

Stochastic representations of model uncertainties at ECMWF: State of the art and future vision

Martin Leutbecher^{a*}, Sarah-Jane Lock^a, Pirkka Ollinaho^b, Simon T. K. Lang^a,
Gianpaolo Balsamo^a, Peter Bechtold^a, Massimo Bonavita^a, H. M. Christensen^c,
Michail Diamantakis^a, Emanuel Dutra^a, Stephen English^a, Michael Fisher^a, Richard M. Forbes^a,
Jacqueline Goddard^a, Thomas Haiden^a, Robin J. Hogan^a, Stephan Juricke^c, Heather Lawrence^a,
Dave MacLeod^c, Linus Magnusson^a, Sylvie Malardel^a, Sebastien Massart^a, Irina Sandu^a,
Piotr K. Smolarkiewicz^a, Aneesh Subramanian^c, Frédéric Vitart^a, Nils Wedi^a, Antje Weisheimer^{a,c,d}

^aEuropean Centre for Medium-Range Weather Forecasts, Reading, UK

^bFinnish Meteorological Institute, Helsinki, Finland

^cAtmospheric, Oceanic and Planetary Physics, University of Oxford, UK

^dDepartment of Physics, National Centre for Atmospheric Science, University of Oxford, UK

*Correspondence to: M. Leutbecher, ECMWF, Shinfield Park, Reading, RG2 9AX, UK; E-mail: M.Leutbecher@ecmwf.int

Members in ensemble forecasts differ due to the representations of initial uncertainties and model uncertainties. The inclusion of stochastic schemes to represent model uncertainties has improved the probabilistic skill of the ECMWF ensemble by increasing reliability and reducing the error of the ensemble mean. Recent progress, challenges and future directions regarding stochastic representations of model uncertainties at ECMWF are described in this paper. The coming years are likely to see a further increase in the use of ensemble methods in forecasts and assimilation. This will put increasing demands on the methods used to perturb the forecast model. An area that is receiving a greater attention than 5 to 10 years ago is the physical consistency of the perturbations. Other areas where future efforts will be directed are the expansion of uncertainty representations to the dynamical core and to other components of the Earth system as well as the overall computational efficiency of representing model uncertainty.

Key Words: ensemble forecasts, ensemble data assimilation, weak-constraint 4D-Var, numerical weather prediction, dynamical core, Earth system model, model uncertainty, stochastic parametrization

Received ...

1. Introduction

Weather forecasting is an initial value problem solved with numerical models that describe the evolution of the state of the atmosphere and other Earth system components interacting with it. Since the seminal work of Edward Lorenz in the 1960s, it is understood that the equations describing the atmospheric flow exhibit sensitive dependence on initial conditions that leads to forecast error growth and eventual loss of predictability (Lorenz 1969; Buizza and Leutbecher 2015).

Some forecast error will be due to model imperfections even if the model had been initialised with an initial state that corresponded exactly to the true state of the Earth system. We propose to define *model error* as the error in forecasts and model climate that would be observed had the model been initialised with the initial state corresponding exactly to the true state. This is not without subtlety as the map from the physical system to the state space of the model involves some filter operation on the scales that are absent in the model. This will render the state of the model that corresponds exactly to the true state ambiguous. By assuming

there is a model state corresponding to the true state of the system, we include the filtering operation of the initial conditions in our definition of model error.

Due to the chaotic nature of geophysical fluid dynamics, model error will limit predictability in addition to initial error. We expect that for systems exhibiting sensitive dependence on initial conditions, solutions will also exhibit sensitive dependence on the models used in the numerical integrations. This becomes obvious when considering an n -day forecast as an $(n - \ell)$ -day forecast initialised from an ℓ -day forecast ($\ell < n$). The $(n - \ell)$ -day forecast depends sensitively on its initial condition, which is the ℓ -day forecast, which in turn is model-dependent.

The process of translating the laws of physics governing the evolution of the Earth system to a numerical code, i.e. the “model”, inevitably requires many simplifications and approximations. These arise from incomplete knowledge of a physical process and its ancillary data, reduced complexity to limit computational costs, omission or mis-representation of processes and system components, from uncertain parameters in

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/qj.3094

parametrizations that do not have a directly observable equivalent, and last but not least from discretisation. Related to the latter are errors due to the omission of fluctuations on the unresolved scales as discussed for instance by Palmer (2001).

Quantification of forecast uncertainties became established in operational numerical weather prediction (NWP) in the 1990s using ensemble forecasting techniques (Lewis 2005). The physical equations describing the evolution of the atmosphere are nonlinear and therefore ensemble-based Monte-Carlo approaches appear as the only feasible way of estimating the future probability distribution of the state of the atmosphere. Building an ensemble forecast system requires the specification of the sources of uncertainties. It seems useful to distinguish between uncertainties in the initial conditions and uncertainties in the forecast model both in terms of the methods as well as the implications for predictability. From now on, for brevity, we will refer to them as initial uncertainties and model uncertainties, respectively. In this paper, we distinguish between actual model errors, where there is only one realisation per model and forecast, and model uncertainty representations, which sample model perturbations from some distribution to make a prediction of the forecast uncertainty. The perturbations should ideally have the same statistics as the error but are likely to have different statistics in practical applications. An ensemble that represents only initial uncertainties consistent with the true distribution of initial condition errors is known to be underdispersive and therefore will lack reliability (e.g. Wilks 2005; Palmer *et al.* 2005). This motivates the development of methodologies to represent model uncertainties.

In the following, we will focus on stochastic representations of model uncertainties (e.g. Palmer 2012). These are schemes that sample perturbations from some underlying distribution that defines the scheme that represents model uncertainties. A key advantage is that these schemes involve a distribution that can be adjusted to control the characteristics of the model error representation. Generally, the aim of stochastic representations of model uncertainty is to simulate the effect of the random component of the errors of the model tendencies. However, this may include flow-dependent systematic errors that appear like random errors.

ECMWF first introduced a stochastic representation of model uncertainties in the medium-range ensemble in October 1998 using a scheme that multiplies the total parametrized physics tendencies with a random number (Buizza *et al.* 1999). The scheme was originally referred to as “stochastic physics”. It is now commonly referred to as the Stochastically Perturbed Parametrization Tendency scheme (SPPT). Major revisions of the stochastic representation of model uncertainties took place in September 2009 and November 2010 (Palmer *et al.* 2009; Shutts *et al.* 2011). These changed aspects of the probability distribution sampled by the SPPT scheme. In addition, a Stochastic Kinetic Energy Backscatter scheme (SKEB) was activated in ECMWF ensemble forecasts in the November 2010 upgrade.

As ensembles are becoming more widely used in forecasting and assimilation, the need for representations of model uncertainties is increasingly recognised and this fuels an expansion of research on this topic. In April 2016, ECMWF and the World Weather Research Programme (WWRP) jointly organised a workshop on model uncertainty with over 80 participants. The workshop proceedings (ECMWF/WWRP 2016) contain summaries of the talks as well as the recommendations from the three working groups. Looking at the research presented at the workshop and reviewing the literature reveals that a range of alternative approaches to represent model uncertainties have been explored and are currently developed across the wider community (see references in Secs. 2, 3 and 7.2). The main reasons for the multitude of approaches are that: (i) it is difficult

to accurately characterise model error and (ii) there are many different sources of model error. Thus, it is not straightforward to decide whether one way of representing model error is better than another. A related challenge is the difficulty in disentangling initial uncertainties from model uncertainties as the estimation of the initial state involves the use of a forecast model.

This paper is based on a *special topic paper* prepared for the ECMWF Scientific Advisory Committee held at ECMWF in October 2016. Its purpose is to (i) report on progress that has been made since the last special topic paper on stochastic parametrization prepared for the ECMWF Scientific Advisory Committee in 2009 (Palmer *et al.* 2009) and to (ii) discuss ideas that influence plans for future work on the representation of model uncertainties in ECMWF’s prediction system. Forecast errors also arise from systematic errors that lead to biases. Such errors and work on identifying their root causes are beyond the scope of this paper. However, the impact of stochastic representations of model uncertainties on the model climate will be covered. This paper will also look at the use of model uncertainty representations in data assimilation.

The outline of the paper is as follows. The operational stochastic representations of model uncertainties used at ECMWF are described in Section 2. Work on developing a process-oriented representation of model uncertainties is summarized in Section 3. The impact of model uncertainty representations on a range of applications at ECMWF is documented in Section 4. Unrepresented sources of model uncertainty in the Earth system are discussed in Section 5. Future directions for work on model uncertainty are presented in Section 6. We conclude with a brief discussion and summary in Sections 7 and 8.

2. Operational methods in the IFS

The Integrated Forecasting System (IFS) is ECMWF’s operational NWP model. This section is dedicated to the operational stochastic parametrizations used in the IFS, i.e. SPPT and SKEB.

2.1. The Stochastically Perturbed Parametrization Tendency scheme

The Stochastically Perturbed Parametrization Tendency scheme (SPPT) makes the assumption that the dominant error of the parametrized physics tendency is proportional to the net physics tendency (Buizza *et al.* 1999). SPPT generates perturbed parametrization tendencies \mathbf{p} stochastically by multiplying the net physics tendency \mathbf{p}_D provided by the physics package with a 2D random field r

$$\mathbf{p} = (1 + \mu r)\mathbf{p}_D. \quad (1)$$

Here, \mathbf{p} denotes the vector of perturbed tendencies of temperature, specific humidity, and wind components in a model column. The factor μ is an optional tapering function that depends on the model level only and has values in the range $[0, 1]$.

The number of NWP models in which a SPPT scheme has been implemented keeps rising. It is used operationally in global ensembles by Environment Canada (Charron *et al.* 2010; Separovic 2016, ECMWF/WWRP workshop) and by Japan Meteorological Agency. Météo-France uses SPPT in the regional ensemble based on the Application of Research to Operations at Mesoscale convection-permitting model (AROME, Bouttier *et al.* 2012). It has been tested as well in the Weather Research and Forecasting model (WRF) in ensembles with parametrized convection and in convection-permitting ensembles (Berner *et al.* 2015; Romine *et al.* 2014). Pégion (2016, ECMWF/WWRP workshop) reported recent tests of SPPT in the United States National Centers for Environmental Prediction’s Global Forecast System model (GFS) and Sanchez *et al.* (2016) describe tests

in the Met Office Unified Model. In summary, SPPT has been found to be effective in generating additional ensemble spread and improving probabilistic skill in a range of NWP ensembles.

Shutts and Pallarès (2014) diagnosed coarse-grained differences of tendencies from integrations with different horizontal resolution and examined the relationship between the width of the probability distribution and the mean value of the tendency of the coarser resolution model. They find different relationships for different physical processes, e.g. convection and radiation. This finding motivated work by Arnold (2013), who explored the effect of applying noise *independently* to the tendencies from radiation, vertical mixing and orographic drag, convection, cloud processes, non-orographic drag, and methane oxidation. Christensen *et al.* (2016) summarize more recent tests with this so-called independent SPPT (iSPPT), which results in an increase in ensemble spread in regions with significant convective activity. The iSPPT approach also offers a larger degree of flexibility, which could facilitate sensitivity studies.

Recently, studies by Davies *et al.* (2013) and Peters *et al.* (2013) raised doubts whether the multiplicative ansatz in SPPT was consistent with observations of the relationship between deep convection and the large-scale state. Watson *et al.* (2015) examined this relationship for the IFS and found that it can reproduce the observed relationships in a deterministic integration as well as in an integration perturbed with SPPT.

2.1.1. The ECMWF methodology

The random field r in (1) is obtained through first order autoregressive processes in spectral space. A multi-scale approach is implemented in IFS with

$$r = \sum_{j=1}^J r_j \quad (2)$$

where the component random fields r_j are independent and represent different scales. Table 1 lists the standard deviation in grid point space, the spatial auto-correlation scale L and the time decorrelation scale τ of the components of the three-scale pattern ($J = 3$) used in the medium-range, extended-range and seasonal (System 4) ensemble forecasts since 2010. Shutts *et al.* (2011) show an example of a realisation of the three components. The ensemble of 4D-Var data assimilations (EDA) uses the single-scale pattern with the fast small-scale pattern only ($J = 1$) mainly for reasons of technical simplicity and lack of investigations assessing the impact of the slower and larger-scale correlations on the background error structures. A first assessment of using the three-scale pattern in the EDA is summarized in Section 4.4.

The implementation of SPPT in the Integrated Forecast System (IFS) uses a tapering function μ which is 1 in the free troposphere and reduces the amplitude gradually to 0 close to the surface and in the stratosphere. The latter is done to avoid large amplitude perturbations of the radiative tendencies in the stratosphere, which are presumed to be more accurate, and the former to avoid numerical instabilities in the boundary layer.

Apart from the limiters mentioned below, r samples a Gaussian distribution with a grid-point space variance of $\sigma^2 = \sum_{j=1}^J \sigma_j^2$, with σ_j^2 being the grid-point variance of the component pattern r_j . When r falls outside of the range $[-1, 1]$, the magnitude of r is reduced to 1. This is required for numerical stability and avoids sign reversal of the tendencies. In addition, a supersaturation limiter reduces the magnitude of r to avoid excessive supersaturation due to the perturbation. The supersaturation limiter determines a vertically consistent reduction of the amplitude of the T and q perturbations based on a mean over the 200 hPa layer requiring the largest reductions.

scale j	σ	L (km)	τ (d)
1	0.52	500	0.25
2	0.18	1000	3
3	0.06	2000	30

Table 1. Characteristics of the three-scale random field used in SPPT: Standard deviation σ , horizontal decorrelation length scale L , time decorrelation scale τ . The operational EDA uses only the first scale $j = 1$.

2.1.2. Global fix for tendency perturbations

In long runs with the EC-Earth climate model version 3.1, SPPT caused substantial imbalances in the radiative fluxes, the surface fluxes of precipitation (P) and evaporation (E) and for the radiation budgets (Davini *et al.* 2017). The SPPT scheme systematically reduced humidity in the atmosphere which was compensated by overly strong evaporation. The global P–E imbalance in the stochastically perturbed simulations was increased from about -0.016 mm d^{-1} in the runs without SPPT to about -0.16 mm d^{-1} . This is based on a 10-year mean of simulations with and without SPPT at TL255L91 resolution*. In 30-day forecasts with IFS CY41R1, a qualitatively similar signal is diagnosed with an increase of the P–E imbalance from a value of 0.03 mm d^{-1} to a value of -0.15 mm d^{-1} when SPPT is activated. It is worth noting that P and E are the raw model output as provided to users and consistent with the practices in the forecast evaluation without accounting explicitly for imbalances introduced by SPPT (see also Sec. 6.2.2).

SPPT also had a significant impact on the energy fluxes at the surface and the top of the atmosphere. The 10-year mean of global top-of-the-atmosphere net flux changed from -1.71 to -2.77 W m^{-2} through the activation of SPPT while the surface net flux changed from -0.46 to -3.66 W m^{-2} in the aforementioned EC-Earth runs. Imbalances exceeding $\sim 1 \text{ W m}^{-2}$ are considered unacceptable for climate models (Mauritsen *et al.* 2012). For comparison, the present-day radiative forcings due to anthropogenic well-mixed greenhouse gases is about 3 W m^{-2} (Myhre *et al.* 2013).

To address these imbalances, a modification of the SPPT scheme was developed. A correction is added to the perturbed tendency, which results in the global integral of the perturbed tendency being equal to that of the unperturbed tendency. Details are described in Appendix A. In terms of medium-range and extended-range ensemble forecast scores, it was found that this SPPT modification has a neutral to slightly positive impact. In view of this impact and the fact that it makes the climate of the forecasts perturbed with SPPT more similar to the climate of unperturbed forecasts, this modification of SPPT was activated in the model upgrade (CY43R1) in November 2016.

2.1.3. Discussion

SPPT can be viewed as a “holistic approach” that maintains the overall balance between the tendencies due to different physical processes (Tim Palmer, pers. comm.). In the presence of compensating biases of the tendencies from different processes, the bias of the total tendency can be smaller than the biases of individual tendencies. A potential strength of SPPT is that its formulation results in unbiased perturbations of the physics when supersaturation limiter, interactions and subsequent nonlinearities are neglected. The problems with the systematic impact on the humidity tendencies discussed above indicate that in practice it is difficult to obtain unbiased perturbations with SPPT.

*Henceforth, we will refer to triangular truncation at wavenumber NNN using a linear grid by TLNNN; the L91 refers to the 91 level vertical discretisation with a model top at 1 Pa.

The mean and the perturbation structure of the tendencies are controlled by the IFS physics. In the limit of vanishing perturbation variance, the scheme converges to the deterministic IFS physics. The scheme is relatively simple and it is computationally cheap: less than 2% of total runtime, or less than 4% with the fix described in Sec. 2.1.2. Furthermore, the scheme requires only modest resources for maintenance. As will be seen later, it is efficient in generating ensemble spread and contributes positively to the probabilistic skill of the ECMWF ensemble forecasts. There are also noticeable impacts on the model climate (cf. Sec. 4.3) and on the EDA and thus data assimilation in general (cf. Sec. 4.4).

However, there are also several limitations of SPPT. It assumes that the error of the physics tendency is always proportional to the deterministic tendency \mathbf{p}_D while the true uncertainty is likely to have also variance in many directions not parallel to \mathbf{p}_D . Considering a single time step, the perturbations are confined to a one-dimensional subspace in a space with a dimension of a few hundred (4 variables times the number of model levels). For example, uncertainty in the shape of a heating profile cannot be captured with SPPT. The validity of the simple multiplicative ansatz has previously been questioned by Shutts and Palmer (2007) as well as Shutts and Pallarès (2014). Their coarse-grained tendencies show evidence of non-vanishing uncertainty when the parameterised physics tendency vanishes.

The modulation of the amplitude of the perturbation with the total tendency implies that the same level of uncertainty is assigned to all processes and all atmospheric situations. This may not reflect actual variations in model uncertainty well. For instance, one would expect that on average longwave radiative cooling in a clear-sky situation is more certain than the heating associated with parametrized deep convection. Moreover, the amplitude modulation implies that the error vanishes where the total tendency is zero while the true error distribution will also be influenced by non-compensating errors of the individual processes.

Another drawback of the current formulation of SPPT is that it does not respect conservation laws. Fluxes at the top of the atmosphere and the surface are not perturbed. This implies that local budgets of energy and moisture that are satisfied by the deterministic parametrizations are violated by the SPPT perturbations.

All implementations of SPPT share the basic principle that the parametrized total physics tendency is multiplied by a random number. However, there are potentially important differences in detail. These differences include the variance of the perturbations, the space and time auto-correlation of the random pattern, the shape of the distribution that is sampled and whether perturbations are suppressed in some regions of the atmosphere or for some processes. For example, the implementations of SPPT in the model of the COntsortium for Small-scale MOdelling (COSMO) and the WRF model do not use the tapering to zero of the perturbations in the boundary layer. While these differences will limit the ability to transfer conclusions obtained with one model to other models, they may motivate a range of sensitivity experiments in the future.

2.2. The Stochastic Kinetic Energy Backscatter scheme

Stochastic Kinetic Energy Backscatter scheme (SKEB) aims to represent model uncertainties associated with scale interactions that take place in the real atmosphere but are absent in a truncated numerical model. Motions on scales that would be fully resolved in the model interact with motions on scales that would be near grid-scale or subgrid-scale in the model.

Following ideas in Large Eddy Simulation (LES), Shutts (2005) and Berner *et al.* (2009) developed a stochastic forcing for the IFS

model targeting this uncertainty. The SKEB scheme is also used in the Environment Canada and in the United Kingdom MetOffice (UKMO) global ensembles (Charron *et al.* 2010; Tennant *et al.* 2011). Sanchez *et al.* (2016) propose improvements to the SKEB scheme used by UKMO. Berner *et al.* (2011) and Berner *et al.* (2015) study the impact of SKEB on WRF ensembles. Pegion (2016, ECMWF/WWRP workshop) summarizes initial tests of SKEB in global GFS ensembles. We note in passing that there are potentially important differences in the detail of the various implementations of SKEB in different models. For example, the versions in the Canadian model and in the Met Office model perturb different ranges of wavenumbers and the implementation in WRF is purely additive, i.e. state-independent.

Recently, Shutts (2015) introduced a stochastic convective backscatter scheme, which focusses entirely on the random model error arising from the interaction of parametrized deep convection with the model dynamics near the grid scale.

2.2.1. The ECMWF methodology

In the IFS, SKEB introduces a stochastic streamfunction forcing $\partial\Psi/\partial t$ given by an evolving 3-dimensional pattern F with an amplitude modulation determined by a horizontally smoothed local estimate of kinetic energy sources at the subgrid-scale, referred to as the dissipation rate D

$$\left. \frac{\partial\Psi}{\partial t} \right|_{\text{SKEB}} = [bD]^{1/2} F \quad (3)$$

Here b is the backscatter ratio that controls the global amplitude of the perturbations. The horizontal structure of the pattern is determined by a power-law variance distribution in spectral space while the vertical structure is obtained from random-phase shifts that decorrelate the structure in the vertical. The dissipation rate is a 3D field diagnosed from the model state. In the original version of the scheme, the dissipation rate estimate is obtained as the sum of three terms describing subgrid-scale energy sources due to orographic gravity wave drag, numerical dissipation and parametrized deep convection. The streamfunction perturbations are tapered to zero in the boundary layer similar to the approach used by ECMWF for SPPT.

