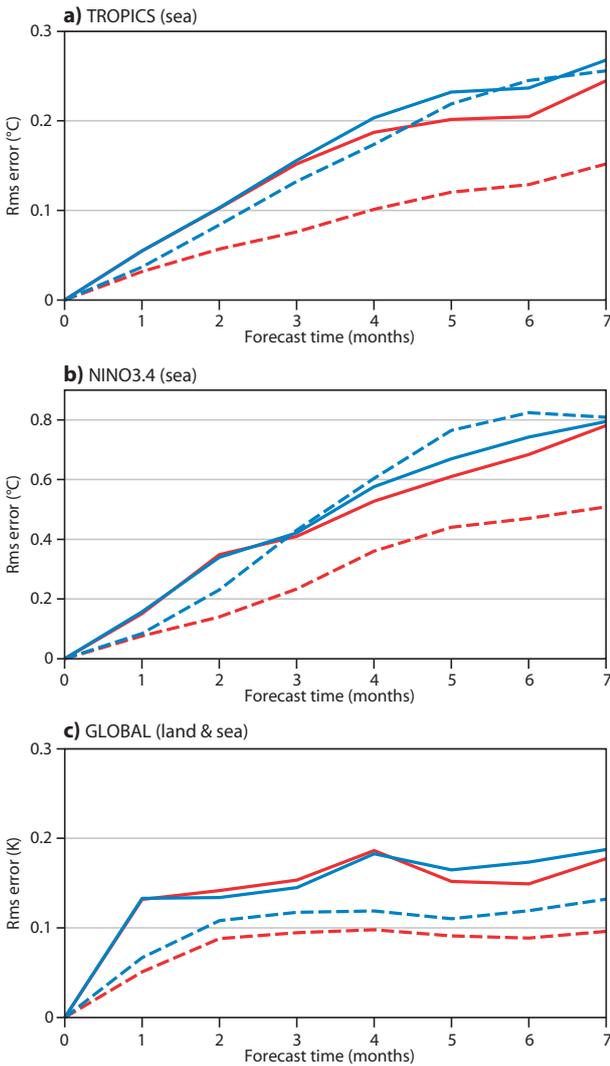


Figure 18: Tropical SST (a), Nino3.4 SST (b) and global mean T2m (c): ensemble spread and RMSE as a function of lead time in coupled seasonal hindcasts for May and Nov start dates 1991-2005. Red: Ctrl, blue: SP2 + SPBS. Solid lines indicate RMSEs and dashed coloured lines the spread.

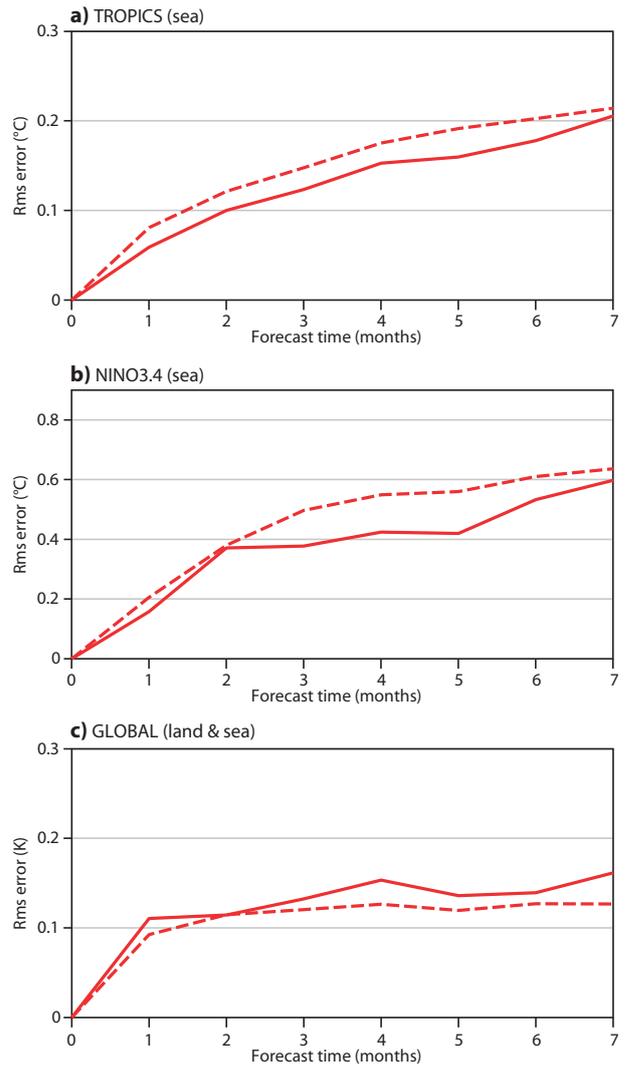


Figure 19: As in Fig. 18 but for the ENSEMBLES multi-model

4.2.2 Impact on spread and skill

Results of the ensemble-mean forecast RMSE and ensemble spread (standard deviation around the ensemble mean) are shown in Fig. 18. For tropical SSTs in Fig. 18a, the *Ctrl* version of the coupled model is underdispersive from the first month of the integration with a clear underestimation of ensemble spread throughout the forecast range. The new stochastic parametrization schemes have a major impact on the ensemble spread with a strong increase of the spread after the first month. Even though the forecast RMSE of *SP2 + SPBS* is somewhat larger than in the control after month 3, the reliability of the forecasts in terms of spread-skill relationship is significantly improved. This improvement is mainly due to the impact of the *SP2* scheme on increasing the spread. The combination of *SP2* and *SPBS* leads to a further increase in spread, but also to a larger RMSE in the second half of the simulation range if compared with the two individual schemes (not shown).

In the tropical Pacific (Fig. 18b), the increase of ensemble spread due to the new stochastic schemes is even more pronounced. Here the forecasts become slightly overdispersive on longer lead times. The increased spread has a positive impact on probabilistic forecasts scores that show a higher reliability component of the Brier skill scores than the *Ctrl*, especially on lead times beyond 3 months (not shown).

The impact of a much increased ensemble spread in *SP2 + SPBS* can also be seen in global mean 2m temperature forecasts, see Fig. 18c. The spread increases by roughly one quarter over the *Ctrl*, but is still too small if compared with the forecast error. The RMSE in *SP2 + SPBS* remains unchanged vs *Ctrl* up to forecast month 4 and increases slightly beyond that range. The underestimation of ensemble spread is mainly associated with the oceanic areas (not shown), whereas over land the match between spread and error is, in general, good.

4.2.3 Comparison with the ENSEMBLES multi-model ensemble

We here compare the latest stochastic parametrization results for seasonal forecasting with the multi-model ensemble predictions of the ENSEMBLES project. In ENSEMBLES, 5 coupled climate models from different institutions across Europe have completed a set of 7-month long seasonal hindcasts over the period 1960–2005 with 4 start dates per year (Weisheimer *et al.*, 2009).

Fig. 19a shows, similar to Fig. 18a, the evolution of the tropical SST RMSE and ensemble spread for the ENSEMBLES multi-model using the same start dates and hindcast period as for the stochastic parametrization experiment. While the multi-model ensemble is slightly overdispersive from the first month onwards, the forecast error is smaller than for both *Ctrl* and *SP2 + SPBS* in Fig. 18a.

Similar conclusions hold for the tropical Pacific SSTs (Fig. 19b). While in the IFS/HOPE coupled system the RMSE continues to grow over lead time, the multi-model ensemble forecast error saturates between months 2 and 5 at a level that is below the IFS/HOPE level. Again, the multi-model ensemble is slightly over dispersive. In terms of probabilistic skill scores (not shown), the multi-model ensemble performs, on average, better than the *Ctrl* and the *SP2 + SPBS* versions of IFS/HOPE. However, for certain events and seasons (upper tercile SST predictions starting in May) the improved reliability components in *SP2 + SPBS* perform equally well as for the multi-model ensemble.

