Tellus (2002), 54A, 350–362
Printed in UK. All rights reserved

TELLUS
ISSN 0280–6495

Role of the metric in forecast error growth: how chaotic is the weather?

By D. ORRELL*, *Centre for Nonlinear Dynamics, University College London, Gower Street, London WC1E 6BT, UK*

(Manuscript received 31 January 2002; in final form 8 February 2002)

ABSTRACT

The atmosphere is often cited as an archetypal example of a chaotic system, where prediction is limited by the model's sensitivity to initial conditions. Experiments have indeed shown that forecast errors, as measured in 500 hPa heights, can double in 1.5 d or less. Recent work, however, has shown that, when errors are measured in total energy, model error is the primary contributor to forecast inaccuracy. In this paper we attempt to reconcile these apparently conflicting sets of results by examining the role of the chosen metric. Using a simple medium-dimensional model for illustration, it is found that the metric has a strong effect, not just on apparent error growth, but on the perceived causes of error. If an insufficiently global metric is used, then it may appear that error is due to sensitivity to initial condition, when in fact it is caused by sensitivity to error in the other variables. If the goal is to diagnose the causes of error, a sufficiently global metric must be used. The simple model is used to predict the internal rate of growth of the ECMWF operational model, and preliminary results compared. It is found that both 500 hPa and total energy results are consistent with high model error and a relatively low internal rate of growth. Experiments are suggested to further verify the results for weather models.

1. Introduction

Error in weather forecast models arises from both the initial condition and the model (Bjerknes, 1911). Since weather models are thought to be chaotic, and therefore sensitive to initial condition, it follows that a small perturbation to the initial condition may grow rapidly (Lorenz, 1963). This hypothesis has been tested for weather models in a number of experiments, by launching model forecasts from perturbed initial conditions and measuring the doubling time of error growth. Early estimates of doubling times were for 5 d or more (Mintz, 1965; Smagorinsky, 1963), but more recent results, based on the lagged forecast method discussed below, have indicated a much more rapid doubling time in the 500 hPa metric of 1.5 d or less (Simmons et al., 1995), which implies that most forecast error is a result of chaos.

On the other hand, results in the total energy metric have shown that model error plays a significant role. In fact, the error attributable to the model accounts for the majority of forecast error out to 3 d (Orrell et al., 2001).

Of course, it is well known that different metrics give different error growth (Mullen and Buizza, 2001; Rabier et al., 2000); it is easier, for example, to predict the 500 hPa height than it is to predict the wind at ground level. However, the 500 hPa results seem to imply that model error is relatively low, so most error is due to initial condition, while the total energy results indicate that most error is due to the model. It is not possible for both of these scenarios to be true at the same time, because any model error in the wind and temperature

* e-mail: d.orrell@ucl.ac.uk

variables that make up the total energy metric will affect the 500 hPa height after a period of time.

The question of whether forecast error is caused predominately by sensitivity to initial condition or model errors is of more than theoretical importance. The weather's initial state can never be perfectly known, so the model's sensitivity to initial condition essentially places an upper limit on the atmosphere's predictability, and to some extent the potential economic benefit of weather forecasts (Ehrendorfer, 1997). It also serves as a guide to whether resources should be allocated to improving the initial condition or the model.

In this paper we attempt to reconcile the 500 hPa and total energy results by examining the role of the metric. It is shown that error growth can appear very different when viewed in a non-global, rather than a global, metric; and that errors in a non-global metric can be misattributed to initial condition when in fact they are due to the model. We begin by examining typical patterns of error growth in the 500 hPa and total energy metrics. A medium-dimensional system, based on the Lorenz '96 system, is then constructed which manages to simulate weather model error growth. The likely impact of the metric on experimental results is determined, and the model used to predict the internal rate of growth of an operational model. The possible effects of errors transferring from small scale to large scale is discussed. Finally, preliminary results for the operational model are presented, and further experiments proposed.

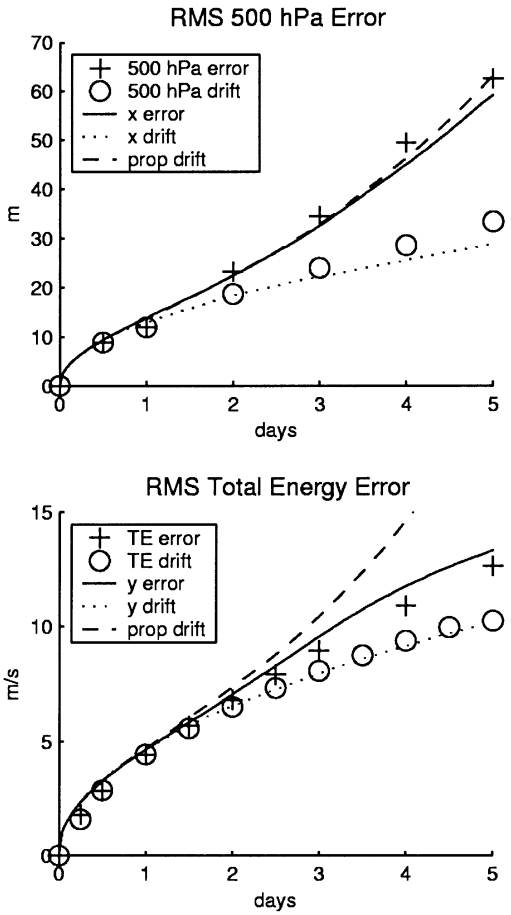

Fig. 1. Plot of r.m.s. forecast errors and drift. Upper panel shows GCM forecast error (+) and drift (○ symbol) in 500 hPa metric over Northern hemisphere for 10 d in October 1999. Lower panel shows total energy error and drift integrated globally over a 15 d period in December 2000. All errors are relative to the analysis. Solid and dotted lines show error and drift for the two-level system over the large-scale x variables (upper) and small-scale y variables (lower), r.m.s. over 500 forecasts, as discussed in text. The dashed lines show the propagated drift.

