

A Spectral Stochastic Kinetic Energy Backscatter Scheme and Its Impact on Flow-Dependent Predictability in the ECMWF Ensemble Prediction System

J. BERNER

ECMWF, Reading, United Kingdom

G. J. SHUTTS

Met Office, Exeter, United Kingdom

M. LEUTBECHER AND T. N. PALMER

ECMWF, Reading, United Kingdom

(Manuscript received 14 November 2007, in final form 4 August 2008)

ABSTRACT

Understanding model error in state-of-the-art numerical weather prediction models and representing its impact on flow-dependent predictability remains a complex and mostly unsolved problem. Here, a spectral stochastic kinetic energy backscatter scheme is used to simulate upscale-propagating errors caused by unresolved subgrid-scale processes. For this purpose, stochastic streamfunction perturbations are generated by autoregressive processes in spectral space and injected into regions where numerical integration schemes and parameterizations in the model lead to excessive systematic kinetic energy loss. It is demonstrated how output from coarse-grained high-resolution models can be used to inform the parameters of such a scheme. The performance of the spectral backscatter scheme is evaluated in the ensemble prediction system of the European Centre for Medium-Range Weather Forecasts. Its implementation in conjunction with reduced initial perturbations results in a better spread-error relationship, more realistic kinetic-energy spectra, a better representation of forecast-error growth, improved flow-dependent predictability, improved rainfall forecasts, and better probabilistic skill. The improvement is most pronounced in the tropics and for large-anomaly events.

It is found that whereas a simplified scheme assuming a constant dissipation rate already has some positive impact, the best results are obtained for flow-dependent formulations of the unresolved processes.

1. Introduction

It is the central concern of numerical weather prediction (NWP) to predict synoptic scales of atmospheric motion as accurately as possible. However, since Lorenz's seminal work in the 1960s it has been believed that even very small uncertainties in the initial conditions or the prediction model will develop over time to synoptic-scale errors and limit the predictability of detailed weather evolution to 2 weeks or so (Lorenz 1963, 1969). These small uncertainties in the initial state and model formulation lead to forecast error and flow-dependent

predictability. To account for initial uncertainty it is now common practice to run not a single forecast but an ensemble of forecasts with slightly perturbed initial conditions (Molteni and Palmer 1993; Toth and Kalnay 1993; Buizza and Palmer 1995). For a perfectly reliable ensemble, we expect the true atmospheric state to be statistically indistinguishable from the ensemble members. However, one robust signature of all currently operational ensemble systems is that they are underdispersive; that is, if realistic initial perturbations are chosen, the best estimate of the true atmospheric state is on average more often outside the range of predicted states than statistically expected (e.g., Buizza et al. 2005). In other words, the trajectories of the individual ensemble members do not diverge rapidly enough to represent forecast error. This underdispersiveness might

Corresponding author address: Dr. Judith Berner, NCAR, P.O. Box 3000, Boulder, CO 80307-3000.
E-mail: berner@ucar.edu

arise in part from a misrepresentation of unresolved subgrid-scale processes (e.g., Palmer 2001). In a time of increased focus on forecasting extreme weather events, this underdispersiveness is certainly undesirable and needs to be remedied.

A source of model error is that conventional bulk parameterizations represent subgrid-scale processes with a single deterministic tendency, slaved to the resolved-scale flow, and neglect statistical fluctuations in fluxes and direct coupling between the resolved flow and unresolved subgrid-scale processes.

One manifestation of insufficiently represented subgrid-scale variability is that the kinetic-energy spectra of NWP and climate models drop off too steeply for wavelengths of less than 400 km and do not produce the observed $n^{-5/3}$ inertial-range power spectrum (Nastrom and Gage 1985), although other aspects such as the details of the numerical integration scheme may play a role (Hamilton et al. 1995; Skamarock 2004). Another source of model error is that some parameterizations, such as those of convection and mountain drag, have problems in their design and implementation. In nature, the flow over a mountain is associated with an increase in kinetic energy from turbulent eddies in the lee of the mountain, whereas in the mountain drag parameterization this effect is modeled by a drag coefficient, thus reducing the turbulent kinetic energy (see Palmer 2001). Furthermore, aspects of organized convection in the tropics and extratropics such as the convective outflow at upper tropospheric levels are often represented poorly by climate and NWP models, although these processes are associated with scales of hundreds of kilometers and should be well resolved (Shutts and Gray 1994).

The idea of stochastic kinetic energy backscatter of subgrid-scale fluctuations was originally developed in the context of large-eddy simulation (LES; Leith 1978; Mason and Thomson 1992) and is based on the notion that the turbulent dissipation rate is the difference between upscale and downscale spectral transfer, with the upscale component being available to the resolved flow as a kinetic energy source.

Frederiksen and Davies (1997, 2004) and Frederiksen and Kepert (2006) formulated dynamical subgrid-scale parameterizations based on eddy-damped quasi-normal Markovian (EDQNM), direct-interaction approximation (DIA) closure models and also a Markov model for the subgrid scale. They found that the kinetic energy spectra of their LES with subgrid-scale parameterization, including a stochastic backscatter term, agree well with those of the direct numerical simulations. Their work stresses the cusp behavior of the forcing at the truncation wavenumber.

Shutts (2005) argued that systematic kinetic energy loss is an underlying theme in both numerical integration schemes and parameterizations and adapted the backscatter concept to NWP use. For instance, errors in departure-point interpolation in semi-Lagrangian advection cause a net energy sink in NWP models, and kinetic energy released in deep convection does not sufficiently find its way into balanced flows and gravity wave generation. By injecting kinetic energy in regions of excessive dissipation due to processes associated with upscale error growth, this systematic kinetic energy loss is counteracted. This leads not only to a more skillful ensemble system in the medium range but also to improved seasonal forecasts and a reduction in model error in coupled simulations (Berner et al. 2008).

Here, we follow Shutts (2005) and propose the use of an improved stochastic kinetic energy backscatter scheme to account for model error in the European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble prediction system. This study assesses the impact of the new scheme on the kinetic-energy spectra, forecast error growth, flow-dependent predictability, precipitation forecasts, and probabilistic skill scores.

To simulate a stochastic kinetic energy source, we introduce random streamfunction perturbations with a prescribed power spectrum and use the local instantaneous dissipation rate as an amplitude function. Different from the Cellular Automaton Stochastic Backscatter Scheme (CASBS) of Shutts (2005), the random pattern generator is not based on a cellular automaton; rather, each spherical harmonic of the streamfunction forcing is evolved by a first-order autoregressive process. This way, we have full control not only over the spatial and temporal correlations but also over the spectral characteristics of the perturbations, such as the power law of the kinetic energy spectrum of the streamfunction forcing. In CASBS, these correlations were implicit in the cellular automaton rules, which we found difficult to manipulate. Here we propose to use output from coarse-grained cloud-resolving models to inform the parameters of the backscatter scheme (e.g., the power-law exponent of the forcing streamfunction).

Because we do not know the phase dependencies between the different modes in the subgrid-scale spectrum, we let each spherical harmonic evolve separately. However, multiplying the streamfunction pattern by the dissipation rate introduces phase dependencies into the effective forcing and brakes the scale invariance.

Due to the spectral nature of the pattern generator, we will refer to the scheme also as the spectral stochastic kinetic energy backscatter scheme, or SSBS. Because there are spatial and temporal correlations in the forcing pattern and dissipation rate, this

parameterization is nonlocal in space and time and thus represents a flow-dependent stochastic parameterization.

The manuscript is organized as follows: The SSBS algorithm is introduced in section 2 and the details of the numerical experiment setup are given in section 3. Section 4 describes the impact of the backscatter scheme on kinetic-energy spectra, the spread–error relationship, forecast error growth, skill, flow-dependent predictability, and precipitation forecasts. Section 5 summarizes the results from experiments with simplified dissipation rates. The sensitivity of the scheme to the details of the computation of the dissipation rate and the backscatter ratio are discussed in section 6. Final conclusions are drawn in section 7.

2. The spectral stochastic kinetic energy backscatter algorithm

This section describes the generation of the SSBS's effective streamfunction forcing, which is added at each time step to the right-hand side of the momentum equation. The effective streamfunction forcing is composed of a random homogenous streamfunction pattern weighted with the instantaneous total dissipation rate. The generation of the pattern is described in section 2a, the computation of the dissipation rates in section 2b, and the computation of the effective streamfunction in section 2c.

a. The stochastic streamfunction pattern

Let ψ' be a streamfunction forcing function expressed in a triangularly truncated spherical harmonics expansion given by

$$\psi'(\phi, \lambda, t) = \sum_{n=0}^N \sum_{m=-n}^n \psi_n^m(t) P_n^m(\cos \phi) e^{im\lambda}, \quad (1)$$

where λ and ϕ denote longitude and latitude in physical space and t time. In spherical harmonics space, m and n denote the zonal and total wavenumbers, N is the truncation wavenumber of the numerical model, and P_n^m is the associated Legendre function of degree n and order m . The spherical harmonics $Y_n^m = P_n^m e^{im\lambda}$ form an orthogonal set of basis functions on the sphere. If the ψ_n^m are nonvanishing for at least one $n < N$ and do not follow a white-noise spectrum, the streamfunction perturbations will be spatially correlated in physical space. Because the physical processes mimicked by this streamfunction forcing have finite correlation times, we introduce temporal correlations by evolving each spectral coefficient by a first-order autoregressive (AR1) process:

$$\psi_n^m(t + \Delta t) = (1 - \alpha) \psi_n^m(t) + g_n \sqrt{\alpha} \epsilon(t), \quad (2)$$

where $1 - \alpha$ is the linear autoregressive parameter, g_n the wavenumber-dependent noise amplitude, and ϵ a Gaussian white-noise process with mean zero and variance σ_z . In addition, we assume $\alpha \in (0, 1]$ (i.e., we exclude the case of a nonfluctuating forcing). The variance and autocorrelation of the AR1 in (2) are well-known quantities (e.g., von Storch and Zwiers 1999) and are given for the Markov process in (2) by

$$\begin{aligned} \langle \psi_n^m(t) \psi_n^m(t) \rangle &= \frac{g_n^2 \sigma_z}{2 - \alpha} \quad \text{and} \\ \frac{\langle \psi_n^m(t + \Delta t) \psi_n^m(t) \rangle}{\langle \psi_n^m(t) \psi_n^m(t) \rangle} &= 1 - \alpha. \end{aligned} \quad (3)$$

Here we interpret (2) as the discrete approximation of a Stratonovitch stochastic differential equation with an exponentially decaying autocorrelation function and a decorrelation time $\tau = \Delta t / \alpha$. The Stratonovitch interpretation is valid for systems where the noise represents continuous processes with decorrelation times smaller than the time increment. For such systems, the noise variance σ_z and α depend implicitly on the time increment, and the $\sqrt{\alpha}$ in front of the noise term guarantees that the noise decorrelates faster than the time step and fulfills the fluctuation–dissipation relationship. For a detailed discussion see, for example, Penland (2003a,b) and references therein.

We furthermore assume that the noise amplitudes follow the power law

$$g_n = b n^p, \quad (4)$$

where n^p is nondimensional and b the amplitude:

$$b = \left(\frac{4\pi a^2}{\sigma_z \Gamma} \alpha \Delta E' \right)^{\frac{1}{2}},$$

$$\text{where } \Gamma = \sum_{n=1}^N n(n+1)(2n+1)n^{2p}. \quad (5)$$

Here, a denotes the radius of the earth. As derived in the appendix, this choice of b is such that at each time step Δt a fixed globally averaged kinetic energy per unit mass

$$\Delta E' = \frac{1}{4\pi} \sum_{n=0}^N \sum_{m=-n}^n \frac{n(n+1)}{a^2} \langle |\psi_n^m(t + \Delta t)|^2 - |\psi_n^m(t)|^2 \rangle \quad (6)$$

$$= \frac{1}{4\pi} \sum_{n=0}^N \sum_{m=-n}^n \left(\frac{2}{\alpha} - 1 \right) \frac{n(n+1)}{a^2} \langle |\psi_n^m(t)|^2 \rangle \quad (7)$$

$$= \frac{\sigma_z}{4\pi a^2} \frac{1}{\alpha} \sum_{n=1}^N n(n+1)(2n+1)g_n^2 \quad (8)$$

is injected into the flow. As discussed in appendix, this derivation is only exact if certain assumptions about the interactions of the forcing with the dynamical source terms are made. These assumptions will not hold in general, but they are necessary to make the problem analytically tractable. However, because our numerical results agree very well with the analytically derived results (see next section), we are not overly concerned about the neglected terms.

In summary, a perturbation of the form (2) with noise amplitude (4) will generate streamfunction perturbations with the kinetic-energy spectrum:

$$E_n = \frac{\sigma_z}{4\pi a^2} \frac{1}{\alpha} n(n+1)(2n+1)g_n^2. \quad (9)$$

In the ensemble-mean sense, these perturbations will inject the globally averaged kinetic energy $\Delta E'$ given in (8) into the flow. We note that the change of total kinetic energy (7) does not solely consist of the injected kinetic energy $(1/4\pi) \sum_{n=0}^N \sum_{m=-n}^n [n(n+1)/a^2] \langle |\psi_n^m(t)|^2 \rangle$ but is modified by the factor $(2/\alpha) - 1$. In the appendix it is shown that this modification is noise induced and reflects the correlations between the total streamfunction and the streamfunction forcing at time t due to their mutual dependence on the streamfunction forcing at the previous time $t - \Delta t$. If there are no such correlations [i.e., $\alpha = 1$ in the evolution Eq. (2)], this factor equals one and the change in total kinetic energy equals that of the injected energy, assuming that the forcing increments are instantaneously injected at each time step. Secondly, we remark that if (4) is inserted into the equation for the kinetic-energy spectrum (9), the leading term in the expansion of the power law is

$$E_n \sim n^3 g_n^2 = b^2 n^{2p+3}, \quad (10)$$

so that a streamfunction forcing with power law n^p will result in a kinetic-energy spectrum of power law n^{2p+3} .

