Ensemble forecasting in a system with model error

David Orrell

Institute for Systems Biology,

1441 North 34th Street, Seattle, USA

Abstract

Error in weather forecasting is due to inaccuracy both in the models used, and in the estimate of the current atmospheric state at which the model is initiated. Because weather models are thought to be chaotic, and therefore sensitive to initial condition, the technique of ensemble forecasting has been developed in part to address the latter effect. An ensemble of forecasts is made with perturbed initial conditions, the aim being to produce an estimate of the probability distribution function for the future state of the weather. Some ensemble schemes also include changes to the model. While the ensemble approach is quite widely adopted, however, its verification is complicated, and the effect of model error on ensemble performance is not clear. In this paper, we investigate the effect of model error on ensemble behavior for a version of the Lorenz '96 system. It is shown that estimates of the model's ability to shadow the observations, obtained using the model drift, are robust to observational error and smoothing schemes such as 4DVAR, and help reveal the effect of model error on ensemble performance. Comparisons are made with full weather models. The aim is to provide a study of ensemble error in the context of the Lorenz '96 system, which may be useful in formulating questions and experiments for weather models.

1 Introduction

Ensemble techniques have become established in recent years as a method for generating probabilistic weather forecasts. By running forward an array of slightly perturbed initial conditions, the ensemble forecast is intended to provide an approximation to the probability density function of the weather's future state (Palmer, 2000). While ensemble schemes have proved to be useful tools in understanding the role of initial condition error for weather models, their verification can be complicated (Ehrendorfer, 1997), and the influence of model error on ensemble results is not clear. As ensemble techniques are extended to related fields such as biological oceanography (Robinson et al., 1999), the need to evaluate the effect of model error becomes increasingly clear.

Ensemble schemes have evolved considerably over the years, and now often include perturbations to the model (Buizza et al., 1999); however the original motivation for their use (Toth and Kalnay, 1993) was to counter the effect of sensitivity to initial condition (Lorenz, 1963). It was assumed that model error should be relatively small, at least for short forecast times (Buizza et al., 2000), so that forecast error would be dominated by the initial condition rather than the model (Toth et al., 1996). In the spirit of ensemble forecasting, this paper starts from a different initial condition, or set of assumptions, and examines ensemble fore-

casting for a model/system pair, based on the Lorenz '96 system (Lorenz, 1996), where the model is not particularly sensitive to initial condition, and model error is large.

Comparisons of the relatively simple system used here and weather models are made for the purpose of motivating the choice of system parameters. An advantage of simple systems is that they are easy to experiment with and understand (cf for example (Anderson, 1996)); however they are of course no substitute for full weather models. The aim is therefore to ask how robust ensemble schemes are in general to the effects of model error, and motivate experiments, rather than draw specific conclusions about weather models or provide a survey of current forecasting techniques.

2   Error growth in the stochastic system

The system used here is a 16D scaled version of the one-level Lorenz '96 system (Lorenz, 1996; Orrell and Smith, 2003), with an additional stochastic forcing term. In this section, we introduce the system and model and study the dynamics of error growth, before examining the ensemble behavior.

The equations for the *stochastic system* are

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F + \Delta F \epsilon(t) \tag{1}$$

for $i = 1$ to 16. The index i is cyclic so that $x_{i-16} = x_{i+16} = x_i$. The x_i's are scaled by a factor $c_x = 22$ to put in units of ms^{-1}, and t by a factor $c_t = 16$ to put in units of days. The parameter F is a constant forcing term, set to $F = 10$. $\epsilon(t)$ is a piecewise constant forcing term, that is updated each $\Delta t = 1$ hr, by selecting from a Gaussian random variable with variance 1. It is multiplied by the factor ΔF here set to 1. The equation can therefore be viewed as a numerical implementation of a Wiener process. The system is calculated using a Runge-Kutta scheme with a timestep of one hour, and observed with a stochastic error with standard deviation $S = 0.2ms^{-1}$. The model used to approximate the system has the same equations, but without the stochastic forcing error or observation error. The dimension was chosen as 16 to provide a sufficiently high dimension space, while the scaling was chosen to allow comparison with weather models in both sensitivity to initial condition (as measured by doubling time) and error growth (as measured by deviation from the target orbit), as seen below.

