Using a Stochastic Kinetic Energy Backscatter Scheme to Improve MOGREPS Probabilistic Forecast Skill

WARREN J. TENNANT, GLENN J. SHUTTS, ALBERTO ARRIBAS, AND SIMON A. THOMPSON
Met Office, Exeter, United Kingdom

(Manuscript received 25 March 2010, in final form 24 September 2010)

ABSTRACT

An improved stochastic kinetic energy backscatter scheme, version 2 (SKEB2) has been developed for the Met Office Global and Regional Ensemble Prediction System (MOGREPS). Wind increments at each model time step are derived from a streamfunction forcing pattern that is modulated by a locally diagnosed field of likely energy loss due to numerical smoothing and unrepresented convective sources of kinetic energy near the grid scale. The scheme has a positive impact on the root-mean-square error of the ensemble mean and spread of the ensemble. An improved growth rate of spread results in a better match with ensemble-mean forecast error at all forecast lead times, with a corresponding improvement in probabilistic forecast skill from a more realistic representation of model error. Other examples of positive impact include improved forecast blocking frequency and reduced forecast jumpiness. The paper describes the formulation of the SKEB2 and its assessment in various experiments.

1. Introduction

Ensemble prediction systems (EPSs) provide an objective way to estimate uncertainty in weather and climate forecasts (Buizza et al. 2005). The sensitivity of forecasts to changes in initial states discovered by Lorenz (1963) aided in the development of ideas to generate multiple realizations of a numerical model forecast by perturbing the initial conditions (Leith 1974). The basis of the approach was to use Monte Carlo approximations of the stochastic dynamic forecasting technique suggested by Epstein (1969). This shifted the focus from a purely deterministic view of forecasting the weather to the idea of trying to quantify the amount of uncertainty in a forecast. However, forecast uncertainty was not only attributed to initial condition errors. A second cause of forecast uncertainty was ascribed to upscaling errors that occurred because of the limited variability in the model phase space. To address this problem, Leith (1978) suggested using empirical correction terms directly in forecast models to represent the unresolved scales of motion. This idea could be used to generate an ensemble suite by forcing forecast runs to diverge during the course of the forecast integration, even if they started from the same initial conditions. Given a solid scientific basis to generate ensembles of forecasts, it now seemed possible to estimate forecast reliability a priori and thus include confidence as an actual forecast parameter (Bengtsson 1991).

With sufficient improvements to the atmospheric observing system, increased computing power, and more sophisticated models, the development of operational EPS suites took place at, inter alia, the National Centers for Environmental Prediction (NCEP; Toth and Kalnay 1993), the Meteorological Service of Canada (MSC; Houtekamer et al. 1996), and the European Centre for Medium-Range Weather Forecasts (ECMWF; Buizza and Palmer 1995; Molteni et al. 1996). Some centers focused on estimating uncertainty in initial conditions of the forecasts, although some success has also been demonstrated in using a lagged forecasting approach (Hoffman and Kalnay 1983). This method is often used in extended-range forecasts (e.g., Vitart et al. 2008), but it has also recently shown utility in short-range convective-scale ensemble systems (Mittermaier 2007). Another popular method to construct an ensemble is to use a multimodel approach. This has a number of unique advantages as discussed, for example, in Park et al. (2008), Candille (2009), and Hagedorn et al. (2005).

Notwithstanding the various methods, and combinations thereof, used to generate ensemble suites, it is clear
that not fully representing the variability of the small scales in numerical weather prediction models can still add to errors of the model-mean state (Penland 2003). This has led to a drive to include some method of addressing model error in forecast systems. Some of the more direct approaches include, for example, introducing a stochastic element into atmospheric models by randomly perturbing the increments or tendencies from parameterization schemes (e.g., Buizza et al. 1999; Palmer et al. 2009), while other approaches seek to formulate the parameterization schemes in a stochastic way (Palmer 2001; Palmer and Williams 2008; Plant and Craig 2008). However, as noted by Buizza et al. (2005), representing forecast uncertainty in imperfect models can be more challenging than simulating initial condition errors, and as a result there are still a range of schemes with varying levels of success being tested to address this problem.

One such scheme, which may be considered as a stochastic parameterization of missing processes in a numerical model, is the backscatter of kinetic energy (Frederiksen and Davies 1997; Shutts 2005). Variations of this scheme have been tested in the ECMWF EPS (Berner et al. 2009) and MSC Global Environmental Multiscale (GEM) model (Houtekamer et al. 2009; Charron et al. 2010). The objective of our paper is to evaluate the implementation of a second version of the kinetic energy backscatter scheme (SKEB2) in the Met Office global EPS with the aim of improving the spread of the ensemble and consequently the probabilistic forecast skill and associated estimate of uncertainty.

2. MOGREPS description

The Met Office Global and Regional Ensemble Prediction System (MOGREPS) has been designed to tackle forecast uncertainty at short time scales (Bowler et al. 2008). The system is based on the Met Office Unified Model (UM; Davies et al. 2005) and uses the ensemble transform Kalman filter (ETKF; Bishop et al. 2001) to calculate initial condition perturbations for 23 ensemble members. A control run at the same resolution without perturbations completes a 24-member suite that runs twice a day (global at 0000 and 1200 UTC, and regional at 0600 and 1800 UTC). The regional suite is driven by lateral boundary conditions output from the global suite initialized 6 h earlier and using initial perturbations that are rescaled from the global system (Bowler and Mylne 2009).

Recent improvements to MOGREPS include a global model resolution upgrade from 1.25° × 0.83° × 38 levels (lid at 39 km) (N144L38) to 0.83° × 0.55° × 70 levels (lid at 80 km) (N216L70) in early 2010. The regional model was upgraded from 24kmL38 to 18kmL38 at the same time, with a further upgrade to 18kmL70 in mid-2010. Both upgrades using the same vertical level set as the global model.