The scheme evolved recently to a version with only the contribution from deep convection. The dissipation rate contribution from orographic gravity-wave drag was abandoned as it generated excessive ensemble spread in lower tropospheric wind near steep orography, i.e. the Andes. The term linked to numerical dissipation estimated from the horizontal diffusion resulted in spurious kinetic energy spectra with the new cubic octahedral grid (Malardel *et al.* 2016) implemented in March 2016 due to the inconsistency in representing horizontal mixing in the model and in SKEB. Therefore, numerical dissipation was deactivated in SKEB.

The computational cost of SKEB is higher than that of SPPT due to requiring spectral transforms of two 3D fields: the pattern F and the dissipation rate estimate D . The operational version does not update these fields every model time step to save computational time.

2.2.2. Discussion

As described above, the IFS implementation of SKEB has evolved to a scheme that is conceptually similar to the Stochastic Convective Backscatter algorithm (SCB) proposed by Shutts (2015). The latter scheme is effective in generating ensemble spread and introduces strong wind perturbations in the boundary layer. However, at present it is unknown whether these low-level perturbations are a good or a bad characteristic and whether they

target the main model uncertainties associated with deep moist convection. Another difference between the operational version of SKEB and Shutts' SCB is that the latter perturbs the velocity potential and the streamfunction on synoptic scales in the vicinity of parametrized deep convection while the former is limited to streamfunction perturbations.

With increasing spatial resolution, the model is able to represent more of the scale interactions explicitly and the need for SKEB should be reduced. This, could be at least partly a justification for deactivating the numerical dissipation rate contribution with the higher resolution TCo639 ensemble[†]. Nevertheless, until deep convection is explicitly resolved, one may see a need for including a stochastic convective backscatter. However, the stochastic convective momentum transport perturbations can also be introduced in an alternative manner (cf. Section 3). Recent work by [Malardel and Wedi \(2016\)](#) looking at the energy transfer in high-resolution, convection-permitting IFS integrations indicates that backscatter happens through available potential energy.

There is a generic challenge in keeping the actual perturbations used by SKEB in line with the spatial scales that can be justified theoretically — while large-scale perturbation structures appear to be most effective in generating ensemble spread, the dominant uncertainty due to the interaction with the unresolved scales should be at scales close to the truncation scale.

3. Towards process-level representation of model uncertainties

Parametrizations of physical processes include a number of tunable parameters, which quantify efficiencies, rates of change, etc. of phenomena that the parametrizations seek to represent. Uncertainty in the parameter values leads to a source of model uncertainty. Sampling values of uncertain key parameters provides a way to represent uncertainties at their sources and thus links to individual physical processes.

A scheme with stochastically perturbed parameters has been developed in the UKMO global ensemble ([Bowler et al. 2008b](#)) and later, applied in the UKMO convection-permitting ensemble ([Baker et al. 2014](#)). The original Random Parameters (RP) scheme was based on parameters that vary stochastically, but discontinuously, in time. [McCabe et al. \(2016\)](#) report improvements of ensemble fog forecasts in the 2.2 km regional UKMO ensemble from implementing the RP scheme. In their experiments, the parameters vary more gradually in time ("RP2") than in previous implementations of the RP scheme. Highlighting future developments, Tennant (2016, ECMWF/WWRP workshop) demonstrated further improvement in global ensemble forecasts from enabling the stochastic parameters to vary in time and space ("RP3").

Stochastic parameter perturbations are beginning to be considered beyond the atmosphere in other Earth system components. [Juricke et al. \(2014\)](#) have explored the impact of stochastic perturbations of sea ice strength on sea ice predictability. [Brankart et al. \(2015\)](#) introduce generic uncertainty representations in the Nucleus for European Modelling of the Ocean model (NEMO) that include a form of SPPT as well as stochastic parameter perturbations. They propose exploring stochastic parameter perturbations to represent uncertainty in marine ecosystem modelling that arises from restricting the diversity of species. In addition, they also propose perturbing the sea ice strength parameter. Stochastic and fixed parameter perturbations also start to be explored in the H-TESSSEL[‡] land

Table 2. The parameters and variables perturbed by SPP in the physical process parametrization schemes.

TURBULENT DIFFUSION & SUBGRID OROGRAPHY
transfer coefficient for momentum (ocean/land)
coefficient in turbulent orographic form drag scheme
stdev. of subgrid orography
length scale for vertical mixing in stable boundary layer
CONVECTION
entrainment rate
shallow entrainment rate
detrainment rate for penetrative convection
conversion coefficient cloud to rain
zonal convective momentum transport
meridional convective momentum transport
adjustment time scale in CAPE closure
CLOUD & LARGE-SCALE PRECIPITATION
RH threshold for onset of stratiform cond.
diffusion coefficient for evaporation by turbulent mixing
critical cloud water content for autoconversion
threshold for snow autoconversion
RADIATION
cloud vertical decorrelation height in McICA
fractional stdev. of horizontal distribution of water content
effective radius of cloud water and ice
scale height of aerosol normalised vertical distribution
optical thickness of aerosol

surface model with the aim of representing model uncertainties ([MacLeod et al. 2016](#); [Orth et al. 2016](#)).

3.1. ECMWF's Stochastically Perturbed Parametrization methodology

At ECMWF, work has started on stochastic parameter perturbations in the framework of the Stochastically Perturbed Parametrization scheme (SPP, [Ollinaho et al. 2017](#)). SPP provides a framework in the IFS code to represent some of the key random errors of the parametrized tendencies close to their sources within the physical processes. Like SPPT, SPP is strongly guided by the existing deterministic parametrizations. The proximity of the uncertainty model to the processes permits the exploitation of physically consistent relationships between different variables. For instance, SPP can produce perturbations to the fluxes at the top of the atmosphere and the surface that are consistent with the tendency perturbation in the model column. Thus local budgets of moisture, momentum and energy, which are respected by the deterministic parametrizations, remain closed. SPP can also represent uncertainty beyond a simple amplitude error, e.g. the uncertainty in the shape of a heating profile.

SPP can be seen as a generalisation of the concept of perturbed parameters as it introduces local stochastic perturbations to parameters and variables in the parametrizations. [Ollinaho et al. \(2017\)](#) describe the methodology in detail and give a justification for the selection of parameters and variables. During the initial development phase of SPP, experts working on the IFS parametrization of individual processes identified 20 parameters and variables that are considered uncertain and when changed introduce significant changes in the forecast (Tab. 2).

Figure 1 displays the distributions sampled by SPP in the parametrizations of (a) turbulent diffusion and subgrid orographic drag, (b) convection, (c) cloud and large-scale precipitation and (d) radiation. SPP samples the distributions independently for each parameter and variable. Thus, the perturbations are uncorrelated. The perturbations evolve using the same type of auto-regressive AR(1) pattern generator in spectral space as

[†]We refer to triangular truncation at wavenumber NNN using a cubic octahedral grid by TCoNNN.

[‡]Revised Hydrology for the Tiled ECMWF Scheme for Surface Exchanges over Land

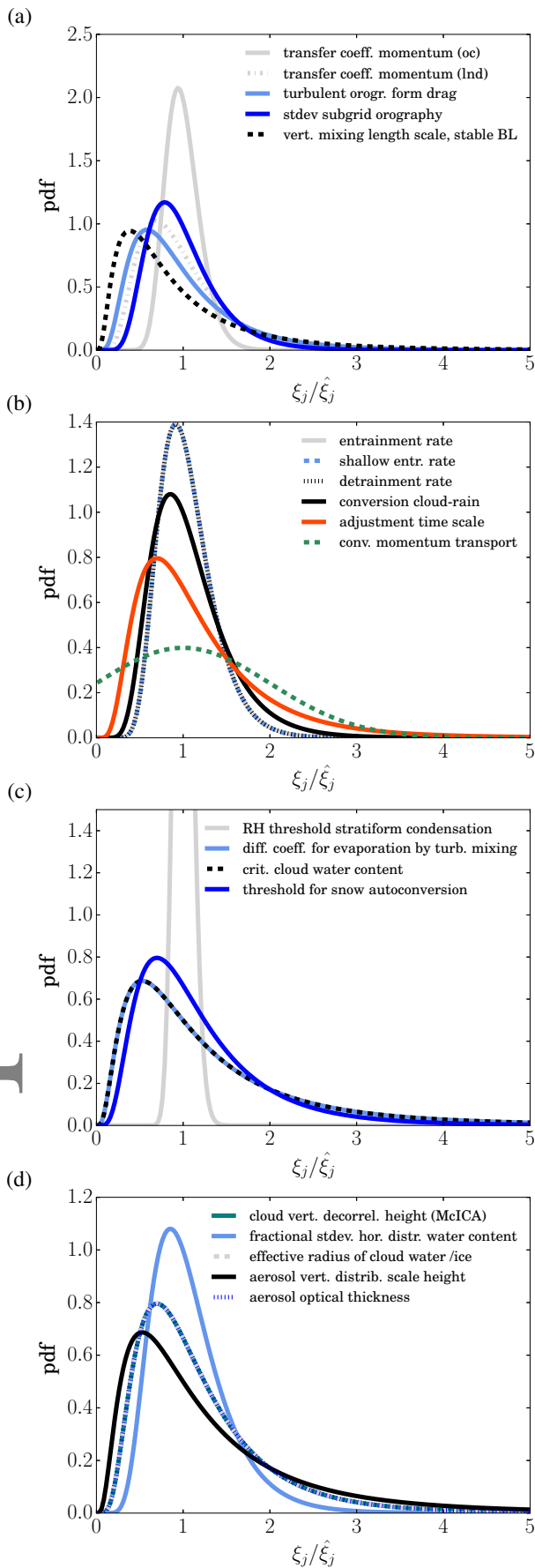


Figure 1. Distributions of parameters and variables sampled by SPP in the parametrizations of (a) turbulent diffusion and subgrid orographic drag, (b) convection, (c) cloud and large-scale precipitation and (d) radiation. The x -axis shows the ratio of the perturbed value ξ_j to the unperturbed value $\hat{\xi}_j$ with the index j referring to the different parameters. The left tail of the distribution for convective momentum transport extends to negative values and is not shown. The realisations of different parameters are independent and have prescribed horizontal and temporal correlations.

SPPT but with correlation scales of 2000 km and 72 h. The perturbed parameters sample a log-normal distribution except for the convective momentum transport. The latter samples a normal distribution that permits sign reversal, which corresponds to an upgradient momentum transport that can occur in mesoscale convective systems. For all log-normal distributions, the probability density for reaching a ratio of zero vanishes. The parametrization experts helped to define an initial guess for the degree of uncertainty of these variables. Numerical experimentation with the medium-range ensemble was used to refine the estimates.

Ollinaho *et al.* (2017) report sensitivity experiments with different spatial and temporal correlation scales; these indicate the largest impact on ensemble spread for scales of 2000 km and 72 h. Shorter decorrelation scales (500 km and 6 h) as well as globally constant parameter perturbations are less effective in generating ensemble spread and led to lower probabilistic skill in the medium-range ensemble.

SPP is a more complex scheme than SPPT and therefore requires more resources to maintain and improve it, which may be seen as a disadvantage. However, there may also be synergies as the development of the uncertainty component and the development of the deterministic component can mutually inform each other. There is a potential to improve the uncertainty models by studying uncertainty at the process level which is not meaningful for SPPT, which follows a uniform approach for all processes. The work on SPP is still in the early stages of development and several potential future extensions are being considered. These will be described in Sections 5.1 and 5.2.

3.2. A look at the tendency perturbations

Representations of model uncertainty generate differences between ensemble member forecasts through the modification of the model tendencies. Examining the tendency perturbations induced by different representations of model uncertainty offers a way to quantitatively compare schemes. Initially, we have focussed on a comparison of SPPT and SPP to better understand in what ways the two schemes differ and to guide future developments (see also Ollinaho *et al.* 2017). Ensemble experiments without initial perturbations, one with SPPT and one with SPP were run with the IFS at TL399 resolution with 91 levels. Tendency output was produced 3-hourly on model levels.

Figure 2 shows maps of the ensemble standard deviation of the accumulated temperature tendencies in the first 3 hours of the forecast for 4 December 2013, 3 UTC. In the mid-troposphere, the regions with highest standard deviation coincide for SPPT and SPP (Fig. 2a,b). These regions are those that are convectively active as can be seen from the ensemble mean convective precipitation (Fig. 3a). Outside of the convectively active regions, the standard deviation drops below $0.05 \text{ K}(3 \text{ h})^{-1}$ in the experiment with SPP while it is above that value in a large fraction of the night-time globe with SPPT. This day-night modulation of the uncertainty in SPPT is consistent with the unperturbed tendency as well as the ensemble mean tendency. During night-time, there is a prevalent cooling in the mid-troposphere with values ranging between 0.1 and $1 \text{ K}(3 \text{ h})^{-1}$. During day-time, the ensemble mean tendency outside the convectively active regions is closer to zero. This day-night contrast is due to the radiative tendencies (Fig. 3c). During the day, the heating due to the solar radiation and the long-wave thermal emission approximately balance each other while the latter is dominating the tendencies in a large fraction of the globe at night. The fact that SPPT attributes a significant level of uncertainty to the night-time clear-sky cooling is considered as an unrealistic description of model uncertainty.

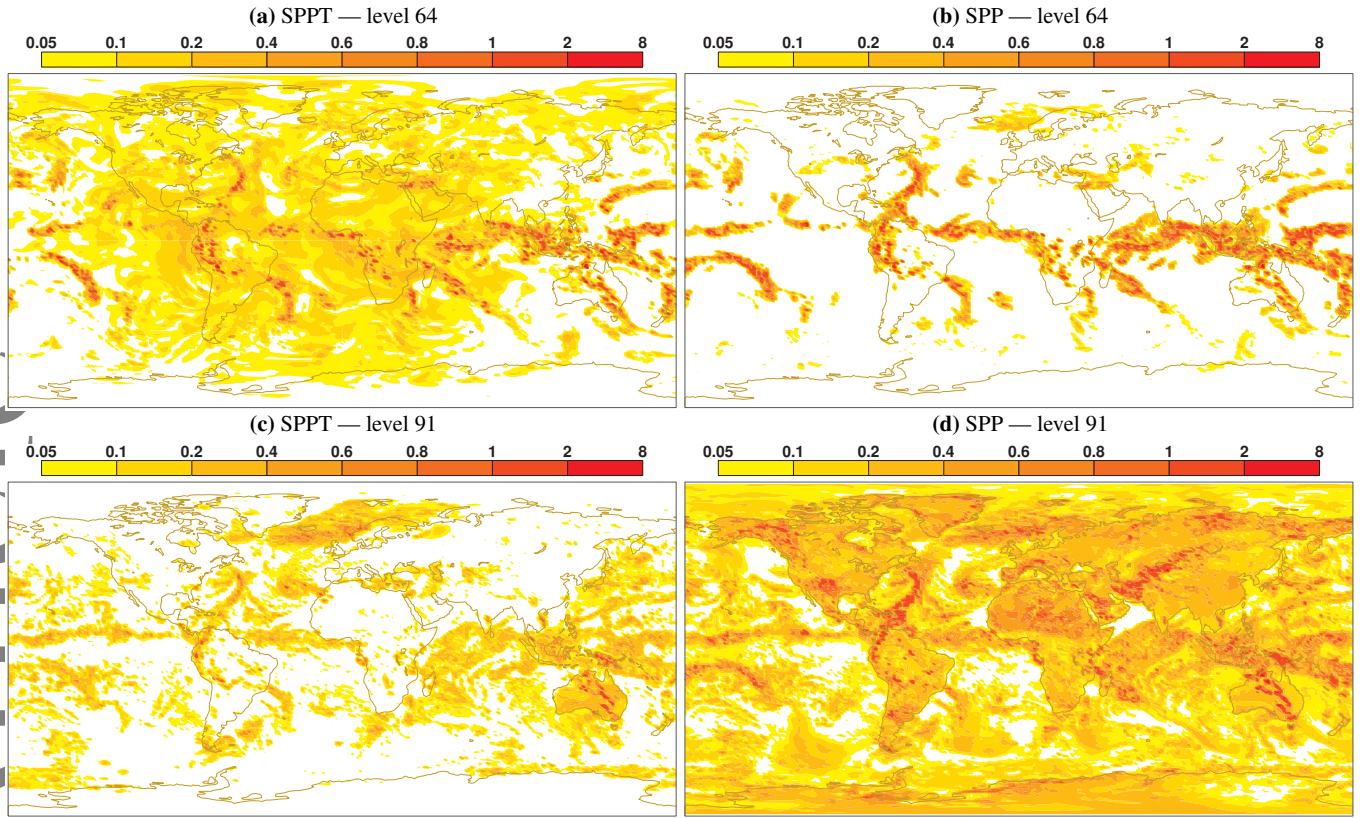


Figure 2. Ensemble standard deviation of 0–3 h accumulated tendencies of temperature (unit: K (3 h)^{-1}) at model levels 64 (~500 hPa) and 91 (10 m above ground) for experiments perturbed with SPPT (a,c) and SPP (b,d), 4 December 2013, 0 UTC, lead time 3 h.

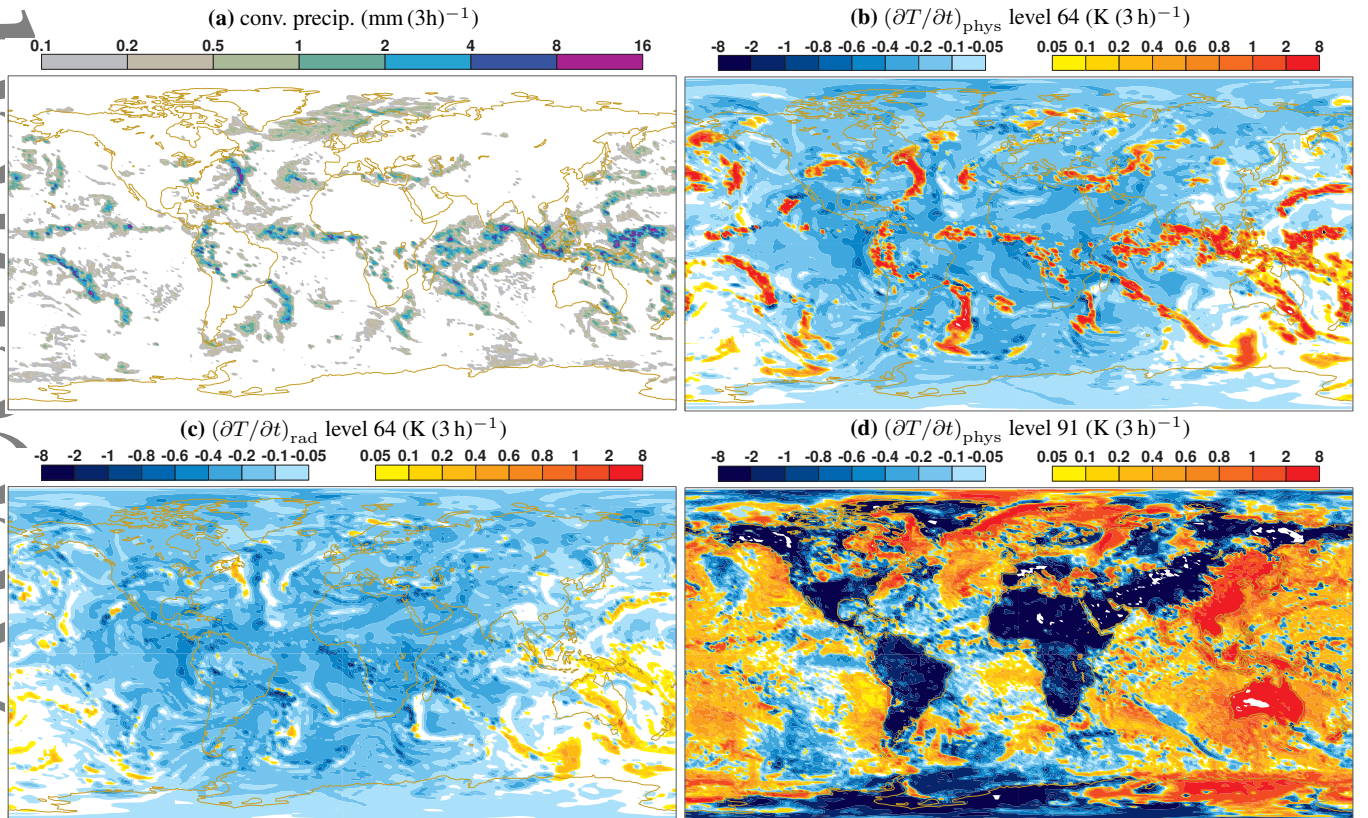


Figure 3. Ensemble mean of (a) convective precipitation, (b–d) temperature tendencies: (b,d) total physics temperature tendency at levels 64 and 91; (c) radiative temperature tendency at level 64. Experiment with SPPT, 4 December 2013, 0 UTC, lead time 3 h. The experiment with SPP shows very similar ensemble means.

In the first three hours of the forecast, SPP generates much larger tendency perturbations close to the surface than SPPT everywhere (Fig. 2c,d). The fact that SPPT assumes a high level of certainty close to the surface is also considered as an unrealistic aspect. The ratio between the standard deviation of the

perturbations and the mean tendency at low levels is less than 2% for SPPT in a large fraction of the domain (compare Fig. 2c with Fig. 3d). Future modifications of SPPT regarding the tapering in the boundary layer and the treatment of clear-sky radiances may address these apparent deficiencies.