2. Patterns of error growth

In nonlinear systems, errors due to initial condition typically grow in an exponentially-on-average fashion (Smith et al., 1999). If models were highly sensitive to initial condition, were a good approximation to the real weather, and also grew in a reasonably exponential manner, then doubling times could be determined simply by looking at the error growth relative to the analysis of the observations. In practice, however, errors do not grow exponentially. The upper panel of Fig. 1 shows root-mean-square 500 hPa errors (+ symbols) for a General Circulation Model (GCM), in this case the ECMWF operational model. Error growth is fairly linear, instead of exponential. The situation is even more pronounced in the lower panel showing errors in the total energy metric, which is more global and includes wind and temperature errors over all model levels (Buizza and Palmer, 1995). The error actually has negative curvature.

Orrell et al. (2001) argued that the reason for the negative curvature is because the equations

create model errors which grow in a square-root fashion. To see this, suppose the target orbit (in this case the analysis) is $\tilde{\mathbf{s}}(t)$, and we write the model equations in the form

$$\frac{d\mathbf{s}(t)}{dt} = \mathbf{G}[\mathbf{s}(t)], \tag{1}$$

where $\mathbf{G}$ is C^1 and $\mathbf{s}$ is the state space vector. The forecast error is

$$\mathbf{e}(t) = \mathbf{s}(t) - \tilde{\mathbf{s}}(t), \tag{2}$$

and the tendency error (Klinker and Sardeshmukh, 1992; Schubert and Schang, 1996), is the difference between the model tendency and that of the target:

$$\mathbf{G}_e[\tilde{\mathbf{s}}(t)] = \mathbf{G}[\tilde{\mathbf{s}}(t)] - \frac{d\tilde{\mathbf{s}}(t)}{dt}. \tag{3}$$

The time dependence of the forecast error is then

$$\begin{aligned}\frac{d\mathbf{e}(t)}{dt} &= \frac{d\mathbf{s}(t)}{dt} - \frac{d\tilde{\mathbf{s}}(t)}{dt} \\ &= \mathbf{G}[\tilde{\mathbf{s}}(t) + \mathbf{e}(t)] - \mathbf{G}[\tilde{\mathbf{s}}(t)] + \mathbf{G}_e[\tilde{\mathbf{s}}(t)] \\ &\approx \mathbf{J}[\tilde{\mathbf{s}}(t)]\mathbf{e}(t) + \mathbf{G}_e[\tilde{\mathbf{s}}(t)]. \end{aligned} \tag{4}$$

A first-order solution to this equation is

$$\mathbf{e}(\tau) \approx \mathbf{M}(\tau, 0)\mathbf{e}(0) + \mathbf{d}(\tau), \tag{5}$$

where $\mathbf{M}(\tau, 0)$ is the linear propagator (Strang, 1986) evaluated along the target orbit from time 0 to τ, and the drift vector is

$$\begin{aligned}\mathbf{d}(t) &= \int_0^\tau \mathbf{G}_e[\tilde{\mathbf{s}}(t)]\, dt \\ &= \int_0^t \mathbf{G}[\tilde{\mathbf{s}}(t)]\, dt - \tilde{\mathbf{s}}(\tau) + \tilde{\mathbf{s}}(0). \end{aligned} \tag{6}$$

The drift is equal to the average tendency error multiplied by time, while the propagator term accounts for the average effect of the gradient. The approximation is valid for short times τ, and particularly for shadow orbits, where the shadow time is limited by the shadow radius.

A more exact solution to eq. (4) (Palmer, 1999; Vannitsem and Toth, 2002) is given by

$$\mathbf{e}(\tau) \approx \mathbf{M}(\tau, 0)\mathbf{e}(0) + \mathbf{d}_p(\tau) \tag{7}$$

where the propagated drift vector is

$$\mathbf{d}_p(\tau) = \int_0^\tau \mathbf{M}(\tau, t)\mathbf{G}_e[\tilde{\mathbf{s}}(t)]\, dt, \tag{8}$$

and $\mathbf{M}(\tau, t)$ is the model linear propagator evaluated along the target orbit from time t to time τ. The propagated drift can be viewed as a version of the drift that also incorporates higher-order terms.

For weather models, the propagated drift is not easily calculated. An attempt to estimate its effect was made in Orrell et al. (2001) by considering the exponential growth of the forecast errors in the drift computation. We will see below, though, that if the action of the propagator is on average to expand propagations, then we can usually assume that

$$d(t) \leqslant \|\mathbf{d}_p(t)\|, \tag{9}$$

so the drift is an underestimate of the propagated drift, and approaches the propagated value for short times.

Because the linear propagator increases exponentially, the propagated drift includes higher-order terms, which tend to blow up at large times. The drift, by contrast, is numerically stable, and easy to calculate. It is therefore a useful measure of model error, which will be zero if the model agrees with the target orbit. When the drift is calculated numerically for a finite step size, the integral is performed by summing a series of short forecast errors. For example, suppose the target point at time $t_j = t_0 + j\Delta$ is $\tilde{\mathbf{s}}(t_j)$, and let $\mathbf{s}_j(t)$ for $t \geqslant t_j$ be the model trajectory initiated at the target point $\tilde{\mathbf{s}}(t_j)$. The forward difference approximation to the model tendency at this point is

$$\frac{\mathbf{s}_j(t_{j+1}) - \tilde{\mathbf{s}}(t_j)}{\Delta} \tag{10}$$

and to the system tendency is

$$\frac{\tilde{\mathbf{s}}_j(t_{j+1}) - \tilde{\mathbf{s}}_j(t_j)}{\Delta}. \tag{11}$$

The tendency error is the difference between these two:

$$\frac{\mathbf{s}_j(t_{j+1}) - \tilde{\mathbf{s}}_j(t_{j+1})}{\Delta}. \tag{12}$$

The drift at time t_K is the numerical integral of the tendency error over the target orbit, which is a sum of short forecast errors:

$$d(t_k) = \left\| \sum_{j=0}^{K-1} [\mathbf{s}_j(t_{j+1}) - \tilde{\mathbf{s}}(t_{j+1})] \right\|. \tag{13}$$

The timestep Δ should be chosen sufficiently small

that the calculation converges. If the forecast errors have little correlation, as tends to be the case with weather models, the drift increases like a random walk (Chatfield, 1989), which lends the error curve a square-root shape. Figure 1 shows the drift measured in both 500 hPa and total energy. It can be compared with the forecast error $\mathbf{e}(t)$, which is for a model forecast initiated on the target orbit, so $\mathbf{e}(0) = \mathbf{0}$ in eq. (7).