Looking at the global scales, the comparison of global mean 2m temperature forecasts between the ENSEMBLES multi-model ensemble in Fig. 19c and the corresponding figure for IFS/HOPE with stochastic parametrization in Fig. 18c reveals that the multi-model ensemble performs better by generating smaller RMSEs and a very good match between the ensemble spread and error on all lead times.

5 Coarse-graining studies

5.1 Cloud-resolving Model

5.1.1 Methodology

Ideally, observational datasets from intensive field experiments would provide sufficient information about the statistical nature of physical processes in the atmosphere for improving their representation in NWP and climate modelling. For instance, probability distribution functions (pdfs) for convective cloud population density (conditioned on CAPE, vertical wind shear, underlying land/sea, topography etc) could be determined and used to devise stochastic convection parametrization schemes. Such datasets that do exist are generally insufficient for this purpose and it is necessary (and more convenient) to use numerical model simulation data. For instance [Shutts and Palmer \(2007\)](#) used a cloud-resolving model (see Section 5.1.2), configured to simulate deep convection in a square domain of side length 7680 km, to compute pdfs of coarse-grained convective warming rate. These pdfs were examined as a function of the strength of *parametrized* convective warming based on the coarse-grained field values and used to validate the statistical assumption upon which the SPPT scheme is based e.g. that the standard deviation is proportional to mean parametrized tendency.

The coarse-graining procedure used for the cloud-resolving model involves averaging model fields and their associated tendencies to a grid resolution typical of current NWP or climate models. By the Reynolds-averaging approach outlined below it is possible to compute the non-advective part of the total tendency that the corresponding coarse-grid model would see. Consider the thermodynamic equation written in its potential temperature form:

$$\frac{D\theta}{Dt} = \frac{\partial\theta}{\partial t} + \mathbf{V} \cdot \nabla\theta = Q \quad (6)$$

where Q represents the diabatic source term and re-write as:

$$\frac{\partial\theta}{\partial t} = -\mathbf{V} \cdot \nabla\theta + Q. \quad (7)$$

Average eq.(7) to a coarse grid and let an overbar denote this operation giving:

$$\frac{\partial\bar{\theta}}{\partial t} = -\overline{\mathbf{V} \cdot \nabla\theta} + \bar{Q} \quad (8)$$

and then add the term $\overline{\mathbf{V} \cdot \nabla\bar{\theta}}$ to each side so that:

$$\frac{\partial\bar{\theta}}{\partial t} + \overline{\mathbf{V} \cdot \nabla\bar{\theta}} = \left[\overline{\mathbf{V} \cdot \nabla\theta} - \overline{\mathbf{V} \cdot \nabla\bar{\theta}} \right] + \bar{Q} = \tilde{Q} \quad (9)$$

where the term in square brackets represents an eddy heat flux divergence and \tilde{Q} is the apparent source of θ on the coarse grid. [Shutts and Palmer \(2007\)](#) sampled the coarse grid-columns by feeding temperature, humidity and wind profiles into a convective parametrization scheme ([Bechtold et al., 2001](#)) and using the predicted convective temperature tendency to group the columns into distinct ranges e.g. -0.1 to 0.1, 0.1 to 9, 9 to 18, 18 to 27, 27 to 36 and 36 to 45 K/day. The mean of \tilde{Q} is then plotted against standard deviation for each sample and repeated for different coarse-grain box sizes. Figure 20 shows these points together with the mean versus standard deviation lines implied by $SP1_M$, $SP1_L$ and SP_{BMP} . For any particular value of mean(\tilde{Q}) the standard deviation of \tilde{Q} increases with decreasing coarse-graining box size and for a chosen box-size the

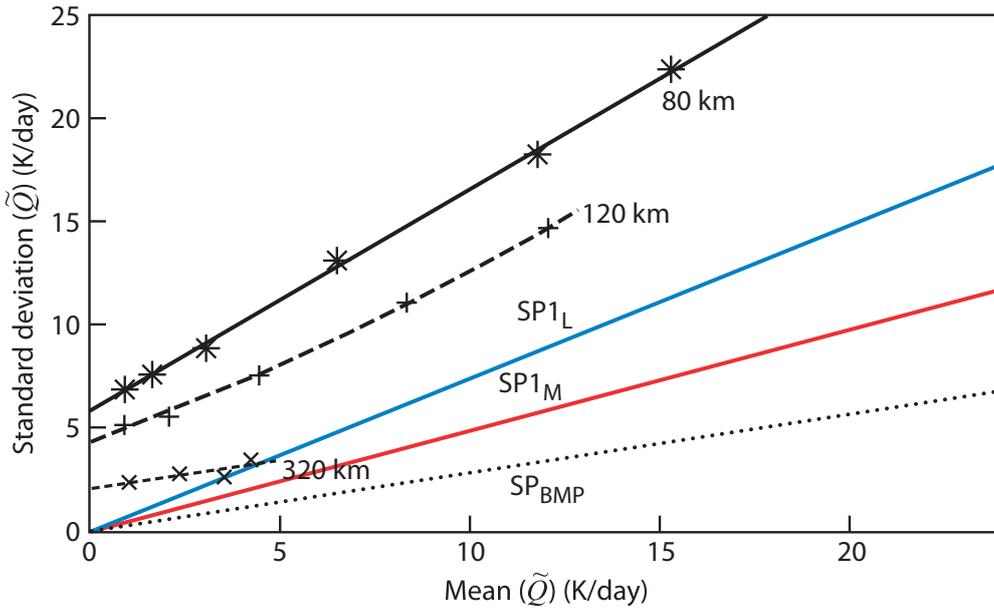


Figure 20: Standard deviation of \tilde{Q} versus mean(\tilde{Q}) inferred from an idealized cloud-resolving model run. Lines linking crosses correspond to different coarse-grain box sizes used to coarse-grain the model. Lines labelled $SP1_M$, $SP1_L$ and SP_{BMP} indicate the implied relationships for the corresponding variants of SPPT. This figure is adapted from Fig. 12 of Shutts and Palmer (2007).

standard deviation of \tilde{Q} is a linear function of the mean but non-zero when the mean is zero. Apart from this offset, the results support the ansatz underpinning SPPT. Non-zero variance in \tilde{Q} when the mean is close to zero reflects the counterbalancing of latent heat release in updraughts with the cooling due to the evaporation of condensate in or beneath mesoscale convective anvils.

The decorrelation scale of $SP1_M$ and $SP1_L$ is currently set to 500 km which implies a somewhat larger coarse-graining scale. Since the standard deviation of \tilde{Q} varies roughly as the inverse of the coarse-grain box size, one might anticipate that the standard deviation appropriate to SPPT is about one half of the values shown in Figure 20 for a 320 km box size. This suggests that the current choices of standard deviation for $SP1_M$ and $SP1_L$ are plausible. It is planned to repeat this coarse-graining analysis with data from the new simulation described below, but this time the fields will be coarse-grained with a spatial filter consistent with the Gaussian auto-correlation function of the revised SPPT scheme.

For the purposes of calibrating SPBS a similar technique has been applied to compute an effective vorticity or streamfunction forcing. The above coarse-graining procedure is applied to wind components u and v and the resulting effective momentum forcing can be expressed as a vorticity forcing function by taking the vector Curl. The vorticity forcing function is converted to a streamfunction forcing by expanding the vorticity forcing in a double-Fourier series and dividing each Fourier coefficient by $-K^2$ where K is the modulus of the associated wavevector.

