Filtering chaos: A technique to estimate dynamical and observational noise in nonlinear systems

David Orrell

Institute for Systems Biology, 1441 North 34th Street, Seattle WA 98103

Abstract

Nonlinear dynamical models are frequently used to approximate and predict observed physical, biological, and economic systems. Such models will be subject to errors both in the model dynamics, and the observations of the underlying system. In order to improve models, it is necessary to understand the causes of error growth. A complication with chaotic models is that small errors may be amplified by the model dynamics. This paper proposes a technique for estimating levels of both dynamical and observational noise, based on the model drift. The method is demonstrated for a number of models, for cases with both stochastic and non-stochastic dynamical errors. The effect of smoothing or treating the observations is also considered. It is shown that use of variational smoothing techniques in the presence of dynamical model errors can lead to potentially deceptive patterns of error growth.

1. Introduction

Dynamical models of observed physical systems, be it the movements of the atmosphere or the dynamics of a living cell, will be subject to two types of noise: that caused by imperfect observations of the physical system, and that caused by imperfections in the model itself. These errors may also be amplified by the effects of chaos. In order to improve model predictions, it is necessary to separate out the levels of observational and dynamical noise, and so determine the causes of error. One approach for estimating noise levels is to begin with the observed data, and then determine a model trajectory which minimises some weighted combination of the dynamical and observational errors. This may be based for example on noise reduction techniques (Kostelich & Yorke, 1990) or a likelihood method (Ozaki et al., 2000; Heald & Stark, 2000). The noise levels can then be estimated from the optimal trajectory. This paper supplements these methods by providing a technique to estimate noise levels from the growth of root-mean-square (RMS) forecast errors, based on the model drift.

We begin in Section 2 by developing the basic theory for the case where the model experiences errors of a general form both in the model equations (dynamical noise or model error) and in the observations. The situation where model errors are stochastic or approximately stochastic, as in models of a physical system where all apparent correlations and biases in errors have been balanced out, is considered in detail. A number of examples using low and medium-dimensional systems are presented in Section 3, for both stochastic and non-stochastic dynamical errors. Section 4 examines the effect on error growth of data smoothing techniques. Finally, the results are summarised and discussed in Section 5.

2. The model drift

Suppose we have a model

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

of some physical process, and wish to compare the model predictions with an equally spaced series of observations $\mathbf{y}(t_j)$, at times $t_j = j\Delta t$, of the underlying system $\mathbf{x}(t)$ given by

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{G}\big(\mathbf{x}(t)\big) + \boldsymbol{\varepsilon}^m(t)$$
$$\mathbf{y}\big(t_j\big) = \mathbf{x}\big(t_j\big) + \boldsymbol{\varepsilon}_y^o\big(t_j\big) \tag{2}$$

Relative to the underlying system, the model is therefore subject to both dynamical (or model) errors, given by $\boldsymbol{\varepsilon}^m$, and observational errors given by $\boldsymbol{\varepsilon}_y^o$. The subscript in the observational errors $\boldsymbol{\varepsilon}_y^o$ reflects the fact that they are dependent on the choice of $\mathbf{y}$. It might seem preferable that errors be measured relative to the underlying system $\mathbf{x}$, but in a real physical system we normally do not have access to $\mathbf{x}$. (Perhaps a better word for error in this context would be discrepancy, since zero error does not mean perfect.)

Let $\mathbf{s}_j(t)$ for $t \geq t_j$ be the model trajectory initiated at time t_j on the point $\mathbf{y}(t_j)$. The drift vector is defined as a sum of short forecast errors:

$$\mathbf{d}_y(t_K) = \sum_{j=0}^{K-1}\big[\mathbf{s}_j\big(t_{j+1}\big) - \mathbf{y}\big(t_{j+1}\big)\big] \tag{3}$$

The drift magnitude gives an approximation of the total error growth at short times, and for most chaotic systems an underestimate at higher times (Orrell, 2002). Because it is generated from short forecasts, it is relatively unaffected by forecast error due to chaos. The

aim here is to show how it can be used to separate out the errors due to dynamical and observational noise. Using Eq. 1 and the $O(\Delta t^2)$ or $O(S_y^2)$ approximations

$$\mathbf{s}_j(t_{j+1}) \cong \mathbf{y}(t_j) + \Delta t \mathbf{G}(\mathbf{y}(t_j))$$
$$\mathbf{x}(t_{j+1}) \cong \mathbf{x}(t_j) + \Delta t \mathbf{G}(\mathbf{x}(t_j)) + \Delta t \boldsymbol{\varepsilon}^m(t_j)$$
$$\mathbf{G}(\mathbf{x}(t_j) + \boldsymbol{\varepsilon}_y^o(t_j)) \cong \mathbf{G}(\mathbf{x}(t_j)) + \mathbf{J}(\mathbf{x}(t_j))\boldsymbol{\varepsilon}_y^o(t_j)$$