Judging from the drift in Fig. 1, it appears that the impact of model error is metric-dependent, and accounts for much of the total energy error, but rather less of the 500 hPa error. However, interpretation is complicated by the fact that 500 hPa fields, being a non-global metric, are affected by errors in other variables as time progresses. For example, miscalculations of winds at sea level [or near-surface temperature, as discussed in Viterbo et al. (1999)] may not immediately influence the 500 hPa height, but after a time the error may propagate through the entire system. Therefore the true impact of model error on 500 hPa height is not clear.

In fact, the choice of metric always has an effect, not just on perceived error growth, but perceived causes of error growth, even for simple systems. We will here try to quantify this effect for weather models. Before proceeding further, it may first be worth discussing an alternative explanation for the negative curvature of the growth curve, which is that it is caused, not by model error, but by a growth rate which varies with scale. It is known that, in weather models, small-scale errors tend to saturate more quickly than large-scale errors (Simmons et al., 1995; Savijarvi, 1995; Toth and Kalnay, 1993). Suppose that the initial error growth is very fast, but then saturates, to be replaced by a slower growth rate, which again saturates, and so on, so that the net effect is an error curve which grows with decreasing slope. This picture might be plausible, and would produce a negative curvature slope, but the fact that the drift varies with the square root of time (rather than simply having negative curvature) would have to be considered a fluke. Also, if the growth rate varied with time as described, then one would expect the drift calculation to be sensitive to the step size used in the calculation, as demonstrated for simple models in Orrell (2001). In fact, the 6 h and 24 h time steps give a similar result (i.e., the drift calculation has converged). Finally, while error growth may vary with physical scale, there does not appear to be experimental evidence to suggest that it varies with time in such a drastic fashion. Nevertheless, it is worth bearing in mind that internal error growth will not be perfectly exponential, and there is undoubtedly some variation in growth rate. For example, the growth eventually saturates after a couple of weeks when errors reach a certain limit.

Another possible explanation is that the error is an artefact of analysis error, which is itself a complex combination of observation error and model error. Again, though, this would not explain the scaling behaviour of the drift, nor the fact that the sum of short forecast errors in the drift equals the total error up to 2 d.

Our aim here, however, is not primarily to demonstrate the importance of model error [see also Orrell et al. (2001) for a discussion of this subject], but to examine the effect of the metric on forecast error growth, point out some of the experimental dangers, and show that the 500 hPa and total energy results are consistent with each other. Since experiments on weather models are expensive and sometimes difficult to interpret, we begin by looking at a simple medium-dimensional system, which is capable of simulating many properties of weather model error growth. The results are then verified against a weather model, and further tests proposed.

3. The two-level system

The two-level system used to simulate weather errors is a variant of the Lorenz '96 system, which was previously used to model forecast error growth in Lorenz (1996). There are eight large-scale variables x_i and 32 coupled small-scale variables $y_{i,j}$, which can be viewed as atmospheric variables around a circle, as shown in Fig. 2. The equations, which are given in the Appendix, simulate properties such as advection, damping, and forcing. Model error is provided by stochastic forcing terms which are present in the system, but absent in the model. In addition, the stochastic component of analysis errors is simulated by adding a random noise component to each observation of the system.

Because the x variables are large-scale and slow-varying, while the y variables are small-scale and

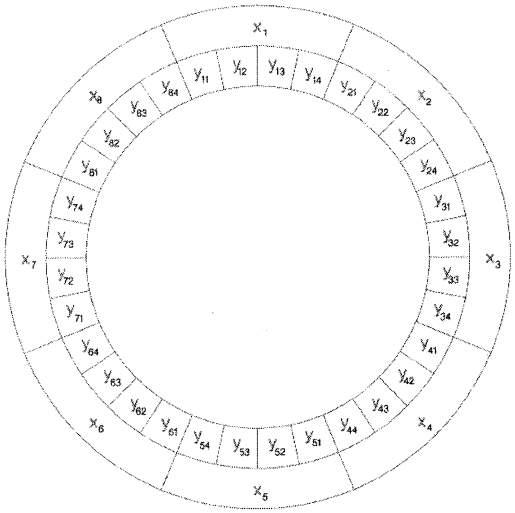

Fig. 2. The two-level system variables. The x variables are large-scale and slow-varying, and can be compared with 500 hPa fields. The y variables are more energetic, and can be compared with wind and temperature fields. The forcing of each x_i variable depends in part on the adjoining $y_{i,j}$ variables.

fast-varying, the former resemble variables such as 500 hPa height, while the latter resemble more energetic variables like wind and temperature. By using an appropriate choice of scaling, the errors can be brought to match their weather model counterparts. The results are shown in Fig. 1: the forecast errors and drift of the large-scale x (solid line, upper panel) agree reasonably well with the GCM 500 hPa errors, while the small-scale y errors (lower panel) match the GCM total energy errors.

The stochastic forcing terms in the model were chosen to give the correct drift (dotted lines in the figure), and as a result are larger for the small-scale y variables than for the large-scale x. They are simply a convenient method to produce the required square-root shape drift, and are not meant to imply that actual forecast errors are stochastic. The observation errors were added to reflect the stochastic component of analysis errors. The real analysis errors are unknown: furthermore, estimates of their magnitude will be affected by model error, due to use of the model in the analysis procedure. Therefore the observation error terms might be of the wrong size, or the wrong sort. In general, though, the observation error has surprisingly little effect on the drift calculation. This can be seen in Fig. 3, which shows the forecast error and drift calculated with a 6 h step length. If the noise term is increased by a factor of 3, then the drift changes a relatively small amount.