The root-mean-square (RMS) error growth for the model is shown in the top

5

left panel of Figure 1, along with the drift and the propagated drift. As discussed in the Appendix, the drift is a sum of short forecast errors, which approximates the total error for short times, in this case out to about 2 days. The propagated drift accounts for the growth of the short forecast errors under the model dynamics, and is a good approximation to the forecast error to beyond 4 days.

The drift has components related to the effect of both observation error and the error in the equations. For the stochastic system, the two can be computed directly (Orrell, 2005). The tendency error due to the model equations alone is $\Delta F \epsilon(t)$. Because of the stochastic nature of the model error, the drift due to the model equations grows in a square-root fashion, like a random walk, with magnitude

$$d^m(\tau) = \Delta F \sqrt{\Delta t \tau}. \tag{2}$$

Here $\Delta t = 1$ hr is the time at which the stochastic forcing is updated. In cases as here where the observation error is uncorrelated with the model error, it contributes a drift of

$$d^o(\tau) \approx \sqrt{2}S, \tag{3}$$

6

and the two components can be summed orthogonally to give

$$d(\tau) \approx \sqrt{d^m(\tau)^2 + d^o(\tau)^2}.$$
(4)

This partitioning of the drift into two components is shown in the top panel of Figure 2. For this model, the partitioning shows that most of the error for times between 1 and 4 days is due to the model equations. In the next sections, the drift is used as a tool for determining the effect of model error on shadow behavior and ensemble performance.

3 Ensemble performance

The above analysis showed that most of the forecast error for this model/system pair, over the time period of interest, is introduced by model error. The doubling time of initial errors for the model, due to the effect of sensitivity to initial condition alone, is around three days, which is about the same as the doubling time in total energy of weather models (Orrell, 2002). In this section, we consider the effect of model error on ensemble performance.

The lower left panel of Figure 1 shows ensemble errors for the stochastic system. The ensemble was formed by taking +/- perturbations of magnitude $0.5ms^{-1}$

in the directions of the leading four singular vectors, with the singular vectors optimized for a time of two days. For this small ensemble, the increasing errors show that the ensemble members are moving away from truth. The effect is more evident in the top left panel of Figure 4, which shows an ensemble of 1000 randomly perturbed members.

For a higher-dimension system, a suitably large ensemble would probably need millions of members to properly sample the space in a similar way. Another approach might be statistical techniques such as rank histograms. As an example, the lower left panel of Figure 4 shows a rank histogram diagram for an 8-member ensemble, which is formed again from perturbations in the positive and negative directions of the leading 4 singular vectors, computed for a time of 48 hours. These diagrams, which are discussed for example in (Ehrendorfer, 1997; Wilks, 1995), provide a statistical test of the ensemble by counting the distribution of the true system relative to the ensemble predictions. Ideally, the distribution should be flat, but here there is a distinct U-shape which indicates that the true values are often falling above or below the ensemble's range. The same effect is typically seen with weather models (Strauss and Lanzinger, 1996). However, while it is obvious that the ensemble has a problem, it is hard to determine whether this is due to model error or just an inappropriate choice of ensemble members.

8

The best method to determine whether a model ensemble can contain a member which stays near truth is by searching for such model orbits directly. Given a time τ and radius r, we define a shadow orbit as a model orbit $\mathbf{s}(t)$ for which the error vector

$$\mathbf{e}(t) = \mathbf{s}(t) - \tilde{\mathbf{s}}(t) \tag{5}$$

satisfies $\|\mathbf{e}(t)\| \leq r$ for $0 \leq t \leq \tau$. (Note that under this definition, there is no question of whether a model *can* shadow a target orbit, only within which radius; any model will shadow for time τ if the shadow radius is set to the maximum error over the time period 0 to τ.) For a particular radius r, model orbits which shadow a trajectory of the true system for the longest possible time τ can be found by optimization methods which choose the optimal initial condition.