Specifically the coarse-grained u momentum equation becomes:

$$\frac{\partial \bar{u}}{\partial t} + \bar{\mathbf{V}} \cdot \nabla \bar{u} - \beta y \bar{v} + \frac{\partial}{\partial x} \left(\frac{p'}{\rho_0} \right) = \left[\bar{\mathbf{V}} \cdot \nabla \bar{u} - \bar{\mathbf{V}} \cdot \nabla u \right] + \bar{F}_u = \tilde{F}_u \quad (10)$$

where β is the meridional gradient of the Coriolis parameter at the equator; p' is the perturbation pressure and $\rho_0(z)$ is a height-dependent reference density in this quasi-Boussinesq approximation of the full momentum

equation. F_u represents the combined effects of a Smagorinsky-Lilly turbulent diffusion scheme and a supplementary horizontal Laplacian diffusion. The first two terms on the right-hand side represent the divergence of a Reynolds stress appropriate to the coarse-graining scale. In this nomenclature the streamfunction forcing can be defined as:

$$F_\psi = \nabla^{-2} \left(\frac{\partial \tilde{F}_v}{\partial x} - \frac{\partial \tilde{F}_u}{\partial y} \right) \quad (11)$$

5.1.2 Model configuration

The simulations to be described have been carried out with version 2.3 of the Met Office LEM (Large Eddy Model). The model uses an Arakawa C-grid staggering in horizontal planes and Lorenz grid in the vertical with periodicity assumed in both horizontal directions. In all experiments, 50 vertical levels are deployed non-uniformly with height with resolution ranging from 150 m in the boundary layer to 500 m in the free troposphere to 800 m in the stratosphere. The Coriolis parameter is made a linear function of y with the equator at $y = 0$ thereby achieving an equatorial beta-plane. The domain is 10,000 km in the x direction and 5000 km in the y direction with horizontal gridlength $\Delta x = \Delta y = 2.44$ km. The lower surface is treated as sea with a temperature $SST(y)$ that varies parabolically in y according to:

$$SST(y) = 301.0 - a(y/y_0)^2 \quad (12)$$

where y_0 is the meridional domain half-width (equal to 2500 km) and $a = 3.8125$. Convection is forced by imposing a horizontally-uniform cooling function of -1.5 K/day up to a height of about 11 km and then tailing off to zero by 15 km. This cooling function is used in place of the model's radiation scheme in order to reduce the considerable computational cost. The initial state is horizontally-stratified with a uniform geostrophic easterly wind of -5 ms^{-1} . The imposed meridional pressure gradient that ensures initial balance is held fixed during the integration and so acts as a kind of Trade wind forcing function. Convection is initiated with some small lower tropospheric temperature perturbations. The model configuration used here, together with some others that use anisotropic horizontal grids, is more fully described in [Shutts \(2006\)](#).

5.1.3 Results

Convection develops in an equatorial band but after a few days splits into the familiar 'double-ITCZ' pattern with precipitation maxima at about 15 degrees from the equator (ITCZ is the Inter-Tropical Convergence Zone). Figure 21 shows a Hovmöller plot of rain-rate averaged between 10 degrees north and south. Rainfall rates increase up to day 3 but fall afterwards as its double-ITCZ structure develops and moves outside of the Hovmöller averaging zone. Westward propagating rain cells dominate up to day 3 but thereafter, eastward propagating cells are evident with a speed equal to 15 ms^{-1} . As in the study of Shutts (2006), these are most likely to be driven by Kelvin waves.

The initial easterly flow evolves into a pair of upper tropospheric westerly jets near the meridional limits of the domain and easterly jets near 8 degrees north and south (Figure 22). These easterly jets appear to result from the Coriolis torque acting on equatorward flowing air driven by the outflow from the two ITCZs.

The streamfunction forcing, computed on a 40 km grid at a height of 10.5 km, is shown in Figure 23. The strongest variance in streamfunction forcing occurs in the jetstream regions and typical gradients of about 20 m^2s^{-2} per 1000 km imply flow accelerations of the order of a few ms^{-1} per day. Details of the spectral power distribution of the streamfunction forcing will be given in the next section together with a similar estimate obtained from the IFS. These in turn will be related to the streamfunction forcing used in SPBS.

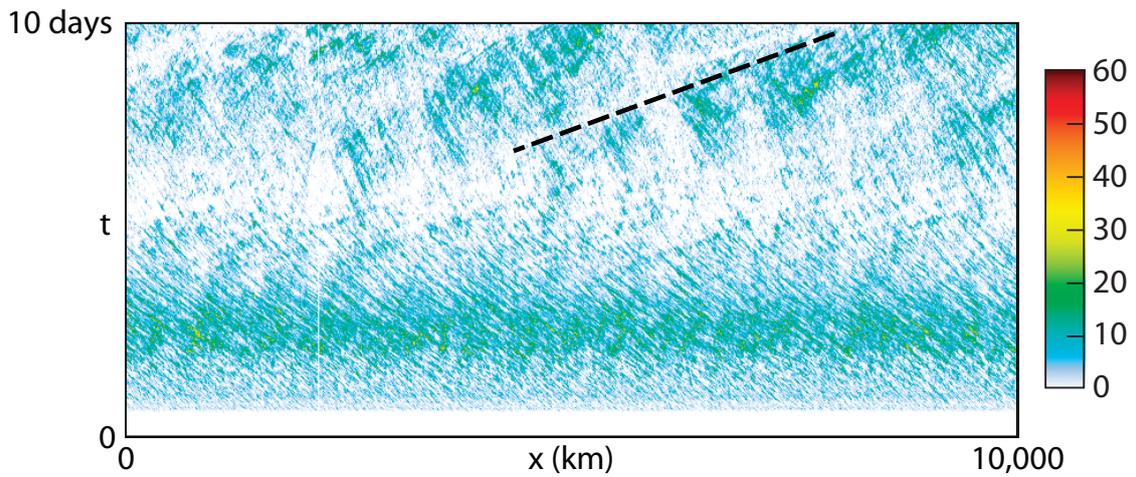


Figure 21: Hovmöller plot of the surface rain-rate averaged within a 20 degree band centred on the equator (Units:mm/hr). The diagonal line has a slope matching the rate of movement of convective systems moving eastwards and corresponds to a speed of 15 ms^{-1}