(4)

where $\mathbf{J}$ is the Jacobian of $\mathbf{G}$, we find that

$$\mathbf{d}_y(t_K) \cong \mathbf{d}_y^m(t_K) + \mathbf{d}_y^o(t_K)$$

(5)

where

$$\mathbf{d}_y^m(t_K) = -\Delta t \sum_{j=0}^{K-1} \boldsymbol{\varepsilon}^m(t_j)$$

(6)

is the drift due to the dynamical error in the model equations, and

$$\mathbf{d}_y^o(t_K) = \Delta t \sum_{j=0}^{K-1} \mathbf{J}(\mathbf{x}(t_j))\boldsymbol{\varepsilon}_y^o(t_j) + \boldsymbol{\varepsilon}_y^o(t_0) - \boldsymbol{\varepsilon}_y^o(t_K)$$

(7)

is the drift due to the observational error.

Suppose that the observational error term $\boldsymbol{\varepsilon}_y^o$ is a vector whose components are independent identically distributed (iid) Gaussian noise processes, updated each Δt, with vector magnitude of mean 0 and standard deviation S_y. If the standard deviation of $\mathbf{J}(\mathbf{x}(t_j))\boldsymbol{\varepsilon}_y^o(t_j)$ is βS_y, then the expected RMS magnitude of the observational drift for $t \geq \Delta t$ is

$$d_y^o(t_K) = \sqrt{2S_y^2 + S_y^2 \beta^2 \Delta t \, t_K} \ .$$

(8)

Therefore

$$d_y(t_K) = \sqrt{d_y^m(t_K)^2 + S_y^2 \beta^2 \Delta t\, t_K + 2S_y^2} \qquad (9)$$

which as t_K goes to zero approaches a value, denoted $d_y(0)$, of

$$d_y(0) = S_y \sqrt{2}. \qquad (10)$$

This suggests that the observational noise level S_y can be estimated simply by plotting the RMS drift versus time, and finding the intercept with the vertical axis by interpolation. The drift due to the dynamical error can be determined from Eq. 9, and used to deduce properties of the model error as discussed in the next section.

The situation is further simplified if the dynamical term ε^m is also assumed to be a vector whose components are iid Gaussian noise processes, updated each Δt, with vector magnitude of mean 0 and standard deviation ΔF. Note that for this to be the case, at least in an approximate sense, it is not required that the model perfectly reproduce the deterministic portion of the system, only that it be balanced in such a way that correlations over time have been eliminated and the remaining errors are essentially Gaussian. From Eq. 6, the component of drift due to the dynamical error then has expected RMS magnitude

$$d_y^m(t_K) = \Delta F \sqrt{\Delta t} \sqrt{t_K}. \qquad (11)$$

Combining the separate sources as in Eq. 9, the total drift vector has an expected RMS magnitude which satisfies

$$d_y(t_K)^2 = \left(\Delta F^2 \Delta t + S_y^2 \beta^2 \Delta t\right) t_K + 2S_y^2. \qquad (12)$$

The error noise levels ΔF and S_y can be estimated by plotting the RMS drift squared versus time. The result should be an approximately straight line, with a slope of $\left(\Delta F^2 + S_y^2 \beta^2\right)\Delta t$ and an intercept $d_y^2(0) = 2S_y^2$.

The term β reflects the model's sensitivity to initial condition. It can be measured in a number of ways. One method is to calculate it directly from the Jacobian. Another, as seen below, is to perform a perfect model experiment, in which Gaussian observational error is artificially imposed on an orbit of the model, and the drift relative to this orbit is computed. Since the drift in this case is entirely due to the observational error, Eq. 8 will apply. A plot of the drift squared versus time should then have slope $S^2 \beta^2 \Delta T$. In many cases the term is small enough to be ignored. The next section illustrates use of the method for two systems due to Lorenz, and the Chua circuit.

3. Numerical examples

The technique for measuring error components using the drift is demonstrated in Figure 1 for two chaotic systems. The first is the Lorenz '63 system (Lorenz, 1963) with equations

$$\frac{dx}{dt} = -\sigma x + \sigma y$$
$$\frac{dy}{dt} = -xz + Rx - y \qquad (13)$$
$$\frac{dz}{dt} = xy - Bz$$

where the parameters are set to $\sigma=10$, $B=8/3$, and $R=28$. The second, higher-dimension system is the 8D Lorenz '96 system (Lorenz, 1996; Orrell and Smith, 2003; Orrell, 2003) for which the equations are

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F \tag{14}$$

where $i=1$ to 8 is a cyclic index, and the forcing is $F=13.2$. In either case, the model is as above, while the underlying system has stochastic dynamical noise of standard deviation ΔF added to the equations, updated each $\Delta t=0.025$ time units. The system is observed with a stochastic observational error of standard deviation S_y. The noise levels are different for the two systems.