It might seem counter-intuitive that observation errors have little effect on a chaotic model. It is often thought that any small perturbation will grow rapidly. However, in a dissipative system such as this, the reality is that small perturbations

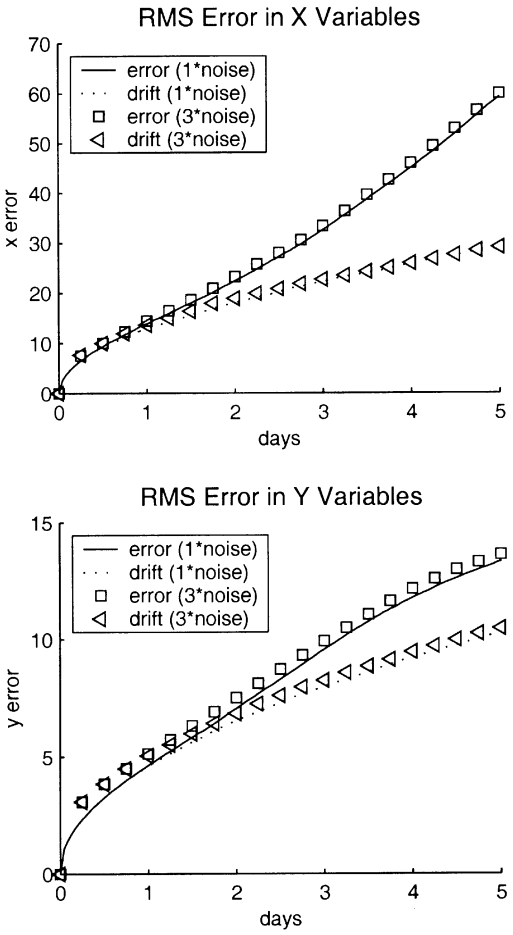

Fig. 3. Plot showing the effect of noise on calculation of r.m.s. forecast errors (solid lines) and drift (dotted lines) for the two-level system. Upper panel shows error in the large-scale x variables, lower panel shows small-scale y variables, r.m.s. over 500 forecasts. When the observation noise term is increased by a factor 3, the effect on either forecast error or drift is small. A 6 h time step was used to calculate the drift.

are at least as likely to shrink as expand over the short term. Errors only grow exponentially over the longer term. Therefore the drift calculation, which is a summed series of short forecasts, is not strongly affected. Even the total forecast error is little changed; we return to discuss this point below.

The propagated drift for the two-level model is also shown in Fig. 1. As in eq. (13), it can again be calculated as a sum of short forecast errors, each of which is this time multiplied by a linear propagator term:

$$\|d_{\mathrm{p}}(t_K)\| = \left\| \sum_{j=0}^{K-1} \mathbf{M}(t_K, t_{j+1})[\mathbf{s}_j(t_{j+1}) - \tilde{\mathbf{s}}(t_{j+1})] \right\|. \tag{14}$$

The effect of the linear propagator $\mathbf{M}(t_K, t_{j+1})$ in the integral will be to rotate each short error term, and (in most cases) stretch it a small amount. Since the direction of the model error is assumed to be random, the rotation has little effect on the total magnitude of the sum; but the stretching will mean that the drift underestimates the true magnitude of the propagated drift, as stated in eq. (9).

In the y variables, the propagated drift is a slightly better estimate of the forecast error than the drift for times up to 4 d. However, it rapidly blows up for larger times due to the presence of the higher-order terms (this could perhaps be corrected by accounting for saturation effects). It is interesting to note, though, that the propagated drift does a much better job of predicting the x errors. It appears that, for the two-level model, the proportion of x error not due to drift in the x terms can be understood as model error rotating in from the y variables through the action of the linear propagator. In other words, it is primarily a result of model error rather than chaos.

Although the two-level parameters were chosen to give realistic error growth, the equations also simulate a number of properties of GCM behaviour. For example, Fig. 4 compares errors for the two-level system $y_{1,1}$ variable (top) with the average 230 hPa temperature over Europe for a 6 d period (bottom). A model trajectory is initiated each 24 h; the resulting errors are indicated by the dashed lines. The plots have a similar nature in terms of their variability and relative error magnitude. Figure 5 shows ensemble errors in the large-scale x variables (upper panel) and small-scale y

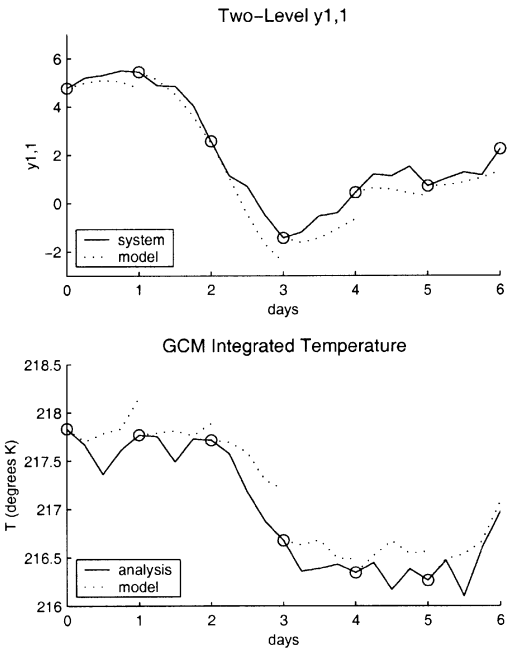

Fig. 4. Figure comparing the model errors for the two-level system $y_{1,1}$ variable (top) with the average 230 hPa temperature over Europe over a 6 d period from 15 October 1999 (bottom) in K. A model trajectory is initiated every 12 h; the resulting errors are indicated by the dashed lines.

(lower panel). The upper panel can be compared with any 500 hPa GCM ensemble. The total energy ensemble is similar, but has a relatively smaller spread due to the larger model error (as for weather models, the spread will also depend to an extent on the initial perturbations).

4. Calculating doubling times

To assess the internal doubling times of error growth due to a small change in initial condition, one must compare not the model trajectory with the true weather, but instead a model trajectory with another from a perturbed initial condition. A question is how to choose the perturbation in a physically realistic manner. Suppose for example that errors are measured in 500 hPa heights: how should the other variables be perturbed? One approach is the Monte-Carlo technique, where random perturbations, with a magnitude con-