To calibrate the distribution of spectral power in the streamfunction forcing as represented by g_n in (2), we have adopted the coarse-graining methodology of Shutts and Palmer (2007) and Shutts (2008). For the most part, these papers focused on computing an effective coarse-grained term and the divergence of an eddy flux of sensible heat. In Shutts (2008) the coarse-grained momentum forcing was similarly computed as a Reynolds stress divergence term and used to estimate the kinetic energy generation in equatorially trapped disturbances due to deep convection. Here the same approach is applied to the simulation data analyzed in Shutts and Palmer (2007) and the resulting momentum forcing function is converted into a streamfunction

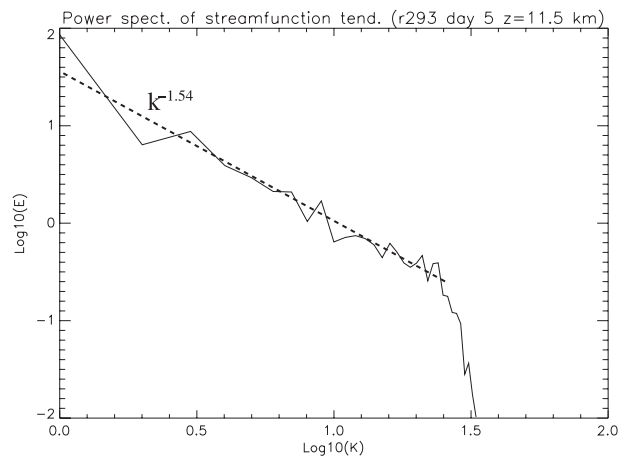


FIG. 1. Spectral power distribution for the streamfunction forcing evaluated at a height of 11.5 km on day 5 of the simulation described in Shutts and Palmer (2007). The $\log_{10} k$ axis is defined such that k is an integer wavenumber and $k = 1$ represents a wavelength equal to the domain size of 7680 km.

forcing. The resulting distribution of spectral power in the streamfunction forcing, plotted with respect to the zonal wavenumber, is given in Fig. 1. As can be seen, the assumed power-law behavior of g_n in (4) is consistent with the spectral power variation and translates to an exponent $p = -1.27$ when the different geometries of the simulations are taken into account.

Choosing this value for the noise amplitudes in (2), the forcing kinetic energy spectrum follows $E_n \sim n^{2p+3} = n^{0.46}$ (10). The perturbation energy input rate per unit mass $\Delta E/\Delta t$ is chosen to be $10^{-4} \text{ m}^2 \text{ s}^{-3}$, which roughly corresponds to a vertically integrated energy-forcing rate per unit area of 1 W m^{-2} .

The noise variance is set to $\sigma_z = 1/12$ and the autoregressive parameter to $1 - \alpha = 0.875$, so that each wavenumber has a decorrelation time scale of $\Delta t/\alpha = 6 \text{ h}$, where Δt is the model time step $\Delta t = 2700 \text{ s}$. Although the scheme allows different autoregressive parameters for different wavenumbers, reflecting the fact that small scales decorrelate faster than large scales, we use here for simplicity a wavenumber-independent α . Here, we pick plausible values for the parameters σ_z and α , but we plan to extend this work to obtain all parameters for SSBS scheme from coarse-grained high-resolution models as demonstrated for the forcing streamfunction slope.

The resulting streamfunction pattern (Fig. 2a) is isotropic and inherits roughly the spatial correlation scale from the largest wavenumber forced and the temporal correlations from the autoregressive parameter. Because of the positive power-law exponent, the kinetic energy injection is largest for large wavenumbers (Fig. 2b) but lacks the cusp behavior emphasized in Frederiksen and Davies (2004). In our experience, the hyperdiffusion of

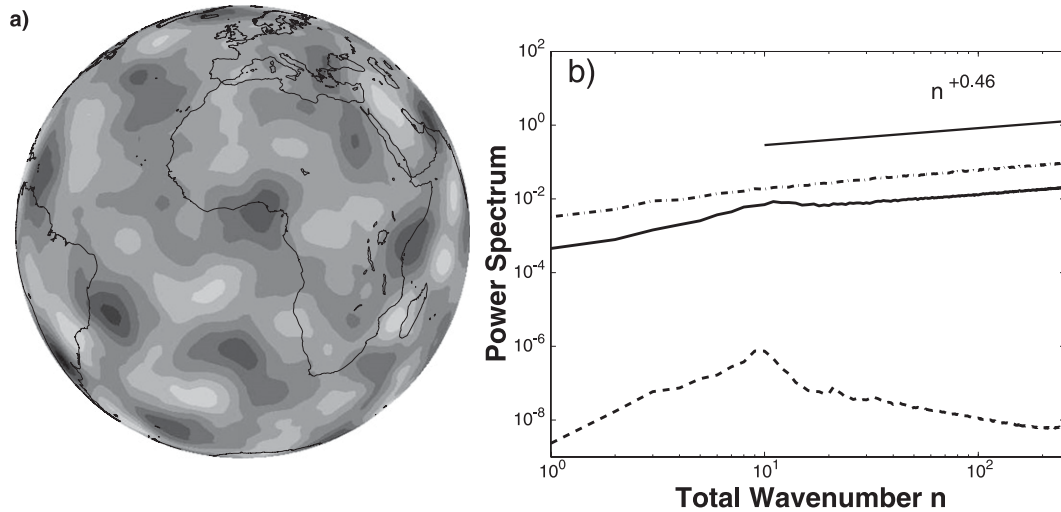


FIG. 2. (a) Instantaneous pattern of the stochastic streamfunction forcing. (b) Power spectrum of the total dissipation (dashed), the kinetic energy spectrum of the streamfunction pattern (dashed-dotted), and the effective streamfunction forcing (solid) after weighting with the total dissipation rate. Line denotes power-law behavior with a slope of +0.46.

the ECMWF model damps energy at the highest wavenumbers very efficiently, so such a cusp would be eliminated by the horizontal diffusion.

b. Computation of the total dissipation rate

True to the underlying backscatter idea, the streamfunction perturbations are weighted with the total instantaneous dissipation rate $D(\phi, \lambda, z, t)$, where z is the height. We include here dissipative processes on the synoptic and subsynoptic scale that are associated with systematic energy loss and upscale error growth, namely numerical dissipation, gravity/mountain wave drag, and deep convection. These are the same processes used in CASBS and discussed in greater detail in Shutts (2005). Here, we include only a brief summary and point out where changes have been made.

The dissipation from gravity/mountain wave drag is the only model variable and is directly taken from the physical parameterization package. The numerical dissipation is estimated as

$$D_N = -k \mathbf{u} \cdot \mathbf{u}', \quad (11)$$

where k is a factor greater than unity, \mathbf{u} is the horizontal wind vector, and \mathbf{u}' is the difference between the wind vector before and after the biharmonic horizontal diffusion is applied. The factor k augments the dissipation rate due to the model's hyperdiffusion scheme so that it notionally includes the contribution arising from the cubic interpolation in the semi-Lagrangian advection scheme.

Different from Shutts (2005), the dissipation rate from deep convection is now based on a convective mass-flux formulation:

$$D_{DC} = \frac{\delta M^2}{\rho^3 \beta^2}, \quad (12)$$

where M is the updraft convective mass flux rate in $\text{kg m}^{-3} \text{s}^{-1}$, δ the updraft detrainment rate in $\text{kg m}^{-3} \text{s}^{-1}$, ρ the density, and β an assumed detrainment cloud fraction of $\beta = 2.6 \times 10^{-2}$. The detrainment cloud fraction is chosen in such a way that the global vertically integrated dissipation per unit time and unit area from deep convection is about 2 W m^{-2} . To focus on the large-scale structure, the dissipation rate is subsequently smoothed by applying a spectral filter that completely retains the spherical coefficients for $n \leq 10$ and gradually reduces to zero coefficients for $10 < n < 30$.

The vertically integrated and zonally averaged annual mean total dissipation rate and its contributions are shown in Figs. 3 and 4. The zonally averaged dissipation rates are given in $\text{m}^2 \text{s}^{-2}$, whereas—for display only—the vertically integrated rates are weighted in such a way that their units, W m^{-2} , express power per unit area. The dominating contributor with an annual global mean of 1.99 W m^{-2} is deep convection. Its maxima are in the deep convective regions of the tropics, especially over Indonesia, but also in the vicinity of the jet stream at around 40°N and 40°S . The maxima around jet stream level are especially pronounced downstream of the Rocky Mountains and the Andes (Fig. 3) and are the

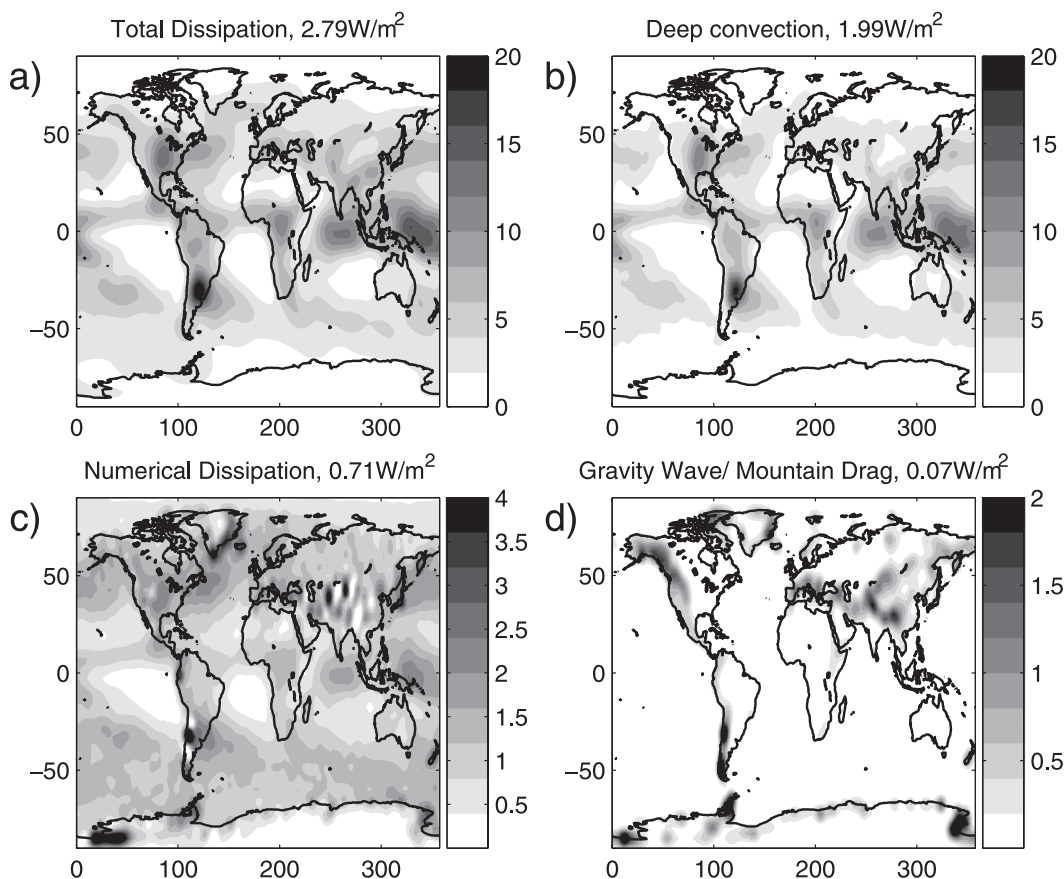


FIG. 3. (a) Vertically integrated annual-mean total dissipation rate per unit area, also shown with contributions from (b) deep convection, (c) numerical dissipation, and (d) gravity/mountain wave drag (W m^{-2}).

result of using mass fluxes for the computation of the dissipation rate. The annual global mean value of agrees well with that estimated by Steinheimer et al. (2007), who estimated a subgrid-scale energy conversion rate of 1.7 W m^{-2} for deep convection using the ECMWF convective parameterization scheme. At 0.71 W m^{-2} , the second-largest contribution comes from numerical dissipation. It is largest in the vicinity of the jet streams and over high orography like Antarctica, Greenland, the Himalayas, and the Andes. The dissipation from gravity and mountain wave drag is much smaller and occurs mainly in the lower tropospheric levels over orography.

c. The effective streamfunction forcing

To compute the effective streamfunction forcing, the spectral streamfunction perturbations are transformed to gridpoint space and then multiplied by the square root of the normalized dissipation rate $D(\phi, \lambda, z, t)$:

$$\psi'_{\text{eff}}(\lambda, \phi, z, t) = \sqrt{r \frac{D(\lambda, \phi, z, t) \Delta t}{\Delta E'}} \psi'(\lambda, \phi, t). \quad (13)$$

The backscatter rate r determines the fraction of the total dissipation rate that is scattered upscale and is effectively a tuning parameter. **It is chosen in such a way as to improve the slope of kinetic-energy spectrum, which will be further discussed in section 4a.** The weighting with the dissipation rate will modify the power-law behavior (10) of the pure streamfunction pattern. This is inevitable because the flow dependence of the dissipation rate will favor different scales at different times and hence break the scale invariance of the streamfunction pattern.

The dissipation spectrum and kinetic-energy spectrum of the effective streamfunction forcing are shown in Fig. 2b. Because of the spectral filtering, the dissipation spectrum peaks at wavenumber $n = 10$ and falls off for higher and lower wavenumbers. For wavenumbers $n > 20$ the spectrum of the effective streamfunction forcing inherits the power-law behavior of the pattern. For smaller wavenumbers it is a convolution of the pattern and the dissipation rate and has a shallow peak around $n = 10$.