The left panel shows the RMS forecast error and drift for each model/system pair. The total error growth after amplification by the effects of chaos is quite similar for either system, so it would be hard to deduce the levels of dynamical and observational noise from the forecast error alone. The right panel shows a plot of the drift squared, along with a linear interpolation. The difference between the two cases is now clearly visible, and the estimates of ΔF and S_y from the linear fit are close to the correct values.

The value of β was determined by a perfect model experiment. Figure 2 is a plot of RMS drift squared for $S=0.1$ and 0.2 for the Lorenz '63 and '96 models. Using Eq. 8, the slope in either case gives $\beta = 11.5$. This term has a relatively small effect on the total drift if, as here, $\beta << \dfrac{\Delta F}{S_y}$, in which case Eq. 12 reduces to

$$d_y(t_K)^2 \cong \left(\Delta F^2 \Delta t\right) t_K + 2S_y^2. \tag{15}$$

In applying this technique, the time step Δt should be sufficiently small that the $O(\Delta t^2)$ approximations in Eq. 4 are valid and the drift calculation converges (it usually suffices to be 3 or 4 times smaller than the doubling time t_d). For this reason, the method does not generalise well to maps where the variables may experience large changes between iterations. The metric should be global, and account for all major error sources (see (Orrell, 2002) for a discussion). Results here were RMS over 500 initial conditions, though a smaller number can also suffice.

Of course, the system need not be chaotic for the method to be used. As an example, Figure 3 demonstrates the technique for the Chua circuit (R. Madan, 1993; T. Matsumoto et al., 1985). The model equations are:

$$
\begin{aligned}
\frac{dx}{dt} &= c_1\left(y - x - g(x)\right) \\
\frac{dy}{dt} &= c_2\left(x - y + z\right) \\
\frac{dz}{dt} &= -c_3 y
\end{aligned}
\tag{16}
$$

where

$$
g(x) = m_1 x + \frac{m_0 - m_1}{2}\left(|x+1| - |x-1|\right).
\tag{17}
$$

The parameter values are here set to $c_1 = 15.6$, $c_2 = 1$, $m_0 = \dfrac{-8}{7}$, $m_1 = \dfrac{-5}{7}$. Two values of the parameter c_3 were used: at $c_3 = 50$ the model is periodic, while at $c_3 = 25$ the model is chaotic. The error terms were set smaller than for the Lorenz systems, corresponding to the smaller attractor dimension. The left panel shows the calculation of β, which is 30 for the periodic case and 16 for the chaotic case. The values of ΔF and S_y estimated from the slope and intercept in the right panel are in accord with the actual values.

The method can also be used when the errors are not Gaussian. Consider for example a model of the Lorenz '63 system Eqs. 13, where the model is the same as the system but has a Reynold's parameter R_m=28.28, an error of one percent. The effect is to introduce a model error of $(R_m - R)x$ in the equation for $\frac{dy}{dt}$. The resulting drift is shown in the top left panel of Figure 4, for an observational error of standard deviation $S_y = 0.2$. Because the errors are non-stochastic, the drift does not grow in a square-root fashion, but almost linearly. However, since observational errors are Gaussian, it still follows from Eq. 10 that the expected intercept is $d_y(0) = S_y\sqrt{2}$. The observational error can therefore be estimated by interpolating the drift, using an appropriate interpolation scheme. For example, a linear interpolation using the first two nonzero points gave an estimate of 0.19. The drift due to the model is then $d_y^m(t_K) \cong \sqrt{d_y(t_K)^2 - 2S_y^2 - S_y^2\beta^2\Delta t\, t_K}$, as shown by the dashed line.

Since the error is parametric rather than stochastic, it is possible to use the drift as a measure to optimize the model parameters. The top right panel shows the RMS forecast error, total drift and model drift at time t=0.2 for various values of R_m. The minimum of either curve occurs at R_m=28 which agrees with the system. However, because the model drift filters out the effects of observational error and chaos, and focuses on the component of error due to the model equations, it provides a more distinct minimum and therefore more reliable optimization (the remaining model drift after optimization is mostly due to the small error in estimating S_y). For more complex cases where none of the parameters are known, the drift technique can be combined with multivariate optimization techniques to determine the optimal parameters. The time t at which the drift is evaluated can be chosen to correspond to a desired forecast time. In principle, the one-step forecast $t=\Delta t$ could be used, however it is preferable to use a longer time since this allows the model drift to accumulate and provide a stronger signal relative to the observational component.