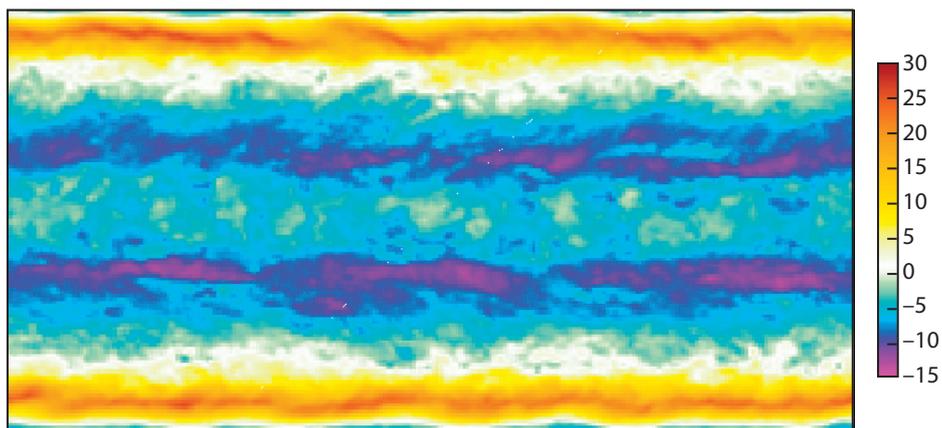


Figure 22: u at a height of 10.5 km on day 7. Units: ms^{-1}

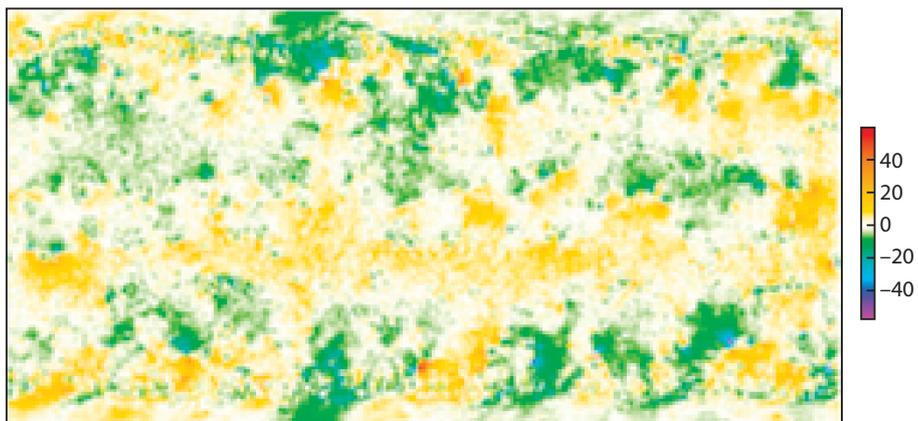


Figure 23: The streamfunction forcing F_ψ computed from the effective momentum forcing function at a height of 10.5 km on day 7 found when the momentum equation is coarse-grained to a 40 km grid. Units: $\text{m}^2 \text{ s}^{-2}$

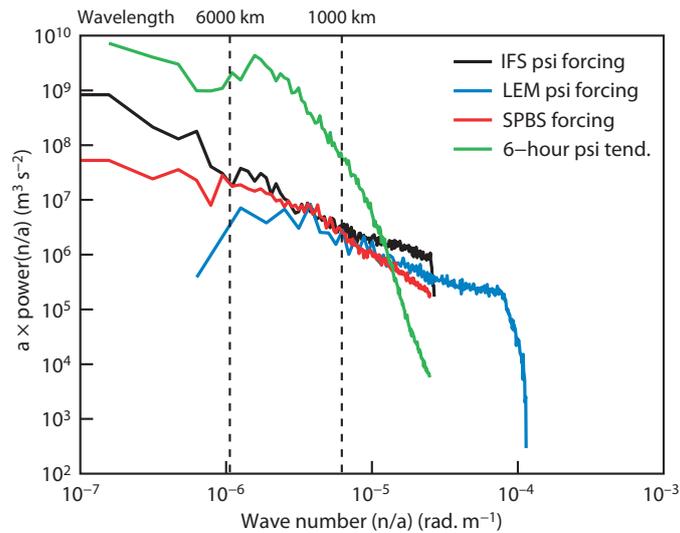


Figure 24: Power spectrum of the streamfunction forcing at level 53 (about 256 hPa) deduced from IFS (black line) and cloud-resolving model (blue line) coarse-graining analysis. The red line shows the power spectrum for the streamfunction forcing used in SPBS (with backscatter ratio of 0.1) from a T_L159 forecast. For reference the green line shows the power spectrum of the six-hour mean streamfunction tendency in the IFS for level 53.

5.2 Coarse-graining NWP model tendencies

5.2.1 Methodology

The approach taken here is to compute differences between forecasts made with the IFS at two radically different horizontal resolutions. The high resolution run (here taken to be T_L1279) is regarded as ‘truth’ and its total (dynamical + physical) tendencies are spectrally-smoothed to the resolution of the low resolution model (here T_L159). The difference between the coarse-grained tendency and the low-resolution model tendency can be considered to be a tendency error (see [Hermanson et al., 2009](#) who carried out a related analysis using the IFS). As with the cloud-resolving model, the aim will be to compute the streamfunction forcing that is implied by the model coarse-grained tendency errors in u and v .

At a detailed level the procedure is as follows. Global gridpoint tendencies fields of u and v are transformed in a single spectral transform routine call to spectral (Spherical Harmonic (SH)) coefficients of vorticity and divergence tendency. The vorticity tendency coefficients are divided by $-n(n+1)/a^2$ (where n is the spherical harmonic degree and a is the Earth’s mean radius) to give streamfunction forcing SH coefficients. Subtracting the streamfunction forcing coefficients for T_L1279 and T_L159 (and neglecting those with $n > 159$) defines the streamfunction forcing ‘error’ of the T_L159 forecast. Since the timestep of the T_L1279 run is 8 times smaller than the 1 hour timestep of the T_L159 run, the T_L1279 streamfunction forcing is averaged over the single step of the T_L159 run.

5.2.2 Results

Streamfunction forcing power spectra have been computed from T+2 hour fields of forecasts starting at 12Z August 17 2006. Figure 24 shows these spectra together with:

- the power spectrum computed on day 7 of the cloud-resolving model simulation

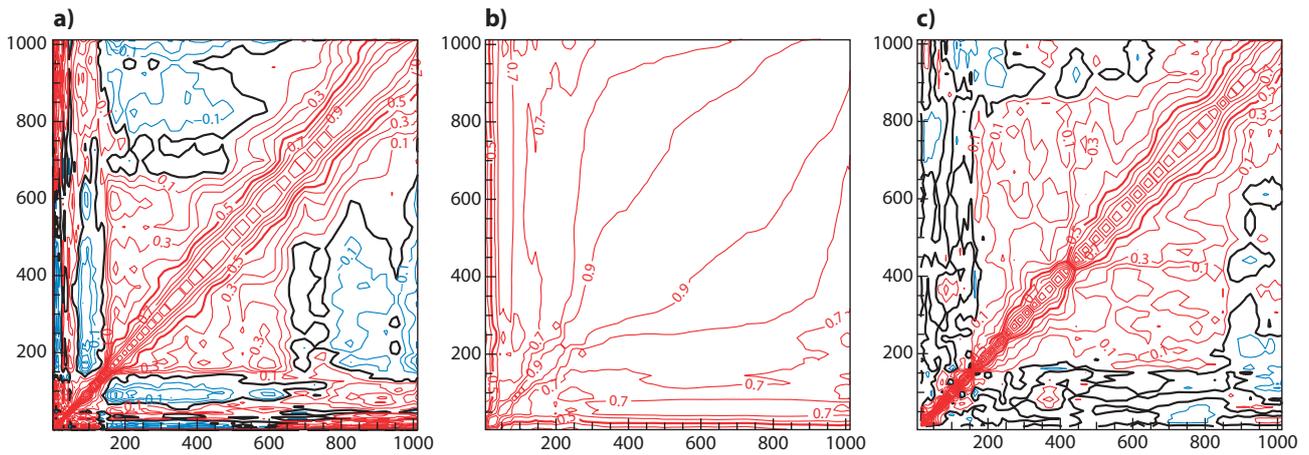


Figure 25: Vertical correlations of the stream function forcing calculated in gridpoint space. The correlation for the IFS coarse-graining are shown in a), b) shows the correlation for SPBS using a vertically constant pattern and c) shows the results for SPBS using the random phase shift scheme (for details see Appendix 8.2).

- the power spectrum computed from SPBS at the same model level as the IFS coarse-graining calculation (backscatter ratio=0.1).

The cloud-resolving model curve has somewhat less power than that implied by the IFS coarse-graining but similar spectral slope. Considering the differences in the LEM and IFS simulations (global forecast versus idealized tropical simulation), it is hardly surprising to find some spectral power differences in the computed streamfunction forcing. The streamfunction forcing in SPBS (for a T_L159 forecast using backscatter ratio of 0.1) has a spectral power distribution lying between the IFS and LEM coarse-graining estimates in the wavelength range 1000 to 6000 km. The current formulation appears to underestimate the power required at low wavenumbers although one should note that SPPT also makes a contribution to the streamfunction forcing (through convective momentum transport and gravity wave drag) and this, together with the SPBS contribution, would make a more appropriate comparison with the IFS coarse-grain streamfunction forcing.