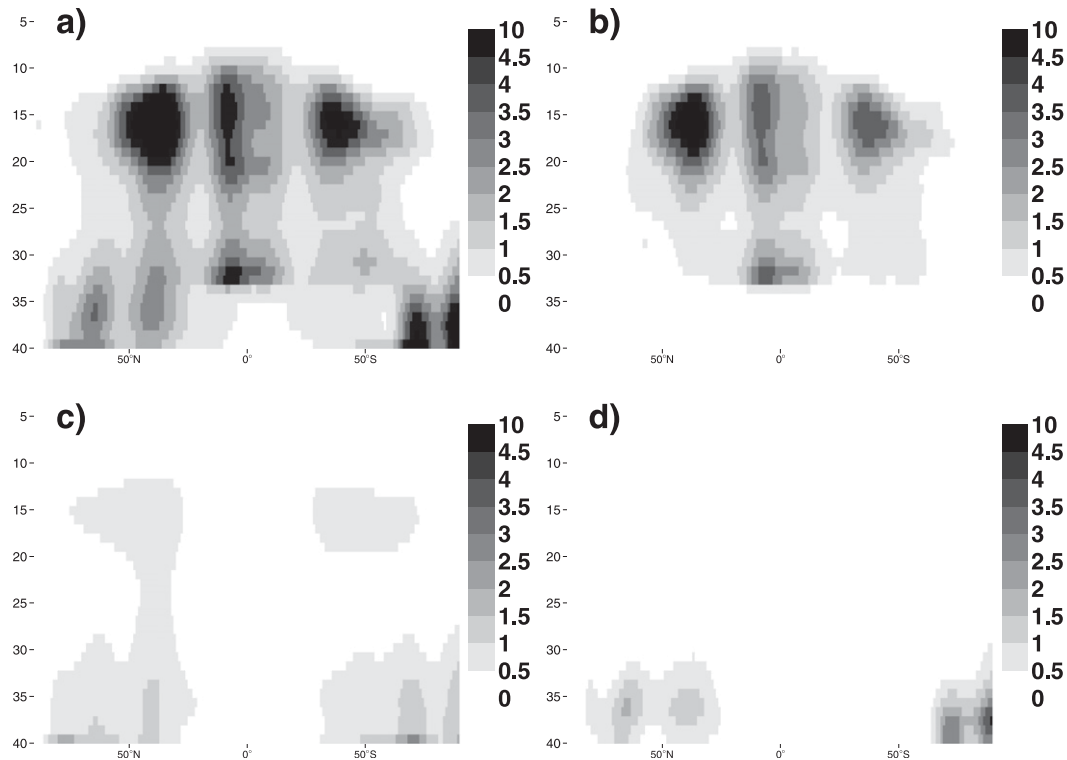


FIG. 4. (a) Zonally averaged annual-mean dissipation rate ($\text{m}^2 \text{s}^{-3}$), also shown with contributions from (b) deep convection, (c) numerical dissipation, and (d) gravity/mountain wave drag. The height is given in model levels, where levels 10, 20, and 30 correspond approximately to geopotential heights of 200, 540, and 880 hPa, respectively.

3. Description of forecast experiments

All results presented are produced by ECMWF's integrated forecasting system (IFS), model version CY31R1, which was operational from 21 November 2005 until 1 February 2006. The ensemble prediction

system was run with 50 ensemble members and started every eighth day for the period between 1 May 2004 and 26 April 2005, resulting in ensemble forecasts for 46 dates or "cases." Each forecast was run for 10 days at horizontal resolution T_L255 and with 40 vertical levels.

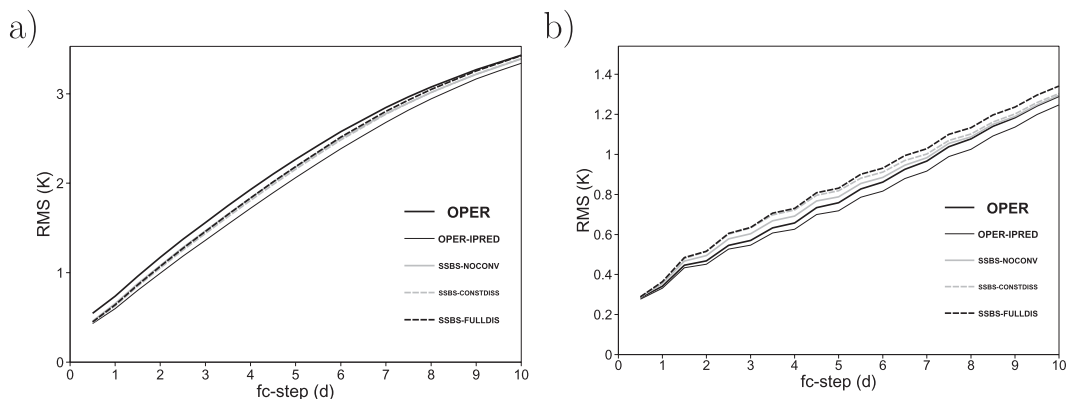


FIG. 5. Ensemble standard deviation in the (a) NH extratropics and (b) tropics for the experiments OPER (black thick solid line), OPER-IPRED (black thin solid line), SSBS-FULLDISS (black thick dashed line), SSBS-NOCONV (gray thick solid line), and SSBS-CONSTDISS (gray thick dashed line). See text and Table 1 for details on experiment setup.

TABLE 1. Experimental setup.

	Initial perturbations	SSBS	Dissipation rates
OPER	Operational	Off	NA
OPER-IPRED	Reduced by 15%	Off	NA
SSBS-FULLDISS	Reduced by 15%	On	Flow-dependent
SSBS-NOCONV	Reduced by 15%	On	Flow-dependent, no contribution from convection
SSBS-CONSTDISS	Reduced by 15%	On	Constant

The initial perturbation methodology is described in detail by Leutbecher and Palmer (2008). The initial perturbations are based on the leading singular vectors of the forecast model's tangent-linear propagator over a 48-h interval. Initial perturbations are obtained through sampling an isotropic Gaussian distribution in the subspace spanned by the leading singular vectors. In the extratropics of each hemisphere, the perturbations are based on the leading 50 singular vectors optimized for the region poleward of 30° latitude. In the tropics, initial perturbations are generally absent except for a region in the Caribbean and localized perturbations targeted on tropical cyclones (Puri et al. 2001). So far, attempts to obtain general perturbations for the tropical region with singular vectors were hampered by the fact that some of the leading singular vectors identified structures that do not result in growing perturbations in the nonlinear model (Barkmeijer et al. 2001).

The ensemble system was run in five configurations, which are summarized in Table 1. For each experiment, the spread around the ensemble mean is shown in Fig. 5 to motivate the experimental setup. Spread and error will be discussed in detail in section 4b. As a first experiment, the ensemble was run in its operational configuration (OPER). To ensure sufficient spread at the extended medium range, the initial perturbations in operational mode are somewhat too large, as is evident from the overdispersion of the ensemble in the early forecast ranges (see section 4b).

To test our representation of model error, the ensemble system was rerun with the full backscatter scheme, in which the dissipation rate includes contributions from deep convection, numerical dissipation, and gravity/mountain wave drag (SSBS-FULLDISS). Because the SSBS generates spread, the initial perturbations were gradually reduced in a series of experiments until the spread in the Northern Hemispheric extratropics at day 10 was approximately the same as in OPER (Fig. 5a). This allowed a reduction of the initial perturbations by 15%. To be able to assess the effect of the reduced initial perturbations separately from that of the stochastic backscatter scheme, an ensemble without SSBS was started from the same reduced initial per-

turbations as SSBS-FULLDISS. We refer to this experiment as OPER-IPRED.

In the OPER configuration the stochastic diabatic tendency (SDT) scheme of Buizza et al. (1999) is active, which attempts to sample subgrid-scale variability by perturbing the tendencies from the physical parameterizations. For a fair comparison this scheme is also activated in all other experiments, so that they only differ in the backscatter scheme. This is no contradiction because the two schemes represent different aspects of model error.

With the aim of understanding which aspects of the dissipation rate play an important role, two more experiments were performed. In SSBS-NOCONV only contributions from numerical dissipation and gravity/mountain wave drag but not from convection were included, and in SSBS-CONSTDISS a globally constant dissipation rate was assumed. For a fair comparison, these experiments were also started from the 15% reduced initial perturbations.

Because different geographical regions are dominated by different physical processes, we focus on three geographical areas: the Northern Hemispheric extratropics (20°–90°N), the Southern Hemispheric extratropics (20°–90°S), and the tropical band (20°S–20°N). The dynamical variables investigated are geopotential height in 500 hPa (Z_{500}), temperature in 850 hPa (T_{850}), and the zonal wind component in 850 hPa (u_{850}).

4. Impact of spectral backscatter scheme with full dissipation rate on flow-dependent predictability

In this section we propose the use of the spectral backscatter scheme in conjunction with reduced initial perturbations as an alternative to the operational ensemble configuration. Because the initial perturbations in OPER are tuned for optimal performance, this is the most challenging comparison. The following section will hence concentrate on the comparison of SSBS-FULLDISS and OPER and describe the impact of SSBS-FULLDISS on different aspects of model error, such as kinetic-energy spectra, the spread–error relationship, forecast error growth, probabilistic skill,

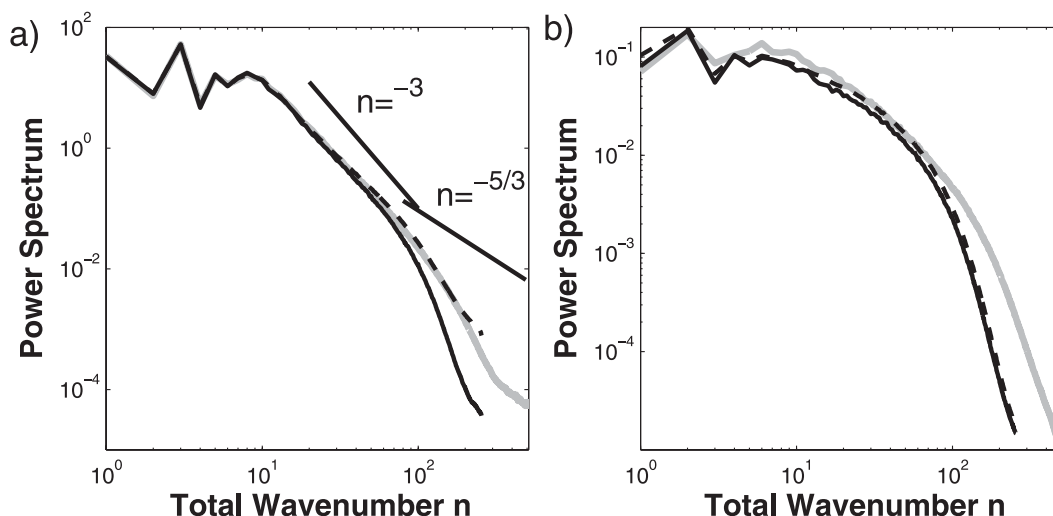


FIG. 6. Kinetic energy spectra for the (a) rotational and (b) divergent component of the flow for T_L511 analysis (gray solid), forecasts with the operational ensemble configuration (OPER; black solid), and the ensemble system with stochastic backscatter (SSBS-FULLDISS; black dashed). Lines denote power-law behavior with slopes of -3 and $-5/3$.

flow-dependent predictability, and skill of precipitation forecasts. To demonstrate that the improvements are indeed due to the SSBS and not the reduction of initial perturbations, section 4e contains a comparison of SSBS-FULLDISS to OPER-IPRED.

a. Kinetic-energy spectra

One manifestation of model error is that many global numerical weather prediction models do not capture the transition from an n^{-3} to an $n^{-5/3}$ kinetic-energy spectrum at wavelengths of 400 km—or equivalently total wavenumbers of $n \sim 100$ —observed in nature (Nastrom and Gage 1985). The ECMWF model is no exception to this and the verdict is open as to whether the cause is excessive numerical dissipation or the misrepresentation of unresolved physical processes, especially because some models capture this transition (e.g., Hamilton et al. 1995; Skamarock 2004; Janssen 2004, personal communication). This detail might be important because the limitations of predictability in some simplified models are entirely determined by the slope of the kinetic-energy spectrum (Lorenz 1969; Tribbia and Baumhefner 2004; Rotunno and Snyder 2008).

In the following we separately analyze the rotational and the divergent component of the total kinetic-energy spectrum. The rotational component is roughly two orders of magnitude larger than the divergent part and is the dominating contributor in the extratropics. For our best guess of the atmospheric state, the T_L511 analysis, this component has an n^{-3} spectrum for $n < 100$ and shows a moderate flattening for $n > 300$ (Fig. 6a). For OPER—run at horizontal resolution T_L255 —the

rotational part of the spectrum has a pronounced drop already at wavenumbers of around $n = 40$. Although SSBS-FULLDISS is run at the same resolution, the spectrum agrees now very well with the T_L511 analysis over a wide wavenumber range. At $n > 150$ there is a hint of a flatter spectrum, which is consistent with the observed spectrum. Although only the rotational part of the flow is forced, the kinetic energy spectrum of the divergent part is also slightly changed (Fig. 6b). Although it still drops off too steeply for larger wavenumbers $n > 80$, it agrees now very well with that of the T_L511 analysis for wavenumbers $20 < n < 80$. The improvement of the spectrum's divergent part in the synoptic scales is an indication that the model is now able to better produce and maintain divergent modes, which is especially important for tropical variability.

Our choice for the streamfunction forcing exponent, $p = -1.27$, from a coarse-grained high-resolution model seems to be justified empirically by noticing the good agreement between the kinetic-energy spectrum in the T_L255 forecast and the T_L511 analysis. Other choices of p in the range of $-3 < p < -5/3$ lead to a slightly worse agreement in terms of spectra but had no influence on the skill or spread-error relationship of the ensemble system.

The backscatter ratio r describes the fraction of the backscattered energy and hence determines the overall amplitude of the streamfunction forcing. It is difficult to measure this parameter from observations and is here primarily used as a tuning parameter. After fixing all other parameters, it is chosen to be as large as possible to remedy the underdispersiveness but as small as

necessary so as not to render the kinetic-energy spectrum unrealistic. Thus, the kinetic-energy spectrum can be used as a guide to determine the upper limit of the backscatter ratio. This strategy leads to a backscatter ratio of 2%, which is small compared to conventional backscatter formulations in LES. This will be further discussed in section 6.

b. Spread–error relationship as a function of forecast lead time

Probabilistic forecasting accounts for initial uncertainty by initializing an ensemble of forecasts from various perturbed states. The aim of this study is to create an ensemble system that is able to represent forecast uncertainty accurately at all lead times, that is, a system in which the difference between the perturbed forecast of the i th ensemble member $p_{i,j}$ and the ensemble mean forecast $\langle p \rangle_j$ will grow at the same rate as the difference between the ensemble mean forecast and the true atmospheric state a_j (e.g., Palmer et al. 2006). We will refer to $\|\langle p \rangle_j - a_j\|^2$ averaged over all cases j as the RMS error of the ensemble mean (or forecast error) and to $\|\langle p \rangle_j - P_{i,j}\|^2$, averaged over all cases j and members i , as spread. As an estimate of the true atmospheric state we will use the analysis, which is a very good estimator for the synoptic and larger scales but has significant errors for smaller scales (e.g., the lack of a $n^{-5/3}$ spectrum) as discussed in the last section.

