Systematic Model Error: The Impact of Increased Horizontal Resolution versus Improved Stochastic and Deterministic Parameterizations

J. BERNER

National Center for Atmospheric Research, Boulder, Colorado*

T. JUNG

Alfred-Wegener-Institut für Polar- und Meeresforschung, Bremerhaven, Germany

T. N. PALMER

ECMWF, Shinfield Park, Reading, and Atmospheric, Oceanic and Planetary Physics, University of Oxford, Oxford, United Kingdom

(Manuscript received 27 May 2011, in final form 9 February 2012)

ABSTRACT

Long-standing systematic model errors in both tropics and extratropics of the ECMWF model run at a horizontal resolution typical for climate models are investigated. Based on the hypothesis that the misrepresentation of unresolved scales contributes to the systematic model error, three model refinements aimed at their representation—fluctuating or deterministically—are investigated.

Increasing horizontal resolution to explicitly simulate smaller-scale features, representing subgrid-scale fluctuations by a stochastic parameterization, and improving the deterministic physics parameterizations all lead to a decrease in the systematic bias of the Northern Hemispheric circulation. These refinements reduce the overly zonal flow and improve the model's ability to capture the frequency of blocking. However, the model refinements differ greatly in their impact in the tropics. While improving the deterministic and introducing stochastic parameterizations reduces the systematic precipitation bias and improves the characteristics of convectively coupled waves and tropical variability in general, increasing horizontal resolution has little impact.

The fact that different model refinements can lead to reductions in systematic model error is consistent with the hypothesis that unresolved scales play an important role. At the same time, this degeneracy of the response to different forcings can lead to compensating model errors. Hence, if one takes the view that stochastic parameterization should be an important element of next-generation climate models, if only to provide reliable estimates of model uncertainty, then a fundamental conclusion of this study is that stochasticity should be incorporated within the design of physical process parameterizations and improvements of the dynamical core and not added a posteriori.

1. Introduction

The construction of a comprehensive climate model, one that simulates realistically Earth's climate, is one of the greatest challenges in computational science. Such

models can be characterized in terms of both their dynamical cores and their physical parameterizations. The dynamical core describes the truncation of the underlying—essentially infinite-dimensional—differential equations onto a finite grid (or in the case of spectral dynamical cores the smallest resolved wavelength) and the corresponding algorithms that evolve the resolved scales of motion. Complementary to this, the physical parameterizations are traditionally deterministic formulas that provide estimates of the grid-scale effect of processes (e.g., deep convection) that cannot be resolved by the dynamical core.

These formulas necessarily involve a number of free parameters that in principle should be determined by

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: Dr. Judith Berner, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000.
E-mail: berner@ucar.edu

observations, but in practice may not be. This may be partially because of insufficient observations, but more typically it is because the formulas themselves are not accurate bulk representations of the subgrid processes. This necessitates a certain amount of climate-model “tuning”: fixing parameters by the fit of the climate model as a whole, to some large-scale climatic metric, such as globally averaged net outgoing longwave radiation.

In recent years, the development of methods to estimate flow-dependent uncertainty in climate model predictions has become an important addendum to the development of climate models. Different techniques have been developed, the most commonplace being multimodel ensembles, for example as described in Assessment Report 4 of the Intergovernmental Panel on Climate Change (AR4 IPCC; Solomon et al. 2007), and perturbed parameter ensembles (e.g., Stainforth et al. 2005). However, a third technique has emerged whose origin lies in numerical weather prediction. Specifically, the basis of stochastic-dynamic parameterization lies in the treatment of tendencies due to unresolved processes by stochastic rather than bulk-formula deterministic representations (Epstein and Pitcher 1972; Lorenz 1975; Pitcher 1977; Palmer 2001).

In this approach, the effect of uncertainties due to the finite truncation of the model can be estimated by running ensembles of integrations, whose members are based on independent realizations of a stochastic processes that describe these truncation uncertainties. One such stochastic parameterization, the focus of attention in this paper, is the so-called stochastic kinetic energy backscatter scheme whose origin lies in large-eddy simulation modeling (Mason and Thomson 1992) and has recently been extended to weather and climate scales (Shutts and Palmer 2003; Shutts 2005). The key idea is that energy associated with subgrid processes, instead of being dissipated within the grid box, is injected back onto the grid using a stochastic pattern generator. This method has been successfully used for operational and research forecasts across a range of scales from short-range to seasonal and annual predictions (Berner et al. 2008, 2009, 2011; Charron et al. 2010; Doblas-Reyes et al. 2009; Tennant et al. 2011; Weisheimer et al. 2011).

It is well known that certain aspects of systematic model error seem to be persistent not only in different model versions of the same model, but also across models developed at different climate centers. For example, many climate models underpredict the observed frequency of blocking (e.g., D’Andrea et al. 1998) and the Northern Hemispheric circulation is overly zonal (e.g., Branković et al. 2002). Similarly, most climate models are still not able to adequately capture the temporal-spatial variability of convectively coupled waves, and in particular the Madden-Julian oscillation (e.g., Lin et al. 2006). The

persistence of these shortcomings over decades (e.g., Jung et al. 2010) brings up the question of whether this could be the consequence of a common source of model error, namely the absence of a representation of the subgrid-scale variability (Palmer 2001).

This hypothesis is further underpinned by emerging studies demonstrating that these long-standing problems can indeed be partly remedied by model refinements to include unresolved scales. A partial improvement in the occurrence of Northern Hemispheric blocking has been reported in simulations with stochastic parameterizations (Jung et al. 2005; Berner et al. 2008) or greatly increased horizontal resolution (Jung et al. 2012). Tropical variability, and in particular the Madden-Julian oscillation, is much better captured in the multiscale modeling framework (Randall et al. 2003), where the traditional convective parameterizations are replaced by 2D cloud-permitting models in each grid box.

Unlike multimodel, multiphysics, or perturbed-parameter schemes, using a stochastic parameterization has the potential to alter the climate of the core model. In the present work, we demonstrate that this potential is realized both in the tropics and the extratropics. Although the impact is not universally beneficial in terms of model bias reduction, it is generally beneficial across a range of variables and diagnostics. Indeed, here we show that its overall impact is comparable to that of significant improvements both to the model’s dynamical core (through a quintupling of horizontal resolution) or with major revisions to the traditional parameterization schemes. More than that, we show that for certain climatological aspects, the impact of adding a stochastic parameterization is actually very similar not only in magnitude, but also in pattern, with the impact of increased resolution or a revised physics package. That is to say, there is a certain degeneracy in the response of the model to three completely different model refinements, in particular in the extratropics. However, for other physical phenomena, such as tropical variability, this degeneracy is not observed and the signature of the different model refinements is more unique. We develop these ideas, and discuss reasons for the degeneracy of the response in the conclusions.