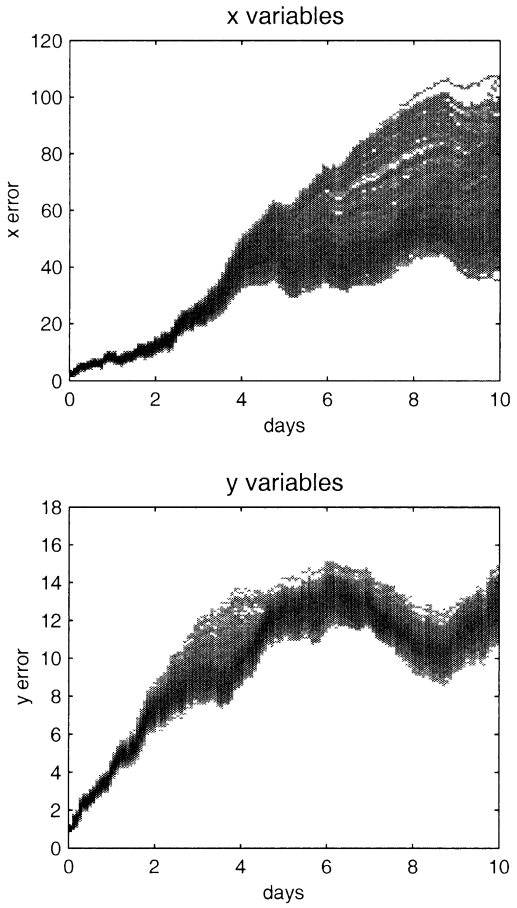

Fig. 5. Plot showing the two-level model ensemble behavior. Upper panel shows an ensemble of large-scale x variables, which can be compared with any 500 hPa GCM ensemble. Lower panel shows a similar ensemble in the small-scale y variables. Note the smaller spread of the y ensemble. The ensembles contain 500 points, perturbed in random directions with magnitude 2 m in x and 1 m s^{-1} in y.

sistent with the observation errors, are added to the initial condition. A potential disadvantage to the Monte-Carlo approach is that the exact characteristics of the analysis error can never be known.

Another widely adopted approach is to let the perturbed initial condition be the one-day forecast generated a day earlier (Lorenz, 1982; Dalcher and Kalnay, 1987). The difference between this perturbed initial condition and the unperturbed control is then just the one-day forecast error. An advantage of this lagged forecast method is that the initial perturbation has a degree of physical relevance: it may not approximate the analysis error, but it does reflect the forecast error after the lag time, and therefore is relevant to the errors experienced as the forecast progresses with time.

Figure 6 shows the lagged forecast technique for the two-level model. Two forecasts are run,

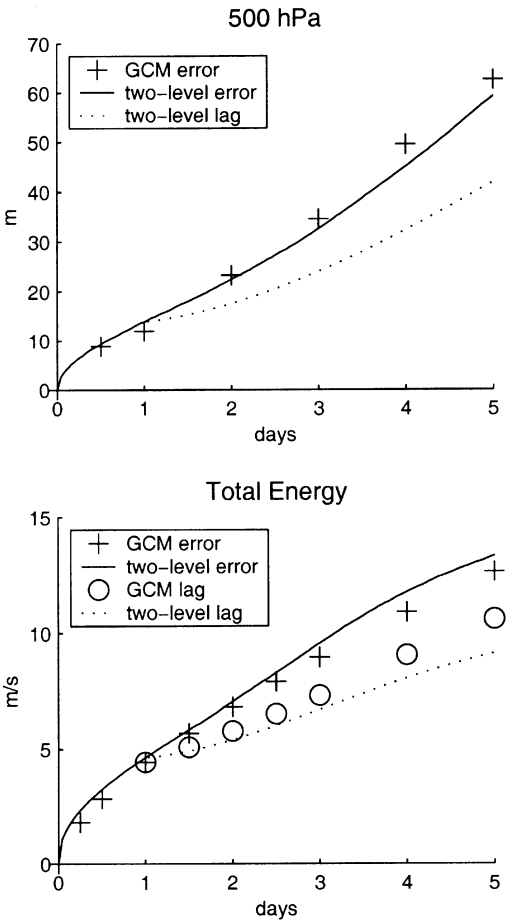

Fig. 6. Plot of r.m.s. forecast errors, normal and lagged. Upper panel shows GCM forecast error (+) in the 500 hPa metric, lower panel shows total energy. Solid line shows errors for the two-level system over the large-scale x variables (upper) and small-scale y variables (lower), r.m.s. over 500 forecasts. Also shown are the difference between consecutive model forecasts with a 1 d lag, which are used to measure sensitivity to initial condition. The two-level results are shown by dotted lines; the GCM results are indicated in the lower panel by (○).

initiated a day apart. The difference between the two is the lagged forecast error. This process is repeated over 500 initial conditions. The root-mean-square errors are shown by the dotted lines starting at time 1 d. When applied to the large-scale x variables of the two-level model (upper panel), it appears that errors double in slightly less than 2 d. When applied to the small-scale y variables (lower panel), however, the growth rate is much slower, and the error takes more than 3 d to double. There is clearly a metric-dependent effect.

The problem is that, in a non-global metric, the concept of a doubling time is not well defined. If the metric only measures errors in certain variables, then the apparent growth rate will depend on errors in the other, unseen variables. For example, from Fig. 6 the x variables appear to grow in a more exponential fashion than the y variables, so it could seem that the model is more sensitive to perturbations in the former than the latter. In fact, the oppposite is true. This is demonstrated in Fig. 7, which shows the result of a Monte-Carlo experiment when applied to the two-level system. The upper panel shows error growth in x when perturbations are made to all variables (+ symbol), and to x variables only (dotted line). Similarly the lower panel shows error growth in y when perturbations are made to all variables, and to y variables only. Perturbations in x were chosen for convenience to be consistent with the 24 h forecast errors, while perturbations in y were chosen to be consistent with the 6 h forecast error. Also shown for comparison are the lagged forecasts, rebased so they begin at $t = 0$.

In the upper panel, when perturbations are made only to x variables, the error growth is essentially flat. In other words, the model is insensitive to changes in x. When perturbations are made to the y variables, the apparent growth rate in x increases. The lagged forecast, however, gives the fastest growth rate. The reason is that, in the lagged forecast experiment, the initial perturbation to the y variables is determined by the 1-d error. Because the 1-d error is large in those variables (due to model error), but relatively small in the x variables, the initial perturbation appears smaller than it would in a more global metric, and the resulting error growth faster. In fact, if model error is significant, the lagged forecast technique is almost unique in its ability to inject a large amount of error into the unseen variables, and create a rapid but apparently plausible rate of growth.

