QUARTERLY JOURNAL
OF THE
ROYAL METEOROLOGICAL SOCIETY

Vol. 127 | JANUARY 2001 Part B | No. 572

Q. J. R. Meteorol. Soc. (2001), **127**, pp. 279–304

A nonlinear dynamical perspective on model error: A proposal for non-local stochastic–dynamic parametrization in weather and climate prediction models*

By T. N. PALMER†

European Centre for Medium-Range Weather Forecasts, UK

(Received 2 February 2000; revised 11 September 2000)

SUMMARY

Conventional parametrization schemes in weather and climate prediction models describe the effects of subgrid-scale processes by deterministic bulk formulae which depend on local resolved-scale variables and a number of adjustable parameters. Despite the unquestionable success of such models for weather and climate prediction, it is impossible to justify the use of such formulae from first principles. Using low-order dynamical-systems models, and elementary results from dynamical-systems and turbulence theory, it is shown that even if unresolved scales only describe a small fraction of the total variance of the system, neglecting their variability can, in some circumstances, lead to gross errors in the climatology of the dominant scales. It is suggested that some of the remaining errors in weather and climate prediction models may have their origin in the neglect of subgrid-scale variability, and that such variability should be parametrized by non-local dynamically based stochastic parametrization schemes. Results from existing schemes are described, and mechanisms which might account for the impact of random parametrization error on planetary-scale motions are discussed. Proposals for the development of non-local stochastic–dynamic parametrization schemes are outlined, based on potential-vorticity diagnosis, singular-vector analysis and a simple stochastic cellular automaton model.

KEYWORDS: Nonlinear dynamics Parametrizations Stochastic models Unresolved scales

For want of a nail the shoe was lost
For want of a shoe the horse was lost
For want of a horse the rider was lost
For want of a rider the battle was lost
For want of a battle the kingdom was lost
And all for the want of a horseshoe nail!
(Anon.)

1. INTRODUCTION

What has been the single most important development in the atmospheric sciences over the last 50 years? There cannot be any disagreement that high on the list, if not at the very top, is the numerical model of the global climate system. Global atmospheric models have transformed the daily weather forecast, and, coupled to global ocean models, are now used routinely to make seasonal predictions and climate change projections. As a result, seven-day forecasts are as skilful today as two-day forecasts were 30-years ago (e.g. Bengtsson 1999); the onset of El Niño and its impact on global weather patterns have been successfully predicted six months in advance (e.g. WMO 1999); and quantitative projections of anthropogenic climate change provide the

* Based on a lecture given at the 1999 European Centre for Medium-Range Weather Forecasts Seminar on model diagnosis.

† Corresponding address: ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, UK.

principal scientific basis for major international protocols on reducing the burning of fossil fuels (e.g. IPCC 1996). The success of global weather and climate models derives not only from the mind-boggling development of computers over the last 50 years, but also from the ingenuity of scientists in devising accurate and efficient computational representations of the equations that govern climate.

Yet, today these models are far from being perfect representations of reality. In the short and medium range, model error is not a negligible source of forecast error (Harrison *et al.* 1998), and the effects of model error must somehow be included in ensemble prediction systems to prevent forecast ensembles becoming systematically under-dispersive in the late medium range. Whilst mean systematic error is quite small in the medium range, it is interesting to note that the pattern of mean systematic error has hardly changed over the last couple of decades (Č. Branković, personal communication). On longer time-scales, systematic model error is a dominant source of forecast inaccuracy. For example, based on data from the European Union project PROVOST*, seasonal-mean systematic error in simulating mid-tropospheric circulation patterns with observed sea surface temperature (SST) is comparable in magnitude with observed interannual variability (Branković and Palmer (2000) and other papers in the Dynamical Seasonal Prediction/PROVOST special issue of the *Quarterly Journal of the Royal Meteorological Society*). An example is illustrated in Fig. 1, which shows such systematic errors to be associated with an erroneous strengthening of the zonal flow and a weakening of planetary-wave activity. Similarly, with interactive oceans the magnitude of the systematic error in tropical east Pacific SST is comparable with the magnitude of typical El Niño SST anomalies (Stockdale *et al.* 1998).

Why are model errors still an important issue in weather and climate prediction? The standard approach to modelling is based on the use of the explicit equations of motion truncated at some prescribed scale, and on the representation of scales below this by a number of deterministic bulk formulae which depend on the resolved flow and some adjustable parameters. Perhaps the right parametrizations have yet to be formulated. Perhaps the right combination of existing parametrizations has not yet been found. Perhaps, with current parameter settings, substantial reduction of systematic error is just around the corner when global models can be run with increased resolution. However, there is another possibility: perhaps the very methodology used to approximate the equations of motion for climate and weather prediction models is itself a source of large-scale systematic error. This possibility is not commonly discussed in the climate modelling community (though see for example: Schertzer and Lovejoy 1993; Mapes 1997; Lander and Hoskins 1997).

The practical difficulty in modelling climate lies in the fact that the governing equations describe a nonlinear coupling of scales of motion that potentially range from tens of thousands of kilometres to the viscous dissipation scale. The role of nonlinearity in generating scale-invariant geometries is vividly illustrated through the generation of fractal sets (such as the Mandelbrot set) by simple nonlinear iteration (e.g. Gulick 1992); such fractal structure has been found in the atmosphere (e.g. Lovejoy 1982; Cahalan *et al.* 1994; Tuck and Hovde 1999). The rhyme at the beginning of the paper captures the flavour of scale invariance as seen in the Mandelbrot set: if losing a kingdom is a tragedy, then losing a nail can also be. Any approximate scale invariance that exists in the real climate is decisively broken in a conventional climate model. Does this matter? Are conventional parametrizations good enough, or are we missing important 'nails' in the formulation of our weather and climate models?

* PRediction Of climate Variations On Seasonal to interannual Time-scales.

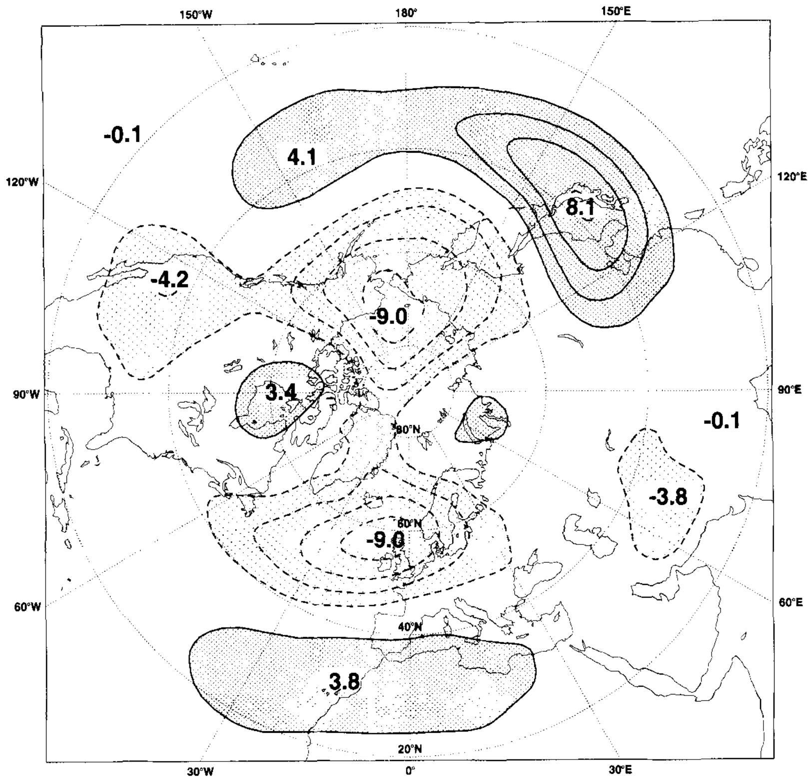

Figure 1. Days 31–120 mean 500 hPa height systematic error during northern winter from a set of 14 winter season integrations of the ECMWF atmosphere model with observed prescribed sea surface temperature, made as part of the PROVOST (see text) project dataset (Branković and Palmer 2000). Contour interval 2 dam.

The purpose of this paper is to make the case that explicit representation of subgrid-scale variability should be considered as part of the parametrization process. More generally, it is proposed that the effects of unresolved scales should be represented by relatively simple stochastic–dynamical systems coupled to the resolved system over a range of scales, rather than by deterministic bulk formulae slaved to the resolved dynamics at precisely the truncation scale. By coupling over a range of scales, such stochastic parametrizations are inherently non-local.

In section 2 we discuss the concept of parametrization in the context of chaotic low-order dynamical systems models whose properties are known exactly. An example is shown (based on the Lorenz (1963) system) of how the neglect of even an energetically unimportant component of a dynamical system's state vector (a nail!) can lead to a major systematic error in the energetically dominant component of the state vector (a kingdom!), and of how a simple stochastic representation of the energetically weak component substantially improves the representation of the dominant component. An approach to the likely existence of an accurate parametrization is discussed from the perspective of a dynamical system, using geometric embedding theorems. With this approach, and using elementary scaling ideas from turbulence theory, it is suggested in section 3 that non-local stochastic parametrization may be needed in climate and weather prediction models, even taking foreseeable increases in model resolution into account. In section 4, simple existing parametrizations for weather and climate models are described, and their impact on weather and climate simulations discussed. Mechanisms based on meteorology describing potential up-scale cascades of error from the subgrid-scale to planetary scales are discussed in section 5. Two sets of mechanisms are discussed: the historical tendency for models to over-populate the more stable high zonal index circulation regimes; and the possible link between organized convection in the warm pool, the Madden–Julian Oscillation (MJO), El Niño, and global-mean temperature. In section 6, a potential-vorticity (PV) perspective on the stochastic representation of unresolved mesoscale organisation associated with convective and orographic systems is put forward. Based on this, two dynamically based possible approaches to non-local stochastic parametrization are outlined in section 7; the first is based on an extension of the singular-velocity methodology used in medium-range ensemble prediction, the second is based on a simple cellular automaton (CA) model.

2. Parametrization and low-dimensional dynamical systems

For the purposes of this paper, a model is a finite representation of a set of partial differential equations which govern the climate system, or the atmosphere in particular, in a form which can be integrated numerically. Let us write the unapproximated equations for climate schematically as:

$$\dot{\widetilde{X}} = \widetilde{F}[\widetilde{X}], \tag{1}$$

where $\widetilde{X}$ is (effectively) infinite dimensional, and $\widetilde{F}$ is some nonlinear functional. A climate or weather prediction model is conventionally constructed by performing some Galerkin decomposition of (1) to produce a set of N deterministic ordinary differential equations:

$$\dot{X} = F[X] + P[X; \alpha], \tag{2}$$

where $F[X]$ represents terms retained in the Galerkin decomposition, and $P[X; \alpha]$ represents some parametrized or bulk representation of the effects of the unrepresented components of $\widetilde{X}$. Conceptually, a parametrization (with parameters α) is based on the notion that there exists a statistical ensemble of unresolved subgrid-scale processes within a grid box x_j, in some secular equilibrium with the grid-box mean flow. For example, to parametrize mixing through the boundary layer, α could be a diffusion coefficient, possibly dependent on the Richardson number of the resolved flow X. If the resolved-scale vertical temperature gradient associated with X is convectively unstable in the vertical column x_j at some grid point j, then over a prescribed time-scale (given by α) P would represent the effect of an ensemble of subgrid-scale convective plumes which operate to relax X back to stability in x_j (Betts and Miller 1986). For flow over unresolved topography, P could represent an ensemble of subgrid-scale orographic gravity waves, propagating vertically, and breaking at some height above the surface, leading to a drag on the resolved flow associated with X, at these heights (e.g. Palmer *et al.* 1986; McFarlane 1987). In these and most other commonly used parametrizations, the bulk representation of small-scale processes within x_j is assumed to be a (horizontally) local deterministic function of X in x_j, i.e.

$$\dot{X}_j = F_j[X] + P[X_j; \alpha], \tag{3}$$

where X_j and F_j represent projection into the subspace associated with x_j.