The lower panel of Figure 2 shows the result of a series of shadow experiments for the stochastic Lorenz system. For a particular shadow radius r, the maximum, minimum and median shadow times were determined by use of an optimization program. The results were then plotted in a reflected sense as radius versus time. Also shown is a plot of the drift divided by two. As discussed in (Orrell et al., 2001), the shadow behaviour is limited by the forecast error and drift. In particular, an approximate lower bound for the expected radius at which the model can shadow for time τ is given by $\frac{d(\tau)}{2}$. More loosely, a curve of the median shadow

9

time should lie close to or above a plot of the drift divided by two, as in the figure. The performance of the ensemble in terms of shadow behavior is therefore directly related to the degree of model error as measured by the drift.

For comparison with weather models, and to better motivate the above results, the top right panel of Figure 1 shows error growth for an experiment in which the ECMWF T42 model was compared with the TL159 model. The metric is total energy (Rabier et al., 1996; Buizza and Palmer, 1995), scaled so that units are ms^{-1}. In this experiment, which was described in (Orrell et al., 2001), the model error is created by the difference between the two sets of equations, while the observation error, clearly visible in the initial error, is caused by the truncation from high-resolution to low-resolution. The two-day drift for T42 was estimated to be $1.8ms^{-1}$ (compared to $1.6ms^{-1}$ for the stochastic system). The top right panel of Figure 1 shows a T42 shadow orbit of the high-resolution trajectory, found using a sensitivity code (Rabier et al., 1996) which minimised the error at 48 hours. Also shown is the bound from the drift over two: as expected, the code was incapable of finding an initial condition which shadowed within this radius. Of course, this does not mean that such an orbit does not exist, but the result is consistent with the behavior expected from the drift calculation.

4 Errors relative to an analysis

Experiments which compare different models, such as weather models of different resolution, are interesting in that they reveal whether or not models are converging to a solution. Because they involve the comparison of one model with another, the cause of the errors can be attributed either to the difference in equations, or the mismatch in translating data from one model to the other. Error growth in actual weather forecasts is more complicated because the errors are measured relative not to another model but to the analysis, a smoothed version of atmospheric observations. Furthermore, the assimilation procedure used to smooth the observations often involves making them compatible with model predictions. Errors in the model can therefore affect both the analysis, and estimates of observational error obtained by comparing the analysis with the original observations. In this section, we show that estimates of model error and shadow behavior obtained using the drift are robust to such smoothing schemes for the system studied here.

To simulate the effect of the analysis procedure on forecast errors, the magnitude of the stochastic terms in the system was first increased to $\Delta F = 1.5$ and $\Delta t = 6$ so as to give errors comparable (for the same scaling as before) with operational weather models. From Eq. 2, this represents an effective increase of about 3.7 in the model drift. The observation error was increased to $2 m s^{-1}$. The up-

per panel of Figure 3 shows a plot of the resulting forecast error, when measured relative to the untreated observations.

The next step was to treat the observations using 4DVAR (Cohn, 1997; Courtier and Talagrand, 1997; Lewis and Derber, 1988). Each point of the analysis $\mathbf{u}_j = \mathbf{u}(t_j)$ was determined by minimising the cost function

$$\sum_{i=1}^{n} \left(\mathbf{s}_j(t_{j+k}) - \mathbf{y}(t_{j+k}) \right)^2 . \tag{6}$$

Here $n = 2$ represents the number of points in the 'assimilation window', which is the time period over which the model fit is optimized, and $\mathbf{s}_j(t_{j+k})$ for $t \geq t_j$ is the model trajectory initiated at time t_j on the point $\mathbf{u}_j$. Operational 4DVAR schemes usually also contain an additional term $\left(\mathbf{u}_j - \mathbf{x}^b(t_j) \right)^2$, where $\mathbf{x}^b$ is some prior estimate of the background state, so the version here is somewhat simplified.