A similar calculation is shown in the lower panels for the Chua circuit with $c_3 = 25$, where the model now contains two types of dynamical error, parametric and stochastic. The model equations have the c_3 parameter set to $c_{3m} = 25.5$, an error of two percent. There is also a stochastic error $\Delta F = 0.5$, updated each $\Delta t = 0.01$. The observational error is $S_y = 0.2$, which can again be estimated by interpolating the drift curve. The drift at time t=0.1 (lower right panel) has a minimum at the correct value of $c_{3m} = 25$. The stochastic component of error which remains cannot be corrected by optimizing the parameters. Its magnitude can be estimated from the minimum of the model drift in the lower right panel, which is 0.0147. This corresponds to the stochastic drift at t=0.1. From Eq. 11 we obtain $\Delta F \cong 0.465$ which is close to the correct value of 0.5.

The main advantage of the drift technique is that it links noise levels to RMS error growth, and does not require treatment of the observations or an estimate of the trajectory **x**. It can therefore be easily applied to large models with many variables. In such models, nonetheless, the observational data will often be smoothed or treated before use. Because errors are measured relative to the observations **y**, the errors will change if **y** changes to some altered version **u**, and the observational errors may become correlated with the dynamical errors. In the next section, we use the drift to determine the (sometimes peculiar) effect of smoothing observations in the presence of dynamical model errors.

4. The effect of smoothing schemes

Smoothing schemes can be global over all data, or local to a particular time frame specified by an assimilation window. Beginning with the former, the essence of most

techniques based on noise reduction or likelihood methods is to minimize a weighted sum

of forecast and observational errors

$$Q(\mathbf{u}) = \sum_{j=0}^{K-1} \left[\left(\mathbf{s}_j\left(t_{j+1}\right) - \mathbf{u}\left(t_{j+1}\right) \right)^2 + w^2 \left(\mathbf{y}\left(t_{j+1}\right) - \mathbf{u}\left(t_{j+1}\right) \right)^2 \right] \qquad (18)$$

where w is a weighting factor, and $\mathbf{s}_j(t)$ for $t \geq t_j$ is the model trajectory initiated at time t_j on

the point $\mathbf{u}(t_j)$ (the dependence of $\mathbf{s}_j(t)$ on the choice of $\mathbf{u}(t_j)$ has been omitted for clarity).

Other schemes may use both the forward and backward dynamics to give an optimal fit

(Kantz and Schreiber, 1997). While the methods differ in their details, their aim is to yield a

$\mathbf{u}(t_j)$ which is close to the trajectory $\mathbf{x}(t_j)$ of the presumed, idealised underlying system

(which, if derived from observations of a physical system, is not the same thing as the true

system (Smith, 2000)).

If errors are measured relative to the smoothed trajectory $\mathbf{u}(t_j)$, then Eq. 1 becomes

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{G}\big(\mathbf{x}(t)\big) + \boldsymbol{\varepsilon}^m(t)$$
$$\mathbf{u}\big(t_j\big) = \mathbf{x}\big(t_j\big) + \boldsymbol{\varepsilon}_u^o\big(t_j\big) \qquad (19)$$

where $\boldsymbol{\varepsilon}_u^o\big(t_j\big)$ with expected RMS magnitude S_u represents the difference between $\mathbf{u}(t_j)$ and

$\mathbf{x}(t_j)$. In this system, the quantity S_u does not then correspond to errors in the observations of

the physical system, but to the artificial residual of an optimization calculation. If the

smoothing technique is working, then this residual should be small: in particular, $S_u < S_y$.

The effect on a plot of the drift squared will be to reduce the intercept, but the estimate of

dynamical errors, determined from the slope, should be relatively unaffected (and can be

compared with any estimate obtained from the smoothing procedure). The accuracy of the

calculation should also be improved, since the data is cleaner. However errors which were

originally iid Gaussian may not remain so after smoothing, which might need to be taken into account.

Local smoothing techniques, such as the 4DVAR (Cohn, 1997) variational scheme used in geophysics, present a more complicated type of data treatment. There are many different flavours of 4DVAR; we will use the term to mean the very simplest form, in which each point of the trajectory $\mathbf{u}_j = \mathbf{u}(t_j)$ is determined by minimising the cost function

$$C\left(\mathbf{u}_j\right) = \sum_{k=0}^{n} \left(\mathbf{s}_j\left(t_{j+k}\right) - \mathbf{y}\left(t_{j+k}\right)\right)^2 . \tag{20}$$

Here n represents the number of points in the local time frame specified by the assimilation window, and as before $\mathbf{s}_j(t)$ 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 $\left(\mathbf{u}_j - \mathbf{x}^b\left(t_j\right)\right)^2$, where $\mathbf{x}^b$ is some prior estimate of the background state.