By aggregating statistics over a number of cases and over a region, systematic differences between the two model uncertainty schemes can be quantified. Ollinaho *et al.* (2017) compare SPP and SPPT in this way. Above level 80 (~ 900 hPa), the SPPT scheme produces more spread, especially in the extratropics (consistent with Fig. 2a,b). For tendency perturbations later in the forecast, say 21–24 h, the difference between the two schemes in the free-troposphere has increased with SPPT clearly generating larger tendency perturbations. In the lowest model levels, SPP still generates larger spread though the standard deviations are more similar than in the first 3 hours. Even though SPPT does not directly introduce perturbations in the boundary layer, the differences in the free-troposphere have by 24 h forced the boundary layer differently.

4. Impact of stochastic model uncertainty representations in the IFS

This section summarises the impact of the stochastic model uncertainty representations SPPT, SKEB and SPP on a range of applications at ECMWF from the medium-range and extended-range ensemble forecasts, to the model climate, the EDA and weak-constraint 4D-Var.

4.1. Medium-range forecasts

First, the impact of the operational versions of SPPT and SKEB on ensemble forecasts in the medium-range and extended-range is briefly summarized. Results are based on a sample of 46 30-day forecasts at TCo255/159L91 resolution and 20 members with a resolution change from TCo255 to TCo159 at Day 15. Using a lower resolution and smaller ensemble size than the operational configuration reduces the cost of numerical experimentation by a factor of about 25 and makes a range of sensitivity experiments affordable. The impact of SPPT and SKEB is expected to be not highly sensitive to the spatial resolution as the fragile numerical dissipation rate estimate in SKEB had been deactivated. Future experimentation is planned to quantify the resolution dependence of the impact of the model uncertainty representations. This will be useful for guiding the development strategy for model uncertainty.

The operational model uncertainty representation generates additional spread in the ensemble with the largest absolute gain after about 1 week. Figure 4a,b shows changes of ensemble standard deviation for 200 hPa zonal wind. The impact is larger in the tropics than in the extratropics reaching more than 20% spread increase over the experiment with initial perturbations only (IP-only) in the former and less than 10% in the latter. The spread increase is statistically significant in the first 2 weeks (beyond 4 weeks) in the extratropics (tropics). Most of the spread increase is due to SPPT with SKEB adding little additional spread. An experiment that uses SKEB and initial perturbations but does not use SPPT confirms the marginal impact of the current configuration of SKEB.

The model uncertainty representations reduce the ensemble mean root-mean-square (RMS) error and the continuous ranked probability score (CRPS) markedly in the tropics with relative improvements of about 10% (Fig. 4d,f). In the extratropics, skill is improved moderately but still statistically significantly reaching reductions of CRPS of up to 2% (Fig. 4c,e). The additional skill improvements due to SKEB are negligible.

The impact of SPP has been compared with that of SPPT using the same TCo255/159 setup. The results are consistent with those of Ollinaho *et al.* (2017), who used a different resolution TL399 configuration with a forecast range to 15 days. Figure 5 shows the impact of SPP and SPPT on ensemble spread and CRPS for 200 hPa zonal wind in the Northern extratropics and

the tropics. SPP is effective in generating considerable additional ensemble spread. The gain in spread compared to experiment IP-only reaches 65% (more than 100%) of the gain in spread due to SPPT in the extratropics (tropics). The probabilistic skill measured in terms of CRPS improves with SPP though the skill improvements are generally somewhat lower than those obtained with SPPT. This is the case also for other upper air variables. A notable exception is 2-metre temperature in the early lead times where the ensemble using SPP is more skilful than the ensemble using SPPT (cf. Ollinaho *et al.* 2017). This is consistent with the impact of the schemes on the ensemble spread of the tendencies discussed in the previous section.

The fact that uncertainty representations for the model physics are having a more substantial impact on ensemble spread and skill in the tropics than elsewhere is consistent with earlier results (Buizza *et al.* 1999) and more recently with the results of Reynolds *et al.* (2011), who investigated constant in time and space parameter perturbations in the convection scheme and the boundary layer parametrization.

4.2. Extended-range forecasts and the MJO

At the extended time range (more than 2 weeks, but less than a season), a key source of predictability is the Madden Julian Oscillation (MJO), which impacts not only the tropical but also the European extended-range skill scores (Vitart and Molteni 2010). This section discusses the impact of stochastic representations of model uncertainty on predictions of the MJO using the leading 2 principal components (PCs) developed by Wheeler and Hendon (2004). The PCs are based on combined empirical orthogonal functions of outgoing long-wave radiation and zonal wind at 850 and 200 hPa. The MJO PCs computed from each experiment are then verified against the ECMWF global atmospheric reanalysis ERA-Interim.

Figure 6 shows the evolution with lead time of the ensemble mean RMS error and ensemble spread of the two MJO PCs for an experiment with initial perturbations only (IP-only), SPPT using the global conservation fix, SPP, and SPP using correlation scales of 500 km and 6 h (SPP_short). The spread is computed as $[v_1 + v_2]^{1/2}$, where the v_j denote the ensemble variances of the two PCs. The ensemble mean RMS error is computed as $[e_1^2 + e_2^2]^{1/2}$, where the $e_j = \bar{x}_j - y_j$ denote the errors of the two PCs of the ensemble mean with ensemble mean PCs \bar{x}_j and analysis PCs y_j . The MJO ensemble spread is generally significantly smaller than the RMS error. The experiment with SPP displays a much larger spread than the other experiments, but also a slightly increased RMS error, although the difference at day 26–32 between SPP and IP-only is not statistically significant at the 90% level of confidence. When the spatial and temporal correlation scales in SPP are reduced (experiment SPP_short), little spread is added compared to IP-only. Experiment SPPT adds considerable spread and decreases RMS error compared to IP-only between day 10 and 20, although the difference in RMS error is not statistically significant.

Ranked Probability Skill Scores (RPSS, not shown) for SPP and SPPT are generally higher than for IP-only, particularly for week 1 in the Tropics. Experiment SPP has a lower skill than Experiment SPPT for most variables and in particular at 50 hPa and for velocity potential at 200 hPa.

When SKEB is added to SPPT, its impact is not statistically significant in terms of the MJO PCs and RPSS scores (not shown). This is consistent with the minor differences in spread and probabilistic skill seen in the medium-range (Sec. 4.1).

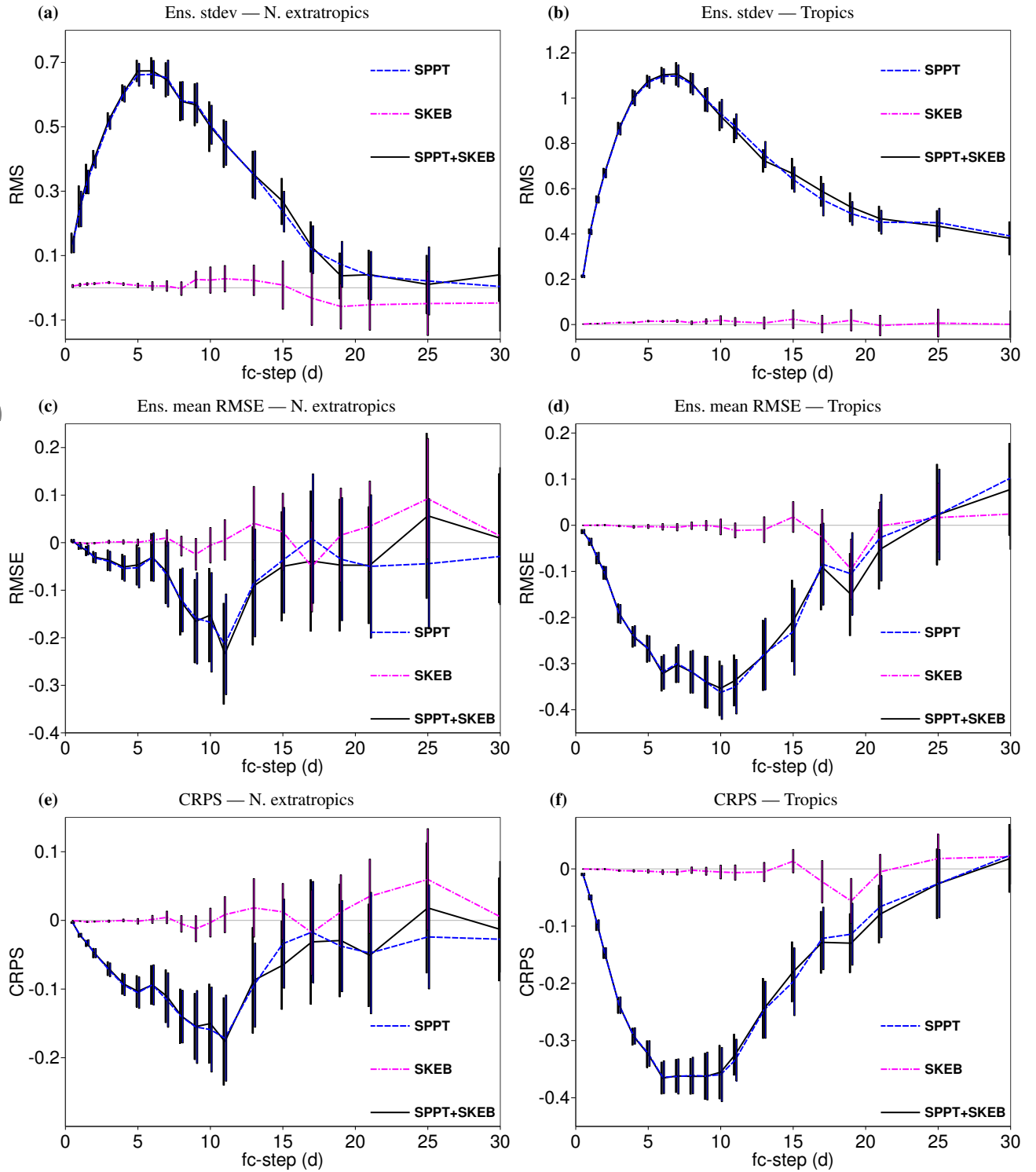


Figure 4. Impact of SPPT, SKEB and SPPT+SKEB on 200 hPa zonal wind (m s^{-1}) relative to an experiment with initial perturbations only (IP only). Top panels: ensemble stdev, middle panels: ensemble mean RMS error, bottom panels: continuous ranked probability score (CRPS). Based on 46 cases covering Dec 2014 to Nov 2015, 20 members, TCo255/159L91 resolution, cycle 42r1. Bars show 95% confidence intervals based on a paired t-test.

4.3. Model climate and systematic errors

In this section, we will look into the impact on the model climate of SPPT and SPP separately. Although the perturbations applied in both schemes are drawn from distributions that are centred around zero, it is not guaranteed that the mean response from the model will be zero. Non-linear physical processes such as saturation of humidity (generation of clouds and precipitation) can give rise to asymmetries in the model response.

Weisheimer *et al.* (2014) and Subramanian *et al.* (2017) documented the impact of SPPT on the ECMWF seasonal forecasting system 4 (Sys4). They found a positive impact during DJF on the precipitation bias over the Maritime continent, which

led to an improved Walker circulation and MJO statistics. Since the implementation of Sys4, several model changes have been introduced, and the impact of SPPT might have changed. In this section, we use seasonal re-forecasts from 1981 to 2010 with 3 ensemble members to investigate the impact of the stochastic schemes on the mean model state. The forecasts have been initialised on 1 May and 1 November and run for 4 months in order to enable verification of JJA and DJF. All simulations are performed with the atmospheric model at TCo255 resolution with 91 vertical levels, coupled to the 1 degree ocean model NEMO. The model cycle used is 42r1. The SPPT_gfix experiment includes the global conservation fix (discussed in Section 2.1.2). The fix

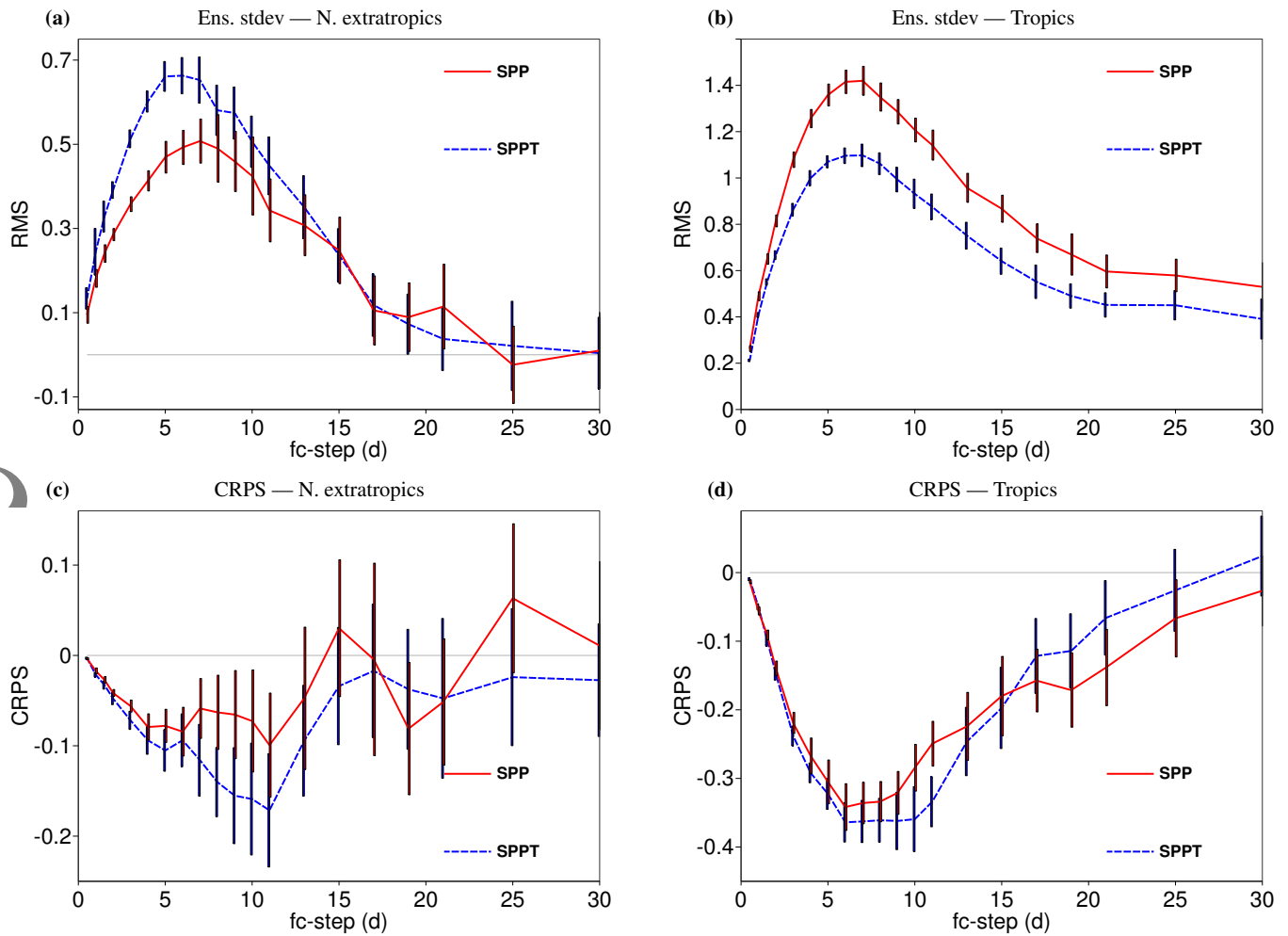


Figure 5. Impact of SPP and SPPT on 200 hPa zonal wind (m s^{-1}) relative to an experiment with initial perturbations only (IP only). Top panels: ensemble stdev, bottom panels: continuous ranked probability score (CRPS). Based on 46 cases covering Dec 2014 to Nov 2015, 20 members, TCo255/159L91 resolution, cycle 42r1.

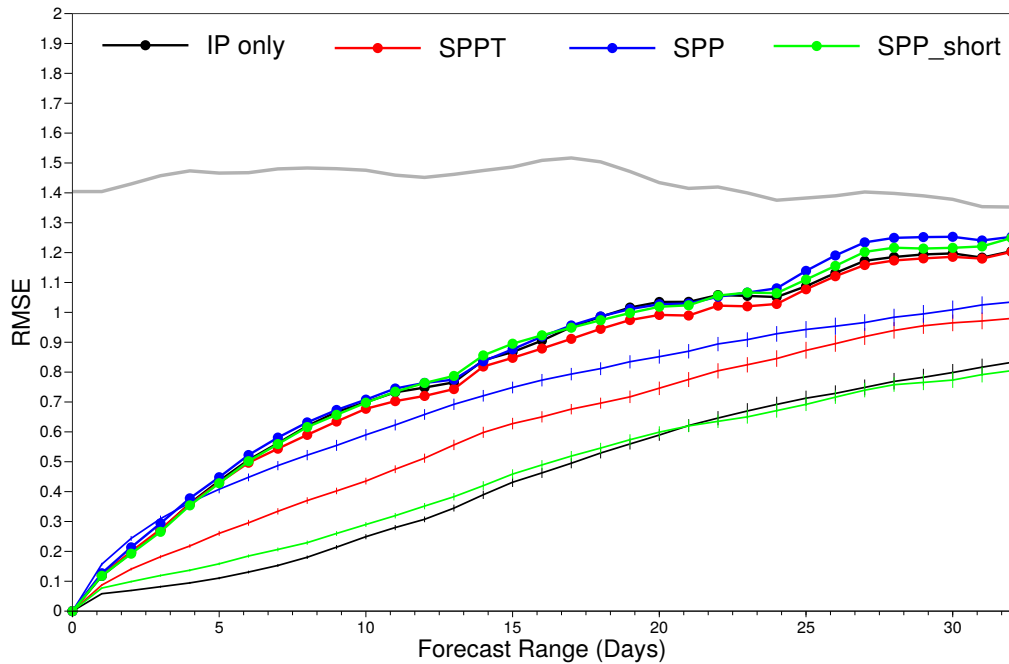


Figure 6. Evolution of RMS error (solid lines with dots) and ensemble spread (thin lines) for the leading 2 MJO PCs. Initial perturbations only (black), SPPT with global conservation fix (red), SPP (blue) and SPP with short correlation scales (green). The grey solid line represents the RMS error of ERA-Interim climatological mean. The vertical bars represent the 90% level of confidence of the ensemble spread. 32-day reforecasts starting 4 times a year (1st February, May, August and November) over the period 1989–2014 (104 start dates). The ensemble size is 15, and the horizontal resolution is TCo159 with 91 vertical levels.

reduces the drying of the atmosphere due to standard SPPT and leads to increased precipitation and also a net warming of the troposphere.

SPPT, SPPT_gfix and SPP have a mixed impact on the model climate compared to the unperturbed model (Table 3). For example, SPP clearly improves the 50hPa geopotential (Z50).

	SPPT	SPPT _{gfix}	SPP
MSLP (eint)	-1.0	-1.0	0.3
Z500 (eint)	-0.4	-2.5	0.1
Z200 (eint)	3.7	0.4	1.4
Z50 (eint)	15.0	0.6	-9.2
Z10 (eint)	12.2	3.8	-2.8
U925 (eint)	-1.9	-1.9	-0.1
U850 (eint)	-2.2	-2.2	-0.7
U500 (eint)	-1.0	-1.0	-0.1
U200 (eint)	-0.8	-0.4	1.0
U100 (eint)	-0.8	4.1	0.1
U50 (eint)	1.5	3.0	2.7
U10 (eint)	16.7	14.2	5.9
T925 (eint)	-0.5	-1.2	-0.2
T850 (eint)	-2.7	-1.0	-1.4
T500 (eint)	-2.2	0.9	1.4
T200 (eint)	2.2	-2.0	-3.8
T100 (eint)	-1.5	-0.7	-5.6
T50 (eint)	6.5	12.9	0.7
T10 (eint)	-5.3	-5.8	-2.0
Precip (gpcp)	-4.0	-1.4	-2.3
SSHF (NOC)	3.9	1.3	-4.2
SLHF (NOC)	2.5	0.0	-2.0
TSR (CERES)	-0.0	-0.2	1.1
TTR (CERES)	-4.7	-4.9	-4.7
TCWV (SSMI)	-0.7	-1.9	-1.7
T2m (cru)	0.2	-1.0	0.5
SST (eint)	-3.1	-2.4	-1.6

Table 3. Relative difference (in percent) in mean absolute error of seasonal means with respect to mean absolute error of seasonal means in unperturbed experiment. Mean sea level pressure (MSLP), geopotential (Z), zonal wind (U), temperature (T), precipitation (Precip), surface sensible heat flux (SSHF), surface latent heat flux (SLHF), top of the atmosphere solar radiative flux (TSR), top of the atmosphere thermal radiative flux (TTR), total column water vapour (TCWV), 2-metre temperature (T2m) and sea surface temperature (SST) — all parameters verified against ERA-Interim (eint) and GPCP are a global mean while against satellite products (NOC, CERES and SSMI) are from 65°N–65°S. The values for CRU and GPCP depend on the distribution of observations. SSMI and NOC are only over ocean.

However, this improvement is caused by a warming in the troposphere during JJA on the northern hemisphere. The warming is degrading the model climate (temperature at 500 hPa) but compensates a cold bias (in the unperturbed model) in the upper troposphere and the net effect on Z50 is positive. The warming of the summer hemisphere has been traced to the perturbations of the radiation parameters. It is an open question whether introducing perturbations to an independently tuned model will generally worsen biases and whether additional tuning of the perturbed model could result in a less biased model. Superposition of biases of an unperturbed model with biases caused by perturbation methods could also explain to some extent how stochastic schemes may behave differently in different models.