Despite weather and climate models being formulated in this (local) way, it is generally agreed that the notion of unresolved scales in secular equilibrium with resolved scales is not rigorously justifiable; there is no known gap in spatio–temporal spectra of atmospheric circulations. Without such a gap, the notion of a local parametrization makes little sense. In this respect, it would be tempting to describe as non-local, parametrization schemes P which do not have the local form of (3). However, this would rule out horizontal diffusive parametrizations, which are not really non-local since they merely describe interfacial interactions between adjacent grid boxes. Hence, non-local parametrizations are defined below as those that involve a range of scales, and depend on more than nearest-neighbour interactions.

However, notwithstanding these arguments about locality, it might be argued that by truncating the equations on a sufficiently small scale, the subgrid motions will be so energetically weak compared with the large-scale circulations (in which we are primarily interested) that it should be possible to describe the effect of small scales, to reasonable accuracy, by conventional parametrization methodology. Is this a reasonable argument? Is it sufficient to say that just because a particular scale of motion only explains a small amount of variance, its effect can be represented in a truncated model where that scale is not explicitly represented, by a local bulk formula?

The following example shows that this idea can be manifestly false. Consider the Lorenz (1963) model:

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

The familiar Lorenz attractor is illustrated in Fig. 2(a), using Lorenz's original choice of parameters ($\sigma = 10, b = 8/3, r = 28$). Based on this parameter setting, the governing equations can be written in terms of the three empirical orthogonal functions (EOFs) of the Lorenz model ($\widetilde{a}_1, \widetilde{a}_2, \widetilde{a}_3$), so that (4) is transformed to (Selten 1995):

$$\begin{aligned} \dot{\widetilde{a}}_1 &= 2.3\widetilde{a}_1 - 6.2\widetilde{a}_3 - 0.49\widetilde{a}_1\widetilde{a}_2 - 0.57\widetilde{a}_2\widetilde{a}_3 \\ \dot{\widetilde{a}}_2 &= -62 - 2.7\widetilde{a}_2 + 0.49\widetilde{a}_1^2 - 0.49\widetilde{a}_3^2 + 0.14\widetilde{a}_1\widetilde{a}_3 \\ \dot{\widetilde{a}}_3 &= -0.63\widetilde{a}_1 - 13\widetilde{a}_3 + 0.43\widetilde{a}_1\widetilde{a}_2 + 0.49\widetilde{a}_2\widetilde{a}_3. \end{aligned} \tag{5}$$

The third EOF only explains 4% of the total variance of the system. Hence we might consider that a reasonable approximation to the full model could be obtained by truncating the system to two EOFs, i.e.

$$\begin{aligned} \dot{a}_1 &= 2.3a_1 - 0.49a_1a_2 \\ \dot{a}_2 &= -62 - 2.7a_2 + 0.49a_1^2. \end{aligned} \tag{6}$$

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 (5) and (6) agree well with $\dot{\widetilde{a}}_1$ and $\dot{\widetilde{a}}_2$ (respectively) for points on the Lorenz attractor. However, (for reasons discussed immediately below) the climatology of (6) bears no relation to the climatology of the full model. Instead of exhibiting chaotic variability, from any initial state the truncated model evolves to one of two fixed points (corresponding to the two regime centroids shown in Fig. 2(a)). The climatology of this truncated model has gross systematic errors in both its mean state, and its internal variability.

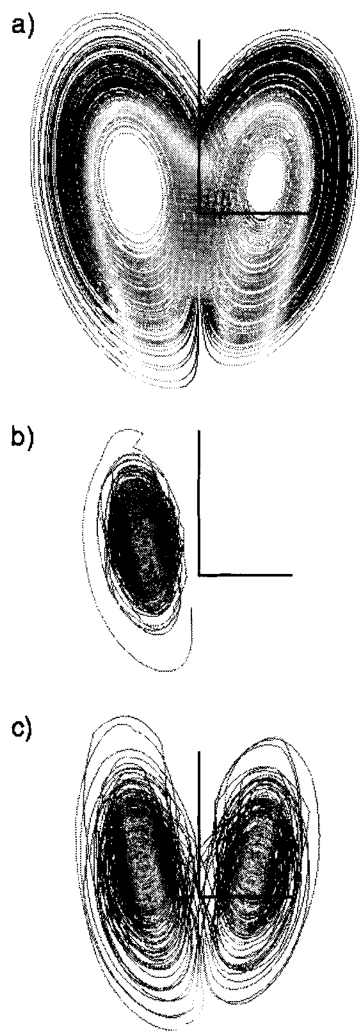

Figure 2. (a) Lorenz attractor, (b) truncated Lorenz equations (see (8) in text) using stochastic forcing with random forcing updated every 0.05 Lorenz time units, (c) truncated Lorenz equations using random forcing updated every 0.1 Lorenz time units (F. Selten, personal communication).

We might naively consider, instead of neglecting the third EOF, parametrizing it in terms of the first two EOFs, i.e.

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

Whilst an astute choice of parametrization, f, might well improve the skill of this model as a short-range forecast model (producing trajectories which shadow the full system for short periods of time), (7) like (6) is climatologically doomed to failure! The reason we can be sure of this is (e.g. Gulick 1992) the Poincaré–Bendixon Theorem: *Consider a two-dimensional (2-D) autonomous system: $\dot{x} = f(x, y)$, $\dot{y} = g(x, y)$, and a trajectory* *which starts at some point p and is confined to some bounded region of phase space; then the trajectory must either: (i) terminate at a fixed point, (ii) return to p, or (iii) approach a limit cycle.*

Because of the Poincaré–Bendixon Theorem, a 2-D autonomous system of ordinary differential equations cannot be chaotic. In the Lorenz (1963) system, the mean state is intrinsically linked with the stability properties of the attractor. The Poincaré–Bendixon theorem provides a counter example to the claim that good short-range forecasting models necessarily make good climate models. (A different argument for the failure of models with severe EOF truncation can be made using generalized stability theory: Farrell and Ioannou 1996, 1999, 2001.)

Let us return to the truncated Lorenz (1963) system. Suppose the dynamical equation for the third EOF is unknown. From the discussion above, it cannot be parametrized as a deterministic function of the dominant EOFs; we somehow need to represent its variability. Consider, therefore, the representation:

$$\begin{aligned} \dot{a}_1 &= 2.3a_1 - 6.2a_3 - 0.49a_1a_2 - 0.57a_2a_3 \\ \dot{a}_2 &= -62 - 2.7a_2 + 0.49a_1^2 - 0.49a_3^2 + 0.14a_1a_3 \\ a_3 &= \beta, \end{aligned} \tag{8}$$

where $\beta(t)$ is a stochastic variable randomly drawn from a Gaussian probability-density function (p.d.f.) whose variance is equal to the explained variance associated with a_3. As such a stochastic model is not autonomous, it is constrained by the Poincaré–Bendixon theorem.

Figure 2(b) and (c) shows the impact of such a stochastic parametrization (F. Selten, personal communication). Figure 2(b) shows results using the parametrized model (represented by (8) where the stochastic forcing is updated every 0.05 nondimensional Lorenz time units). The static vector oscillates irregularly about one of the regime centroids with unimodal p.d.f. Hence, both the mean and the variance in the space S_{EOF} of the dominant EOFs from this stochastic model are still seriously in error (though clearly the error in the variance is not as dire as with no stochastic parametrization). On the other hand, Fig. 2(c) shows the climate of the model when the stochastic variable is updated every 0.1 Lorenz time units. Now the model p.d.f. is clearly bimodal, similar to the exact Lorenz attractor (though the state vector tends to reside in a particular regime too long; see the additional remarks in section 5). Hence, through a simple stochastic representation of the energetically weak third EOF, we have been able to reduce fundamentally the time-mean systematic error in S_{EOF}. This result depends fundamentally on the nonlinearity of the underlying system in S_{EOF}. (In fact, the use of multiplicative noise is not necessary to produce a bimodal p.d.f., the result illustrated in Fig. 2(c) can be qualitatively replicated merely by adding noise terms to the right-hand sides of (6) (F. Selten, personal communication).)

A more general way of looking at this problem is provided by a classic theorem in differential geometry (e.g. Dodson and Poston 1979), the Whitney Embedding Theorem: *Consider an m-dimensional manifold M; then M can be embedded in* $\mathbf{R}^n$ *providing* $n > 2m$.

This theorem has been generalized by Takens (1981) so that M includes chaotic attractors (see also Sauer *et al.* 1991). Note that $n > 2m$ is a sufficient condition for embedding; a necessary condition is that $n \geqslant m$.

For the case of Lorenz (1963) the attractor dimension is greater than two; hence the attractor cannot be embedded in the space spanned by the two dominant EOFs. However, there are dynamical models where these embedding ideas can point to the