As an example of 4DVAR, Figure 5 shows a zoomed view of the Lorenz '96 errors recalculated relative to a trajectory that has been treated by 4DVAR over an assimilation window of $2\Delta t = 0.05$, so $n=2$ in Eq. 20. A peculiar feature is that the intercept $d_u^2(0)$ in the right panel (line with circle symbols) is near zero. A similar effect has been noted with weather forecasts in a global metric, where the error increases in a near-exact square-root fashion over the first two days (Orrell et al., 2001), even though observational errors are known to be large. It doesn't seem plausible that 4DVAR has perfectly recovered a trajectory with $S_u = 0$; so where did all the observational error go?

To address this question, we need to consider more carefully how 4DVAR affects the drift. Referring to Eq. 5, the main contribution of non-dynamical errors to the drift $d_u(t_K)$, calculated now relative to $\mathbf{u}$, is from $\boldsymbol{\varepsilon}_u^o(t_0)$ and $\boldsymbol{\varepsilon}_u^o(t_K)$. These terms represent the

difference between $\mathbf{u}(t_j)$ and the idealised trajectory $\mathbf{x}(t_j)$, as opposed to $\mathbf{u}(t_j)$ and the observations. To obtain Eq. 12, it was assumed that they are uncorrelated with the dynamical errors, and so add in an orthogonal sense to the RMS error. This will still be true of $\boldsymbol{\varepsilon}_u^o(t_K)$, however 4DVAR will choose $\boldsymbol{\varepsilon}_u^o(t_0)$ in a direction to offset the initial errors, and so will be correlated to an extent with the dynamical errors in the assimilation window. Therefore it will not add in an orthogonal sense to the RMS drift, and the intercept $d_u^2(0)$ of the drift squared can not be used directly as an estimate for S_u. The terms $\boldsymbol{\varepsilon}_u^o(t_j)$, though, are chosen to minimize the cost function Eq. 19. If $n=1$, and the approximations in Eq. 4 are valid, their magnitude can be solved directly, and we find

$$S_u^2 \cong \frac{1}{(n+1)}S_y^2 + \frac{n}{4}\Delta t^2 \Delta F^2 \tag{21}$$

which also holds approximately when $n=2$. Because $\boldsymbol{\varepsilon}_u^o(t_0)$ is chosen to offset the dynamical errors in the initial assimilation window, the intercept of the drift squared is reduced to a first-order estimate of

$$d_u^2(0) \cong S_y^2 - \frac{1}{2}\Delta F^2 \Delta t^2 \tag{22}$$

for $n=1$, and somewhat lower for higher n.

While 4DVAR affects the estimate of S_u, it has less impact on the dynamical component of the drift. So long as $j>n$, $\boldsymbol{\varepsilon}_u^o(t_0)$ will be independent of the dynamical terms $\boldsymbol{\varepsilon}^m(t_j)$, so the slope of the drift squared remains relatively unchanged. Indeed the estimate of ΔF in Figure 5, based on a linear interpolation over points outside the initial assimilation window, is within five percent of its correct value.

One way to interpret this is that 4DVAR uses the dynamical model error terms to rebase **u** so as to reduce the short-term forecast errors relative to **u**. Since the dynamical errors remain much the same, the contribution of the drift to total error depends on the intercept $d_u^2(0)$. If the model were perfect, then from Eq. 22 the intercept would be about $d_u^2(0) = S_y^2$, which is an improvement on the unsmoothed value of $d_y^2(0) = 2S_y^2$. If the dynamical errors are large, however, the intercept can be reduced to near zero, thus reducing the total errors relative to **u** (the intercept is unlikely to become significantly negative, because this would require large adjustments to the observations, which would normally be limited by the assimilation procedure).

The key difference between the 4DVAR and untreated cases is therefore that in the former, 4DVAR has selected a reference orbit for which all errors appear as dynamical errors. Given the choice between the forecast errors in the left panel, one would say that the smoothed errors represent a huge improvement. However the 4DVAR version of **u**, with $S_u=0.39$, is a relatively small improvement on the untreated observations **y** with $S_y =0.5$. Thus **u** is only much better than the observations as a representation of truth if model errors are small compared to observational errors. This points to an interesting contradiction about the simplified 4DVAR scheme phrased above: it assumes that the model is perfect, and adjusts the observations to fit model predictions; achieves the greatest reduction in error S_u if model error is small; yet yields the best *apparent* improvement in forecast accuracy when model error is large.