The present work asks how different representations of unresolved and underresolved scales will influence the systematic model error of the atmospheric component of the climate system. To this end we conduct a set of 5-month integrations started from 1 November of each year between 1962 and 2005 but discard the first 30 days of each integration to avoid possible spinup effects. Jung (2005) performed a set of experiments very similar to those reported here and found that in boreal winter this is long enough for systematic atmospheric circulation errors to have reached their asymptotic

TABLE 1. Description of experiments.

Expt	Model version	Model error scheme	Horizontal resolution	Vertical resolution	Integration period
LOWRES	CY32R1	—	T95	L91	1962–2005
STOCH	CY32R1	Stochastic kinetic energy backscatter	T95	L91	1962–2005
HIGHRES	CY32R1	—	T511	L91	1990–2005
PHYS	CY36R1	—	T95	L91	1962–2005

seasonal mean values. Additionally, multiyear simulations following the Atmospheric Model Intercomparison Project (AMIP) protocol (e.g., Gates et al. 1999) were carried out with the core model of this study and it was confirmed that the systematic errors are indeed very similar for the 5-month and multiyear simulations (Jung et al. 2012). These findings, together with the growing body of literature that demonstrates that many systematic errors on climatic scales are already evident in the first few time steps of a model simulation (Klinker and Sardeshmukh 1992; Rodwell and Palmer 2007; Hannay et al. 2009), assure us that we are indeed analyzing asymptotic seasonal biases and not transitional behavior.

The paper is organized as follows. The experimental setup and data are described in section 2. Section 3 summarizes the improvements to the conventional physics parameterizations and the stochastic kinetic energy backscatter scheme. The impacts of increasing resolution, improving the deterministic parameterizations, and including a stochastic parameterization on systematic model error are presented in section 4 and discussed in section 5.

2. Experiment setup

To investigate the impact of model refinement on systematic model error, four experiments were performed: The control integration LOWRES uses model cycle CY32R1 at a horizontal resolution of T_L95 (about 210 km) and 91 vertical levels. This model version is currently used operationally at horizontal resolution T_L159 (about 125 km) in system 3 for seasonal predictions (Anderson et al. 2007). A second experiment, HIGHRES, uses the same model version and number of vertical levels, but a much higher horizontal resolution of T_L511 (about 40 km). While the horizontal resolution of spectral models is nominally the spatial scale Δx associated with twice the Nyquist frequency, it has been suggested that the effective resolution might be as much as 10 times Δx (Skamarock 2004).

All integrations use observed sea surface temperatures and sea ice fields as lower boundary conditions. To see the impact of including a stochastic subgrid model and improving the deterministic parameterizations, two

additional experiments were conducted: STOCH, based also on CY32R1, utilizes a stochastic kinetic energy backscatter stochastic backscatter scheme; and PHYS uses the—at the time of writing—most recent set of deterministic physical parameterizations as part of version CY36R1 (Bechtold et al. 2008; Jung et al. 2010). A summary of the model experiments is given in Table 1.

For experiments LOWRES, STOCH, and PHYS, the model was initialized on 1 November of each of the years 1962–2005 and integrated over 5 months. To remove the influence of the initial condition, the first month was discarded and the analysis done on the four subsequent months of extended boreal winter, from December through March. Because of the computational costs, the integrations of HIGHRES had to be restricted to 16 winters from 1990 to 2005.

The European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS) uses a semi-Lagrangian advection scheme which is a dissipative numerical scheme and does not conserve either mass or energy. For the seasonal runs presented here, a mass fixer was therefore applied every 24 h. Because of its primary use as an NWP model, the vertical momentum transport by deep convection is parameterized and the heating effects associated with the corresponding dissipation terms are not included in the model. To make sure that none of the above is the reason for the systematic errors reported here, AMIP-type studies with the IFS were carried out and it was confirmed that the systematic errors are indeed very similar for the 5-month and multiyear integrations (Jung et al. 2012) and that the IFS's circulation statistics are comparable to that of other climate models (Hazeleger et al. 2010).

Model biases are defined with regard to our best estimate of the true atmospheric state, which is given by the 40-yr ECMWF Re-Analysis (ERA-40; Uppala et al. 2005) and covers the period 1962–2001. After 2001 we use as reference the operational analysis truncated to T95. We will refer to this combined dataset as “analyses” throughout this study.

To facilitate a direct comparison between the different experiments, most figures and statistical significance

tests in this manuscript are based on the 16-yr period 1990–2005. Unless mentioned otherwise, the results for the 16- and 44-yr periods are very similar. One advantage of taking a more recent but shorter verification period is that because of the improved data coverage and quality (e.g., through satellites) in the recent decades, the analyses are more trustworthy, especially in relatively data-sparse regions such as the tropics.

3. Model refinements

a. Improvements to the physical parameterization package

There have been a number of profound improvements to the ECMWF model in the cycles between CY32R1 (operational use started in September 2006) and CY36R1 (operational use started in January 2010). Details on the improvements to the deterministic parameterizations are summarized in Jung et al. (2010) and references therein. The most important changes relevant for modeling the atmospheric circulation are a new formulation of convective entrainment and relaxation time in the convection scheme and a reduced vertical diffusion in the free atmosphere (Bechtold et al. 2008), which resulted in increased activity and better variability, especially in the tropical atmosphere. Other modifications include changes to the gravity wave drag parameterization, a new soil hydrology scheme, revisions to the cloud scheme, and a new radiation transfer package (Morcrette et al. 2008) consisting of the Monte Carlo Independent Column Approximation (McICA; Pincus et al. 2003) and the Rapid Radiative Transfer Model (RRTM; Iacono et al. 2008). In short, between 2006 and 2010, almost all physical parameterizations were modified in some manner, leading to an essentially different model. In general an improvement to the physical parameterizations will go into the next model release if it is physically justified and does not deteriorate the forecast skill significantly.

b. The stochastic kinetic energy backscatter scheme

The stochastic kinetic energy backscatter scheme aims at representing kinetic energy injections caused by interactions with unresolved scales and is based on the notion that the turbulent dissipation rate is the difference between upscale and downscale spectral transfer, with the parameterized upscale component being available to the resolved flow as a kinetic energy source (Mason and Thomson 1992; Shutts 2005). Since the scheme implemented here is exactly the scheme described in detail in Berner et al. (2009), we only give a short summary here.

To calculate a stochastic kinetic energy source, streamfunction perturbations of the form

$$\Psi'(\lambda, \phi, z, t) = rD(\lambda, \phi, z, t)\psi'(\lambda, \phi, t) \quad (1)$$

are introduced at each model time step t and λ, ϕ , and z denote longitude, latitude, and height in gridpoint space. Here $D(\lambda, \phi, z, t)$ is the local, instantaneous dissipation rate as a function of longitude, latitude, and height and $\psi'(\lambda, \phi, t)$ is an instantaneous 2D random streamfunction pattern with a barotropic vertical structure. The “backscatter ratio” r determines which fraction of the dissipation rate D will be backscattered onto the resolved scales.

