QUARTERLY JOURNAL
OF THE
ROYAL METEOROLOGICAL SOCIETY

Vol. 131 JANUARY 2005 Part B No. 606

Q. J. R. Meteorol. Soc. (2005), **131**, pp. 389–407 doi: 10.1256/qj.04.03

Effects of stochastic parametrizations in the Lorenz '96 system

By DANIEL S. WILKS†
Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, USA

(Received 12 January 2004; revised 28 June 2004)

SUMMARY

Stochastic parametrization of the effects of unresolved variables is studied in the context of the Lorenz '96 system. These parametrizations are found to produce clear improvements in correspondence between the model and 'true' climatologies; they similarly provide clear improvements in all ensemble forecast verification measures investigated, including accuracy of ensemble means and ensemble probability estimation, and including measures operating on both scalar (each resolved forecast variable evaluated individually) and vector (all forecast variables evaluated simultaneously) predictands. Scalar accuracy measures for non-ensemble (i.e. single integration) forecasts are, however, degraded. The results depend very strongly on both the amplitude (standard deviation) and time-scale of the stochastic forcing, but only weakly on its spatial scale. In general there seems not to be a single clear optimum combination of time-scale and amplitude, but rather there exists a range of combinations producing similar results.

KEYWORDS: Ensemble prediction Probability forecasting Spread–skill relationship

1. INTRODUCTION

The accuracy of numerical weather prediction is limited by a combination of initial-condition uncertainty and model error. The problem of initial-condition uncertainty is currently addressed using ensemble forecasts (e.g. Houtekamer *et al.* 1996; Molteni *et al.* 1996; Toth and Kalnay 1997). The significance of model error to forecast uncertainty is illustrated by studies that show distinct phase-space separation for ensemble forecasts using different models initialized at the same ensemble of initial conditions (Harrison *et al.* 1999). As the physics of large-scale atmospheric motion is basically well known, much of current model error almost certainly derives from the practice of representing the effects of processes occurring at unresolved scales by using comparatively simple deterministic functions of the resolved variables, or 'parametrizations'.

A typical parametrization consists of a function that attempts to summarize the effects of a small-scale process or processes in terms of some larger-scale, resolved, prognostic variable or variables. Ideally, the data underlying these parametrizations would exhibit little scatter. But in general any number of combinations of small-scale (i.e. unresolved) events can combine to yield a given larger-scale result, and this multiplicity essentially guarantees that there will not be a one-to-one correspondence between values of resolved variables and the effects of parametrized processes on the

† Corresponding address: Department of Earth and Atmospheric Sciences, Cornell University, Ithaca NY, 14853 USA. e-mail: dsw5@cornell.edu
© Royal Meteorological Society, 2005.

resolved scales. Consequently there is an inherent indeterminacy in the parametrization exercise, and failure to represent these uncertainties (as is conventional practice) is itself a component of model error. Conventional parametrizations may return something like a mean parametrized quantity, conditional on the relevant large-scale variable or variables, whereas nature supplies a quantity drawn from the point cloud of possibilities scattered around 'the' parametrized value. The resulting model rendition of atmospheric behaviour is thus impoverished relative to the much richer set of behaviours of which nature is capable.

Some movement in the direction of variable parametrizations has occurred experimentally and at the operational forecast centres. Houtekamer *et al.* (1996) and Stensrud *et al.* (2000) describe the use of forecast ensembles in which each member uses a slightly different set of parametrizations. Multi-model (or 'poor-man's') ensembles (Evans *et al.* 2000; Ziehmann 2000; Mylne *et al.* 2002) also include this kind of variation in the treatment of subgrid-scale processes, by construction. In all of these cases, however, a fixed collection of parametrizations is used at all time steps in a particular forecast integration.

The forecast ensembles at the European Centre for Medium-Range Weather Forecasts (ECMWF) have included an explicit stochastic perturbation to the net effect of parametrized physical processes, called 'stochastic physics,' in their ensemble prediction system since October 1998 (Buizza *et al.* 1999). These are applied rather crudely, with diabatic tendencies (i.e. the net effect of multiple parametrizations) scaled multiplicatively by uniform (0.5, 1.5) random numbers, which are applied equally to all points over 5×5 degree horizontal boxes, for 3-hour periods. Adjacent space and time regions are perturbed with independent random forcing. Although these specific parameter choices were arrived at after only limited tuning, the result is that the ECMWF forecasts are improved with respect to ensemble spread and RMS error of the ensemble mean, among other measures of forecast performance (Buizza *et al.* 1999). However, this operational stochastic physics scheme can scarcely be regarded as being optimal, and in the empirical tuning process it was found to be very sensitive to space- and time-scales of the stochastic forcing (Buizza *et al.* 1999).

Ensemble forecasts are often found to be 'under-dispersive', i.e. too frequently not containing the subsequent analysis (e.g. Buizza 1997; Toth and Kalnay 1997; Hamill and Colucci 1997, 1998; Stensrud *et al.* 1999). Some ensemble under-dispersion probably derives from inadequacies in the selection of the ensemble of initial states; but almost certainly some of the problem is also due to model errors, which cause the forecast ensemble to systematically diverge from, and not 'shadow,' the analysis in phase space (e.g. Orrell *et al.* 2001; Smith 2001). One function that the stochastic components of parametrizations may perform, is to allow the forecast ensemble to explore important nearby regions of phase space that more restricted deterministic parametrizations do not reach.

The idea of stochastic parametrization in atmospheric models may have been suggested first in the 1970s (Lorenz 1975; Pitcher 1977). However, even now questions relating to stochastic physics are only beginning to receive intensive study. In addition to the work at ECMWF (Buizza *et al.* 1999), stochastic parametrization of convection has been found to improve the fidelity of climate in general circulation models (Lin and Neelin 2000, 2002). Here the Lorenz '96 system is used as a vehicle to explore the effects of stochastic parametrizations on model climate and forecast performance. Section 2 describes the model systems, section 3 reports consequences for the model climatology, section 4 details the effects on various attributes of the resulting forecasts, and section 5 presents conclusions.

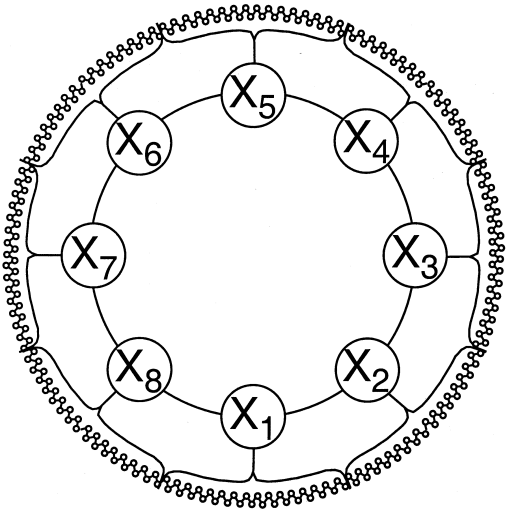

Figure 1. Schematic illustration of the Lorenz '96 system (Eq. (1)) with $K = 8$ resolved variables X_k (large circles), each associated with $J = 32$ unresolved variables Y_j (unlabeled small circles grouped according to the X variable to which they belong), so that there are $JK = 256$ Y variables in total. The forecast model (Eq. (2)) represents explicitly only the X variables, with contributions to each tendency that are due to the unresolved scales being parametrized in terms of the local resolved variable only.

2. MODEL SYSTEMS

The Lorenz '96 system (Lorenz 1996) is given by:

$$\frac{\mathrm{d}X_k}{\mathrm{d}t} = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F - \frac{hc}{b} \sum_{j=J(k-1)+1}^{kJ} Y_j; \quad k = 1, \ldots, K \tag{1a}$$

$$\frac{\mathrm{d}Y_j}{\mathrm{d}t} = -cbY_{j+1}(Y_{j+2} - Y_{j-1}) - cY_j + \frac{hc}{b} X_{\mathrm{int}[(j-1)/J]+1}; \quad j = 1, \ldots, JK. \tag{1b}$$

It is is used here to define 'truth,' i.e. the quantities to be predicted. This system has been used in several previous studies as a metaphor for the atmosphere (Lorenz 1996; Palmer 2001; Smith 2001; Orrell 2002, 2003; Vannitsem and Toth 2002; Roulston and Smith 2003), although with slightly different notation. Equation (1a) describes the linked dynamics of a set of K slow, large-amplitude variables X_k, each of which is associated with J fast, small-amplitude variables Y_j whose dynamics are described by Eq. (1b). Here $K = 8$ and $J = 32$, so that there are $JK = 256$ Y variables in total, as illustrated in Fig. 1. The scaling constants h, c, and b are taken to be 1, 10, and 10, respectively, as is conventional; and F is a forcing taken in the following to be either 18 or 20. The subscripts are cyclic so, for example, $X_0 = X_K$, $X_{-1} = X_{K-1}$, etc. and likewise for the Y variables.