Recognizing the need for incorporating model uncertainties in an EPS setup, MOGREPS includes the following stochastic physics schemes. The first, available in both the global and regional models, is a random parameters (RP) scheme, where a number of selected parameters controlling the large-scale precipitation, convection, boundary layer, and gravity wave drag parameterization schemes are stochastically perturbed during the model forecast. The perturbations evolve according to a first-order autoregression process (Bowler et al. 2008). A similar scheme is used in the Consortium for Small Scale Modeling—Deutsch (COSMO-DE) system (Gebhardt et al. 2008), except here each ensemble member has a predefined set of perturbed parameters that remain fixed during the forecast run.

The second set of stochastic schemes aims to address missing physics processes in the MOGREPS forecast model. The development began with a stochastic convective vorticity (SCV) scheme that constructs potential vorticity (PV) anomaly dipoles that are typically associated with mesoscale convective systems (Bowler et al. 2008). The combined impact of the SCV and RP schemes on the model climate accuracy is small, but there is a contribution to increasing EPS spread growth during the model forecast. Following this, the first version of the SKEB (SKEB1) was implemented in the MOGREPS global model (Bowler et al. 2009). Technically, SKEB1 is similar to the SKEB2 described in this paper. A three-dimensional random pattern with prescribed spatial and temporal characteristics is generated, and is modulated by a field that represents the spatial structure of energy sinks in the model. In SKEB1 the modulating field is based on the wind field kinetic energy and scaled to a global average value of 0.75 W m⁻². Results from this scheme showed an increased spread in wind variables and an increased rate of growth of spread in the upper troposphere.

SKEB2 has been designed to incorporate and expand upon the benefits of SCV and SKEB1. It continues with the idea of identifying areas of excessive energy dissipation by the model and introduces other possible sources of kinetic energy from processes such as convection, which are not accounted for by the model. This energy is then scattered back to the larger resolved scales as horizontal wind increments at each forecast time step. The details of the scheme are described in the next section.

3. The stochastic kinetic energy backscatter scheme

The rationale of backscatter originates with large-eddy simulations (LESs) of the turbulent boundary layer. To
maintain the momentum flux by the large-scale eddies and reduce errors in the near-surface flow. Mason and Thomson (1992) used a stochastic momentum forcing at scales close to the model resolution limit. This forcing consisted of random stresses and scalar fluxes to provide backscatter of energy and scalar variance to account for missing stochastic subgrid stress variations. Frederiksen and Davies (1997) and Shutts (2005) suggested that a similar backscatter scheme could be developed to inject energy into a numerical weather prediction model to offset the excessive energy dissipation by numerical advection and horizontal diffusion. There is also scope for such a scheme to incorporate additional sources of kinetic energy from processes such as convection, which may not be fully modeled. To simulate this process in the UM, we use a similar strategy to Mason and Thomson (1992) but use a random streamfunction forcing field modulated by the square root of the local estimated energy dissipation, so that areas of high dissipation receive the highest energy input in the form of equivalent vorticity. The spectral stochastic kinetic energy backscatter scheme (SSBS) being tested at ECMWF (Berner et al. 2009; Palmer et al. 2009) is quite similar to SKEB2, and so we will draw a number of comparisons between these two systems in this paper.

a. The stochastic streamfunction forcing pattern

The scheme starts by generating a random streamfunction forcing pattern, similar to Berner et al. (2009), which is described by a triangular spherical harmonic expansion in the horizontal, that is,

\[ F^i = \sum_{n_{1}, m_{1} = n_{1}}^{n_{2}} \sum_{m_{2} = m_{1}}^{m_{2}} f^i_{m_{1} m_{2}} P^m_n(\mu) \exp(i[m\lambda + \epsilon(n, z)]), \]

where \( F^i \) is the streamfunction forcing and \( f^i_{m_{1} m_{2}} \) are complex spectral amplitudes at step \( i; P^m_n(\cdot) \) is the associated Legendre function for modes with zonal wavenumber \( m \) and degree \( n \) (noting that the number of zeroes between the poles is \( n - m \)); \( \epsilon(n, z) \) is a prescribed height-dependent phase shift; \( \lambda \) is the longitude; and \( \mu \) is sine of latitude.

Over time, each spectral mode is evolved following a first-order autoregressive process according to

\[ f^{i+1}_{m_{1} m_{2}} = [1 - \alpha(n)] f^i_{m_{1} m_{2}} + \sqrt{\alpha(n) g(n)} r_{m_{1} m_{2}}, \]

where \( \alpha(n) \) is a parameter between 0 (no stochastic forcing) and 1 (white-noise forcing) and is set as \( \alpha(n) = 1 - \exp[-\Delta t/\tau(n)] \), with the model time step \( \Delta t \). This type of application of the autoregressive technique with spectral modes was first reported in Li et al. (2008), although development of the method had also taken place in parallel at the Met Office. Our implementation of this scheme uses a uniform decorrelation time of \( \tau(n) = 2 \times 10^4 \) s (6 h) over all wavenumbers \( n \). This time scale is suitable for short- and medium-range forecasts; however, there is some evidence that a wave-dependent decorrelation time would be beneficial at these forecast time scales (Berner et al. 2009) and possibly also in applications of backscatter in longer-range forecasts as trialed by Doblas-Reyes et al. (2009).