For a perfect model, the flow-dependent initial uncertainty would be fully represented by the ensemble spread and thus spread and RMS error should grow at the same rate. However, when the initial perturbations are chosen to reflect our best knowledge of the magnitude of the analysis error, all current operational ensemble systems suffer from underdispersiveness (e.g., Buizza et al. 2005). Because insufficient spread is degrading the skill of an ensemble system, it is common practice to inflate the initial perturbations to ensure sufficient spread at the extended medium range. This is seen when the spread around the ensemble-mean forecast is compared to the RMS error of the ensemble mean as function of forecast lead time (Fig. 7). The ensemble system in the Northern Hemispheric extratropics is overdispersive¹ up to day 6 for Z_{500} and day 4 for u_{850} and underdispersive thereafter. For T_{850} , spread and RMS error agree well up to day 5, but the ensemble system is underdispersive for larger forecast lead times. The same holds for the Southern Hemispheric extra-

tropics although the crossover occurs at slightly different lead times. In the tropics, we focus on T_{850} and u_{850} , for which the ensemble system is markedly underdispersive (Figs. 7e,f). This is due to the absence of initial perturbations except in the vicinity of tropical cyclones and a general underactivity in the dynamics, probably caused by the lack of convectively coupled waves.

By including a representation of model error through the stochastic backscatter scheme, the error growth in the model is better at capturing model uncertainty and the initial perturbations can be reduced to values that are closer to our best estimate of the analysis error. This was done in such a way that the spread of Z_{500} at day 10 in the Northern Hemispheric extratropics is roughly the same in the ensemble with and without SSBS (Fig. 5a). In the extratropics, this reduces the spread of Z_{500} for the first 5 days and brings it closer to the ensemble-mean error (Fig. 7a). For u_{850} , the spread after day 5 is increased and is again in better agreement with the RMS error (Fig. 7b). For T_{850} , however, the spread introduced by the SSBS is not large enough to counteract the reduction in the initial perturbations, leaving the ensemble even more underdispersive for the entire forecast range. Interestingly, at day 10 the spread in T_{850} of the ensemble with SSBS has caught up to the spread of OPER.

In the tropical band, the spread increases at all forecast lead times, so that it is now in better agreement with the RMS error of the ensemble mean (Figs. 7e,f). This improvement is especially pronounced for u_{850} , in which the spread at day 2 has increased by 50% compared to OPER and the spread–error match is now almost perfect. It is worth noting that although the ensemble is now more dispersive, the RMS error is significantly reduced, especially for T_{850} . This is confirmed when running the scheme in multiyear integrations, which show a marked reduction in systematic error and a better representation of tropical waves (unpublished results and Berner et al. 2008).

A summary of the dispersion for SSBS-FULLDISS and OPER is given in Table 2 for the different geographical regions. The first column summarizes the dispersion of Z_{500} with regard to the RMS error of the ensemble mean and the second column describes the impact of SSBS-FULLDISS on the spread. Analogously, the other columns summarize spread and impact for other variables. Note that (as explained in the caption) the symbols change meaning depending on whether they describe spread or impact. Although the SSBS has overall improved the spread–error relationship, it is unable to introduce enough spread for a perfect match. This is an indication that there are aspects of model error that the current stochastic backscatter scheme does not account for.

¹ Here, the adjective “overdispersive” is and will be used in the sense of having a spread-to-error ratio of more than 1, not in the sense of an inherent characteristic of diverging trajectories.

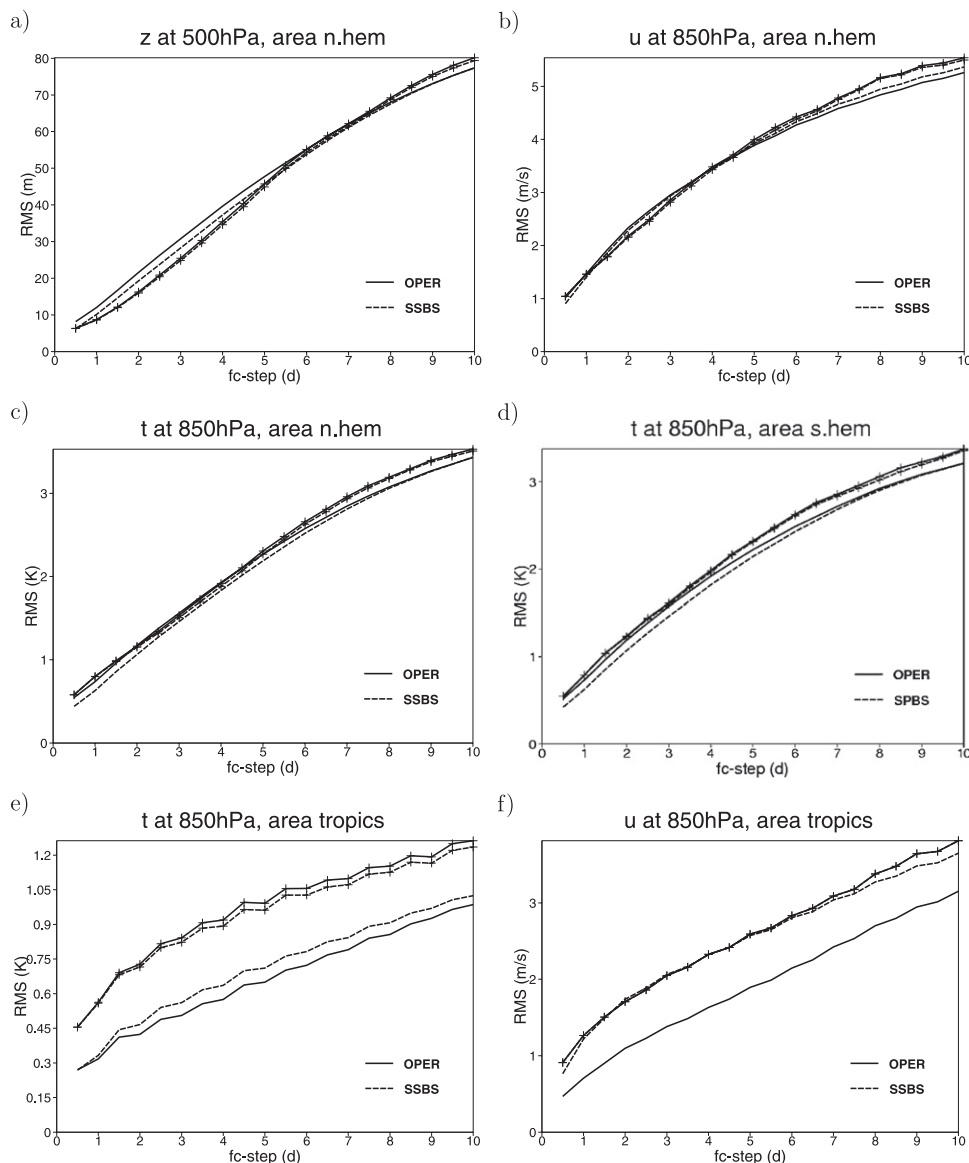


FIG. 7. RMSE of ensemble mean (lines with + symbols) and spread around ensemble mean (lines without symbols) for the operational ensemble (OPER; solid) and ensemble with stochastic backscatter (SSBS; dashed) for (a) Z_{500} , (b) u_{850} , and (c) T_{850} (all in the NH extratropics); (d) T_{850} in the SH extratropics; and (e) T_{850} and (f) u_{850} in the tropics. SSBS is short for SSBS-FULLDISS.

c. Error growth as function of wavenumber

To see to what degree the scale-dependent error characteristics of the ensemble-mean forecast can be captured by the model, we analyze the spread and RMS error for fixed forecast lead times as function of wavenumber. This points to the interpretation of the spread as model uncertainty and emphasizes that the goal of a perfect ensemble system is to capture the scale-dependent (i.e., wavenumber-dependent) forecast error by scale-dependent model uncertainty. If the model per-

fectly captured the characteristics of error growth, forecast error and spread would match for each wavenumber.

For OPER, the power spectrum of RMS error of the ensemble-mean forecast in 500 hPa is plotted as function of the total wavenumber n and for four different forecast lead times: 12 h, 2 days, 5 days, and 10 days (thin solid lines in Fig. 8). The spectra are averaged over all 46 cases. The maximum of the error spectrum moves with increasing forecast time from wavenumbers 20–30 at 12 h to larger scales and peaks at day 10 at wavenumbers 8 and 9. The spectrum is initially shallow,

TABLE 2. Dispersion of spread (first, third, and fifth columns) and impact (second, fourth, and sixth columns) of the stochastic backscatter scheme in an ensemble with reduced initial perturbations (SSBS-FULLDISS). In the columns summarizing spread, the symbols are as follows: + = overdispersive at all forecast lead times; − = underdispersive at all forecast lead times; < = underdispersive for shorter but overdispersive for longer forecast lead times; > = overdispersive for shorter but underdispersive for longer forecast lead times. In the columns summarizing impact, the symbols describe the impact of the scheme: + = positive impact at all forecast lead times; − = negative impact all forecast lead times; < = negative impact for shorter but positive impact for longer forecast lead times; > = positive impact for shorter but negative impact for longer forecast lead times. The rows are for different geographical regions: the NH extratropics (20°–90°N), the SH extratropics (20°–90°S), and the tropics (20°S–20°N).

	Z_{500}		u_{850}		T_{850}	
	Spread	Impact	Spread	Impact	Spread	Impact
NH	>	+	>	+	−	−
SH	>	>	>	+	−	−
Tropics	−	+	−	+	−	+

with a broad maximum across wavenumbers 10–100, but becomes more peaked for increasing forecast time. The RMS error of the ensemble mean grows for each wavenumber until a forecast lead time of 5 days. For a forecast lead time of 10 days, the error decreases compared to that for day 5 for wavenumbers of $20 < n < 200$. This is probably a consequence of analyzing the statistics of the ensemble mean; at longer forecast times, errors present in the individual ensemble members in

this wavenumber range seem to be averaged out by taking the ensemble mean.

The spectrum of spread in OPER (thick solid lines) also shows upscale error growth but is much more peaked and has more amplitude for shorter forecast lead times. The maximum for 12 h is shifted to wavenumbers 10–20, and for day 2 the spectrum has a peak at $n = 15$ rather than a broad spectrum. For longer forecast lead times the spectra agree better, but it is evident that at day 10 the spread underestimates the RMS error for all wavenumbers.

There is a large discrepancy between RMS error and spread for small wavenumbers $n < 5$, where the spread underestimates the forecast error. This feature is robust and might be a signature of systematic model error, possibly in the tropics. A more detailed analysis of this feature is beyond the scope of this study and will be discussed elsewhere.

There is also a marked difference for large wavenumbers $n > 80$ or so, where the spectra of the spread fall off much more steeply than those of the RMS error. Although analysis errors are not negligible in this wavenumber range, this discrepancy is attributable to model error, namely the fact that the kinetic-energy spectra for large wavenumbers drop off too steeply.

For SSBS-FULLDISS, the spectra of spread and forecast error agree much better for both the synoptic and subsynoptic range (Fig. 8b). There is still too much power in the spectrum of spread (thick dashed lines) at forecast lead times 12 h and 2 days, but it has now a

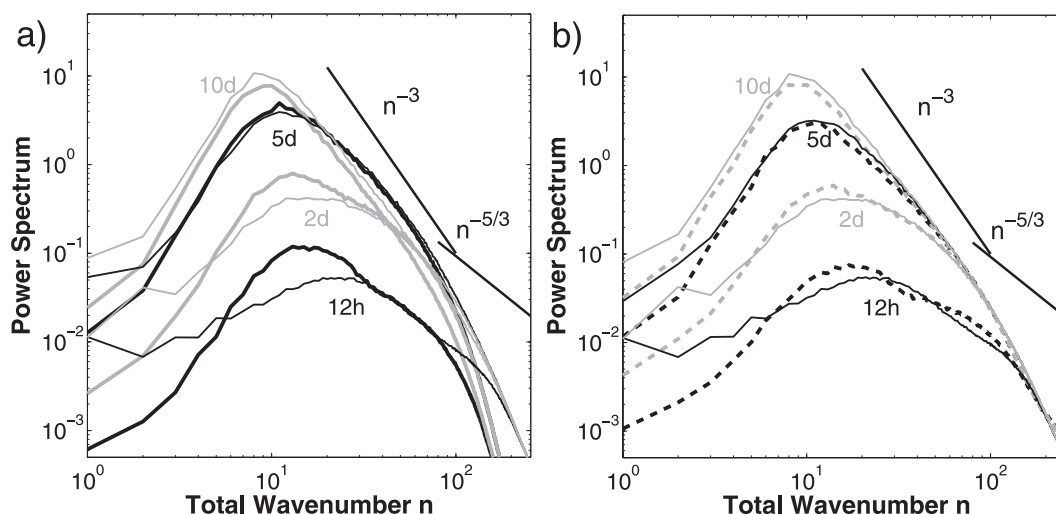


FIG. 8. Power spectrum of the error of the ensemble-mean forecast (thin solid lines) and spread (thick lines) in 500 hPa for fixed forecast lead times of 12 h, 2 days, 5 days, and 10 days for (a) the operational ensemble configuration (spread in OPER: thick solid line) and (b) the ensemble system with a stochastic backscatter scheme and reduced initial perturbations (spread in SSBS: thick dashed line). SSBS is short for SSBS-FULLDISS. Lines for forecast lead times of 12 h and 5 days are shown in black and for 2 days and 10 days in gray. See text for details.

TABLE 3. Definition and meaning of different scores and a general definition of skill score. In the definition of the Brier score, p_i denotes the forecast probability, o_i the dichotomous (yes/no) occurrence of the event i , and N the number of events. In the definition of the ranked probability score, K is the number of forecast categories; $P_{i,k} = \sum_{j=1}^k p_{i,j}$ is the predicted probability cumulative distribution function (CDF) for the forecast category k ; and $O_{i,k} = \sum_{j=1}^k o_{i,j}$ is the observed probability CDF, where $p_{i,j}$ is the probabilistic forecast for the i th event to happen in category k and $o_{i,j}$ is an indicator (0 = no, 1 = yes) for the observation falling into category k . In the definition of the ignorance score, $p_{i,j}$ is the probability of outcome j according to the probabilistic forecast i ; o_i is the corresponding actual outcome. In the definition for skill score, “clim” stands for the climatology of an event and $\text{Score}_{\text{clim}}$ denotes the score when using climatology as the forecast.