Equations (1) are used to investigate the effects of stochastic parametrizations by regarding the X variables in Eq. (1a) as 'resolved', and the Y variables in Eq. (1b) as 'unresolved'. The tendencies of the resolved variables depend on the local and also nearby values of the resolved variables, and in addition depend on the net effect of the J unresolved variables that 'belong' to the resolved variable X_k in question. For example, in Fig. 1 the leftmost small circle within the bracket connected to X_1 is Y_1, and the rightmost small circle within this bracket is Y_{32}; while the rightmost unresolved variable

belonging to X_8 (and located just to the left of Y_1) is Y_{256}. The unresolved dynamics in Eq. (1b) are analogous to those for the resolved variables in Eq. (1a), although in reverse, and each of the J unresolved variables in a given group is affected equally by the value of the X variable to which it belongs: the notation int$[(j-1)/J]$ in Eq. (1b) indicates integer truncation of the quotient in the square brackets. Equation (1) will be integrated in the following using the fourth order Runge–Kutta formula, with time step $\Delta t = 0.0001$.

A forecast model for the resolved variables in Eq. (1a) is constructed by parametrizing the effects of the unresolved variables in terms of the local values of the resolved variables:

$$\frac{\mathrm{d}X_k^*}{\mathrm{d}t} = -X_{k-1}^*(X_{k-2}^* - X_{k+1}^*) - X_k^* + F - g_U(X_k^*); \quad k = 1, \ldots, K. \tag{2}$$

That is, the 'physical laws' governing the resolved dynamics are assumed to be known exactly, but the effects $(hc/b)\Sigma Y_j$ due to the unresolved variables are represented through parametrizations $g_U(X_k^*)$ that depend only on the value of the resolved variable whose tendency is being calculated, plus additional random forcing. Equation (2) will be integrated in the following using the second order Runge–Kutta formula, with model time step $\Delta t = 0.005$.

The parametrizations $g_U(X_k^*)$ are developed from integrations of Eq. (1) yielding 2000 pairs of points $X(t)$ and $X(t + 0.005)$, with each of the 2000 pairs separated by 5 time units. The unresolved tendencies $U(t)$ to be parametrized are then calculated as the differences between the resolved and actual tendencies at the model time step of $\Delta t = 0.005$:

$$U(t) = [-X_{k-1}(t)\{X_{k-2}(t) - X_{k+1}(t)\} - X_k(t) + F] - \left\{\frac{X_k(t+\Delta t) - X_k(t)}{\Delta t}\right\}. \tag{3}$$

The results, as a function of the resolved variable $X(t)$, are presented in Fig. 2(a) for $F = 18$ and Fig. 2(b) for $F = 20$, and apply equally to each of the X_k because of the symmetry in Eqs. (1a) and (2). The parameter values $J = 32$ and $F = 18$ or 20 are larger than those often chosen for work with the Lorenz '96 system (e.g. Palmer 2001; Smith 2001; Orrell 2002, 2003; Vannitsem and Toth 2002), but are used here because they yield a smaller scatter of points in Fig. 2, so that the conventional non-stochastic parametrizations can be more competitive.

Also shown in Fig. 2 are polynomial regressions of the form:

$$g_U(X_k) = b_0 + b_1 X_k + b_2 X_k^2 + b_3 X_k^3 + b_4 X_k^4 + e_k, \tag{4}$$

with specific parameter values given in Table 1. Clearly the unresolved tendencies depend strongly and nonlinearly on the value of the resolved variable, and the fitted curves represent a large fraction of this variation (R^2 in Table 1). However, for a given value of the resolved variable there is a range of unresolved tendencies that can occur, with specific realizations of the unresolved tendency corresponding to different values for the stochastic component, e_k, in Eq. (4). These plausible ranges are centred at the fitted curves in Fig. 2, with standard deviations (aggregated over the full range of X_k) given as s_e in Table 1. Even though the parametrizations used here are curve fits, rather than analogues of physically based parametrizations, this experimental setting is still highly relevant to the actual practice of atmospheric modelling. The key point is that there is a multiplicity of small-scale states (and therefore also a multiplicity of effects on the large-scale tendencies) consistent with any given large-scale state, so the effects of

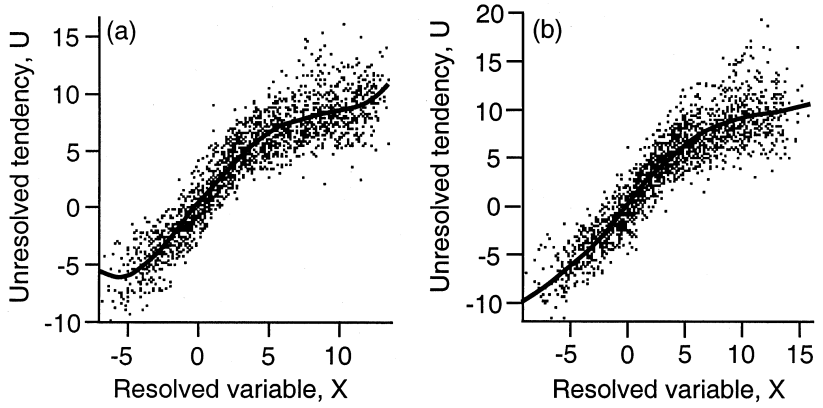

Figure 2. Scatterplots of the unresolved tendency U (Eq. (3)) as a function of the resolved variable X to which it applies, together with the regression functions constituting the deterministic part of the parametrizations $g_U(X)$ (Table 1 and Eq. (4)), for forcings: (a) $F = 18$ and (b) $F = 20$.

TABLE 1. PARAMETRIZATION REGRESSION PARAMETERS AND DIAGNOSTICS

	b_0	b_1	b_2	b_3	b_4	s_e	R^2
$F = 18$	0.275	1.59	−0.0190	−0.0130	0.000707	1.74	87.6%
$F = 20$	0.262	1.45	−0.0121	−0.00713	0.000296	1.99	86.3%

See text for details.

the unresolved processes are inherently uncertain; even if they are not truly random (as is the case here, since Eq. (1b) is fully deterministic), they are effectively random from the perspective of knowing the large-scale variables only.

Use of Eq. (4) as a conventional deterministic parametrization would consist of setting the stochastic portion $e_k = 0$, evaluating the polynomial portion at the current value of the resolved variable $X_k^*(t)$, substituting the resulting $g_U(X_k^*)$ into Eq. (2), and integrating forward in time. Of course this procedure results in an abbreviated representation of the effects of the unresolved variables, and can be enriched by representing the deviations e_k from the fitted curves in Fig. 2 as realizations from a stochastic process; for example:

$$e_k(t) = \phi e_k(t - \Delta) + \sigma_e(1 - \phi^2)^{1/2} z_k(t). \tag{5}$$

Equation (5) is a simple first-order autoregression (e.g. Wilks 1995), in which the autoregressive parameter ϕ is equal to the lag-1 (i.e. one time step of length Δ) autocorrelation of the $e_k(t)$ series; the $z_k(t)$ are independent draws from a probability distribution with zero mean and unit variance, and the resulting e_k have standard deviation σ_e. For $\phi > 0$ the resulting $e_k(t)$ series is a red-noise (positively persistent) process, with 'decorrelation' (i.e. e-folding) time-scale of $-1/\ln(\phi)$ time steps. For $\phi = 0$ the $e_k(t)$ consist of white noise (i.e. are serially independent). For $\sigma_e = 0$ the $e_k(t)$ series is zero for all t, with the result that Eq. (4) reduces to a conventional deterministic parametrization.