Here \( g(n) \) controls the power in a single spectral mode \( n \), so as to give a net kinetic energy backscatter rate of unity when summed over all modes. If we let

\[ g(n) = F_0 \chi(n), \]

where \( \chi \) is a nondimensional function that has been deduced using the coarse-graining methodology applied to a cloud-resolving model of Shutts and Palmer (2007), to give the power in a single mode as \( \chi(n) = n^{-1.27} \), the amplitude \( F_0 \) is defined as

\[ F_0 = \left( \frac{4\pi a^2}{\Delta \Sigma^2 (\Gamma_{n_{1} n_{2}})} \right)^{1/2}, \]

where \( a \) is the radius of the earth, \( \Sigma^2 \) is the noise variance, and the sum \( \Gamma_{n_{1} n_{2}} \) is over the input wavenumber range \( [n_1; n_2] \) given by

\[ \Gamma_{n_{1} n_{2}} = \sum_{n = n_1}^{n_2} \frac{n(n + 1)(2n + 1)\chi(n)^2}{\alpha(n)}. \]

An example of the random streamfunction forcing pattern and its power spectrum is shown in Figs. 1a–1c. This forcing implies a typical global-mean wind speed tendency of around 2 m s\(^{-1}\) per day. For comparison, we also include the power spectra from two coarse-graining studies: the cloud-resolving model mentioned above (39 km \( \times \) 39 km) and using differences between high-resolution (T1279) and low-resolution (T159) forecasts made with the ECMWF Integrated Forecast System (IFS). These spectra in Fig. 1c imply direct forcing of synoptic and planetary scales by subgrid-scale eddies and are worthy of some comment.

It is likely that there are several different dynamical processes that contribute to this upscale energy transfer. In the context of tropical convection, Shutts (2008) showed...
that in a big domain, cloud-resolving model simulation, energy appears to grow spontaneously at all zonal wavenumbers after the onset of convection. The zonal extent of the model domain was about 40,000 km and thus represented the bulk of the tropical atmosphere. Energy growth was primarily in zonal wavenumbers less than 10. Similar numerical simulations (not described in Shutts 2008) that used a convective parameterization scheme failed to generate significant equatorially trapped wave motion.

Midlatitude convection is strongly constrained by the earth’s rotation, but mesoscale convective systems that form in regions of large convective available potential energy are associated with upper-level cloud “anvils” with diameters of order 1000 km. These cloud systems are characterized by strong anticyclonic vorticity, and their associated circulations are of synoptic scale. Since these weather systems are imperfectly resolved or parameterized at current EPS resolutions, uncertainty in their representation must be accounted for.

Another dynamical process that could be contributing to the streamfunction forcing power in Fig. 1c is direct, spectrally nonlocal energy transfer in the spirit of two-dimensional turbulence. The eddy-straining hypothesis for the forcing of large-scale blocking is an example of this and involves the transfer of energy directly from meridionally elongated troughs embedded in diffluent flows to blocking flow dipoles (Shutts 1983). Although this process is quite well represented at current EPS resolutions, there is still some potential need to account for uncertainty in its...
representation. In general, subfilter-scale energy can be absorbed directly into the synoptic- and planetary-scale flows through the action of flow deformation on eddies in a quasi-two-dimensional flow environment.

Variation of the spectral pattern in the vertical is described through a phase-shift function \( e(n, z) \) for each spectral mode independently, given by

\[
e(n, z) = \frac{N - n}{N} \frac{z}{z_{\text{ref}}},
\]

where \( N \) is the total input wavenumber range \( N = n_2 - n_1 \).

Maximum phase shift is achieved for the synoptic-scale waves between the surface and the reference level \( z_{\text{ref}} \), which is nominally chosen near the global-average tropopause height of 12 km. The variation in phase of the waves with height does not alter the amplitude of the individual spectral modes and thus the overall energy in the streamfunction forcing pattern is not affected. However, the assumed westward tilt with height (Fig. 1b) loosely matches the observed phase tilt of midlatitude baroclinic wave systems (Ebisuzaki 1991) and equatorial, convectively coupled waves in the troposphere. The benefit of this vertical structure in the streamfunction forcing is to support baroclinic instability in the midlatitudes and to promote vertical motion in the tropics with near-surface convergence below upper-troposphere divergence and vice versa. Similar ideas lie behind the development of initial perturbations based on singular vectors (e.g., Buizza and Palmer 1995) that exhibit rapid wave growth when vertical wind shear tends to steepen the phase line slope.

The streamfunction forcing pattern defined by (1) provides a unit rate of energy input and so should be modulated by a local energy dissipation rate calculated at each model time step. In this way, areas of high diagnosed energy dissipation receive the largest streamfunction perturbations. A small percentage of this dissipation rate, called the backscatter ratio (currently around 2.5%–3.0%), is assumed to be injected back into the explicitly resolved flow. To achieve this, the final streamfunction forcing field is calculated as the product of the square root of the dissipation rate and the original stochastically generated streamfunction forcing pattern.

\( b. \) Local dissipation rate calculation

The local dissipation rate comprises contributions from numerical advection schemes and convection parameterization. Numerical dissipation, occurring explicitly through diffusion and implicitly through interpolation in numerical advection schemes, destroys eddy kinetic energy near the grid scale of an atmospheric model (Shutts 2005). Semi-Lagrangian advection algorithms involve the interpolation of fields to the departure point, thus smoothing model fields and removing energy. Additionally, horizontal diffusion employed to explicitly smooth model fields is unlikely to result in a realistic energy drain rate given its artificial nature. The backscatter methodology simulates an upscale energy flow using the original Smagorinsky nonlinear diffusion scheme (Smagorinsky 1963) to estimate the numerical energy dissipation rate. This is a two-dimensional form of the more general Smagorinsky–Lilly turbulence closure equation used in large-eddy simulations of boundary layer turbulence (Mason and Thomson 1992). The true energy dissipation rate is the difference between the diffusive energy “drain” and backscatter energy transfer rate. The drain term associated with numerical dissipation \( D_{\text{num}} \) is given by

\[
D_{\text{num}}(\lambda, \phi, z) = (k_H \Delta)^2 D^3,
\]

with the shearing and tension strains

\[
D_S = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \quad \text{and}
\]

\[
D_T = \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y},
\]

such that

\[
D = \sqrt{D_S^2 + D_T^2}.
\]

Here \( \Delta \) is the model grid length and \( k_H \) is a numerical factor that is used to tune the dissipation rate so that the global-mean rate matches an estimated energy loss (approximately 0.7 W m\(^{-2}\) for model resolutions of 90 km/N144 and 60 km/N216) due to interpolation in the semi-Lagrangian algorithm. This scheme differs from the previous SKEB1 version, where the dissipation field was crudely assumed to be proportional to the local kinetic energy of the flow. Here, we target areas of high shear and tension strain (Fig. 1d).