5. Conclusions

The above results show that, by filtering the effects of chaos, the model drift makes it possible to separate out the effects of dynamical and observational errors. It can be applied not only to small models such as those considered here, but in principle at least to large models such as those used in geophysics, biological oceanography (Robinson and Lermusiaux, 2002), or economics. It is of course necessary that observations be sufficiently frequent, however there is no need for the timestep Δt to approach zero. It is only required that the approximations in Eq. 4 be valid, which can be tested on a case-by-case basis. The drift calculation works not as a result of taking vanishingly small time steps, but by integrating the short forecast errors over time, so that the dynamical model errors accumulate relative to the observational term. The method can be applied to situations where the errors are both stochastic or non-stochastic, and used as a criterion for optimization of model parameters.

A common feature in models of observed physical systems is the use of data smoothing techniques, such as 4DVAR, which must be taken into account when interpreting the results as shown in Section 4. Of course, operational assimilation schemes vary significantly, and do not necessarily assume that the model is perfect. However, two things are shown. Firstly, 4DVAR techniques must be used with caution if there exist large dynamical errors, for which square-root growth of the kind seen in Figure 5 may be an indicator. Secondly, the drift technique, when used to measure dynamical errors, is quite robust to data manipulation, including that from noise reduction or variational methods. In general, because dynamical errors act in a cumulative manner, they are resistant to attempts at smoothing or masking. Indeed, perhaps the strongest evidence of dynamical errors is when procedures such as 4DVAR are unable to reduce them.

Error growth in nonlinear systems is a complex convolution of dynamical errors, observational errors, and magnification by the nonlinear dynamics. Its interpretation is further complicated by the smoothing or treating of observations. By effectively filtering out the effects of chaos, however, the drift calculation provides a simple method to estimate the levels of dynamical and (treated or untreated) observational noise. Applied to physical, biological, or economical models, it has the potential to provide a check on whether the total error growth is primarily due to the accumulation of dynamical model errors, or the effects of chaos. This is especially useful in the modeling of large, complex systems, where dynamical model errors can be expected to be significant.

References

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

J.P.M. Heald and J. Stark. [2000] "Estimations of noise levels for models of chaotic dynamical systems," *Phys. Rev. Lett.*, **84**, 2366-2369.

H. Kantz and T. Schreiber. [1997] "*Nonlinear Time Series Analysis*," Cambridge University Press, Cambridge, 1997.

E.J. Kostelich and J.A. Yorke. [1990] "Noise reduction: finding the simplest dynamical system consistent with the data," *Physica*, **41D**, 421.

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

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

R. Madan. [1993] "Chua's Circuit: A Paradigm for Chaos," World Scientific, Singapore.

T. Matsumoto, L.O. Chua, M. Komuro. [1985] "The double scroll," *IEEE Trans. Circuits Syst.*, **23**, 798-817.

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

D. Orrell. [2002] "Role of the metric in forecast error growth: how chaotic is the weather?," *Tellus*, **54A**, 350-362.

D. Orrell. [2003] "Predictability over different timsescales in the Lorenz '96 systems," *J. Atmos. Sci.*, **60**, 2219-2228.

D. Orrell and L.A. Smith. [2003] "Visualising bifurcations in high dimensional systems: The spectral bifurcation diagram," *Int. J. Bifurcations and Chaos*, **13**, 2003.

T. Ozaki, J.C. Jimenez, and V. Haggan-Ozaki. [2000] "The role of the likelihood function in the estimation of chaos models," *J. Time Ser. Anal.*, **21**, 363-387.

A.R. Robinson and P.F.J. Lermusiaux. [2002] "Data Assimilation for Modeling and Predicting Coupled Physical-Biological Interactions in the Sea," Chapter 12 of *The Sea*, **12**, edited by Allan R. Robinson, James J. McCarthy and Brian J. Rothschild, John Wiley & Sons, Inc., New York.