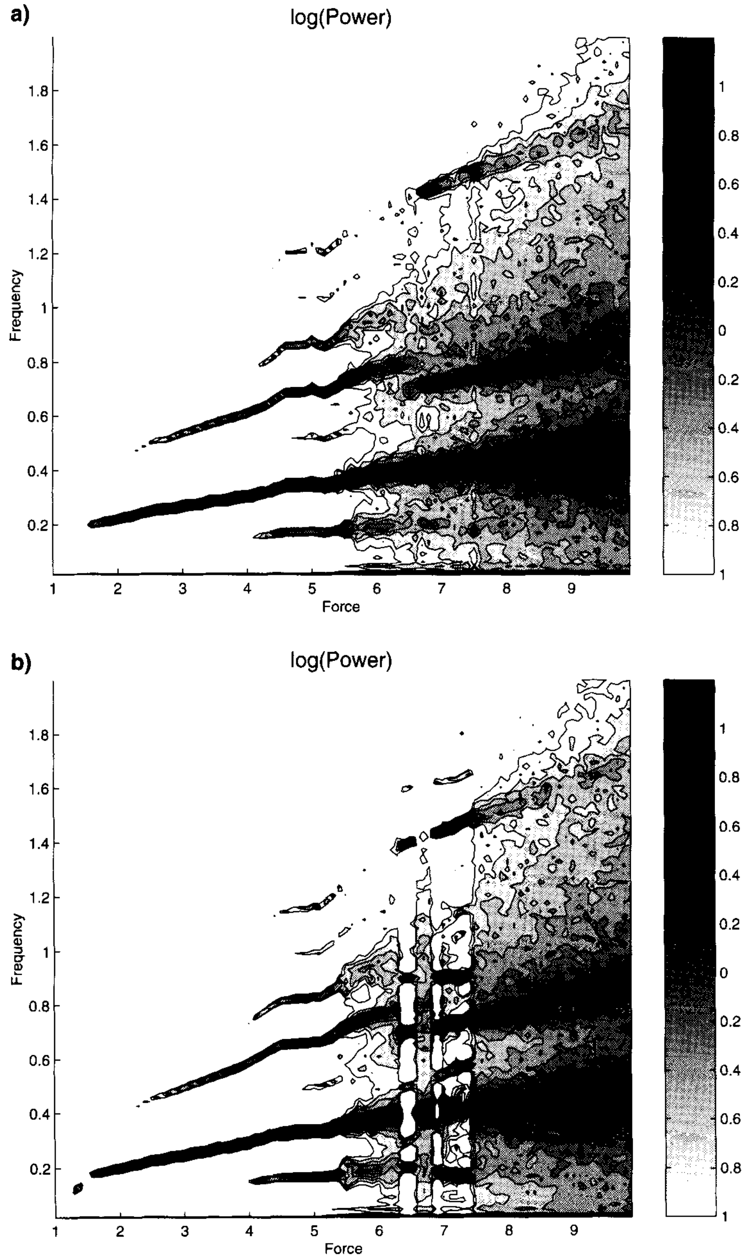

Figure 3. Power spectrum bifurcation diagram for the X_1 variable in Lorenz (1996) system. (a) Exact system (see (9) and (10) in text); (b) parametrized system (see (11) and (12) in text): from Orrell (1999).

likelihood of an accurate parametrization. Consider for example the hierarchical Lorenz (1996) system:

$$\dot{\widetilde{X}}_i = -\widetilde{X}_{i-2}\widetilde{X}_{i-1} + \widetilde{X}_{i-1}\widetilde{x}_{i+1} - \widetilde{X}_i + F - \frac{c}{b}\sum_{j=1}^{N}\widetilde{x}_{j,i} \tag{9}$$

$$\dot{\widetilde{x}}_{j,i} = -cb\widetilde{x}_{j+1}\widetilde{x}_{j+2,i} + cb\widetilde{x}_{j-1,i}\widetilde{x}_{j+1,i} - c\widetilde{x}_{j,i} + \frac{c}{b}\widetilde{X}_i, \tag{10}$$

where the $\widetilde{X}_i$ are large-scale variables, the $\widetilde{x}_{j,i}$ are small-scale variables and b and c are constants. For $N = 8$, the attractor of the Lorenz (1996) system has dimension ~ 4 (Orrell 1999; Smith 2000). Hence, by the Whitney/Takens theorem, an embedding of the attractor into the space of large-scale variables $X_1, X_2, \ldots X_8$, and hence an accurate parametrized model,

$$\begin{aligned} \dot{X}_i &= -X_{i-2}X_{i-1} + X_{i-1}X_{i+1} - X_i + F + P_i \\ P_i &= P_i(X_1, X_2, \ldots, X_8), \end{aligned} \tag{11}$$

would appear to be possible. Of course there is no guarantee that the resulting P_i are in any sense small. In fact, Orrell (1999) has studied the local linear parametrization

$$P_i = \alpha_0 + \alpha_1 X_i, \tag{12}$$

where the parameters α_0 and α_1 are determined by linear regression. Figure 3 shows a spectral bifurcation diagram for the exact system (Fig. 3(a)) and for the linearly parametrized model (Fig. 3(b)). (The spectral bifurcation diagram gives a power spectrum for the X_1 variable, for different values of F.) For small values of F, both the exact system and the parametrized model show most of the power on a set of discrete frequencies; conversely, for large F the power is distributed over a continuum of frequencies for both the exact system and the parametrized model. In general the model compares well with the system. However, there are values of F where the parametrized model fails. For example, near $F = 7$ the model's spectrum is dominated by power on discrete frequencies; in the exact system the power is distributed more uniformly with frequency. Orrell (personal communication) has found parametrizations P_i^{NL} which fit the exact system better than the parametrization in (12). However, these P_i^{NL} are non-local in the sense that the parametrized tendency for X_1, say, is a function not only of X_1, but also of the other large-scale variables $X_2 \ldots$ etc. Note that the Whitney/Takens embedding theorem does not require the parametrizations to be local ($P_i = P_i(X_i)$) in any way. This latter point is relevant when stochastic parametrization in weather and climate models is considered.

3. PARAMETRIZATION AND HIGH-DIMENSIONAL DYNAMICAL SYSTEMS

In the previous section, the concept of parametrization was discussed from the perspective of a dynamical system. If the dimension of a system's attractor is small compared with the size N of a model truncation of the system, then conventional parametrization of unresolved scales may be possible.

Chaotic atmospheric behaviour can be simulated in low-order models (e.g. Ghil and Childress 1987). The attractor dimension, O(10), of such models is therefore much less than the number of resolved variables in a typical weather or climate prediction model. However, with an intermediate-resolution quasi-geostrophic model with O(1000) degrees of freedom, the dimension of the simulated attractor appears to increase to O(100) (Palmer 1996; Reynolds and Errico 1999). One can, therefore, ask the question: how does the dimension of the simulated climate attractor increase as the number of resolved variables increases to values typical of weather and climate prediction models?

As discussed above, the dimension of the attractor cannot be assessed from the number of dominant EOFs of the flow. By the arguments in section 2 above, it may be that 100 EOFs describe 99% of the variance of the flow; however, both the variability and the mean state of the system within the space S_{EOF} of dominant EOFs may depend fundamentally on the remaining 1% of explained variance.

If we consider the whole range of atmospheric motions, then the atmosphere can be viewed as a high-Reynolds-number turbulent fluid. A simple representation of such a high-dimension system is given by the hierarchical Gledzer, Ohkitani and Yamada (GOY) shell system (e.g. Bohr *et al.* 1998):

$$\dot{u}_n + \nu k_n^2 = i k_n \left(u_{n+1}^* u_{n+2}^* - \frac{\delta}{2} u_{n-1}^* u_{n+1}^* - \frac{1-\delta}{4} u_{n-1}^* u_{n-2}^* \right) + f_n, \tag{13}$$

where ν is viscosity and $*$ denotes complex conjugation; $n = 1, 2, \dots, \widehat{N}$, $k_n = 2^n k_0$, and u_n is a complex variable. For $\delta < 1$ the GOY model has both energy and helicity-like invariants. For such a system, the Lyapunov dimension, D, of the attractor increases proportionally with the truncation shell $\widehat{N}$ (Bohr *et al.* 1998). Such a system cannot, therefore, be described as 'low-order' in the limit of high Reynolds number ($\widehat{N} \to \infty$). Hence the discussion above suggests it will be impossible to find an accurate parametrized model of the shell system with truncation scale $N \ll \widehat{N}$. On the other hand, it could be imagined that the systematic effects of parametrization error on the mean and variance of the large-scale ($n \sim 1$) flow might be small enough if $N \gg 1$, even if $N \ll \widehat{N}$. However, the scaling argument below suggests that this cannot be guaranteed.

A fundamental characteristic of error growth in climate and weather prediction is the up-scale transformation from small-scale to large-scale error. This characterizes the 'butterfly effect' paradigm as much as does amplitude growth.

The evolution of a small-amplitude initial perturbation $\delta x(t_0)$ to an initial state X is determined by linearizing (3) about some nonlinear ('basic state') solution. This can be written as:

$$\delta x(t) = M(t, t_0)\delta x(t_0), \tag{14}$$

where M is the tangent propagator (or tangent-linear model) associated with (3). The initial perturbations with largest amplitude at time t are the dominant singular vectors of M (e.g. Farrell and Ioannou 1996). For atmospheric flows the dominant energy-norm singular vectors of M often describe an up-scale transformation from sub-cyclone scales to cyclone and planetary scales (e.g. Molteni and Palmer 1993; Buizza and Palmer 1995). This non-modal behaviour arises because of the non-normality of M.

The 'butterfly effect' can describe not only small-scale error in the initial state, but also model error in representing subgrid-scale activity by bulk formulae. For example, consider (3) with an additional weak imposed forcing $f(t)$, representing random errors in the bulk formula P. The influence of f on the resolved scale over some finite interval $\Delta t = [t_0, t]$ can be written (cf. (14)) as:

$$\delta x(t) = \int_{t_0}^{t} M(t, t') f(t')\,\mathrm{d}t'. \tag{15}$$

If f was constant over Δt, then forcing perturbations with largest impact on the flow at time t would be given by the dominant singular vectors of

$$\mathcal{M}(t, t_0) \equiv \int_{t_0}^{t} M(t, t')\,\mathrm{d}t'. \tag{16}$$

As with initial perturbations, the non-normality of this operator means that small-scale variability in f could efficiently force variability in the large-scale components X of the resolved flow.

To take this argument further, consider the following well-known scaling argument for error propagation in homogeneous isotropic turbulence. Consider a model of such a system with truncation wave number k_N within the inertial subrange, and let f denote model truncation error at k_N. Let $E(k)$ be kinetic energy per unit wave number of the atmosphere at wave number k. Following Lorenz (1969) and Lilly (1973) (see also Vallis (1985)) let us assume that the time it takes error at wave number $2k$ to infect wave number k (i.e. to propagate one 'octave') to be proportional to the 'eddy turn-over time' $\tau(k) = k^{-3/2}\{E(k)\}^{-1/2}$. Then the time $\Omega(k_N)$ taken for uncertainty to propagate N_o octaves from wave number k_N to some large-scale k_L of interest is given by:

$$\Omega(k_N) \equiv \sum_{n=0}^{N_o-1} \tau(2^n k_\mathrm{L}). \tag{17}$$

In the case of 2-D turbulence in the enstrophy-cascading inertial subrange between some large-scale (e.g. baroclinic) forcing scale and dissipation scale, then $E(k) \sim k^{-3}$, τ is independent of k, and $\Omega(k_N) \sim N_o$ which diverges as $k_N \to \infty$. By contrast, if $E(k) \sim k^{-5/3}$ so that $\tau \sim k^{-2/3}$ then $\Omega(k_N)$ tends to a finite limit,

$$\Omega(k_N) \sim \tau(k_\mathrm{L}), \tag{18}$$

as $k_N \to \infty$.

From a physical point of view, this analysis suggests that for $k^{-5/3}$ flow the effect of neglecting unresolved-scale variance on the (mean and variance of the) large-scale flow cannot necessarily be made arbitrarily small by resolving more of the inertial subrange. Hence, whilst up-scale error growth implies that stochastic parametrization may be important in a truncated system with a k^{-3} spectrum, it may be inevitable in truncated systems with partial $k^{-5/3}$ behaviour. (Mathematically, it can be noted that Ω^{-1} is related to the dominant singular value of $\mathcal{M}$, where the norm on the right singular-vector space involves a projection onto wave number k_N, and the norm on the left singular-vector space maximizes growth onto wave number k_L. By contrast, $\tau^{-1}(k_N)$ is associated with the dominant Lyapunov exponent of the system. This demonstrates a fundamental difference between Lyapunov and singular vectors in a multi-scale system.)