The green curve in figure 24 gives some indication of the strength of the streamfunction forcing relative to the power in the total model streamfunction tendency. It confirms that the power in backscatter forcing is orders of magnitude smaller than the power in the streamfunction tendency except for scales less than about 500 km. T_L159 ensemble forecasts made with this backscatter ratio have been shown to have a positive impact on skill scores.

In addition to tuning the streamfunction forcing spectrum, the coarse-grained data was also used for tuning the vertical correlation scale in SPBS. In the original implementation of SPBS the same spectral pattern was used at all model levels, so the resulting vertical correlation of the stream function forcing was determined by the dissipation rate alone. Those correlations (Fig. 25b) turned out to be problematic in ensemble data assimilation where high long-range correlations have a damaging effect. The coarse-graining results (Fig. 25a) do not show these long-range correlations. The random vertical phase scheme (RVP) used in SPBS (for details see Appendix 8.2) applies a phase shift to the random numbers used in the spectral pattern generator between any two layers. The coarse-graining results suggest that the variance of these phase-shifts, hence the correlation scale, depends on wavenumber and pressure. These functional dependencies are used in RVP to improve the vertical structure. Figure 25c shows correlations of the SPBS streamfunction forcing using RVP. They are overall in good agreement with the coarse-graining results, showing comparable correlation scales at all pressure levels.

5.2.3 Conclusions

The coarse-graining technique provides a way of designing and calibrating stochastic parametrization schemes. These preliminary results suggest that the current formulation of stochastic backscatter is reasonably well supported by the coarse-graining analyses. Future work will be directed at estimating the spectral power of F_ψ in the wavenumber-frequency domain so that the decorrelation time can be made a function of wavenumber. It will also be necessary to be more careful in distinguishing between systematic and random contributions to the streamfunction forcing. The methodology can also be applied to temperature and humidity tendencies and it is planned to calibrate the revised SPPT scheme in this way.

6 What are other centres doing?

In the operational UK Met Office Global and Regional Prediction System (MOGREPS, [Bowler et al., 2008](#)) three schemes have been used to simulate the effect of model errors. The first one, a random parameter scheme, has been designed to simulate the fact that many key parameters in model physical parametrization schemes are uncertain. The second one, the stochastic convective vorticity scheme, provides a kinetic energy source that compensates for unresolved mesoscale convective systems (MCSs) in areas of high convective available potential energy. It does this by injecting anticyclonic vorticity in the upper troposphere and cyclonic vorticity in the mid-lower troposphere - consistent with the observed structure of MCSs ([Gray, 2001](#)). The third scheme, Stochastic Kinetic Energy Backscatter (SKEB), is similar to SPBS except that only numerical dissipation is accounted for and this is assumed to be proportional to the local kinetic energy ([Bowler et al., 2009](#)).

In the ensemble prediction system developed at the Meteorological Service of Canada (MSC), a multi-model approach was followed initially to simulate the effect of model uncertainty (MSC, see [Houtekamer et al., 1996](#)). In the 1996 MSC ensemble, two different dynamical cores with different orography were used, and each ensemble member was integrated by calling different parametrization schemes of horizontal diffusion, convection, radiation and gravity wave drag. In 2007, a perturbed tendency perturbation scheme and a stochastic backscatter scheme were also introduced ([Gagnon et al., 2007](#); [Houtekamer et al., 2007](#)). Since then, the MSC-EPS has been using four different methods to simulate model error: a parametrized system error scheme, the use of different parametrization schemes in each ensemble member, a stochastic tendency perturbation scheme and a stochastic backscatter scheme. The parametrized system error scheme ([Houtekamer and Mitchell, 2005](#)) simulates the effect of model error sources that degrade the quality of the initial conditions and that are not taken into account by the MSC procedure used to generate the ensemble initial conditions, by adding a random perturbation field to the initial conditions. The stochastic tendency perturbation scheme was developed following the ideas of [Buizza et al. \(1999\)](#) with patterns defined as in [Li et al. \(2008\)](#), with time-space correlations simulated using first-order Markov processes. The stochastic backscatter scheme was developed following [Shutts \(2005\)](#), with numerical dissipation estimated from the activity of the gravity wave drag parametrization and of the explicit diffusion. [Charron et al. \(2009\)](#), who discussed the most recent changes of the Canadian ensemble system, concluded that the use of the MSC stochastic tendency perturbation and stochastic backscatter schemes ‘improve the ensemble forecast reliability, by acting on the bias and the ensemble dispersion’.

At NCEP, there is ongoing research to develop a practical and effective stochastic parametrization scheme with total model tendencies perturbed with stochastic forcing sampled from the differences in the conventional tendencies between the ensemble perturbed members and the control ([Hou et al., 2006, 2008](#)). Compared to the scheme used at ECMWF ([Buizza et al., 1999](#)), the NCEP stochastic forcing is proportional to the whole model tendency instead of only the tendency due to physical parametrizations. [Hou et al. \(2006\)](#) discussed results based on an earlier, simplified version of this scheme applied every 6 hours to the NCEP ensemble system, and showed that it led to a substantially larger and better tuned spread, lower ensemble-mean error and higher

probabilistic scores. More recent results ([Hou et al., 2008](#)) indicated that a refined version of the scheme can significantly improve the performance of the NCEP ensemble system.

In the US Navy Operational Global Atmospheric Prediction System (NOGAPS) ensemble system, a ‘stochastic convection’ approach ([Teixeira and Reynolds, 2008](#)) is used to simulate model error uncertainties mainly due to convection parametrization. Within this approach, parametrization schemes are used with control parameters sampled from a distribution of probable values instead of with a unique, most likely value. They showed that this approach led to a substantial increase of the NOGAPS ensemble spread in the tropics, and a small increase in the extra-tropics, with positive impacts on some measurements of ensemble forecast skill. [Reynolds et al. \(2008\)](#) showed that the use of this scheme helps to mitigate a fundamental problem, namely, that ensemble systems with initial conditions defined using Ensemble Transformed methods have too little initial variance in the tropics and too much variance in the extra-tropics.

7 Conclusions and Discussion

Two complementary stochastic parametrization schemes have been developed and their impacts have been studied on medium-range and seasonal forecast timescales.

Research will continue to develop and extend these stochastic parametrizations. For example, results for the seasonal predictions indicate that, despite SPPT and SPBS, the EPS remains underdispersive for 2m temperature. This may indicate that a stochastic element should be included in the land surface component of the forecast model. However, more likely, such underdispersion suggests that variables associated with cloud amount should be treated stochastically. Work to extend the stochastic parametrizations in this direction will begin shortly.

Currently, the stochastic parametrizations discussed in this paper are applied only to the atmospheric model. However, there is no reason in principle why they should be so restricted, and further developments should see stochastic parametrization schemes being applied to the land and oceanic components of the ECMWF forecast system.

Both stochastic schemes discussed in Section 2 use spectral stochastic pattern generators. There are other schemes which allow one to focus on spatially confined features of the flow. For example, [Palmer \(1997, 2001\)](#) suggested the use of stochastic cellular automata as providing pattern generators to represent individual mesoscale convective complexes. Early results using a cellular automaton pattern generator in seasonal forecast mode have been promising ([Bernier et al., 2008](#)). There are some advantages to such cellular automata schemes. For example, it is straightforward to represent the advection of a convective complex by some steering level flow. In this way, information about convective instability can propagate from one grid box to a neighbouring one, in a way which a conventional convective parametrization cannot do. It is possible that the representation of advection of convective systems by the trade winds may be an important process missing in conventional schemes, for the correct development of the Madden-Julian Oscillation and research is underway to test this.