For the surface heat-fluxes (SSHF, SLHF), SPPT (without fix) is worse than the unperturbed experiment over ocean, while SPPT_{gfix} degrades the fluxes less and SPP improves the fluxes. For all experiments, the sensible upward heat-fluxes are too strong. The difference between the experiments could at least partly be related to wind speeds (and ocean wave-height) over the storm tracks. Here SPPT gave a positive bias in the wind speed that leads to increased (upward) surface-fluxes. On the other hand, SPP shows reduced winds and weaker fluxes. Whether the sensitivity in wind speeds are a direct effect of the perturbation schemes or a result of changes in the general circulation is an open question.

The largest degradation of the model climate with SPPT and SPPT_{gfix} is found in the stratospheric winds, especially at 10hPa. The SPPT scheme and to some degree SPP increases the westerly bias over the equator, where the quasi-biennial oscillation (QBO) is the main source of variability. The degraded modelling of the QBO is of concern for seasonal forecasting as it has been suggested in the literature that the QBO has teleconnections to the northern extra-tropical winter temperatures.

The improvement in total precipitation (Figure 7) and top-of-the-atmosphere thermal radiation (TTR) with both schemes is linked to a reduction of biases in the tropics. Reduction of the magnitude of the bias of optically thick clouds is found over the Maritime continent, the Amazon and tropical Africa, alongside a reduction in precipitation. A reduction of the precipitation biases is also present over the north-eastern Indian ocean during the monsoon season (JJA) for both SPPT and SPPT_{gfix}. However, the improvement in the precipitation over the Maritime continent from SPPT_{gfix} is less than was found for Sys4 (Weisheimer *et al.* 2014). The difference in impact from SPPT_{gfix} is at least partly explained by the introduction of the global tendency fix (Sec. 6.2.2) that increases the precipitation over the Maritime continent.

The schemes also affect the variability in the model. For example, both SPPT and SPP increase and thereby improve the MJO activity, especially over the western Pacific. The improvement of the MJO by SPPT (without the conservation fix) was discussed in Weisheimer *et al.* (2014). Inness *et al.* (2003) documented the sensitivity of the MJO to the model mean state over the Western Pacific and DeMott *et al.* (2015) provide a review. SPPT and to a lesser degree SPP have also an impact on the overall variability, which is evident in the kinetic energy spectra at 200 hPa (not shown). SPPT increases the activity on all scales while SPP increases the activity mainly in divergent planetary scales (probably related to the MJO activity). The increased activity due to SPPT in synoptic scale and mesoscale is detrimental — leading to overactivity in comparison to ERA-Interim. For SPPT_{gfix}, the increase in model activity is not confined to any specific region but is strongest in the convectively active tropical regions (Maritime continent, Amazon, central Africa).

4.4. Ensemble of data assimilations

ECMWF uses an ensemble of data assimilations (EDA) to estimate background error covariances for 4D-Var and to sample initial uncertainties for the ensemble forecasts (Fisher 2003a; Buizza *et al.* 2008; Isaksen *et al.* 2010; Bonavita *et al.* 2014). Model uncertainty affects forecasts at all lead times and we think that there is very limited physical justification for using different approaches for representing model uncertainty at different lead times, i.e. in assimilation ensembles, medium-range, extended-range and seasonal forecasts. This does not preclude though that the relative importance of uncertainty representations may change for different lead times. Representing uncertainty in slow processes may be more crucial for the longer lead times. Insufficient error growth in a model may require additional perturbations that are relevant for longer lead times and are not required for the shorter lead times. The overall development of model uncertainty representations can benefit by being able to transfer lessons learnt in one application to other applications. Therefore, the ECMWF strategy is to work towards using the same representation of model uncertainties in forecasts and assimilation. The plan is to first explore a consistent representation in forecasts and assimilation using SPPT and SPP.

Here, recent experimentation is summarized that assesses the impact of SPP and two versions of SPPT on the EDA. The

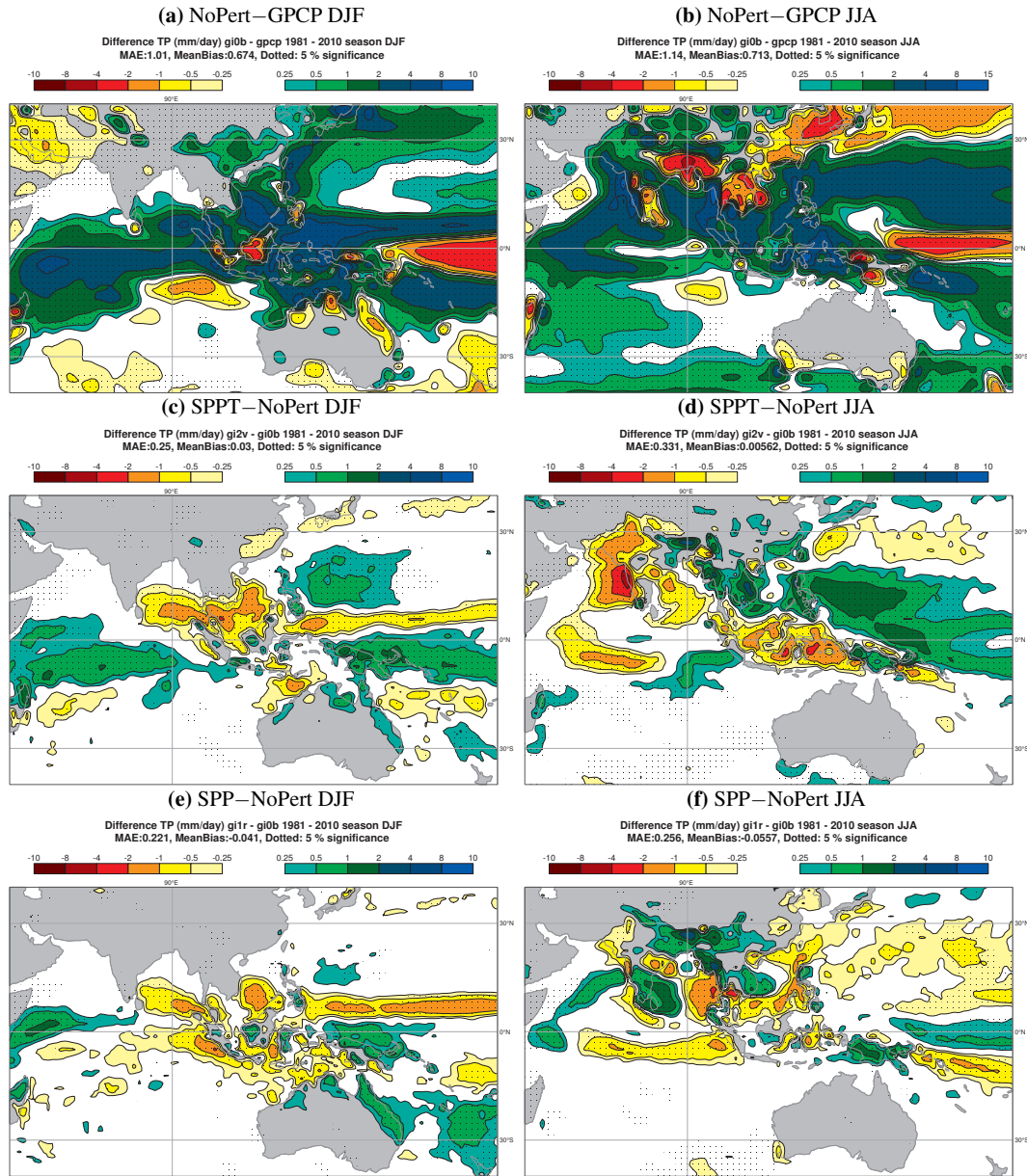


Figure 7. Seasonal mean precipitation error in unperturbed simulation (NoPert) verified against GPCP: (a) DJF, (b) JJA. Mean precipitation difference of experiments SPPT with global conservation fix and NoPert (c,d) and precipitation difference of experiments SPP and NoPert (e,f) for DJF and JJA, respectively.

SPPT experiments shown here do not include the global fix for the tendencies yet. Four EDA experiments are run in total: one experiment without model uncertainty representation (NoPert), one experiment with the 1-scale SPPT (SPPT1) operational in the EDA, one experiment with 3-scale SPPT (SPPT3) and one experiment with SPP (SPP). The outer loop resolution is TL399 with 137 vertical levels. The experiments span the period 1 June to 11 July 2015.

In experiments SPPT3 and SPP, the random pattern evolves continuously across consecutive assimilation windows. Due to the technical overhead, this aspect had been missing in the SPPT1 version that is operational in the EDA. In SPPT1, all perturbed member nonlinear forecasts initialise the pattern from a random number seed ignoring the state of the pattern in the preceding assimilation cycle. However, with the longer de-correlation time scales of SPPT3 and SPP it was deemed important to correctly propagate the pattern in the assimilation cycles.

All three representations of model uncertainty lead to a considerable increase in first guess ensemble variances overall, e.g. in terms of meridional wind (Fig. 8). The inclusion of the slower time scales in SPPT and the cycling of the pattern (Exp. SPPT3) increases the spread compared to the operational version

of SPPT while the spatial distribution remains similar (Fig. 8b and c).

The SPP scheme is more effective in generating spread in the tropics and close to the surface than the operational version of SPPT (Fig. 8b and d). However in the extra-tropics in the free troposphere, it generates less spread than SPPT. This is consistent with the differences seen in the standard deviation of the tendencies (cf. section 3.2). Future plans include an assessment of the impact of the different model error schemes on the reliability of EDA variances via observation space diagnostics, e.g. following the work of Rodwell *et al.* (2016). The interpretation of observation space diagnostics applied to the EDA with and without model error schemes is expected to benefit from improvements in the characterisation of observation error statistics, e.g. the efforts described in section 6.1.2.

Representing model uncertainty in the EDA may not only affect the estimated background error variances, it can also alter the background error correlation length scales (see Fisher 2003b, for a description of the covariance model used at ECMWF). Activating any of the three model uncertainty representations moderately increases the horizontal correlation length scales. Experiments SPPT1 and SPPT3 have up to $\sim 10\%$ longer correlation length

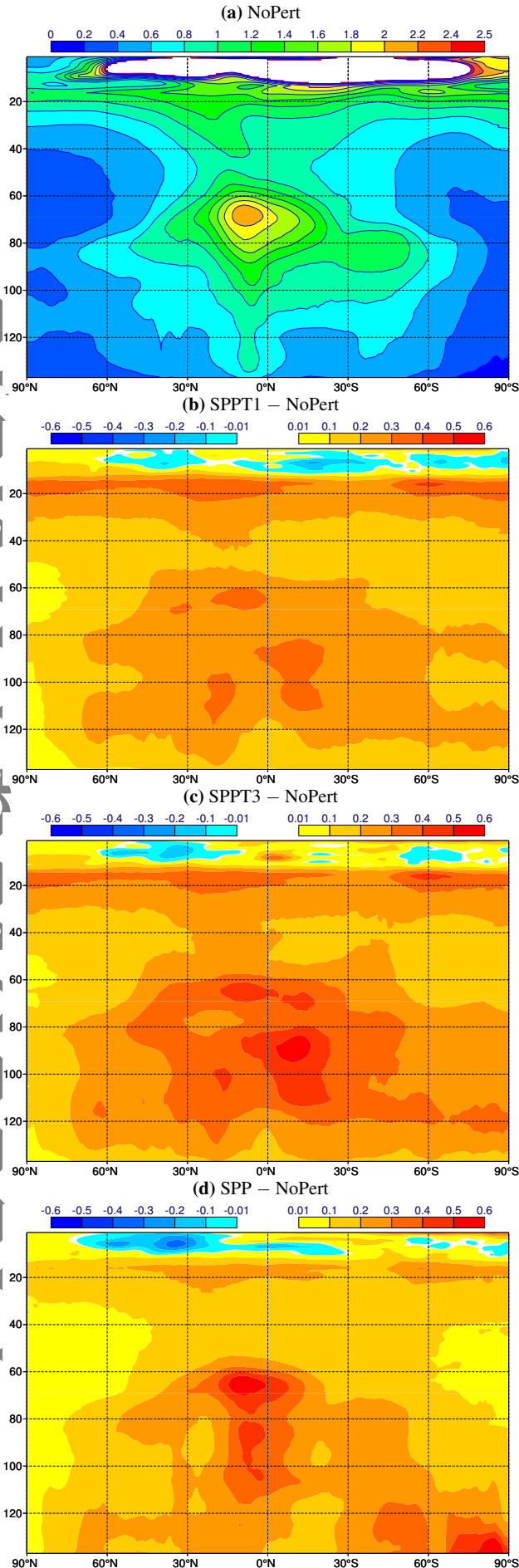


Figure 8. Latitude model level section of zonal mean background error standard deviation for meridional wind (m s^{-1}) estimated from four EDA experiments: Panel (a) shows the stdev of NoPert while the other panels show differences between the three experiments with a model error representation and NoPert. Average from 1 June to 11 July 2015.

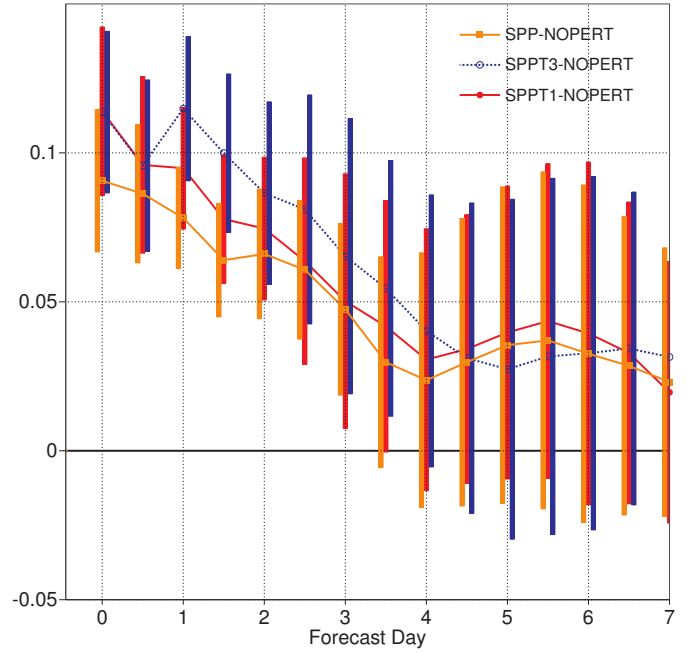


Figure 9. Difference of anomaly correlation coefficient between experiments with model uncertainty representation and NoPert for 500 hPa geopotential in the northern extratropics. The difference is normalised by one minus the anomaly correlation of the NoPert experiment. Positive values indicate an improvement with respect to NoPert; SPP (orange), SPPT3 (blue) and SPPT1 (red). The bars show 95% confidence intervals. The forecasts are verified against the operational high-resolution analysis.

scales than NoPert for the unbalanced part of temperature with the largest increases occurring between 900 and 300 hPa (not shown). The SPP scheme increases the length scales by up to 15% with the maximum increase in the boundary layer. The impact of the model uncertainty representations on the background error vorticity length scales is smaller — an increase up to 5% (not shown).

The background error covariances predicted by the EDA can be globally evaluated by using them in a 4D-Var assimilation and quantifying the impact on forecast skill. In the operational EDA and the experiments discussed here, each member uses the background error covariances predicted in the EDA. The background error covariance matrix is based on a mixture of a “climatological” part and a flow-dependent part with errors of the day (Bonavita *et al.* 2014). Thus, the deterministic skill of the 10-day forecast initialized from the control member every 24 hours can be used to indirectly measure the global quality of the predicted background error covariances. All three model error representation schemes improve the overall forecast scores compared to NoPert considerably, e.g. 500 hPa geopotential in the extra-tropics (Fig. 9). The differences in skill between the three experiments with a model uncertainty representation are minor in comparison.

The global fix for the SPPT tendency perturbations was tested in the EDA separately (not shown). The impact on the background error variances and correlation length scales computed from the EDA is minor. In line with this, the impact on the deterministic forecast skill of the EDA control member is neutral. The fix brings average values of global integrated moisture and cloud cover of the perturbed members closer to the unperturbed control member.

4.5. Model error covariances for weak-constraint 4D-Var

In November 2016 (model version CY43R1), ECMWF reintroduced the model error term in 4D-Var. It had been turned off in 2013 due to issues with the continuous slow growth of the model error forcing term over many subsequent assimilation cycles. This

issue was linked to the vertical resolution upgrade from 91 to 137 levels. In the interim, the operational assimilation scheme at ECMWF was based on strong-constraint 4D-Var (Rabier *et al.* 2000; Mahfouf and Rabier 2000; Klinker *et al.* 2000). As other aspects of data assimilation processes have advanced, the validity of the perfect model assumption used in strong-constraint 4D-Var becomes more questionable and limits the length of the analysis window to roughly 12 hours. Weak-constraint 4D-Var relaxes the perfect model assumption by explicitly representing model error as part of the 4D-Var control variable (Trémolet 2007).

Model error contains both random and systematic (or even constant) components. To simplify the problem, we consider the model error to be constant during the 12-hour assimilation window. The weak-constraint 4D-Var cost function includes a term that penalises updates of the model forcing η

$$\frac{1}{2}(\eta - \eta_b)^T \mathbf{Q}^{-1}(\eta - \eta_b), \quad (4)$$

where η_b is a model error background value that changes slowly from cycle to cycle and \mathbf{Q} is the model error covariance matrix. Observations and model error are assumed uncorrelated in time (Trémolet 2007).

In order to calculate the new model error covariance matrix \mathbf{Q} , statistics are generated from a dedicated TL399 ensemble experiment with identical initial conditions but using the stochastic model uncertainty representations SKEB and SPPT. The differences between members after 12 hours of model integration give an estimate of the integrated effect of model error over 12 hours; from which statistics appropriate for use in 4D-Var can be calculated. These statistics are used to construct a covariance model similar to that described by Derber and Bouttier (1999) for background error covariances. This method of generating model error covariance statistics provides greater consistency between the approaches to represent model error in 4D-Var and in ensemble forecasts than previous methods. There was a need to scale the estimate of \mathbf{Q} with a factor of 0.2 to ensure that the model error \mathbf{Q} term has an influence in the cost function but does not dominate. The minimisation was run several times with scaling factors between 0 and 1. By looking at the contribution of each term to the cost function, a value of 0.2 was chosen for the scaling factor.

Figure 10b shows the average vertical correlations of posteriori model error estimate (η) for divergence. Its structure disagrees markedly from the structure of the model error covariance \mathbf{Q} predicted by the ensemble using SPPT+SKEB, (panel a); for this reason the deep correlations from the posteriori model error estimate (panel b) were deemed to be suspicious and triggered an investigation.

By looking at the geographical location of large covariances between distant levels, we saw a clear pattern over North America and Europe corresponding to areas with a high number of aircraft observations. This suggests that 4D-Var is misinterpreting aircraft observation error or bias as model error. In order to avoid this interaction with aircraft observations, subsequent experiments restricted the effect of model error to be active only above 40hPa (model levels 1 to 44).

A CY41R2 TCol279 experiment was run to test weak-constraint 4D-Var active above 40 hPa. Forecast skill scores were verified against own analysis and also against Global Positioning System (GPS) Radio Occultation (GPSRO) observations. The verification against own analysis in the northern hemisphere showed a significant reduction in RMS error at 100hPa. GPSRO verification in the stratosphere showed a change in the bias structure throughout the forecast. In the northern hemisphere, bias is slightly improved at all levels, in the tropics it is largely unchanged and in the southern hemisphere the results are mixed

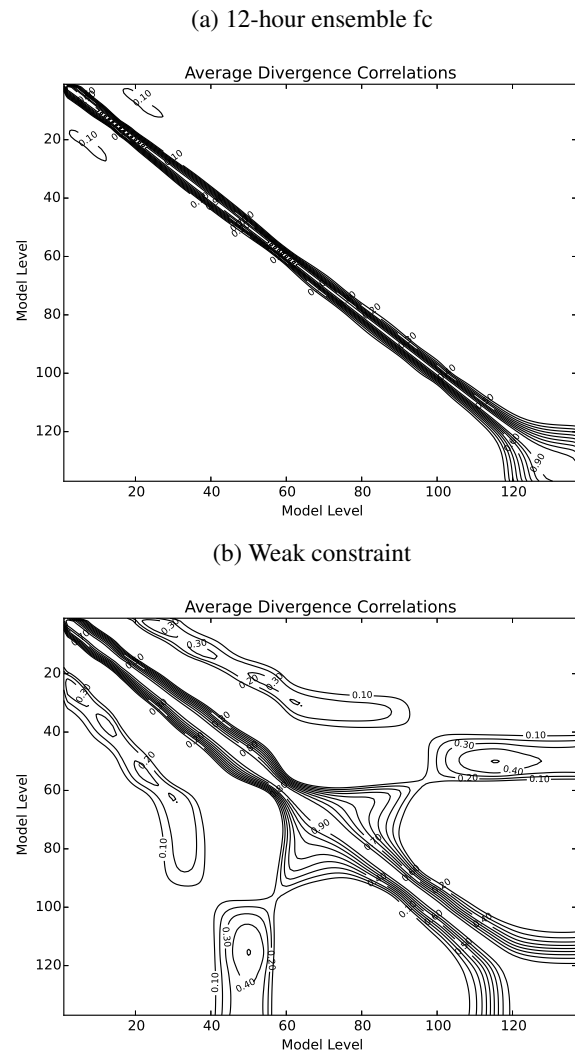


Figure 10. Divergence model error average vertical correlations for (a) model error estimate from 12-hour ensemble forecasts perturbed with SKEB and SPPT and (b) a posteriori weak-constraint 4D-Var model error estimate. Contour interval is 0.1 for both figures.