The resulting forecast error and drift relative to the analysis is shown in the top panel of Figure 3 by the solid line. The forecast error compares with errors for the ECMWF operational model (circle symbol). The drift is still a good approximation to the error up to about two days. Note that the effect of the analysis procedure is primarily to reduce the observational component, but the square-root component due to model error remains. If model error is large, then the assimi-

12

lation procedure adjusts the observations to fit the faulty model prediction. The result is an apparent decrease in short-term errors; however the dynamical model errors continue to assert themselves over the longer term. In general, due to their cumulative nature, dynamical errors are not susceptible to smoothing procedures. Also, if model error is large compared to observational errors, the analysis errors can actually be worse than the original observational errors. Operational assimilation schemes are of course more complicated than the simple scheme presented above, and assign weights to the model and observational errors; but any scheme which places too much weight on model predictions will suffer if those predictions are affected by model error. This problem goes away if the forecast is compared directly with the observations.

The lower panel of Figure 3 shows the result of a series of shadow experiments performed with the stochastic system relative to the observations treated by the 4DVAR scheme. The median shadow time was again determined as a function of radius, and the results plotted as radius versus time. Also shown is a plot of the drift divided by two. The shadow plot is above or near the drift curve, as expected. The drift is therefore still a reliable indicator of shadow performance, despite the assimilation procedure.

In fact, it is interesting to note that the optimization procedure involved in

minimizing the cost function Eq. 6 is similar to that of finding shadow orbits. Perhaps the best evidence that a model can not shadow the observations, within a tolerance equal to the observational error, is that significant forecast errors remain after the observations are treated with 4DVAR.

5   Improving ensemble performance

Given the fact that the stochastic Lorenz system experiences large errors which affect ensemble performance, the obvious question is whether ensemble performance can be improved. One approach might be to add stochastic terms to the model. The idea would be to compensate for the 'missing' terms, and thus increase the ensemble spread.

For the Lorenz system, we can actually go further, and add exactly the same stochastic terms, so that the model is identical to the system, except for different realizations of the stochastic forcing. The effect on ensemble performance is shown in the upper panels of Figure 4. The ensemble spread is larger, as expected, but the error of the ensemble mean (not shown) is little changed. Similar behavior has been noted when stochastic perturbations are made to weather models (Buizza et al., 1999), though note that the stochastic perturbations made here are relatively

14

speaking much larger than normally would be made to an operational model.

The middle panels show the RMS forecast error for the model with stochastic terms. The tendency error between the stochastic model and system is now the difference between two stochastic terms of equal magnitude, so its expected magnitude increases by $\sqrt{2}$. This similarly affects the drift and therefore the RMS forecast error, which is about a factor of $\sqrt{2}$ higher in the middle right panel.

Despite the fact that the RMS drift has increased, some of the 1000 ensemble members shown have reduced errors in the top right panel. This might appear to contradict the relationship between expected shadow performance and the drift. However, this ensemble effectively consists of 1000 different models, each with a different stochastic forcing and consequently a different drift. Therefore increased RMS drift does not translate in this case to worsened shadow performance for an individual member; only to the expected shadow performance of a typical member. In fact, if the number of ensemble members is taken sufficiently large, one member will eventually replicate the true system over any time period.

Because the model now has the same climatology as the system, the rank histogram diagram in the lower right panel gives near-perfect statistics. Other statistical tests of course exist; however this serves to emphasise the point that getting the ensemble statistics right does not guarantee that performance is improved if

the aim is short to medium term prediction.