Equations (4) and (5) will be used to define parametrizations in the following, which are deliberately imperfect as a consequence of several simplifying assumptions. First, the functional form of Eq. (5) implies that the temporal autocorrelation function for the e_k's decreases exponentially with time lag l, as $r_l = \phi^l$. As indicated in Fig. 3(a), the actual temporal autocorrelation function for the stochastic part of the unresolved

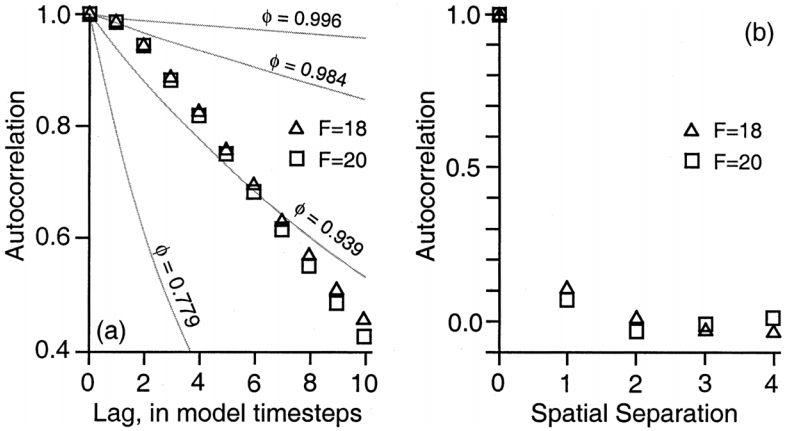

Figure 3. (a) Temporal and (b) spatial autocorrelation functions for the stochastic term in the unresolved tendencies (e_k in Eq. (4)), for forcings $F = 18$ (triangles) and $F = 20$ (squares). Also shown in (a) are autocorrelation functions implied by Eq. (5) for selected values of the parameter ϕ (grey lines).

tendencies is more complicated, and a more elaborate model than Eq. (5) would be required to capture it fully. Second, for the specific results that will be presented, the stochastic forcings, e_k, for each of the $K = 8$ resolved variables will be mutually independent. Figure 3(b) shows the actual 'spatial' autocorrelation functions, which are small but not exactly zero for non-zero spatial separation. Third, the white-noise realizations $z_k(t)$ in Eq. (5) will be implemented as independent standard Gaussian variates, implying that the distributions of the simulated e_k are also Gaussian; whereas the distributions of the actual e_k values are positively skewed, with skewness coefficients of 0.41 and 0.44 for $F = 18$ and 20, respectively. Finally, the standard deviations, σ_e, will be assumed not to depend on the value of the resolved variable, whereas it is evident from Fig. 2 that these increase as X_k increases in absolute value.

3. Stochastic effects on the model climate

The form of the parametrization chosen to represent the unresolved tendencies affects the resemblance of the model climate to the 'true' climate. Table 2 compares means and standard deviations for the true (Eq. (1)) probability distributions of the resolved variables with those produced by three choices for the parametrization of the unresolved tendencies: conventional non-stochastic parametrizations ($\sigma_e = \phi = 0$ in Eq. (5)); serially independent stochastic forcing of the stochastic parametrizations ($\sigma_e = s_e$; $\phi = 0$); and serially persistent stochastic forcing ($\sigma_e = s_e$; $\phi = 0.984$). For both of the stochastic parametrizations the standard deviation σ_e is equal to the value obtained in the fitting of Eq. (4) (Table 1). The choice of ϕ in the latter case corresponds, for both $F = 18$ and $F = 20$, to the autoregressive model that would conventionally be fitted to the first observed autocorrelation of the $e_k(t)$ series (cf. Fig. 3(a)), and produces a decorrelation time-scale of 64 model time steps, or 0.32 time units.

Table 2 shows that all three model versions produce means and standard deviations that are close to the true values. The modelled means are slightly too high for the non-stochastic and independent parametrizations, while the time-dependent stochastic parametrization yields the correct mean for $F = 18$ and an improved mean for $F = 20$. Modelled standard deviations are all too small, and differences among the three parametrizations are slight.

TABLE 2. CLIMATOLOGICAL MEANS AND STANDARD DEVIATIONS FOR 'TRUTH' (EQ. (1)) AND THREE MODEL PARAMETRIZATIONS (EQ. (5))

Forcing		'Truth' (Eq. (1))	Non-stochastic $\sigma_e = \phi = 0$	Independent $\sigma_e = s_e; \phi = 0$	Persistent $\sigma_e = s_e; \phi = 0.984$
$F = 18$	μ_{clim}	3.69	3.76	3.76	3.69
	σ_{clim}	4.55	4.45	4.45	4.45
$F = 20$	μ_{clim}	3.78	3.92	3.92	3.83
	σ_{clim}	5.07	4.92	4.92	4.91

See text for details.

A fuller picture of the accuracy of the model climate is provided by the probability density functions for the resolved variables. Figure 4 shows these for the same cases defined in Table 2. For $F = 18$ (Fig. 4(a)) results for the non-stochastic and independent stochastic parametrizations are quite similar. However, both are clearly inferior to the persistent stochastic parametrization in approximating the true probability density function, particularly with respect to correctly reproducing the secondary maxima near $X = -1$ and $X = 8$. The differences for $F = 20$ in Fig. 2(b) are more subtle, but again the non-stochastic and independent parametrizations yield similar probability densities, while the persistent parametrization produces a probability density that is more similar to the truth. The result that stochastic parametrizations improve the representation of the climate relative to the conventional deterministic alternative, indicates that potential problems related to numerical integration of the stochastic differential equations have not produced computational artifacts (Penland 2003; Ewald *et al.* 2004).

An informative single-number summary of the correspondence between the true probability densities $\psi(x)$ for the resolved variables and those produced by the model integrations is the Kolmogorov–Smirnov statistic:

$$D_n = \max_X |\Psi_{\text{true}}(X) - \Psi_{\text{model}}(X)| = \max_X \left| \int_{-\infty}^{X} \psi_{\text{true}}(x)\, \mathrm{d}x - \int_{-\infty}^{X} \psi_{\text{model}}(x)\, \mathrm{d}x \right|, \tag{6}$$

which is the maximum absolute difference, over the full range of X, between the cumulative probabilities $\Psi(x)$ according to the true and model distributions. Each panel in Fig. 4 also includes the D_n statistics for the three model distributions in each case, confirming the qualitative observations made in the previous paragraph.

More broadly, Fig. 5 shows the dependence of the quality of the model climates in terms of D_n, as a function of the decorrelation time-scale (horizontal) and amplitude (standard deviation, relative to the regression RMSE, s_e, in Table 1) of the stochastic part of the unresolved tendencies. The accuracy of the model climate with non-stochastic parametrization ($\sigma_e/s_e = \phi = 0$) is clearly inferior to those of models with stochastic parametrizations having all but the most extreme combinations of time-scale and amplitude. For both $F = 18$ and $F = 20$ there is a broad area of relatively small D_n values (relatively good representation of the truth climatology); these range from combinations of relatively short decorrelation times paired with relatively large amplitudes, to relatively long decorrelation times paired with smaller amplitudes. For $F = 18$ the minimum D_n is achieved at a point very close to the amplitude specified in Table 1, paired with the time-scale defined by the lag-1 autocorrelation in Fig. 3(a). For $F = 20$ the minimum is achieved at the same time-scale, but paired with a relative amplitude between 1.5 and 2.