Consistent with the idea of the production of balanced potential vorticity in organized convective systems (Shutts and Gray 1994), here only the rotational component of the spectrum was forced. An alternative approach is reported in Bowler et al. (2009), who also force divergent modes in the tropics.

The local total dissipation rate $D(\lambda, \phi, z, t)$ together with its contributions from deep convection, numerical dissipation, and gravity/mountain wave drag (Shutts 2005) are shown in Fig. 1 for the boreal winter. Detailed formulations of the dissipation rates are given in Berner et al. (2009) and not repeated here for brevity. For display purposes the rates were weighted by density and vertically averaged, so that the units in Fig. 1 are W m^{-2} . The boreal winter mean of the total dissipation rate is 1.65 W m^{-2} . Its dominating contributor is deep convection with a mean of 0.99 W m^{-2} . Its maxima are in the deep convective regions of the tropics. There is also a maximum downstream of the Andes (with maximal amplitude in the height of the jet stream level, not shown), which is likely the result of using a mass-flux formulation to estimate the dissipation from deep convection (Berner et al. 2009). The annual global mean value is slightly smaller than that 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.

With a value of 0.56 W m^{-2} the second largest contribution comes from numerical dissipation. It is largest over high orography such as Greenland, the Himalayas, the Andes, and the storm-track regions. The dissipation from gravity and mountain wave drag is with 0.09 W m^{-2} the smallest contributor to the annual mean but it can be locally very large, especially over orography. Furthermore, we note that the dissipation rates computed here at a resolution of T95 and averaged over 16 years agree within a factor of 2 with those reported in Berner et al. (2009), which were computed for shorter simulation periods and at a higher resolution of T255. This points to a weak resolution dependence of this calculation.

The 2D random streamfunction pattern $\psi'(\lambda, \phi, t)$ is temporally and spatially correlated and follows

FIG. 1. Vertically averaged (a) mean total dissipation rate per unit area (W m^{-2}) for extended boreal winter (December–March) with contributions from (b) deep convection, (c) numerical dissipation, and (d) gravity–mountain wave drag. Global averages for the boreal mean averaged over 40 years and its components are 1.65, 0.99, 0.56, and 0.09 W m^{-2} , respectively.

a prescribed kinetic energy spectrum. Each level uses the same streamfunction pattern, implying a barotropic vertical structure. The spatial and temporal characteristics of the random pattern are controlled by expanding it in spherical harmonics and evolving each wavenumber as a first-order autoregressive process. The amplitudes for each wavenumber are chosen in such a way that the injected kinetic energy and its spectrum follow prescribed formulations [see Eqs. (4) and (5) in Berner et al. 2009]. We confirmed that the injected kinetic energy is compensated by an increased numerical dissipation and that the total energy is not increasing unphysically over the simulation period.

We stress that the current scheme is identical to the one used in Berner et al. (2009) for medium-range weather forecasts, with the one exception of the backscatter ratio parameter, which was set here to a value of $r = 0.2$, which is 10 times the value used for the medium-range ensemble system. This choice was made to yield a reasonable spread in the seasonal ensemble prediction system described in Doblas-Reyes et al. (2009). However, even with this relatively high value of the backscatter ratio parameter, indicating that 20% of the dissipated energy is scattered upscale, the ensemble system continues to be underdispersive for seasonal forecasts.

4. Impact of model refinements on systematic model error

All model refinements discussed in the previous section have been shown to improve forecast skill not only over the short and medium range (Berner et al. 2009; Tennant et al. 2011; Berner et al. 2011) but also for seasonal to annual predictions (Doblas-Reyes et al. 2009; Weisheimer et al. 2011). Here, we focus on the question of whether a reduced systematic model error and better circulation statistics might have contributed to these improvements. To this end we study the impact of the different model refinements on a number of tropical and extratropical climate fields and statistics such as kinetic energy spectra, systematic biases in the geopotential height field and precipitation, the occurrence of blocking, convectively coupled tropical waves, and more generally tropical variability.

a. Energy spectra

Figure 2 shows the rotational (upper thick lines) and divergent part (lower thin lines) of the kinetic energy spectrum of wind at 200 hPa for analyses and all model experiments. There is a clear maximum in the spectra for synoptic wavenumbers with a maximum between wavenumber 8 and 10. LOWRES lacks kinetic energy

FIG. 2. Kinetic energy spectra of rotational (thick upper lines) and divergent (thin lower lines) wind anomalies at 200 hPa averaged over the boreal winters of the period 1990–2005. Shown are the reference spectra from the analyses (black solid) together with those from the model experiments: (a) LOWRES (gray solid) and STOCH (gray dotted) and (b) HIGHRES (black dashed) and PHYS (gray dashed). See Table 1 for a summary of the different model experiments.

in the rotational component of the spectrum for wave-numbers 20 and higher throughout the troposphere (Fig. 2a). The divergent part of the spectrum drops off right at the peak, suggesting that divergent modes are underrepresented even in the most powerful, synoptic part of the spectrum. The stochastic backscatter scheme injects kinetic energy and there is an increase in both the rotational and divergent part of the energy spectrum (Fig. 2a). However, STOCH has now slightly too much power in the kinetic energy spectrum, especially in the rotational modes. Since the backscatter scheme injects by design only kinetic energy in the rotational part of the spectrum, the increase in the divergent component of the spectrum is an indirect effect (e.g., by changing convection) and confirms that the scheme interacts with the model physics and dynamics. Increasing horizontal resolution also leads to an increase in kinetic energy (Fig. 2b). The match between the spectrum for analyses and HIGHRES is excellent for divergent modes at 200 hPa. But there is slightly too much power in the rotational component at 200 hPa. The changes in the deterministic parameterizations of PHYS are also characterized by overall increased levels of kinetic energy (Fig. 2b). This results in too much power at 200 hPa, especially for the divergent modes.

The analysis of the kinetic energy spectrum shows that all model refinements lead to an increase in rotational and divergent kinetic energy, which overall is positive. However, discrepancies with the analyses continue to exist, and in some cases the spectra are characterized by too much power.