Another approach might be to incorporate changes in the model parameters, or even completely different models (Harrison et al., 1999; Hansen, 2002). However the model used is actually optimal in the sense that it minimises the expected drift. The only contributors to the drift, by construction, are stochastic terms which by definition can not be predicted or eliminated (weather models would also be expected to have some systematic component to the error).

This points to a fundamental difference between initial condition and model errors. The ensemble approach is well-suited to initial condition error, because the errors only occur at the initial time, have a magnitude assumed to be smaller than some limit, and exist in the space of model variables. Model error, however, is altogether different. The errors do not occur only at the initial time, but are cumulative and state-dependent. Thus it is more challenging to apply the ensemble approach to model errors (though see Barkmeijer et al., 2003).

6   Discussion - how long are the shadows?

Through a detailed study of ensemble performance in the context of the Lorenz '96 system, we have been able to establish or confirm a number of points about

ensemble behavior. The ability for the model to shadow an observed target orbit was limited by the drift, which in this case was dominated by model error. The drift could therefore be used to quantify the extent to which ensemble performance was affected by model error. The result was robust both to the presence of observational error, and the effect of smoothing techniques such as 4DVAR. Statistical tests in themselves were not enough to describe an ensemble's ability to provide useful predictions over the short to medium range. Adding random forcing terms to the equations improved the spread, but depending on one's aims, did not necessarily make the ensemble much more useful as a predictive tool.

The stochastic system in this paper is particularly difficult to model, and is intended only as a rather extreme example. Fortunately, for weather models the errors are not completely stochastic, but are the result of unparameterized physical processes. Even if it is not possible to perfectly model these processes, it is always possible to improve the parameterizations, which of course is what modelers continuously do. Techniques such as calculating the drift will not directly improve the models, but may help identify flaws.

Of course, shadowing is metric-dependent, so it is possible that a weather model could shadow a local variable such as temperature in a specific location, even though it fails to shadow in a more general metric. An ensemble will there-

fore generate a spread of temperatures, and the correct answer can be expected to lie within that range. Our point, though, is that, if a model does not shadow in a global metric, then that is because of the model drift. Therefore, while perturbing the initial condition will result in a certain spread (as will any kind of perturbation), and stand a good chance of including the correct temperature, it will not address the underlying problem. An alternative method to obtain a similar result might be to simply add error bars to the predicted temperature, where the size of the error bars is determined from error statistics.

While weather models are very different from the highly simplified stochastic system considered here, the present study may provide a framework to address similar questions in more complicated models. For example, the drift technique could be applied to errors relative to both the analysis and the untreated observations, in order to estimate the components of error due to the model equations and the observations, and to estimate shadow times. Techniques such as the sensitivity code used here for intermodel experiments could be applied to find model trajectories which shadow the analysis, and the results compared with estimates from the drift. The experiments with simple models suggest the question: how long are the shadows? It is hoped that similar experiments with weather models in the coming years will help answer this question.

Appendix

This appendix collates an overview of the drift techniques presented in (Orrell et al., 2001; Orrell, 2002, 2003, 2005); further details can be found in those references. Suppose the observed target orbit, which includes the observation error, is $\tilde{\mathbf{s}}(t)$, and the model equations are written in the form

$$\frac{d\mathbf{s}}{dt} = \mathbf{G}(\mathbf{s}(t)), \tag{A1}$$

where $\mathbf{s}$ is the state space vector. If observations are made at discrete time points $t_j = j\Delta t$, then the observed orbit is only known at these times. For convenience, we first assume that $\tilde{\mathbf{s}}(t)$ is a continuous interpolation through the data points. Alternatively, the equations can be formulated in a discrete fashion as seen below.