Score	Definition	Meaning
Brier score (BS)	$\text{BS} = \frac{1}{N} \sum_{i=1}^N (p_i - o_i)^2$	Measures the mean squared error of the probability forecast
Area under the relative operating characteristics (ROC)	Plot of the hit rate vs the false alarm rate; compute the area under the curve	Measures the ability of the forecast to discriminate between two alternative outcomes, thus measuring resolution
Ranked probability score (RPS)	$\text{RPS} = \frac{1}{N} \sum_{i=1}^N \frac{1}{K-1} \sum_{k=1}^K (P_{i,k} - O_{i,k})^2$	Measures the squared difference in probability space when there are multiple categories
Ignorance score (IGN)	$\text{IGN} = -\frac{1}{N} \sum_{i=1}^N \log_2 p_{i,o_i}$	Logarithmic score that measures the amount of data compression required for the forecast to represent the truth
Skill score	$\text{Skill Score} = 1 - \frac{\text{Score}}{\text{Score}_{\text{clim}}}$	Relative skill of the forecast compared to climatology

broader structure and the maxima are closer to the wavenumber band where the RMS error has its maxima. This is mainly the consequence of the reduction in the initial perturbations. For wavenumbers $n > 80$, the drop in the spectrum of spread has been remedied, consistent with the change in kinetic-energy spectra reported above. The error in the smaller scales saturates to that at day 10 and the error and spread spectra agree now almost perfectly for $n > 30$ (Fig. 8b). In summary, the characteristics of error growth are better captured by an ensemble with reduced initial perturbations and stochastic backscatter.

d. Skill of ensemble system

As demonstrated, the stochastic backscatter algorithm increases the spread and thus partly remedies the inherent underdispersiveness of the ensemble system. However, increasing the spread by introducing stochastic perturbations is an almost trivial task because at each time step each ensemble member is randomly perturbed and thus the ensemble trajectories diverge more quickly.²

Hence, the ultimate test is to evaluate if the skill of the stochastically perturbed ensemble system is better

than that of the ensemble without SSBS. For this purpose we compare in this section the skill of the operational ensemble (OPER) with that of the spectral backscatter scheme in conjunction with reduced initial perturbations (SSBS-FULLDISS). To separate the effects of the backscatter scheme from those of the reduced initial perturbation, the next section compares the operational ensemble and stochastic backscatter schemes started from the same reduced initial perturbations. This will confirm that the improvement in skill is indeed the consequence of the scheme and not the reduced initial perturbations.

Because different skill scores focus on different aspects of the ensemble system, it can lead to a one-sided verdict to base the evaluation on a single measure and variable. Hence, we evaluate the scheme by a whole range of different skill scores and compute them for more than one dynamical variable. A brief summary of the different scores and the general definition of a skill score are given in Table 3. For further details, we refer the reader to the general verification literature (e.g., Katz and Murphy 1997; Roulston and Smith 2002; Jolliffe and Stephenson 2003).

Table 4 compares the skill of SSBS-FULLDISS to OPER with regard to the Brier skill score (BSS), area under the relative operating characteristics (ROC), the ignorance skill score (ISS), the ranked probability skill score (RPSS), the percentage of outliers, and the rank histogram at 48 h. The Brier skill score, ROC area, and ignorance skill score are defined with regard to anomalies exceeding a threshold. Here, we pick a threshold of

² Theoretically, it is possible to damp a system by introducing flow-dependent random perturbations; then the noise-induced drift would be equal and opposite to the deterministic drift (e.g., Sardeshmukh et al. 2001; Berner 2005). But generally this is rare and the stochastic parameterization tends to act as a net forcing for the resolved flow.

TABLE 4. Impact of stochastic backscatter scheme and reduced initial perturbations (SSBS-FULLDISS) when compared to OPER for NH and SH extratropics and tropics. Probabilistic skill scores are the BSS, ROC area, ISS, RPSS, percentage of outliers, and rank histogram at 48 h (RH48). Symbols represent the following impacts: +++ = very strong positive impact at all forecast lead times; ++ = strong positive impact at all forecast lead times; + = positive impact at all forecast lead times; ○ = neutral impact; − = negative impact all forecast lead times; < = negative impact for shorter but positive impact for longer forecast lead times; > = positive impact for shorter but negative impact for longer forecast lead times.

Score	Threshold	NH			SH			Tropics		
		Z_{500}	u_{850}	T_{850}	Z_{500}	u_{850}	T_{850}	Z_{500}	u_{850}	T_{850}
BSS	$> +1.5\sigma$	+	+	+	+	+	+	++	++	++
	$< -1.5\sigma$	+	+	+	+	+	+	++	++	++
ROC	$> +1.5\sigma$	+	+	+	+	+	+	+	+	+
	$< -1.5\sigma$	+	+	+	+	+	+	+	+	+
ISS	$> +1.5\sigma$	++	++	++	+	+	+	++	++	++
	$< -1.5\sigma$	++	++	++	+	+	+	++	++	++
RPSS		+	+	+	+	+	+	++	++	++
% outliers		+	++	<	+	+	<	+	++	++
RH48		++	+	+	+	+	○	+	+	+

1.5σ , where σ denotes the climatological standard deviation, and look at the frequency of events above the $+1.5\sigma$ threshold. Qualitatively, the results are not sensitive to the threshold chosen for the computation of the skill scores, and although the improvements are smaller for smaller thresholds, they are still evident for thresholds of 0σ (correctness of sign of anomaly). We also find the results to be symmetric (i.e., the impact on positive and negative anomalies are very similar). The distinction between a positive (+), a strongly positive (++), and a very strong positive (+++) impact is made subjectively, but quantitative results for some of the skill scores are shown later. Note that the impact is only classified as positive if the skill improvement is seen at all forecast lead times. If the skill is decreased for shorter but increased for longer forecast lead times, the impact is denoted by the symbol <. Analogously, the symbol > indicates that the skill is increased for shorter but decreased for longer forecast times. For further definitions see the caption for Table 4.

Overall the impact of the stochastic scheme in combination with reduced initial perturbations is good to very good. The best results are obtained for the ignorance skill score in the Northern Hemispheric extratropics (for all variables) and the tropical region (for all variables and all skill measures). The impact on skill is throughout positive, the only exception being the percentage of outliers of T_{850} in the extratropics at small forecast lead times of less than 2 days.

In the Northern Hemispheric extratropics the results show uniformly a small but significant improvement.

For well-tuned ensemble prediction systems, such as the one used here, even small improvements are difficult to achieve. Figure 9 shows the (left) RPSS and (right) ISS for the variables Z_{500} , u_{850} , and T_{850} as function of forecast time. The improvement is especially visible in the ISS but is also evident in the RPSS, BSS, and ROC area (latter not shown). The scheme has a slightly larger positive impact on T_{850} and u_{850} than on Z_{500} . Importantly, the skill of the ensemble improves at all forecast lead times.

To assess the statistical significance of the results, confidence intervals have been computed for the score differences of RPSS and ISS. The confidence intervals are estimated with a bootstrap method that obtains many different realizations of N cases by randomly sampling with replacement from the set of $N = 46$ start dates. The increased uncertainty due to temporal correlation of the score differences is represented by a moving block resampling procedure in which the block size depends on the lag-1 and lag-2 correlations of score differences for individual dates, following a proposal of Wilks (1997). For each N -case sample, the average scores are computed and the confidence intervals are based on the empirical distribution of score differences for the 5000 realizations. A confidence level of 95%, which is used here, implies that if the mean score difference lies above the confidence interval, the probability that this (i.e., picking a particular set of cases) happened by chance is below 5%. Using this approach, all improvements in Fig. 9 are statistically significant. The error bars for the RPSS increase as function of forecast lead time and for a lead time of 10 days are about as wide as the line width, which is why they are not plotted.

The results for the Southern Hemispheric extratropical skill scores are similar to those presented for the Northern Hemispheric extratropics. As an example we show the rank histograms for Z_{500} in Figs. 10a,b. For a perfectly reliable ensemble, the histogram would follow the uniform distribution shown by the thin dotted line. The rank histogram at day 2 has the typical U shape caused by the fact that the analysis falls disproportionately often in the most extreme bins. For Z_{500} , there is an additional peak for the middle ranks, which reflects the overdispersion of the ensemble system at this forecast range. For the ensemble with SSBS the overpopulation for the most extreme bins is slightly reduced; more markedly, the double-well structure of the middle bins is flattened out and the histogram is very close to the theoretically desired uniform distribution. The same is seen for the rank histograms in the Southern Hemisphere extratropics (Fig. 10b).

The percentage of outliers as function of forecast lead time measures the occupancy in the two most extreme

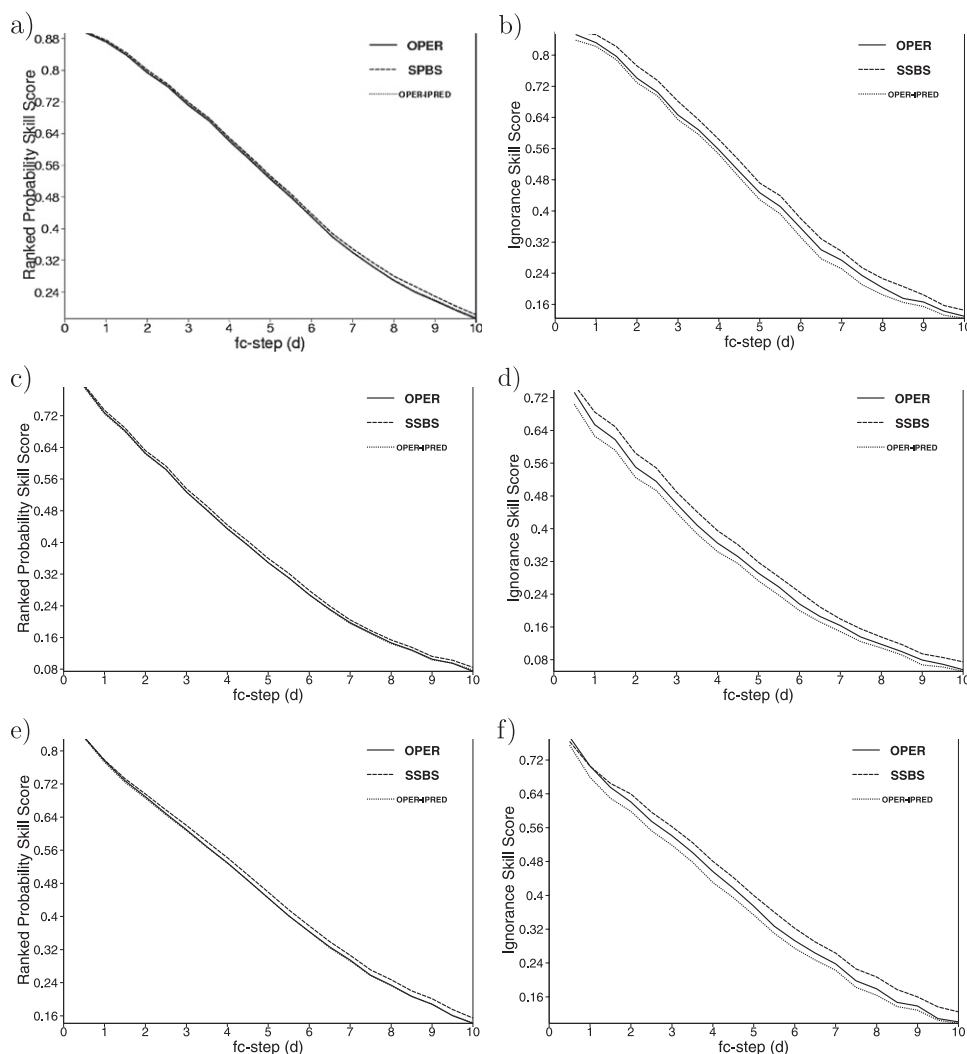


FIG. 9. (left) RPSS and (right) ISS for (a),(b) Z_{500} , (c),(d) u_{850} , and (e),(f) T_{850} in the NH extratropics. The ISS is computed for events exceeding a threshold of $+1.5\sigma$, where σ denotes the climatological standard deviation. The solid line indicates the operational ensemble (OPER); dashed, the ensemble with stochastic backscatter (SSBS, short for SSBS-FULLDISS); and dotted, the operational configuration with reduced initial perturbations (OPER-IPRED). In (a), (c), and (e) the solid and dotted lines are hardly distinguishable.

bins. For a perfectly reliable 50-member ensemble the expected number of outliers would be 2/51 at all forecast lead times (denoted again by the dotted line). For T_{850} , the percentage of outliers increases for the first 36 h in both the Northern and Southern Hemisphere extratropics (Figs. 10c,d), resulting in the only $<$ impact in Table 4. The reason for this is the reduction in spread for T_{850} discussed above. However, the percentage of outliers is reduced after the first two days although the ensemble system is still underdispersive at these forecast lead times (Fig. 7c). This is an indication of a more skillful ensemble, and indeed the skill scores for T_{850} are improved even at the shortest forecast lead times (Fig. 9).