Whilst the atmosphere has a k^{-3} spectrum on cyclone scales, there is evidence for a $k^{-5/3}$ spectrum on scales up to a few hundred kilometres (Nastrom and Gage 1985; Gage and Nastrom 1986; Cho *et al.* 1999). Lilly (1983) has suggested that this is at least partly associated, not with 3-D motion, but with up-scale energy transformations forced by organized mesoscale activity (including mesoscale convective complexes, MCCs). More recently, Lilly *et al.* (1998) outline a 'PVs-spreading' mechanism for up-scale transformations, associated with a direct effect of mass outflow from MCCs (see section 6). The truncation-scale of some global climate and weather prediction models is within this range of observed $k^{-5/3}$ activity. The European Centre for Medium-Range Weather Forecasts (ECMWF) model, even with a truncation-scale of tens of kilometres, shows little sign of the shallower $k^{-5/3}$ spectrum (M. Hortal, personal communication). This could be taken as evidence of an inability of global weather and climate prediction models to simulate mesoscale variability. According to the scaling argument above, such a shortcoming could have an impact on simulations of scales k_L within the k^{-3} range.

On the other hand, the Geophysical Fluid Dynamics Laboratory SKYHI model has unambiguous evidence of $k^{-5/3}$ variability near its truncation scale (Koshyk *et al.* 1999). Recent diagnosis (Koshyk and Hamilton 2001) suggests that the shallow mesoscale regime in this model is not associated with an inverse energy cascade, and much of the energy is contained in gravity-wave components near the truncation scale, consistent with VanZandt's (1982) analysis of the observations. As such, it would appear that there is still some ambiguity in the interpretation of the existence of the observed $k^{-5/3}$ spectra, and the mere existence of this shallow spectrum cannot be taken as unambiguous evidence of an up-scale energy cascade. If the behaviour of the SKYHI model is realistic, then misrepresentation of mesoscale organization in weather and climate models may not have a substantial effect on large-scale variability. On the other hand, if the processes suggested by Lilly (1983) and Lilly *et al.* (1998) are relevant, then the results of Koshyk and Hamilton (2000) warn that the injection of unbalanced stochastic noise near the truncation scale of a climate model cannot be guaranteed to efficiently propagate energy up-scale. In section 5 it is proposed to project the stochastic forcing onto 'balanced' PV structures, from which up-scale energy transfer may more readily be excited.

4. STOCHASTIC PARAMETRIZATION IN CLIMATE AND WEATHER PREDICTION MODELS

The use of stochastic noise to represent unpredictable small-scale variability is familiar in a number of geophysical models (e.g. Hasselmann 1976; Farrell and Ioannou 1993; DelSole and Farrell 1995; Penland 1996; Newman *et al.* 1997; Moore and Kleeman 1999). Moreover, following Leith (1990), a stochastic representation of subgrid-scale stress variations has also been used in comprehensive models of 3-D turbulent flow, leading to improvement in the simulation of the resolved flow near rigid surfaces, and to a more accurate logarithmic flow profile in particular (Mason and Thompson 1992).

A version of this stochastic parametrization ('stochastic backscatter') has been applied to the Met Office's global weather prediction model (Frederiksen and Davies 1997; Evans *et al.* 1998). The parametrization was adapted to produce quasi nondivergent horizontal velocity increments with fixed amplitude, at all grid points, and at one model level in the lower troposphere. It was shown that an ensemble of integrations with different realizations of the stochastic parametrization could produce significantly different cyclone- and planetary-scale variability by the late medium range. The effect of this scheme on model systematic error has not yet been quantified.

A somewhat different stochastic formulation ('stochastic physics') was proposed by Buizza *et al.* (1999), linking stochastic forcing to regions in the atmosphere where conventional subgrid parametrization is active, specifically:

$$\dot{X}_j = F_j[X] + \beta P[X_j; \alpha], \tag{19}$$

where β is a stochastic variable drawn from a uniform distribution in [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 spatio–temporal autocorrelations strongly influences the performance of the scheme.

A dramatic effect of this stochastic representation (distinct from the impact of initial singular-vector perturbations) was found in the simulation of isolated atmospheric vortices (Puri *et al.* 2001). An example is given in Fig. 4 which shows sea-level pressure over part of Australia and the west Pacific from four two-day integrations of the ECMWF model. The integrations have identical starting conditions, but different realisations of β. The figure shows two tropical cyclones, the intensity of which can be seen to be very sensitive to the realization of the stochastic parametrization. Hence if the stochastic parametrization in (19) is a fair representation of uncertainty in the

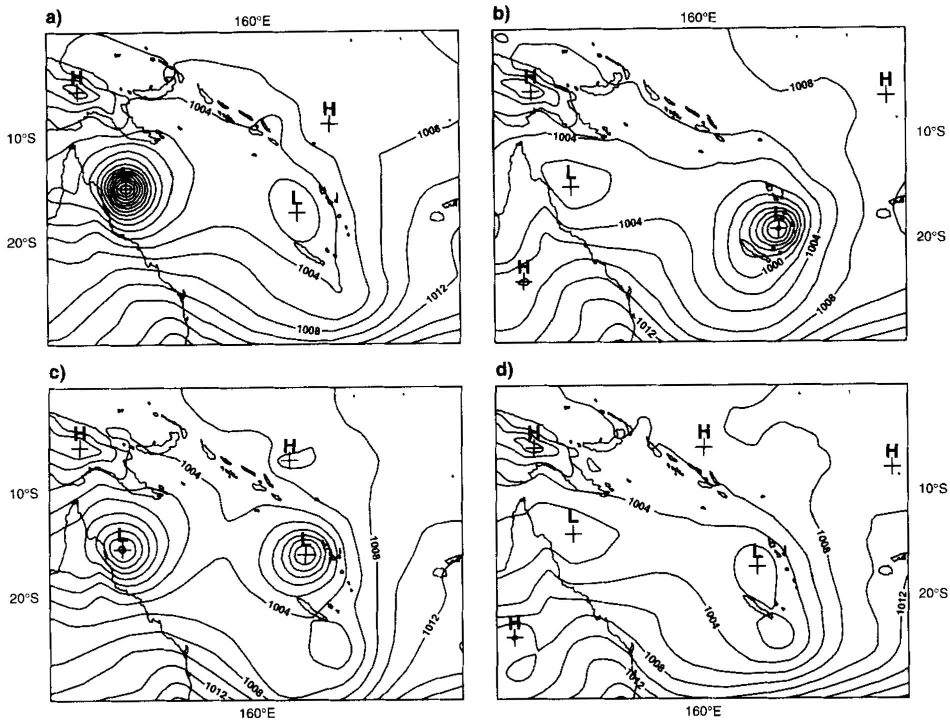

Figure 4. Four two-day integrations of the ECMWF model from identical starting conditions but different realizations of the stochastic parametrization scheme represented by (19) in text with parameter settings as given in Buizza *et al.* (1999). The field shown is sea-level pressure (hPa) over parts of Australia and the west Pacific. The depressions in the pressure field represent potential tropical cyclones. K. Puri, personal communication.

parametrization of diabatic processes, then, especially in the tropics, this leads to significant uncertainty in the resolved-scale flow, even on relatively short time-scales.

As such, the stochastic physics scheme has the potential to impact on synoptic variability in the climatology of the model. This is illustrated in Fig. 5, which shows synoptic time-scale variability in the warm-pool area based on ECMWF analyses (Fig. 5(a)), and two ensemble pairs of six-month ECMWF T63 coupled-model integrations (Stockdale *et al.* 1998), one without stochastic physics (Fig. 5(b)), the other with stochastic physics (Fig. 5(c)). The results show that there is an increase in synoptic time-scale variability using stochastic physics but that, even with stochastic physics, the overall level of variability is too small compared with the analyses. This in turn suggests that the current stochastic physics scheme may still be deficient—see the discussion in section 6.

Stochastic parametrization appears to have a beneficial effect on the skill of probabilistic forecasts of rainfall, as given by the ECMWF ensemble prediction system (Buizza *et al.* 1999). Figure 6 shows a probabilistic measure of forecast skill (the area under the relative operating characteristic curve) for the dichotomous event: 12 h accumulated precipitation is greater than 20 mm, taken over individual grid points in the northern hemisphere (NH). It can be seen that there is a fairly substantial improvement in skill in both summertime and wintertime, when the stochastic physics parametrization is included.

A full assessment of the impact of the stochastic physics parametrization on the systematic error of the ECMWF coupled model is currently in progress. Preliminary results showing the impact of stochastic physics on six-month mean tropical SST is

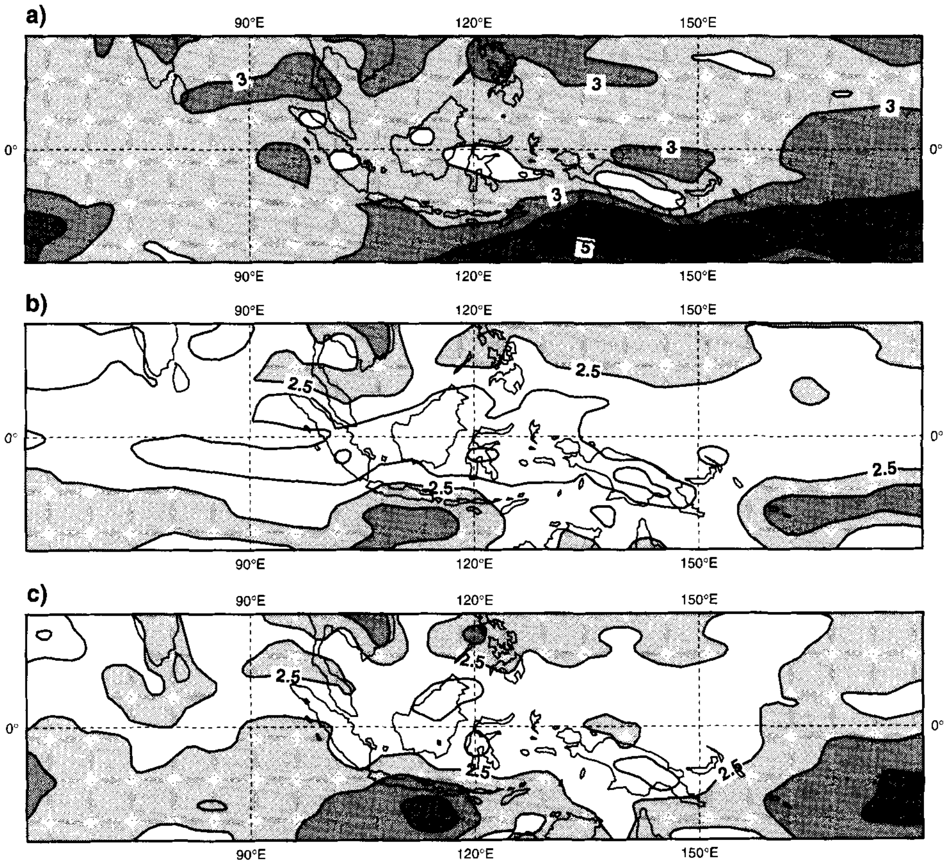

Figure 5. Standard deviation of 850 hPa wind (m s^{-1}), filtered to keep only time-scales less than six days for: (a) ECMWF analyses for December to February 1997/98; (b) December to February of an ensemble of five integrations of the ECMWF coupled model, without stochastic physics, initialized in October 1997; (c) as (b) but with stochastic physics. Contour interval 0.5 m s^{-1}. Shading above 2.5 m s^{-1}.

shown in Fig. 7, based on two ensembles of six-month coupled-model integrations (F. Vitart, personal communication). The shaded region indicates areas where the impact of stochastic physics is statistically significant at the 95% level. Across the western Pacific and western Atlantic, the impact is such as to reduce the systematic error in SST (a general cold bias) found without stochastic physics.