Results presented in this section have been confined to the case of (spatially) independent random forcing for each of the $K = 8$ resolved variables; results under

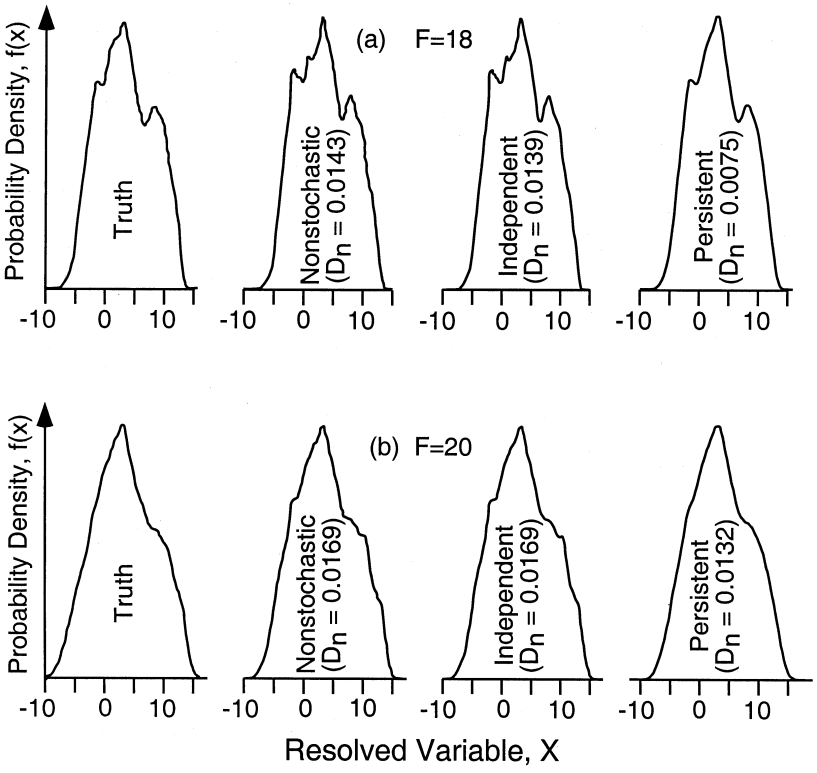

Figure 4. Probability density functions for the resolved variable X_k, for: the true climate, and (as defined in Table 2) non-stochastic parametrizations, stochastic parametrizations with independent random components, and stochastic parametrizations with persistent random components, for forcings: (a) $F = 18$, (b) $F = 20$. The D_n statistic (Eq. (6)) compares the fit of the three model distributions to the respective 'truth' distribution (see text).

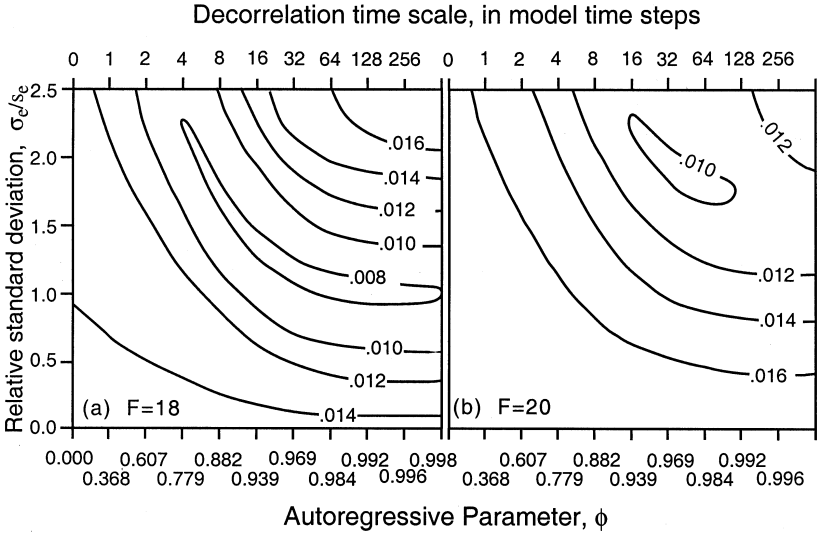

Figure 5. Kolmogorov–Smirnov statistic D_n (Eq. (6)) as a function of the autoregressive parameter ϕ and the amplitude (relative standard deviation) of the stochastic part of the unresolved forcing, σ_e/s_e, for forcings: (a) $F = 18$ and (b) $F = 20$.

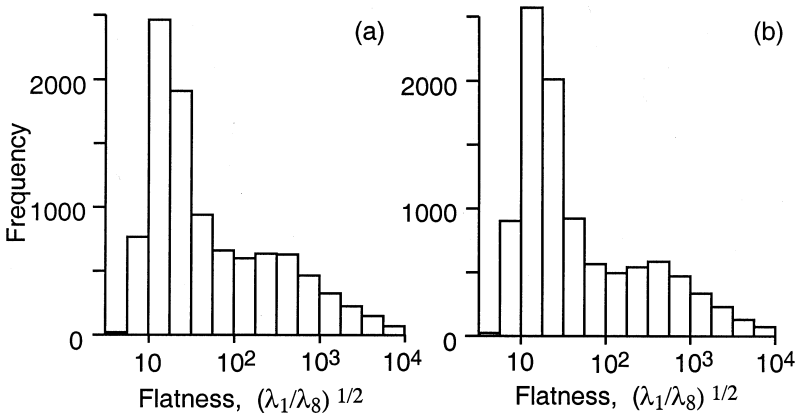

Figure 6. Histograms of the 'flatness index', or square root of the ratio of the largest to the smallest of the eight eigenvalues of the 10 000 local covariance matrices [S_{local}], for forcings: (a) $F = 18$ and (b) $F = 20$. Note the logarithmic scale.

non-zero spatial correlations (not shown) are similar, but they are noticeably worse for large spatial correlations.

4. Stochastic effects on forecast performance

(*a*) *Initialization of ensembles*

Results presented in this section pertain to forecasts beginning at 10 000 initial points computed using Eq. (1), with each consecutive pair being separated by 50 time units. Initial ensembles are chosen randomly, but from distributions approximating the local shape of the attractor, and so reasonably simulating a realistic initialization (Anderson 1996a). These initial distributions were defined through compilations of analogue states, again through long integrations of Eq. (1), within small distances of each of the 10 000 initial points. Here 'small' is interpreted to mean within an (eight-dimensional hyper-) cube with sides of length approximately 5% of the climatological range of X in each dimension (1.0 and 1.2 for $F = 18$ and $F = 20$, respectively, cf. Fig. 4), centred on each of the initial points. These analogue-generating integrations were extended sufficiently for each of the 10 000 cubes to be populated with at least 100 points, although the median number is approximately 900.

The (8×8) covariance matrix of $\mathbf{X}$ for the analogues within each of the 10 000 boxes is defined as [S_{local}]. The local-attractor geometry within the boxes is generally quite 'flat', in the sense that there are one or more directions in the eight-dimensional phase space in which the scatter of analogue states has very little thickness. Figure 6 illustrates the frequency distribution of the local-attractor flatness, expressed as the square root of the ratio of the largest to the smallest of the eigenvalues of [S_{local}], $(\lambda_1/\lambda_8)^{1/2}$. Clearly the local-attractor shape varies considerably depending on the state, but nearly all of these distributions are at least an order of magnitude thinner in their thinnest direction, and the median flatnesses are 25.9 and 29.5 for $F = 18$ and $F = 20$, respectively.

The covariance structure [S_{init}] of the initial ensemble distributions is defined to have the same shape (i.e. same correlations and same eigenvectors) as [S_{local}], but scaled such that on average the standard deviation in each of the $K = 8$ directions is 5% of the

climatological standard deviation, σ_{clim} (Table 2):

$$[\mathrm{S}_{\mathrm{init}}] = \frac{0.05^2\,\sigma_{\mathrm{clim}}^2}{\frac{1}{8}\sum_{k=1}^{8}\lambda_k}[\mathrm{S}_{\mathrm{local}}], \tag{7}$$

where, again, λ_k are the eigenvalues of $[\mathrm{S}_{\mathrm{local}}]$. The results in the following are fairly insensitive to the magnitude of the initial ensemble dispersion, but 5% of the climatological standard deviation is reasonably consistent with current practice in weather forecasting (Simmons and Hollingsworth 2002; Kalnay 2003). The actual ensemble dispersion will be larger in the directions of the leading eigenvectors of $[\mathrm{S}_{\mathrm{local}}]$, and small or nearly zero in the direction(s) of the last eigenvector(s).

The centre, in a population sense, of each ensemble is the point $\mathbf{X}_{\mathrm{anal}}$, simulating an analysis which is removed from each of the 10 000 initial true states $\mathbf{X}_{\mathrm{true}}$ according to a random draw from the multivariate Gaussian distribution with variance matrix $[\mathrm{S}_{\mathrm{init}}]$:

$$\mathbf{X}_{\mathrm{anal}} = \mathbf{X}_{\mathrm{true}} + [\mathrm{S}_{\mathrm{init}}]_i^{1/2}\,\mathbf{z}_i, \quad i = 1, \ldots, 10\,000. \tag{8}$$

Here the square root $[\mathrm{S}_{\mathrm{init}}]^{1/2}$ can be computed through the Cholesky decomposition (e.g. Press *et al.* 1986) of $[\mathrm{S}_{\mathrm{init}}]$, and $\mathbf{z}_i$ is an eight-dimensional vector of independent standard Gaussian variates that is different for each of the 10 000 cases. The ensemble members are then drawn from a multivariate Gaussian distribution centred on $\mathbf{X}_{\mathrm{anal}}$ and with covariance matrix $[\mathrm{S}_{\mathrm{init}}]$:

$$\mathbf{X}_j = \mathbf{X}_{\mathrm{anal}} + [\mathrm{S}_{\mathrm{init}}]_i^{1/2}\,\mathbf{z}_j, \quad j = 1, \ldots, N_{\mathrm{ens}}, \tag{9}$$

where the eight-dimensional independent Gaussian random draws $\mathbf{z}_j$ are different for each of the 10 000 forecasts and N_{ens} ensemble members.