The forecast error

$$\mathbf{e}(t) = \mathbf{s}(t) - \tilde{\mathbf{s}}(t) \tag{A2}$$

then satisfies

$$\frac{d\mathbf{e}(t)}{dt} \approx \mathbf{J}(\tilde{\mathbf{s}}(t))\mathbf{e}(t) + \mathbf{G}_e(\tilde{\mathbf{s}}(t)), \tag{A3}$$

19

where

$$\mathbf{G}_e(\tilde{\mathbf{s}}(t)) = \mathbf{G}(\tilde{\mathbf{s}}(t)) - \frac{d\tilde{\mathbf{s}}(t)}{dt} \qquad \text{(A4)}$$

is the model's tendency error, and $\mathbf{J}$ is the Jacobian of $\mathbf{G}$. As can be seen by direct substitution, a solution to Eq. A3 is given by

$$\mathbf{d_p}(\tau) = \int_0^\tau \mathbf{M}(\tau, t)\mathbf{G}_e(\tilde{\mathbf{s}}(t))dt, \qquad \text{(A5)}$$

where $\mathbf{M}(\tau, t)$ is the model linear propagator (Strang, 1986) evaluated along the target orbit from time t to time τ. The error can therefore be approximated over the short to medium-term by

$$\mathbf{e}(\tau) \approx \mathbf{d_p}(\tau) \qquad \text{(A6)}$$

which we denote the propagated drift vector (continuous form). The $\mathbf{G}_e$ term represents the tendency error, while the linear propagator term $\mathbf{M}(\tau, t)$ reflects the amplification of the tendency error by the dynamics.

A disadvantage of the propagated drift is that it requires computation of the linear propagator, and may therefore be impractical to calculate for large models such as weather models (although its effect can be estimated by considering the

exponential magnification of the forecast errors that constitute the drift calculation as in (Orrell et al., 2001)). If the effect of the linear propagator is neglected, we obtain the *drift vector*, given in continuous form by

$$\mathbf{d}(\tau) = \int_0^\tau \mathbf{G}_e(\tilde{\mathbf{s}}(t))dt \qquad (A7)$$

For sufficiently short times τ, the drift will be approximately equal to the propagated drift, where what is meant by short depends on the particular model/system pair. However, because the action of the propagator in most chaotic models is on average to expand propagations (though this needs to be checked since contractive effects may dominate at some time scales), we can usually assume to good approximation

$$\|\mathbf{d}(\tau)\| \leq \|\mathbf{d_p}(\tau)\|. \qquad (A8)$$

The drift $d(\tau) = \|\mathbf{d}(\tau)\|$ therefore provides either an underestimate, or at shorter times an estimate, of the propagated drift.

In practice, the drift and propagated drift are evaluated in a discrete fashion at points corresponding to the observation times $t_j = j\Delta t$. Let $\mathbf{s}_j(t)$ for $t \geq t_j$ be the model trajectory initiated at time t_j on the observed point $\tilde{\mathbf{s}}(t_j)$. Using a forward difference approximation to the tendency error in Eq. A7, the drift becomes a sum

21

of short forecast errors

$$\mathbf{d}(t_K) = \sum_{j=0}^{K-1} \mathbf{f_j}, \tag{A9}$$

where $\mathbf{f_j} = \mathbf{s}_j(t_{j+1}) - \tilde{\mathbf{s}}(t_{j+1})$ (Orrell, 2002). Similarly the expression for the propagated drift in discrete form is

$$\mathbf{d_p}(t_K) = \sum_{j=0}^{K-1} \mathbf{M}(t_K, t_{j+1})\mathbf{f_j}. \tag{A10}$$

The timestep Δt should be small enough that the calculation converges; one check is to calculate the drift with a larger timestep of $2\Delta t$ and see if the result is significantly different.

References

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

Barkmeijer, J., Iversen, T., Palmer, T., 2003, Forcing singular vectors and other sensitive model structures, *Quart. J. Roy. Meterol. Soc.*, **129**, 2401-2423.

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

Buizza, R., Miller, M., and Palmer, T., 1999, Stochastic simulation of model uncertainties in the ECMWF Ensemble Prediction Scheme, *Quart. J. Roy. Meterol. Soc.*, **125**, 2887-2908.