(but the differences from the control in this region are very small). This configuration of weak-constraint 4D-Var with model error forcing above 40 hPa has been implemented in the operational assimilation in November 2016.

5. Unrepresented uncertainties in the Earth system

Extending and refining the model uncertainty representations in ECMWF's ensembles will be guided by knowledge of the presumed relevant sources of model uncertainties. This section is an inventory of known sources of uncertainties that are not yet explicitly represented in ECMWF's ensembles. Sections 5.1 and 5.2 describe uncertainties in the atmosphere and land surface that could be included in future extensions of the SPP scheme. Then, model uncertainties in the ocean and in the dynamical core are discussed in Sections 5.3 and 5.4, respectively.

5.1. Parametrized atmospheric physical processes

5.1.1. Microphysics and large-scale precipitation

The microphysics perturbations introduced under the initial SPP implementation do not target the large uncertainties associated with phase changes, particularly for precipitation and for the ice phase. Applying SPP perturbations to the uncertain parameters that affect the evaporation rates for rain and snow, and deposition and riming rates for the ice phase would modify the phase changes

and thereby directly perturb humidity and temperature tendencies. Ideally, perturbations should represent uncertainties as close as possible to their sources, for example particle size distributions and ice particle properties in the microphysics, but there is also a need to represent uncertainties in subgrid variance and covariance of cloud, precipitation and humidity as well as uncertainties due to numerical approximations.

5.1.2. Radiation

Atmospheric heating rates due to gaseous absorption in clear-skies are an unrepresented source of uncertainty in SPP while they are poorly represented in SPPT (cf. 3.2). Errors arise due to misrepresentation of the concentrations of atmospheric gases, an imperfect knowledge of the spectroscopy of the atmosphere and the use of discrete quadrature points to represent the longwave (LW) and shortwave (SW) spectra, necessary for computational efficiency. Comparisons to benchmark high-spectral resolution radiative transfer models enable estimates of the heating rate errors in the current configuration

Recent work revealed model sensitivity to the choice of mixing ratio and horizontal distribution of the background aerosol in the troposphere. The latest Copernicus Atmosphere Monitoring Service (CAMS) reanalysis (Flemming *et al.* 2016) shows that perturbations of up to 2–3 times the global constant background aerosol maintain the total aerosol optical thickness within the limits of the constraints provided by satellite observations.

5.1.3. Vertical mixing

In the initial SPP implementation, perturbations were only applied to turbulent mixing within stable boundary layers, where large uncertainties in the degree of mixing are partly responsible for the notorious difficulties that NWP models have in representing such conditions (Sandu *et al.* 2013).

For the convectively unstable boundary layer, the asymptotic mixing length and the entrainment rate at the top of the boundary layer have been identified as uncertain parameters. Above the boundary layer, the addition of perturbations to turbulent mixing is anticipated to address some of the “missing” spread observed from SPP in the free troposphere when compared to SPPT.

5.1.4. Atmospheric composition

Given the importance of atmospheric composition for air-quality forecasts, the IFS was extended by Fleming *et al.* (2009) to include atmospheric chemistry and interactive aerosols. This so-called Composition-IFS (C-IFS) shares the same infrastructure as IFS except that tracers are added together with meteorological variables. Work is underway to account for uncertainties in the modelling of atmospheric tracers in C-IFS. Tracers in C-IFS are subject to surface fluxes and chemical reactions, and they are transported by the IFS wind and diffusion. Uncertainty in atmospheric tracer prediction arises from each of these processes. SPPT perturbs wind, temperature and humidity and thus IFS transport is perturbed either directly or indirectly through geostrophic adjustment. We expect that these perturbations can at least partially represent tracer transport uncertainties. Uncertainty in the chemistry is represented with SPPT by perturbing the tracer chemistry tendency with multiplicative noise. Unlike in standard SPPT, the perturbations to the tracer tendencies are applied also throughout the boundary layer.

Massart and Bonavita (2016) report on numerical experimentation of C-IFS with the EDA. An EDA experiment including SPPT perturbations of two species of the anthropogenic long-lived greenhouse gas family — carbon dioxide (CO₂) and methane (CH₄) — has demonstrated that the ensemble spread for CO₂ and

CH₄ concentrations is too small compared to the ensemble mean RMS error, especially in the boundary layer and free troposphere. Experiments are planned to explore the impact of SPP for C-IFS. It is expected that the SPP scheme will yield greater spread and reliability, in particular due to the perturbations of vertical mixing in the boundary layer and surface fluxes.

5.2. Land surface uncertainties

Ensemble forecasts exhibit a lack of spread in the near-surface variables (e.g. Balsamo *et al.* 2014). Recent work has shown potential for improving ensemble reliability by simulating uncertainty due to the representation of the land-surface (MacLeod *et al.* 2016; Orth *et al.* 2016).

Uncertainties in the energy exchange between atmosphere and land surface are not yet represented. In the IFS, the energy exchange between the lowest atmospheric model level and the land-surface is simulated via a “skin layer”, which takes account of heterogeneity in the land-surface type via the H-TESEL scheme (IFS documentation CY43R1, Part IV, Chapter 8). For each of the H-TESEL tiles, a skin layer conductivity is defined to control the level of thermal conductivity between the uppermost level of the surface model (e.g. top soil or snow level) and the skin layer — the weakest coupling being for e.g. grassland and the strongest for inland water masses. Stochastic perturbations of the H-TESEL skin layer conductivities could account for these uncertainties.

Uncertainties in the humidity fluxes at the land surface have the potential to affect atmospheric predictability during events with considerable land-atmosphere coupling. Some experience in representing soil moisture related uncertainties has been gained by MacLeod *et al.* (2016) using perturbations to hydrological parameters in the van-Genuchten formulation.

Uncertainties in the representation of vegetation can influence atmospheric forecast skill through, e.g. evapotranspiration and albedo errors. At present, vegetation cover is quantified with a climatological leaf area index. Deviations of the actual vegetation state from the climatology are a source of error.

5.3. Ocean uncertainties

Ocean model uncertainty is not represented at all in the ECMWF ensembles. The medium-range, extended-range and seasonal forecasts are coupled to the NEMO ocean model. Sources of model uncertainty encompass a wide range from sea ice rheology (Juricke *et al.* 2014), to unresolved scales and even missing diversity in marine ecosystems (Brankart *et al.* 2015).

The horizontal model resolution of current state-of-the-art ocean models, especially on timescales of seasons to decades and including the model version of NEMO at 1° horizontal resolution, is of the order of 100 km (used in the operational ensemble until November 2016). Mesoscale eddies in the ocean are, however, about an order of magnitude smaller and are therefore not resolved. As a consequence of this, the effects of unresolved eddies on the resolved, large scale circulation need to be parametrized.

Eddies are just one of many sub-grid scale ocean processes that are not explicitly resolved in ocean models. Most of the commonly implemented ocean parametrizations deal with unresolved horizontal and vertical sub-grid scale mixing processes that can vary strongly with time and location. Often the parameters in these schemes are imperfectly constrained due to missing process understanding or unavailable observations.

Juricke *et al.* (2017) studied stochastic perturbations to three mixing parametrizations in NEMO. The perturbations to sub-grid scale mixing parametrizations in the uncoupled experiments lead to an increase in low frequency variability in eddy-active

regions for a variety of variables, even though the perturbations themselves exhibit high frequency variability. The interannual variability for sea surface temperature, sea surface height, integrated heat content and zonally averaged streamfunction increases predominantly in the Southern Ocean and along western boundary currents such as the Kuroshio region. This is in accordance with missing low frequency variability in these regions when compared to observations and reanalysis products. Hence, including high frequency perturbations in parametrizations of horizontal and vertical mixing improves the representation of low frequency variability in the ocean model. This could also be achieved by increased resolution but at greater computational cost. However, the effect of the stochastic perturbations is not sufficient to fully compensate for the effects of the missing eddy variability in the 1° model.

First test experiments have recently been run with the coupled ECMWF seasonal forecasting system based on CY41R1 and horizontal resolution TL255/ 1° NEMO. As in the uncoupled case, the stochastic perturbations increase the ensemble spread (see also Andrejczuk *et al.* 2016), especially in the eddy-active regions of the Southern Ocean and the western boundary currents.

Williams *et al.* (2016) report the reduction of biases in a non-eddy-permitting ocean model (FAMOUS) due to the introduction of stochastic perturbations to the temperature in the ocean and improved variability of the strength of the thermohaline circulation. The noise for the perturbations was informed by diagnostics of eddy statistics in an eddy-permitting ocean model (HiGEM).

Another source of model uncertainty that deserves attention is the coupling method for the atmosphere-ocean-(sea-ice) system.

5.4. Dynamical core

Wedi *et al.* (2015) have discussed recent advances in the dynamical core at ECMWF and plans for its further development with the main focus being on the unperturbed model. Significant advances have been made in better resolving meso-scale circulations with the IFS and other global models. However, it is expected that global and regional NWP models will never have sufficient spatial resolution to resolve all relevant motions in the atmosphere. Resolving the motions in the stable boundary layer may require spatial resolutions of the order of 1 m as indicated by studies with large eddy simulation (LES) models (Sullivan *et al.* 2016). Regional and global models will not reach such resolutions in the foreseeable future and one should expect forecast errors that originate from partly resolved or un-resolved aspects of the flow. SKEB can be seen as an attempt to represent such errors. However, there is no strong theoretical link between SKEB and the numerical approximations involved in the dynamical core.

Ultimately, representing model error associated with a dynamical core could become part of its formulation. Hodyss *et al.* (2014) describe a stochastic horizontal advection scheme in the US Navy Global Environmental Model primitive equation model (NAVGEN) that introduces spread in tropical cyclone position. The method results in Brownian motion of the cyclone position and the noise can be informed by past tropical cyclone track errors. It is too early to judge whether such ideas will be of benefit in operational ensemble forecasting. The variational framework to derive stochastic partial differential equations for fluid dynamics introduced by Holm (2015) could be seen as a step towards establishing a theoretical basis for developing stochastic dynamical cores. In Section 6.4, plans for introducing stochasticity in the IFS dynamical core will be discussed.

Projections of numerical weather forecasting and climate science envision global cloud-permitting simulations at ≈ 1 km horizontal grid intervals in the next decade (Palmer 2014;

Bauer *et al.* 2015). With deep convection being simulated rather than parametrized, NWP moves towards the large-eddy simulation (LES) regime, characterised by substantial flow-related uncertainty in turbulent small scales. The latter may be correctly captured as local fluctuations of physically realisable model solutions or, to the contrary, may represent a new source of numerical model errors. Consequently, it may be of interest to quantify the solution adequacy in small scales, convergence behaviour of the model, and the relative role of small scales explicitly or implicitly represented in local and global predictions.

Natural stochasticity in small scales is closely linked to shear-gravitational instabilities and turbulence. Therefore, its strength can be measured with the turbulent kinetic energy (TKE) underlying LES subgrid-scale models (Clark 1979). Alternatively, local-flow physical realisability can be assessed from the flow Jacobian (Cullen *et al.* 2000), or the related “Lipschitz” number (Cossette *et al.* 2014). Technically, all these measures represent norms of the 3×3 Jacobi matrices of local flow derivatives; namely, the strain rate tensor for TKE, the Lagrangian displacement gradient tensor for the flow Jacobian, or the velocity gradient tensor for the Lipschitz number. Although not entirely exclusive, the three norms carry distinct information and complement each other.

Regardless of the potential utility of the flow related tensor norms, evaluating the required tensor entries is not an easily affordable computational task for the IFS. Generally, semi-implicit semi-Lagrangian (SISL) spectral transform based NWP models are not equipped to enable efficient local calculations of these tensor quantities at high resolution where they may matter, due to the communication overhead and the non-locality associated with global spectral transformations. However, this is a fairly straightforward task with the technologies applied in the autonomous, all-scale global finite-volume module (FVM) recently developed under the auspices of the ERC-funded project PantaRhei hosted at ECMWF (Smolarkiewicz *et al.* 2015, 2016). The FVM infrastructure facilitates evaluation of local measures of IFS solution uncertainty and model errors using methods familiar from cloud-resolving research models, applied at resolutions beyond what is possible in global NWP.

For illustration, Fig. 11a displays the magnitude of instantaneous subgrid-scale (sgs) velocity fluctuations derived from the Smagorinsky eddy viscosity (cf. Clark 1979) diagnostically evaluated in an inviscid FVM simulation of the moist precipitating Held-Suarez climate with ~ 60 km resolution (Kurowski *et al.* 2015); for reference the vertical component of velocity is shown in the bottom panel (Fig. 11b). Not surprisingly, local “uncertainties” as measured by sgs fluctuations are correlated with collapsing fronts of baroclinic eddies in mid latitudes and with convective motions in the tropics. While the magnitude of such local uncertainties can be substantial, it remains confined to disjoint areas with small radii of influence.

The three flow related tensor norms discussed earlier derive from physical arguments on flow topology, conservativeness and integrability of the underlying equations (Cossette *et al.* 2014). These ideas closely link to the derivation of stochastic partial differential equations that preserve fundamental invariants of the flow (Holm 2015). In particular, a promising and practical approach to maintain physical realisability is to exploit the uncertainty associated with determining the semi-Lagrangian departure point (cf. Sec. 6.4). Like semi-Lagrangian schemes, the FVM transport algorithm integrates the governing partial differential equations (PDEs) over an element of the 4D time-space continuum, effectively viewing the governing equations as physical constraints of arbitrary path or volume integrals of differential forms in the 4D space (cf. Smolarkiewicz and Pudykiewicz 1992). As a result, FVM naturally provides measures

of truncation and residual, spatio-temporal error patterns that are congruent with the underlying flow statistics. Hence, it naturally satisfies for example the “proportionality of scales” (Tsyrlunikov and Gayfulin 2016), i.e. larger spatial perturbation scales that persist in time longer than smaller spatial scales, and more generally preserving the non-separability of spatio-temporal covariances, typical of geophysical data (Gneiting 2002). It is planned to exploit these features as a means for driving future stochastic parametrizations towards improving sharpness and reliability of ensemble forecasts.

6. Future directions

While the previous Section listed so far unrepresented uncertainties in the Earth system, we will now turn to future directions for work aimed at improving the representation of model uncertainties. The topics covered now will be priorities where resources will be allocated at ECMWF in the coming years. Section 6.1 describes new efforts in diagnostics and verification. Then, ideas for improving SPPT and SPP are summarized in Sections 6.2 and 6.3, respectively. Section 6.4 proposes work on introducing stochastic perturbations in the semi-Lagrangian advection to represent uncertainties associated with the dynamics. Section 6.5 makes suggestions for potential improvements to the random fields used by the stochastic model uncertainty representations.

6.1. Diagnostics and verification

Examples of routine diagnostics for ensemble forecasts and model climate appeared in Section 4. These diagnostics will play a key role in improving the stochastic model uncertainty representations. Here, additional efforts are described that are thought to provide further insight and that have not received much attention to date.

Experiments with the EDA are a useful framework to evaluate model uncertainty representations at early lead times using observations following Rodwell *et al.* (2016). An advantage of the EDA configuration at ECMWF is that the spread is a function of the perturbations to the observations and the representation of model uncertainties only, there is no additional artificial inflation. Observation space diagnostics can be applied also to short-range ensemble forecasts following the work of Yamaguchi *et al.* (2016). However, more technical work would be required to enable this as a routine diagnostic at ECMWF.

6.1.1. Process-oriented evaluation of ensemble forecasts

Stochastic perturbations of physical processes can have a quite indirect impact on the forecast. For instance, SPPT or SPP can perturb heating rates in a convective region; through geostrophic adjustment, this leads to wind perturbations later in the forecast within a distance that scales with the Rossby radius of deformation. To better constrain the stochastic perturbations, it may help to validate the changes in precipitation that are closely connected to the directly perturbed convection process. Along the same lines, if SPP introduces perturbations to the cloud-radiation interaction, the radiative fluxes at the top of the atmosphere will be immediately affected offering an avenue to better constrain the perturbations via verification of the fluxes.

Verification procedures developed at ECMWF could aid in the process-oriented evaluation of ensemble forecasts. These are the verification of radiative fluxes based on satellite and surface station data, and the verification of precipitation based on radar and raingauge data.

Daily averages of the reflected shortwave flux at the top of the atmosphere (TOA) and the downward shortwave flux at the surface from the Climate Monitoring Satellite Application Facility

(CM SAF) are used to evaluate forecasts of these fluxes in the operational model version, and in experimental suites. The metrics used are common error measures such as RMSE and anomaly correlation for individual forecasts, and CRPS and spread-error for ensemble forecasts. Since the evaluation is done for gridded fields, maps of the above quantities can be generated, highlighting areas of over- and under-dispersion, the latter most notably in the tropics along the Inter-Tropical Convergence Zone (ITCZ) (Haident *et al.* 2015). In addition to the verification against satellite data, downward radiation fluxes (both shortwave and longwave) at the surface are evaluated against Baseline Surface Radiation Network (BSRN) observations. While there is limited coverage (only about 50 active stations globally, some of them reporting with considerable delay), this high-quality dataset is suitable for assessing the reliability of the satellite derived downward radiation products.

Routine verification of precipitation at ECMWF is mostly based on SYNOP observations. However, recently this has been enhanced by high-density observations from member states, providing improved coverage for large parts of Europe (Haident and Duffy 2016). The increased observation density reduces the noise in the computed scores and allows up-scaling of the point observations onto a regular grid which in turn opens up the possibility of evaluating skill as a function of spatial scale, for example by using the Fractions Skill Score (FSS). This has been tested using Next-Generation Radar (NEXRAD) Stage IV precipitation observations (radar+raingauge combined) from the US (Rodwell *et al.* 2015).

The procedures described above have been demonstrated to provide useful results but are not yet operational in the sense of fully-automated verification suites. Also, more development work is needed to provide a broader range of scores and metrics specifically relevant for model uncertainty.

Ollinaho *et al.* (2017) identified differences in the impact of SPP and SPPT in terms of CRPS of 2-metre temperature verified against SYNOPs. Additional developments should also include evaluation of skin temperature using satellite data. This may help in the improvement of forecasts of 2-metre temperature, which are diagnostically derived from lowest model level and skin temperature forecasts, and which currently exhibit substantial situation-dependent biases.

The effect of observation errors (measurement errors and representativeness errors) on the verification is not yet being accounted for in routine verification of the ensemble forecasts. Methods which have been proposed in the literature (Saetra *et al.* 2004; Yamaguchi *et al.* 2016) should be tested and included in the future process-oriented ensemble verification in order to facilitate interpretation of the results.

Another avenue that may help to better constrain the random fields used in the stochastic schemes is to develop and apply multi-variate verification. For instance, one could explore whether the horizontal correlation scales in the random fields have an influence on the predicted ensemble covariances for precipitation or radiative fluxes. If this is the case, a verification of the ensemble covariances could help estimate correlation scales used in the random fields.

6.1.2. Use of an SI traceable observation network to characterise uncertainty

The Global Climate Observing System (GCOS) Reference Upper Air Network (GRUAN) aims to provide highly accurate and well characterised temperature and humidity radiosonde observations, whose calibration is traceable to Système International (SI) standards. ECMWF is part of the 18 partner Horizon 2020 GAIA-CLIM project, primarily contributing to a work package

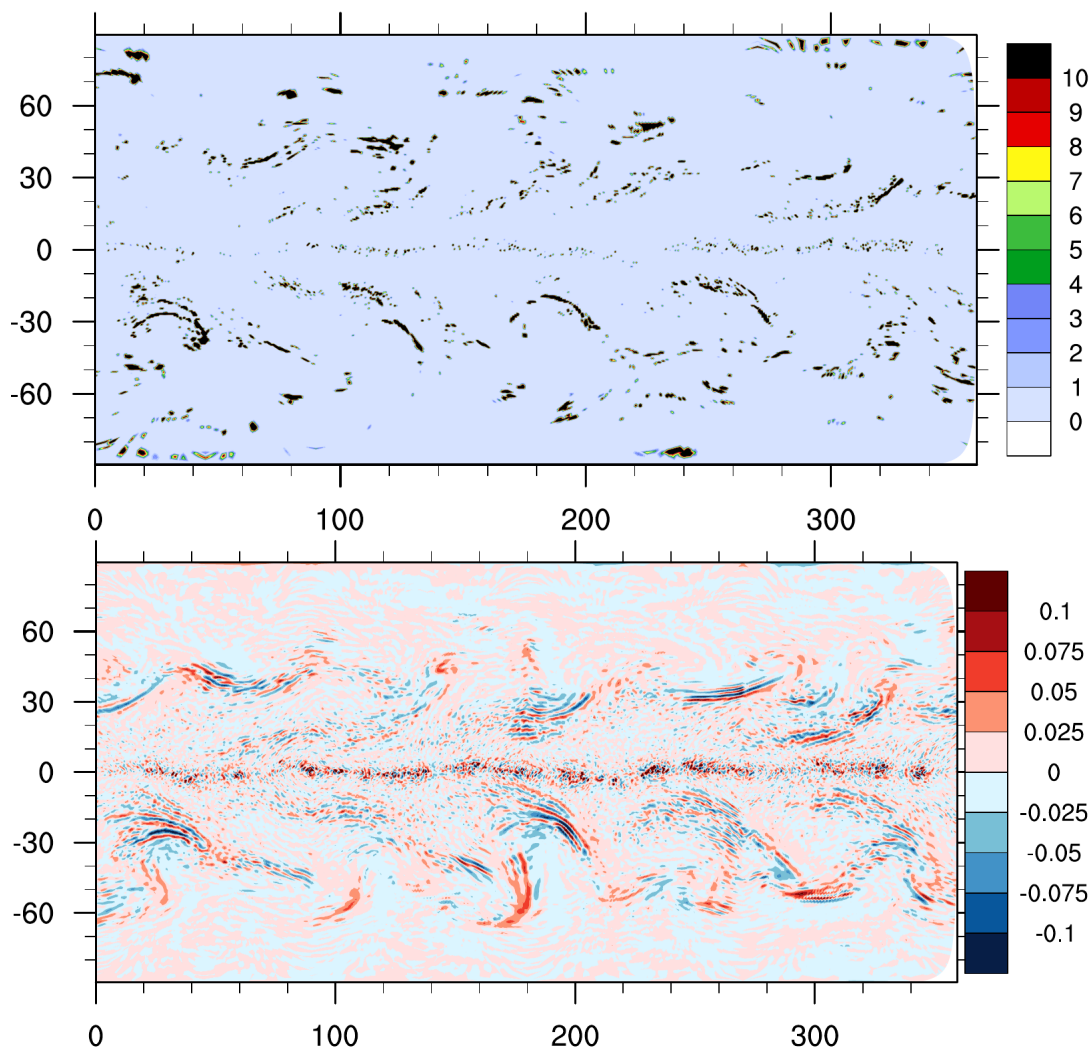


Figure 11. Magnitude of instantaneous subgrid-scale (sgs) velocity fluctuations (top panel) and the vertical component of the velocity vector (bottom panel) after 3 months of a FVM N160 simulation of a moist precipitating Held-Suarez climate. Both panels show the result in the xy plane at 1.3 km above sea level; the colour bar show values in m s^{-1} .

led by the UK Met Office. This aims (amongst other things) to use these reference observations to provide an absolute calibration for the ECMWF and UK Met Office NWP background fields, which should aid the interpretation of biases found in the calibration/validation of new satellite instruments. More generally, this should lead to an improved characterisation of uncertainty across the entire Global Observing System, which is relevant to model error characterisation through the link to using data assimilation to estimate model error. One example is the evaluation of the reliability of the EDA using observation space diagnostics (cf. Sec. 4.4). The usefulness of the diagnostic depends on how well observation error statistics are known (Rodwell *et al.* 2016). Another example is the weak constraint formulation described in Section 4.5, which relies on the availability of so called anchor observations. Some of the limitations of weak constraint to estimate model error arise from insufficient knowledge of systematic errors in observations (e.g. aircraft). Therefore although the goal of GAIA-CLIM is to characterise observation and background errors, this knowledge should contribute indirectly to improved characterisation of model error. The GAIA-CLIM project ends in February 2018, and will demonstrate how reference in-situ observations can be exploited to improve satellite instrument calibration and validation. The project will also identify geographical and temporal gaps in these reference observations, making the case both for the maintenance and the improvement of such networks.