(*b*) *Effects on accuracy of ensemble means*

Figure 7(a) shows RMSE and Fig. 7(b) anomaly correlations for ensemble-mean forecasts as functions of lead time, for the same three parametrizations defined in Table 2, and three choices of ensemble size, N_{ens}. As was also the case for the portrayal of the climatology, the non-stochastic and independent stochastic parametrizations yield quite similar results with respect to these two measures of forecast accuracy. This result is also consistent with that reported by Buizza *et al.* (1999). For single-integration (i.e. non-ensemble, $N_{\mathrm{ens}} = 1$) forecasts, stochastic parametrization accuracy is generally worse than that for the conventional non-stochastic parametrization, and is substantially worse at all lead times for the persistent stochastic parametrization. For lead times of 0.5 time units and 1 time unit the persistent parametrization also yields less accurate mean forecasts for $N_{\mathrm{ens}} = 5$ and $N_{\mathrm{ens}} = 20$ than the non-stochastic parametrizations according to these two measures, although only slightly so.

For longer lead times stochastic parametrizations can provide clear improvements in ensemble-mean accuracy over the non-stochastic parametrization; the independent stochastic parametrization is only slightly better, but the persistent stochastic parametrization provides substantial accuracy improvements. One clear measure of this improvement is provided by the 0.6 anomaly correlation in Fig. 7(b) (grey horizontal line); this level defines a conventional rule of thumb for sufficient forecast accuracy to be useful. For $F = 18$ and $N_{\mathrm{ens}} = 20$, forecasts using the non-stochastic parametrization cross this level at a lead time of 3.5 units, while those computed using the persistent stochastic parametrization cross it at 4.5 time units, yielding a 29% extension of the lead times for useful forecasts. Relative to a single integration with deterministic

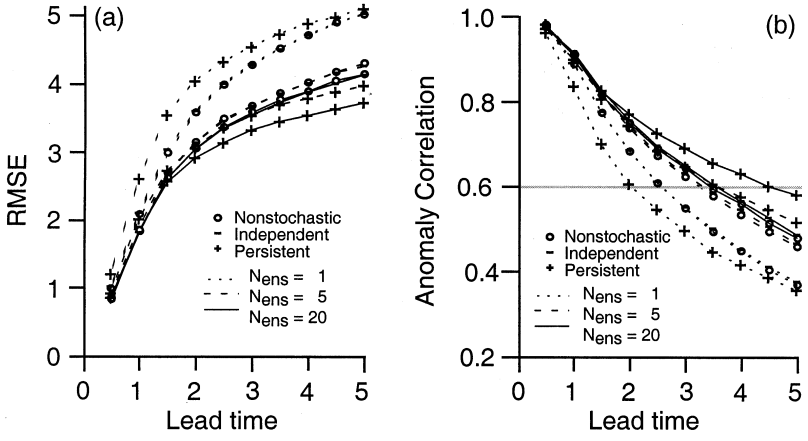

Figure 7. (a) RMSE, and (b) anomaly correlation, given ensemble sizes N_{ens} of 1, 5 and 20, for the three parametrizations defined in Table 2, for forcing $F = 18$.

parametrization, an ensemble size of 20 with the persistent stochastic parametrization results, approximately, in an 80% extension of this critical lead time. Results for larger ensemble sizes are similar but slightly better than those shown for $N_{\text{ens}} = 20$. Corresponding results for $F = 20$ (not shown) are also comparable.

Also apparent from Fig. 7 is the favourable trade-off, with respect to ensemble-mean accuracy, between increasing ensemble size and increasing model complexity through the introduction of stochastic parametrization. Ensemble-mean accuracy, as measured both by RMSE and anomaly correlation, is as good or better for $N_{\text{ens}} = 5$ with the persistent stochastic parametrization than for $N_{\text{ens}} = 20$ ensemble members parametrized deterministically.

The dependence of ensemble-mean accuracy on the amplitude and time-scale of the stochastic forcing is illustrated in Fig. 8, which displays RMSE at a 2 time unit lead time, for $N_{\text{ens}} = 20$. As was the case for accuracy in the portrayal of the climatological distribution (Fig. 5), stochastic parametrizations that are not too extreme provide clear improvements over the non-stochastic parametrization ($\sigma_e/s_e = \phi = 0$). Another similarity is the existence of a broad range of combinations of amplitude and time-scale for which RMSE is close to a minimum, which again is consistent with the results of the more limited tuning exercise described in Buizza *et al.* (1999). For $F = 18$ (Fig. 8(a)) this region is broadly aligned with that corresponding to the best climatological distributions in Fig. 5(a). For $F = 20$ (Fig. 8(b)), this region is nearly the same as for $F = 18$ in Fig. 8(a), and comprises shorter time-scales and smaller amplitudes than the combinations yielding the best climatological distributions in Fig. 5(b). Corresponding figures for stochastic forcing exhibiting spatial correlation (not shown) are again very similar.

(*c*) *Effects on other ensemble properties*

Figure 9 illustrates the effects of stochastic parametrizations on rank histograms (Anderson 1996b; Hamill and Colucci 1997); these tabulate the ranks of the true state within the $N_{\text{ens}} + 1$ member distribution comprised of the union of the ensemble and the true state, over 80 000 ensembles ($K = 8$ forecasts of $\mathbf{X}_k$ for each of 10 000 forecast realizations). Figure 9 pertains to $F = 18$, $N_{\text{ens}} = 20$, and forecast lead times of two units; however, these results are also representative of $F = 20$, other ensemble

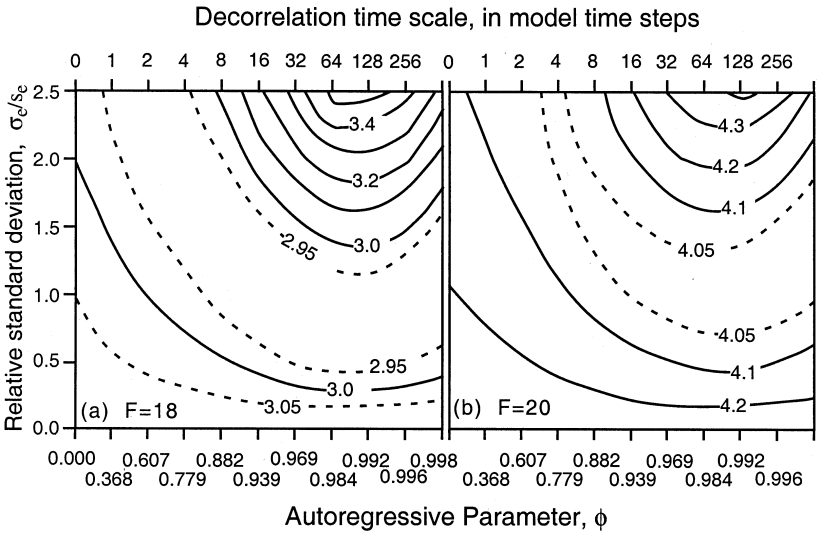

Figure 8. Ensemble-mean RMSE, for ensemble size 20 at a lead time of 2 units, as a function of the autoregressive parameter ϕ and the amplitude (relative standard deviation) of the stochastic part of the unresolved forcing, σ_e/s_e, for forcings: (a) $F = 18$ and (b) $F = 20$.