Convective energy “dissipation” in the model is considered in a somewhat different way, with the emphasis being on the upscale energy transfer following kinetic energy production in buoyant updrafts. Convective parameterization is not concerned with the fate of this released kinetic energy, which for mesoscale convective systems may span the filter scale of the forecast model (Shutts 2005). Shutts and Palmer (2007) applied a coarse-graining methodology to cloud-resolving model simulations of deep convection and were able to characterize the dependence of the probability distribution function of convective warming on the strength of convective forcing. The coarse-grained momentum forcing has also been determined and used to compute an effective streamfunction forcing.
Building on this idea and the original SCV routine previously running in the MOGREPS suite, we have designed a scheme to estimate the amount of kinetic energy that might be injected by convective processes into the atmosphere at the mesoscale. The estimated dissipation rate relates changes in the vertical mass flux with height to work done on the environment. As convection schemes differ between models, our dissipation rate calculation in the UM is different from Berner et al. (2009). We calculate the product of the change in the density-weighted vertical mass flux between two model levels and the locally diagnosed column convective available potential energy (CAPE) using

\[
D_{\text{con}}(\lambda, \phi, z) = \frac{|M_{flx}^{k+1} - M_{flx}^{k-1}|}{\rho g \Delta z} \text{CAPE},
\]

where \(M_{flx}^k\) is the residual vertical mass flux at model level \(k\), \(\rho\) is the density, \(g\) is the gravitational constant, and \(\Delta z\) is the thickness between model levels \((k - 1)\) and \((k + 1)\). The CAPE factor offsets variations in the mass flux field between time steps because it generally evolves over longer time scales, typically tied to the diurnal cycle. As it is a vertically integrated quantity, it also produces a more coherent vertical structure in the energy dissipation field. These are important because convective complexes can move relatively slowly, compared to individual convective cells, and the release of kinetic energy should retain a reasonable spatiotemporal coherence.

The local numerical and convective energy dissipation rates are then combined into a single total dissipation rate using

\[
D_{\text{tot}}(\lambda, \phi, z) = \beta_n D_{\text{num}}(\lambda, \phi, z) + \beta_c D_{\text{con}}(\lambda, \phi, z).
\]

Scaling factors \(\beta_n\) and \(\beta_c\) are used to control the relative contribution of each dissipation rate to the final rate. Such flexibility is required to tune the response of the scheme under varying model resolutions, as the relative energy dissipation calculations can alter in a different way with model changes.

In the final step, the total dissipation rate is iterated through a 1–2–1 spatial smoother to smear out dissipation patterns. This becomes more important at higher model resolutions, as the streamfunction can become too fine-scale or noisy and not project suitably onto the streamfunction forcing pattern. A similar smoothing strategy was employed by Berner et al. (2009).

c. Modulated streamfunction forcing

The energy input rate of the streamfunction pattern \(F_\psi\) defined in (1) is modulated locally in grid space by the total dissipation rate calculated in (3) using

\[
\tilde{F}_\psi(\lambda, \phi, z) = [b_R \cdot D_{\text{tot}}(\lambda, \phi, z)]^{1/2} F_\psi(\lambda, \phi, z),
\]

where the backscatter rate \(b_R\) controls the local energy input rate that results in the largest streamfunction (or equivalently vorticity) being injected in areas of high diagnosed dissipation. Our experiments suggested that \(b_R\) should be increased with additional iterations of the spatial smoothing of \(D_{\text{tot}}\), to counteract the reduction in the dissipation rate through smoothing, as seen in the lower spectral power in the modulated streamfunction forcing compared to the random forcing pattern (Fig. 1c). The modulated streamfunction forcing field is then used to derive rotational wind components that are passed back to the model dynamics as part of the physics wind increment for each time step. Note that the kinetic energy spectrum of these wind increments is quite flat but peaks toward the small-scale end of the input wavenumber range, namely, \(n \sim 60\) (Fig. 1c).

To maintain model stability and physically sensible wind increments, the following extra steps are built into the scheme. A ramped damping filter is applied to the local energy dissipation fields near the poles. As these fields are calculated in grid space and the east–west grid spacing approaches zero near the poles in the UM, some spurious values typically appear and need to be constrained within reasonable limits. In the vertical dimension, we found that the vertical mass flux was often large in the boundary layer and led to a maximum in the energy dissipation field near the top of the boundary layer. A logarithmic-shaped damping filter below 2 km ensures that wind increments reduce to zero at the surface, eliminating instabilities or excessive noise in the boundary layer. Palmer et al. (2009) applied a similar filtering in the vertical in their updated stochastically perturbed parameterization tendencies (SPPT) scheme in the ECMWF EPS. They also included a filter in the stratosphere, which we apply by setting a top model level for the SKEB2 wind increments. Lastly, to avoid injecting too much rotational energy into the tropics, where the Coriolis parameter is weak, the rotational wind increments from SKEB2 are weighted by sine of latitude.

d. Velocity potential forcing

Divergent motion is also an important component of the atmospheric energy cascade. Hamilton et al. (2008) found enhanced divergence fields associated with organized weather systems when studying the mesoscale energy spectra of very fine-resolution global general circulation models. It has also been long understood that large-scale circulations in the tropics are generally initiated and driven by divergent outflow from deep convection (Trenberth et al. 1998). Despite there being little consensus in the literature about the relative magnitude of downscale
energy cascade into the mesoscale, as highlighted by Tung and Orlando (2003), it does seem prudent to include a divergent component in the backscatter scheme, as confirmed by our results discussed in this paper.