It should be noted that whilst the scheme given in (19) has shown some positive impact on the skill of short- and medium-range forecasts, and on systematic error, it is not energetically consistent with associated surface fluxes of heat and momentum. Indeed, if the surface fluxes were perturbed in such a way as to be consistent with stochastic perturbations to the parametrized tendencies, the impact on SST could be larger than that shown in Fig. 7. The problem of energetic consistency could be addressed straightforwardly if, instead of the parametrization tendency, the parametrization input fields were stochastically perturbed, i.e.

$$\dot{X}_j = F_j[X] + P[X_j + \beta; \alpha]. \tag{20}$$

Experimentation with such a scheme is in preparation.

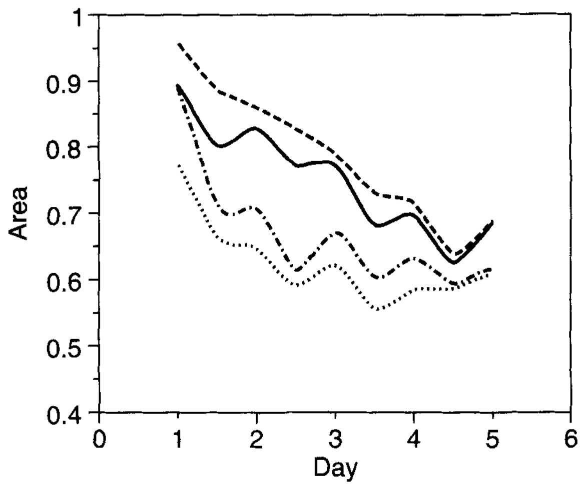

Figure 6. Area under the relative operating characteristic curve for the event: 12 h accumulated precipitation greater than 20 mm. Based on 50-member ensemble integrations of the ECMWF ensemble prediction system: without stochastic physics (solid) and with stochastic physics (dashed) for the period 16 to 22 December 1997; without stochastic physics (dotted) and with stochastic physics (chain-dashed) for the period 29 June to 5 July 1997. From Buizza *et al.* (1999).

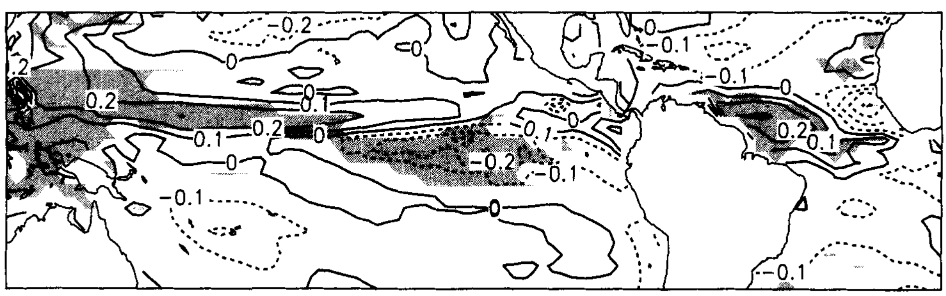

Figure 7. The impact of stochastic physics on the systematic error of sea surface temperature in the ECMWF coupled model, based on 30 pairs of 6-month integrations started one day apart in the spring of 1997; contour interval 0.1 K. Regions where the impact is statistically significant at the 95% level are shaded. F. Vitart, personal communication.

Indeed, there is much scope for development of such schemes, and (19) and (20) are probably simplistic. Suggestions for more dynamically based stochastic parametrizations are outlined in section 7. However, before concluding this section, it should be noted that there is another commonly used technique for representing model uncertainty in ensemble predictions: the multi-model ensemble. This is achieved by incorporating within the ensemble, integrations from a number of quasi-independent weather or climate prediction models. Results suggest that probabilistic skill scores for the multi-model ensemble can exceed the mean skill of individual-model ensembles (Harrison *et al.* 1998; Palmer *et al.* 2000). A variant on the multi-model technique has been proposed by Houtekamer *et al.* (1996) where ensemble members are integrated within a common numerical framework, but with different parametrizations, P, or different values of the parameters α. This is similar to the multi-model technique insofar that the parametrizations and parameters are held fixed within a particular integration. These

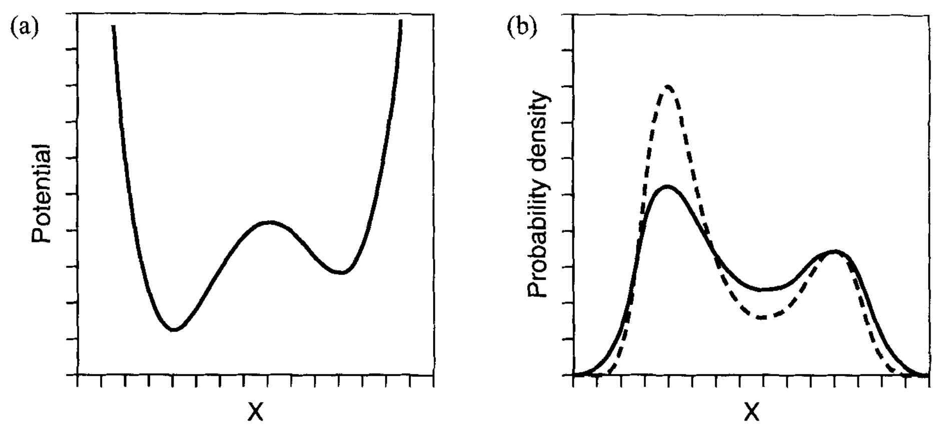

Figure 8. (a) Potential-well function, and (b) probability-density functions obtained using (a) and given white-noise stochastic forcing, based on a solution of the Fokker–Planck equation. The white noise for the dashed-line solution is half that for the solid-line solution. Relative to the solid-line solution, the mean state for the dashed-line solution is biased towards the more populated regime. From Molteni and Tibaldi (1990).

techniques are conceptually distinct from the type of stochastic physics scheme described schematically in (19) and (20). In multi-model ensembles, the effective model perturbations account for the fact that, even in the circumstance where secular equilibrium might be a reasonable assumption, the expectation value of the p.d.f. of subgrid-scale processes is not itself well known. By contrast, the schemes described in (19) and (20) are attempts to account for the fact that in circumstances of mesoscale organisation, the p.d.f. of subgrid-scale processes, even if it was well known, would not be sharp around the mean.

5. Possible scenarios for an up-scale cascade of model error

In studying the truncated Lorenz (1963) models in section 2, it was shown that the neglect of variability of energetically weak components of a dynamical system can have a systematic impact on the mean state in the space S_{EOF} of dominant EOFs. Is there any evidence that this argument applies to climate dynamics? It was noted in section 4 that, despite a positive impact on model performance, the schemes developed so far may be simplistic and somewhat energetically inconsistent. Moreover, as discussed in section 6 below, these schemes may not explicitly perturb the most dynamically relevant variables. As such there is a need for further development of such schemes before extensive integrations and detailed diagnostic analyses are performed (see section 7 below). Nevertheless, it is perhaps worth speculating on possible meteorological mechanisms in the real climate system whereby stochastic physics can influence the largest scales.

Consider first the extratropics. As discussed in many papers (see e.g. Ghil and Childress 1987; Molteni *et al.* 1990; Kimoto and Ghil 1993), there is evidence that in the northern extratropics, the p.d.f. of S_{EOF} has non-Gaussian, and possible multi-modal properties. Let us assume that a climate model can correctly simulate the circulation regimes as diagnosed by Corti *et al.* (1999) from operational analyses, but with inadequate small-scale variability to trigger regime transitions. In such a model, as suggested by Molteni and Tibaldi (1990) the more stable circulation regimes will become overly populated. Figure 8 illustrates this effect. Figure 8(a) shows a hypothetical double potential well; Fig. 8(b) shows two bi-modal p.d.f.s computed using a Fokker–Planck equation for white-noise stochastic perturbations evolving in

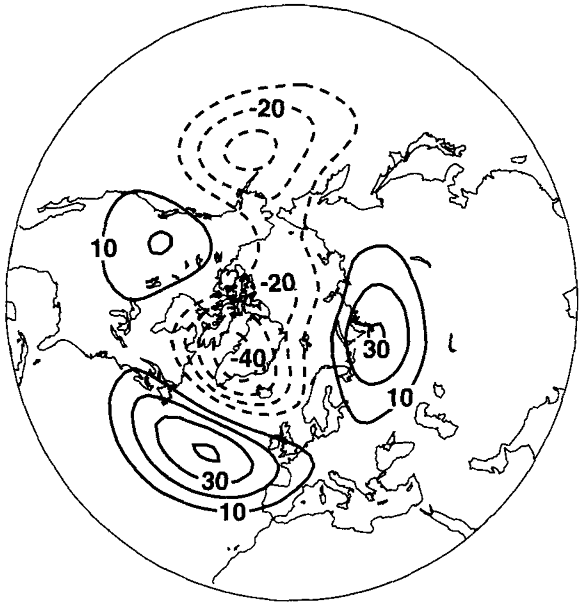

Figure 9. Geographical patterns of the most populated hemispheric circulation regime in the period 1971–94, based on a regime analysis of Corti *et al.* 1999. Based on 500 hPa height. Contour interval 10 m.

the double potential well. The dashed-line results are based on a stochastic variance which has been reduced by a factor of two compared with the solid-line results. With reduced stochastic variance, the mean state shifts towards the more stable and populated regime. According to the analysis in Corti *et al.* (1999), the most populated regime in the real atmosphere in the last couple of decades (Fig. 9) corresponds to a relatively strong zonal flow with weak planetary waves (and is equivalent to the 'cold ocean/warm land' (COWL) pattern of Wallace *et al.* (1996)). In the Pacific/North American sector this circulation regime projects onto the positive phase of the Pacific/North American pattern (Wallace and Gutzler 1981) which is known to be a relatively stable pattern (Palmer 1988). In this nonlinear perspective, therefore, the systematic error of a climate model without stochastic physics would tend to be partially correlated with the anomaly field of the dominant cluster. Comparing Figs. 1 and 9 reveals such a partial correlation.