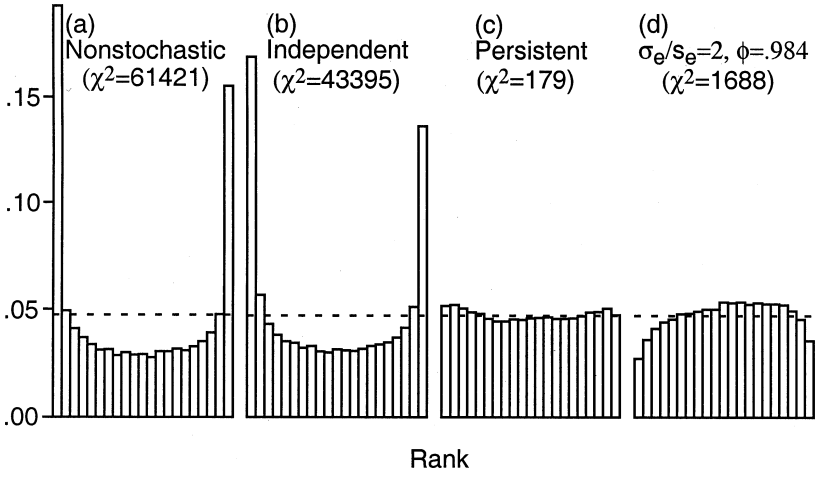

Figure 9. Selected rank histograms, for forcings $F = 18$, and ensemble size $N_{\text{ens}} = 20$, at a forecast lead time of 2 units. Dashed lines indicate the fraction expected in each bin under rank uniformity. Parameters defining cases (a), (b) and (c) are listed in Table 2; (d) shows results from a parametrization with excessive amplitude (see text).

sizes, and other lead times. Results for the three parametrizations defined in Table 2 (Figs. 9(a) to (c)), plus a fourth parametrization with excessive amplitude (Fig. 9(d)) are shown. The U-shaped rank histogram in Fig. 9(a) indicates that ensembles based on the non-stochastic parametrization are strongly under-dispersed, i.e. that forecast ensemble members are much more similar to each other relative to the true state than could be expected by chance (Hamill 2001). The stochastic parametrization with independent random increments (Fig. 9(b)) improves this under-dispersion only slightly, but parametrization with an appropriate combination of amplitude and decorrelation

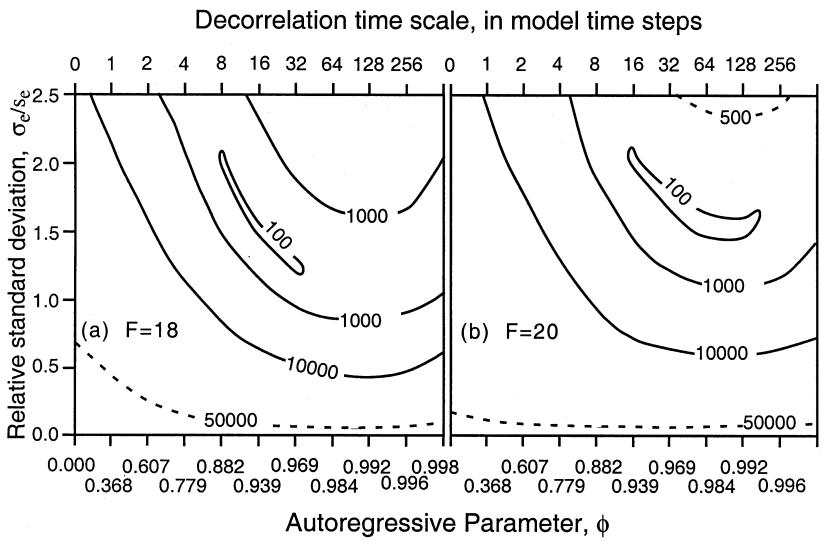

Figure 10. χ^2 statistic for uniformity of rank histograms, for ensemble size $N_{\text{ens}} = 20$, at a forecast lead time of 2 units, as a function of the autoregressive parameter ϕ, and the amplitude (relative standard deviation) of the stochastic part of the unresolved forcing σ_e/s_e, for: (a) forcing $F = 18$ and (b) $F = 20$.

time-scale (Fig. 9(c)) results in a nearly uniform rank histogram. The same decorrelation time-scale coupled with a doubling of the amplitude (Fig. 9(d)) results in a mound-shaped rank histogram, indicating somewhat over-dispersed ensembles on average (Hamill 2001), as the true state is the most extreme of the $N_{\text{ens}} + 1$ values too rarely.

Also shown in Fig. 9 are values of the χ^2 statistic that is conventionally used to evaluate the null hypothesis—that an observed rank histogram is a realization from a generating process in which the true state occupies any of the $N_{\text{ens}} + 1$ ranks with equal probability (dashed lines). This null hypothesis would be rejected for all of these four cases, although the very large sample size makes this a very strict test indeed, since for a given distribution of relative frequencies the χ^2 statistic scales linearly with the sample size. For perspective, note that the nearly flat histogram in Fig. 9(c) would yield $\chi^2 = 17.9$ if the sample size were 8000, and in that hypothetical case would be regarded as clearly consistent with the null hypothesis of uniformity.

Figure 10 shows the dependence of the χ^2 statistic for rank uniformity on the amplitude and time-scale of the random component of the unresolved forcing. As before, a broad minimum of good model performance can be seen for a wide range of combinations of time-scale and amplitude, as exemplified in Fig. 10(a) by $\phi = 0.984$ and $\sigma_e/s_e = 1$, corresponding to Fig. 9(c). For smaller amplitudes and/or time-scales the rank histograms are less flat, and in particular indicate ensemble under-dispersion as exemplified by Fig. 9(a) ($\phi = \sigma_e/s_e = 0$) and Fig. 9(b) ($\phi = 0$ and $\sigma_e/s_e = 1$). Persistent stochastic parametrizations forced with excessively variable random components produce mound-shaped rank histograms, as exemplified by Fig. 9(d) ($\phi = 0.984$ and $\sigma_e/s_e = 2$), indicating ensemble over-dispersion. The results in Fig. 10 are representative of those for other ensemble sizes. Similarly, the shapes of MST (minimum spanning tree) histograms (Smith and Hansen 2004; Wilks 2004), which analogously reflect the multivariate (all $K = 8$ variables simultaneously) relationship between the ensemble and truth, reveal the same pattern of: under-dispersed ensembles for insufficient stochastic

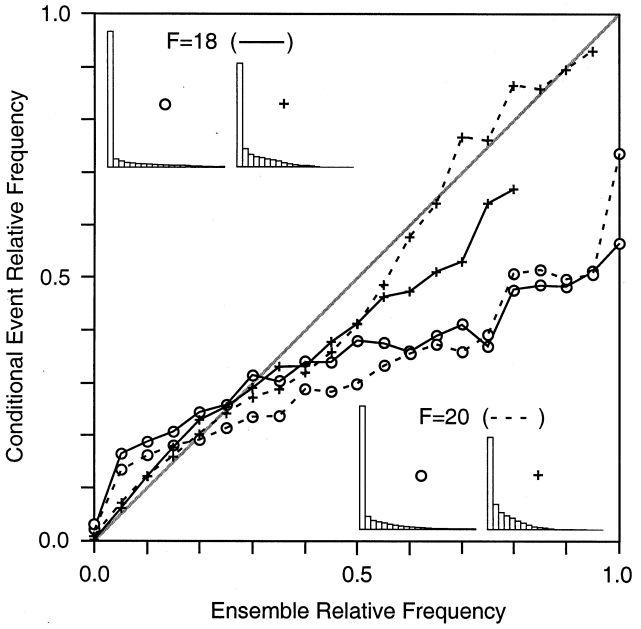

Figure 11. Reliability diagrams for raw ensemble probability forecasts of **X** below the first decile (smallest 10%) of its climatological distribution, for non-stochastic (open circles) and persistent (+ signs) parametrizations as defined in Table 2; for ensemble size $N_{\text{ens}} = 20$, at a forecast lead time of 2 units, for forcings: $F = 18$ (solid lines) and $F = 20$ (dashed lines). Insets show relative frequencies of the $N_{\text{ens}} + 1$ probability values.

amplitude and/or time-scale; a broad optimum producing nearly flat MST histograms; and over-dispersed ensembles for excessive amplitudes (not shown).