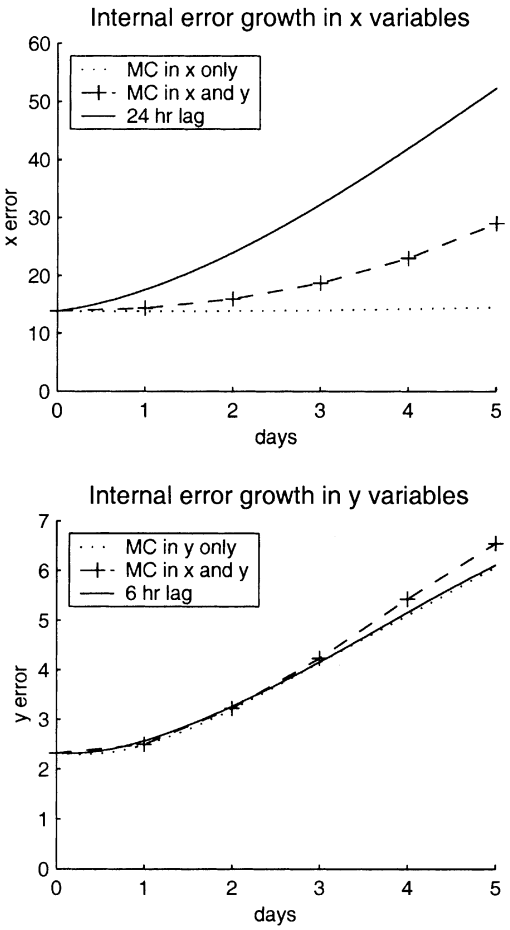

Fig. 7. Plot of internal error growth for the two-level model, as measured by a Monte-Carlo experiment. Upper panel shows r.m.s. errors over 500 runs for the x variables, lower panel shows the y variables. Initial perturbation in x was equal to the 24 h forecast error, the perturbation in y was equal to the 6 h error. Shown are the x errors where x variables alone, or both x and y, are perturbed; and the y errors where y variables alone, or both x and y, are perturbed. Also shown are the errors from lagged forecasts. The results vary depending on method for the x error growth, with the lagged forecast giving the fastest growth. The three methods give similar results when errors are measured in the y variables.

This variability of growth rate may explain why estimates of 500 hPa doubling times depend on the method. For example, some Monte-Carlo tests

in 500 hPa showed no systematic growth at all (Leith, 1964). The apparent growth rate will be a function of which perturbations are made to the unseen variables.

So what is the actual internal doubling time of the two-level model, and what is a realistic value for weather model doubling times? For a representative picture of error growth, the metric must be reasonably global, and it must contain those variables which introduce the largest errors into the system. This latter is especially relevant if certain variables experience large model errors. In the case of the two-level system, the small-scale y variables are responsible for most of the error, due to the large model error component. The lagged forecast in these variables gave a doubling time of slightly over 3 d. As shown in the lower panel of Fig. 7, the Monte-Carlo approach gives the same result, regardless of whether perturbations are made to the y variables only, or both x and y. Because the lagged forecast and Monte-Carlo techniques give the same answer, the doubling time represents a robust measure of the model's internal rate of growth.

It is now clear that the reason why the forecast error in Fig. 3 was unaffected by noise in the observations of the target orbit is that, with such a slow doubling time, error growth is dominated by model error. The noise term in the y observations has standard deviation 0.5 units. Such an error in the initial condition grows by about a factor $2^{5/3}$ to 1.6 units after 5 d. The total error, meanwhile, is 12.8 units. Assuming that the initial condition term is orthogonal to the total error (a safe assumption in a high-dimension space), then a rough idea of the likely combined error is given by the orthogonal sum $\sqrt{(1.6^2 + 12.8^2)} = 12.9$, a negligible difference of 0.7%.

Since the two-level model agrees in other aspects of error growth with GCMs, it seems reasonable to expect that they would have a similar internal rate of growth. Figure 6 shows the result of a total energy lagged forecast experiment performed at ECMWF using the operational model (circle symbols). Apart from the fact that the two-level model saturates earlier, the growth curves are quite similar, and the weather model has a doubling time of about 3 d.

Viewed naively, the lagged forecast in Fig. 6 could be interpreted as indicating that most of the GCM forecast error is a result, not of model error, but simply the growth of the 1-d error. However, this is no more true of the GCM than it is of the two-level model. In a high-dimension system, error due to initial condition is expected to be orthogonal to error due to the model, so that only a portion contributes to the total error. Furthermore, saturation effects mean that the contributions from different sources are not additive even in an orthogonal sense. In general, the convolution of model error and initial condition error is complex, and it is difficult to infer the contribution of each from the forecast error curve alone.

While the lagged experiment is identical, apart from metric, to that used for the 500 hPa lagged forecast experiments, it should still only be considered a preliminary result. We discuss in the conclusions further tests that can be carried out.

5. Other systems

The two-level system has a special structure that combines large-scale x parameters with small scale y parameters, which is useful because it aids comparison with the 500 hPa and total energy metrics. However, it leads to questions about the transfer of error from small-scale to large-scale, which as mentioned in Section 2 is known to occur with weather models. Is the difference that we see between x and y forecast error growth due to a similar effect?

In fact, the difference between x and y error growth is due primarily to the disparity in their stochastic model error terms. This can be seen by looking at simpler systems. For example, the one-level system is similar to the two-level system but has only x variables. The equations, which are given in the Appendix, are similar to the large-scale equations, with the difference that the forcing $F = 10$ is a constant function of time. The model is the same as the system, but the forcing term for variables x_2 through x_8 is $F = 12$, an error of 20%, while x_1 has the same forcing as the system. Therefore x_1 is now the low-error variable, but there is no difference in scale between it and the other variables.

The model's tendency error is equal to the constant forcing error. When measured over all variables seven of which have forcing error 2, the magnitude of the tendency error is equal to $2\sqrt{7}$.

If measured over the first variable x_1 only, the forcing error for that variable is zero, so the tendency error is zero. The drift, which is the integral of the tendency error over the target orbit, will then be equal to $2\sqrt{7}t$ over all variables (i.e., linear), but zero over x_1. The effect of the drift can be seen in the forecast errors in the left panels of Fig. 8. The x_1 error (solid line, top left panel) has an initial slope of zero, because tendency error and drift in that variable is zero. The error over all variables (lower left) grows linearly for small t, with a slope equal to $2\sqrt{7}$ as expected.