Ultimately, stochastic representations should be incorporated ab initio into the development of parametrization schemes; the schemes described in Section 2 are essentially “bolt on” extras, to be added to standard deterministic parametrization schemes. An example of an approach where a parametrization is considered stochastic ab initio is the scheme developed by [Plant and Craig \(2008\)](#), where one assumes a statistical ensemble of convective plumes with inherently small sample size, leading to significant dispersion of some underlying probability distribution.

As discussed in the text, very high resolution cloud-resolving models provide a rigorous approach to the development of stochastic parametrization and will become an increasingly important tool for the future.

This paper has focused on the application of stochastic parametrization on prediction timescales from the medium range to the seasonal. However, on the basis that insights and constraints on these timescales can inform prediction on longer timescales where verification data does not yet exist - the seamless prediction paradigm - it would seem reasonable to conjecture that stochastic parametrization should play a key role in the development of Earth System Models for multi-decadal and centennial climate prediction. Indeed since an Earth System Model purports to be a comprehensive representation of processes and aspects needed to predict climate, it should by definition be capable of predicting uncertainty. Currently, individual Earth System Models are not so capable, and standard ways of predicting uncertainty (e.g. in IPCC assessment reports) make use of the multi-model ensemble. Additionally, as discussed in the context of seasonal prediction timescales, stochastic parametrizations can reduce systematic biases, and therefore in some sense could be considered a “poor man’s” alternative to higher resolution. This consideration may be particularly important when limited computer resources are needed to represent biogeochemistry, the dynamic cryosphere, and other aspects of Earth-System complexity. It seems reasonable therefore to propose that next-generation Earth System Models should be explicitly stochastic (*Palmer et al., 2009*).

8 Appendix

8.1 Stochastically Perturbed Parametrization Tendencies

Now, the details of the pattern generator for the multiplicative noise used in the revised stochastically perturbed parametrization tendencies scheme are described. Let r denote, as before, the pattern in grid point space and \hat{r} its spectral transform, i.e.

$$r = \sum_{mn} \hat{r}_{mn} Y_{mn}, \quad (13)$$

where Y_{mn} , m and n denote a spherical harmonic, the zonal wavenumber and the total wavenumber, respectively. The spectral coefficients $\hat{r}_{mn} \in \mathbb{C}$ evolve according to an AR(1) process

$$\hat{r}_{mn}(t + \Delta t) = \phi \hat{r}_{mn}(t) + \sigma_n \eta_{mn}(t), \quad (14)$$

where the $\eta_{mn} \in \mathbb{C}$ denote random numbers. Real part and imaginary part of η_{mn} are independent Gaussian random numbers with unit variance and zero mean. Furthermore, η_{mn} are also independent for different spherical harmonics and white in time. The temporal correlation of \hat{r}_{mn} is controlled by ϕ , which is the correlation over one timestep Δt . Given a decorrelation timescale τ , the one-timestep correlation is set to

$$\phi = \exp(-\Delta t / \tau). \quad (15)$$

The variance of \hat{r}_{mn} depends on the one-timestep correlation according to

$$\text{var}(\text{Re } \hat{r}_{mn}) = \text{var}(\text{Im } \hat{r}_{mn}) = \frac{\sigma_n^2}{1 - \phi^2}. \quad (16)$$

The standard deviation in (14) is set to

$$\sigma_n = F_0 \exp(-\kappa T n(n+1)/2) \quad \text{with} \quad (17)$$

$$F_0 = \left(\frac{\text{var}(r) (1 - \phi^2)}{2 \sum_{n=1}^N (2n+1) \exp(-\kappa T n(n+1))} \right)^{1/2} \quad (18)$$

The choice of the normalisation factor F_0 ensures that the variance at any point on the sphere is given by $\text{var}(r)$ as the variance in grid point space is uniform and equals the total variance in spectral space.

The variance spectrum of \hat{r}_{mn} depends on total wavenumber as $\exp(-\kappa T n(n+1))$, where parameter κT determines the horizontal correlation length. Therefore, the pattern has a spatial autocorrelation corresponding to the equivalent of a Gaussian on the sphere according to [Weaver and Courtier \(2001\)](#). The correlation length scale is approximately given by $(2\kappa T)^{1/2} R_E$, where R_E denotes the earth's radius.

At initial time, the spectral random numbers are initialised by

$$\hat{r}_{mn}(0) = (1 - \phi^2)^{-1/2} \sigma_n \eta_{mn}(0). \quad (19)$$

Thus, a statistically steady state is achieved immediately without having to spin up the AR(1) processes.

In order to avoid arithmetic overflow, the random numbers η_{mn} and the spectral coefficients \hat{r}_{mn} are bounded to the range of ± 10 standard deviations. This has no practical impact on the distribution but makes the code numerically safe. Furthermore, the pattern is bounded in grid point space in order to avoid numerical instabilities. The range is limited to ± 3 standard deviations for SP1_M and to ± 2 standard deviations for SP1_L.

8.2 Stochastic backscatter

The streamfunction forcing (F_ψ) is defined to be the product of a pattern field and the square root of a total dissipation rate function (D_{tot}). The pattern field evolves in time and contains information about the power spectrum of the forcing with the dissipation rate factor acting as an amplitude modulator. The horizontal structure of the pattern field ($F_{\psi*}$) is represented by:

$$F_{\psi*,k}^j = \sum_{m=-N}^N \sum_{n=|m|}^N f_{m,n,k}^j P_n^{|m|}(\mu) \exp(im\lambda) \quad (20)$$

where m is the zonal wavenumber, n is the degree of the associated Legendre function $P_n^m(\mu)$, $f_{m,n,k}^j$ is the spectral coefficient satisfying the reality condition

$$f_{m,n,k}^j = f_{-m,n,k}^{j*}$$

where the superscript $*$ denotes the complex conjugate, μ is the sine of the latitude, λ is the longitude, k is the model level number and j is the number of timesteps of length Δt . This series expansion represents a triangular spectral truncation at spherical harmonic degree N . The time evolution of the spectral coefficients $f_{m,n,k}^j$ is assumed to be governed by a first-order auto-regressive process defined by:

$$f_{m,n,k}^{j+1} = [1 - \alpha(n)] f_{m,n,k}^j + \sqrt{\alpha(n)} g(n) r_{m,n,k}^j \quad (21)$$

where $\alpha(n)$ is a wavenumber-dependent parameter lying between 0 (no stochastic forcing) and 1 (white noise forcing); $g(n)$ is a real function controlling the spectral power in each mode and $r_{m,n,k}^j = a_{m,n,k}^j + ib_{m,n,k}^j$ where $a_{m,n,k}^j$ and $b_{m,n,k}^j$ are random numbers with zero mean and variance Σ^2 . Therefore, each spherical harmonic mode, characterized by (m, n) , is associated with a discrete time series of complex random numbers indexed by j . Inspection of the 'memory term' in equation (21) suggests the definition $\alpha(n) = 1 - \exp(-\Delta t / \tau(n))$ where $\tau(n)$ is an exponential time scale. In general the decorrelation time should be dependent on spatial scale and so allowance is made for that in the definition of τ .