As a starting point, we consider the streamfunction forcing field also as a velocity potential forcing field, and thus we can generate divergent wind increments to complement the rotational components from the streamfunction field. It is important to note that the vertical structure of the stochastic streamfunction forcing field described above produces divergent (convergent) flow in the upper troposphere above convergent (divergent) flow near the surface. Furthermore, the divergent wind component is derived such that divergent (convergent) flow is centered on anticyclonic (cyclonic) rotational flow. To ensure that this field impacts mostly in the tropics, the divergent wind component is multiplied by cosine of latitude.

e. Model experiments and verification

A number of month-long trials were designed to test the impact of SKEB2 on the MOGREPS performance. The main aim was to quantify the improvement in forecast spread and determine whether the skill of the ensemble-mean (EM) forecast also improved. By using the full MOGREPS system in the trials, the impact of SKEB2 on the cycling of the ETKF was also tested. To separately assess the relative contribution of model error estimates from SKEB2 and RP, and initial condition uncertainty estimates via the ETKF, we also did various extra trials, each with a selection of these components switched on.

Two main periods were chosen for the trials, namely, a boreal spring/summer 30-day period during May 2008 and a winter period during December 2008. Runs were completed at N144L38 (FULL) and N216L70 (FULLHR) model resolution. The forecast length of these runs was set to 3 and 15 days, to mimic the MOGREPS-G and MOGREPS-15 suites. The latter comprises the Met Office contribution to the Global Interactive Forecast System–The Observing System Research and Predictability Experiment Interactive Grand Global Ensemble (GIFS-TIGGE) database. Bougeault et al. (2010) describe the GIFS-TIGGE archive, and Park et al. (2008) do an interesting comparison of the model forecasts from the various contributing centers. Given the current international interest in inter-EPS comparison, we attempt to gauge the impact of SKEB2 on MOGREPS skill by comparing the skill changes in our trials to those of the ECMWF EPS. This was done for a two-week period of parallel suite testing during October 2009. Lastly, a short two-week trial during February 2010 was done to compare the impact of SKEB2 (FULL_FLIP) to an “initial condition only” suite (ETKF_FLIP) during a period of high forecast uncertainty due to a blocking episode in northwest Europe.

For clarity, the names and details of the various experiments and their components are listed in Table 1. The same experiment naming convention is also used in the figures for ease of reference.

### Table 1. List of experiments and details.

<table>
<thead>
<tr>
<th>Name</th>
<th>Model resolution</th>
<th>Cases</th>
<th>Period</th>
<th>Forecast length (days)</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>FULL</td>
<td>N144L38</td>
<td>61</td>
<td>May 2008</td>
<td>3</td>
<td>ETKF, SKEB2, RP2</td>
</tr>
<tr>
<td>SKEB1</td>
<td>N144L38</td>
<td>61</td>
<td>May 2008</td>
<td>3</td>
<td>ETKF, SKEB1, RP2</td>
</tr>
<tr>
<td>SKEB2</td>
<td>N144L38</td>
<td>61</td>
<td>May 2008</td>
<td>3</td>
<td>SKEB2</td>
</tr>
<tr>
<td>SKEB2ETKF</td>
<td>N144L38</td>
<td>61</td>
<td>May 2008</td>
<td>3</td>
<td>ETKF, SKEB2</td>
</tr>
<tr>
<td>RP2</td>
<td>N144L38</td>
<td>61</td>
<td>May 2008</td>
<td>3</td>
<td>RP2</td>
</tr>
<tr>
<td>FULL15</td>
<td>N144L38</td>
<td>59</td>
<td>Dec 2008</td>
<td>15</td>
<td>ETKF, SKEB2, RP2</td>
</tr>
<tr>
<td>ETKF15</td>
<td>N144L38</td>
<td>59</td>
<td>Dec 2008</td>
<td>15</td>
<td>ETKF</td>
</tr>
<tr>
<td>SKEB215</td>
<td>N144L38</td>
<td>59</td>
<td>Dec 2008</td>
<td>15</td>
<td>SKEB2</td>
</tr>
<tr>
<td>FULLHR</td>
<td>N216L70</td>
<td>59</td>
<td>Dec 2008</td>
<td>3</td>
<td>ETKF, SKEB2, RP2</td>
</tr>
<tr>
<td>FULLHR_NOVP</td>
<td>N216L70</td>
<td>59</td>
<td>Dec 2008</td>
<td>3</td>
<td>ETKF, SKEB2, RP2</td>
</tr>
<tr>
<td>FULLHR_SSCL</td>
<td>N216L70</td>
<td>59</td>
<td>Dec 2008</td>
<td>3</td>
<td>ETKF, SKEB2 small scales only, RP2</td>
</tr>
<tr>
<td>NOSKEB</td>
<td>N216L70</td>
<td>59</td>
<td>Dec 2008</td>
<td>3</td>
<td>ETKF, RP2</td>
</tr>
<tr>
<td>FULL_FLIP</td>
<td>N144L38</td>
<td>29</td>
<td>Feb 2010</td>
<td>15</td>
<td>ETKF, SKEB2, RP2</td>
</tr>
<tr>
<td>ETKF_FLIP</td>
<td>N144L38</td>
<td>29</td>
<td>Feb 2010</td>
<td>15</td>
<td>ETKF</td>
</tr>
</tbody>
</table>

4. Results and discussion

Verification of spatial fields is presented by comparing the root-mean-square error (RMSE) of the control and EM forecasts as measured against surface station and radiosonde observations. Results from using model analyses as truth were consistent with the results shown below, so the discussion will focus on forecasts verified against observations. Probabilistic forecasts are verified against observations through the use of Brier scores and rank histograms.

a. Skill and spread of EPS

Spread of the EPS is calculated as a root-mean-square difference about the ensemble mean. Ideally, the magnitude of spread should match the error of the ensemble...
mean throughout the forecast range for all variables and levels. However, EPS spread tends to be too low in many systems (Buizza et al. 2005) and MOGREPS does exhibit this tendency (e.g., Bowler et al. 2009). A large part of the problem is that although the initial perturbations generate sufficient spread at 12-h lead time, the growth rate of spread in MOGREPS is too small to match the growth of EM forecast error at later lead times. This paper therefore examines the issue of increasing the EPS spread and its rate of growth during the forecast, through attempts to simulate model error more accurately.