6.1.3. Upscale perturbation growth and spectral energy cascade

Since Lorenz (1969) examined predictability in a fluid with many scales of motion, it is known that the slope of the kinetic energy spectrum has implications for the theoretical predictability limit. In a fluid with a k^{-3} spectrum, the forecast range when predictability is lost could be extended in principle to infinity by reducing the magnitude of initial error further and further while in a system with a $k^{-5/3}$ spectrum there is a finite predictability limit regardless how small the initial error is. Whether errors in the small scales that grow upscale and in amplitude have practical implications for predictability given present-day initial errors is a matter of active scientific debate (Durrant and Gingrich 2014; Selz and Craig 2015a,b; Buizza and Leutbecher 2015).

Perturbation growth experiments similar to the study of Selz and Craig (2015b) permit understanding of the interaction between model uncertainty representations that trigger small-scale perturbations and the dynamics of the model. It is planned to conduct such perturbation growth experiments with the IFS using current and future model uncertainty representations.

Malardel and Wedi (2016) diagnosed the transfer of kinetic energy (KE) and available potential energy (APE) in the IFS following the method proposed by Augier and Lindborg (2013). The planetary and synoptic scales constitute an energetically closed system with very little energy transfer towards or from the mesoscale. At the mesoscale, the KE cascade in the IFS is

downscale, the APE cascade is upscale. Both terms are small in amplitude and partially balance each other for parametrized convection. In contrast, for permitted convection, both terms are larger in amplitude and balance each other more. The energy transfer at different scales is not directly observable. The connection to observations are atmospheric spectra of wind or temperature in the free atmosphere (e.g. [Nastrom and Gage 1985](#)). These are limited too as they are based on one-dimensional observations along aircraft tracks in the upper troposphere and lower stratosphere in mid-latitudes.

Sensitivity experiments showed that the parametrizations contribute to the gain and dissipation of energy directly at all scales. They also showed that the parametrizations strongly influence the energy transfers across scales in particular through the control of the surface forcings (orography, friction) in the boundary layer and associated initiation of divergent and interacting divergent and rotational modes. At the mesoscale at wavenumbers larger than about 20, divergent motions and gravity waves are the main contributors of the upscale APE cascade and downscale KE cascade. Modifications in the physics, which increase the activity of the model in terms of divergent modes, modify substantially the non-linear energy transfers at the mesoscale. The overall impact of the physical parametrization leads to a kinetic energy spectrum in IFS that is closer to a k^{-3} than a $k^{-5/3}$ spectral slope. Further work will start to evaluate the potential of these diagnostics to understand the current limitations of NWP models in their representation of the energy cascades across scales with a view to optimize the stochastic representation of model errors.

6.2. Improvements to SPPT

6.2.1. Clear-sky radiation

Analyses of model temperature tendencies from experiments with SPPT and SPP exposed a relative weakness in the SPPT formulation with respect to clear-skies radiation (Sec. 3.2). SPPT applies perturbations, which are proportional to the net tendencies from the physics parametrizations. Hence, small (large) net physics tendencies are taken to imply small (large) uncertainty. In the day-time, outside convectively-active regions, temperature tendencies due to physics tend to be small due to the balance between incoming solar radiation and out-going thermal radiation. Meanwhile, at night-time, thermal radiation in clear-skies regions dominates yielding relatively large (negative) temperature tendencies. Consequently, SPPT attributes relatively large uncertainty to the night-time cooling in clear-sky conditions, which appears to be unrelated to any real source of model error.

Work is underway to explore the impact of applying SPPT perturbations to total temperature tendencies that either *exclude* the contribution due to clear-sky radiative forcing or assign it a considerably reduced level of uncertainty. There is good physical justification for this approach, since the greatest uncertainty in the radiative forcing integration is due to the treatment of clouds (see e.g. [Morcrette et al. 2008a](#)), while the contribution due to clear-skies is understood to be relatively accurate unless it is due to water vapour (or other trace gas) errors. Furthermore, SPPT already recognises crudely this weakness in its formulation: the tapering function that reduces uncertainty perturbations to zero in the stratosphere was imposed for the very purpose of excluding relatively large tendencies that are dominated by relatively accurate contributions from clear-skies radiation. Experiments will determine the impact on ensemble performance from revising SPPT to remove or reduce contributions due to clear-skies radiation, thereby enabling the removal of the stratospheric tapering. [Sanchez et al. \(2016\)](#) report tests of removing the clear-sky radiation tendencies from their version of SPPT (referred to

as SPT by them). They found a substantial reduction of ensemble spread in the tropics and extra-tropics (their Fig. 9). We expect that these findings might not be directly transferrable as the MetOffice SPT scheme differs in its configuration. For instance, it uses random patterns with a vertical structure.

6.2.2. Supersaturation treatment

Work is in progress to understand the underlying causes for the drying of the atmosphere caused by SPPT without the global fix. There are three elements in SPPT that could be contributing to its enhancement of the dry bias:

1. the supersaturation limiter (cf. Sec. 2.1.1)
2. the boundary layer tapering (cf. Sec. 2.1.1),
3. the imbalance between perturbed tendencies and unperturbed surface fluxes.

Work has started to explore whether SPPT could be improved by deactivating the supersaturation limiter. Moreover, the deactivation of the boundary layer tapering is tested. While the combination of these two modifications shows a promising change for the vertically integrated moisture, a strong drying immediately above the boundary layer emerges.

By removing the boundary layer tapering, the imbalance between the perturbed tendencies and the surface fluxes (point 3 above) is perhaps exacerbated. In the unperturbed model, conservation of water in each column is achieved by the parametrized physical processes. When SPPT acts, an imbalance arises since perturbations are applied to the physics tendencies without consistent perturbations being applied to the surface fluxes. It seems plausible that this imbalance will be most apparent when SPPT perturbations are applied near the surface. Future efforts will be directed at finding a means to address this inconsistency. It is expected that humidity conservation should be achievable if a linear scaling of the surface fluxes consistent with the random number used for the upper air perturbation could be implemented and if all tapering of the perturbations could be deactivated.

6.3. Implementation of new SPP scheme

The SPP scheme has potential as a method to represent model uncertainties as it is able to generate considerable additional ensemble spread at all lead times from the EDA to the extended-range ensemble forecasts (cf. Sec. 4). In general, versions of SPP tested so far are less active than SPPT in the extratropics but generate significantly more spread in the MJO principal components than SPPT. At early lead times, SPP appears to introduce even too much spread in the large-scale MJO state and this may cause a potential degradation of extended-range predictions. It is planned to conduct a range of sensitivity experiments to better understand how the configuration of SPP affects MJO predictions.

Initial results at TL399 and TCo255 resolutions also indicate that SPP is more effective in generating spread in tropical cyclone intensity than SPPT. Work is planned to evaluate SPP at the operational resolution of TCo639 and to examine how the reliability of tropical cyclone intensity predictions is affected. Gaining a deeper understanding will require a number of sensitivity experiments with different configurations of SPP.

Further understanding of how SPP modifies forecasts could be obtained from analysing the tendency perturbations in case studies and examining the sensitivity of the tendency perturbations to the SPP configuration. This work could be linked to cases of, for instance, MJO events, tropical cyclones, meso-scale convective systems or extra-tropical wind storms.

Alongside the work to improve on the current configuration of SPP, work is planned to extend the scope of SPP to also cover the unrepresented sources of model uncertainties summarized in Section 5.1.

It may also be necessary to consider a combination of SPP and SPPT as a representation of model uncertainty to generate sufficient spread before SPP has included sufficient sources of uncertainty. However, a combination of both schemes would still suffer from the physical inconsistencies present in SPPT.

6.4. Semi-Lagrangian advection schemes

There are two sources of error in a semi-Lagrangian advection scheme: (i) errors associated with poor estimation of the departure points (DP); and (ii) errors associated with the interpolation method used. The former would give rise to phase errors while interpolation “adds” numerical diffusion or unwanted oscillations resulting in underestimation or overestimation of the amplitude of the advected waves.

The interpolation error depends on the order of the method used and the model mesh resolution and it is of purely numerical origin. Errors associated with the departure points are more complex. The reason for this is that the accurate estimation of the DP requires an accurate numerical scheme for solving the DP kinematic equation and an accurate “prescribed” wind field. The latter is subject to model errors arising from any model process contributing to its prediction – inaccuracies due to the transport scheme itself, as well as from the parametrized physics processes. Uncertainty in these processes implies a degree of inherent uncertainty in the DP estimation which would be desirable to represent in ensembles.

In the IFS, the accuracy of the DP estimation is sensitive to the complexity of the local flow. Convergence of the iterative DP algorithm is slowest for regions of strong shear and large vorticity where the norm of the difference between successive DP estimates shrinks slowly (Diamantakis and Magnusson 2016). Hence, as such differences are indicative of the magnitude of model error, flow-dependent uncertainty can be incorporated into the semi-Lagrangian advection scheme by adding random perturbations to the estimated position of the DP scaled by the difference from two different iterations. Appendix B describes the proposed method in more detail. Work is planned to develop the methodology for stochastic DP perturbations in the IFS.

DP perturbations have previously been explored with the UK Met Office forecast model. Bowler *et al.* (2008a) demonstrated that by using a multianalysis (2-member) ensemble, the RMSE at early lead times (up to 3 days) could be significantly improved compared to a single high-resolution forecast. Of the perturbation methods explored, the greatest skill improvement was found from including a member with uncorrelated random perturbations to the DP estimation. Though their approach was considerably different to that proposed here, their work confirms the sensitivity of forecast skill to the DP estimation; and hence, further highlights the potential for representing model uncertainty associated with DP estimation.

6.5. Random fields

Most stochastic representations of model uncertainty require noise that is correlated in space and time. At present, time series of (pseudo-)random fields (“patterns”) are generated in the IFS using AR(1) processes in spectral space. Spectral transforms provide the random fields in grid point space. This results in homogeneous isotropic correlations on the sphere with a correlation function on the sphere that approximate a Gaussian with a specified correlation length scale (Weaver and Courtier 2001). Sensitivity experiments such as those described by Ollinaho *et al.* (2017) together with advances in diagnostics and ensemble verification

may help to better constrain the space-time correlations of the random fields.

Work on the random field generators is considered important both for reasons of increasing computational efficiency as well as the need for exploring wider classes of space-time correlation models. Schlather (2012) and Kroese and Botev (2013) give recent overviews of fast algorithms to simulate stationary or isotropic random fields. The efficient generation of non-stationary Gaussian random fields at high spatial resolution is discussed by Kleiber (2016).

One potential avenue to reduce the cost of generating random fields is to exploit multi-grid approaches. Both SPPT as well as SPP specify random fields with a high correlation at distances of at least 5 times the model grid spacing Δx . A multi-scale approach that generates the fields first on a coarse grid, say with a resolution of $5\Delta x$, and then uses horizontal interpolation to obtain the random field on the model grid would appear viable and may reduce the computational cost of pattern generation significantly. Moreover, there is scope to further reduce cost by evolving the random fields with a time step that is longer than the model time step as the decorrelation time is much longer than a typical model time step.

Previous studies in other fields have made the link between random processes in space and time and solutions of stochastic partial differential equations. The potential advantages of viewing random fields as approximate solutions of stochastic partial differential equations is that a wider class of space-time correlations can be introduced and that more efficient solution methods may emerge. This approach could also deal with complex geometries such as those encountered in ocean models. Lindgren *et al.* (2011) discuss how the solution of the linear stochastic PDE for advection-diffusion can be used to generate a large class of non-separable non-stationary models. Approaches based on solutions to stochastic PDEs also start to be explored in the field of post-processing ensemble forecasts (Sigrist *et al.* 2015). Even with the current spectral random field generator, it may be possible to explore some non-separable covariance models, e.g. ones that obey the “proportionality of scales” advocated by Tsyrlunikov and Gayfulin (2016).

Development of the infrastructure for the future dynamical core will facilitate efficient solution algorithms for implementing random fields based on solutions of stochastic partial differential equations. The Atlas library (Deconinck *et al.* 2016) will provide the framework for local derivatives and more generally dynamically evolving local differences on coarse and fine resolution representations of the same parameter. This could be exploited for coarse-graining of tendencies as well (Shutts 2013).

7. Discussion

7.1. Initial uncertainties

In Section 4, results were based on a fixed representation of initial uncertainties corresponding to the operational configuration of the ECMWF medium-range/extended-range ensemble. The initial perturbations are based on perturbations constructed from the EDA and singular vectors (Buizza *et al.* 2008; Leutbecher and Lang 2014). The perceived optimal model uncertainty representation will under(over)-estimate the effect of actual model errors if the initial perturbations sample a wider (tighter) distribution than the distribution of actual initial errors. Following an approach that uses the same representation of model uncertainties at all lead times and evaluating its impact from analyses to extended-range will help to avoid gross compensating errors between the ensemble variance generated by initial perturbations and the ensemble variance generated by the model uncertainty representation.

However, future changes to the representation of initial uncertainties are likely to influence the development of the model uncertainty representation. For instance the MJO prediction with an ensemble using initial perturbations only is quite underdispersive (Sec. 4.2). The MJO spread may also be sensitive to the ocean initial perturbations, which are currently provided by uncoupled ocean assimilations. The effectiveness of ocean initial perturbations may change when the initial perturbations are produced with a coupled data assimilation system. This development for the initial conditions may then influence the future development of the model uncertainty representation.

Work on the representation of uncertainties in data assimilation can improve the seamlessness of the prediction from initial state to short-range and medium-range forecasts in three ways: firstly, by developing a consistent representation of model uncertainties for the EDA and the ensemble forecasts. This aspect could be further strengthened by initialising the random fields used in the model error representation of the ensemble forecasts from their state in the EDA at initial time. Secondly, by starting the ensemble forecasts directly from an EDA instead of re-centring the initial conditions on the unperturbed high-resolution analysis (Lang *et al.* 2015). Thirdly, improvements in the description of the sources of uncertainty in the EDA may lead to a reduced need for singular vector initial perturbations, this could pave the way for creating an ensemble prediction system that starts directly from the EDA without a need for further additional or inflated initial perturbations and yet still exhibiting a good calibration of the ensemble spread.

7.2. Alternatives for representing model uncertainties

Across the research community and NWP centres additional methods to represent model uncertainties are pursued. A short overview follows, which for the sake of brevity cannot be comprehensive. Broadly speaking, the alternatives fall into four categories: (i) multi-model and multi-parametrization, (ii) schemes that account for the stochasticity of the subgrid-scale forcing for a given resolved-scale state, (iii) flow-independent additive perturbations or multiplicative inflation informed by data assimilation, (iv) ensemble postprocessing/calibration techniques.

(i) Multi-model and multi-parametrization A pragmatic way of accounting for model uncertainty is to construct ensembles with members that use different numerical models (Iversen *et al.* 2011; Weisheimer *et al.* 2011) or different parametrizations for physical processes. The global ensemble forecasts at Environment Canada, Météo-France and the regional US Air-force Weather Agency ensemble employ such a “multi-physics” approach (Berner *et al.* 2011; Charron *et al.* 2010; Descamps *et al.* 2015). Using different models or different physics parametrizations can increase ensemble spread but the sampling of uncertainty is discrete and some of the spread increase is simply due to different model biases, which is not desirable as different biases can lead to a clustering of members by model or by parametrization scheme. For instance, in the Canadian ensemble prediction system, some members use the Kain-Fritsch convection scheme and others the Kuo convection scheme. For some variables, e.g. 250 hPa temperature in the tropics this leads to a noticeable bi-modality of the ensemble distribution with all members using the Kuo scheme being colder than the members using the Kain-Fritsch scheme (Separovic *et al.* 2016). Related to multi-parametrization approaches are parameter perturbations that are fixed in time and space. The limited-area model ensemble of the German Weather Service (DWD) follows this approach to represent model uncertainties (Theis *et al.* 2016; Gebhardt *et al.* 2011). Similar to a multi-physics ensemble, members are not exchangeable and

are likely to exhibit different biases. Some members may be on average considerably less skillful than others.

(ii) Schemes accounting for subgrid-scale stochasticity Plant and Craig (2008) introduced a stochastic scheme for deep convection based on the Kain-Fritsch scheme. Separovic *et al.* (2016) reported tests that extended this work to a stochastic version of the Tiedtke-Bechtold scheme, which is the convection scheme used in the IFS (Bechtold *et al.* 2008). The results indicated that the scheme introduces white noise at the grid-scale that does not show significant upscale growth compared to perturbations from SPPT.

Kober and Craig (2016) developed stochastic perturbations of the tendencies due to vertical mixing in the boundary layer to account for the stochasticity of large eddies therein. Their scheme is targeted at initiating convection in convection-permitting ensembles.

If new evidence emerges that perturbations from such schemes can have upscale effects that are still relevant in the presence of realistic initial perturbations, it could be explored whether the perturbation methodologies of these schemes could be incorporated into SPP. An alternative avenue to address subgrid-scale stochasticity of convection could be in the iSPPT framework using a kind of square-root scaling proposed by Shutts and Pallarès (2014).

(iii) Flow-independent additive perturbations and multiplicative inflation Driven by the needs of data assimilation, numerous methods for additive or multiplicative inflation have emerged. Météo-France uses a multiplicative inflation in their EDA system to account for lack of spread (Raynaud *et al.* 2012). Cardinali *et al.* (2014) have tested an additive model error representation based on perturbations sampling a Gaussian distribution with a covariance given by a scaled background error covariance matrix. Piccolo and Cullen (2016) explore using a random sample from weak-constraint 4D-Var analysis increments as representation of model uncertainties.

While these methods are efficient in generating spread, there appears to be no obvious way to make them dependent on the resolved-scale state or to ensure physical consistency, e.g. energy and moisture budget of fluxes and state perturbations.

(iv) Ensemble postprocessing/calibration Reliability could also be obtained through post-processing alone (Gneiting 2014). Stochastic representations of model uncertainty could be viewed as in-model-ensemble-calibration. It is an emerging area of research to examine the added benefits of a stochastic representation of model uncertainty after ensemble outputs have been calibrated. Berner *et al.* (2015) examine this question for ensemble forecasts with the WRF model.

One would expect that perturbations in the model have the potential to generate flow-dependent physically consistent spatial fields and multi-variate covariances. This will make it easier for post-processing techniques as they will have to add less flow-independent variance and can then exploit the multi-variate dependency structure of the ensembles using e.g. ensemble copula coupling (Scheffzik *et al.* 2013).

7.3. Estimating model uncertainties

Initial work on using the stochastic representation of model uncertainty to estimate a model error covariance matrix for weak-constraint 4D-Var was reported in Sec. 4.5. The posteriori model error estimates obtained with weak-constraint 4D-Var may provide useful information to further refine stochastic representations of model uncertainty. Using idealized systems,

work is on-going to establish under what conditions weak-constraint 4D-Var is able to estimate the true statistics of model error that is prescribed and therefore known in the idealized experiments (Todling 2015a,b; Bowler 2017; Bishop 2016).