It is also possible that a systematic misrepresentation of mesoscale variability can have a systematic effect on larger-scale climatic variability in the tropics. Wang and Schlesinger (1999) have shown that a climate model's ability to simulate the MJO is sensitive to certain thresholds in the convection scheme, particularly the value of threshold relative humidity below which the convective parametrization scheme will not trigger. Similar experiments confirming this sensitivity have been performed at ECMWF using convective available potential energy (CAPE) to define a threshold for convective triggering (L. Ferranti, D. Gregory, C. Jakob, personal communication). Specifically, with small threshold CAPE the ECMWF model has a rather poor simulation of the MJO. With convection triggering only when CAPE exceeds 600 J kg^{-1}, the simulation of the MJO is much improved.

Diagnosis of these integrations (not illustrated) shows that large-scale rain dominates over convective rain in the tropics with the 600 J kg^{-1} CAPE threshold. Hence, the explicit dynamics is playing a significant role in releasing the convective instability and hence generating kinetic energy when the convection parametrization scheme is suppressed. Indeed, diagnosis of these results shows that excessive divergent kinetic energy is aliased onto the model grid, creating an overly strong Hadley circulation, for example. Conversely, when the threshold for triggering the convective parametrization is low, convective rain dominates over large-scale rain, and the parametrization adjusts the convectively unstable temperature profiles back to neutrality without any explicit production of kinetic energy. It is speculated that if, instead of relying on the CAPE threshold to generate convectively forced balanced kinetic energy, the model was stochastically forced (in particular with stochastic PV-dipole forcing, see section 6 below) especially in the warm-pool region where MCCs are common, a more satisfactory simulation of both the time-mean flow and MJO variability would ensue.

One can readily speculate about further possible up-scale effects. For example, as suggested by Moore and Kleeman (1999), the (initial-time) singular vectors of the El Niño/Southern Oscillation event (ENSO) have a strong projection onto the MJO, suggesting that models with a poor representation of the MJO may also have an excessively weak and/or regular El Niño climatology. Again, because El Niño itself is a nonlinear phenomenon (Münnich *et al.* 1991), an excessively weak or overly periodic ENSO may give rise to a systematic error in mean tropical Pacific SST. Finally, note that anomalously high tropical east Pacific SST has a tendency to enhance the frequency of occurrence of the COWL cluster shown in Fig. 9, and that this is associated with warm extratropical hemispheric-mean surface temperature (e.g. Palmer 1996). The dynamical coupling of this speculative but plausible up-scale error cascade is reminiscent of the poem at the beginning of the paper: the nail represents stochastic parametrization, the shoe is the MCC, the horse is the MJO, the rider is ENSO, the battle is global-mean temperature, the kingdom our credibility!

6. PV AND STOCHASTIC PARAMETRIZATION

The specific representations, as given in (19) and (20) are of a rather restricted form, where only the tendencies associated with conventional parametrizations are perturbed. Is this a physically justified restriction?

Consider for example the MCC. A characteristic of such systems is the balanced mesoscale circulation field which exist on scales much larger than the component cumulonimbi. These circulations can be conveniently described in terms of PV: cold upper-level anticyclone perturbations associated with a region of near-zero PV, and warm mid-level convergent cyclonic perturbations associated with a significant positive PV anomaly (Shutts and Gray 1994). The existence of such PV anomalies is in part a consequence of the generation of balanced kinetic energy associated with horizontal variations in convective heating. On the basis of high-resolution modelling in which 3-D convection is explicitly simulated, Shutts (1997) finds that if the mass convected in an MCC is equal to M_{c}, then the balanced energy generated in the MCC is proportional to $M_{\mathrm{c}}^{5/3}$. Hence, for a given M_{c} being convected in a typical global climate model grid box, the ratio of balanced energy produced by a single MCC, to the balanced energy produced by an ensemble of, say, 100 convecting elements, is of the order of $100^{2/3} \sim 22$. On the basis that conventional parametrization schemes are designed to describe the latter situation (where the convective kinetic energy is implicitly assumed to be dissipated within a grid box), such schemes may misrepresent situations of mesoscale organization, missing a potentially important PV-source of kinetic energy cascading up-scale.

Gray (2001) has attempted to quantify the impact of such MCC PV forcing in a series of forecast experiments with the Met Office weather prediction model. A control forecast was run using the operational weather prediction model, and a second

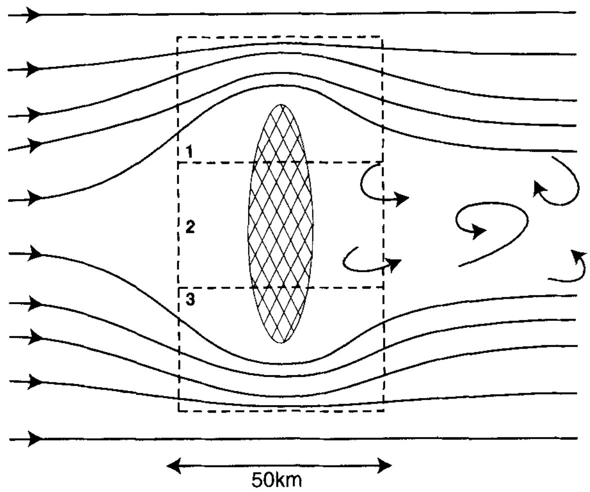

Figure 10. A schematic illustration of flow around topography with mesoscale structure but coherence across model grid boxes (shown dashed). Gravity-wave orographic parametrization would diagnose subgrid-scale orographic variance in boxes 1, 2 and 3, but erroneously apply a drag to the grid-box mean flow in boxes 1 and 3. A non-local and possibly stochastic parametrization based on a horizontal potential-vorticity-dipole forcing might generate a more realistic response.

forecast was run using an initial analysis modified to include PV anomalies associated with MCCs diagnosed from satellite imagery. The PV anomalies were determined by idealized conceptual models and observational studies. Results were either significantly positive in terms of reducing forecast error, or at worst neutral.

Sensitivity studies reported by Gray (2001) suggest that the mid-level positive PV anomaly has a larger impact on forecast evolution than the upper-level negative PV anomaly. This is consistent with the singular-vector analysis of Molteni and Palmer (1993) and Buizza and Palmer (1995) who show that the large-scale extratropical flow is sensitive to perturbations near the baroclinic steering level. Such perturbations can propagate vertically and lead to rapid energy growth in the upper troposphere. Although the dominant singular vectors can have strong baroclinic tilt, Badger and Hoskins (2001) have shown that a monopole low-level PV perturbation, which itself has no such tilt, can nevertheless have sufficient projection onto these rapidly growing structures to give impressive growth characteristics.

Non-local PV forcing may also be relevant to the problem of partially resolved orography in global weather and climate models. Figure 10 shows, schematically, 2-D flow around a poorly resolved obstacle. With inviscid flow, a dipole pair of vortex sheets is created at the fluid boundary on either side of the obstacle. With viscosity, the vorticity perturbations can advect and diffuse outwards in the wake behind the cylinder (specific examples in the case of Greenland are shown in Doyle and Shapiro (1999)). If the obstacle is only partially resolved, as in Fig. 10, then this PV dipole will be misrepresented, both by the explicit dynamics, and by the subgrid-scale orographic parametrizations. For example, orographic gravity-wave drag will apply a negative tendency to the grid-box mean flow in all three grid boxes (since subgrid-scale orographic variance exists in all three grid boxes). However, in reality the flow in grid boxes 1 and 3 is enhanced (compared with an unperturbed flow) as it moves past the obstacle (these are ‘tip jets’, as discussed in Doyle and Shapiro (1999)). Similar to the convective case, the effect of such mesoscale topographic organization might best be represented by some non-local PV dipole forcing, but oriented in the horizontal rather than the vertical. In the case of orography, it is not obvious that such PV dipoles should be represented stochastically; unlike convection, the orographic forcing itself is known precisely. Nevertheless, it is well known that the response to mesoscale orography can depend extremely sensitively on the direction of the large-scale upstream flow. In these circumstances, the strength of the imposed PV dipole might indeed be only representable as a p.d.f.

The downstream spreading of PV anomalies associated with localized orographic forcing and ambient downstream wind shear, is a process which can generate an up-scale transformation of PV and associated wind field. Similar downstream and up-scale spreading effects associated with convectively forced PV have been suggested by Lilly *et al.* (1998) as contributing to the observed $k^{-5/3}$ spectrum as discussed in section 3.

7. Dynamical approaches to stochastic parametrization

From a dynamical systems perspective, one can imagine the scales above and below the truncation scale as represented by two coupled dynamical nonlinear systems, $S_{\leq N}$ and $S_{>N}$. The system $S_{\leq N}$ is given by the finite-N Galerkin representation of (1). Based on the discussion above, it is proposed to represent $S_{>N}$ by a simple dynamically based stochastic system coupled to $S_{\leq N}$ over a range of scales, rather than as a 'lifeless' bulk formula depending on $S_{\leq N}$ only at the truncation scale. The weakness of coupling the parametrized processes to the resolved dynamics at precisely the truncation scale has already been exposed by Lander and Hoskins (1997). Based on the PV perspective outlined above, we consider two complementary approaches which make more explicit use of the underlying dynamics in the formulation of $S_{>N}$. The first follows the approach used at ECMWF to initialize medium-range ensembles, the second utilizes a CA approach to parametrization. Consistent with the comments made in section 2 regarding the Whitney/Takens theorem, these stochastic–dynamic parametrizations are necessarily non-local.

(*a*) *Stochastic forcing in singular-vector space*

As discussed in section 2, a number of geophysical phenomena have been modelled using a background stochastic forcing to excite the singular vectors of a stable but non-normal linearized operator. This raises the possibility that the stochastic forcing could be directly computed from the relevant singular vectors.

It is worth recalling the philosophy used to justify the singular-vector strategy for generating initial perturbations in the ECMWF ensemble prediction system (e.g. Palmer 2000). In principle, an unbiased ensemble of initial states should ideally be created by randomly sampling the initial p.d.f. $\rho_i(X, t=0)$ which determines the probability that the true state lies in some neighbourhood of a point X in phase space. There are two related problems that complicate such a procedure. Firstly, ρ_i is not well known; there are many assumptions in data assimilation (e.g. in quality control, representativity of observations, the tangent approximation, the role of model error in the data assimilation process, and so on) whose contribution to ρ_i is not well quantified. Secondly, for a contemporary weather prediction model, if $\rho_i(X, t=0)$ is Gaussian, then $O(10^{14})$ numbers are needed to specify it, orders of magnitude more than the maximum available sample size. In practice, an inadequately sampled p.d.f. can lead to an underdispersive ensemble and overconfident probability forecasts. For these reasons, the initial ensemble prediction system perturbations are based on the dominant singular vectors of the relevant linearized dynamical operators.

A similar problem arises for the problem of stochastic forcing. A p.d.f. ρ_m can be defined, giving the probability that the actual grid-box tendency due to unresolved scales lies within some neighbourhood of the true tendency. We have argued, for example, that this p.d.f. should be relatively broad in circumstances where mesoscale organization is likely to occur. However, this hardly fixes the actual p.d.f. in any precise sense. Hence, like the initial p.d.f. ρ_i, the subgrid p.d.f. ρ_m is not well-known.

It is, in principle, possible to define a strategy for determining a stochastic sampling of ρ_m in the space of dominant singular vectors of $\mathcal{M}$ (see (16)) optimized over the autocorrelation time associated with mesoscale subgrid-scale processes. A metric can be used to constrain the (right) singular vectors to regions where the parametrized tendencies are large (in the same way that the initial ensemble perturbations are weighted towards regions where observation errors are likely to be large), and to scales close to the truncation scale. The stochastic forcing perturbation can be taken as a random linear combination of these singular vectors. As discussed in Hoskins *et al.* (2000), the PV structure of typical fast-growing singular vectors is peaked in the low to middle troposphere. As such, the components of uncertainty in mesoscale balanced energy which have the potential to cascade up to the synoptic flow, could be efficiently excited by forcing singular vectors.

(*b*) *A cellular automaton model*

An alternative approach to the singular-vector method discussed above, would be to try to model ρ_m more explicitly. One possible type of dynamical system on which to base ρ_m is the CA from which coherent structures (like tesselating hexagons reminiscent of organized Rayleigh–Bénard convection) can readily be produced (see e.g. Adamatsky 1994). The CA model, first applied by von Neumann to biological problems (von Neumann 1966), is a dynamical system with a state vector which takes on a number of discrete (often just two) states. This CA state vector is defined on a discrete grid of points in space and time. There is a rule (either deterministic or stochastic) which determines the state at some space point, as a function of the state of the CA at surrounding points, and at the concurrent and earlier times.

It is possible that a stochastic CA model could form the basis of a simple representation of the MCC; indeed such an approach has already been suggested by Randall and Huffman (1980). Figure 11 is a snapshot of an example of an extremely simple CA representation of mesoscale organization (Palmer 1997). The dashed-line grid is presumed to be equivalent to the grid of a climate or weather prediction model. The CA grid is much finer than the climate model grid, and the CA is presumed to have bivalent states, black representing a convecting (or 'on') state, and white representing a non-convecting state. The CA grid could possibly be initialized from high-resolution satellite imagery, otherwise by some random seed. A rule is needed to determine whether the CA is 'on' or 'off' at the next CA time step, much shorter than the GCM time step. In Fig. 11, a stochastic rule is applied: the probability of the CA being 'on' depends both on the large-scale CAPE (interpolated to a particular CA grid point) and on the number of adjacent CAs which were 'on' at the previous time step. For given CAPE, isolated 'on' cells represent individual cumulonimbi and have relatively short lifetimes, whilst the aggregates of 'on' cells, represent MCCs and have relatively long lifetimes. (In addition to the evolution rules described above, the CA cells can be made to advect across model grid-box boundaries with the resolved-scale wind, and the ability to organize can be made dependent on the resolved-scale wind shear.)

The feedback of the CA onto the large-scale flow could be based on the nonlinear formula of Shutts (1997) discussed in section 6. If we imagine that each 'on' state is

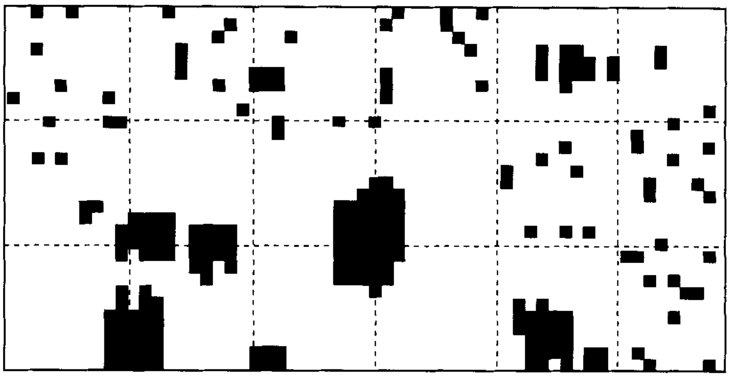

Figure 11. A snapshot of a stochastic cellular automaton (CA) model of organized convection, from Palmer (1997). The dashed-line grid is presumed to be equivalent to the grid of a climate or weather prediction model. The probability of a cell remaining 'on' (shown black) would depend on the large-scale convective available potential energy, and the number of adjacent 'on' cells. The CAs would feed potential vorticity (PV) back to the resolved-scale flow. The strength of a PV-dipole forcing would depend nonlinearly on the number of adjacent 'on' cells.

convecting a mass M, then the 'blobs' with N connected elements represent an organised MCC convecting a mass NM. From the discussion in section 6, this determines a specific PV forcing onto the large-scale flow (proportional to $(NM)^{5/3}$). The total PV-forcing field can be put through PV-inversion software to determine the wind and temperature forcing at each GCM grid point (as was done by Gray (2001)).

Similar ideas could be applied to the orography problem. For example, let h be the mean height of the orography within a CA cell. The probability of a CA element being in an 'on' or blocked state would depend on Nh/U, where N and U are resolved-scale values of static stability and wind speed, interpolated to a CA grid point. A PV-forcing dipole would be centred on the centroid of aggregates of 'on' CAs, representing mesoscale orography. The strength of the PV dipole would depend on the number of connected 'on' cells of an aggregated CA; its width would depend on the dimensions of the aggregate 'on' blob, and the dipole would be oriented relative to the large-scale flow.

8. Discussion

On the rare occasions when meteorologists and economists get together, prediction techniques used in the two disciplines are often compared. At some stage, the meteorologists will raise the fact that whilst the governing equations of climate are known from underlying theoretical principles, the equations which describe the economy are only known from empirical analysis.

Whilst this may well be an excellent debating ploy, there is some dishonesty in making too much of this point. Whilst we know Newton's laws of motion extremely well, there is no unique prescription for representing the governing equations of climate computationally, since the process of parametrization is not a rigorously (or even heuristically) justifiable procedure in regions of mesoscale organisation. Obviously these uncertainties in parametrization in no way invalidate meteorological prediction, after all global weather and climate models over the last 50 years have been one of the most important and fruitful products of our field of research. However, these uncertainties may impede future development, unless they are recognized explicitly.

On the other hand, it could be argued that once models have sufficient resolution, then the effect of such uncertainties on scales of interest will be minimal. However, arguments presented in this paper suggest that under-representation of subgrid variability could have an impact on large-scale systematic error, for any foreseeable resolution. (In any case, climate and weather forecasts must include estimates of uncertainty, and this requires significant utilization of computer time for producing forecast ensembles. This requirement will limit the extent to which very high-resolution climate integrations are possible.)

The real problem is that climate is a complex nonlinear system with many interacting scales. The generic procedure of truncating the equations to some 'hard' limit, and parametrizing physical processes associated with the unresolved scales as local deterministic bulk formulae depending on the resolved scales (at the truncation limit), may well be an underlying factor why some model systematic errors (see Fig. 1) have so stubbornly resisted upgrades in model resolution and parametrization complexity.

The issues, characterized by the words 'deterministic' and 'local', might in some sense be thought of as separate. However, it is argued here that they are manifestations of the same basic problem: that the spectrum of motions has no clear gap, and moreover becomes shallow on scales of hundreds of kilometres and less. The weakness of coupling the parametrized processes to the resolved dynamics at precisely the truncation scale has already been exposed by Lander and Hoskins (1997).

Recognizing the interrelationship of the issues of locality and determinism, a case has been put for the generalization of parametrization schemes from local deterministic systems, slave to the large-scale flow at precisely the truncation scale, to stochastic non-local nonlinear dynamical systems, weakly coupled to the large-scale flow over a range of scales. Examples of such systems, based on CAs, and singular-vector analysis, were outlined in the paper.

The migration away from a purely deterministic approach to modelling, emphasises further that all meteorological prediction problems, from weather forecasting to climate-change projection, are essentially probabilistic.

ACKNOWLEDGEMENTS

I am very grateful to Laura Ferranti, Kamal Puri, Frank Selten and Frederick Vitart for many useful discussions and for supplying me with unpublished material (Figs. 5, 4, 2 and 7 respectively) for this paper; to David Orrell, Cecile Penland and Lenny Smith for discussions on the applications of stochastic and deterministic dynamical systems theory to the parametrization problem; to David Gregory and Christian Jakob for discussions on convective parametrization and the MJO; to Roberto Buizza, Brian Hoskins, Martin Miller, Franco Molteni for discussions on development of stochastic physics schemes, and to Paul Mason, Mike Gray and Glenn Shutts for discussions on stochastic backscatter and PV generation by mesoscale convection. I am also grateful to an anonymous reviewer for many helpful suggestions to improve the first submitted version of this paper.

REFERENCES

Adamatzky, A. 1994 *Identification of cellular automata.* Taylor and Francis, London, UK

Badger, J. and Hoskins, B. J. 2001 Simple initial value problems and mechanisms for baroclinic growth. *J. Atmos. Sci.*, in press

Bengtsson, L. 1999 From short-range barotropic modelling to extended-range global weather prediction: a 40-year perspective. *Tellus*, **51A-B**, 13–32

Betts, A. K. and Miller, M. J. 1986 A new convective adjustment scheme. II: Single column tests using GATE wave, BOMEX, ATEX and arctic air-mass data sets. *Q. J. R. Meteorol. Soc.*, **112**, 693–709

Bohr, T., Jensen, M. H., Paladin, G. and Vulpiani, A. 1998 *Dynamical systems approach to turbulence*. Cambridge nonlinear science series 8. Cambridge University Press, Cambridge, UK

Branković, Č. and Palmer, T. N. 2000 Seasonal skill and predictability of ECMWF PROVOST ensembles *Q. J. R. Meteorol. Soc.*, **126**, 2035–2067

Buizza, R. and Palmer, T. N. 1995 The singular vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, **52,** 1434–1456

Buizza, R, Miller M. J. and Palmer, T. N. 1999 Stochastic simulation of model uncertainties in the ECMWF ensemble prediction system. *Q. J. Meteorol. Soc.*, **125,** 2887–2908

Cahalan, R. F., Ridgway, W., Wiscombe, W. J. and Bell, T. L. 1994 The albedo of fractal stratocumulus clouds. *J. Atmos. Sci.*, **51,** 2434–2455

Cho, J. Y. N., Zhu, Y., Newell, R. E., Anderson, B. E., Barrick, J. D., Gregory, G. L., Sachse, G. W., Carroll, M. A. and Albercook, G. M. 1999 Horizontal wavenumber spectra of winds, temperature and trave gases during the Pacific Exploratory Missions: 1. Climatology. *J. Geophys. Res.*, **104,** 5697–5716

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

DelSole, T. M. and Farrell, B. F. 1995 A stochastically excited linear system as a model for quasi-geostrophic turbulence: analytic results for one- and two-layer fluids. *J. Atmos. Sci.*, **52,** 2531–2547

Dodson, C. T. J. and Poston, T. 1979 *Tensor Geometry*. Pitman, London, UK

Doyle, J. and Shapiro, M. A. 1999 Flow response to large-scale topography: the Greenland tip jet. *Tellus*, **51A,** 728–748

Evans, R. E., Graham, R. J., Harrison, M. S. J. and Shutts, G. J. 1998 'Preliminary investigations into the effect of adding stochastic backscatter to the unified model'. Forecasting Research Technical Report No. 241. The Met Office, London Road, Bracknell, Berkshire, UK

Farrell, B. F. and Ioannou, P. J. 1993 Stochastic dynamics of baroclinic waves. *J. Atmos. Sci.*, **50,** 4044–4057

1996 Generalised stability theory. Part I: Autonomous operators. *J. Atmos. Sci.*, **53,** 2025–2040

1999 Perturbation growth and structure in time dependent flows. *J. Atmos. Sci.*, **56,** 3622–3639

2001 Accurate low-dimensional approximation of the linear dynamics of fluid flow. *J. Atmos. Sci.*, in press

Frederiksen, J. S. and Davies, A. G. 1997 Eddy viscosity and stochastic backscatter parametrisations on the sphere for atmospheric circulation models. *J. Atmos. Sci.*, **54,** 2475–2492

Gage, K. S. and Nastrom G. D. 1986 Theoretical interpretation of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft during GASP. *J. Atmos. Sci.*, **43,** 729–739

Ghil, M. and Childress, S. 1987 *Topics in geophysical fluid dynamics: atmospheric dynamics, dynamo theory and climate dynamics*. Applied Mathematical Sciences, 60. Springer-Verlag, New York, USA

Gray, M. E. B. 2001 The impact of mesoscale convective system potential-vorticity anomalies in numerical weather prediction forecasts. *Q. J. R. Meteorol. Soc.* in press

Gulick, D. 1992 *Encounters with chaos*. McGraw-Hill, New York, USA

Harrison, M., Palmer, T. N., Richardson, D. S. and Buizza, R. 1998 Analysis and model dependencies in medium-range ensembles: Two transplant cast-studies. *Q. J. R. Meteorol. Soc.*, **125,** 2487–2515

Hasselmann, K. 1976 Stochastic climate models. Part I. Theory. *Tellus*, **28,** 473–85

Hoskins, B. J., Buizza, R. and Badger, J. 2000 The nature of singular vector growth and structure. *Q. J. R. Meteorol. Soc.*, **126,** 1565–1580

Houtekamer, P. L., Lefaivre, L., Derome, J., Richie, H. and Mitchell, H. L. 1996 A system simulation approach to ensemble prediction. *Mon. Weather Rev.*, **124,** 1225–1242

IPCC 1996 *Climate Change 1995: The science of climate change*. Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK

Kimoto, M. and Ghil, M. 1993 Multiple flow regimes in the northern hemispheric winter. Part I: Methodology and hemispheric regimes. *J. Atmos. Sci.*, **50**, 2625–2643

Koshyk, J. N. and Hamilton, K. 2001 The horizontal kinetic energy spectrum and spectral budget simulated by a high-resolution troposphere–stratosphere GCM. *J. Atmos. Sci.*, in press

Koshyk, J. N., Hamilton, K. and Mahlman, J. D. 1999 Simulation of the $k^{-5/3}$ mesoscale spectral regime in the GFDL SHYHI general circulation model. *Geophys. Res. Lett.*, **26**, 843–846

Lander, J. and Hoskins, B. J. 1997 Believable scales and parametrisations in a spectral transform model. *Mon. Weather. Rev.*, **125**, 292–303

Leith, C. E. 1990 Stochastic backscatter in a subgrid-scale model: plane shear mixing layer. *Phys. Fluids* **A**, **2**, 297–299

Lilly, D. K. 1973 Dynamical meteorology. Pp. 353–418 in *Lectures in sub-synoptic scales of motions and two-dimensional turbulence*. Ed. P. Morel. Reidel, Boston, USA

1983 Stratified turbulence and the mesoscale variability of the atmosphere. *J. Atmos. Sci.*, **40**, 749–761

Lilly, D. K., Bassett, G., Droegemeier, K. and Bartello, P. 1998 Stratified turbulence in the atmospheric mesoscales. *Theoret. Comput. Fluid Dyn.*, **11**, 139–153

Lorenz, E. N. 1963 Deterministic nonperiodic flow. *J. Atmos. Sci.*, **42**, 433–471

1969 The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307

1996 'Predictability—a problem partly solved.' In 'Predictability', proceedings of the 1995 ECMWF Seminar. Shinfield Park, Reading, UK

Lovejoy, S. 1982 Area–perimeter relation for rain and cloud areas. *Science*, **216**, 185–187

Mapes, B. E. 1997 Equilibrium vs. activation control of large-scale variations of tropical deep convection. Pp. 321–58 in *The physics and parameterization of moist atmospheric convection*. Ed. R. K. Smith. Kluwer Academic Publishers, the Netherlands

Mason P. J. and Thompson, D. J. 1992 Stochastic backscatter in large-eddy simulations of boundary layers. *J. Fluid. Mech.*, **242**, 51–78

McFarlane, N. A. 1987 The effect of orographically excited gravity-wave drag on the general circulation of the lower stratosphere and troposphere. *J. Atmos. Sci.*, **44**, 1775–1800

Molteni, F. and Palmer, T. N. 1993 Predictability and finite-time instability of the northern winter circulation. *Q. J. R. Meteorol. Soc.*, **119**, 269–298

Molteni, F. and Tibaldi, S. 1990 Regimes in the wintertime circulation over northern extratropics. II: Consequences for dynamical predictability. *Q. J. R. Meteorol. Soc.*, **116**, 1263–1288

Molteni, F., Tibaldi, S. and Palmer, T. N. 1990 Regimes in the wintertime circulation over northern extratropics. I: Observational evidence. *Q. J. R. Meteorol. Soc.*, **116**, 31–67

Moore, A. M. and Kleeman, R. 1999 Stochastic forcing of ENSO by the intraseasonal oscillation. *J. Climate*, **12**, 1199–1220

Münnich, M., Cane, M. A., and Zebiak, S. E. 1991 A study of self-excited oscillations of the tropical ocean–atmosphere system. Part II: Nonlinear cases. *J. Atmos. Sci.*, **48**, 1238–1248

Nastrom, G. D. and Gage, K. S. 1985 A climatology of atmospheric wavenumber spectra observed by commercial aircraft. *J. Atmos. Sci.*, **42**, 950–960

Newman, M., Sardeshmukh, P. D. and Penland, C. 1997 Stochastic forcing of the wintertime extratropical flow. *J. Atmos. Sci.*, **54**, 435–455

Orrell, D. 1999 'A shadow of a doubt: model error, uncertainty and shadowing in nonlinear dynamical systems'. Thesis, University of Oxford

Palmer, T. N. 1988 Medium and extended-range predictability, and stability of the PNA mode. *Q. J. R. Meteorol. Soc.*, **114**, 691–713

1996 Predictability of the atmosphere and oceans: from days to decades. NATO Advanced Study Institute Series, Subseries I 'Global environmental change'. Vol 44 *Decadal climate variability: dynamics and predictability*. Eds. D. Anderson and J. Willebrand. Springer, New York, USA

Palmer, T. N. 1997 'On parametrizing scales that are only somewhat smaller than the smallest resolved scales, with application to convection and orography'. In proceedings of the 1996 ECMWF workshop on convection. ECMWF, Shinfield Park, Reading, UK

2000 Predicting uncertainty in forecasts of weather and climate. *Rep. Prog. Phys.*, **62,** 1–47

Palmer, T. N., Shutts, G. J. and Swinbank, R. 1986 Alleviation of a systematic westerly bias in general circulation and numerical weather prediction models through an orographic gravity-wave drag parametrization. *Q. J. R. Meteorol. Soc.*, **112,** 1001–1031

Palmer, T. N., Branković, Č. and Richardson, D. S. 2000 A probability and decision model analysis of PROVOST seasonal multi-model ensemble integrations. *Q. J. R. Meteorol. Soc.*, **126,** 2013–2033

Penland, C. 1996 A stochastic model of IndoPacific sea surface temperature anomalies. *Physica D*, **98,** 534–558

Puri, K., Barkmeijer, J. and Palmer, T. N. 2001 Ensemble prediction of tropical cyclones using targeted singular vectors. *Q. J. R. Meteor. Soc.*, in press

Randall, D. A. and Huffman, G. J. 1980 A stochastic model of cumulus clumping. *J. Atmos. Sci.*, **37,** 2068–2078

Reynolds, C. A. and Errico, R. 1999 Convergence of singular vectors towards Lyapunov vectors. *Mon. Weather. Rev.*, **127,** 2309–2323

Sauer, T., Yorke, J. A. and Casdagli, M. 1991 Embedology. *J. Stat. Phys.*, **65,** 579–616

Schertzer, D. and Lovejoy, S. 1993 'Nonlinear variability in geophysics 3: scaling and multifractal processes'. Lecture Notes. Institut D'Etudes Scientifiques de Cargése

Selten, F. M. 1995 'An efficient empirical description of large-scale atmospheric dynamics'. PhD Thesis, Vrije Universiteit te Amsterdam

Shutts, G. J. 1997 'A balanced flow perspective on the effect of deep convection'. In proceedings of the 1996 ECMWF workshop on convection. ECMWF, Shinfield Park, Reading, UK

Shutts, G. J. and Gray, M. E. B. 1994 A numerical modelling study of the geostrophic adjustment process following deep convection. *Q. J. R. Meteorol. Soc.*, **120,** 1145–1178

Smith L. A. 2000 Disentangling uncertainty and error: on the predictability of nonlinear systems. In *Nonlinear dynamics and statistics.* Ed. A. Mees, Birkhauser, New York, USA

Stockdale, T. N., Anderson, D. L. T., Alves, J. O. S. and Balmaseda, M. A. 1998 Global seasonal rainfall forecasts using a coupled ocean–atmosphere model. *Nature*, **392,** 370–373

Takens, F. 1981 Detecting strange attractors in turbulence. Pp. 336–81 in *Dynamical systems and turbulence: Lecture notes in mathematics.* Vol 898. Eds. D. A. Rand and L.-S. Young. Springer-Verlag, New York, USA

Tuck, A. F. and Hovde, S. J. 1999 Fractal behavior of ozone, wind and temperature in the lower stratosphere. *Geophys. Res. Lett.*, **26,** 1271–1274

Vallis, G. K. 1985 Remarks on the predictability properties of two- and three-dimensional flow. *Q. J. R. Meteorol. Soc.*, **111,** 1039–1047

VanZandt, T. E. 1982 A universal spectrum of buoyancy waves in the atmosphere. *Geophys. Res. Lett.*, **9,** 575–578

von Neumann, J. 1966 *Theory of self-reproducing automata.* University of Illinois Press. Urbana, USA

Wallace, J. M. and Gutzler, D. S. 1981 Teleconnections in the geopotential height field during the northern hemisphere winter. *Mon. Weather Rev.*, **109,** 784–812

Wallace, J. M., Zhang, Y. and Bajuk, L. 1996 Interpretation of interdecadal trends in northern hemisphere surface air temperature. *J. Climate*, **9,** 249–259

Wang, W. and Schlesinger, M. E. 1999 The dependence on convection parametrization of the tropical intraseasonal oscillation simulated by the UIUC 11-layer atmospheric GCM. *J. Climate*, **12,** 1423–1457

WMO 1999 'The 1997–1998 El Niño event: a scientific and technical perspective.' WMO-No.905. World Meteorological Organization. Avenue de la Paix. PO Box 2300. CH-1211 Geneva-2, Switzerland