Since the forecast error in x_1 grows in a quasi-exponential fashion, it is possible to try to fit it with one of the standard error growth formulas from the literature. An example from Leith (1978) is

$$\frac{dE(t)}{dt} = \alpha E(t) + S \tag{15}$$

where $E(t)$ denotes the error variance, S is a (constant) model error term, and α specifies the internal growth rate due to initial condition error. The dashed line shows this fit with $S = 0.0204$ and a doubling time of 0.082 time units. One would therefore conclude from this fit that model error was small, and the doubling time was fast. This conclusion would be backed up by the x_1 lagged forecast, made with a lag of 0.1 time units, which is shown by the dotted line.

In the right panels, the lagged forecasts are compared with error growth from a Monte-Carlo experiment where perturbations are made to all variables. The actual doubling time over all variables, using either the lagged technique or the Monte-Carlo method, is about 0.27. The reason for the difference in estimated doubling time is again that the x_1 error is due not to sensitivity to initial condition, but to error in other variables; one can think of the model error as advecting in from the other x_i (or drift rotating in due to the action of the linear propagator).

Such curve fits should only be applied with extreme caution, since the convolution of model error with initial conditional error will be complex. [Note also that eq. (15) is not perfectly compatible with eq. (7).] For the case considered here, use of the formula with a non-global metric led to a factor 3.3 error in estimated doubling time.

This example shows that the difference in error growth between high-error and low-error parameters is not a unique property of the two-level system, or due to the disparity in scale between variables. Rather, it is a fairly generic property of model/system pairs where model error is distributed unevenly among the variables. It is easy to come up with versions of even simpler systems, such as Lorenz '63, which behave similarly.

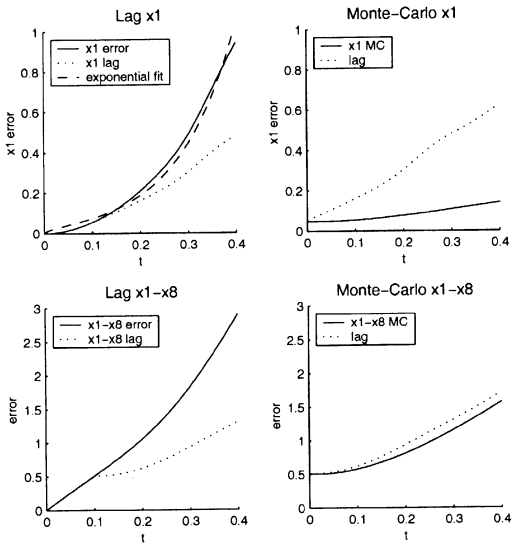

Fig. 8. Plot of error growth for the one-level model. Top panels show errors in the low-error variable x_1 only, lower panels show errors over all variables. Left panels show r.m.s. error growth and a lagged forecast with a lag time of 0.1. The x_1 error (top left) is compared with a fit using eq. (15) with doubling time 0.082. The right-hand panels compare the lagged forecasts (rebased to start at time zero) with Monte-Carlo experiments where errors are made to all variables. The actual internal doubling time is 0.27 time units.

6. Conclusions and future work

In this paper we have aimed to show that experimental results in 500 hPa and total energy, which appear to have conflicting error patterns, are in fact consistent with one another. A medium-dimensional system was constructed which managed to simulate both types of error, with the result depending only on the chosen metric. The internal doubling time of the model was about 3 d. It was shown that error growth can only be accurately diagnosed if the metric is sufficiently

global, and in particular contains the high-error variables. This requirement is common to all dynamical systems, from the Lorenz '63 system up to full weather models. Experiments with the simpler one-level system showed that the difference in error growth was due to the uneven distribution of model error, rather than the disparity in scale of the variables.

In weather models, the total energy metric is better suited for the purposes of error diagnosis than the 500 hPa metric. By integrating wind and temperature errors over the entire space, it offers the most global, complete and physically relevant description of the atmospheric state (after all, you cannot feel the 500 hPa height on your face, especially if you are standing at sea level). Most importantly, these variables are also responsible for the majority of the error. The 500 hPa height, by contrast, is too strongly affected by unseen error in other variables to be a reliable measure of error growth. For example, it tends to reflect the average temperature below the layer, so is intrinsically smoother and less variable than the actual temperature measurements, and tends to mask the causes of error. Apparent rapid error growth indicates sensitivity to errors in the unseen variables rather than sensitivity to initial condition. In particular, experiments with 500 hPa lagged forecasts, on which many current estimates of doubling time are based, give misleadingly fast results.

This is not to say that the 500 hPa field is not a useful metric for forecasting. A forecaster may well prefer to use the 500 hPa height, which by its nature will smooth out some of the short-term errors in, for example, temperature. However, the aims on the one hand of making a forecast, and on the other of diagnosing the cause of forecast error in a numerical model, are completely different. For the former, a model output which experiences low error in the near term is obviously preferable; but one cannot determine the cause of forecast error by ignoring the contribution of high-error variables as time unfolds.

Preliminary tests using a lagged forecast in total energy indicate a doubling time of about 3 d, as for the two-level model. This result is not particularly unusual: historically it is in the intermediate range of doubling time estimates, and if results based on 500 hPa lagged forecasts are discounted it is even on the faster end of the scale. Compared, though, to a doubling time of 1.5 d, it represents an order of magnitude difference in the amount of error attributable to initial condition over a 10-d forecast. This finding is also consistent with other localised studies where sensitivity to initial condition was seen to be small (Shukla, 1998).

Because an accurate estimate of internal doubling times is useful for allocating resources and estimating the potential economic worth of forecasts, more experiments should be carried out using a global metric such as total energy. The lagged forecast test can be performed with a variety of lag times to determine whether the internal growth rate changes with time. Monte-Carlo experiments provide an independent test of growth rates, and can also be used to determine the dependence of growth rate on the size of the initial perturbation. Together, lagged forecasts and Monte-Carlo experiments should give a robust and complete picture of a weather model's sensitivity to initial condition.

As for the weather itself: by searching a 5-yr climate record for reasonably near analogs, and extrapolating to small errors, Lorenz (1969) estimated an error doubling time in the region of 2.5 d. However, the analogs were not very close. Van den Dool (1994) calculated that it would take an observational record of order 10^{30} yr before the atmospheric state repeated itself to within the current observational error, so we will have to wait a long time for a more accurate figure. In the meantime, if the models are anything to go by, it may be the case that the atmosphere is more predictable, and the potential economic worth of forecasting higher, than some of the more pessimistic estimates have indicated.