The best improvements by far are seen for the tropical band. Figure 11 shows the BSS for events exceeding a threshold of $+1.5\sigma$, RPSS, rank histogram, and percentage of outliers for u_{850} . Independent of the nature of the measure, the stochastic backscatter scheme improves the skill of the forecast markedly. As discussed, tropical variability and spread are significantly underestimated by the ensemble without SSBS and the improvements seen in the tropics are due to both a greatly increased ensemble spread and a reduction in the RMS error (Figs. 7e,f). The increased spread results in fewer occupations of the most extreme bins, and the rank histogram at 48 h follows the theoretical distribution for a

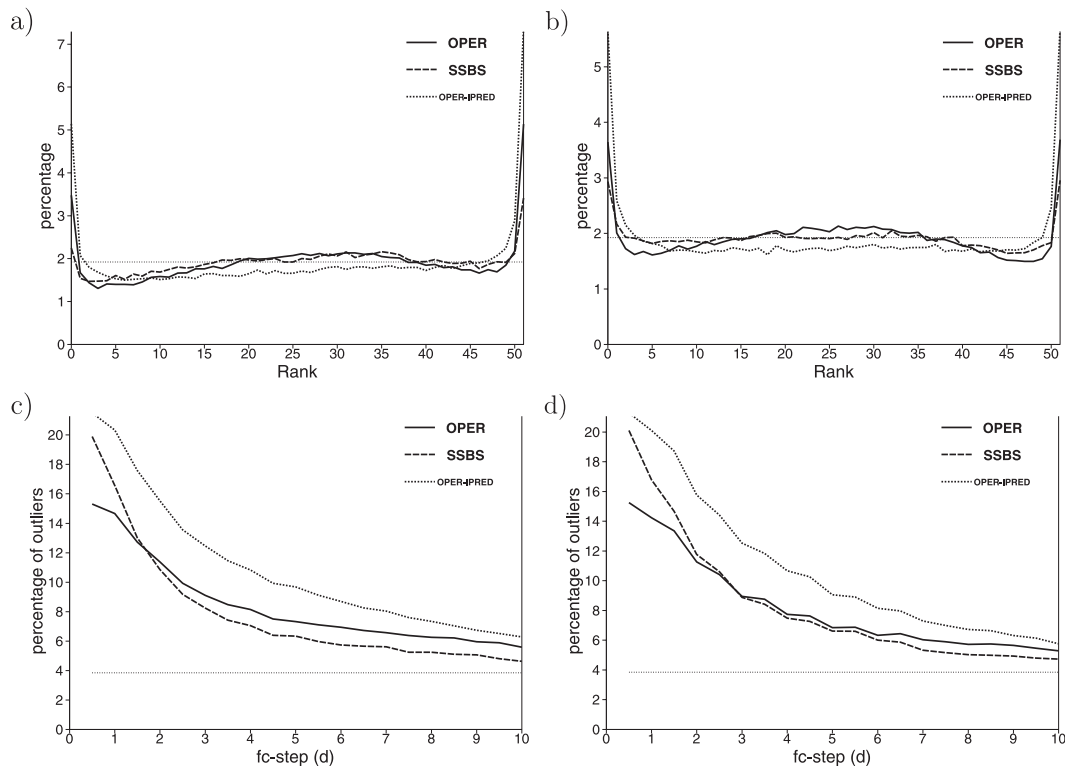


FIG. 10. (a),(b) Rank histograms at 48 h for Z_{500} and (c),(d) percentage of outliers for T_{850} in the (left) NH and (right) SH extratropics. Lines indicate ensembles as in Fig. 9.

perfectly reliable ensemble markedly well (Fig. 11c). This is reflected by the percentage of outliers as function of forecast lead time. At day 2, the percentage of outliers has been reduced from 24% to 7% and at day 10 from 12% to 5%, which is a reduction of more than 50% (Figs. 11d,e).

It should be mentioned that this study does not take analysis error into account, which can influence the percentage of outliers and the shape of rank histograms (Saetra et al. 2004). Currently, work at ECMWF is under way to account in the future for the uncertainty in the analysis (Leutbecher and Hagel 2006, personal communication).

e. Reduction of initial perturbations versus SSBS

To separate the impact of the stochastic backscatter scheme from that of the initial perturbations, we compare in this section the operational and stochastic backscatter scheme started from the same reduced initial perturbations: OPER-IPRED versus SSBS-FULLDISS. As seen in Fig. 5 the spread in both experiments is very similar for a forecast lead time of 12 h but increases more rapidly in the experiment with SSBS-FULLDISS. Examples for the improvement in skill of SSBS-FULLDISS over OPER-IPRED are shown in Figs. 9–11 and summarized in Table 5. The performance of SSBS-FULLDISS is very positive for all skill scores, variables, and regions. The

improvement is most pronounced for the ignorance skill score (Figs. 9b,d,f), the percentage of outliers (Figs. 10c,d and 11d), the rank probability skill score (Figs. 9a,c,e), and, in the tropics, the Brier skill score (Fig. 11a).

The improvement in skill is even larger than for the comparison of SSBS-FULLDISS to OPER analyzed in the last section. In part, this can be explained by the difference in spread between SSBS-FULLDISS and OPER-IPRED. Wherever OPER-IPRED is significantly underdispersive, the additional spread in SSBS-FULLDISS will improve the skill as long as the RMS error is not increased. This, for example, is the reason why the percentage of outliers is so positively affected (Table 5). In general, the relationship between skill, error, and spread is more complex and would need to be analyzed in detail for a thorough understanding of why a particular skill score improves.

The results in this section clearly show that the increase in skill of SSBS-FULLDISS reported in section 4d is the consequence of the stochastic backscatter scheme and not of the reduced initial perturbations.

f. Flow dependence of spread

Ideally, a perfectly reliable ensemble would predict predictability (i.e., the flow-dependent error growth). If

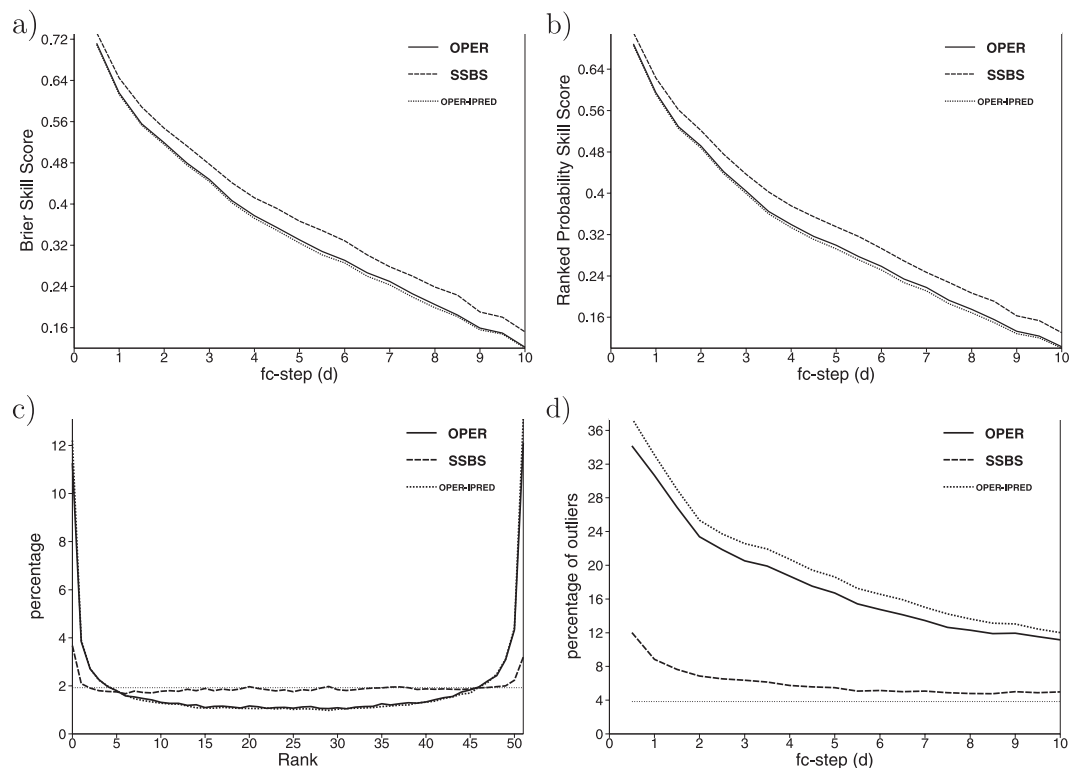


FIG. 11. (a) BSS, (b) RPSS, (c) rank histogram at 48 h, and (d) percentage of outliers for u_{850} in the tropics. The BSS is computed for events exceeding a threshold of $+1.5\sigma$, where σ denotes the climatological standard deviation. Lines indicate ensembles as in Fig. 9.

the flow is predictable, it should be associated with a small ensemble-mean error and small spread. If it is unpredictable, we would expect a large spread, reflecting a large ensemble-mean error. To see to what degree the ensemble system is able to model flow-dependent predictability, we compute spread-reliability diagrams (Leutbecher and Palmer 2008).

For this purpose, all data for a fixed forecast lead time are categorized by the predicted ensemble standard deviation and partitioned in equally populated bins. Then, the ensemble-mean error is computed and for each

bin plotted against the ensemble standard deviation (Fig. 12). To focus on the flow-dependent variation of the spread and not regional variations, the spread-reliability diagrams for SSBS-FULLDISS and OPER are computed for limited geographical regions. For a perfectly reliable ensemble system, the flow-dependent spread would model the RMS error perfectly and all points would lie on the diagonal.

Figure 12 displays the spread-reliability diagrams of Z_{500} in the Northern Hemispheric extratropics and T_{850} in the tropics for forecast lead times of 48 h and 120 h. In

TABLE 5. As in Table 4, but for the impact of the ensemble with stochastic backscatter and reduced initial perturbations (SSBS-FULLDISS) compared to the operational ensemble started from the same reduced initial perturbations (OPER-IPRED).

Score	Threshold	NH			SH			Tropics		
		Z_{500}	u_{850}	T_{850}	Z_{500}	u_{850}	T_{850}	Z_{500}	u_{850}	T_{850}
BSS	$> +1.5\sigma$	+	+	+	+	+	++	++	+++	+++
	$< -1.5\sigma$	+	+	+	+	+	++	++	+++	+++
ROC	$> +1.5\sigma$	+	+	+	+	+	+	+	+	+
	$< -1.5\sigma$	+	+	+	+	+	+	+	+	+
ISS	$> +1.5\sigma$	+++	+++	+++	+++	++	++	+++	+++	+++
	$< -1.5\sigma$	+++	+++	+++	+++	++	++	+++	+++	+++
RPSS		+	+	++	+	++	++	+++	+++	++
% outliers		+++	+++	+++	+++	+++	+++	++	+++	+++
RH48		+	+++	++	+	+++	++	+	+	+

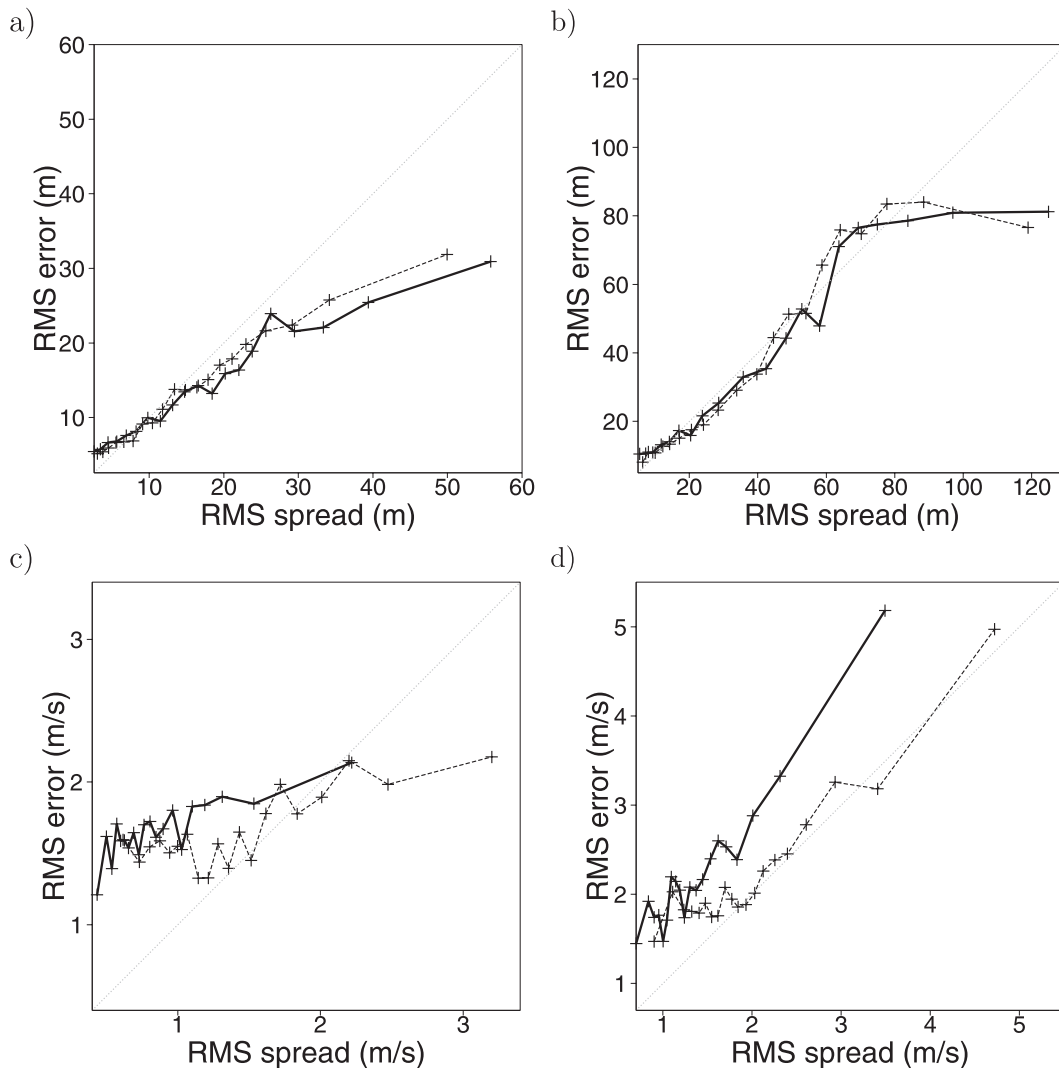


FIG. 12. Spread-reliability diagrams for (a),(b) Z_{500} in the NH and (c),(d) u_{850} in the tropics at the forecast lead times: (left) 48 and (right) 120 h; solid = operational ensemble (OPER); dashed = ensemble with stochastic backscatter (SSBS-FULLDISS).

the Northern Hemispheric extratropics, the relationship between spread and RMS error improves with forecast lead time and matches well for a lead time of 120 h except for the largest-spread bins. For shorter lead times, however, OPER is significantly overdispersive for the large-spread bins (curves lie below the diagonal). For SSBS-FULLDISS, we see a small improvement in the spread reliability at 48 h in that the points are slightly closer to the diagonal.

In the tropics, we see that for u_{850} OPER is underdispersive for all spread bins and forecast times. With the stochastic backscatter, this underdispersion is greatly reduced for all but the smallest spread bins. Many points now fall onto the diagonal, signifying a markedly improved flow-dependent predictability.

g. Precipitation skill

Although the main focus of this study is the verification of dynamical variables, it is of practical interest if the improvement in skill also carries over to precipitation. For this purpose we include a verification of 24-h accumulated precipitation forecasts as described in Rodwell (2006). To avoid the challenges associated with computing area averages from station data, the forecast precipitation is bilinearly interpolated from the model grid to individual station locations.