It is clear that including SKEB2 in MOGREPS (FULL trial) increases the absolute ensemble spread (significant beyond the 90% level) as well as the growth rate of spread in all regions when compared to the current operational system SKEB1 (Fig. 2). See Bowler et al. (2009) for more details of the SKEB1 performance in MOGREPS.

FIG. 2. RMSE of control (light gray), EM (black), and ensemble spread (medium gray) for 3-day 250-hPa wind speed forecasts against radiosonde observations from the FULL (solid) and SKEB1 (dashed) trials during May 2008. Error bars show confidence estimated at 90% using the Monte Carlo method.
Although the impact of errors in observations can exaggerate spread deficiency in EPSs (Hamill 2001), the growth rate of spread in MOGREPS is clearly deficient. This is illustrated when comparing the MOGREPS system to the ECMWF EPS for a two-week period in October 2009 (Fig. 3). For this forecast parameter (500-hPa heights), the spread and forecast error values at 24-h lead time in the two EPS suites are similar, but the growth of spread in the ECMWF EPS is faster than the MOGREPS EPS, though this field in the ECMWF system appears somewhat overdispersive between 96- and 192-h lead time. Although SKEB2 (MOGREPS-15_PS) does improve the spread growth and EM error relative to SKEB1 (MOGREPS-15), the MOGREPS system remains underdispersive. Comparison of other variables [e.g., 850-hPa temperature, which is largely underdispersive in both EPS suites in the tropics (not shown)] illustrates the difficulty in both systems of universally matching EPS spread to EM forecast error.

The impact of SKEB2, in addition to increasing spread, is to achieve a ubiquitous decrease in the RMSE of the EM, most notable in the upper-tropospheric winds in the tropics and Southern Hemisphere (Fig. 2), approaching 90% significance at 3-days’ lead in the tropics. This could be partly a result from replacing the SKEB1 where perturbations were based solely on the kinetic energy of the flow and may not have always been physically realistic, thus adding excessive noise to the system. However, the results do show that the SKEB2 perturbations have made a widespread overall improvement to MOGREPS, similar to the findings in Berner et al. (2009) with the SSBS in the ECMWF EPS. This suggests that the assumptions underpinning SKEB2 lead to reasonable representations of model error and the useful generation of spread in the EPS.

b. Comparing the SKEB2 and ETKF contribution to ensemble spread

A desirable consequence of including a better representation of model error is that the initial condition (IC) perturbations do not need to be as large to generate the required spread (Berner et al. 2009). In MOGREPS the IC perturbations are generated by the ETKF system and their size is controlled by an inflation factor that ensures that the spread overall matches the RMSE of the EM in observation space at 12 h into the forecast (Bowler et al. 2008). Thus, the ETKF will adapt to changes in forecast error, and if the system works correctly, there should be an increase in the growth of spread when adding SKEB2 to the ETKF. A compensating reduction of RMSE of the EM will also be realized if the spread samples the forecast uncertainty realistically. Any degradation in EM RMSE

![Figure 3: RMSE of the control (black), EM (medium gray), and spread (light gray) of the MOGREPS-15 (solid line) and ECMWF EPS operational forecasts (dashed line with crosses) of 500-hPa heights initialized between 22 Sep and 6 Oct 2009 compared to the MOGREPS-15_PS parallel suite trial (dotted line with diamonds) running for the same period. Verification is done against radiosonde observations in the Northern Hemisphere.](image-url)
is undesirable, as this indicates that the spread may be associated with adding too much random noise to the system, for example, with the SKEB1 discussed above. In the trial of 3-day forecasts run twice daily from 1 to 31 May 2008, it is clear that representing model error in MOGREPS (FULL run) significantly improves the spread compared to the run with only ETKF initial perturbations for 250-hPa temperature forecasts over the Northern Hemisphere and tropics (Fig. 4). The ETKF on its own does not produce much growth of spread in the tropics, confirming the need for further work on initial condition perturbations in this region (Bowler et al. 2009).

The relative contribution of RP2 to ensemble spread was tested by running a separate trial using only RP2 (i.e., no IC perturbations) and another with RP2 removed from the FULL suite (SKEB2ETKF). Figure 4 shows that the RP2 scheme adds a significant component to the spread in the tropics when included in the full suite. The impact of RP2 elsewhere is generally quite small, which is consistent with the results shown in Bowler et al. (2008). However, the FULL system (which includes RP2) does have the best forecast verification throughout, supporting the case for including both SKEB2 and RP2. A case with no IC perturbations (SKEB2 only) is included in the figure for comparison. Similar results to those shown in Fig. 4 were found for other forecast parameters and also for the 15-day forecasts in the December trial (not shown).

It is instructive to understand the nature of the ensemble spread and its impact on probabilistic forecast skill. Figure 5 shows the rank histogram of the 250-hPa wind speed forecasts against observations from different experiments from the 15-day forecast trial in December 2008, namely, ETKF15, SKEB215, and FULL15. MOGREPS is generally underdispersive at all lead times out to 15 days, as illustrated by the U-shaped histogram, which is similar to the short-range results in Bowler et al. (2009). Including stochastic physics in the run (FULL15) results in a flatter rank histogram as expected. Observation error and conditional systematic forecast error can contribute somewhat to the U shape of rank histograms (Hamill 2001). The former issue is somewhat evident when comparing rank histograms equivalent to Fig. 5 but using model analyses as truth (not shown), as these are slightly flatter at short lead times (<48 h), that is, when forecast errors are still small and the verification is relatively more sensitive to errors in observations. However, the overall pattern and relative position of the graphs from the different experiments are consistent for both truth types.