The magnitude of the stochastic forcing used here was originally chosen to introduce sufficient spread in

seasonal ensemble forecasts presented in Doblas-Reyes et al. (2009). However, some of the diagnostics suggest that this backscatter ratio is slightly too large and that a smaller value might have been preferable, especially in the tropics, where the present scheme exhibits too much activity in the upper tropospheric wind divergence (see section 4g and Fig. 8). However, even with this choice, the scheme still results in an underdispersive ensemble system (Doblas-Reyes et al. 2009), suggesting that there are additional sources of model uncertainty that are not captured by the present scheme.

b. Systematic biases in the geopotential height field

The mean bias in 500-hPa geopotential height fields (Z500 hereafter) in the Northern Hemisphere extratropics during boreal winter is shown in Fig. 3. The geopotential heights are averaged over four winter months and all years, and the bias is calculated as the difference between the mean winter climate in the analysis and each experiment. The error in LOWRES reflects an overly zonal flow, especially in the storm-track regions. This systematic error has been persistent across a number of model versions for more than a decade (e.g., Branković et al. 2002; Jung et al. 2010). Including a stochastic model error representation reduces the systematic Z500 bias in both the North Pacific–North American and the North Atlantic regions substantially, but the circulation continues to be somewhat too zonal. Increasing horizontal resolution leads to a very similar improvement in systematic model bias in Z500 (Fig. 3c). Note that the increased horizontal resolution of HIGHRES not only affects the atmospheric dynamics, but also the representation of orographic details. Resolving orography in itself

FIG. 3. Mean systematic error of 500-hPa geopotential height fields (shading; m) for extended boreal winters of the period 1990–2005. Errors are defined with regard to the observed mean field (contour lines), consisting of a combination of ERA-40 (1990–2001) and operational ECMWF analyses (2002–05). Shown are the systematic error of experiments (a) LOWRES, (b) STOCH, (c) HIGHRES), and (d) PHYS. Significant differences at the 95% confidence level based on a Student's t test are hatched.

is expected to lead to a better representation of the Northern Hemispheric circulation (Jung et al. 2010). The biggest reduction in systematic bias is achieved by improving the physical parameterizations (Fig. 3d): the dipole pattern characterizing the error over the western Pacific has disappeared completely and over the European/North Atlantic region only a residual error remains. To determine the statistical significance of the bias, a Student's t test at each grid point was performed. Hatched areas denote the regions where the model climate differs from the observed climate at the 95% confidence level.

Our analysis shows that the mean bias in Z500 can be reduced in more than one way: increasing horizontal resolution, introducing stochastic perturbations, or changing the physics packages all lead to a significant improvement. While the improved physical parameterizations of the

experiment PHYS result overall in the closest match to the analysis, differences on a regional level exist. For example, the large bias of LOWRES over the western United States only completely disappears in the experiment STOCH (Fig. 3b). It is of interest to determine whether these improvements are the result of a better representation of local processes or if they are a teleconnection response to a better representation of tropical variability. This is discussed further in section 5.

The impact on the Southern Hemispheric circulation (not shown) differs quite substantially among the different experiments. HIGHRES and STOCH introduce a large positive bias over Antarctica, which probably has its origin in overactive synoptic variability in the Southern Hemispheric midlatitudes. PHYS reduces this bias over both the Antarctic continent and the Southern Ocean.

FIG. 4. Frequency of wintertime Northern Hemisphere blocking events for analyses (black solid) and model experiments LOWRES (gray solid), STOCH (gray dotted), HIGHRES (black dashed) and PHYS (gray dashed) for the periods (a) 1990–2005 and (b) 1962–2005. No curve is shown for HIGHRES since this experiment encompasses only 15 years. The gray shading indicates the 95% confidence intervals of the blocking index from analyses.

c. Northern Hemispheric blocking index

Another manifestation of the overly zonal flow in the Northern Hemisphere extratropics is the inability of LOWRES to represent the observed frequency of blocking events. Given the importance of blocking for local weather conditions, in particular over Europe, it is crucial that models capture the onset and frequency of blocking. It is well known that many climate models underpredict the observed frequency of blocking, particularly in the Euro-Atlantic region (e.g., D'Andrea et al. 1998). Linked to the Z500 error described above, this has been an outstanding issue for some time.

An index for the observed and simulated frequency of Northern Hemisphere blocking as function of longitude was computed using the algorithm proposed by Tibaldi and Molteni (1990) (Fig. 4). The observations show two maxima: one in the North Pacific and one in the Euro-Atlantic region. The gray shading indicates the 95% confidence intervals of the blocking index from analyses. The model captures the bimodal character of the blocking index, but the location of the North Atlantic maximum is shifted by 30° to the east. Since the occurrence of blocking is quantitatively somewhat different depending on the verification period, Fig. 4 shows the blocking index based on 16 and 44 years, separately. LOWRES substantially underestimates the occurrence of blocked states by about half in the North Atlantic and by one-third in the North Pacific region. All model refinements increase the frequency of blocking in both sectors. STOCH and HIGHRES increase the frequency of blocking at the maximum in the Pacific sector by more than 50% and in the Atlantic by 10%–20%. Similar improvements over the North Pacific due to a stochastic parameterization scheme were reported by Jung et al. (2005) and Berner et al. (2008).

That body of work used the cellular automaton scheme backscatter scheme (CASBS; Shutts 2005; Berner et al. 2008), the predecessor to the current stochastic backscatter scheme. The best results are obtained by PHYS. Over the North Pacific region the blocking index falls into the 95% confidence interval of the observed values, but the maximum is shifted slightly to the west. Over the North Atlantic the occurrence of blocking remains underestimated by 25%–30% and the eastward shift remains.

d. Systematic biases in precipitation

Figure 5 shows the systematic error in global precipitation as difference between the observed winter precipitation climatology estimated by the Global Precipitation Climatology Project (GPCP; Adler et al. 2003) and that of the model experiments. The pattern of systematic precipitation errors is very similar across the various model experiments: In the tropics, precipitation is generally too weak over the continents and too strong over oceans. Positive precipitation biases are particularly problematic in the tropical Atlantic, the southern parts of the Caribbean, the Indian Ocean, and in the tropical Pacific north of the equator. The model has a double intratropical convergence zone in the east Pacific, which in turn leads to two bands of precipitation error: one north and one south of the equator. In the North Atlantic–European region precipitation biases are indicative of an underestimation of Euro-Atlantic blocking events and a too weak storm track in higher latitudes. Again, these precipitation biases have been a robust feature of the ECMWF model for many years (e.g., Jung and Tompkins 2003; Jung et al. 2010).

In the experiment STOCH the general pattern of error remains, but there are substantial improvements in

FIG. 5. Mean systematic error in precipitation (mm day^{-1}) for extended boreal winters of the period 1990–2005. Errors are defined with regard to observations from the GPCP. Shown are systematic error of experiments (a) LOWRES, (b) STOCH, (c) HIGHRES, and (d) PHYS.

some regions (Fig. 5b). In particular, the wet bias north of the tropical Pacific, the Indian Ocean, and the Arabian Sea is greatly reduced and the dry bias over northern Australia improved. However, some of the excess precipitation north of the equator is shifted toward the south (e.g., in the central and eastern Pacific). A decomposition into convective and stratiform precipitation (not shown) indicates that the change in precipitation is associated mostly with a change in parameterized convective rainfall. In the Northern Hemispheric extratropics, the biggest improvement is seen over the Mediterranean region, where the dry bias is reduced.

Increasing horizontal resolution generally has a relatively small impact on the systematic precipitation errors in the tropics (Fig. 5c). The circulation over the Mediterranean region is improved to a similar degree as in STOCH, and probably linked to the improved representation of Atlantic blocking. In certain regions such as the Atlantic or North Pacific the precipitation bias even increases slightly.