The Brier skill scores for events in which 24-h accumulated precipitation exceeded 10 mm in the Northern Hemisphere extratropics or 20 mm in the tropics are shown in Figs. 13a,b. A significance test using a two-sided

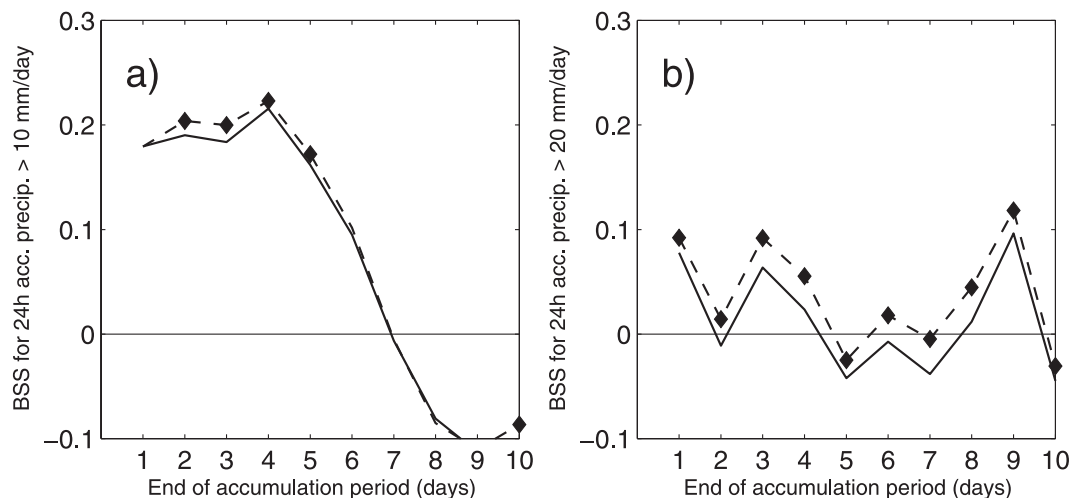


FIG. 13. The Brier skill scores for the event in which 24-h accumulated station/location point precipitation exceeds (a) 10 mm in the NH extratropics or (b) 20 mm in the tropics. Diamonds indicate statistical significance at the 90% confidence level; solid = operational ensemble (OPER); dashed = ensemble with stochastic backscatter (SSBS-FULLDISS).

first-order autoregressive t test was performed. With the SSBS there is a significant improvement in the Brier skill score in the Northern Hemisphere extratropics from day 2 until day 5 at the 90% confidence level. For longer forecast lead times, the skill of precipitation forecasts in the ensemble with SSBS is also better but has less confidence. The Brier skill scores for other precipitation thresholds and other geographical regions are also improved (not shown). In the tropics, there is generally little skill in precipitation forecasts and the Brier skill score is often less than zero. However, for extreme precipitation events in which the accumulated 24 h-precipitation exceeds 20 mm, the SSBS improves the skill significantly at all lead times (Fig. 13b). In some instances, the SSBS is able to shift the forecast from an unskillful to a skillful prediction (in the sense of $BSS > 0$; e.g., in the accumulation period ending at day 6).

5. Spectral backscatter schemes with simplified dissipation rates

One fundamental strength of the backscatter scheme is its physical motivation to link the streamfunction perturbations to the flow-dependent dissipation rates. However, in reality very little is known about the dissipation rates of various processes in nature and to what degree the model is able to and should capture them. Research on how to best compute the dissipation rates and which contributions should be included is necessary and currently underway.

To get a first insight into the importance of the various contributions to the dissipation rate and its spatial structure, we have performed additional experiments.

We are interested in the importance of including dissipation from deep convective processes and would like to know if the flow dependence of the streamfunction forcing is essential for a good performance of the backscatter scheme. For this purpose, the ensemble system is rerun twice more: first by including only contributions from numerical dissipation and gravity/mountain wave drag, not from deep convection (SSBS-NOCONV); and secondly with a constant dissipation rate that has no horizontal or vertical structure (SSBS-CONSTDISS; see also Table 1). The experiments are started from 15% reduced initial perturbations, and the backscatter ratio r is chosen in such a way that the spread at day 10 in the Northern Hemispheric extratropics is as close as possible to that of SSBS-FULLDISS (Fig. 5a). For the no-convection experiment this strategy results in a backscatter ratio of $r = 3\%$ (all other parameters are kept the same) and for SSBS-CONSTDISS in a globally constant dissipation backscatter rate of $rD = 6 \times 10^{-3} \text{ m}^2 \text{ s}^{-3}$.

Generally, the performance of both SSBS-NOCONV and SSBS-CONSTDISS is good and both have a positive impact on skill when compared to OPER and even more so when compared to OPER-IPRED (not shown). In the extratropics, the difference among SSBS-NOCONV, SSBS-CONSTDISS, and SSBS-FULLDISS is small but significant at the 95% confidence level. Figure 14a shows the ignorance skill score for SSBS-FULLDISS, SSBS-NOCONV, SSBS-CONSTDISS, and OPER for Z500 in the Northern Hemispheric extratropics. Both SSBS-NOCONV and SSBS-CONSTDISS outperform the operational ensemble at least up to day 6, but neither of them performs as well as the backscatter scheme with the

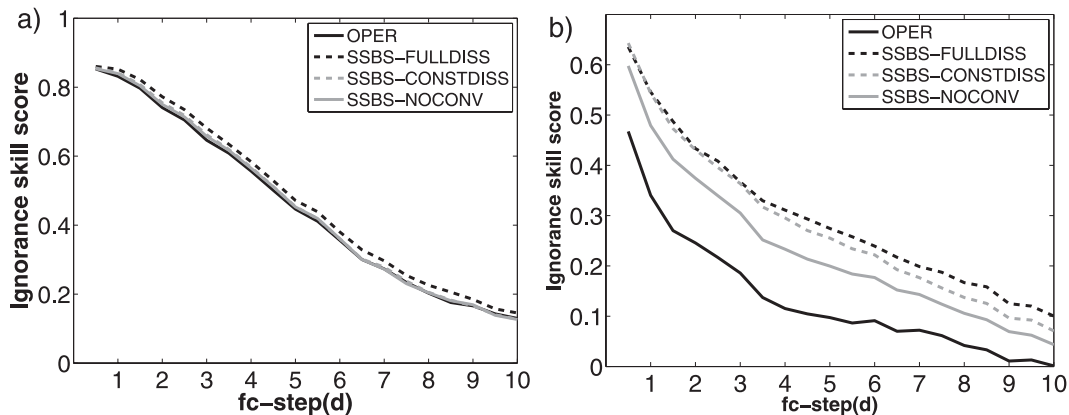


FIG. 14. ISS for (a) Z_{500} in the NH extratropics and (b) u_{850} in the tropics for the operational ensemble and experiments with simplified dissipation rates (see text and Table 1): OPER (black solid), SSBS_FULLDISS (black dashed), SSBS_CONSTDISS (gray dashed), and SSBS_NOCONV (gray solid). The ISS is computed for events exceeding a threshold of $+1.5\sigma$, where σ denotes the climatological standard deviation.

full dissipation rate. In the tropics, we find the same: the ignorance skill score for u_{850} is higher for SSBS-NOCONV and SSBS-CONSTDISS than for the operational ensemble but is not as high as for SSBS-FULLDISS (Fig. 14).

Of the two simplified schemes, SSBS-NOCONV tends to perform better than SSBS-CONSTDISS in the extratropics but not as well in the tropics. This might be an indication that in the extratropics a flow-dependent stochastic parameterizations is desirable, whereas in the tropics the ensemble is so underdispersive that the scheme with most spread will perform best. As seen in Fig. 5b, SSBS-CONSTDISS is more dispersive in the tropical band than SSBS-NOCONV because latter has no contribution from deep convection.

As discussed for SSBS-FULLDISS in section 4e, the improvement in skill is even more pronounced if the simplified schemes are compared against an ensemble with reduced initial perturbations. No general conclusion can be drawn as to whether SSBS-NOCONV or SSBS-CONSTDISS lead to a bigger improvement in skill because this depends on the details of the verification (e.g., geographic location). Although the simplified backscatter schemes are an improvement over the operational model, neither of them has as much skill as the scheme with a full dissipation rate.

6. Discussion

The notion of a kinetic energy backscatter scheme features two novel ideas. First, it goes beyond the idea of merely sampling subgrid-scale variability by picking realizations from a distribution centered on the value of the deterministic bulk parameterization (Buizza et al. 1999; Lin and Neelin 2002) by adding perturbations that

mimic the influence of altogether unrepresented subgrid-scale processes. Secondly, the perturbations are introduced as streamfunction forcing and are not directly applied to the physical tendencies. This implies that the state of the model with all its parameterizations can adjust to a slightly perturbed dynamical state, whereas directly perturbing the physical tendencies does not allow for an adjustment between the dynamical and parameterized parts of the model.

The two approaches address different aspects of model error: the stochastic diabatic tendencies account for model error caused by neglecting subgrid-scale sampling uncertainty whereas the stochastic backscatter approach accounts for model error from neglecting net kinetic energy injection from subgrid-scale processes. In principle, both forms of model error (and others not discussed here) should be represented within the same model, although this might not be practical from the perspective of developing deterministic parameterizations.

An intriguing aspect of the backscatter scheme is the use of the total kinetic energy dissipation rate to link the random perturbations to the large-scale state. This choice is physically motivated by the fact that in regions of large dissipation, a fraction of the diagnosed energy “drain” is transferred upscale and available as kinetic energy forcing for the resolved flow. These tend to be the regions of large parameterization uncertainty and thus model error. The connection of the stochastic perturbations to the instantaneous dissipation rate makes the spectral backscatter scheme a physically motivated stochastic parameterization that takes flow dependence into account.

Although relating the stochastic streamfunction perturbations to the total dissipation rate is a strength of kinetic energy backscatter schemes, it is also their

biggest challenge. In reality, very little is known about the dissipation rate of various processes in the atmosphere and even less about how much of their energy is scattered upscale. Guided by the impact on the kinetic-energy spectrum, we suggest here a backscatter ratio of 2%, which is small compared to the values typically used in LES (usually close to 100%). However, in simulations of two-dimensional turbulence the near-grid-scale disturbances are balanced, whereas some of the processes included in the SSBS are mostly unbalanced (e.g., convective systems). If the backscatter ratio represents the fraction of the near-grid-scale kinetic energy that projects onto scales close to the Rossby radius of deformation L_R , then we anticipate a factor on the order of $(\Delta/L_R)^2$ in its definition (where Δ is the grid length). For $\Delta \sim 100$ km and $L_R \sim 1000$ km this implies $r \sim 1\%$. Indeed, this is consistent with the analysis of Schubert et al. (1980), who analyzed the geostrophic adjustment of disturbances with circular symmetry and found that just a few percent of the convective energy was captured in balanced flow. Lilly (1983) also suggested that only a few percent of the energy released in deep convection was likely to escape to large scales and that this might be enough to account for the $-5/3$ mesoscale energy spectrum. These considerations are therefore consistent with the choice of $r \sim 2\%$ discussed in section 4a. In contrast, if the near-grid-scale disturbance is balanced, then one anticipates backscatter ratios similar to those used in LES and it is therefore possible that the small value of r used here is inappropriate for the numerical dissipation component of the backscatter.

In addition to the choice of the backscatter ratio, it needs to be decided which contributions to the total dissipation rate should be included. Here, we included processes that are associated with upscale error propagation, like deep convection and gravity/mountain wave drag, but there are many more dissipative processes that we have assumed are not associated with upscale error growth, such as boundary layer processes. Although in LES there might be sound theoretical assumptions for the rate of dissipated energy, the estimation of dissipation in NWP models will in the foreseeable future depend on the use of model variables from the conventional parameterization routines. A shortcoming is that errors in the conventional parameterizations can potentially be amplified by the backscatter scheme. New means of illuminating these issues might come from the analysis of coarse-grained cloud-resolving models and observational studies.

Other parameter choices in the SSBS are the forcing wavenumber band and the smoothness of the dissipation rate. Here, they have been picked so that the rep-

resentation of wavenumber-dependent error growth is improved (section 4c). We found that when forcing only wavenumbers close to the truncation scale—as in LES—the errors did not cascade upscale fast enough. This could be the result of not forcing the right spatial multivariate structures or it could reflect the fact that certain aspects of model error are associated with larger scales (e.g., the organized convection in the tropics and the convective outflow from mesoscale systems in the extratropics). This might justify perturbations across all wavenumbers (Fig. 2). Our findings are consistent with the work of Tribbia and Baumhefner (2004), who find that for exponential error growth with a realistic time scale, disturbances with a spectral peak in the synoptic scales have to be forced directly rather than seeded from smaller scales.

7. Conclusions

In this study we introduced a new stochastic spectral kinetic energy backscatter scheme and evaluated its impact on flow-dependent predictability in the ECMWF ensemble system. Extensive analysis of a large number of skill scores in a 50-member ensemble at resolution T_L255 showed that the scheme produced more skillful probabilistic forecasts and allowed us to reduce the somewhat too large initial perturbations by 15%. It was also shown that the improvement was not due to the reduction of the initial perturbations alone but was a direct consequence of the scheme itself. The positive impact is evident not only in the probabilistic skill scores but also in a better match of the kinetic-energy spectra with the T_L511 analysis and a better representation of flow-dependent predictability. The spread–error relationship as function of forecast lead time is improved and the model is better at capturing error growth as function of wavenumber. The scheme also positively impacted precipitation forecasts.

The positive effects were especially pronounced in the tropics, where the operational ensemble is traditionally underdispersive because it has no initial perturbations based on singular vectors except those around tropical cyclones. In this region, the stochastic backscatter perturbations not only greatly increased the spread but also reduced the RMS error of the ensemble mean, which lead to a marked improvement in probabilistic forecast skill. Although it is plausible that a stochastic scheme will introduce additional spread, this will not necessarily lead to an improvement in forecast skill. For instance, introducing humidity perturbations in the same model increases not only the spread but also the RMS error, so that the skill of the ensemble system is reduced (Tompkins and Berner 2008).

To understand which processes are responsible for the improvement, we performed experiments with simplified dissipation rates. It was shown that the backscatter scheme also has a positive impact if no contributions from deep convection were included, although the improvement is not as large as with the full dissipation rate. Because there is very little forcing in the tropics in this experiment (not shown), we can conclude that the positive impact of the backscatter scheme in the extratropics is not solely due to an improvement in tropical activity. Of interest was the result that a scheme with constant dissipation performed also well, especially in the tropics. Because such a scheme is much simpler and foregoes some of the difficulties of computing the full dissipation, it might be a good alternative for studying the potential of kinetic-energy backscatter schemes in models of intermediate complexity. However, neither of the simplified schemes performed as well as the scheme with the full dissipation rate, suggesting that the use of a flow-dependent stochastic backscatter scheme is best suited for improving flow-dependent predictability.