Buizza, R., Barkmeijer, J., Palmer, T., and Richardson, D., 2000, Current status and future developments of the ECMWF Ensemble Prediction System, *Meteorol. Applic.*, **7**, 163-175.

Cohn, S., 1997, An introduction to estimation theory, *J. Meteorol. Soc. Japan, 75*, 257-288.

Courtier, R. and Talagrand, 0., 1997, Variational assimilation of meteorological

observations with the adjoint vorticity equation. Part 2: Numerical results., *Q.J.R. Meteorol. Soc.*, **113**, 1329-1347.

Ehrendorfer, M., 1997, Predicting the uncertainty of numerical weather forecasts: a review, *Meteorologische Zeitschrift*, **6**, 147-183.

Hansen, J., 2002, Accounting for model error in ensemble-based state estimation and forecasting, *Mon. Wea. Rev.*, **130**, 2373-2391.

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

Lewis, J. and Derber, J., 1985, The use of adjoint equations to solve a variational adjustment problem with advective constraints, *Tellus*, **37A**, 309-322.

Lorenz, E., 1996, Predictability - a problem partly solved, in *Predictability*, edited by T. Palmer, European Centre for Medium-Range Weather Forecasting, Shinfield Park, Reading UK.

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

Mukougawa, H., Kimoto, M., and Yoden, S., 1991, A relationship between local error growth and quasi-stationary states - case-study in the Lorenz system, *J. Atmos. Sci.*, **48**, 1231-1237.

Orrell, D., 2002, Role of the metric in forecast error growth: how chaotic is the

weather?, *Tellus*, **54A**, 350-362.

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

Orrell, D. and Smith, L., 2003, Visualizing bifurcations in the Lorenz '96 systems, *Int. J. Bifurcation and Chaos*, **13**, 3015-3027.

Orrell, D., 2003, Model error and predictability over different time scales in the Lorenz '96 systems, *J. Atmos. Sci.*, **60**, 2219–2228.

Orrell, D., 2005, Filtering chaos: A technique to estimate dynamical and observational noise in nonlinear systems, *Int. J. Bifurcation and Chaos*, (in press)

Ott, E., 1993, *Chaos in dynamical systems*, Cambridge University Press.

Palmer, T., 2000, Predicting uncertainty in forecasts of weather and climate, *Reports on Progress in Physics*, **63**, 71-116.

Rabier, F., Klinker, E., Courtier, P., and Hollingsworth, A., 1996, Sensitivity of forecast errors to initial conditions, *Q.J.R. Meteorol. Soc.*, **122**, 121-150.

Robinson, A.R., McCarthy, J.J., and Rothschild, B.J., 1999, Interdisciplinary Ocean Science is Evolving and a Systems Approach is Essential, *Journal of Marine Systems*, **22**, 231-239.

Smith, L., 1996, Accountability in ensemble prediction, in *Predictability*, edited by T. Palmer, European Centre for Medium-Range Weather Forecasting, Shinfield

Park, Reading UK.

Smith, L., 2000, Disentangling uncertainty and error: On the predictability of nonlinear systems, in *Nonlinear Dynamics and Statistics*, edited by A.I. Mees, pp. 31-64, Birkhauser, Boston.

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

Strauss, B. and Lanzinger, A., 1996, Validation of the ECMWF prediction system, in *Proc. 1995 ECMWF Seminar on Predictability*, edited by T. Palmer, European Centre for Medium-Range Weather Forecasting, Shinfield Park, Reading UK.

Toth, Z. and Kalnay, E., 1993, Ensemble forecasting at NMC: the generation of perturbations, *Bull. Am. Meteorol. Soc.*, **74**, 2317-2330.