7. Acknowledgements

Thanks to M. Leutbecher for performing the total energy experiments, and to J. Barkmeijer for helping with other calculations. Thanks also to L. Smith for useful discussions.

8. Appendix: the systems

The two-level system is a scaled version of the Lorenz '96 system, which was used in Lorenz (1996) to simulate error growth, with stochastic

terms added. The equations are

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F - \sum_{j=1}^{4} y_{i,j} + N_x \quad \text{(A1)}$$

$$\frac{dy_{i,j}}{dt} = c^2 y_{i,j+1}(y_{i,j-1} - y_{i,j+2}) - c y_{i,j} + x_i + c N_y \quad \text{(A2)}$$

for $i = 1$ to 8, and $j = 1$ to 4. The indices are cyclic, so for example $x_{i+8} = x_i$ and $y_{i,j+4} = y_{i+1,j}$, and the variables can be viewed as atmospheric quantities around a circle. The parameter c is set to 10. The x variables are scaled by a factor 900 to put in units of m for comparison with GCM 500 hPa results, while the y variables are scaled by a factor 5.3 to put in units of m s^{-1} for comparison with total energy. Time is scaled by a factor 100 to put in days. $F = 14$ is a constant forcing term, while N_x and N_y are random variables with variance 2.5 and 7.5 respectively, updated every hour. In addition, the x and y variables are observed each hour with a stochastic error O_x and O_y, which have standard deviation 1.0 m and 0.5 m s^{-1}, respectively. These terms are meant to simulate the random component of the analysis errors.

The model has the same equations, but with no stochastic forcing, so $N_x = N_y = 0$, and no observation error so $O_x = O_y = 0$. The difference between the model and the system is therefore the stochastic forcing terms, and the observation error. Equations are solved using a Runge–Kutta scheme with time step of 1 h. A long transient of 100 000 h is run before making calculations.

The one-level system has only x variables. The equations are

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F_i \quad \text{(A3)}$$

for $i = 1$ to 8. The indices are again cyclic, so $x_{i+8} = x_i$. In the system, $F_i = 10$ for all i. In the model, $F_1 = 10$, so there is no forcing error in the first variable, but $F_i = 12$ for $i > 1$ which introduces model error. The variables are left unscaled, and there is no stochastic observation error. Equations are solved using a Runge–Kutta scheme with time step of 0.005.

REFERENCES

Bjerknes, V. 1911. *Dynamic meteorology and hydrography, Part II. Kinematics*. Gibson Bros., Carnegie Institute, New York.

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

Chatfield, C. 1989. *The analysis of time series*. Chapman and Hall, New York.

Dalcher, A. and Kalnay, E. 1987. Error growth and predictability in operational ECMWF forecasts. *Tellus* **39A**, 474–491.

Ehrendorfer, M. 1997. Predicting the uncertainty of numerical weather forecasts: a review. *Meteorol. Z.* **6**, 147–183.

Klinker, E. and Sardeshmukh, P. D. 1992. The diagnosis of mechanical dissipation in the atmosphere from large-scale balance requirements. *J. Atmos. Sci.* **49**, 608–627.

Leith, C. 1964. Numerical simulation of the earth's atmosphere. *Meth. Comput. Phys.* **4**, 1–28.

Leith, C. 1978. Objective methods of weather prediction. *Ann. Rev. Fluid Mech.* **10**, 107–128.

Lorenz, E. N. 1963. Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141.

Lorenz, E. N. 1969. Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.* **26**, 636–646.

Lorenz, E. N. 1982. Atmospheric predictability experiments with a large numerical model. *Tellus* **34**, 505–513.

Lorenz, E. 1996. Predictability — a problem partly solved. In *Predictability* (ed. T. Palmer). European Centre for Medium-Range Weather Forecasting, Shinfield Park, Reading, UK.

Mintz, Y. 1965. Very long-term global integration of the primitive equations of atmospheric motion, *Proc. WMO/IUGG Symposium on Research and Development Aspects of Long-Range Forecasting,* **66**, 141–167.

Mullen, S. and Buizza, R. 2001. Quantitative precipitation forecasts over the United States by the ECMWF ensemble prediction system. *Mon. Wea. Rev.* **129**, 638–663.

Orrell, D. 2001. Modelling nonlinear dynamical systems: chaos, error, and uncertainty. *D.Phil. Thesis*, Oxford University.

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

Palmer, T. N. 1999. A nonlinear dynamical perspective on climate prediction. *J. Climate* **12**, 575–591.

Rabier, P., Jarvinen, H., Mahfouf, J. F. and Simmons, A. 2000. The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Q. J .R. Meteorol. Soc.* **126**, 1143–1170.

Savijarvi, H. 1995. Error growth in a large numerical forecast system. *Mon. Wea. Rev.* **123**, 212–221.

Schubert, S. and Schang, Y. 1996. An objective method for inferring sources of model error. *Mon. Wea. Rev.* **124**, 325–340.

Shukla, J. 1998. Predictability in the midst of chaos: a scientific basis for climate forecasting. *Science* **282**, 728–731.

Simmons, A., Mureau, R. and Petroliagis, T. 1995. Error growth and estimates of predictability from the ECMWF forecasting system. *Q. J. R. Meteorol. Soc.* **121**, 1739–1771.

Smagorinsky, J. 1963. General circulation experiments with the primitive equations. *Mon. Wea. Rev.* **91**, 99–164.

Smith, L. A., Ziehmann, C. and Fraedrich, K. 1999. Uncertainty dynamics and predictability in chaotic systems. *Q. J. R. Meteorol. Soc.* **125**, 2855–2886.

Strang, G. 1986. *Introduction to applied mathematics*, Wellesley-Cambridge Press.

Van den Dool, H. M. 1994. Searching for analogues, how long must we wait? *Tellus* **46A**, 314–324.

Toth, Z. and Kalnay, E. 1993. Ensemble forecasting at NMC: the generation of perturbations. *Bull. Am. Meteorol. Soc.* **74**, 2317–2339.

Vannitsem, S. and Toth, Z. 2002. Personal communication.

Viterbo, P., Beljaars, A., Mahfouf, J. and Teixeira, J. 1999. The representation of soil moisture freezing and its impact on the stable boundary layer. *Q. J. R. Meteorol. Soc.* **125**, 2401–2426.