L.A. Smith. [2000] "Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems," in *Nonlinear Dynamics and Statistics*, edited by A. I. Mees, Birkhauser, Boston, 31-64.
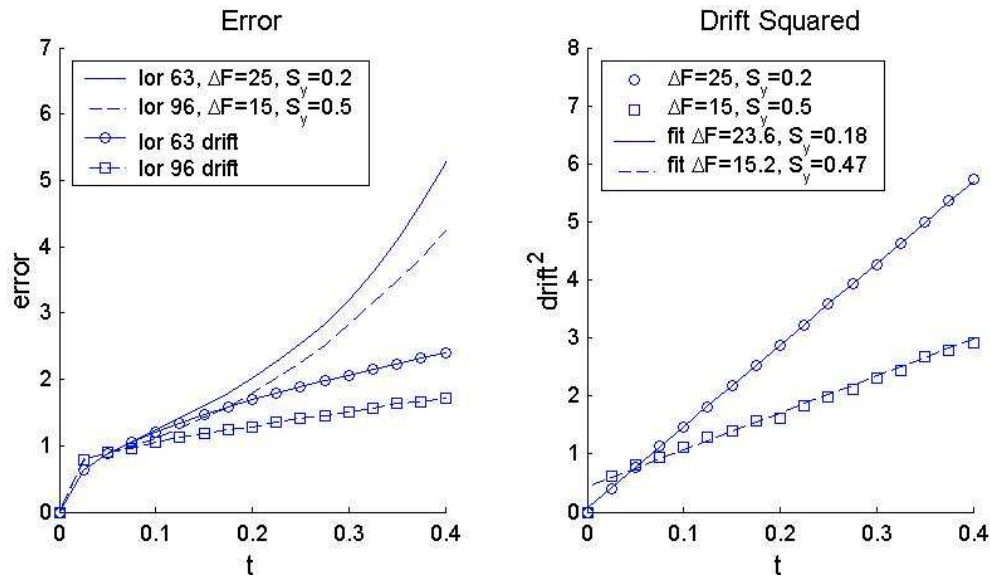