The most substantial reduction in systematic error is obtained by improving the physical parameterizations (Fig. 5d). The dry bias over the continents in the tropics is reduced and the wet biases over the eastern tropical Pacific and the tropical Atlantic/Caribbean are both markedly reduced. However, large biases over the Indian Ocean and western Pacific remain.

e. Convectively coupled tropical waves

Previous studies show that current state-of-the-art general circulation models display a wide range of skill

in simulating convectively coupled tropical waves (e.g., Lin et al. 2006). Following Wheeler and Kiladis (1999), wavenumber–frequency spectra for observed and simulated anomalies of outgoing longwave radiation (OLR) in the latitude band 5°S – 5°N have been computed. Here, we focus on the component that is symmetric about the equator.

The total symmetric spectrum in the observations is characterized by strong eastward propagating and weaker westward propagating Rossby waves (Fig. 6a). Generally the opposite is true for most model experiments, where there is more power in the westward propagating than in the eastward propagating modes (Figs. 6c,e,g). The exception is the experiment PHYS, which exhibits strongly eastward propagating waves. However, even in this experiment the spectra remain too steep, indicating too little power in the higher frequencies.

In the observations, tropical variability associated with the Madden–Julian oscillation (MJO; Madden and Julian 1971) leads to a distinct peak in the spectrum of eastward propagating modes of wavenumber 1–3, which has markedly more power than its counterpart in the westward modes (Fig. 6a). None of the model experiments captures the MJO peak correctly, which will be analyzed in more detail in the next section.

Following Wheeler and Kiladis (1999), we define a signal-to-noise spectrum by removing the background spectrum for each experiment separately. The erroneous dominance of westward propagating modes and missing MJO variability is even more evident in these signal-to-noise spectra (Fig. 6, right column). The signal-to-noise

FIG. 6. Mean wavenumber–frequency spectra for the symmetric component of daily tropical outgoing longwave radiation anomalies (K^2) during December–March for the period (1990–2005) (left) before and (right) after a background spectrum has been removed (see text for details). The mean annual cycle has been removed prior to the computation of the wavenumber–frequency spectra. The wavenumber–frequency spectra are depicted for (a),(b) satellite data from NOAA and for experiments (c),(d) LOWRES, (e),(f) STOCH, (g),(h) HIGHRES, and (i),(j) PHYS. Contour intervals increase quadratically from 0.4 to 400.

spectra of experiments STOCH and PHYS resemble the observed signal closest, although they both misrepresent the MJO peaks. PHYS is slightly better at capturing the amplitude of eastward propagating modes, whereas STOCH is more successful at damping westward propagating Kelvin waves. HIGHRES shows no improvement over the wavenumber–frequency spectrum of LOWRES (Figs. 6g,h).

While we have not conducted a formal significance test, the wavenumber–frequency spectra from the 44-yr integrations are very similar to those from 16 years. Hence we believe that the described features are robust.

In summary, increasing horizontal resolution did not improve the wavenumber–frequency spectra of OLR, but introducing a stochastic kinetic energy backscatter scheme or improving the deterministic parameterizations resulted in less westward and more eastward propagating modes and an overall better match to the observed spectra. The closest match was obtained by refining the physical parameterization, especially the convective parameterization scheme (Jung et al. 2010).

f. The Madden–Julian oscillation

The Madden–Julian oscillation is an important contributor to tropical variability and capturing its characteristics is important for a number of reasons: First, the MJO may trigger El Niño–Southern Oscillation (ENSO) events (e.g., Vitart et al. 2003; Vitart and Molteni 2010); second, there is evidence that it can influence Northern Hemispheric extratropical medium-range predictability (Ferranti et al. 1990). Third, since the MJO has a characteristic time scale of 30–60 days, capturing MJO events has the potential to increase predictive skill on monthly time scales (Vitart 2004; Vitart and Jung 2010). However, both climate and weather models have great difficulties capturing not only individual events but also the statistics of the MJO (e.g., Lin et al. 2006), and the ECMWF model is no exception to this (Bechtold et al. 2008; Vitart and Jung, 2010).

To determine the model's ability to capture MJO variability, the power spectra of tropical velocity potential anomalies at 200 hPa in model and observations are computed (Fig. 7). Prior to the computation of the spectra the mean annual cycle has been removed. The spectra are averaged in time and over the latitudes band 5°S–5°N, but the longitudinal dependence is kept. Unlike the results presented so far, the details of these power spectra depend on the length of the analyzed time period. Hence, Fig. 7 depicts the power spectra, separately for the periods 1990–2005 (Figs. 7a,c,e,g,h) and 1962–2005 (Figs. 7b,d,f,i).

For the period 1990–2005, the analyses show a clear spectral maximum at a period of 40 days in the Eastern

Hemisphere, particularly between 60°E and 180° (Fig. 7a). When taking all 44 years of data into account, this peak shifts to a period of 60 days (Fig. 7b). LOWRES, on the other hand, merely produces red power spectra with no distinct spectral peak (Figs. 7c,d). In the model, two longitude bands with enhanced power emerge, between which is a region of low variability, around 120°E. The two-banded structure remains prevalent in all model experiments but is not present in the observations. In the period 1990–2005, a distinct peak at wavenumber 60 emerges for STOCH at longitudes of 120°E–180° (Fig. 7e). However, if the spectra are averaged over more years, this peak gives way to a red spectrum, albeit shallower than that of LOWRES. This problem is not remedied by increasing horizontal resolution. The spectrum of HIGHRES is still too red and does not have enough power in the MJO band. However, for longitudinal wavenumbers between 120° and 160°E there is a weak local maximum at a period of 40 days. The revised physical parameterizations of PHYS led to a large increase in power, resulting in too much amplitude for wavenumbers ≥ 70 in the west Pacific. However, the spectrum continues to have too much power for long periods. For the shorter period, there is evidence of a slight peak around 60°E (Fig. 7h), but the maximum is shifted toward longer periods of 60–70 days. However, for the longer period of 44 years, this peak disappears.

g. Tropical divergent wind activity

The tropical variability described in the previous sections is closely linked to the divergence of upper-tropospheric wind activity. To focus on variability in the synoptic frequency band (“synoptic activity”), the wind was high-pass filtered, and the filter was selected to retain variability on time scales shorter than 10 days (for details, see Jung 2005). The difference in the divergence of high-pass filtered winds between observations and model experiments is depicted in Fig. 8. A Student's *t* test at each grid point was performed, and regions of statistical significance at the 95% confidence level are hatched. LOWRES is characterized by too weak synoptic divergent wind activity (or too much convergent activity) in the tropics, subtropics, and into the mid-latitudes (Fig. 8a). There is slightly too little divergent wind activity over the poles. However, the wind activity is too divergent poleward of the midlatitudes and in particular over the Southern Hemispheric midlatitudes.