Vertical structure is introduced to the pattern generator through a random phase shift scheme known as RVP (Random Vertical Phase). For any particular wave mode, the phase of the complex random number $r_{m,n,k}^j$ is evolved by an autoregressive process in the vertical:

$$\varphi_{m,n,k}^j = \varphi_{m,n,k-1}^j + \beta_{m,n,k} \varepsilon_{m,n,k}^j \quad (22)$$

where $\beta_{m,n,k}$ is a scaling parameter, whose properties depend on n and pressure, and $\varepsilon_{m,n,k}^j$ is a random number drawn from a Laplace distribution with zero mean. The modulus of $r_{m,n,k}^j$ and the phase on the first level $\varphi_{m,n,1}^j$ are given by random numbers $a_{m,n,1}^j, b_{m,n,1}^j$. The phase-shift between any two levels is given by $\beta_{m,n,k} \varepsilon_{m,n,k}^j$ and governs the vertical correlation of F_{ψ^*} . The form of the structure function $\beta_{m,n,k}$ is determined from the coarse-graining method applied to both cloud-resolving model and IFS forecast fields (see Section 5). The phase shifts applied do not affect the variance of $r_{m,n,k}^j$, so the following analysis of the energy input rate is not affected by the vertical structure and the model level index k is omitted.

The ensemble-mean variance of the spectral coefficients $f(m,n)$ can be shown to be given by:

$$\langle |f_{m,n}|^2 \rangle = 2 \frac{g(n)^2 \Sigma^2}{(2 - \alpha)} \quad (23)$$

where angle brackets denote the ensemble average and this implies an energy input rate $B_{m,n}$ into wavenumber (m,n) given by:

$$B_{m,n} = \Delta t \frac{n(n+1)}{4\pi a^2} \frac{g(n)^2 \Sigma^2}{\alpha(n)} \quad (24)$$

Summing $B_{m,n}$ over all wavenumbers gives a total energy input rate (B_{tot}) of:

$$B_{tot} = \frac{\Sigma^2 \Delta t}{4\pi a^2} \sum_{m=-N}^N \sum_{n=|m|}^N \frac{n(n+1)g(n)^2}{\alpha(n)}$$

which can be rewritten as:

$$B_{tot} = \frac{\Sigma^2 \Delta t}{8\pi a^2} \sum_{n=0}^N \frac{n(n+1)(2n+1)g(n)^2}{\alpha(n)} \quad (25)$$

since for each value of n in the summation there are $2n+1$ equal contributions from the sum over m .

Now let

$$g(n) = F_0 \chi(n) \quad (26)$$

where $\chi(n)$ is a non-dimensional function and also define the sum $\Gamma(N)$ according to

$$\Gamma(N) = \sum_{n=0}^N \frac{n(n+1)(2n+1)\chi(n)^2}{\alpha(n)}$$

Substituting into the above definitions into equation (25) gives

$$B_{tot} = \frac{F_0^2 \Delta t \Sigma^2}{8\pi a^2} \Gamma(N)$$

which on rearranging for F_0 gives:

$$F_0 = \left(\frac{8\pi a^2 B_{tot}}{\Delta t \Gamma(N) \Sigma^2} \right)^{1/2}. \quad (27)$$

The coarse-graining studies to be discussed in Section 5 suggest that $\chi(n)$ may behave like a power law for large n and our current implementation assumes:

$$\chi(n) = (1+n)^{-1.27}. \quad (28)$$

Therefore the energy input rate can be controlled through F_0 and equated to a diagnosed *total dissipation rate* multiplied by a *backscatter ratio* b_R . The full streamfunction forcing is therefore given by:

$$F_\psi = \left(\frac{b_R D_{tot}}{B_{tot}} \right)^{1/2} F_{\psi^*} \quad (29)$$

which expresses the fact that if F_{ψ^*} gives an energy input rate of B_{tot} per unit mass then F_ψ gives an energy input rate of $b_R D_{tot}$ i.e. a fraction b_R of the total dissipation rate D_{tot} .

Acknowledgements.

We are grateful to several people that contributed to this work through discussions, diagnostics or numerical experimentation: Michael Ball, Peter Bechtold, Anton Beljaars, Judith Berner, Lars Isaksen, Philippe Lopez, Martin Miller, Franco Molteni, Mark Rodwell, Frederic Vitart.

References

- Barkmeijer, J., R. Buizza, and T. N. Palmer (1999), 3D-Var Hessian singular vectors and their potential use in the ECMWF Ensemble Prediction System, *Quart. J. Roy. Meteorol. Soc.*, *125*, 2333–2351.
- Bechtold, P., E. Bazile, F. Guichard, P. Mascart, and E. Richard (2001), A mass flux convection scheme for regional and global models., *Q.J.R.Meteorol.Soc.*, *127*, 869–886.
- Berner, J., F. Doblas-Reyes, T. N. Palmer, G. J. Shutts, and A. Weisheimer (2008), Impact of a quasi-stochastic cellular automaton backscatter scheme on the systematic error and seasonal prediction skill of a global climate model., *Phil. Trans. R. Soc. A*, *366*, 2561–2579, doi:10.1098/rsta.2008.0033.
- Berner, J., T. Jung, and T. N. Palmer (2009a), Influence of stochastic parametrization versus horizontal resolution on systematic model error, manuscript in preparation.
- Berner, J., G. J. Shutts, M. Leutbecher, and T. N. Palmer (2009b), A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system, *J. Atmos. Sci.*, *66*, 603–626.
- Bormann, N., S. Saarinen, G. Kelly, and J. N. Thepaut (2003), The spatial structure of observation errors in atmospheric motion vectors from geostationary satellite data, *Mon. Weather Rev.*, *131*, 706–718.
- Bowler, N. E., A. Arribas, K. R. Mylne, K. B. Robertson, and S. E. Beare (2008), The MOGREPS short-range ensemble prediction system, *Quart. J. Roy. Meteorol. Soc.*, *134*, 703–722.

- Bowler, N. E., A. Arribas, S. E. Beare, K. R. Mylne, and G. J. Shutts (2009), The local ETKF and SKEB: Upgrade to the MOGREPS short-range ensemble prediction system, *Quart. J. Roy. Meteorol. Soc.*, *135*, 767–776.
- Buizza, R., and T. N. Palmer (1995), The singular-vector structure of the atmospheric general circulation, *J. Atmos. Sci.*, *52*, 1434–1456.
- Buizza, R., M. Miller, and T. N. Palmer (1999), Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System, *Quart. J. Roy. Meteorol. Soc.*, *125*, 2887–2908.
- Buizza, R., M. Leutbecher, and L. Isaksen (2008), Potential use of an ensemble of analyses in the ECMWF Ensemble Prediction System, *Quart. J. Roy. Meteorol. Soc.*, *134*, 2051–2066.
- Charron, M., G. Pellerin, L. Spacek, P. L. Houtekamer, N. Gagnon, H. L. Mitchell, and L. Michelin (2009), The Canadian global ensemble prediction system with a single dynamical core and stochastic physical perturbations, *submitted to Mon. Weather Rev.*
- Doblas-Reyes, F., A. Weisheimer, M. Déqué, N. Keenlyside, M. McVean, J. M. Murphy, P. Rogel, D. Smith, and T. N. Palmer (2009), Addressing model uncertainty in seasonal and annual dynamical ensemble forecasts., *Quart. J. Roy. Meteorol. Soc.*, *135*(643), 1538–1559.
- Evans, R. E., R. J. Graham, M. S. J. Harrison, and G. J. Shutts (1998), Preliminary investigations into the effect of adding stochastic backscatter to the Unified Model, *Met Office Forecasting Research Technical Report, No. 241*, Met Office, Exeter, UK.
- Frederiksen, J. S., and A. G. Davies (1997), Eddy viscosity and stochastic backscatter parameterizations on the sphere for atmospheric circulation models., *J. Atmos. Sci.*, *54*, 2475–2492.
- Gagnon, N., G. Pellerin, M. Charron, P. Houtekamer, H. Mitchell, G. Candille, L. Spacek, X.-X. Deng, B. He, and R. Verret (2007), An update on the CMC ensemble medium-range forecast system, in *Proceedings of the ECMWF 11th Workshop on Meteorological Operational Systems, 12-16 November 2007*, pp. 55–59, ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, UK.
- Gray, M. E. B. (2001), The impact of mesoscale convective-system potential-vorticity anomalies on numerical-weather-prediction forecasts, *Q. J. R. Meteorol. Soc.*, *127*, 73–88.
- Hermanson, L., B. J. Hoskins, and T. N. Palmer (2009), A comparative method to evaluate and validate stochastic parametrizations, *Q. J. R. Meteorol. Soc.*, *135*, 10.1002/qj.436.
- Hou, D., Z. Toth, and Y. Zhu (2006), A stochastic parameterisation scheme within NCEP global ensemble forecast system, in *Proceedings of the 18th AMS Conference on Probability and Statistics, 29 January - 2 February 2006, Atlanta, GA*.
- Hou, D., Z. Toth, Y. Zhu, and W. Yang (2008), Impact of a stochastic perturbation scheme on NCEP global ensemble forecast system, in *Proceedings of the 19th AMS Conference on Probability and Statistics, 21-24 January 2008, New Orleans, Louisiana*.
- Houtekamer, P. L., and H. L. Mitchell (2005), Ensemble Kalman filtering, *Quart. J. Roy. Meteorol. Soc.*, *131*, 3269–3289.
- Houtekamer, P. L., L. Lefaiivre, J. Derome, H. Ritchie, and H. L. Mitchell (1996), A system simulation approach to ensemble prediction, *Mon. Weather Rev.*, *124*, 1225–1242.