A perspective on the ensemble dispersion characteristics as they relate to raw ensemble probability forecasts (regarded as literally equal to ensemble relative frequency, with no post-processing) is provided in Fig. 11. This figure shows reliability diagrams (e.g. Wilks 1995) for a forecast that **X** will be in the smallest 10% of its climatological distribution, for $N_{\text{ens}} = 20$ and a lead of 2 time units. The inset histograms, showing relative frequencies of each of the $N_{\text{ens}} + 1$ possible probability forecasts for the event, clearly reflect the lesser ensemble dispersion of the non-stochastic relative to the persistent parametrizations. This smaller ensemble dispersion results in more cases for which none of the ensemble members are in the smallest 10% (thus forecasting zero probability), and also more cases in which a substantial fraction are in the smallest 10% (the longer distribution tails indicate more frequent forecasting of the larger probabilities). However, the conditional relative frequencies in the main body of the diagram show clearly that the forecasts based on the non-stochastic parametrization are overconfident, and resolve the event relatively poorly (the points in aggregate have a slope that is substantially smaller than that of the ideal 1:1 line, shown in grey). For example, in cases where nearly all the non-stochastic parametrization ensemble members are below the 10th percentile, the actual relative frequency of this outcome is only about 0.50. In contrast, the outcome relative frequencies conditional on the persistent stochastic forecasts are much closer to the correct calibration, especially for $F = 20$.

A more generalized perspective on the accuracy of the raw ensemble probability forecasts is provided by the ranked probability skill (RPS) score (Wilks 1995) for forecasts that simultaneously specify probabilities of **X** being in each of the 10 equiprobable climatological classes defined by the distribution deciles, with the reference

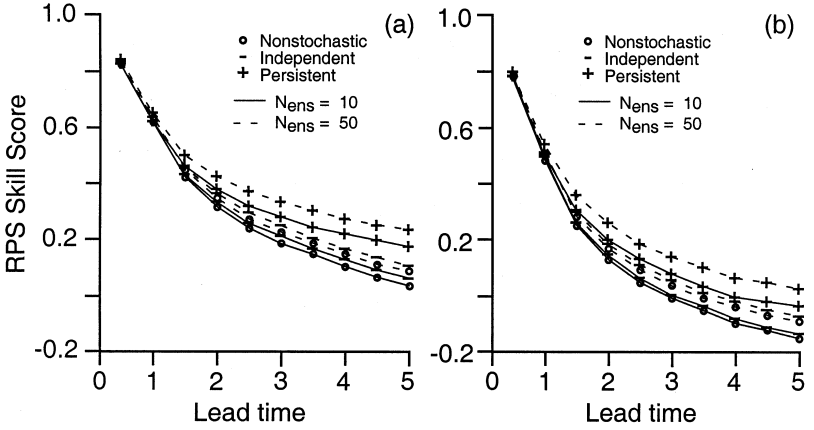

Figure 12. Ranked probability skill (RPS) scores, as functions of forecast lead time, for the parametrizations defined in Table 2, for ensemble sizes N_{ens} of 10 and 50; for forcings: (a) $F = 18$ and (b) $F = 20$.

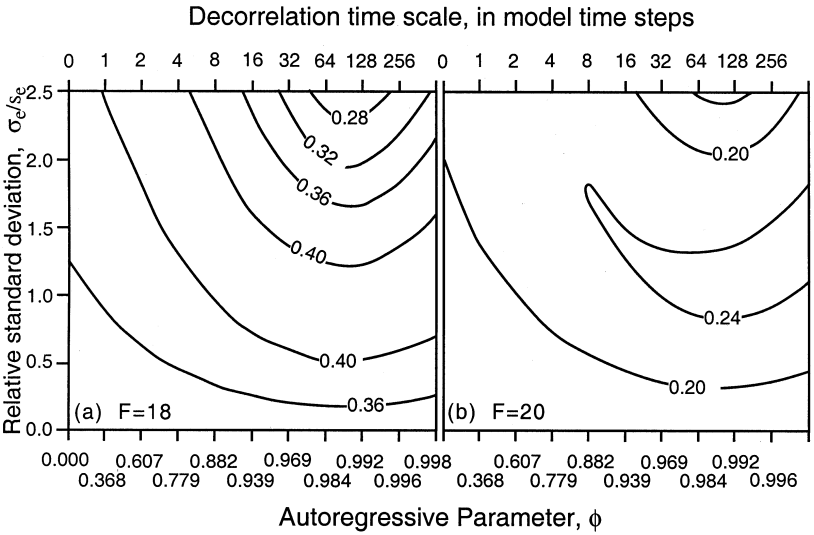

Figure 13. Ranked probability skill (RPS) score, for ensemble size 20 at a lead time of 2 units, as a function of the autoregressive parameter ϕ, and the amplitude (relative standard deviation) of the stochastic part of the unresolved forcing, σ_e/s_e, for forcings: (a) $F = 18$ and (b) $F = 20$.

forecasts taken to be the climatological relative frequencies (= 0.10 for each class). Figure 12 shows the RPS score for two ensemble sizes as a function of lead time. Clearly for both $F = 18$ (Fig. 12(a)) and $F = 20$ (Fig. 12(b)), the independent stochastic parametrization provides a small improvement over the non-stochastic parametrization, while the persistent stochastic parametrization provides a large improvement, and these differences become more pronounced at longer lead times. Again, use of the persistent stochastic parametrization extends substantially the lead times for which particular performance levels are maintained. For example, for $F = 20$ and $N_{\text{ens}} = 10$ the duration of positive skill is extended from 3 to 4 time units, or 33%.

Figure 13 shows the dependence of the RPS score, for $N_{\text{ens}} = 20$ and lead time of 2 units, on the time-scale and amplitude of the stochastic forcing. As before there is a

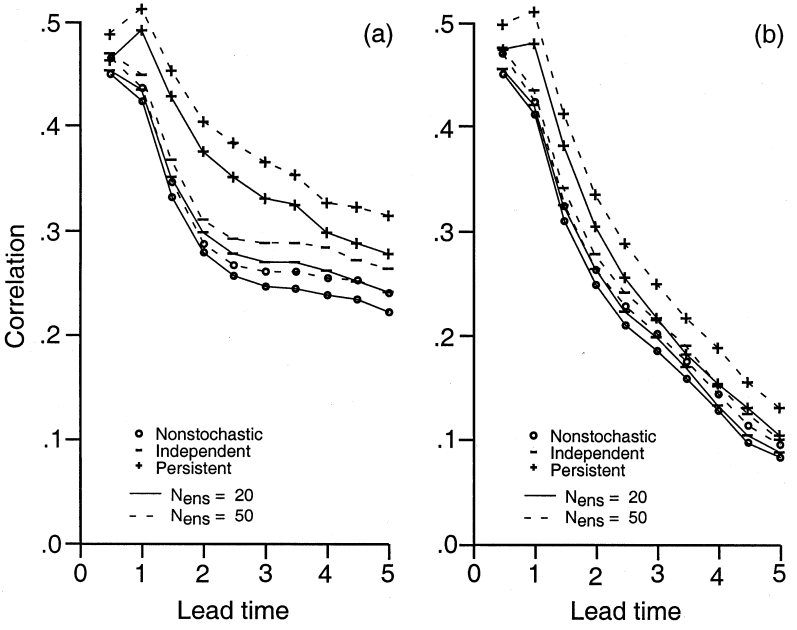

Figure 14. Correlations between ensemble variance and average absolute ensemble-mean error, as functions of forecast lead time, for the parametrizations defined in Table 2, for ensemble sizes N_{ens} of 20 and 50; for forcings: (a) $F = 18$, (b) $F = 20$.

broad optimum, suggesting that good results are achieved for a variety of mutually compatible parameter combinations. Also as before, allowing non-zero 'spatial' correlation for the stochastic parametrizations (not shown) produces a slight degradation of forecast accuracy.

Finally, an attribute of interest is the potential for ensemble dispersion to predict accuracy of the ensemble-mean forecast, in order to 'forecast forecast skill' (e.g. Kalnay and Dalcher 1987; Houtekamer *et al.* 1996; Buizza and Palmer 1998). Figure 14 displays this potential in terms of the correlation between ensemble variance and mean absolute error of the ensemble mean, again for the three parametrizations defined in Table 2, and for two ensemble sizes. Clearly the use of an (at least approximately) correctly parametrized stochastic parametrization improves this correlation, and can substantially extend the lead times for which a given level of correlation is achieved, particularly for $F = 18$ (Fig. 14(a)).

5. Conclusions

This paper has explored the use of stochastic parametrizations in the Lorenz (1996) model. The construction of these parametrizations has been deliberately imperfect (e.g. incorrect form of the time dependence and statistical distribution for the random components) in an effort to mimic the unavoidably imperfect parametrizations that can be derived for real atmospheric models.