Acknowledgments. We thank Peter Bechthold, Roberto Buizza, and Martin Miller for many inspiring discussions and helpful suggestions throughout the duration of this work. Mark Rodwell is acknowledged for his rainfall verification software, which was used for the results in section 4g. Rich Rotunno and Andy Brown are thanked for their insightful comments on earlier versions of this manuscript. We are indebted to Ceçile Penland and an anonymous reviewer for their thoughtful reviews.

APPENDIX

Derivation of Backscattered Energy

This appendix derives the global kinetic energy that streamfunction perturbations of the form (2) inject into the resolved flow. The injected energy per unit mass $\Delta E'$ is the difference between the total kinetic energy per unit mass at time $t + \Delta t$ and t :

$$\Delta E' = \frac{1}{4\pi} \sum_{n=0}^N \sum_{m=-n}^n \frac{n(n+1)}{a^2} \left\langle |\psi_n^m(t + \Delta t)|^2 - |\psi_n^m(t)|^2 \right\rangle. \quad (\text{A1})$$

The change in total kinetic energy is expressed in terms of spherical harmonics and takes the spherical symmetry into account (e.g., Koshyk and Hamilton 2001). The streamfunction at time $t + \Delta t$ is given by

$$\psi_n^m(t + \Delta t) = \psi_n^m(t) + S_n^m(t)\Delta t + \psi_n'^m(t), \quad (\text{A2})$$

where S_n^m is the spectral source term due to advection, diffusion, and physical parameterizations. Using (2), we arrive at

$$\Delta E' = \frac{1}{4\pi} \sum_{n=0}^N \sum_{m=-n}^n \frac{n(n+1)}{a^2} \left\langle |\psi_n^m(t) + S_n^m(t)\Delta t + \psi_n'^m(t)|^2 - |\psi_n^m(t)|^2 \right\rangle \quad (\text{A3})$$

$$= \frac{1}{4\pi} \sum_{n=0}^N \sum_{m=-n}^n \frac{n(n+1)}{a^2} \left\langle 2|\psi_n^m(t)\psi_n'^m(t)| + |\psi_n'^m(t)|^2 \right\rangle \quad (\text{A4})$$

$$+ \frac{1}{4\pi} \sum_{n=0}^N \sum_{m=-n}^n \frac{n(n+1)}{a^2} \left\langle 2|\psi_n^m(t)S_n^m(t)\Delta t| + |S_n^m(t)\Delta t|^2 \right\rangle. \quad (\text{A5})$$

Whereas the first term describes the energy injection due to interactions between the resolved flow and the forcing, the second is directly associated with the energy injected by the streamfunction perturbations. The third and fourth terms contain the spectral source term S_n^m and its interaction with the resolved flow. Because the source term is not correlated with the streamfunction perturbations, we can set $\langle |\psi_n^m(t)S_n^m(t)\Delta t|^2 \rangle = 0$.

Let us have a closer look at the interaction term $\langle 2|\psi_n^m(t)\psi_n'^m(t)| \rangle$. To calculate this term, we start from the propagation Eq. (2), multiply it by $\psi_n'^m(t + \Delta t)$, and take the ensemble average:

$$\begin{aligned} \langle \psi_n^m(t + \Delta t)\psi_n'^m(t + \Delta t) \rangle &= \langle \psi_n^m(t)\psi_n'^m(t + \Delta t) \rangle \\ &+ \langle S_n^m(t)\psi_n'^m(t + \Delta t) \rangle \Delta t \\ &+ \langle \psi_n'^m(t)\psi_n'^m(t + \Delta t) \rangle. \end{aligned} \quad (\text{A6})$$

Using (2) to substitute $\psi_n^m(t + \Delta t)$, and noting that $\langle \psi_n^m(t)\epsilon(t) \rangle = 0$, $\langle \psi_n'^m(t)\epsilon(t) \rangle = 0$, and $\langle |\psi_n^m(t)S_n^m(t)\Delta t|^2 \rangle = 0$, we arrive at

$$\begin{aligned} \langle \psi_n^m(t + \Delta t)\psi_n'^m(t + \Delta t) \rangle &= (1 - \alpha) \\ &\times [\langle \psi_n^m(t)\psi_n'^m(t) \rangle + \langle \psi_n'^m(t)\psi_n'^m(t) \rangle]. \end{aligned} \quad (\text{A7})$$

Using the stationarity condition, $\langle \psi_n^m(t + \Delta t)\psi_n'^m(t + \Delta t) \rangle = \langle \psi_n^m(t)\psi_n'^m(t) \rangle$ yields

$$\langle \psi_n^m(t)\psi_n'^m(t) \rangle = \left(\frac{1}{\alpha} - 1 \right) \langle \psi_n'^m(t)\psi_n'^m(t) \rangle. \quad (\text{A8})$$

Note that for $\alpha = 0$, $\langle \psi_n^m(t)\psi_n'^m(t) \rangle = \text{const}$ and we cannot derive (8) from (7), which is why we exclude this case. Inserting (8) into (5), the total injected kinetic energy is derived to

$$\Delta E' = \frac{1}{4\pi} \sum_{n=0}^N \sum_{m=-n}^n \left(\frac{2}{\alpha} - 1 \right) \frac{n(n+1)}{a^2} \langle |\psi_n^m(t)|^2 \rangle \quad (\text{A9})$$

$$+ \frac{1}{4\pi} \sum_{n=0}^N \sum_{m=-n}^n \frac{n(n+1)}{a^2} \langle 2|\psi_n^m(t)S_n^m(t)\Delta t|^2 + |S_n^m\Delta t|^2 \rangle. \quad (\text{A10})$$

To make the problem tractable, we neglect all terms with the source and sink term $S_n^m(t)$. There is no strict justification for making this assumption, but because the stochastic kinetic backscatter algorithm makes many assumptions and simplifications, we are not overly concerned about neglecting these terms. With this assumptions, the injected kinetic energy is given by

$$\Delta E' = \frac{1}{4\pi} \sum_{n=0}^N \sum_{m=-n}^n \left(\frac{2}{\alpha} - 1 \right) \frac{n(n+1)}{a^2} \langle |\psi_n^m(t)|^2 \rangle. \quad (\text{A11})$$

If the noise process had no temporal memory, $\alpha = 1$, and forcing increments were injected instantaneously at each time step, the injected energy would equal the kinetic energy of the streamfunction perturbations. However, because the noise process has temporal correlations, the streamfunction and streamfunction perturbations are correlated and these correlations lead to an increase in the injected energy (5).

With (3) and noting that $\sum_{n=1}^N \sum_{m=-n}^n n(n+1) = \sum_{n=1}^N n(n+1)(2n+1)$, we arrive at the following analytic expression for the perturbation energy:

$$\Delta E' = \frac{\sigma_z}{4\pi a^2 \alpha} \sum_{n=1}^N n(n+1)(2n+1)g_n^2. \quad (\text{A12})$$

Assuming furthermore that the noise amplitudes g_n follow the power law $g_n = bn^p$ and that the globally averaged kinetic energy increment $\Delta E'$ is given and fixed, we are solving (12) for the amplitude function b :

$$b = \left(\frac{4\pi a^2}{\sigma_z \Gamma} \alpha \Delta E' \right)^{\frac{1}{2p}}, \quad (\text{A13})$$

where $\Gamma = \sum_{n=1}^N n(n+1)(2n+1)n^{2p}$.

REFERENCES

- Barkmeijer, J., R. Buizza, T. N. Palmer, K. Puri, and J.-F. Mahfouf, 2001: Tropical singular vectors computed with linearized diabatic physics. *Quart. J. Roy. Meteor. Soc.*, **127**, 685–708.
- Berner, J., 2005: Linking nonlinearity and non-Gaussianity of planetary wave behavior by the Fokker–Planck equation. *J. Atmos. Sci.*, **62**, 2098–2117.
- , F. J. Doblas-Reyes, T. N. Palmer, G. Shutts, and A. Weisheimer, 2008: Impact of a quasi-stochastic cellular automaton backscatter scheme on the systematic error and seasonal prediction skill of a global climate model. *Philos. Trans. Roy. Soc.*, **366A**, 2561–2579.
- Buizza, R., and T. Palmer, 1995: The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, **52**, 1434–1456.
- , M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908.
- , P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, and Y. Zhu, 2005: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Wea. Rev.*, **133**, 1076–1097.
- Frederiksen, J. S., and A. G. Davies, 1997: Eddy viscosity and stochastic backscatter parameterizations on the sphere for atmospheric circulation models. *J. Atmos. Sci.*, **54**, 2475–2492.
- , and —, 2004: The regularized DIA closure for two-dimensional turbulence. *Geophys. Astrophys. Fluid Dyn.*, **98**, 203–223.
- , and S. M. Kepert, 2006: Dynamical subgrid-scale parameterizations from direct numerical simulations. *J. Atmos. Sci.*, **63**, 3006–3019.
- Hamilton, K., R. J. Wilson, J. Mahlman, and L. Umscheid, 1995: Climatology of the SKYHI troposphere–stratosphere–mesosphere general circulation model. *J. Atmos. Sci.*, **52**, 5–43.
- Jolliffe, I., and D. Stephenson, 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Wiley, 240 pp.
- Katz, R. W., and A. H. Murphy, Eds., 1997: *Economic Value of Weather and Climate Forecasts*. Cambridge University Press, 222 pp.
- Koshyk, J. N., and K. Hamilton, 2001: The horizontal kinetic energy spectrum and spectral budget simulated by a high-resolution troposphere stratosphere mesosphere GCM. *J. Atmos. Sci.*, **58**, 329–348.
- Leith, C. E., 1978: Objective methods for weather prediction. *Annu. Rev. Fluid Mech.*, **10**, 107–128.
- Leutbecher, M., and T. N. Palmer, 2008: Ensemble forecasting. *J. Comput. Phys.*, **227**, 3515–3539.
- Lilly, D. K., 1983: Stratified turbulence and the mesoscale variability of the atmosphere. *J. Atmos. Sci.*, **40**, 749–761.
- Lin, J. W.-B., and J. D. Neelin, 2002: Considerations for stochastic convective parameterization. *J. Atmos. Sci.*, **59**, 959–975.
- Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.
- , 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307.
- Mason, P., and D. Thomson, 1992: Stochastic backscatter in large-eddy simulations of boundary layers. *J. Fluid Mech.*, **242**, 51–78.
- Molteni, F., and T. N. Palmer, 1993: Predictability and finite-time instability of the northern winter circulation. *Quart. J. Roy. Meteor. Soc.*, **119**, 269–298.
- Nastrom, G., and K. Gage, 1985: A climatology of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft. *J. Atmos. Sci.*, **42**, 950–960.
- Palmer, T. N., 2001: A nonlinear dynamical perspective on model error: A proposal for nonlocal stochastic-dynamic parameterization in weather and climate prediction models. *Quart. J. Roy. Meteor. Soc.*, **127**, 279–304.
- , R. Buizza, R. Hagedorn, A. Lawrence, M. Leutbecher, and L. Smith, 2006: Ensemble prediction: A pedagogical perspective. *ECMWF Newsletter*, No. 106, ECMWF, Reading, United Kingdom, 10–17.

- Penland, C., 2003a: Noise out of chaos and why it won't go away. *Bull. Amer. Meteor. Soc.*, **84**, 921–925.
- , 2003b: A stochastic approach to nonlinear dynamics: A review. *Bull. Amer. Meteor. Soc.*, **84**, ES43–ES51.
- Puri, K., J. Barkmeijer, and T. N. Palmer, 2001: Ensemble prediction of tropical cyclones using targeted diabatic singular vectors. *Quart. J. Roy. Meteor. Soc.*, **127**, 709–731.
- Rodwell, M., 2006: Comparing and combining deterministic and ensemble forecasts: How to predict rainfall occurrence better. *ECMWF Newsletter*, No. 106, ECMWF, Reading, United Kingdom, 17–23.
- Rotunno, R., and C. Snyder, 2008: A generalization of Lorenz's model for the predictability of flows with many scales of motion. *J. Atmos. Sci.*, **65**, 1063–1076.
- Roulston, M. S., and L. Smith, 2002: Evaluating probabilistic forecasts using information theory. *Mon. Wea. Rev.*, **130**, 1653–1660.
- Saetra, O., H. Hersbach, J.-R. Bidlot, and D. Richardson, 2004: Effects of observation errors on the statistics for ensemble spread and reliability. *Mon. Wea. Rev.*, **132**, 1487–1501.
- Sardeshmukh, P., C. Penland, and M. Newman, 2001: Rossby waves in a fluctuating medium. *Stochastic Climate Models*, P. Imkeller and J. S. von Storch, Eds., Vol. 49, *Progress in Probability*, Birkhäuser Verlag, 369–384.
- Schubert, W., J. Hack, P. L. Silva Dias, and S. Fulton, 1980: Geostrophic adjustment in an axisymmetric vortex. *J. Atmos. Sci.*, **37**, 1464–1484.
- Shutts, G. J., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **612**, 3079–3102.
- , 2008: The forcing of large-scale waves in an explicit simulation of deep tropical convection. *Dyn. Atmos. Oceans*, **45**, 1–25.
- , and M. E. B. Gray, 1994: A numerical modelling study of the geostrophic adjustment following deep convection. *Quart. J. Roy. Meteor. Soc.*, **120**, 1145–1178.
- , and T. N. Palmer, 2007: Convective forcing fluctuations in a cloud-resolving model: Relevance to the stochastic parameterization problem. *J. Climate*, **20**, 187–202.
- Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.*, **132**, 3019–3032.
- Steinheimer, M., M. Hantel, and P. Bechtold, 2007: Convection in Lorenz's global energy cycle with the ECMWF model. ECMWF Tech. Memo 545, 39 pp. [Available online at <http://www.ecmwf.int/publications/>.]
- Tompkins, A. M., and J. Berner, 2008: A stochastic convective approach to account for model uncertainty due to unresolved humidity variability. *J. Geophys. Res.*, **113**, D18101, doi:10.1029/2007JD009284.
- Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.
- Tribbia, J. J., and D. P. Baumhefner, 2004: Scale interactions and atmospheric predictability: An updated perspective. *Mon. Wea. Rev.*, **132**, 703–713.
- von Storch, H., and F. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 494 pp.
- Wilks, D. S., 1997: Resampling hypothesis tests for autocorrelated fields. *J. Climate*, **10**, 65–82.