A way to estimate characteristics of model error due to parametrized moist convection is to compare integrations of cloud-resolving models with models with parametrized convection. Coarse-graining of tendencies has been proposed to obtain proxies for model error (Shutts and Palmer 2007). Global cloud-resolving simulations are an emerging capability but experimentation will be strongly limited by available computational resources for many years to come.

Super-parametrization (SP) could also provide opportunities to gain insights into the effects of model error associated with parametrized deep convection in the coming years and at a significantly lower computational cost than global cloud-resolving simulations. SP includes effects of moist convection in global models through cloud resolving model arrays embedded within the global model (Grabowski and Smolarkiewicz 1999; Khairoutdinov and Randall 2003; Randall *et al.* 2003). A SP scheme has been implemented and tested in the IFS over the last year. The SP component of the SP-IFS is a modified version of a three-dimensional cloud-resolving model described in detail by Khairoutdinov and Randall (2003).

7.4. Computational efficiency gains

Hierarchies of complexity are a promising idea in the context of scalability, where ensembles clearly offer a high level of parallelisation that can naturally be exploited on many-core supercomputers. Already, coarsening of the radiation calculations in time and in space has been used to reduce the cost of the ensemble (Morcrette *et al.* 2008b; Hogan and Bozzo 2015). An interesting question to investigate is if the knowledge of model error or its existence could be exploited to improve either the scalability and/or the time-to-solution of individual ensemble members. Aspects that are explored in this context are bounds on numerical accuracy for a given process, which would allow lower precision to be used on suitable future hardware; considering some processes to be executed concurrently (hence modifying their coupling to each other and exploring the associated uncertainty); and/or executing processes on a different physical grid. For instance, stochastic approaches for simulating the momentum transport by gravity waves have been proposed by Eckermann (2011) and Lott *et al.* (2012). Single precision for the entire medium-range ensemble was explored by Váňa *et al.* (2017). In the future, there may also be opportunities to harness more computing resources through the use of imprecise hardware (Palmer 2014).

In view of higher resolution ensembles in the future, work will be required to address data mining and data reduction, together with I/O strategies on future high-performance computing platforms.

8. Summary

Advancing the representation of model uncertainties is a challenging task due to the difficulty in characterising model errors and since forecast errors are always due to model error and initial condition error. While simple schemes, such as SPPT, continue to be highly effective, some of their shortcomings, such as lack of conservation and assumed uncertainty of clear-sky radiative cooling, have received more attention. Current and future work will explore ways to at least partially address identified deficiencies. Future progress may also happen through better understanding of these simple schemes via sensitivity experiments, i.e. repeating tests performed with

different models. It seems premature though to initiate a model uncertainty intercomparison project. Considering standard upper air probabilistic skill metrics, the currently operational schemes like SPPT and to a lesser extent SKEB are still providing a challenging target to improve on.

Work is progressing to develop uncertainty representations for the model physics close to the assumed sources of model error with the advantage of obtaining perturbations that maintain the physical consistency imposed by the model's parametrizations. Such process-oriented representations of uncertainty, as in the new SPP scheme, are desirable. However, they will also pose larger challenges during initial development and testing of future model upgrades due to this higher complexity with respect to simpler schemes like SPPT. This may not be a negative aspect though as there is a growing interest in the academic community to work on stochastic representations of model uncertainties. There are many interesting research questions to tackle from understanding how stochastic perturbations to the tendencies grow in the forecast to identifying ways to better constrain their parameters. It is expected that the representation of model uncertainties remains an active area of research for the foreseeable future and that ample opportunities for collaboration between the academic community and operational NWP centres will arise in this context.

In the coming years, ECMWF plans to extend the scope of stochastic representations of model uncertainty. It is proposed to consider stochastic terms in other parts of the forecast model, for instance the dynamical core, the ocean, the land surface and the sea ice components. There are some promising initial efforts within the research community on which the developments at ECMWF can build.

Quantifying uncertainty is a dominant driver of computational cost at ECMWF: about 56% of the computational resources for all operational suites is consumed by the medium-range and extended-range ensemble forecasts. The main reason being the need for many integrations instead of a single one. Therefore, the efficiency of running the ensemble forecasts is important. The aim of current and future developments in stochastic representations of model uncertainty is to develop schemes that are computationally highly efficient and contribute only moderately to the overall computational cost of the operational ensembles. An objective criterion to justify the computational cost of a stochastic representation of model uncertainty could be based on a comparison of the skill increase due to activating it and the skill increase due to adding members to the ensemble with the number of further members determined by matching the computational cost. Efforts will be devoted to identifying areas in the code and algorithms where efficiency gains could be made in the future. The random pattern generators are a promising candidate where savings may be feasible.

Developing a model error representation that is consistent at all lead times appears to remain a viable long-term goal. Work towards having a consistent representation in assimilation and forecasts is taking place in the context of the EDA and weak-constraint 4D-Var. The latter is a field that deserves further investigations. Some progress could be made by research with intermediate complexity models where the truth is known and thus model error is directly observable.

In the next 2 to 4 years, ECMWF resources for the area of model uncertainty will be primarily focused (i) on developing a configuration of SPP that is suitable for operational implementation, (ii) on extending SPP to include more aspects of model uncertainty, (iii) on developing model uncertainty representations for the dynamical core, (iv) on improving SPPT assuming that it will remain a useful element in uncertainty representation and (v) on increasing computational efficiency in the model uncertainty representations.

Acknowledgements

We would like to thank Peter Bauer, Roberto Buizza and Erland Källén who provided many useful comments on an earlier version. The ECMWF Fellows Tim Palmer and Tilmann Gneiting contributed welcome feedback and comments and Tim Palmer is also thanked for a fruitful discussion that concluded that the hard-to-define concept of “physically-based schemes” should be avoided. We are also grateful to two anonymous reviewers for their constructive comments.

The support of Anton Beljaars, Roberto Buizza, Franco Molteni and Erland Källén for putting resources into the development of the SPP scheme is gratefully acknowledged. Anton’s insight also helped tremendously for developing the SPP scheme. Jost von Hardenberg is thanked for diagnosing the $P - E$ imbalance of SPPT and proposing the global tendency fix for SPPT.

PKS acknowledges support by funding received from the European Research Council under the European Unions Seventh Framework Programme (FP7/2012/ERC Grant agreement no. 320375).

D.M. and A.W. were supported by the FP7 project SPECS (grant agreement number 308378). The research of H.M.C., S.J. and A.S. was supported by European Research Council grant number 291406. S.J. was also supported by the UK NERC grant NE/J00586X/1.

References

- Andrejczuk M, Cooper F, Juricke S, Palmer T, Weisheimer A, Zanna L. 2016. Oceanic stochastic parametrisation in a seasonal forecast system. *Mon. Wea. Rev.* **144**: 1867–1875, doi:10.1175/MWR-D-15-0245.1.
- Arnold HM. 2013. Stochastic parametrisation and model uncertainty. PhD thesis, University of Oxford.
- Augier P, Lindborg E. 2013. A new formulation of the spectral energy budget of the atmosphere, with application to two high-resolution general circulation models. *J. Atmos. Sci.* **70**: 2293–2308.
- Baker L, Rudd A, Migliorini S, Bannister R. 2014. Representation of model error in a convective-scale ensemble prediction system. *Nonlinear Processes in Geophysics* **21**: 19–39.
- Balsamo G, Agusti-Panareda A, Albergel C, Beljaars A, Boussetta S, Dutra E, Komori T, Lang S, Munoz-Sabater J, Pappenberger F, Rosnay PD, Sandu I, Wedi N, Weisheimer A, Wetterhall F, Zsoter E. 2014. Representing the Earth surfaces in the Integrated Forecasting System: Recent advances and future challenges. Tech. Memo. 729, ECMWF, Reading, UK.
- Bauer P, Thorpe A, Brunet G. 2015. The quiet revolution of numerical weather prediction. *Nature* **525**(7567): 47–55.
- Bechtold P, Köhler M, Jung T, Doblas-Reyes F, Leutbecher M, Rodwell MJ, Vitart F, Balsamo G. 2008. Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales. *Q. J. R. Meteor. Soc.* **134**: 1337–1351.
- Berner J, Fossell KR, Ha SY, Hacker JP, Snyder C. 2015. Increasing the skill of probabilistic forecasts: Understanding performance improvements from model-error representations. *Mon. Wea. Rev.* **143**: 1295–1320, doi:10.1175/MWR-D-14-00091.1.
- Berner J, Ha SY, Hacker JP, Fournier A, Snyder C. 2011. Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Mon. Wea. Rev.* **139**: 1972–1995, doi:10.1175/2010MWR3595.1.
- Berner J, Shutts GJ, Leutbecher M, Palmer TN. 2009. 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.
- Bishop C. 2016. Using trajectories of ensemble analyses and tools from weak constraint 4dvar to represent model uncertainty. In: *Workshop on model uncertainty*. ECMWF/WWRP: Reading, UK, p. 64, URL www.ecmwf.int/en/elibrary/16551-ecmwf-wwrp-workshop-model-uncertainty-proceedings.
- Bonavita M, Hólm E, Isaksen L, Fisher M. 2014. The evolution of the ECMWF hybrid data assimilation system. Tech. Memo. 743, ECMWF, Reading, UK.
- Bouttier F, Vié B, Nuissier O, Raynaud L. 2012. Impact of stochastic physics in a convection-permitting ensemble. *Mon. Wea. Rev.* **140**: 3706–3721, doi:10.1175/MWR-D-12-00031.1.
- Bowler NE. 2017. On the diagnosis of model error statistics using weak-constraint data assimilation. *Q. J. R. Meteor. Soc.* (accepted).
- Bowler NE, Arribas A, Mylne KR. 2008a. The benefits of multianalysis and poor mans ensembles. *Mon. Wea. Rev.* **136**: 4113–4129.
- Bowler NE, Arribas A, Mylne KR, Robertson KB, Beare SE. 2008b. The MOGREPS short-range ensemble prediction system. *Q. J. R. Meteor. Soc.* **134**: 703–722.
- Brankart JM, Candille G, Garnier F, Calone C, Melet A, Bouttier PA, Brasseur P, Verron J. 2015. A generic approach to explicit simulation of uncertainty in the NEMO ocean model. *Geoscientific Model Development* **8**(5): 1285–1297, doi:10.5194/gmd-8-1285-2015.
- Buizza R, Leutbecher M. 2015. The forecast skill horizon. *Q. J. R. Meteor. Soc.* **141**: 3366–3382, doi:10.1002/qj.2619.
- Buizza R, Leutbecher M, Isaksen L. 2008. Potential use of an ensemble of analyses in the ECMWF ensemble prediction system. *Q. J. R. Meteor. Soc.* **134**: 2051–2066.
- Buizza R, Miller M, Palmer TN. 1999. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Q. J. R. Meteor. Soc.* **125**: 2887–2908.
- Cardinali C, Žagar N, Radnoti G, Buizza R. 2014. Representing model error in ensemble data assimilation. *Nonlinear Processes in Geophysics* **21**(5): 971–985.
- Charron M, Pellerin G, Spacek L, Houtekamer PL, Gagnon N, Mitchell HL, Michelin L. 2010. Toward random sampling of model error in the Canadian ensemble prediction system. *Mon. Wea. Rev.* **138**: 1877–1901.
- Christensen HM, Lock SJ, Moroz IM, Palmer TN. 2016. Introducing independent patterns into the Stochastically Perturbed Parametrisation Tendencies (SPPT) scheme. *Q. J. R. Meteor. Soc.* (in review).
- Clark TL. 1979. Numerical simulations with a three-dimensional cloud model: Lateral boundary condition experiments and multicellular severe storm simulations. *J. Atmos. Sci.* **36**(11): 2191–2215.
- Cossette JF, Smolarkiewicz PK, Charbonneau P. 2014. The Monge-Ampère trajectory correction for semi-Lagrangian schemes. *J. Comp. Phys.* **274**: 208–229.
- Cullen M, Salmond D, Smolarkiewicz P. 2000. Key numerical issues for the future development of the ECMWF model. In: *Proceedings of the ECMWF Workshop on Developments in Numerical Methods for Very High Resolution Global Models*, Reading, UK, ECMWF, pp. 5–7.
- Davies L, Jakob C, May P, Kumar V, Xie S. 2013. Relationships between the large-scale atmosphere and the small-scale convective state for Darwin, Australia. *J. Geophys. Res.* **118**(20).
- Davini P, von Hardenberg J, Corti S, Christensen HM, Juricke S, Subramanian A, Watson PAG, Weisheimer A, Palmer TN. 2017. Climate SPHINX: evaluating the impact of resolution and stochastic physics parameterisations in climate simulations. *Geosci. Model Dev. Discuss.* **10**: 1383–1402, doi:10.5194/gmd-10-1383-2017.
- Deconinck W, Hamrud M, Kühnlein C, Mozdzyński G, Smolarkiewicz PK, Szmelter J, Wedi NP. 2016. Accelerating extreme-scale numerical weather prediction. In: *Parallel Processing and Applied Mathematics*, Springer, pp. 583–593.
- DeMott CA, Klingaman NP, Woolnough SJ. 2015. Atmospheric-ocean coupled processes in the Madden-Julian oscillation. *Reviews of Geophysics* **53**: 1099–1154, doi:10.1002/2014RG000478.
- Derber J, Bouttier F. 1999. A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus A* **51**: 195–221.
- Descamps L, Labadie C, Joly A, Bazile E, Arbogast P, Cébron P. 2015. PEARP, the Météo-France short-range ensemble prediction system. *Q. J. R. Meteor. Soc.* **141**: 1671–1685, doi:10.1002/qj.2469.
- Diamantakis M, Magnusson L. 2016. Sensitivity of the ECMWF model to semi-Lagrangian departure point iterations. *Mon. Wea. Rev.* **144**: 3233–3250, doi:10.1175/MWR-D-15-0432.1.
- Durrán DR, Gingrich M. 2014. Atmospheric predictability: Why butterflies are not of practical importance. *J. Atmos. Sci.* **71**(7): 2476–2488.
- Eckermann SD. 2011. Explicitly stochastic parameterization of nonorographic gravity wave drag. *J. Atmos. Sci.* **68**: 1749–1765, doi:10.1175/2011JAS3684.1.
- ECMWF/WWRP (ed). 2016. *Proceedings of workshop on model uncertainty*, URL www.ecmwf.int/en/elibrary/16551-ecmwf-wwrp-workshop-model-uncertainty-proceedings.
- Fisher M. 2003a. Background error covariance modelling. In: *Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean*. ECMWF: Reading, UK, pp. 45–63.
- Fisher M. 2003b. Background error covariance modelling. In: *Proceedings of the ECMWF Seminar on recent developments in data assimilation for atmosphere and ocean*. ECMWF, pp. 45–63.
- Flemming J, Benedetti A, Inness A, Engelen R, Jones L, Huijnen V, Remy S, Parrington M, Suttie M, Bozzo A, Peuch VH, Akritidis D, Katragkou E. 2016. The CAMS interim reanalysis of carbon monoxide, ozone and aerosol for 2003–2015. *Atmospheric Chemistry and Physics Discussions*

- 2016: 1–82, doi:10.5194/acp-2016-666.
- Flemming J, Inness A, Flentje H, Huijnen V, Moinat P, Schultz M, Stein O. 2009. Coupling global chemistry transport models to ECMWF's Integrated Forecast System. *Geoscientific Model Development* **2**: 253–265.
- Gebhardt C, Theis S, Paulat M, Bouallégué ZB. 2011. Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. *Atmospheric Research* **100**: 168–177, doi: 10.1016/j.atmosres.2010.12.008.
- Gneiting T. 2002. Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association* **97**(458): 590–600.
- Gneiting T. 2014. Calibration of medium-range weather forecasts. Tech. Memo. 719, ECMWF, Reading, UK.
- Grabowski WW, Smolarkiewicz PK. 1999. CRCP: a Cloud Resolving Convection Parameterization for modeling the tropical convecting atmosphere. *Physica D: Nonlinear Phenomena* **133**(1-4): 171–178.
- Haiden T, Duffy S. 2016. Use of high-density observations in precipitation verification. *ECMWF Newsletter* **147**: 20–25.
- Haiden T, Forbes R, Ahlgrimm M, Bozzo A. 2015. The skill of ECMWF cloudiness forecasts. *ECMWF Newsletter* **143**: 14–19.
- Hodyss D, McLay JG, Moskaitis J, Serra EA. 2014. Inducing tropical cyclones to undergo Brownian motion: A comparison between Itô and Stratonovich in a numerical weather prediction model. *Mon. Wea. Rev.* **142**: 1982–1996.
- Hogan RJ, Bozzo A. 2015. Mitigating errors in surface temperature forecasts using approximate radiation updates. *J. Adv. Model. Earth Syst.* **7**(2): 836–853.
- Holm DD. 2015. Variational principles for stochastic fluid dynamics. *Proc. R. Soc. A* **471**(2176), doi:10.1098/rspa.2014.0963.
- IFS Documentation. 2016. Part IV: Physical processes. IFS Documentation CY43R1, ECMWF, URL www.ecmwf.int/en/elixir/17117-part-iv-physical-processes. Operational implementation: 16 November 2016.
- Inness PM, Slingo JM, Guilyardi E, Cole J. 2003. Simulation of the Madden-Julian oscillation in a coupled general circulation model. Part II: The role of the basic state. *J. Clim* **16**: 365–382.
- Isaksen I, Bonavita M, Buizza R, Fisher M, Haseler J, Leutbecher M, Raynaud L. 2010. Ensemble of data assimilations at ECMWF. Tech. Memo. 636, ECMWF.
- Iversen T, Deckmyn A, Santos C, Sattler K, Bremnes JB, Feddersen H, Frogner IL. 2011. Evaluation of 'GLAMEPS' — a proposed multimodel EPS for short range forecasting. *Tellus A* **63**: 513–530, doi:10.1111/j.1600-0870.2010.00507.x.
- Juricke S, Goessling HF, Jung T. 2014. Potential sea ice predictability and the role of stochastic sea ice strength perturbations. *Geophys. Res. Letters* **41**(23): 8396–8403, doi:10.1002/2014GL02081. 2014GL02081.
- Juricke S, Palmer TN, Zanna L. 2017. Stochastic sub-grid scale ocean mixing: Impacts on low frequency variability. *J. Clim.* doi:10.1175/JCLI-D-16-0539.1. (accepted).
- Khairoutdinov MF, Randall DA. 2003. Cloud resolving modeling of the ARM summer 1997 IOP: Model formulation, results, uncertainties, and sensitivities. *J. Atmos. Sci.* **60**(4): 607–625.
- Kleiber W. 2016. High resolution simulation of nonstationary Gaussian random fields. *Comp. Statistics & Data Analysis* **101**: 277 – 288, doi: <http://dx.doi.org/10.1016/j.csda.2016.03.005>.
- Klinker E, Rabier F, Kelly G, Mahfouf JF. 2000. The ECMWF operational implementation of four-dimensional variational assimilation. III: Experimental results and diagnostics with operational configuration. *Q. J. R. Meteor. Soc.* **126**: 1191–1215.
- Kober K, Craig GC. 2016. Physically based stochastic perturbations (PSP) in the boundary layer to represent uncertainty in convective initiation. *J. Atmos. Sci.* **73**: 2893–2911, doi:10.1175/JAS-D-15-0144.1.
- Kröber D, Botev Z. 2013. *Spatial process generation, lectures on stochastic geometry, spatial statistics and random fields, Volume II: Analysis, modeling and simulation of complex structures*. Springer-Verlag: Berlin. (in press).
- Kurowski MJ, Grabowski WW, Smolarkiewicz PK. 2015. Anelastic and compressible simulation of moist dynamics at planetary scales. *J. Atmos. Sci.* **72**: 3975–3995.
- Lang STK, Bonavita M, Leutbecher M. 2015. On the impact of re-centring initial conditions for ensemble forecasts. *Q. J. R. Meteor. Soc.* **141**: 2571–2581, doi:10.1002/qj.2543.
- Leutbecher M, Lang STK. 2014. On the reliability of ensemble variance in subspaces defined by singular vectors. *Q. J. R. Meteor. Soc.* **140**: 1453–1466, doi:10.1002/qj.2229.
- Lewis JM. 2005. Roots of ensemble forecasting. *Mon. Wea. Rev.* **133**: 1865–1885.
- Lindgren F, Rue H, Lindström J. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J. Roy. Statistical Soc. B* **73**(4): 423–498.
- Lorenz EN. 1969. The predictability of a flow which possesses many scales of motion. *Tellus* **21**: 289–307.
- Lott F, Guez L, Maury P. 2012. A stochastic parameterization of non-orographic gravity waves: Formalism and impact on the equatorial stratosphere. *Geophys. Res. Letters* **39**(6): L06 807, doi:10.1029/2012GL051001.
- MacLeod DA, Cloke HL, Pappenberger F, Weisheimer A. 2016. Improved seasonal prediction of the hot summer of 2003 over Europe through better representation of uncertainty in the land surface. *Q. J. R. Meteor. Soc.* **142**: 79–90, doi:10.1002/qj.2631.
- Mahfouf JF, Rabier F. 2000. The ECMWF operational implementation of four-dimensional variational assimilation. II: Experimental results with improved physics. *Q. J. R. Meteor. Soc.* **126**: 1171–1190, doi:10.1002/qj.49712656416.
- Malardel S, Wedi N, Deconinck W, Diamantakis M, Kühnlein C, Mozdzyński G, Hamrud M, Smolarkiewicz P. 2016. A new grid for the IFS. *ECMWF Newsletter* **146**: 23–28.
- Malardel S, Wedi NP. 2016. How does subgrid-scale parametrisation influence nonlinear spectral energy fluxes in global NWP models? *J. Geophys. Res. Atmos.* **121**, doi:10.1002/2015JD02.
- Massart S, Bonavita M. 2016. Ensemble of data assimilations applied to the CAMS greenhouse gases analysis. Tech. Memo. 780, ECMWF, Reading, UK.
- Mauritsen T, Stevens B, Roeckner E, Crueger T, Esch M, Giorgetta M, Haak H, Jungclaus J, Klocke D, Matei D, Mikolajewicz U, Notz D, Pincus R, Schmidt H, Tomassini L. 2012. Tuning the climate of a global model. *Journal of Advances in Modeling Earth Systems* **4**(3), doi:10.1029/2012MS000154.
- McCabe A, Swinbank R, Tennant W, Lock A. 2016. Representing model uncertainty in the Met Office convection-permitting ensemble prediction system and its impact on fog forecasting. *Q. J. R. Meteor. Soc.* **142**: 2897–2910, doi:10.1002/qj.2876.
- Morcrette JJ, Barker HW, Cole JNS, Iacono MJ, Pincus R. 2008a. Impact of a new radiation package, McRad, in the ECMWF Integrated Forecasting System. *Mon. Wea. Rev.* **136**: 4773–4798, doi:10.1175/2008MWR2363.1.
- Morcrette JJ, Mozdzyński G, Leutbecher M. 2008b. A reduced radiation grid for the ECMWF Integrated Forecasting System. *Mon. Wea. Rev.* **136**: 4760–4772.
- Myhre G, Shindell D, Bréon FM, Collins W, Fuglestad J, Huang J, Koch D, Lamarque JF, Lee D, Mendoza B, Nakajima T, Robock A, Stephens G, Takemura T, Zhang H. 2013. *Climate change 2013: The physical science basis*, ch. Anthropogenic and Natural Radiative Forcing. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.
- Nastrom GD, Gage KS. 1985. A climatology of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft. *J. Atmos. Sci.* **42**: 950–960, doi:10.1175/1520-0469(1985)042<0950:ACOWS>2.0.CO;2.
- Ollinaho P, Lock SJ, Leutbecher M, Bechtold P, Beljaars A, Bozza A, Forbes RM, Haiden T, Hogan RJ, Sandu I. 2017. Towards process-level representation of model uncertainties: Stochastically perturbed parametrisations in the ECMWF ensemble. *Q. J. R. Meteor. Soc.* **143**: 408–422, doi:10.1002/qj.2931.
- Orth R, Dutra E, Pappenberger F. 2016. Improving weather predictability by including land surface model parameter uncertainty. *Mon. Wea. Rev.* **144**: 1551–1569, doi:10.1175/MWR-D-15-0283.1.
- Palmer T. 2014. Climate forecasting: Build high-resolution global climate models. *Nature* **515**: 338–339, doi:10.1038/515338a.
- Palmer T, Buizza R, Doblas-Reyes F, Jung T, Leutbecher M, Shutts G, Steinheimer M, Weisheimer A. 2009. Stochastic parameterization and model uncertainty. Tech. Memo. 598, ECMWF, Reading, UK.
- Palmer TN. 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. Meteor. Soc.* **127**: 279–304.
- Palmer TN. 2012. Towards the probabilistic Earth-system simulator: a vision for the future of climate and weather prediction. *Q. J. R. Meteor. Soc.* **138**: 841–861, doi:10.1002/qj.1923.
- Palmer TN, Shutts GJ, Hagedorn R, Doblas-Reyes FJ, Jung T, Leutbecher M. 2005. Representing model uncertainty in weather and climate prediction. *Annu. Rev. Earth Planet. Sci.* **33**: 163–193.
- Peters K, Jakob C, Davies L, Khouider B, Majda AJ. 2013. Stochastic behavior of tropical convection in observations and a multicloud model. *J. Atmos. Sci.* **70**: 3556–3575.
- Piccolo C, Cullen M. 2016. Ensemble data assimilation using a unified representation of model error. *Mon. Wea. Rev.* **144**: 213–224, doi:10.1175/