Toth, Z., Kalnay, E., Tracton, S., Wobus, S., and lrwin, J., 1996, A synoptic evaluation of the NCEP ensemble, in *Proc. 5th Workshop on Meteorological Operating Systems*, edited by T. Palmer, European Centre for Medium-Range Weather Forecasting, Shinfield Park, Reading UK.

Wilks, D., 1995, *Statistical Methods in the Atmospheric Sciences*, Academic Press,.

Captions

Figure 1. Comparison of the stochastic system with intermodel experiments. Top left panel shows error growth, drift, and propagated drift for the stochastic system. Results are RMS over 1000 initial conditions. Lower left panel shows ensemble errors for an ensemble generated from +/- perturbations of the leading four singular vectors, optimized for two days. Right panel shows error growth in square root of total energy for T42 relative to TL159, from (Orrell et al., 2001). Also shown is a shadow orbit, and the bound on shadow radius from the drift calculation. Lower right panel shows a T42 ensemble relative to TL159.

Figure 2. Top panel shows components of drift in the stochastic system. Solid line shows total drift, dashed line is the drift component due to the stochastic forcing error, dotted line is the drift component due to the observation error, '+' symbols is the orthogonal sum of the two components. Lower panel shows a curve of shadow times for the stochastic system. Results are over 40 shadow experiments at each radius. Also shown is the curve of drift over two. The median shadow curve is always above or near the drift bound.

Figure 3. Plot of errors for the high-error stochastic system ($\Delta F = 1.5$, $\Delta t = 6$). Curve with '+' symbol shows errors relative to untreated observations. The solid line is errors relative to the observations smoothed by 4DVAR, dashed line is drift. These compare with a plot of total energy errors for the ECMWF operational model over a 15 day period in December 2000 (circle symbol). Lower panel compares a curve of median shadow times with a plot the drift over two.

Figure 4. Plot comparing 48 hour errors for the stochastic system (normal version), with and without stochastic forcing in the model. In the left panels, the model is the usual model, while in the right panels stochastic forcing is added. The model and the true system are therefore identical, except for different realizations of the stochastic forcing. Upper panels show a density plot of 1000 ensemble members with initial perturbations of 0.5 ms^{-1}; at each time point, the grayscale indicates the number of ensemble members with the error given by the vertical axis. The effect of the stochastic forcing in the model (top right panel) is to increase the spread. The RMS forecast error, measured over 1000 initial conditions, also increases by a factor of about $\sqrt{2}$ (middle panels). Lower panels show rank histogram diagrams for an 8-member ensemble. The diagram at right shows nearly perfect statistics, despite the fact that the model is not better in terms of
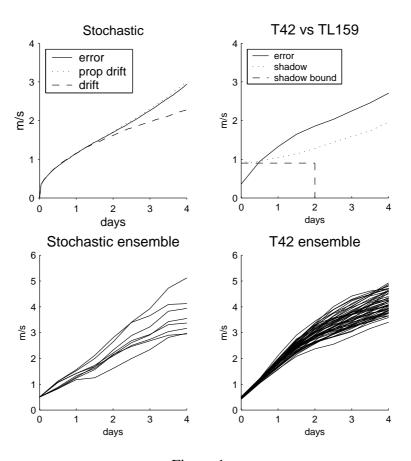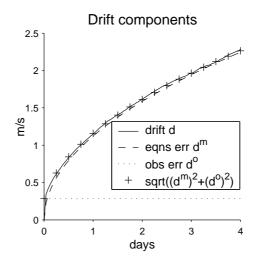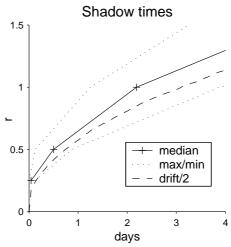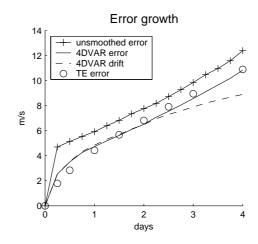
short-term errors.

Figure 1:

Figure 2:

Figure 3:

Figure 4: