Annu. Rev. Earth Planet. Sci. 2005. 33:163–93
doi: 10.1146/annurev.earth.33.092203.122552


# REPRESENTING MODEL UNCERTAINTY IN WEATHER AND CLIMATE PREDICTION

T.N. Palmer, G.J. Shutts, R. Hagedorn, F.J. Doblas-Reyes, T. Jung, and M. Leutbecher

*European Centre for Medium-Range Weather Forecasts (ECMWF), Shinfield Park, Reading RG2 9AX, United Kingdom; email: tim.palmer@ecmwf.int*



**Key Words** multimodel ensembles, stochastic-dynamic parameterization

■ **Abstract** Weather and climate predictions are uncertain, because both forecast initial conditions and the computational representation of the known equations of motion are uncertain. Ensemble prediction systems provide the means to estimate the flow-dependent growth of uncertainty during a forecast. Sources of uncertainty must therefore be represented in such systems. In this paper, methods used to represent model uncertainty are discussed. It is argued that multimodel and related ensembles are vastly superior to corresponding single-model ensembles, but do not provide a comprehensive representation of model uncertainty. A relatively new paradigm is discussed, whereby unresolved processes are represented by computationally efficient stochastic-dynamic schemes.

*I believe that the ultimate climatic models. . .will be stochastic, i.e., random numbers will appear somewhere in the time derivatives.*

Lorenz (1975)

## INTRODUCTION

There can be little doubt that among the finest achievements of twentieth-century climate science are the detailed numerical models of the atmosphere, oceans, and land surface. As we enter the twenty-first century, these models are becoming increasingly comprehensive representations of the so-called Earth-System, and include, for example, a range of biogeochemical feedbacks. Such models, based on possibly tens of millions of degrees of freedom, are used routinely to make forecasts of weather on daily timescales and climate variability on seasonal timescales, and they are the basis of detailed assessments of humankind's impact on climate.

The core of this development is the numerical weather prediction (NWP) model. Weather prediction attained scientific status with the publication of Wilhelm Bjerknes' article (Bjerknes 1904) that discussed the problem from the perspective of scientific determinism: An accurate forecast of the weather can be made given

sufficient knowledge of the atmospheric initial conditions, and sufficient means of integrating the known equations of motion. L.F. Richardson (1922) put mathematical structure to Bjerknes' argument in his seminal treatise on the development of numerical methods for integrating the atmospheric equations of motion. In this work, Richardson proposed the now-familiar spatial discretization grid and immediately faced the problem of how to represent physical processes that were not explicitly resolved by this grid. Motivated by earlier work of Osborne Reynolds, he developed a set of local deterministic bulk formulas that attempted to estimate the mean impact of a supposed statistical ensemble of subgrid processes in secular equilibrium with the resolved-scale flow. This paradigm is illustrated schematically in Figure 1*a*. Although discretization grids have become increasingly finer and parameterizations increasingly complex since Richardson's treatise, the underlying paradigm has not changed over the years.

Of course, weather and climate predictions are uncertain, not only because of uncertainty in the forecast initial conditions. Because of the underlying nonlinearity of the equations of motion, the growth of inevitable uncertainties in forecast initial conditions is flow dependent. To estimate such flow-dependent predictability, it is now commonplace to run ensembles of weather or climate prediction models from a set of initial conditions, each element of which is consistent with the initial observations (see Ensemble Prediction as the Basis for Environmental Risk Assessment, below). However, uncertainty in the initial conditions is not the only source of inaccuracy in weather and climate prediction: The models themselves are inaccurate. Representations of model uncertainty must also be included in ensemble

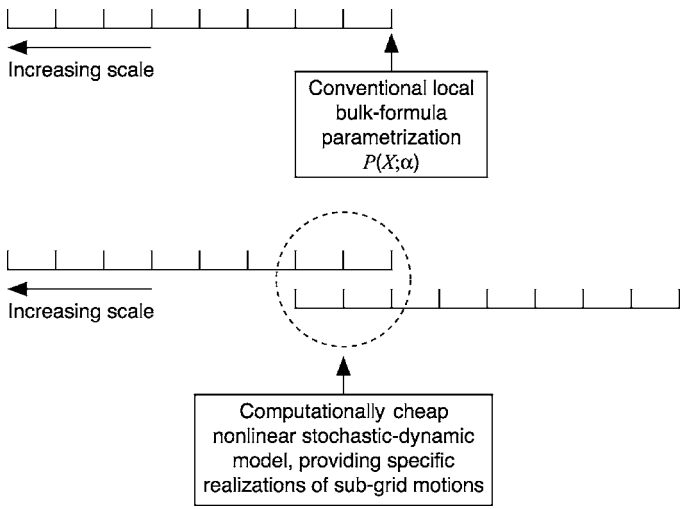



**Figure 1** (*Upper*) Schematic for conventional weather and climate prediction models (the Reynolds/Richardson Paradigm). (*Lower*) Schematic for weather and climate prediction using simplified stochastic-dynamic model representations of unresolved processes.

prediction systems so that ensemble-based forecast probability distributions are not underdispersive or overconfident.

Within the framework of the conventional paradigm shown in Figure 1*a*, a number of approaches used to represent model uncertainty have arisen and are summarized in the following hierarchical way.

- The multimodel ensemble comprises quasi-independent climate models (differing in numerics and parameterization) developed in different institutes around the world. Results from the DEMETER project (Palmer et al. 2004) illustrate both enhanced reliability and limitations of the multimodel ensemble (see Multimodel, Multiparameterization, and Multiparameter Ensembles, below).
- In the multiparameterization ensemble a number of quasi-independent parameterizations that have been installed within a single weather or climate model infrastructure. Houtekamer et al. (1996) use this approach in the Meteorological Service of Canada's operational ensemble prediction system for weather forecasting.
- In the multiparameter ensemble parameters are perturbed within a fixed set of parameterizations in a given climate model. The Hadley Centre QMUP (Quantifying Model Uncertainty Project; Murphy et al. 2004) and the climateprediction.net project (Allen & Stainforth 2002, Stainforth et al. 2005) are examples of the multiparameter approach in the context of climate-change prediction.

In Inherent Errors in the Reynolds/Richardson Paradigm (see below), we discuss whether model uncertainty is indeed fully represented by the approaches listed above. More specifically we discuss whether the paradigm in Figure 1*a* is an adequate framework within which to represent all sources of model uncertainty. We argue that it is not, first by considering the energy spectrum of the atmosphere and the implication of this spectrum for bulk-formula parameterization, and second by explicit calculations, using a cloud-resolving model, of coarse-grained tendencies (Shutts & Palmer 2004). These considerations led us to consider a rather broader paradigm for the development of weather and climate models (Figure 1*b*) in which the subgrid processes are represented by stochastic-dynamic dynamical systems. The key conceptual differences are twofold. First, the subgridscale tendencies are no longer thought of as representing a statistical average, but rather as an actual and specific realization of the subgrid world. Second, the stochastic-dynamic system is coupled with the climate model over a range of scales near the truncation scale, not only at the truncation scale itself. We argue that, in addition to a better realization of model uncertainty, this approach brings with it the prospect of a reduction in some long-standing model systematic errors through a "noise-induced drift" effect.

Specific examples of realizations of the paradigm illustrated in Figure 1*b* are discussed below. In Stochastic Parameterization, we focus on the impact of purely stochastic subgrid models through a hierarchy of models. In Stochastic-Dynamic

Subgrid Models, we discuss more recent work developing and using stochastic-dynamic subgrid models.

## ENSEMBLE PREDICTION AS THE BASIS FOR ENVIRONMENTAL RISK ASSESSMENT

Ensemble prediction provides a means of forecasting uncertainty in weather and climate prediction. The scientific basis for ensemble forecasting is that in a nonlinear system, which the climate system surely is, the finite-time growth of initial error is dependent on the initial state. For example, Figure 2 (see color insert) shows the flow-dependent growth of initial uncertainty associated with three different initial states of the Lorenz (1963) model. Hence, in Figure 2*a* uncertainties grow significantly more slowly than average (where local Lyapunov exponents are negative), and in Figure 2*c* uncertainties grow significantly more rapidly than one would expect on average. Conversely, estimates of forecast uncertainty based on some sort of climatological-average error would be unduly pessimistic in Figure 2*a* and unduly optimistic in Figure 2*c*.

Ensemble prediction has now been operational at the European Centre for Medium-Range Weather Forecasts (ECMWF) and at the United States National Centers for Environmental Prediction (NCEP) (Palmer et al. 1993, Toth & Kalnay 1993) for more than a decade. The Meteorological Service of Canada also routinely runs ensemble weather forecasts (Houtekamer et al. 1996). As an example, Figure 3 (see color insert) shows the individual ensemble members from the ECMWF ensemble initialized 42 h before the devastatingly severe storm Lothar crossed France. The ensemble forecast suggests that the predictability of this weather event was extremely low (akin to the exceptional forecast spread shown in Figure 2*c*). Nevertheless, the ensemble prediction system indicated a significant probability, or risk, that some exceptional event would occur. Such a probabilistic forecast was more preferable to the deterministic prediction from the best-guess initial state, which gave no clue of the likelihood of this severe weather event.

Different approaches to validating ensemble forecasts use, for example, Brier Skill Scores, Ranked Probability Skill Scores, Relative Operating Characteristic, and Potential Economic Value (for example, see Jolliffe & Stephenson 2003). With regard to the last, Figure 4 shows one specific analysis that addresses the economic value of ensemble forecasts. Imagine a casino that accepts bets on the daily maximum temperature at Heathrow airport, at forecast lead times from D + 1 to D + 10 days. A gambler has access to the ECMWF ensemble prediction system and places bets on temperatures forecast by the ensemble, where the size of the bet is proportional to the forecast probability of the occurrence of a specific temperature. The casino, on the other hand, determines the payout using the ECMWF high-resolution model, converted to a simple Gaussian probability density function (PDF) by climatological error statistics. Results shown in Figure 4 indicate that the gambler wins over the house over the course of time, especially for the longer

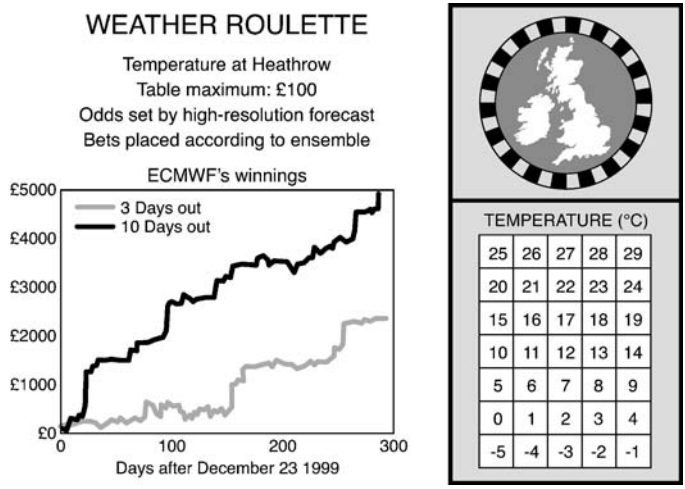



**Figure 4** An assessment of the economic value of the ECMWF ensemble prediction system compared with the deterministic forecast (M. Roulston & L. Smith, personal communication).

lead times. These results are entirely consistent with a general assessment of the potential economic value of ensemble prediction systems in weather forecasting (Palmer 2002).

Figure 5 shows an intercomparison of ensemble skill and ensemble spread from the three operational ensemble systems mentioned above (Buizza et al. 2004). In a well-calibrated system, the spread about the ensemble mean should equal the error of the ensemble mean; the extent to which they are not indicates deficiencies in the ensemble systems. At initial time, results show that the ECMWF ensemble is better calibrated than the other systems—this results from the use of rapidly growing singular vectors (Molteni & Palmer 1993, Buizza & Palmer 1995) as initial perturbations. By contrast, the use of breeding vectors (Toth & Kalnay 1993) or initial perturbations on the basis of ensemble data assimilation (Houtekamer et al. 1996) led to excessively weak growth, compensated for by excessively large initial amplitudes.

Notwithstanding these differences, all three systems show a common deficiency toward the end of the 10-day forecast range: The ensembles are consistently underdispersive. Although the amplitude of initial perturbations could be increased such that spread and error be matched at day 10, this would lead to completely unacceptable performance in the earlier ranges, with a grossly overdispersive ensemble (and correspondingly low Brier Skill Scores). Rather, it would appear more likely that this common deficiency indicates that none of these ensemble systems has an adequate representation of model uncertainty.

This deficiency becomes even more apparent when looking at longer lead times. Figure 6 shows ensemble predictions of El Niño sea surface temperature (SST)

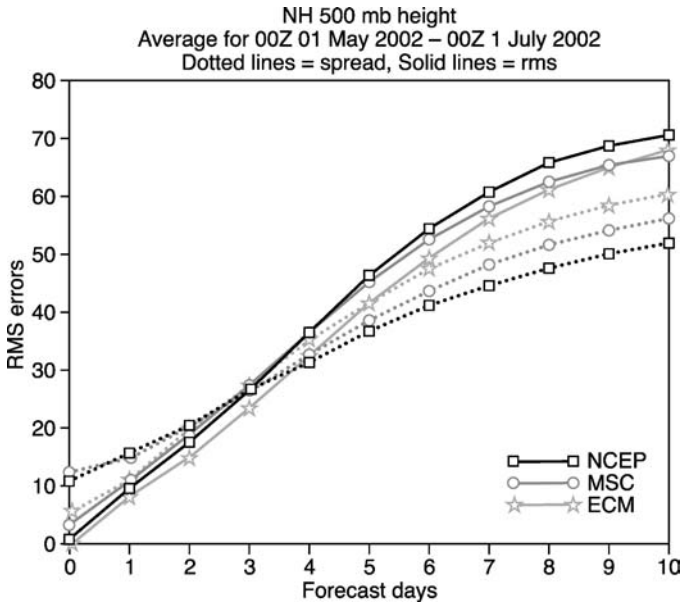



**Figure 5** Solid lines represent the RMS error of 500 hPa height of ensemble mean forecasts from three operational forecast centers [ECMWF, NCEP, and MSC (Meteorological Service of Canada)]. Dashed lines represent the RMS spread from the same three forecast systems. In a well-calibrated system, the mean spread should equal the mean error. All three systems have insufficient spread toward the end of the forecast range. Data are from Buizza et al. 2004.

anomalies based on the ECMWF operational seasonal forecast system for the 1997–1998 El Niño (Stockdale et al. 1998). Although forecast skill exceeds that of persistence, even at short lead times, the forecast ensemble is underdispersive; again it would appear likely that this deficiency arises from the fact that this seasonal prediction system has no representation of model uncertainty (Figure 6 also shows the impact of a stochastic parameterization scheme, discussed below).

These results indicate that single-model ensembles do not yield reliable probabilities. Consider a "dichotomous event" such as E, defined to have occurred if the 2 m temperature is greater than zero, and to have not occurred if the 2 m temperature is not greater than zero. Consider a set of ensemble seasonal forecasts made, for example, over a 20-year period, and consider also a set of individual grid points. Within this period, there might be individual ensemble forecasts in which E was forecast to occur with certainty. Conversely, there might be individual ensemble forecasts in which E was forecast not to occur with certainty. More generally, there will be a subset of forecasts in which E was predicted to occur with probability $p$. If the forecast system is reliable, then within this subset E should be expected to occur a fraction $p$ of occasions. If a forecast ensemble is underdispersive, then

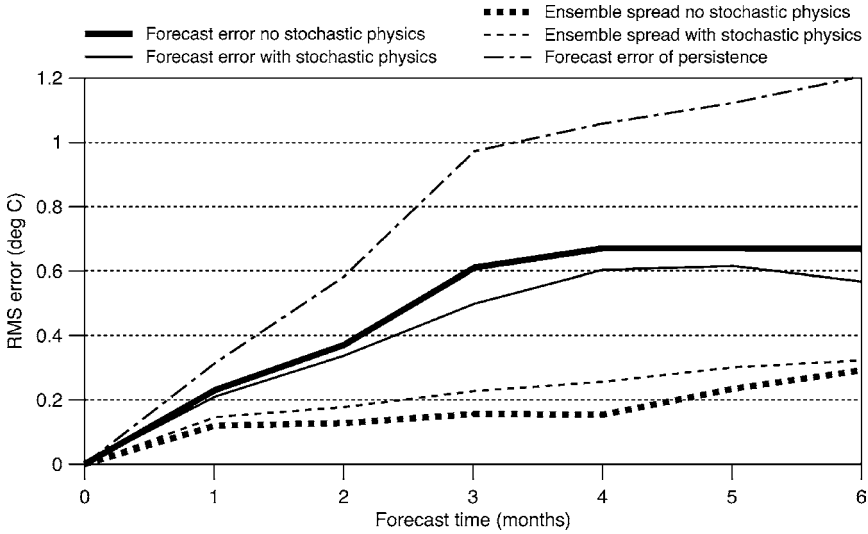



**Figure 6** Solid lines represent the RMS error of NINO3 from the EMCWF seasonal forecast system with and without the stochastic physics scheme of Equation 12. Dashed lines represent the corresponding RMS spread. Skill of a forecast in which initial anomalies are persisted is also shown (T. Stockdale & M. Balmaseda, personal communication).

E will tend to be predicted with probability close to one, or close to zero, too often. As discussed below, seasonal forecasts from single-model ensembles are not reliable.

The results raise fundamental questions concerning the reliability of climate-change prediction based on single-model integrations. If forecast probability distributions based on single-model ensembles are not reliable on the seasonal timescale, it is unlikely that they will be reliable on longer timescales relevant to climate-change assessments.

## MULTIMODEL, MULTIPARAMETERIZATION, AND MULTIPARAMETER ENSEMBLES

As discussed in the Introduction, within the Reynolds/Richardson paradigm, methodologies for representing model uncertainty can be classed in terms of multimodel, multiparameterization, and multiparameter approaches. In this section, we focus on these approaches, emphasizing the multimodel ensemble approach.

### Multimodel Ensembles

Early work was directed at the weather prediction problem (Harrison et al. 1999), for which evidence was found that multimodel systems could be more reliable than

single-model systems. Here we focus attention on seasonal and climate-change timescales.

To date, the most comprehensive assessment of the skill of multimodel ensembles has come from the European Union DEMETER project (Development of a European Multimodel Ensemble for seasonal to inTERannual climate prediction; Palmer et al. 2004, http://www.ecmwf.int/research/demeter). In DEMETER, seven different global coupled ocean-atmosphere models (six of which were installed on a single supercomputer system) were each run over six-month integration periods in ensemble mode, through much of the ERA-40 (http://www.ecmwf.int/research/era/) period. For each single model, an ensemble comprising nine members was made by perturbing the atmosphere-ocean initial conditions consistent with initial uncertainty (only).

Figure 7 shows reliability diagrams for the seven single-model ensembles and for the multimodel ensemble; for the prediction event E, the JJA mean 2 m temperature is above normal for grid points in the tropics. The reliability diagram shows the observed frequency of E for the subset of forecasts where E was predicted with probability $p$. Consistent with the characteristic underdispersion of single-model ensembles discussed in this section, every single-model ensemble proves to be overconfident, characterized by an excessively shallow slope of the reliability curve. By contrast, the reliability curve for the multimodel ensemble fits the diagonal much better.

Reliability is an important attribute for an ensemble forecast; however, it does not guarantee its usefulness. A forecast of the climatological probability of E is reliable. The ability of an ensemble system to discriminate between situations that lead to different events in the future is called resolution. Both reliability and resolution are components of the Brier Skill Score, which assesses performance relative to a forecast of the climatological probability of occurrence of the event. As discussed in Hagedorn et al. (2005), both reliability and resolution components of the multimodel ensemble are improved compared with the single-model performance. This results in a significant overall improvement of the Brier Skill Score.

Of course, it can be asked whether the improvement in the multimodel ensemble arises purely from its larger size. To separate the benefits that are derived from combining models of different formulation from those simply associated with the accompanying increase in ensemble size, a 54-member ensemble hindcast has been generated by the ECMWF model alone for the period of 1987 to 1999 using the May start date. Although the increase in size improved reliability of the single-model ensemble, it did not outperform the multimodel ensemble, and both reliability and resolution and Brier Skill Score are smaller for the single model compared with the multimodel ensemble (see Hagedorn et al. 2005 for details).

Hagedorn et al. (2005) show that the more skillful performance of the multimodel over the single-model ensemble also holds for seasonal temperature events over many different regions (e.g., Europe, North America, and West Africa). However, the performance of the multimodel appears less impressive for precipitation events. For example, forecasts for seasonal precipitation in the upper tercile over

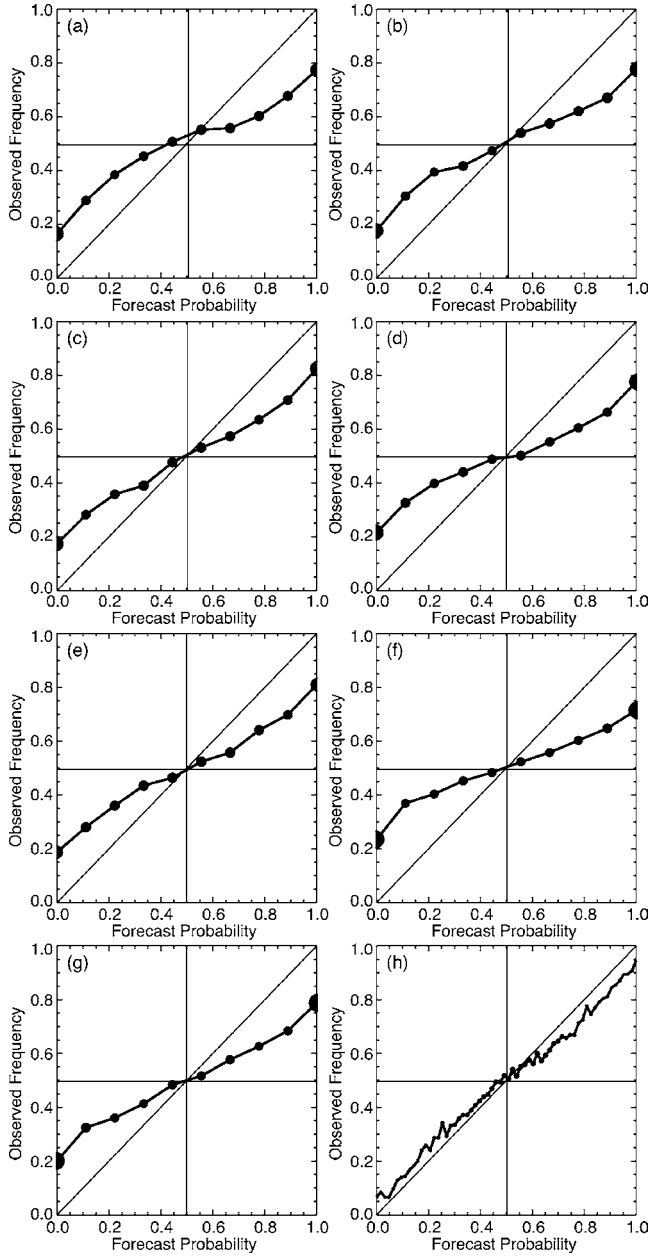



**Figure 7** (*a–g*) Reliability diagram for the event: seasonal mean 2 m temperature for months 2 through 4 is above average for all grid points in the tropics, for single-model ensembles. (*h*) Corresponding reliability diagram for DEMETER multimodel ensemble.

Europe are not as reliable as forecasts for temperature (Figure 8). Here the single-model ensembles are so unreliable that the multimodel combination appears to help little. This result indicates that the multi-model ensemble cannot correct the systematic deficiencies of the component models.

These results are relevant to longer-timescale prediction. Multimodel ensembles have also been used to give a probabilistic assessment of the impact of anthropogenic greenhouse forcing on risk of wintertime flooding. Figure 9 (see color insert) (from Palmer & Räisänen 2002) is based on an analysis of the CMIP2 multimodel ensemble (Meehl et al. 2000). One fundamental question in appraising this analysis concerns the reliability of the CMIP2 multimodel ensemble; in other words, how well is model uncertainty represented in the CMIP2 multimodel ensemble? A partial answer to this question can be provided by evaluating the reliability of the same multi-model ensemble over periods for which verification data exist.

Such analysis has not been performed on the CMIP2 ensemble. However, for the purposes of this paper, let us assume that the CMIP2 ensemble has the same reliability as the DEMETER ensemble. Even for relatively "difficult" areas such as Europe, the seasonal timescale multimodel probability of temperature is remarkably reliable, which suggests that corresponding probabilistic forecasts of climate change might be similarly reliable. By contrast, the multimodel probability forecasts of extreme seasonal precipitation events are much less reliable. This indicates that some caution is needed in drawing firm conclusions about results shown in Figure 9.

## Multiparameterization Ensembles

Houtekamer et al. (1996) present a system simulation approach to ensemble weather prediction in which initial perturbations are obtained by randomly perturbing initial observations. For each perturbation, an independent six-hour assimilation cycle is performed. To obtain model perturbations, different parameterizations of horizontal diffusion, deep convection, radiation, gravity wave drag, and orography were selected. In addition, perturbations to surface fields such as sea surface temperature, roughness length, and albedo were made. Despite this apparently comprehensive approach to the representation of initial and model uncertainty, with realistic parameters, the rms spread in the ensemble is smaller than it should be by a factor of about 2 (also see Figure 5). The authors conclude that their multiparameterization approach may not be adequate to describe model uncertainty and that "less conventional and perhaps more dramatic perturbations" are needed (see below).

## Multiparameter Ensembles

Murphy et al. (2004) and Stainforth (2005) have both used multiparameter ensembles to quantify uncertainty in climate-change projections. In the Met Office climate model HadAM3, there are O(100) uncertain parameters. As reported by

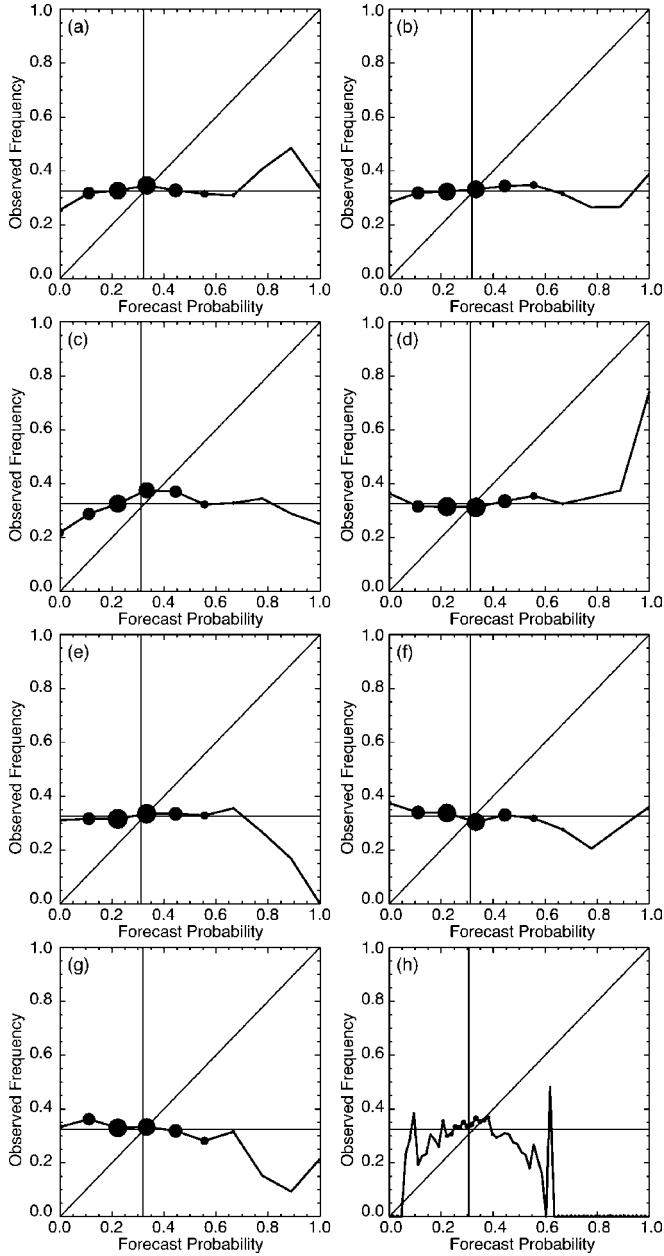



**Figure 8** As in Figure 7, but for the following event: precipitation for months 2 through 4 in the upper tercile for grid points over Europe.

Murphy et al. (2004), 29 of these were identified by modeling experts as controlling key physical characteristics of subgrid atmospheric and surface processes. In Murphy et al. (2004) the parameters were perturbed either by changing a logical switch or by setting a coefficient or threshold to a minimum, intermediate, or maximum value as defined by the experts. The resulting perturbed physics ensemble consisted of 53 model versions (HadAM3 coupled with a mixed-layer ocean model) used to make simulations of present-day and doubled $CO_2$ climate.

Results from a much larger (2578-member) ensemble of multiparameter integrations of HadAM3 coupled with a mixed-layer ocean have been achieved by the distributed computing method of climateprediction.net (Allen & Stainforth 2002, http://www.climateprediction.net), which requires the use of models with a resolution lower than those used in Murphy et al. (2004). The methodology explores combinations of perturbations in six parameters, selected as likely to have an impact on climate sensitivity.

In both studies, integrations over twentieth-century conditions are compared with observations, and those model integrations that fit poorly to the observations are rejected or given small weight in the ensemble.

One primary aim of Murphy et al. (2004) and Stainforth (2005) is to provide PDFs of climate sensitivity: the increase in global temperature associated with a doubling of atmospheric $CO_2$. In Murphy et al. (2004) the corresponding PDF gives a 90% confidence interval of 2.5 to 6 K, and in the climateprediction.net results there are apparently realistic model versions with climate sensitivities ranging from $<2$ to $>11$ K.



## INHERENT ERRORS IN THE REYNOLDS/RICHARDSON PARADIGM

The notion of the bulk-formula parameterization in weather and climate prediction models was largely motivated by the diffusive macroscopic representations of heat and momentum transfer associated with microscopic molecular motion, as determined by conventional statistical mechanics. Such bulk representations of molecular motions are accurate because of the extreme scale separation between the microscale and the macroscale.

For example, parameterization of orographic gravity-wave drag is based on the notion of a statistical ensemble of gravity waves forced by an incoherent spectrum of subgrid orography in approximate equilibrium with the large-scale flow. These waves exert a drag on the large-scale flow at wave-breaking levels. Similarly, parameterization of convection is based on the notion of an incoherent statistical ensemble of convective overturning circulations in approximate equilibrium with the large-scale temperature and humidity fields. In regions of ascent, latent-heat release is approximately balanced by adiabatic cooling, and the net effect of subgrid convective motions on the gridbox is a mean warming tendency arising from the broad regions of descent that are assumed to dominate the area of the gridbox.

Without the notion of an incoherent statistical ensemble of subgrid motions, subgrid orography need not act as a drag and convection need not produce a warming tendency. Examples are shown in Figure 10. For example, in the case in which the orography cannot be resolved by the large-scale model, but is nevertheless coherent across neighboring grids, the effect of subgrid orography may be to accelerate rather than retard the flow within some gridboxes. Similarly, when convection becomes organized and coherent across gridboxes, the effects of subgrid overturning may cool rather than warm the gridbox mean temperature. Also, in such situations, there is no requirement for the vertical momentum transfer to be downgradient (which it frequently is in convective momentum parameterization).

This failure of the conventional notion of parameterization is implied by the lack of any separation between resolved and unresolved scales in the atmosphere, and it is implied by the analysis of Nastrom & Gage (1985), which shows that the atmospheric energy density drops at $k^{-3}$ on scales larger than truncation scales for global Earth-System models; on scales smaller than this, the spectrum shallows to $k^{-5/3}$ (see Figure 11, below). Although the physical interpretation of this $k^{-5/3}$ spectrum is ambiguous, it signals that the notion of deterministic bulk-formula parameterization is not a rigorous concept. As discussed below, the T799 ECMWF model with conventional bulk-formula parameterization has no representation of the $k^{-5/3}$ spectrum.

To quantify better the failure of bulk-formula parameterization, Shutts & Palmer (2004) have performed a coarse-grained budget analysis of a cloud-resolving model, treated as "truth," to compare "exact" with parameterized subgrid tendencies. The cloud-resolving model is a modified form of the Met Office "Large Eddy Model" based on the anelastic, quasi-Boussinesq equations of motion (Shutts & Gray 1994). The model is configured for use in an equatorial beta-plane together with anisotropic grid spacing in the horizontal plane: $\Delta x = 1$ km in the zonal direction and operational NWP model resolution ($\Delta y = 40$ km) in the meridional direction. The model is run in aqua-planet mode with an idealized sea surface temperature field approximating that of the equatorial Pacific Ocean. A coarse-grained grid of $64 \times 80$ km zonal/meridional resolution is considered.

Consider now the thermodynamic equation:

$$\frac{\partial \theta}{\partial t} + V \cdot \nabla \theta = D, \tag{1}$$

where $\theta$ is the potential temperature, $V$ is the wind vector, and $D$ represents all diabatic source terms (e.g., latent heat release), including an imposed tropospheric cooling function equivalent to 1.5K day$^{-1}$. Denoting the average over each coarse-grained gridbox region by an overbar, Equation 1 can be expressed as

$$\frac{\partial \bar{\theta}}{\partial t} + \bar{V} \cdot \nabla \bar{\theta} = \bar{D} + \bar{V} \cdot \nabla \bar{\theta} - \overline{(V \cdot \nabla \theta)} = \widehat{Q}_1, \tag{2}$$

where the right-hand side ($\widehat{Q}_1$) represents the exact subgrid forcing seen by the coarse-grained flow. By contrast, the temperature tendency equation for a

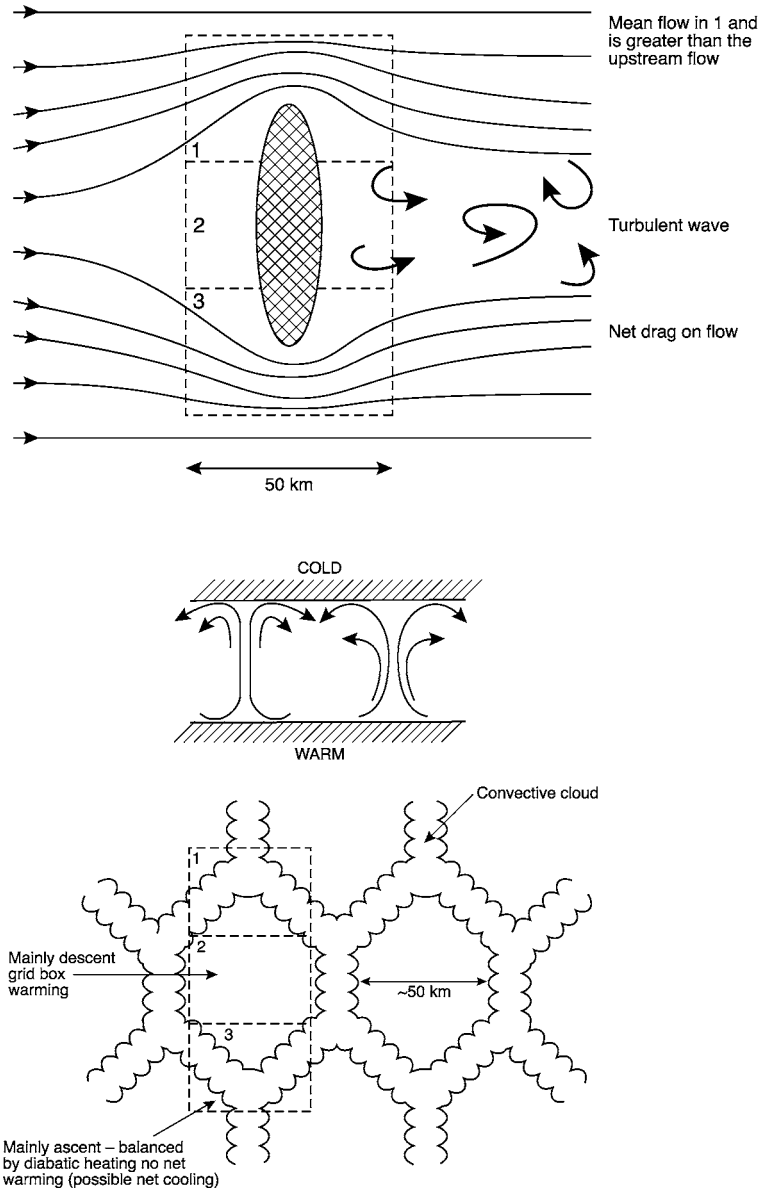



**Figure 10** Schematic examples of the failure of conventional parameterization to account for tendencies associated with subgrid processes. The top panel depicts when the subgrid topographic forcing is coherent across gridboxes, and the bottom panels depict a situation when the convective motions have mesoscale organization.

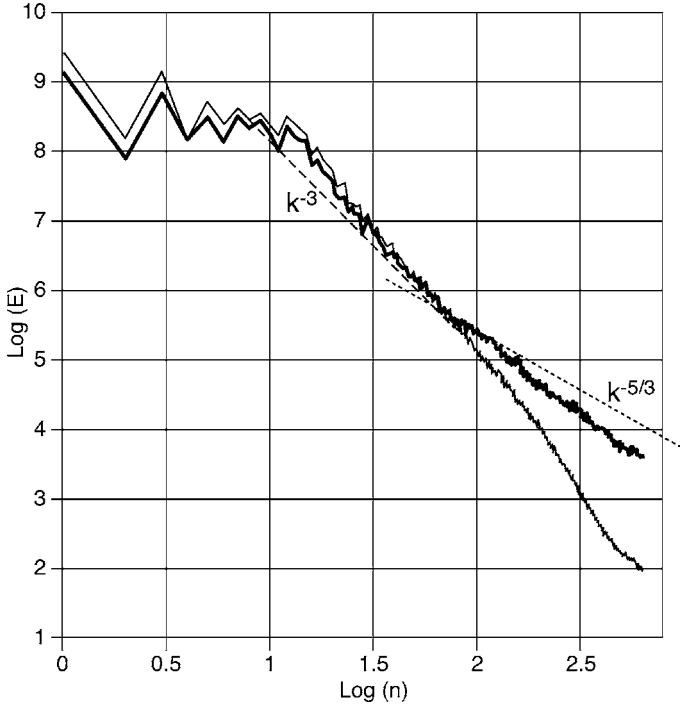



**Figure 11** Energy spectra at day 5 in ECMWF forecasts at T799 resolution: with cellular automaton stochastic backscatter (thick line) and without such backscatter (thin line). log(E) is the logarithm (base 10) of the energy density and $n$ is the spherical harmonic order.

conventional NWP model with truncation scale corresponding to the coarse-grained grid is written as

$$\frac{\partial \bar{\theta}}{\partial t} + \bar{V} \cdot \nabla \bar{\theta} \approx Q_1, \tag{3}$$

where $Q_1$ denotes the temperature tendency assuming conventional NWP parameterizations. What is the relationship between $Q_1$ and $\widehat{Q}_1$?

Figure 12 shows three PDFs of $\widehat{Q}_1$ using the convection scheme of Bechtold et al. (2001). These PDFs are conditioned on three separate ranges of $Q_1$: $-3 < Q_1 < +3$ K day$^{-1}$ (associated with weak convection), $+9 < Q_1 < +15$ K day$^{-1}$ (associated with moderate convection), and $+27 < Q_1 < +33$ K day$^{-1}$ (associated with strong convection).

Broadly speaking, if the bulk-formula parameterization concept were quantitatively valid, each of these conditional PDFs should correspond to a single-bin delta function. This is not the case; indeed, the spread of the conditional PDFs, at least in the strongly convecting cases, is one order of magnitude larger than the spread, 6 K day$^{-1}$, in the corresponding $Q_1$ bins. Indeed there appears to be a monotonic

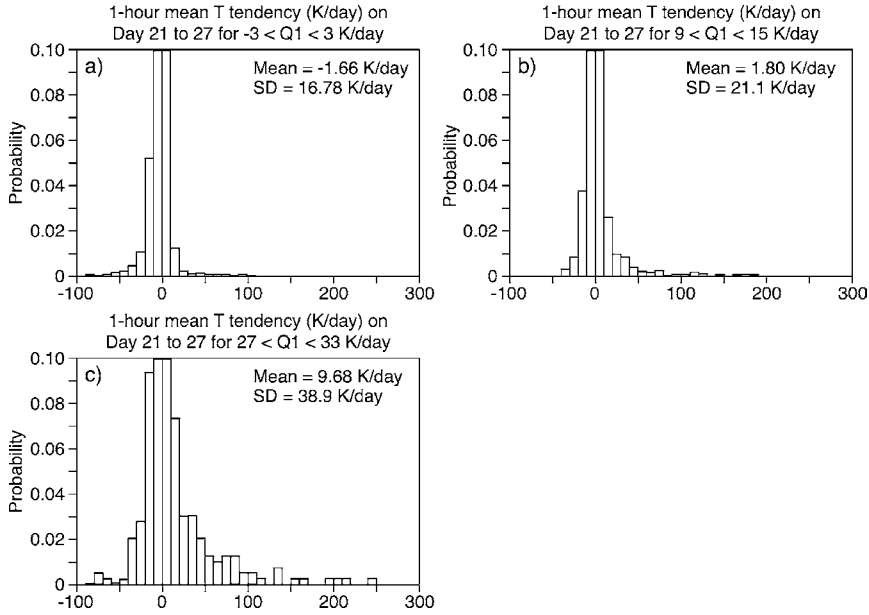



**Figure 12** PDFs of coarse-grained gridbox mean temperature tendencies $\widehat{Q}_1$ binned into three levels of parameterized convective forcing, $Q_1$. (*a*) $-3$ to $+3$ K day$^{-1}$, (*b*) $+9$ to $+15$ K day$^{-1}$, and (*c*) $+27$ to $+33$ K day$^{-1}$. The convective forcing strength $Q_1$ is computed by using the coarse-grained model fields as input to a convective parameterization scheme. The PDFs are derived from subsamples of all model points at z $=$ 5 km selected according to the indicated ranges of $Q_1$. Data are from figure 15 in Shutts & Palmer (2004).

and approximately linear relationship between $Q_1$ and the spread of the conditional PDFs. This provides some motivation for the use of a multiplicative noise term in the ECWMF stochastic parameterization scheme discussed below.

## STOCHASTIC PARAMETERIZATION

The underlying PDF of subgrid variability associated with fixed large-scale forcing is far from being the delta function required by bulk-formula parameterization. Suppose, then, instead of representing the subgrid tendency by the bulk formula, the corresponding PDF was sampled stochastically. What improvement would this bring? Some more deterministically minded readers might wonder how adding random numbers to an otherwise deterministic model could possibly improve forecast performance. Indeed, providing the bulk-formula parameterizations properly describe the mode of the associated distribution of subgrid tendencies, it is unlikely that a stochastic parameterization would improve conventional deterministic skill

scores (Nicolis 2005). However, there are two different ways in which stochastic parameterization might be beneficial. The first relates to issues already discussed, that a stochastic representation of model uncertainty makes the underlying ensembles less underdispersive, and consequently more reliable. A quantitative measure of the possible improvement that stochastic parameterization might bring would be in terms of an increase in ensemble-based probabilistic skill scores, for example, Brier Skill Score, Ranked Probability Skill Score, or Relative Operating Characteristic.

A second way in which stochastic parameterization might improve the performance of an NWP or climate model is by reducing systematic errors in the representation of its climatological mean state through noise-induced drift. Consider, for example, a ball bearing in a skewed potential well. Without noise, the ball bearing sits at the minimum of the well. Forced with Gaussian noise, the ball is most likely found displaced away from the minimum, toward the more shallow-sloping side of the well. This picture already has relevance to the real atmosphere. As shown, for example, in Corti et al. (1999), atmospheric low-frequency variability of the northern winter flow appears to have clear (non-Gaussian) regime behavior, and in recent decades, the dominant regime corresponds to anomalously westerly flow across midlatitudes. By contrast, blocking-type flow patterns correspond to subdominant regimes. As discussed by Molteni & Tibaldi (1990), models with insufficient transient activity would tend to overpopulate the dominant regimes and underpopulate the secondary regimes. In a recent study Jung (2005) suggests that this picture describes the systematic error of the most recent versions of the ECMWF model.

In the section below, we describe the impact that stochastic parameterization can have on model performance through a hierarchical sequence of models, from simple chaotic systems to comprehensive climate models.



## A Stochastic Parameterization of Lorenz (1963)

Consider the Lorenz (1963) model:

$$\begin{aligned} \dot{X} &= -\sigma X + Y \\ \dot{Y} &= -XZ + rX - Y \\ \dot{Z} &= XY - bZ. \end{aligned} \tag{4}$$

The familiar Lorenz attractor is illustrated in Figure 13*a* by Lorenz's original choice of parameters ($\sigma = 10$, b $= 8/3$, r $= 28$). On the basis of this parameter setting, the governing equations can be written in terms of the three empirical orthogonal functions of the Lorenz model, so that Equation 4 is transformed to (Selten 1995)

$$\begin{aligned} \dot{a}_1 &= 2.3a_1 - 6.2a_3 - 0.49a_1a_2 - 0.57a_2a_3 \\ \dot{a}_2 &= -62 - 2.7a_2 + 0.49a_1^2 - 0.49a_3^2 + 0.14a_1a_3 \\ \dot{a}_3 &= -0.63a_1 - 13a_3 + 0.43a_1a_2 + 0.49a_2a_3 \end{aligned} \tag{5}$$

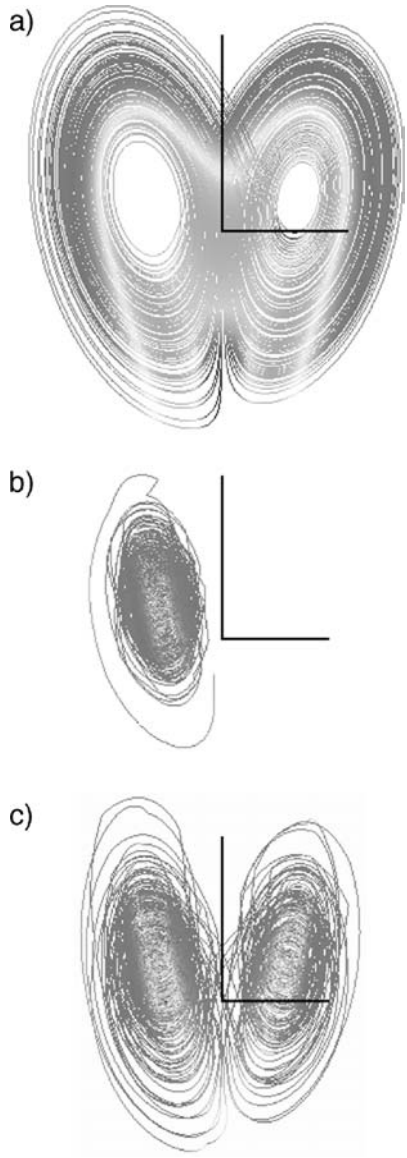



**Figure 13** (*a*) The full Lorenz (1963) attractor. (*b*) A representation of the Lorenz attractor using stochastic parameterization of the third principal component with weak and/or excessively white noise. (*c*) A representation of the Lorenz attractor using stochastic parameterization of the third principal component with realistic noise.

for principal components $a_1$, $a_2$, and $a_3$. The third empirical orthogonal function explains only 4% of the total variance of the system. Hence, it might be imagined (by analogy with bulk-formula parameterization) that a reasonable approximation to the full system could be obtained by truncating the system equations to two principal components, with the third principal component given as a deterministic "parameterized" function of the first two, that is

$$\begin{aligned} \dot{a}_1 &= 2.3a_1 - 6.2a_3 - 0.49a_1a_2 - 0.57a_2a_3 \\ \dot{a}_2 &= -62 - 2.7a_2 + 0.49a_1^2 - 0.49a_3^2 + 0.14a_1a_3. \\ a_3 &= f(a_1, a_2) \end{aligned} \tag{6}$$

Such a truncated model is a reasonable short-range forecast model in the sense that the initial tendencies $\dot{a}_1$ and $\dot{a}_2$ in Equation 6 can be made to agree well with those in Equation 5 for points on the Lorenz attractor. However, because Equation 6 is a system of two autonomous ordinary differential equations, we know from the Poincaré-Bendixon theorem that it cannot exhibit chaotic dynamics and indeed, as shown below, exhibits gross systematic errors.

Consider, then, the representation

$$\begin{aligned} \dot{a}_1 &= 2.3a_1 - 6.2a_3 - 0.49a_1a_2 - 0.57a_2a_3 \\ \dot{a}_2 &= -62 - 2.7a_2 + 0.49a_1^2 - 0.49a_3^2 + 0.14a_1a_3, \\ a_3 &= \beta \end{aligned} \tag{7}$$

where $\beta(t)$ is a stochastic variable randomly drawn from a Gaussian PDF with variance equal to that associated with $a_3$. Figure 13*b*,*c* show the impact of such a stochastic parameterization in which the noise in Figure 13*b* can be interpreted either as having less amplitude than that in Figure 13*c* or, for given amplitude, as being whiter (see Palmer 2001). In Figure 13*b*, the system PDF is monomodal and, compared with Figure 13*a*, has errors in both variability and mean state. In Figure 13*c*, the system PDF is bimodal and the overall PDF agrees in basic shape with the full Lorenz attractor.

## Stochastic Parameterization of Lorenz

We now progress to the more complex Lorenz (1996) system:

$$\begin{aligned} \dot{X}_i &= -X_{i-2}X_{i-1} + X_{i-1}X_{i+1} - X_i + F - \frac{c}{b}\sum_{j=1}^{N} x_{j,i} \\ \dot{x}_{j,i} &= -cbx_{j+1,i}x_{j+2,i} + cbx_{j-1,i}x_{j+1,i} - cx_{j,i} + \frac{c}{b}X_i, \end{aligned} \tag{8}$$

where the $X_i$ variables are considered large-scale variables, and the $x_i$ variables are considered small-scale. Wilks (2005) has investigated how well the dynamics

of the large-scale variables are represented by the parameterized equations

$$\dot{X}_i = -X_{i-2}X_{i-1} + X_{i-1}X_{i+1} - X_i + F + P_i$$
$$P_i = b_0 + b_1 X_i + b_2 X_i^2 + b_3 X_i^3 + b_4 X_i^4 + e_i, \quad (9)$$

where the $b_i$'s are deterministic coefficients, and

$$e_i(t) = \phi e_i(t - \Delta) + \sigma_e(1 - \phi^2)^{1/2} z_i(t) \quad (10)$$

is a simple first-order autoregressive in which the parameter $\phi$ is equal to the lag-1 (i.e., one time step of length $\Delta$) autocorrelation of the $e_i$ time series, the $z_i$ variables are independent draws from a probability distribution with zero mean and unit variance, and the resulting $e_i$ have standard deviation $\sigma_e$.

Wilks (2005) shows that the parameterized model Equation 9 with stochastic parameterization is significantly more skillful as a forecast model of Equation 8 than the deterministically parameterized model with $e_i = 0$. Figure 14 shows the skill of ensemble mean forecasts as a function of the autoregressive parameter $\phi$ and the amplitude of the stochastic forcing $\sigma_e$ expressed in terms of its ratio with the climatological mean standard deviation $s_e$ of the tendency of the small scales on the large scales. The minima of the ensemble mean forecast error occur for a locus of $\phi - \sigma_e$ values. As with the multimodel ensembles, Wilks's results indicate that significantly enhanced reliability is achieved with stochastic rather than deterministic parameterizations.

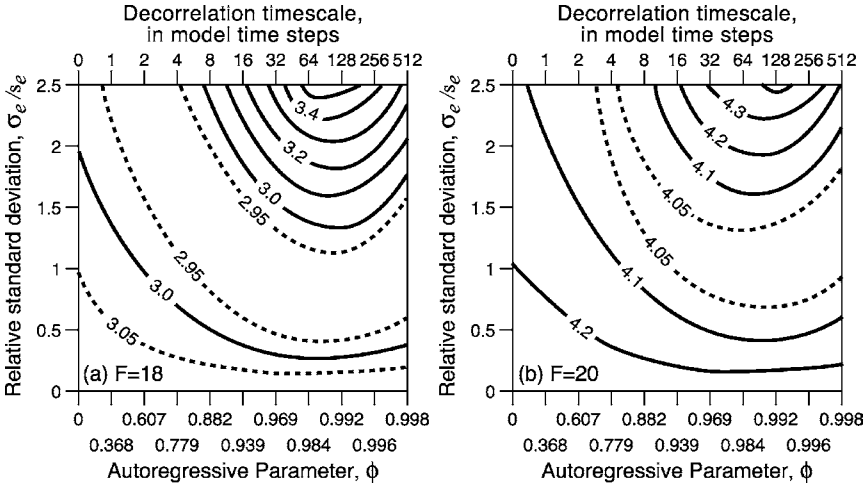



**Figure 14** Ensemble mean skill of the stochastically parameterized Lorenz (1996), validated against the full model, as a function of the amplitude of the noise (x-axis) and its redness (y-axis). Data are from figure 8 in Wilks (2005).

## Barotropic and Quasi-Geostrophic Models

One of the first studies on the use of stochastic noise to represent weather-forecast model error was performed by Pitcher (1977) using a barotropic model. In this paper, Pitcher was studying the evolution of moments of a PDF, based on a set of given dynamical evolution equations (which, following Epstein (1969), Pitcher calls stochastic-dynamic prediction; this should not be confused with the term stochastic-dynamic to refer to the paradigm in Figure 1*b*). Pitcher initially considers the inherent uncertainty in the initial state and growth of this error through what he calls the mechanism of "internal error growth." However, he also noted that a substantial part of real forecast error is attributable to the relative simplicity of the dynamical model, which contributes to what is called external error growth. To represent external error, Pitcher adds random forcing terms to the right-hand side of the barotropic equations of motion, noting that such terms are presumed to represent not specific physical processes, but rather a rational mathematical device for incorporating external error in the stochastic-dynamic prediction system. Two random forcings are tested. In the first, the forcing is proportional to the streamfunction of the barotropic model; in the second, the forcing is purely additive noise. Pitcher (1977) discusses problems of assessing the most appropriate redness for the noise. The (positive) impact of these representations of external error on estimates of the spread of the forecast PDF for three-day forecasts is shown.

In passing, Pitcher notes that in (stochastic-dynamic prediction) studies of this sort, one is faced with the problem of closure, and in his study, third moments are neglected. Even with such truncation, computational cost is significant, and Pitcher acknowledges toward the end of the paper that Monte Carlo methods (now called ensemble methods) may hold the most promise for forecasting uncertainty in weather prediction.

More recently, Williams et al. (2004) discuss a nonlinear quasi-geostrophic model simulating two-layer shear flow of a laboratory annulus experiment. The shear is maintained by differential rotation $\Delta\Omega$ between the base and side walls, both of which rotate with angular speed $\Omega$, and the lid. With conventional deterministic form, there is reasonable agreement between the laboratory and model steady-state zonal wave number regime diagrams in the two-dimensional phase-space defined by $[\Omega, \Delta\Omega]$.

A simple stochastic parameterization of unresolved inertia gravity wave activity is included in the quasi-geostrophic model. This is achieved simply by adding white noise (i.e., uncorrelated in time or space) to the right-hand side of the prognostic equations in each layer. The noise is equal and opposite between layers. Laboratory studies (Williams et al. 2004) indicate that inertia gravity waves, generated by the large-scale modes through spontaneous adjustment, play a key role in transitions between flows of different zonal wave numbers.

This idea was tested by integrating the quasi-geostrophic model with a value of $[\Omega, \Delta\Omega]$, such that the equilibrium baroclinic mode was close to a transition between wave numbers 1 and 2. Initializing the model with the wave number 2 mode, the stochastic parameterization led to a spontaneous transition between

wave number 2 and wave number 1. If the noise was then turned down, the wave number 1 pattern persisted indefinitely. Williams et al. (2004) refer to this as a nonlinear stochastic resonance.

## Large-Eddy Simulation Models

Large-eddy simulation of turbulent flow involves the explicit calculation of the large-scale resolved motions using the spatially filtered Navier-Stokes equations. Studies have shown that the transfer of energy from the resolved scales to the subgridscales is bidirectional and that nonlinear interactions between resolved and subgridscales cause both "forward scatter" as well as "backscatter" (i.e., in the latter, the transfer of energy is from the subgridscales to the resolved scales). The use of stochastic parameterizations to represent this backscatter was first implemented by Leith (1990) in studies of the plane turbulent shear-mixing layer. Mason & Thomson (1992) incorporated a stochastic backscatter similar to that of Leith in a simulation of boundary-layer flow (see also Frederiksen & Davis 1997). The cellular automaton stochastic backscatter (CASB) scheme for NWP and climate models described below makes use of the notion of stochastic backscatter.

## Global Circulation Models

Stochastic parameterizations have begun to be developed and tested for Earth-System models. Consistent with the basic result shown in Figure 10, in which the standard deviation of the constrained PDF of subgrid tendencies scaled approximately with the parameterized subgrid tendencies, Buizza et al. (1999) proposed the rather simple stochastic parameterization:

$$\dot{X} = D + P + \varepsilon P, \tag{11}$$

where $\dot{X}$ denotes some model variable (e.g., temperature or a component of velocity), $D$ denotes the terms from the dynamical core (e.g., the advection and Coriolis terms), $P$ denotes the parameterized tendency associated with subgrid processes (e.g., convective heating or gravity-wave drag), and $\varepsilon$ is a stochastic variable drawn from a uniform distribution in the interval $[-0.5,1.5]$. The random drawings were constant over a time range of 6 h and a spatial domain of $10 \times 10$ latitude/longitude. As with the stochastic forcing in the truncated Lorenz (1963) model above, the choice of spatiotemporal autocorrelations strongly influences the performance of the scheme. Buizza et al. (1999) showed that this scheme had a positive impact on medium-range probability forecast skill scores for precipitation, mainly by increasing ensemble spread for this variable.

In this standard stochastic physics scheme implemented operationally in the ECMWF ensemble prediction scheme, there was no representation of uncertainties directly associated with the numerics of the model, including, for example, the effective diffusivity associated with the semi-Lagrangian scheme or indeed the horizontal diffusion term itself, which is traditionally (in the NWP community) not treated as part of "$P$." A simple way of taking these latter uncertainties into account was through a modification of Equation 11 given by

$$\dot{X} = D + P - \varepsilon D + \varepsilon P. \quad (12)$$

Notice that when $D + P \approx 0$, for example, in the tropics, the impact of the stochastic perturbations in Equation 12 is approximately twice that in Equation 11. Figure 6 shows the impact of the stochastic parameterization in equation 12 on rms error and spread of El Niño prediction. Figure 15*a* (see color insert) shows the precipitation systematic error in a version of the ECMWF coupled model with no stochastic parameterization (i.e., $\varepsilon = 0$); Figure 15*b* shows the impact of the stochastic physics scheme in Equation 12. There is substantial noise-induced drift, and the stochastic parameterization scheme has reduced significantly the model systematic error. By contrast (not shown), the impact of the stochastic representation in Equation 11 on systematic error is much smaller.

In Buizza et al. (1999), the stochastic perturbations are applied to the total tendency (temperature or momentum); consistent with the philosophy of Pitcher (1977), no attempt is made to consider stochastic perturbations associated with individual physical processes. By contrast, Lin & Neelin (2003) propose stochastic parameterizations specifically targeted at deep convection. Two different processes are considered. In the first, the relationship between mass flux and large-scale convective available potential energy is posited to have a stochastic component. In the second, the vertical structure of convective heating is modified by a simple stochastic process, perturbed about a mean structure given by a traditional convective parameterization. The first scheme increases the overall variance of precipitation toward observations with a realistic spatial pattern. The second scheme has smaller impact on precipitation variance but has significant positive impact on large spatial scales and low frequencies (e.g., associated with intraseasonal variability).

Finally, Piani et al. (2005) have studied a generalization of middle-atmosphere gravity wave parameterizations, in which the source term is considered stochastic (justified by the need to represent better the intermittent nature of the convection generation of gravity waves). The most striking effect of the stochastic parameterization is the stabilization of the stratospheric quasi-biennial oscillation in multidecadal simulations.

## STOCHASTIC-DYNAMIC SUBGRID MODELS

So far, the stochastic generalizations of deterministic bulk-formula parameterizations have not been profound—we vary stochastically either the parameters or the tendencies associated with these parameterizations. In this section we take a rather broader approach. Figure 1*a* was a schematic of conventional parameterization. The subgridscale tendencies are determined by gridscale input from the weather or climate model and then fed back to the model again at gridscale. Because the gridscales are the least accurate of all scales in a weather or climate model, the accuracy of this approach has been questioned in the past (Lander & Hoskins 1997), although defenders of the status quo would note that orography is well known right down to the gridscale and that many subgrid processes have a strong orographically forced component. A schematic of an alternative paradigm for subgrid

representation is presented in Figure 1*b*. In it, the subgridscales are represented by a simplified stochastic-dynamic nonlinear dynamical system, coupled with the weather or climate model across a range of scales.

The notion of superparameterization (Grabowski 2001, Randall et al. 2003) can be discussed within this context and involves running a cloud system-resolving model (CSRM) inside a weather or climate model. In its original form, Grabowksi (2001) embedded a two-dimensional (e.g., east-west oriented) CSRM inside each gridbox of a global climate model. There is evidence (Randall et al. 2003) that the performance of climate models in simulating the Madden-Julian Oscillation is significantly improved by superparameterization.



This formulation of superparameterization is a hybrid between the schematics in Figures 1*a* and 1*b*. On the one hand, the conventional bulk-formula parameterization is replaced by a simplified stochastic-dynamic model, a two-dimensional CSRM (where the CSRM initial state can be considered completely stochastically specified). On the other hand, in this original proposal, the coupling is still performed at the gridscale (i.e., similar to that in Figure 1*a*) rather than across a range of scales (i.e., similar to that in Figure 1*b*). Randall et al. (2003) recognize the limitations of this formulation of superparameterization and consider generalizations in which the CSRMs can communicate directly between adjacent gridboxes. However, this increases substantially the computational cost of the scheme, which is already computationally expensive. For example, as discussed by Randall et al. (2003), a climate model with superparameterization is $10^2$ to $10^3$ times more expensive than the same model with conventional parameterization. Going to a three-dimensional superparameterization with communication between gridboxes will increase costs yet further, ultimately approaching the cost of replacing the climate model with a global CSRM (about $10^6$).

Is there a more computationally efficient alternative to superparameterization in which the coupling naturally occurs across a range of scales? One class of stochastic-dynamic models that has gained much prominence in recent years is the cellular automaton (CA), which has been used to represent a bewildering array of physical and biological systems (Wolfram 2002). For example, consider the problem of representing organized convection in a climate model. Consider a two-dimensional grid $(i, j)$, which is much smaller than the climate model grid. A CA $\sigma_{ij}$ takes the value of 1 ("on"), whether convection is active within this CA gridbox, or 0 ("off"), when convection is inhibited there. The key notion that defines a CA is that the rules for evolving $\sigma_{ij}$ are based on (*a*) the value of $\sigma_{ij}$, (*b*) the values of $\sigma_{ij}$ at surrounding gridboxes, and (*c*) the values of climate model variables (considered external forcing). The field of values $\sigma_{ij}$ then defines a forcing back to the climate model grid.

The concept of using a CA field to represent organized convection was proposed by Palmer (1997, 2001) (see Figure 16). In this model, the probability of a CA gridbox remaining "on" is proportional to the number of surrounding "on" cells, and to the convective available potential energy from the climate model gridbox containing the CA gridbox in question. In this way, long-living cells are likely associated with organized agglomerations of "on" cells. The scales on which the

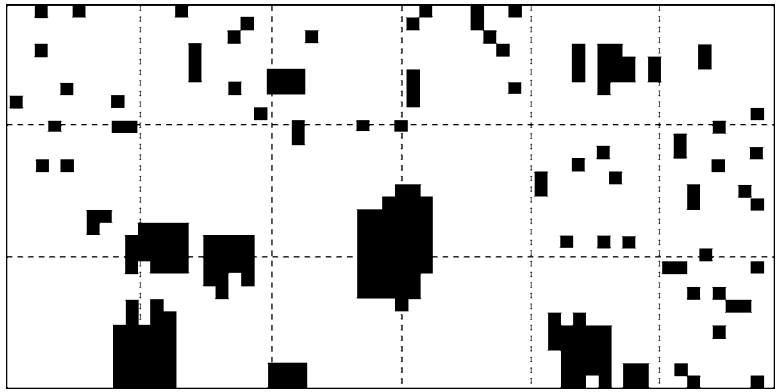


**Figure 16** A snapshot in time from a cellular automaton model for convection. Black squares correspond to convectively active sites, white squares correspond to convectively inhibited sites. From Palmer (2001).



CA feeds back onto the climate model grid are associated with the scale of these agglomerations. For example, for incoherent fields of CA cells, the vorticity forcing back onto the climate model will be small; however, agglomerations of cells might be associated with significant vorticity forcing on scales that are much larger than the climate model grid. In this CA, individual cells can be made to advect with the resolved-scale wind, while larger-scale agglomerations can be made to propagate against the wind with wave-like dynamics. This is something no conventional parameterization would be capable of.

A famous example of a stochastic CA is the Ising model, originally proposed to explain empirically observed facts about ferromagnetism. The Ising model has been used to emulate many diverse systems; the key is that individual elements of the model (e.g., atoms, animals, protein folds, biological membranes, and cumulus clouds) modify their behavior to conform to the behavior of other individuals in their vicinity. In the context of ferromagnetism, the Ising model successfully describes the emergence of macroscopic magnetism when the temperature falls below the Curie temperature.

A prototype coarse-grained stochastic parameterization for the interaction of unresolved tropical convection with the resolved flow, based on the Ising model, has been developed by Majda & Khouider (2001) and Khouider et al. (2003). This idealized model demonstrates how these coarse-grained stochastic parameterizations can significantly impact the climatology of the large-scale flow in the tropics, e.g., the Walker circulation.

A stochastic-dynamic parameterization developed by Shutts (2004) combines the notion of stochastic backscatter with that of the CA. In this scheme, a fixed fraction of the implied energy dissipation from a regular parameterization is "backscattered" onto resolved-scale motions in the model; the patterns onto which the energy is backscattered are determined by the CA rules. In Shutts (2004) a

relatively idealized CA generator is used. More specifically, in the CASB, the streamfunction tendency arising from unresolved or poorly resolved processes is proportional to the square root of the dissipation rate associated with the conventional parameterizations, and the pattern of streamfunction forcing is given by the CA pattern itself. When applied to the T799 ECMWF weather forecast model, Figure 11 shows that CASB can simulate the $k^{-5/3}$ spectrum of Nastrom & Gage (1985), which is totally missing in the standard deterministic version of the model.

Early tests using this scheme show that both model climatology and probabilistic forecast skill are improved. As an example of the former, Figure 17 shows the frequency of occurrence of boreal winter blocking in a version of the ECMWF model, run over six-month timescales, with and without CASB. A stochastic parameterization scheme might increase the frequency of occurrence of otherwise too weakly populated regimes. Figure 17 shows that where the standard model has a significant deficiency in frequency of blocking, particularly over the North Pacific, CASB significantly increases it.



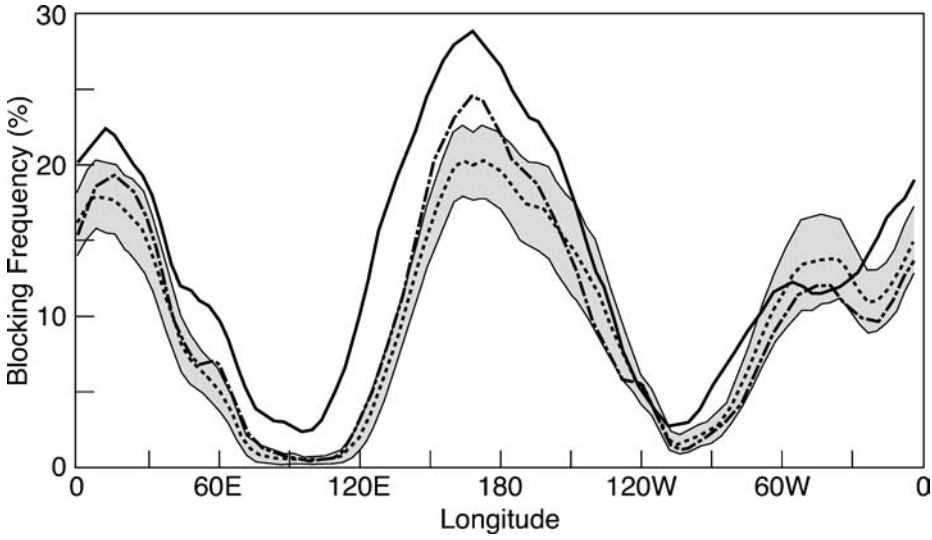



**Figure 17** The solid line represents the climatological frequency of boreal winter blocking based on the definition of Tibaldi & Molteni (1990). The thick solid line represents the observed blocking frequency based on ERA-40 data. The dashed line represents the simulated frequency of blocking based on months 3 through 6 of 40 years of six-month integrations of a standard (cycle 26r3) version of the ECMWF model (T95L60 resolution). The chain-dashed line represents the simulated frequency of blocking based on a similar set of integrations, including the CASB scheme. The shaded area bounded by thin solid lines denotes levels outside which the model with CASB scheme is significantly different from the standard model at the 95% confidence level (e.g., over the North Pacific, where the standard model is most deficient).

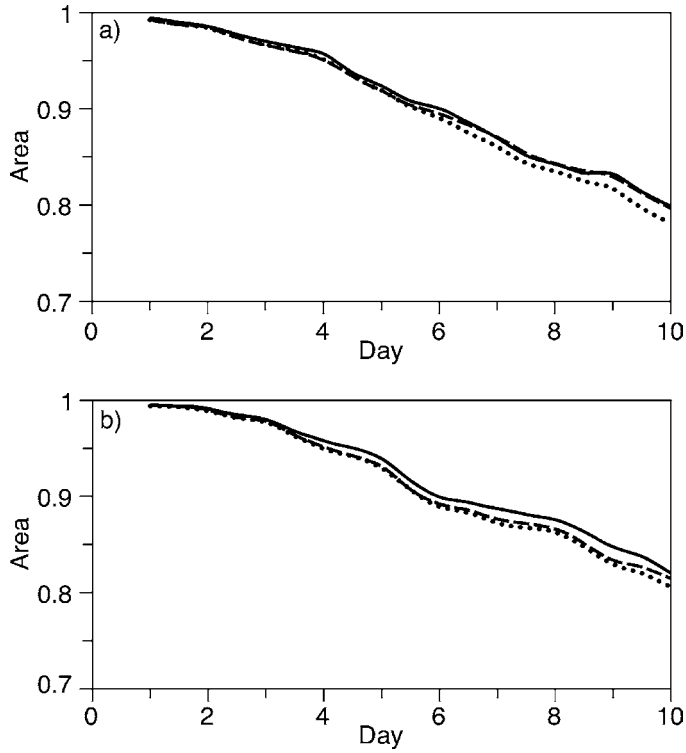



**Figure 18** Mean over 10 medium-range forecasts for boreal winter, of the global ROC area skill score defined for two event thresholds: (*a*) 500 hPa height anomaly is $<-100$ m, (*b*) 500 hPa height anomaly is $>100$ m. Solid line represents CASB; dashed line represents stochastic physics (Equation 11); and dotted line represents forecasts with no estimates of model uncertainty.

As an example of the improvement of probabilistic skill, Figure 18 shows the mean over a set of 10 medium-range forecasts for boreal winter of the skill score known as Area Under the ROC Curve (Jolliffe & Stephenson 2003). Consistent with the improvement in simulation of blocking, the forecast skill for relatively large positive geopotential height events with CASB is improved substantially over ensembles with no representation of model error and ensembles with the simple scheme (Equation 11). For forecasts of relatively large negative height anomalies, the skill of CASB is comparable to that of Equation 11.

## CONCLUSIONS

Chaotic systems are sensitive to initial conditions. For finite-time integrations the degree of sensitivity can be strongly dependent on initial conditions. Ensemble forecast systems are designed to forecast such flow-dependent predictability. This

leads to forecast systems that are more reliable and ultimately more valuable than those based on single, best-guess deterministic forecasts.

However, for weather and climate forecasting, there is a source of uncertainty in addition to that associated with the initial observations: The computational representation of the (partial differential) equations of motion of weather and climate is also uncertain. Some representation of model uncertainty is required if ensemble weather and climate forecasts are to be considered properly reliable.

Weather and climate models are traditionally formulated in terms of explicit projection of the underlying partial differential equations of climate onto some Galerkin basis, with a deterministic bulk-formula parameterization of subgrid processes. As such, a representation of model uncertainty in ensemble prediction systems falls into the hierarchical class of multimodel, multiparameterization, or multiparameter. Multimodel ensembles sample uncertainties in the numerical schemes for discretization and integration, and in the bulk-formula parameterizations. Multiparameterization schemes sample uncertainties in the parameterizations with fixed numerical discretization and integration schemes. Multiparameter schemes sample uncertainty in certain key parameters in fixed parameterizations. In seasonal forecast studies, multimodel schemes give rise to inherently more reliable and skillful probability forecasts of temperature than do those associated with single-model ensembles. For forecasts of extreme precipitation amounts, multimodel ensembles appear less reliable.

With respect to this latter aspect, one cannot expect multimodel ensembles to represent the full spectrum of model uncertainty. The reason for this is simple: The energy spectrum of subgrid motions is generally dominated by scales near the truncation scale and as such cannot be treated by the bulk-formula approach of conventional parameterization. A generalized framework for the formulation of weather and climate prediction models (more generally, Earth-System models) was outlined, whereby specific realizations of subgrid processes were represented by stochastic-dynamic models, coupled with the explicitly resolved part of the model across a range of scales. Uncertainty in such subgrid processes is represented by different realizations of the underlying stochastic generator. Some results from stochastic-dynamic representations of subgridscales were shown. One advantage of stochastic-dynamic parameterization is that the effect of noise-induced drift can lead to a significant reduction in some of the most robust systematic errors, which conventional bulk-formula parameterizations have been unable to reduce substantially over the years. One should also be aware that stochastic parameterization can increase the amount of natural variability within a climate model, impacting detection/attribution studies for climate change.

The only way to assess the relative merits of stochastic-dynamic parameterization against multimodel and other related attempts to represent model uncertainty is through extensive numerical integration. Such an assessment will take place as part of the European Union Sixth Framework Integrated Project ENSEMBLES.

## LITERATURE CITED


Allen MR, Stainforth DA. 2002. Towards objective probabilistic climate forecasting. *Nature* 419:228

Bechtold P, Bazile E, Guichard F, Mascart P, Richard E. 2001. A mass flux convection scheme for regional and global models. *Q. J. R. Meteorol. Soc*. 127:869–86

Bjerknes V. 1904. Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik. *Meteorol. Z.* 21:1–7

Buizza R, Houtekamer PL, Toth Z, Pellerin G, Wei M, Zhu Y. 2004. Assessment of the status of global ensemble prediction. *Mon. Weather Rev.* In press

Buizza R, Miller MJ, Palmer TN. 1999. Stochastic simulation of model uncertainties in the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.* 125:2887–908

Buizza R, Palmer TN. 1995. The singular vector structure of the atmospheric global circulation. *J. Atmos. Sci*. 52:1434–56

Corti S, Molteni F, Palmer TN. 1999. Signature of recent climate change in frequencies of natural atmospheric circulation regimes. *Nature* 398:799–802

Epstein ES. 1969. Stochastic-dynamic prediction. *Tellus* 21:739–59

Frederiksen JS, Davies AG. 1997. Eddy viscosity and stochastic backscatter parameterizations on the sphere for atmospheric circulation models. *J. Atmos. Sci.* 54:2475–92

Grabowski WW. 2001. Coupling cloud processes with the large-scale dynamics using the cloud-resolving convection parameterization. *J. Atmos. Sci.* 58:978–97

Hagedorn R, Doblas-Reyes FJ, Palmer TN. 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting. *Tellus.* In press

Harrison M, Palmer TN, Richardson DS, Buizza R. 1999. Analysis and model dependencies in medium-range ensembles: two transplant case studies. *Q. J. R. Meteorol. Soc*. 125:2487–515

Houtekamer PL, Lefaivre L, Derome J, Richie H, Mitchell HL. 1996. A system simulation approach to ensemble prediction. *Mon. Weather Rev.* 124:1225–42

IPCC. 2001. *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, ed. JT Houghton, Y Ding, DJ Griggs, M Noguer, PJ Van der Linden, et al. Cambridge, UK: Cambridge Univ. Press

Jolliffe IT, Stephenson DB. 2003. *Forecast Verification: A Practitioner's Guide in Atmospheric Science.* Chichester, UK: Wiley. 240 pp.

Jung T. 2005. Systematic errors of the atmospheric circulation in the ECMWF forecasting system. *Q. J. R. Meteorol. Soc.* In press

Khouider B, Majda AJ, Katsoulakis MA. 2003. Coarse-grained stochastic models for tropical convection and climate. *Proc. Natl. Acad. Sci. USA* 100:11941–46

Lander J, Hoskins BJ. 1997. Believable scales and parameterizations in a spectral model. *Mon. Weather Rev*. 125:292–303

Leith C. 1990. Stochastic backscatter in a subgridscale model: plane shear mixing layer. *Phys. Fluids A* 2:297–99

Lin JW-B, Neelin JD. 2003. Toward stochastic deep convective parameterisation in general circulation models. *Geophys. Res. Lett.* 30:11-1–4

Lorenz EN. 1963. Deterministic non-periodic flow. *J. Atmos. Sci.* 42:433–71

Lorenz EN. 1975. Climatic predictability. In *The Physical Basis of Climate and Climate Modelling*, WMO GARP Publ. Ser. No. 16, pp. 132–36. Geneva: World Meteorol. Organ. 265 pp.

Lorenz EN. 1996. Predictability: a problem partly solved. In *Predictability. Proc. 1995 ECMWF Seminar*, pp. 1–18. Reading, UK: ECMWF

Majda AJ, Khouider B. 2002. Stochastic and mesoscopic models for tropical convection. *Proc. Natl. Acad. Sci. USA* 99:1123–28

Mason PJ, Thomson DJ. 1992. Stochastic backscatter in large-eddy simulations of boundary layers. *J. Fluid Mech.* 242:51–78

Meehl GA, Boer GJ, Covey C, Latif M, Stouffer RJ. 2000. The coupled model intercomparison project. *Bull. Am. Meteorol. Soc.* 81:313–18

Molteni F, Palmer TN. 1993. Predictability and finite-time instability of the northern winter circulation. *Q. J. R. Meteorol. Soc.* 119:269–98

Molteni F, Tibaldi S. 1990. Regimes in the wintertime circulation over the northern extratropics. II. Consequences for dynamical predictability. *Q. J. R. Meteorol. Soc.* 116:1263–88

Murphy JM, Sexton DMH, Barnett DN, Jones GS, Webb MJ, et al. 2004. Quantifying uncertainties in climate change using a large ensemble of global climate model predictions. *Nature.* 430:768–72

Nastrom GD, Gage KS. 1985. A climatology of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft. *J. Atmos. Sci*. 42:950–60

Nicolis C. 2004. Dynamics of model error: the role of unresolved scales revisited. *J. Atmos. Sci.* 61:In press

Palmer TN. 1997. On parameterizing scales that are only somewhat smaller than the smallest resolved scales, with application to convection and orography. In *Proc. 1996 ECMWF Workshop on Convection*, pp. 328–37. Reading, UK: ECMWF

Palmer TN. 2001. A nonlinear dynamical perspective on model error: a proposal for nonlocal stochastic-dynamic parameterization in weather and climate prediction models. *Q. J. R. Meteorol. Soc*. 127:279–304

Palmer TN. 2002. The economic value of ensemble forecasts as a tool for risk assessment: from days to decades. *Q. J. R. Meteorol. Soc.* 128:747–74

Palmer TN, Alessandri A, Andersen U, Canteloube P, Davey M, et al. 2004. Development of a European multi-model ensemble system for seasonal to inter-annual prediction. *Bull. Am. Meteorol. Soc.* 85:853–72

Palmer TN, Molteni F, Mureau R, Buizza R, Chapelet P, Tribbia J. 1993. Ensemble prediction. In *Proc. 1992 ECMWF Seminar: Validation of Models over Europe*, pp. 21–66. Reading, UK: ECMWF

Palmer TN, Räisänen J. 2002. Quantifying the risk of extreme seasonal precipitation events in a changing climate. *Nature* 415:512–14

Piani C, Norton WA, Stainforth D. 2005. The equatorial stratospheric response to variations in deterministic and stochastic gravity-wave parameterizations. *J. Geophys. Res.* In press

Pitcher EJ. 1977. Application of stochastic dynamic prediction to real data. *J. Atmos. Sci.* 34:3–21

Randall D, Khairoutdinov M, Arakawa A, Grabowski W. 2003. Breaking the cloud parameterization deadlock. *Bull. Am. Meteorol. Soc*. 84:1547–64

Richardson LF. 1922. *Weather Prediction by Numerical Process*. Cambridge, UK: Cambridge Univ. Press

Selten FM. 1995. *An efficient empirical description of large-scale atmospheric dynamics*. PhD thesis. Vrije Univ. Amsterdam. 169 pp.