Rather than reducing the divergence bias, increasing resolution in HIGHRES does not reduce the error but instead introduces additional overactivity in the storm-track regions (Fig. 8c). The experiment STOCH increases the divergence everywhere. This leads to a reduction in error over regions with too much convergent activity

FIG. 7. Power spectra ($\text{m}^4 \text{s}^{-2}$) of tropical velocity potential anomalies at 200 hPa for each longitude and averaged over the latitudes 5°S – 5°N . The mean annual cycle has been removed and the spectra were computed separately for each year and subsequently averaged over the periods (left) 1990–2005 and (right) 1962–2005. Depicted are power spectra for (a),(b) observations from NOAA and experiments (c),(d) LOWRES, (e),(f) STOCH, (g) HIGHRES, and (i),(j) PHYS. (h) Since the experiment HIGHRES encompasses only 15 years, this panel is left blank.

FIG. 8. Mean systematic error for divergence of bandpass-filtered winds in 200 hPa (shading; s^{-1}) for extended boreal winters of the period 1990–2005. Errors are defined with regard to the observed mean field (contour lines). Shown are systematic error of experiments (a) LOWRES, (b) STOCH, (c) HIGHRES, and (d) PHYS. Significant differences at the 95% confidence level are hatched.

(e.g., over the poles) but increases the error in the regions characterized by too much divergence such as the Southern Hemispheric midlatitudes. Upper-level divergence in the tropics is greatly increased, resulting in an improvement over the tropical Atlantic and the South American and African continents but leading to large errors over the rest of the tropical belt, especially the tropical Pacific. Compared to LOWRES, the experiment PHYS has overall more divergent wind activity, resulting in a smaller error over the tropics but an increased error over the midlatitudes. These results are consistent with the previous results that indicate an increase in tropical wave activity in the experiments STOCH and PHYS. In its current form the stochastic scheme produces too much divergent wind activity, especially in the tropical belt. This, together with the analysis of the kinetic energy spectrum (section 4a), leads us to believe that the amplitude of the stochastic forcing was chosen somewhat too large and a smaller amplitude would have yielded better results.

h. Tropical empirical orthogonal functions

The previous subsections analyzed different physical processes of tropical variability. Next, we attempt to show their combined dynamical impact by calculating empirical orthogonal functions (EOFs). Tropical EOFs of streamfunction anomalies in 200 hPa were computed by restricting the region for the EOF analysis to the tropical band between 20°S and 20°N . A global pattern was obtained by finding composites that highly project

onto the leading tropical EOFs (e.g., Kimoto and Ghil 1993). The first observed EOF is dominated by a dipole pattern over the central Pacific that extends into the northern and southern Pacific to form a north–southward wave train (Fig. 9a). The associated principal component (PC) explains 22% of the tropical variance (Table 2).

Generally, the first EOF in the model experiments is qualitatively similar to that in the observations. However when focusing on the details, we note that the contours over the central Pacific are not closed, indicating a less pronounced dipole structure (Fig. 9b). More importantly, the explained variance of the first PC is with 27% too large. For STOCH, the first tropical EOF resembles that of the observations more closely, recapturing the dipole over the Pacific. In addition, the explained variance of the first PC is now reduced to 23% and closer to that in the observations. The first EOF of HIGHRES has least resemblance to the observed tropical EOF. It lacks the central Pacific dipole pattern and its extratropical extensions and assigns with 27% too much variance to the first PC. In contrast, PHYS captures the pattern and explained variance of the first EOF very well.

The EOFs based on 44 years of winter are very similar to those for the shorter period, especially in the tropical band, and are hence not shown. The extratropical features of the leading EOF are slightly smoother for the longer period, but there are no major qualitative differences. However, the explained variances show a slight dependence on the number of years used and are hence given for the shorter and extended period separately

FIG. 9. Composites of streamfunction (at 200 hPa) with large projection onto the first tropical EOF for extended boreal winters of the period 1990–2005. Shown are the composites for streamfunction from (a) analyses and the experiments (b) LOWRES, (c) STOCH, (d) HIGHRES, and (e) PHYS.

TABLE 2. Explained variances (%) of the first and second PCs of tropical EOFs for the periods 1990–2005 (see Fig. 9) and 1962–2005.

Expt	PC1 (1990–2005)	PC1 (1962–2005)
Analyses	22	21
LOWRES	27	30
STOCH	23	22
HIGHRES	27	27
PHYS	20	21

(Table 2). The values show that LOWRES and HIGHRES consistently overestimate the variance of the first PC, and that both STOCH and PHYS reduce the explained variance to the observed value.

5. Discussion and conclusions

In the Northern Hemisphere a long-standing systematic error of many climate models consists of an overly zonal mean circulation and underestimation of the frequency of blocked events. Here, we demonstrate that there are several ways to improve, albeit not fully remedy, this shortcoming. Increasing horizontal resolution, representing subgrid-scale fluctuations by a stochastic kinetic energy backscatter scheme and improving the conventional deterministic representations all reduce the zonality of the circulation and increase the frequency of blocking. This leads to the conclusions that climate models seem to be able to capture blocking if certain aspects of the unresolved flow are represented in some manner or—more generally—that a better representation of subgrid-fluctuations is crucial to remedy certain aspects of systematic model error. At the same time the degeneracy of the response can lead to compensating model errors and warrants a cautionary note: if a reduction in systematic model bias can be achieved in more than one way, changes to the model formulation must be underpinned by additional physical reasoning.

Such a degeneracy in the model response to different forcings can occur in nonlinear models and another example has recently been reported by Palmer and Weisheimer (2011). They report how an inadequate representation of horizontal baroclinic fluxes resulted in a model error that was equal and opposite to the systematic error caused by insufficiently represented vertical orographic gravity wave fluxes. Improvements to wave drag parameterization without increasing resolution unbalanced the compensating model errors, leading to an increase in systematic model bias. This might be one reason why it is so difficult to isolate and analyze model bias in climate models.

This point is further substantiated by investigating the impact of the three model refinements on the tropical

circulation. While increasing horizontal resolution did not markedly improve the dominance of westward propagating Kelvin waves or the ability to realistically capture MJO variability, both improving the deterministic parameterization and using a stochastic subgrid-scale representation did improve tropical variability. However, this could be a model-specific result; indeed, more recent versions of the ECMWF model indicate some sensitivity in the tropics to horizontal resolution in the presence of a trade wind bias (Jung et al. 2012). Overall, best results were obtained by revising the physical parameterizations, and in particular the introduction of a more active convective parameterization scheme (Bechtold et al. 2008; Jung et al. 2010).

Without doubt, many more resources go into improving physical parameterizations than into developing physically based stochastic parameterizations. With this in mind, we think that the results of implementing a stochastic kinetic energy backscatter scheme are very encouraging, especially when considering how long-standing some of the reported systematic model errors are. In particular, this is the first demonstration of a stochastic parameterization shown to be beneficial across a wide range of scales, providing improved climate statistics (this study), as well as more skillful forecasts from the short and medium range (Berner et al. 2009, 2011; Tennant et al. 2011) to seasonal and annual predictions (Doblas-Reyes et al. 2009; Weisheimer et al. 2011).

A notion often associated with stochastic parameterizations is that they introduce variability at all spatial and temporal scales, which is the expected response of a linear system to white noise forcing. Interestingly, this work shows two examples where the stochastic forcing resulted in the opposite, namely a decrease in variability: 1) The stochastic scheme intermittently breaks up the zonality of the Northern Hemispheric circulation and leads to more blocking (Fig. 4); and 2) in the tropics the scheme reduces the unrealistically large activity of westward propagating convectively coupled waves and reduces their power in the wave-frequency spectrum (Fig. 6). This encouraging result points to the importance of a flow-dependent scheme and/or the nonlinear nature of the climate model. However, a more in-depth analysis is needed to better understand the relevant physical mechanisms.

While the impact of the stochastic backscatter scheme is generally favorable, there are incidences where the backscatter scheme increases the systematic model error. One example is the increase in systematic error in Z500 over the Southern Hemisphere (not shown), which is likely caused by the exasperated tropical divergence and synoptic activity over the Southern Ocean.

While it is beyond the scope of this work to understand the mechanisms of tropical-extratropical interactions in

detail, the following observations are noteworthy: In the high-resolution experiment HIGHRES, the Northern Hemispheric circulation was more blocked, but tropical variability hardly changed. We might therefore hypothesize that the improvement was due to a better representation of local extratropical circulation features. Blocking is enhanced likely because the backscatter contribution from numerical dissipation is greater in the north-central Atlantic and Pacific just upstream of typical blocking locations, which has been shown to support blocking via spatially localized upscale transfers of kinetic energy (Fournier 2003).

In the experiments STOCH and PHYS, tropical convective waves were better represented and the convergence of upper tropospheric winds reduced. Hence, the improvements in the Northern Hemispheric extratropics might partly arise from a better representation of tropical variability as proposed by Ferranti et al. (1990). An additional experiment was conducted where the stochastic forcing was confined to the tropical band only (not shown). This confirmed that the improvements in the Z500 bias over the North American continent in STOCH are due to both: local extratropical refinement and tropical influence by teleconnection patterns. Additional work has been conducted to test the reverse: namely, to which degree the tropics respond to a better extratropical representation. Vitart and Jung (2010) relaxed the Northern Hemisphere to observations and allowed the tropics to evolve freely. Better capturing the Northern Hemispheric extratropical circulation and its variability did not improve tropical variability, suggesting that internal tropical dynamics are responsible for changes in the Kelvin wave propagation and MJO variability.

One can speculate if there will be a need for stochastic parameterizations as computational resources allow us to increase horizontal resolutions of weather and climate models to cloud-resolving levels. To the extent that stochastic parameterizations represent subgrid-scale fluctuations around the equilibrium state, they can be expected to play an even more important role as resolution increases. Insofar as they represent the upscale effect of unresolved dynamical processes, one would hope that their magnitude will become smaller as more and more interactions are explicitly resolved. Presumably this will depend partly on the capability of next-generation climate models to capture the shallower “ $-5/3$ ” slope of the atmospheric energy spectrum in addition to the “ -3 ” geostrophic regime.

In this study, the backscatter term was incorporated into the model after the traditional parameterization development (and tuning) had taken place. However, if one takes the view that stochastic parameterization should be an important element of next-generation climate models

(Epstein and Pitcher 1972; Lorenz 1975; Pitcher 1977; Palmer 2001), if only to provide reliable estimates of model uncertainty, then a fundamental conclusion of this study is that such a posteriori addition of stochasticity to an already tuned model is simply not viable. This in turn suggests that stochasticity must be incorporated at a very basic level within the design of physical process parameterizations and improvements to the dynamical core.

Acknowledgments. We thank Paco Doblas-Reyes, Antje Weisheimer, and Roberto Buizza for many motivating discussions throughout the years. The Newton Institute in Cambridge hosted the first and third author during their program on “Stochastic modeling in the climate sciences” in 2010, which gave them the opportunity to work on this manuscript. We are indebted to Annabel Bowen and Bob Hine for improving the graphics. Thanks to Joe Tribbia and three anonymous reviewers whose insightful suggestions and comments improved earlier versions of this manuscript.

REFERENCES

- Adler, R. F., and Coauthors, 2003: The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeorol.*, **4**, 1147–1167.
- Anderson, D., and Coauthors, 2007: Development of the ECMWF seasonal forecast System 3. ECMWF Tech. Memo. 503, 56 pp. [Available online at http://www.ecmwf.int/publications/library/ecpublications/_pdf/tm/501-600/tm503.pdf.]
- Bechtold, P., M. Köhler, T. Jung, F. Doblas-Reyes, M. Leutbecher, M. Rodwell, F. Vitart, and G. Balsamo, 2008: Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales. *Quart. J. Roy. Meteor. Soc.*, **134**, 1337–1351.
- Berner, J., 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.*, **A366**, 2559–2577.
- , G. Shutts, M. Leutbecher, and T. Palmer, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.*, **66**, 603–626.
- , S.-Y. Ha, J. P. Hacker, A. Fournier, and C. Snyder, 2011: Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Mon. Wea. Rev.*, **139**, 1972–1995.
- Bowler, N. E., A. Arribas, S. E. Beare, K. R. Mylne, and G. J. Shutts, 2009: The local ETKF and SKEB: Upgrades to the MOGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **135**, 767–776.
- Branković, C., C. Jakob, M. Miller, A. Untch, and N. Wedi, 2002: Climate diagnostics of the ECMWF AMIP-2 simulations. ECMWF Tech. Memo. 360, 38 pp. [Available online at http://www.ecmwf.int/publications/library/ecpublications/_pdf/tm/301-400/tm360.pdf.]
- Charron, M., G. Pellerin, L. Spacek, P. L. Houtekamer, N. Gagnon, H. L. Mitchell, and L. Michelin, 2010: Toward random sampling of model error in the Canadian Ensemble Prediction System. *Mon. Wea. Rev.*, **138**, 1877–1901.
- D’Andrea, F., and Coauthors, 1998: Northern Hemisphere atmospheric blocking as simulated by 15 atmospheric general circulation models in the period 1979–1988. *Climate Dyn.*, **14**, 385–407.
- Doblas-Reyes, F., and Coauthors, 2009: Addressing model uncertainty in seasonal and annual dynamical seasonal forecasts. *Quart. J. Roy. Meteor. Soc.*, **135**, 1538–1559.
- Epstein, E. S., and E. J. Pitcher, 1972: Stochastic analysis of meteorological fields. *J. Atmos. Sci.*, **29**, 244–257.
- Ferranti, L., T. N. Palmer, F. Molteni, and F. M. E. Klinker, 1990: Tropical–extratropical interaction associated with the 30–60-day oscillation and its impact on medium and extended range prediction. *J. Atmos. Sci.*, **47**, 2177–2199.
- Fournier, A., 2003: Atmospheric energetics in the wavelet domain. Part II: Time-averaged observed atmospheric blocking. *J. Atmos. Sci.*, **60**, 319–338.
- Gates, W. L., and Coauthors, 1999: An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull. Amer. Meteor. Soc.*, **80**, 29–55.
- Hannay, C., D. Williamson, J. J. Hack, J. T. Kiehl, J. G. Olson, S. A. Klein, C. S. Bretherton, and M. Köhler, 2009: Evaluation of forecasted southeast Pacific stratocumulus in the NCAR, GFDL, and ECMWF models. *J. Climate*, **22**, 2871–2889.
- Hazeleger, W., and Coauthors, 2010: EC-Earth: A seamless earth-system prediction approach in action. *Bull. Amer. Meteor. Soc.*, **91**, 1357–1363.
- Iacono, M., J. Delamere, E. Mlawer, M. Shephard, S. Clough, and W. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, doi:10.1029/2008JD009944.
- Jung, T., 2005: Systematic errors of the atmospheric circulation in the ECMWF forecasting system. *Quart. J. Roy. Meteor. Soc.*, **131**, 1045–1073.
- , and A. M. Tompkins, 2003: Systematic errors in the ECMWF forecasting system. ECMWF Tech. Memo. 422, 72 pp. [Available online at http://www.ecmwf.int/publications/library/ecpublications/_pdf/tm/401-500/tm422.pdf.]
- , T. N. Palmer, and G. J. Shutts, 2005: Influence of a stochastic parameterization on the frequency of occurrence of North Pacific weather regimes in the ECMWF model. *Geophys. Res. Lett.*, **32**, L23811, doi:10.1029/2005GL024248.
- , and Coauthors, 2010: The ECMWF model climate: Recent progress through improved physical parametrizations. *Quart. J. Roy. Meteor. Soc.*, **136**, 1145–1160.
- , and Coauthors, 2012: High-resolution global climate simulations with the ECMWF model in Project Athena: Experimental design, model climate, and seasonal forecast skill. *J. Climate*, **25**, 3155–3172.
- Kimoto, M., and M. Ghil, 1993: Multiple flow regimes in the Northern Hemisphere winter. Part I: Methodology and hemispheric regimes. *J. Atmos. Sci.*, **50**, 2625–2644.
- Klinker, E., and P. Sardeshmukh, 1992: The diagnosis of mechanical dissipation in the atmosphere from large-scale balance requirements. *J. Atmos. Sci.*, **49**, 608–627.
- Lin, J., and Coauthors, 2006: Tropical intraseasonal variability in 14 IPCC AR4 climate models. Part I: Convective signals. *J. Climate*, **19**, 2665–2690.
- Lorenz, E. N., 1975: Climate predictability. *The Physical Basis of Climate and Climate Modelling*, GARP Publ. Series 16, 132–136.