Stochastic parametrizations were found to improve the correspondence between the model and true climates, and these improvements were more clearly evident when examining the Kolmogorov–Smirnov statistic D_n, reflecting the overall climatological

distribution, than when considering climatological means or standard deviations. However, the improvement in portrayal of the climate depends strongly on inclusion of adequate time correlation in the stochastic part of the parametrizations, and in combination with a compatible amplitude (standard deviation) for the stochastic variations. This conclusion is consistent with the results of Lin and Neelin (2000, 2002). Parametrizations with uncorrelated random components produced climates that differed very little from the climate produced by the corresponding non-stochastic parametrization. Including 'spatial' correlations in the random forcings for the stochastic parametrizations had very little effect.

The effects of stochastic parametrization on forecast performance were found to be positive only for ensemble forecasts. Consistent with the results of Orrell (2003), stochastic parametrizations degraded scalar accuracy measures (e.g. RMSE, anomaly correlation) for single-integration forecasts (i.e. $N_{\text{ens}} = 1$). However, this result may not necessarily be universally true, particularly for systems where stochastic parametrization may produce sufficiently large improvements in the model climate. Accuracy of ensemble means was clearly improved through the use of appropriately tuned stochastic parametrizations (in terms of combinations of amplitude and time-scale), with the relative advantage increasing with lead time. Similarly, stochastic parametrizations were found to produce more correct ensemble dispersion, yielding nearly flat rank histograms in contrast to the strongly U-shaped rank histograms resulting from use of non-stochastic parametrizations. The result is that better calibrated and more accurate probability forecasts are produced through the use of stochastic parametrization.

These results suggest that stochastic parametrizations may not be worthwhile in non-ensemble settings, except in those cases such as climate simulation where the best achievable correspondence between model and true climates is an important goal. The results suggest strongly that forecasts based on ensemble methods will benefit from appropriate stochastic parametrizations, with those benefits increasing at longer lead times.

Of course the degree to which the patterns of improvements reported here will transfer to atmospheric models remains to be seen, although the general consistency between the present results and those of Buizza *et al.* (1999) is encouraging in this regard. More work with stochastic parametrizations in models of the real atmosphere is clearly warranted. Those parametrizations will probably require empirical tuning, since atmospheric data analogous to those presented in Fig. 2 may not be readily available, and in any case imperfections in the structure of the model and its parametrizations will probably make some tuning advisable. The experience accumulated here suggests that it may be most fruitful to search for appropriate combinations of time-scale and amplitude for the stochastic components. While corresponding spatial correlations had little effect on the simple system considered here, parametrizations of real atmospheric processes may benefit from these as well.

ACKNOWLEDGEMENTS

I thank Tom Hamill and Tim Palmer for constructive suggestions. This work was supported by NSF under grant ATM-0221542.

REFERENCES

Anderson, J. L. 1996a Selection of initial conditions for ensemble forecasts in a simple perfect model framework. *J. Atmos. Sci.*, **53,** 22–36

Anderson, J. L. 1996b A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate*, **9,** 1518–1530

Buizza, R. 1997 Potential forecast skill of ensemble prediction, and spread and skill distributions of the ECMWF ensemble prediction system. *Mon. Weather Rev.*, **125,** 99–119

Buizza, R. and Palmer, T. N. 1998 Impact of ensemble size on ensemble prediction. *Mon. Weather Rev.*, **126,** 2503–2518

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

Evans, R. E., Harrison, M. S. J., Graham, R. J. and Mylne, K. R. 2000 Joint medium-range ensembles from the Met Office and ECMWF systems. *Mon. Weather Rev.*, **128,** 3104–3127

Ewald, B., Penland, C. and Temem, R. 2004 Accurate integration of stochastic climate models with application to El Niño. *Mon. Weather Rev.*, **132,** 154–164

Hamill, T. M. 2001 Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Weather Rev.*, **129,** 550–560

Hamill, T. M. and Colucci, S. J. 1997 Verification of Eta-RSM short-range ensemble forecasts. *Mon. Weather Rev.*, **125,** 1312–1327

1998 Evaluation of Eta-RSM ensemble probabilistic precipitation forecasts. *Mon. Weather Rev.*, **126,** 711–724

Harrison, M. S. J., Palmer, T. N., Richardson, D. S. and Buizza, R. 1999 Analysis and model dependencies in medium-range ensembles: Two transplant case-studies. *Q. J. R. Meteorol. Soc.*, **125,** 2487–2515

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

Kalnay, E. 2003 *Atmospheric modelling, data assimilation and predictability*. Cambridge University Press, Cambridge, UK

Kalnay, E. and Dalcher, A. 1987 Forecasting forecast skill. *Mon. Weather Rev.*, **115,** 349–356

Lin, J. W.-B. and Neelin, J. D. 2000 Influence of a stochastic moist convective parametrization on tropical climate variability. *Geophys. Res. Lett.*, **27,** 3691–3694

2002 Considerations for stochastic convective parameterizations. *J. Atmos. Sci.*, **59,** 959–975

Lorenz, E. N. 1975 Climate predictability. Pp. 132–136 in: *The physical basis of climate and climate modelling*. GARP Publication Series, vol. 16. WMO, Geneva, Switzerland

1996 'Predictability—A problem partly solved'. Pp. 1–18 in Proceedings of seminar on predictability: Volume 1. ECMWF, Reading, UK

Molteni, F., Buizza, R., Palmer, T. N. and Petroliagis, T. 1996 The ECMWF ensemble prediction system: Methodology and validation. *Q. J. R. Meteorol. Soc.*, **122,** 73–120

Mylne, K. R., Evans, R. E. and Clark, R. T. 2002 Multi-model, multi-analysis ensembles in quasi-operational medium-range forecasting. *Q. J. R. Meteorol. Soc.*, **128,** 361–384

Orrell, D. 2002 Role of the metric in forecast error growth: How chaotic is the weather? *Tellus*, **54A,** 350–362

2003 Model error and predictability over different timescales in the Lorenz '96 Systems. *J. Atmos. Sci.*, **60,** 2219–2228

Orrell, D., Smith, L., Barkmeijer, J. and Palmer, T. 2001 Model error in weather forecasting. *Nonlinear Proc. Geophys.*, **8,** 357–371

Palmer, T. N. 2001 A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Q. J. R. Meteorol. Soc.*, **127,** 279–304

Penland, C. 2003 Noise out of chaos and why it won't go away. *Bull. Am. Meteorol. Soc.*, **84,** 921–925

Pitcher, E. 1977 Application of stochastic prediction to real data. *J. Atmos. Sci.*, **34,** 3–21

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. 1986 *Numerical recipes, the art of scientific computing.* Cambridge University Press, Cambridge, UK

Roulston, M. S. and Smith, L. A. 2003 Combining dynamical and statistical ensembles. *Tellus*, **55A,** 16–30

Simmons, A. J. and Hollingsworth, A. 2002 Some aspects of the improvement in skill of numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **128,** 647–677

Smith, L. A. 2001 Disentangling uncertainty and error: On the predictability of nonlinear systems. Pp. 31–64 in *Nonlinear dynamics and statistics*. Ed. A. I. Mees. Birkhauser, Boston, USA

Smith, L. A. and Hansen, J. A. 2004 Extending the limits of forecast verification with the minimum spanning tree. *Mon. Weather Rev.*, **132,** 1522–1528

Stensrud, D. J., Brooks, H. E., Du, J., Tracton, M. S. and Rogers, E. 1999 Using ensembles for short-range forecasting. *Mon. Weather Rev.*, **127,** 433–446

Stensrud, D. J., Bao, J.-W. and Warner,T. T. 2000 Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Weather Rev.*, **128,** 2077–2107

Toth, Z. and Kalnay, E. 1997 Ensemble forecasting at NCEP and the breeding method. *Mon. Weather Rev.*, **125,** 3297–3319

Vannitsem, S. and Toth, Z. 2002 Short-term dynamics of model errors. *J. Atmos. Sci.*, **59,** 2594–2604

Wilks, D. S. 1995 *Statistical methods in the atmospheric sciences*. Academic Press, San Diego, USA

2004 The minimum spanning tree (MST) histogram as a verification tool for multidimensional ensemble forecasts. *Mon. Weather Rev.*, **132,** 1329–1340

Ziehmann, C. 2000 Comparison of single-model EPS with a multi-model ensemble consisting of a few operational models. *Tellus*, **52A,** 280–299