As forecast lead time increases the spread continues to increase and the number of outliers drops off (Fig. 6). It is interesting that the spread in the tropics initially drops off with the ETKF15 experiment (with a corresponding increase in the number of outliers), consistent with the ETKF experiment for May 2008 shown in Fig. 4. The number of outliers in the FULL15 experiment does not increase as much as the ETKF15 experiment in the early part of the forecast, and then it starts to drop off at least a day earlier (Fig. 6). The utility of the spread in terms of probability forecasts is shown using the decomposed Brier score (Murphy 1973, 1986). Forecasts of surface temperature falling below a threshold of 0°C (high uncertainty in a decomposed Brier score sense), verified against stations in the Northern Hemisphere winter at 0000 and 1200 UTC, are shown in Fig. 7. SKEB2 improves the Brier score of the FULL15 trial at all lead times when compared to the ETKF15 experiment. This improvement is seen in both resolution and reliability components of the score. Although these improvements are small, a ubiquitous positive impact from SKEB2 on both reliability and resolution is seen for all thresholds and parameters studied (10-m wind speed; wind and temperature on standard
pressure levels). Corresponding results from the N216L70 trial (FULLHR vs NOSKEB) of forecast precipitation exceeding various thresholds shows an even greater positive impact of SKEB2 on the Brier score components (Fig. 8). These results are consistent with the increased spread and reduced outliers in the FULL15 experiment, which has a flatter rank histogram and captures more of the observations toward the tails of the distribution (Figs. 5 and 6).

Atger (1999) ascribed better reliability to increased spread (from a larger ensemble size) but also pointed out that improved resolution depends on the daily variation of spread. It is encouraging that resolution is improved in the FULL15 trial for both common and rare events, suggesting that SKEB2 increases the spread of the EPS not only with a wider range of samples but also with new information that provides a more realistic forecast PDF for individual events.

The ranked probability score (RPS) of Northern Hemisphere surface temperature essentially summarizes the Brier score over a range of thresholds (Fig. 9). The

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure5}
\caption{Rank histograms of the FULL15, SKEB215, and ETKF15 trial ensemble forecasts of 250-hPa wind speed measured against radiosondes in the Northern Hemisphere at lead times of (top) 24 and (bottom) 240 h.}
\end{figure}
FULL15 trial has a better RPS than the ETKF trial for all lead times, and the improvement appears to be significant at the 95% confidence level, at least for short lead times. This method of calculating confidence intervals for the median and its application to categorical statistics is described in Brown et al. (1997).

c. Tuning the scheme

Although the streamfunction forcing pattern power law, as deduced from coarse-graining studies, includes the large scales, one may argue that the original backscatter theory of Mason and Thomson (1992) implies that backscatter should only target the small scales. To investigate this issue, we have tested the impact of the input wavenumber range \([n_1; n_2]\) on the wave-dependent power spectrum \(g(n)\) in (2). Results of trials showed that the range \(n \in [5, 60]\) yielded the best impact in terms of ensemble spread growth and probabilistic verification scores. Berner et al. (2009) chose a wavenumber band to optimize the wavenumber-dependent error growth and reached similar conclusions about the backscatter forcing scales. Houtekamer et al. (2009) and Charron et al. (2010) describe a backscatter scheme in which rotational modes are forced between wavenumbers 40 and 128, which improves dispersion and reliability in their EPS. We have tested this same wave-range setting on N216L70 3-day forecasts (FULLHR_SSCL). There is no discernable difference in the RMSE of the EM between the FULLHR and FULLHR_SSCL trials, suggesting that by including the large scales in the backscatter forcing pattern is not detrimental to forecast skill; however, our wavenumber range does produce the greatest growth of ensemble spread (Fig. 10).

The other backscatter option tested here is to include increments of the divergent component of the wind. Our results for the December 2008 N216L70 trial do indeed show a general increased growth rate of spread compared to the experiment without divergent wind increments (FULLHR_NOVP) (Fig. 10). Although the increased spread is not quite significant at the 90% level, the impact is positive across all fields, and there is also a marginal improvement in the RMSE of the EM (Fig. 10) and a reduction in systematic error (not shown). As the proportion of divergent mode kinetic energy relative to the
total kinetic energy has been found to increase with higher wavenumbers (Hamilton et al. 2008) and that models often lack divergent mode kinetic energy at the mesoscales (Berner et al. 2009), it does make sense to also force the divergent modes. This is supported by Lindborg and Brethouwer (2007), who suggested that the forward cascade of energy in stratified turbulence is also produced by forcing in divergent modes. Thus, we believe that the divergent modes make the SKEB2 more physically realistic.

d. Skill of perturbed ensemble members

As pointed out by Hamill et al. (2000), initial condition perturbations added to the EPS should produce realistic samples of the forecast PDF. The same is true of perturbations used to account for model error. So far we have shown how SKEB2 improves the skill of the ensemble mean. However, if backscatter successfully represents missing physical processes, it should not significantly degrade the average verification scores of the individual perturbed ensemble members. This has been investigated by comparing the skill of perturbed member forecasts of various fields. The 250-hPa wind speed (Fig. 11) shows only a small increase in the average RMSE of perturbed forecasts of the SKEB215 experiment against the control (which has no perturbations of any kind). This corresponds to nothing more than the loss of an hour or two of predictive skill on average. This result is consistent with
Buizza et al. (1999) (their Fig. 11). For comparison, the same verification scores for the ETKF15 and FULL15 trials are shown on the same plot. The impact of initial perturbations on the skill of each ensemble member is considerably greater (up to a 12-h loss of predictive skill for this field), but the addition of SKEB2 (FULL15) shows again only a small change in average RMSE compared to ETKF15. The average-mean error results suggest that for this variable, SKEB2 has a positive impact on the small negative bias at all lead times (Fig. 11). This is particularly noticeable when compared to the ETKF15 run during the first 7 days of the forecast.

e. Evaluation of synoptic patterns

In this section we briefly look at some case study examples to investigate the impact of SKEB2 on particular forecast parameters. Generally, the trial datasets in this study are not long enough to draw statistically significant conclusions for some of these parameters; nevertheless, they do provide useful diagnostics of how the scheme is impacting the EPS forecast.