- MWR-D-15-0270.1.
- Plant R, Craig GC. 2008. A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci.* **65**: 87–105.
- Rabier F, H J, Klinker E, Mahfouf JF, Simmons A. 2000. The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Q. J. R. Meteor. Soc.* **126**: 1143–1170.
- Randall D, Khairoutdinov M, Arakawa A, Grabowski W. 2003. Breaking the Cloud Parameterization Deadlock. *Bulletin of the American Meteorological Society* **84**(11): 1547–1564.
- Raynaud L, Berre L, Desroziers G. 2012. Accounting for model error in the Météo-France ensemble data assimilation system. *Q. J. R. Meteor. Soc.* **138**: 249–262.
- Reynolds C, Ridout J, McLay J. 2011. Examination of parameter variations in the U. S. Navy global ensemble. *Tellus A* **63**: 841–857, doi:10.1111/j.1600-0870.2011.00532.x.
- Rodwell MJ, Ferranti L, Haiden T, Magnusson L, Bidlot J, Bormann N, Dahoui M, Chiara GD, Duffy S, Forbes R, Hólm E, Ingleby B, Janousek M, Lang S, Mogensen K, Prates F, Rabier F, Richardson DS, Tsonevsky J, Vitart F, Yamaguchi M. 2015. New developments in the diagnosis and verification of high-impact weather forecasts. Tech. Memo. 759, ECMWF, Reading, UK.
- Rodwell MJ, Lang STK, Ingleby NB, Bormann N, Hólm E, Rabier F, Richardson DS, Yamaguchi M. 2016. Reliability in ensemble data assimilation. *Q. J. R. Meteor. Soc.* **142**: 443–454, doi:10.1002/qj.2663.
- Romine GS, Schwartz CS, Berner J, Fossell KR, Snyder C, Anderson JL, Weisman ML. 2014. Representing forecast error in a convection-permitting ensemble system. *Mon. Wea. Rev.* **142**: 4519–4541, doi:10.1175/MWR-D-14-00100.1.
- Saetia Ø, Hersbach H, Bidlot JR, Richardson DS. 2004. Effects of observation errors on the statistics for ensemble spread and reliability. *Mon. Wea. Rev.* **132**: 1487–1501.
- Sanchez C, Williams KD, Collins M. 2016. Improved stochastic physics schemes for global weather and climate models. *Q. J. R. Meteor. Soc.* **142**: 147–159.
- Sandu I, Beljaars A, Bechtold P, Mauritsen T, Balsamo G. 2013. Why is it so difficult to represent stably stratified conditions in numerical weather prediction (NWP) models? *J. Adv. Model. Earth Syst.* **5**(2): 117–133, doi:10.1002/jame.20013.
- Scheffzik R, Thorarindottir TL, Gneiting T. 2013. Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statist. Sci.* **28**: 616–640, doi:10.1214/13-STS443.
- Schlather M. 2012. *Construction of covariance functions and unconditional simulation of random fields*. Springer-Verlag: Berlin, Heidelberg, pp. 25–54, doi:10.1007/978-3-642-17086-7_2.
- Selz T, Craig GC. 2015a. Simulation of upscale error growth with a stochastic convection scheme. *Geophys. Res. Letters* **42**(8): 3056–3062.
- Selz T, Craig GC. 2015b. Upscale error growth in a high-resolution simulation of a summertime weather event over Europe. *Mon. Wea. Rev.* **143**: 813–827.
- Separovic L, Charron M, Erfani A, Gagnon N, Zadra A, Vaillancourt P. 2016. Model error representation in the Canadian ensemble prediction systems. In: *Workshop on model uncertainty*. ECMWF/WWRP: Reading, UK, p. 48, URL www.ecmwf.int/en/elibrary/16551-ecmwf-wwrp-workshop-model-uncertainty-proceedings.
- Shutts G. 2005. A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Q. J. R. Meteor. Soc.* **131**: 3079–3102.
- Shutts G. 2015. A stochastic convective backscatter scheme for use in ensemble prediction systems. *Q. J. R. Meteor. Soc.* **141**: 2602–2616, doi:10.1002/qj.2547.
- Shutts G, Leutbecher M, Weisheimer A, Stockdale T, Isaksen L, Bonavita M. 2011. Representing model uncertainty: Stochastic parametrizations at ECMWF. *ECMWF Newsletter* **129**: 19–24.
- Shutts G, Pallarès AC. 2014. Assessing parametrization uncertainty associated with horizontal resolution in numerical weather prediction models. *Phil. Trans. R. Soc. A* doi:10.1098/rsta.2013.0284.
- Shutts GJ. 2013. Coarse graining the vorticity equation in the ECMWF integrated forecasting system: The search for kinetic energy backscatter. *J. Atmos. Sci.* **70**: 1233–1241, doi:10.1175/JAS-D-12-0216.1.
- Shutts GJ, Palmer TN. 2007. Convective forcing fluctuations in a cloud-resolving model: Relevance to the stochastic parameterization problem. *J. Clim.* **20**: 187–202.
- Sigrist F, Künsch HR, Stahel WA. 2015. Stochastic partial differential equation based modelling of large space-time data sets. *J. Roy. Statistical Soc. B* **77**(1): 3–33.
- Smolarkiewicz P, Deconinck W, Hamrud M, Kühnlein C, Mozdzyński G, Szmelter J, Wedi N. 2015. An all-scale, finite-volume module for the IFS. *ECMWF Newsletter* **145**: 24–29.
- Smolarkiewicz PK, Deconinck W, Hamrud M, Kühnlein C, Mozdzyński G, Szmelter J, Wedi NP. 2016. A finite-volume module for simulating global all-scale atmospheric flows. *J. Comp. Phys.* **314**: 287–304.
- Smolarkiewicz PK, Pudykiewicz JA. 1992. A class of semi-Lagrangian approximations for fluids. *J. Atmos. Sci.* **49**: 2082–2096.
- Subramanian A, Weisheimer A, Palmer T, Vitart F, Bechtold P. 2017. Impact of stochastic physics on tropical precipitation and climate variability in the ECMWF IFS. *Q. J. R. Meteor. Soc.* **143**: 852–865, doi:10.1002/qj.2970.
- Sullivan PP, Weil JC, Patton EG, Jonker HJ, Mironov DV. 2016. Turbulent winds and temperature fronts in large-eddy simulations of the stable atmospheric boundary layer. *J. Atmos. Sci.* **73**: 1815–1840.
- Tennant WJ, Shutts GJ, Arribas A, Thompson SA. 2011. Using a stochastic kinetic energy backscatter scheme to improve MOGREPS probabilistic forecast skill. *Mon. Wea. Rev.* **139**: 1190–1206.
- Theis S, Buchhold M, Gebhardt C, Kohlhepp R, Röpnack A. 2016. Model uncertainty representation in COSMO-DE-EPS. In: *Workshop on model uncertainty*. ECMWF/WWRP: Reading, UK, pp. 52–53, URL www.ecmwf.int/en/elibrary/16551-ecmwf-wwrp-workshop-model-uncertainty-proceedings.
- Todling R. 2015a. A lag-1 smoother approach to system-error estimation: Sequential method. *Q. J. R. Meteor. Soc.* **141**: 1502–1513.
- Todling R. 2015b. A complementary note to ‘A lag-1 smoother approach to system-error estimation’: The intrinsic limitations of residual diagnostics. *Q. J. R. Meteor. Soc.* **141**: 2917–2922.
- Trémolet Y. 2007. Model-error estimation in 4D-Var. *Q. J. R. Meteor. Soc.* **133**: 1267–1280, doi:10.1002/qj.94.
- Tsyrlunikov M, Gayfulin D. 2016. A limited-area spatio-temporal stochastic pattern generator for simulation of uncertainties in ensemble applications. *Meteor. Z. (in press)*, see also arXiv:1605.02018.
- Vaña F, Düben P, Lang S, Palmer T, Leutbecher M, Salmond D, Carver G. 2017. Single precision in weather forecasting models: An evaluation with the IFS. *Mon. Wea. Rev.* **145**: 495–502, doi:10.1175/MWR-D-16-0228.1.
- Vitart F, Molteni F. 2010. Simulation of the Madden-Julian oscillation and its teleconnections in the ECMWF forecast system. *Q. J. R. Meteor. Soc.* **136**: 842–855.
- Watson PAG, Christensen HM, Palmer TN. 2015. Does the ECMWF IFS convection parameterization with stochastic physics correctly reproduce relationships between convection and the large-scale state? *J. Atmos. Sci.* **72**: 236–242, doi:10.1175/JAS-D-14-0252.1.
- Weaver A, Courtier P. 2001. Correlation modelling on the sphere using a generalized diffusion equation. *Q. J. R. Meteor. Soc.* **127**: 1815–1846.
- Wedi NP, Deconinck W, Diamantakis M, Hamrud M, Kühnlein C, Malardel S, Mogensen K, Mozdzyński G, Smolarkiewicz P. 2015. The modelling infrastructure of the Integrated Forecasting System: Recent advances and future challenges. Tech. Memo. 760, ECMWF, Reading, UK.
- Weisheimer A, Corti S, Palmer T, Vitart F. 2014. Addressing model error through atmospheric stochastic physical parametrizations: impact on the coupled ECMWF seasonal forecasting system. *Phil. Trans. R. Soc. A* **372**(2018), doi:10.1098/rsta.2013.0290.
- Weisheimer A, Palmer TN, Doblas-Reyes FJ. 2011. Assessment of representations of model uncertainty in monthly and seasonal forecast ensembles. *Geophys. Res. Letters* **38**(16), doi:10.1029/2011GL048123. L16703.
- Wheeler MC, Hendon HH. 2004. An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.* **132**: 1917–1932.
- Wilks DS. 2005. Effects of stochastic parametrizations in the Lorenz ‘96 system. *Q. J. R. Meteor. Soc.* **131**: 389–407.
- Williams PD, Howe NJ, Gregory JM, Smith RS, Joshi MM. 2016. Improved climate simulations through a stochastic parameterization of ocean eddies. *J. Clim.* **29**: 8763–8781, doi:10.1175/JCLI-D-15-0746.1.
- Yamaguchi M, Lang STK, Leutbecher M, Rodwell MJ, Radnoti G, Bormann N. 2016. Observation-based evaluation of ensemble reliability. *Q. J. R. Meteor. Soc.* **142**: 506–514, doi:10.1002/qj.2675.

A. Details for global fix of tendency perturbations in SPPT

The P–E problem of the SPPT scheme (operational until November 2016) is demonstrated in Figure 12a for results from monthly forecast experiments. The value of P–E in the ensemble mean is significantly lower than in the control forecast.

To address these imbalances a modification of the SPPT scheme was developed. A correction is added to the perturbed tendency, which results in the global integral of the perturbed tendency being equal to that of the unperturbed tendency. The corrected perturbed

tendency p^* is given by

$$p^* = p_1 + w(\langle p_0 \rangle - \langle p_1 \rangle) \quad (5)$$

$$w = \frac{|p_0 - p_1|}{\langle |p_0 - p_1| \rangle} \quad (6)$$

with p_0 being the unperturbed physics tendency and p_1 being the tendency after applying the stochastic perturbations and the limiters described in Sec. 2.1.1. The operator $\langle \dots \rangle$ denotes a global spatial average of the mass-weighted vertically integrated quantity. The local weights w distribute the correction proportional to the amplitude of the local tendency perturbation.

The modification eliminates the drying of the atmosphere on the global scale (Fig 12b). The globally conserving SPPT scheme also modifies the model climate as diagnosed from a set of seasonal hindcasts. While the unperturbed coupled model without SPPT rains too much compared to GPCP over the convectively active regions in the tropics on seasonal time scales, the SPPT scheme without the global fix reduces the convective activity leading to smaller precipitation biases. With the SPPT scheme using the global fix, the global mean difference between the perturbed and unperturbed simulations become smaller leading to a smaller reduction of the wet precipitation bias. In terms of medium-range forecast scores, it was found that the SPPT modification has a neutral to slightly positive impact. In view of this impact on the ensemble forecasts and the fact that it makes the climate of the forecasts perturbed with SPPT more similar to the climate of unperturbed forecasts, this modification of SPPT was implemented in November 2016.

B. Semi-Lagrangian departure point perturbations

A stochastic scheme to represent departure point (DP) uncertainty could adopt the following algorithm. For advection of a scalar field ϕ ,

$$\frac{D\phi}{Dt} \equiv \frac{\partial \phi}{\partial t} + \mathbf{V} \cdot \nabla \phi = 0, \quad (7)$$

where $\mathbf{V} = (u, v, w)$ is the wind field, the semi-Lagrangian approximation for the solution at time t is

$$\phi(\mathbf{r}_A, t) = \phi(\mathbf{r}_D, t - \Delta t), \quad (8)$$

where \mathbf{r}_A denotes the coordinates of the “arrival points” (the gridpoints of the discretization mesh) and \mathbf{r}_D denotes the coordinates of the irregularly spaced DP mesh which is computed at each timestep.

The proposed method for representing the uncertainty of the advection scheme is to add a small “noise” term ε to the estimated DP \mathbf{r}_D to obtain

$$\phi(\mathbf{r}_A, t) = \phi(\mathbf{r}_D + \varepsilon, t - \Delta t) \quad (9)$$

where the stochastic term ε contains flow-dependent information.

The perturbed DP method of Eq (9) can be regarded as the semi-Lagrangian scheme “analogue” of the method proposed by Hodyss *et al.* (2014). This can be seen by taking a first order Taylor expansion around (\mathbf{r}_A, t) of the stochastically perturbed advected field as follows:

$$\begin{aligned} \phi(\mathbf{r}_A, t) &= \phi(\mathbf{r}_D + \varepsilon, t - \Delta t) \\ &= \phi(\mathbf{r}_A - \mathbf{V}\Delta t + \varepsilon, t - \Delta t) \\ &\approx \phi(\mathbf{r}_A, t) + (\varepsilon - \mathbf{V}\Delta t) \cdot \nabla \phi - \Delta t \frac{\partial \phi}{\partial t}. \end{aligned} \quad (10)$$

Rearranging (10), it is seen that stochastically perturbing the DP estimate can be expressed as a stochastic perturbation to the

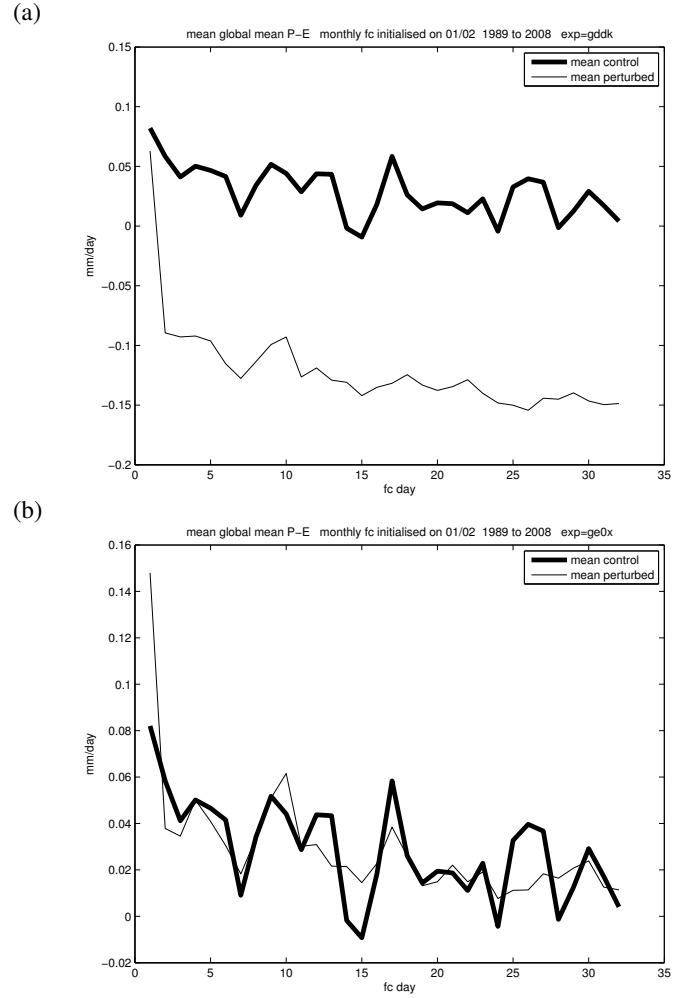


Figure 12. Global mean precipitation (P) minus evaporation (E) in mm d^{-1} for two different configurations of SPPT (a) default, (b) global fix of perturbed tendency. Ensemble mean (thin) and control forecast (thick line); average over 20 years of hindcasts, February start dates 1989–2008, CY41R1.

advection of ϕ :

$$\frac{\partial \phi}{\partial t} + \mathbf{V} \cdot \nabla \phi \approx \frac{1}{\Delta t} \varepsilon \cdot \nabla \phi$$

which shows that the proposed method approximates the “Stratonovich” stochastic parametrization of advection proposed by Hodyss *et al.* (2014).

Flow-dependent uncertainty can be incorporated into the perturbed DP estimate by multiplying the difference between DP estimates from two different iterations with a random field γ :

$$\mathbf{r}_D + \varepsilon = \mathbf{r}_D^{(l)} + \gamma \left(\mathbf{r}_D^{(l)} - \mathbf{r}_D^{(l-m)} \right)$$

where $\mathbf{r}_D^{(j)}$ is the unperturbed estimate of the DP after the j -th iteration; and l refers to final iteration and $(l-m)$ to one of the earlier iterations. In this way, the horizontal and vertical structure of the adjustments to the DP estimate are inherited by the noise term.