- Houtekamer, P. L., M. Charron, H. L. Mitchell, and G. Pellerin (2007), Status of the global EPS at Environment Canada, in *Proceedings of the ECMWF Workshop on Ensemble Prediction, 7-9 November 2007*, pp. 57–68, ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, UK.
- Isaksen, L., M. Fisher, and J. Berner (2007), Use of analysis ensembles in estimating flow-dependent background error variance, in *ECMWF Workshop on "Flow-dependent aspects of data assimilation", 11-13 June 2007*, pp. 65–86, ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, UK.
- Jung, T., T. N. Palmer, and G. J. Shutts (2005), Influence of a stochastic parameterization on the frequency of occurrence of North Pacific weather regimes in the ECMWF model, *Geophys. Res. Lett.*, *32*, L23,811, doi:10.1029/2005GL024248.
- Jung, T., G. Balsamo, P. Bechtold, A. Beljaars, M. Köhler, M. Miller, J.-J. Morcrette, A. Orr, M. J. Rodwell, and A. M. Tompkins (2009), The ECMWF model climate: Recent progress through improved physical parametrizations, in *Proceedings of ECMWF Seminar on Parametrization of Subgrid Physical Processes*, pp. 233–250, ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, UK.
- Leutbecher, M., and T. N. Palmer (2008), Ensemble forecasting, *J. Comp. Phys.*, *227*, 3515–3539.
- Li, X., M. Charron, L. Spacek, and G. Candille (2008), A regional ensemble prediction system based on moist targeted singular vectors and stochastic parameter perturbations, *Mon. Weather Rev.*, *136*, 443–462.
- Mason, P. J., and D. J. Thomson (1992), Stochastic backscatter in large-eddy simulations of boundary layers., *J. Fluid Mech.*, *242*, 51–78.
- Palmer, T. N. (1997), On parametrizing scales that are only somewhat smaller than the smallest resolved scales, with application to convection and orography, in *Proceedings of the ECMWF Workshop on New Insights and Approaches to Convective Parametrization, 4-7 November 1996*, pp. 328–337, ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, UK.
- Palmer, T. N. (2001), A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models, *Q. J. R. Meteorol. Soc.*, *127*, 279–304.
- Palmer, T. N., R. Gelaro, J. Barkmeijer, and R. Buizza (1998), Singular vectors, metrics, and adaptive observations, *J. Atmos. Sci.*, *55*, 633–653.
- Palmer, T. N., A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Décluse, M. Déqué, E. Díez, F. J. Doblas-Reyes, H. Feddersen, R. Graham, S. Gualdi, J. F. Guérémy, R. Hagedorn, M. Hoshen, N. Keenlyside, M. Latif, A. Lazar, E. Maïsonnave, V. Marletto, A. P. Morse, B. Orfila, P. Rogel, J. M. Terres, and M. C. Thomson (2004), Development of a european multimodel ensemble system for seasonal-to-interannual prediction (demeter), *Bull. Amer. Meteor. Soc.*, *85*, 853–872.
- Palmer, T. N., G. J. Shutts, R. Hagedorn, F. Doblas-Reyes, T. Jung, and M. Leutbecher (2005), Representing model uncertainty in weather and climate prediction, *Ann. Rev. Earth Planet. Sci.*, *33*, 163–193.
- Palmer, T. N., F.-J. Doblas-Reyes, A. Weisheimer, G. J. Shutts, J. Berner, and J. M. Murphy (2009), Towards the Probabilistic Earth-System Model, *arxiv.org: 0812.1074*.
- Plant, R. S., and G. C. Craig (2008), A stochastic parameterization for deep convection based on equilibrium statistics, *J. Atmos. Sci.*, *65*, 87–105.
- Reynolds, C. A., J. Teixeira, and J. G. McLay (2008), Impact of stochastic convection on the Ensemble Transform, *Notes and Correspondences of Mon. Wea. Rev.*, *136*, 4517–4526.

- Rodwell, M. J., and T. Jung (2008), The ECMWF ‘Diagnostic Explorer’: A web tool to aid forecast system assessment and development, *ECMWF Newsletter 117*, ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, UK.
- Shutts, G. J. (2005), A kinetic energy backscatter algorithm for use in ensemble prediction systems, *Q. J. R. Meteorol. Soc.*, *131*, 3079–3102.
- Shutts, G. J. (2006), Upscale effects in simulations of tropical convection on an equatorial beta-plane, *Dynamics Of Atmospheres And Oceans*, *42*, 30–58.
- Shutts, G. J., and M. E. B. Gray (1994), A numerical modelling study of the geostrophic adjustment process following deep convection, *Q. J. R. Meteorol. Soc.*, *120*, 1145–1178.
- Shutts, G. J., and T. N. Palmer (2007), Convective forcing fluctuations in a cloud-resolving model: Relevance to the stochastic parameterization problem, *J. Clim.*, *20*, 187–202.
- Teixeira, J., and C. A. Reynolds (2008), Stochastic nature of physical parameterisations in ensemble prediction: a stochastic convection approach, *Mon. Weather Rev.*, *136*, 483–496.
- Tompkins, A. M., K. Gierens, and G. Rädcl (2007), Ice supersaturation in the ECMWF integrated forecast system, *Quart. J. Roy. Meteorol. Soc.*, *133*, 53–63.
- Vallis, G. K., G. J. Shutts, and M. E. B. Gray (1997), Balanced mesoscale motion and stratified turbulence forced by convection, *Quarterly Journal Of The Royal Meteorological Society*, *123*, 1621–1652.
- Vialard, J., F. Vitart, M. Balmaseda, T. Stockdale, and D. Anderson (2005), An ensemble generation method for seasonal forecasting with an ocean atmosphere coupled model, *Mon. Weather Rev.*, *133*, 441–453.
- Weaver, A., and P. Courtier (2001), Correlation modelling on the sphere using a generalized diffusion equation, *Quart. J. Roy. Meteorol. Soc.*, *127*, 1815–1846.
- Weisheimer, A., F. Doblas-Reyes, T. N. Palmer, A. Alessandri, A. Arribas, M. Déqué, N. Keenlyside, M. MacVean, A. Navarra, and P. Rogel (2009), Ensembles - a new multi-model ensemble for seasonal-to-annual predictions: Skill and progress beyond demeter in forecasting tropical pacific ssts., *Geophys. Res. Lett.*, accepted.