Shutts G. 2004. A stochastic kinetic energy backscatter algorithm for use in ensemble prediction systems. *ECMWF Tech. Mem.* 449

Shutts GJ, Gray MEB. 1994. A numerical modeling study of the geostrophic adjustment process following deep convection. *Q. J. R. Meteorol. Soc*. 120:1145–78

Shutts G, Palmer TN. 2004. The use of high-resolution numerical simulations of tropical circulation to calibrate stochastic physics schemes. In *Proc. ECMWF Workshop Intra-Seasonal Variability*, pp. 83–102. Reading, UK: ECMWF

Stainforth DA, Aina T, Christensen C, Collins M, Frame DJ, et al. 2005. Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature* 433:403–6

Stockdale TN, Anderson DLT, Alves JOS, Balmaseda MA. 1998. Global seasonal rainfall forecasts using a coupled ocean-atmosphere model. *Nature* 392:370–73

Tibaldi S, Molteni F. 1990. On the operational predictability of blocking. *Tellus* 42A:343–65

Toth Z, Kalnay E. 1993. Ensemble forecasting at NMC: the generation of perturbations. *Bull. Am. Meteorol. Soc.* 74:2317–30

Wilks DS. 2005. Effects of stochastic parameterizations in the Lorenz '96 system. *Q. J. R. Meteorol. Soc.* In press

Williams PD, Haine TWN, Read PL. 2005. Stochastic resonance in a nonlinear model of a rotating stratified shear flow, with a simple stochastic inertia-gravity wave parameterization. *Nonlinear Process. Geophys.* 11:127–35

Wolfram S. 2002. *A New Kind of Science*. Champaign, IL: Wolfram Media Inc.

**Figure 2** The scientific basis for ensemble forecasting illustrated by the prototypical Lorenz (1963) model of low-order chaos, showing that in a nonlinear system, predictability is flow dependent. (*a*) A forecast with high predictability, (*b*) forecast with moderate predictability, (*c*) forecast with low predictability.

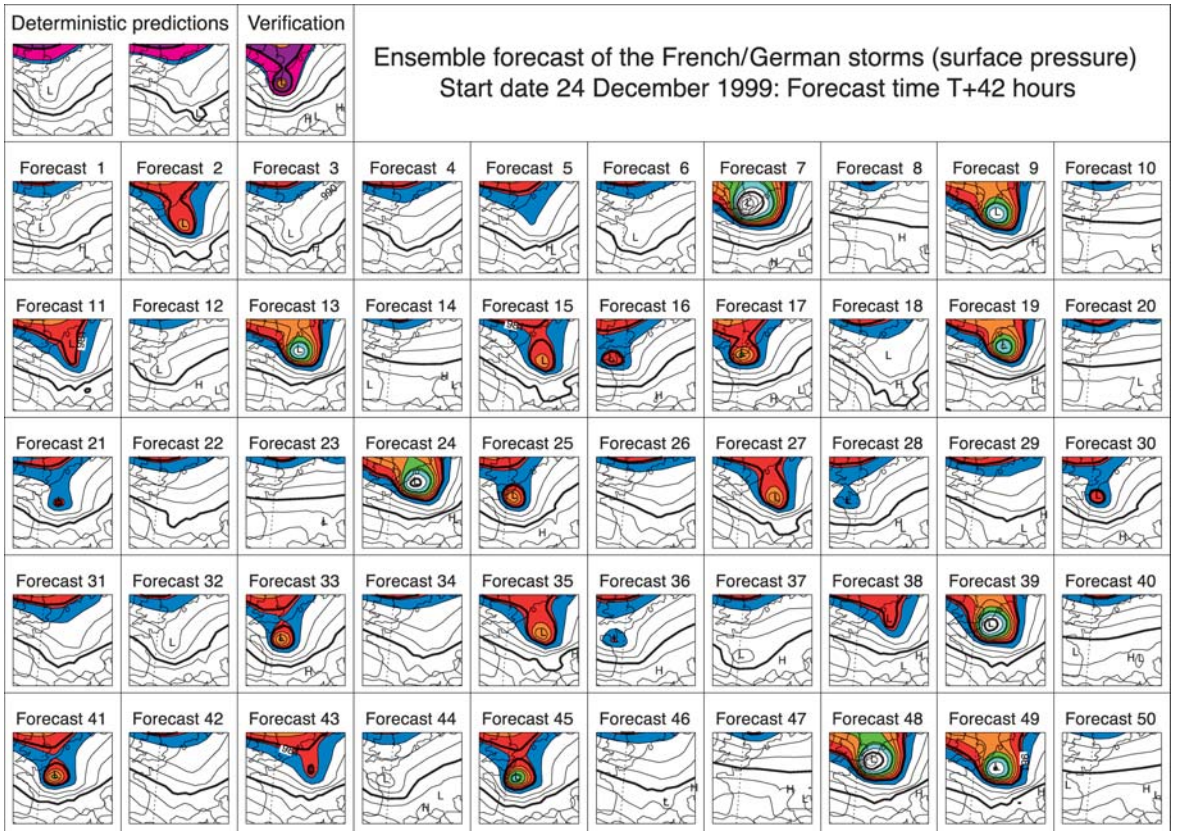
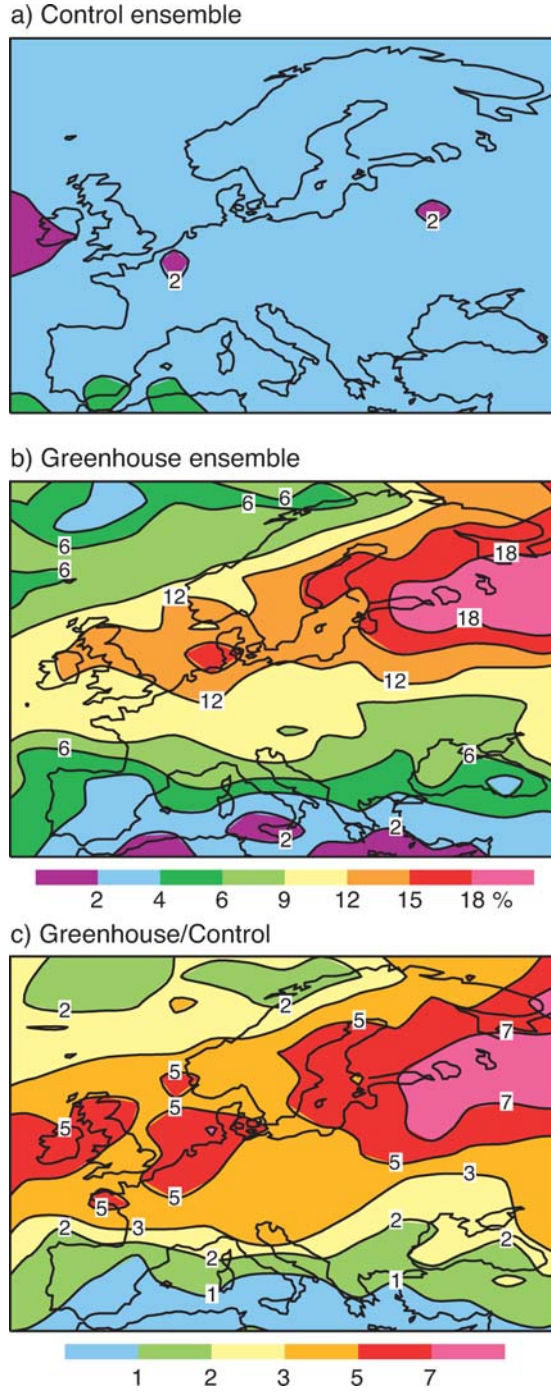



**Figure 3** Forty-two-hour forecast surface pressure "stamp maps" from the ECMWF ensemble prediction system for the severe storm Lothar.




See legend on next page

**Figure 9** The changing probability of extreme seasonal precipitation for Europe in boreal winter. (*a*) The probability (in percentages) of a "very wet" winter defined from the control CMIP2 multimodel ensemble with twentieth-century levels of $CO_2$ and based on the event E: total boreal winter precipitation greater than the mean plus two standard deviations. (*b*) The probability of E but using data from the CMIP multimodel ensemble with transient increase in $CO_2$ and calculated around the time of $CO_2$ doubling (61 to 80 years from present). (*c*) The ratio of values in (*b*) to those in (*a*), giving the change in the risk of a very wet winter arising from human impact on climate. Data are from figure 1 in Palmer & Räisänen (2002).

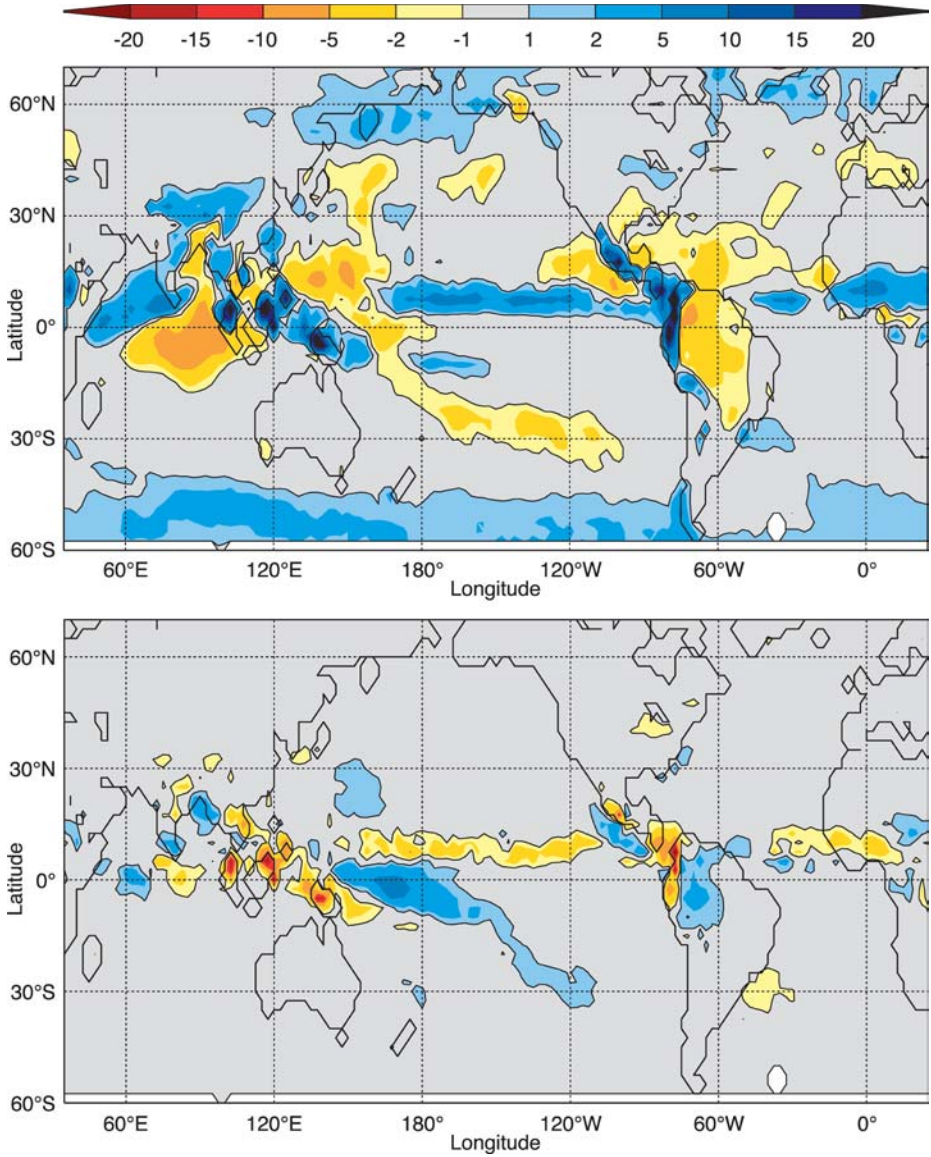



**Figure 15** (*a*) Systematic error for precipitation (mm day$^{-1}$). (*b*) Impact of stochastic physics scheme in Equation 12 on precipitation (mm day$^{-1}$) of ECMWF coupled model. Figures based on month 6 (September) of the T95L40 version of the ECMWF-coupled model (cycle 23r4) from integrations for 1991 through 1998.

*Annual Review of Earth and Planetary Sciences*
*Volume 33, 2005*

# CONTENTS