1) Blocking

An anticipated benefit of stochastic backscatter is improved simulation of tropospheric blocking (Shutts 2005; Berner et al. 2008). To investigate this we have used a PV–θ blocking index (Pelly and Hoskins 2003) that finds the difference between the average potential temperature θ on a potential vorticity surface of PVU = 2 (1 PVU = 10^{-6} m^{-2} K s^{-1} kg^{-1}) in two 15°-latitude bands north and south of 50°N.

The hemispheric values from the December 2008 ETKF15, SKEB215, and FULL15 trials (Fig. 12) agree largely with Fig. 7 in Pelly and Hoskins (2003), with the SKEB215 experiment exhibiting the most instances of blocking. The increased blocking frequency in SKEB215 could be related to the lower spread in this trial, as it does not include initial perturbations that might otherwise force the ensemble members into various regimes not present in the initial conditions. However, the FULL15 blocking frequency appears to match the observed frequency better than the ETKF15 results where the forecasts are poorest, for example, around the date line and 60°–90°E.

2) Forecast Juminess

The second issue is a case in which forecasts for early February 2010 over western Europe were extremely jumpy from one run to the next, which lasted for several days. See Zsoter et al. (2009) for a good discussion on jumpy control and EPS forecasts for the same region in early 2008. After an anomalously cold January in the United Kingdom (Eden 2010), the start of February saw some relatively warmer air being advected eastward over the country from the North Atlantic region. However, the available forecasts for the second week of February 2010 showed signs of a returning cold surface anticyclone, but run-to-run variability and large ensemble spread showed that there was a high uncertainty whether this event would occur.

Forecast area-averaged mean sea level pressure (MSLP) over the United Kingdom for 10 February 2010 (Fig. 13) showed a high probability of anticyclonic conditions about 12 days before, grew rapidly to cyclonic conditions over the next two days, dropped back to anticyclonic conditions at forecast days 8 and 9, returned to cyclonic conditions up until 5 days’ lead time, before finally settling on the anticyclonic pattern. As expected, the EPS EM was less jumpy because of averaging; nevertheless, it followed the trend of the control run. However, the unusually large range in the ensemble spread (more than 40 hPa at a lead time of 6 days
and up to 50 hPa at longer lead times!) gave a clear signal in this case that the forecast was very uncertain up until 4 days before the event. The FULL_FLIP runs do show less run-to-run variability than the ETKF_FLIP runs, showing the positive impact of SKEB2 on the ensemble spread and run-to-run consistency (Fig. 13). Note that the FULL_FLIP runs do not always have the higher spread, yet they still maintain better consistency. Verification of the surface pressure and wind at U.K. stations for this two-week period also indicates that SKEB2 had a positive impact on the ensemble spread and skill of the EM over the United Kingdom for this case (not shown).

5. Summary and conclusions

This paper has described the implementation of a Stochastic kinetic energy backscatter scheme (SKEB2) in the MOGREPS global suite run at the Met Office. The main aim of the scheme is to increase ensemble spread so that probabilistic forecast skill is improved. We found a consistent widespread improvement in nearly all aspects of MOGREPS with the introduction of SKEB2. The interaction between SKEB2 and the ETKF initial perturbation scheme in MOGREPS was also positive.

SKEB2 is calibrated through a power law derived from coarse-graining studies, and our experience with the scheme is that the best results are achieved when we retain the power in the low frequencies of the streamfunction forcing pattern. After this pattern is modulated by the energy dissipation fields, the resulting wind increments deliver a fairly flat power spectrum that peaks near the high-frequency end of the input power band. This backscatter of energy into the MOGREPS trials has improved the RMSE of the ensemble mean and has increased ensemble spread in such a way that probabilistic verification scores are improved for all variables and regions studied.

The original backscatter theory includes backscattering scalar fields, such as temperature and moisture. While work is ongoing at various centers to experiment with such methods, we believe that because geostrophic adjustment proceeds on time scales of less than one day, the wind increments soon grow accompanying temperature increments. Furthermore, the divergence forcing would generate temperature perturbations directly.

As model resolution continues to increase, and more of the atmospheric physical processes are adequately simulated, the utility of a backscatter scheme may change.

**FIG. 11.** Average (left) RMSE and (right) mean error of perturbed ensemble member 250-hPa wind speed forecasts for the FULL15 (dash), ETKF15 (dotted), and SKEB215 (dashed-dotted) experiments compared to the control run (solid) for 59 cases during December 2008 verified against model analyses and averaged over the globe.

**FIG. 12.** Observed instantaneous PV blocking index frequency at 50°N for the period 9 Dec 2008–6 Jan 2009 (gray line), with corresponding values from the ETKF15, SKEB215, and FULL15 trials.
However, there should always be some component of energy upscaling not fully captured by forecast models. To this end, the focus of development in backscatter schemes should be toward finding a generalized formulation that requires less tuning between model versions and provides a more robust framework for addressing missing or incorrect energy cascades within a numerical weather prediction model. It may then also be possible to use this type of scheme to improve estimates of uncertainty for other forecast applications, such as short-range convective-scale ensembles and long-range climate predictions.

Acknowledgments. We wish to acknowledge the contribution of our colleagues Sarah Beare, Neil Bowler, and Richard Swinbank to the code development of this system and for their valuable comments on the content of this paper. We also thank the reviewers of this article for their insightful comments and suggestions.

REFERENCES


Hamilton, K., Y. O. Takahashi, and W. Ohfuchi, 2008: Mesoscale spectrum of atmospheric motions investigated in a very fine