- Madden, R. A., and P. R. Julian, 1971: Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.*, **5**, 702–708.
- Mason, P., and D. Thomson, 1992: Stochastic backscatter in large-eddy simulations of boundary layers. *J. Fluid Mech.*, **242**, 51–78.
- Morcrette, J.-J., H. W. Barker, J. N. S. Cole, M. J. Iacono, and R. Pincus, 2008: Impact of a new radiation package, McRad, in the ECMWF integrated forecasting system. *Mon. Wea. Rev.*, **136**, 4773–4798.
- Palmer, T. N., 2001: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parameterization in weather and climate prediction. *Quart. J. Roy. Meteor. Soc.*, **127**, 279–304.
- , and A. Weisheimer, 2011: Diagnosing the causes of bias in climate models—Why is it so hard? *Geophys. Astrophys. Fluid Dyn.*, **105**, 351–365.
- Pincus, R., H. Barker, and J.-J. Morcrette, 2003: A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous clouds. *J. Geophys. Res.*, **108D**, 4376, doi:10.1029/2002JD003322.
- Pitcher, E. J., 1977: Application of stochastic dynamic prediction to real data. *J. Atmos. Sci.*, **34**, 3–21.
- Randall, D. A., M. Khairoutdinov, A. Arakawa, and W. Grabowski, 2003: Breaking the cloud parameterization deadlock. *Bull. Amer. Meteor. Soc.*, **84**, 1547–1564.
- Rodwell, M., and T. Palmer, 2007: Using numerical weather prediction to assess climate models. *Quart. J. Roy. Meteor. Soc.*, **133**, 129–146.
- Shutts, G. J., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079–3102.
- , 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, 2003: The use of high resolution numerical simulations of tropical circulation to calibrate stochastic physics schemes. *Proc. ECMWF/CLIVAR Workshop on Simulation and Prediction of Intra-Seasonal Variability with Emphasis on the MJO*, Reading, United Kingdom, ECMWF, 83–102.
- Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.*, **132**, 3019–3032.
- Solomon, S., D. Qin, M. Manning, M. Marquis, K. Averyt, M. M. B. Tignor, H. L. Miller Jr., and Z. Chen, Eds., 2007: *Climate Change 2007: The Physical Science Basis*. Cambridge University Press, 996 pp.
- Stainforth, D., and Coauthors, 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433**, 403–406.
- Steinheimer, M., M. Hantel, and P. Bechtold, 2007: Convection in Lorenz's global energy cycle with the ECMWF model. ECMWF Tech. Memo. 545, 37 pp. [Available online at http://www.ecmwf.int/publications/library/ecpublications/_pdf/tm/501-600/tm545_rev.pdf.]
- Tennant, W. J., G. J. Shutts, and S. A. T. A. Arribas, 2011: Using a stochastic kinetic energy backscatter scheme to improve MOGREPS probabilistic forecast skill. *Mon. Wea. Rev.*, **139**, 1190–1206.
- Tibaldi, S., and F. Molteni, 1990: On the operational predictability of blocking. *Tellus*, **42A**, 343–365.
- Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012.
- Vitart, F., 2004: Monthly forecasting at ECMWF. *Mon. Wea. Rev.*, **132**, 2761–2779.
- , and T. Jung, 2010: Impact of the northern extratropical circulation on the skill in predicting the Madden–Julian oscillation. *Geophys. Res. Lett.*, **37**, L23085, doi:10.1029/2010GL045465.
- , and F. Molteni, 2010: Simulation of the Madden–Julian oscillation and its teleconnections in the ECMWF forecast system. *Quart. J. Roy. Meteor. Soc.*, **136**, 842–855.
- , M. A. Balmaseda, L. Ferranti, and D. Anderson, 2003: Westerly wind events and the 1997/98 El Niño event in the ECMWF seasonal forecasting system: A case study. *J. Climate*, **16**, 3153–3170.
- Weisheimer, A., T. N. Palmer, and F. J. Doblas-Reyes, 2011: Assessment of representations of model uncertainty in monthly and seasonal forecast ensembles. *Geophys. Res. Lett.*, **38**, L16703, doi:10.1029/2011GL048123.
- Wheeler, M., and G. N. Kiladis, 1999: Convectively coupled equatorial waves: Analysis of clouds and temperature in the wavenumber–frequency domain. *J. Atmos. Sci.*, **56**, 374–399.