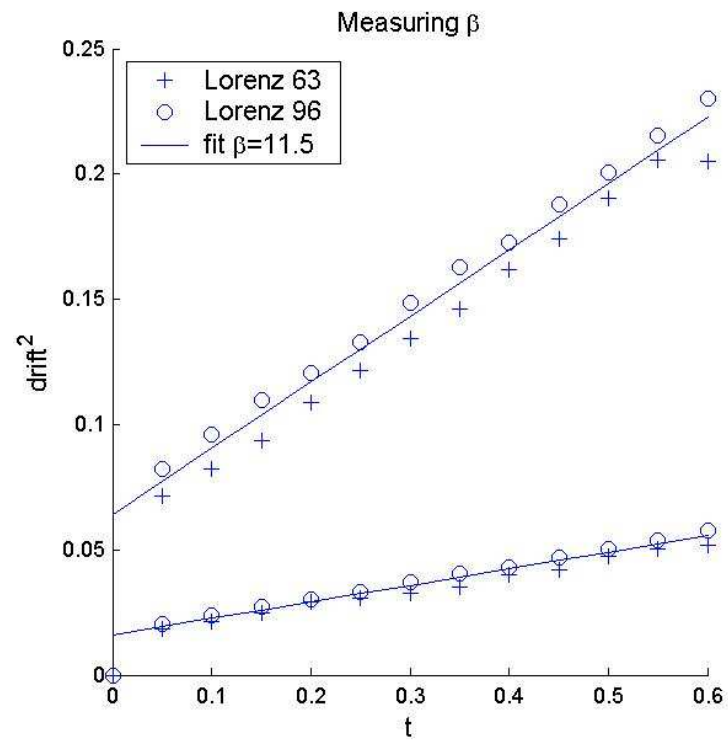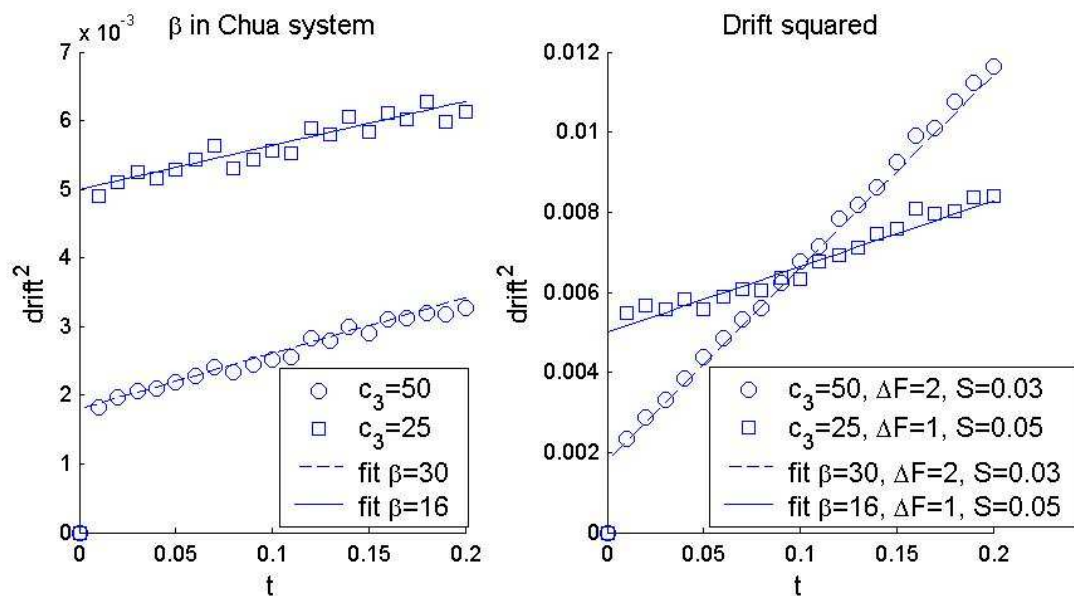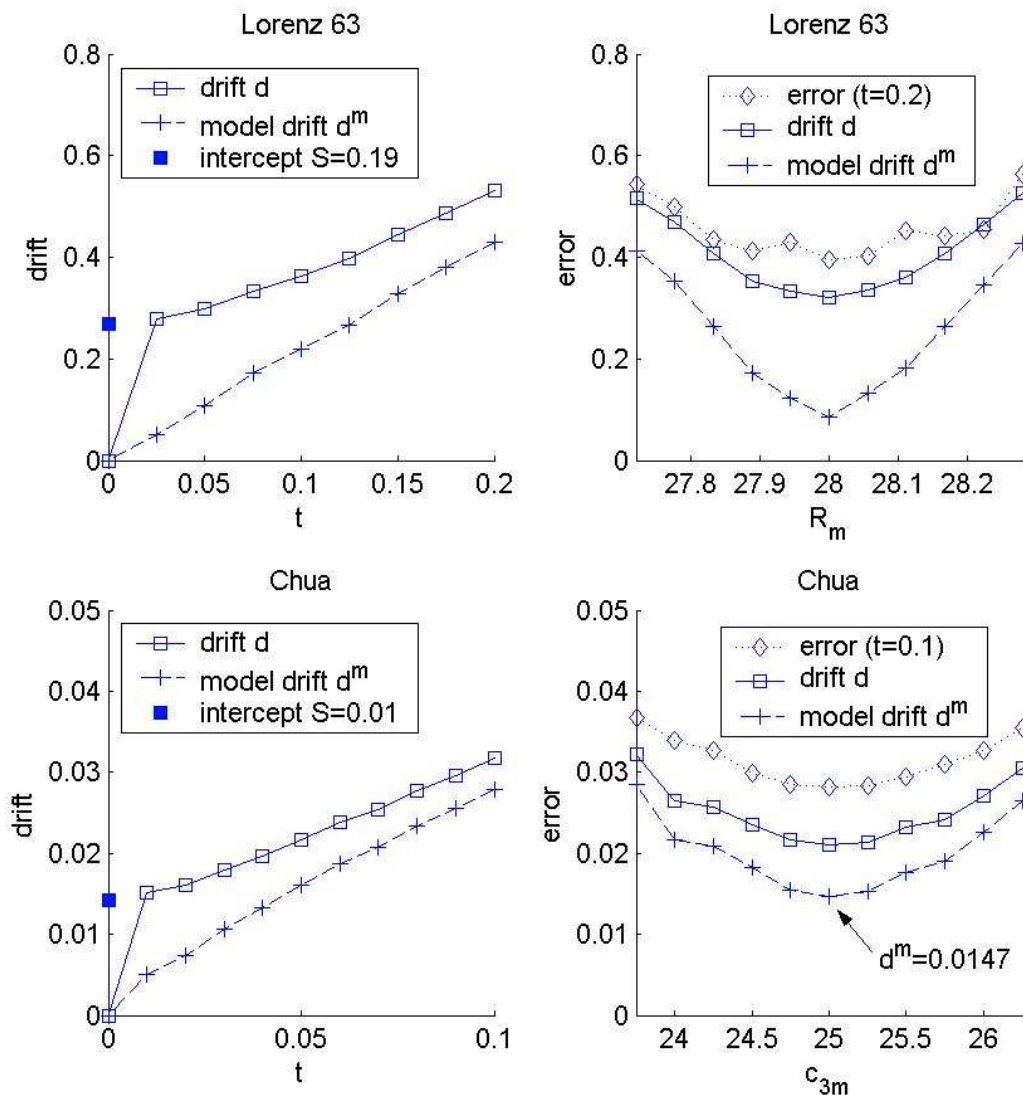
Figures

Figure 1. Left panel shows forecast errors and drift for two cases: the Lorenz '63 system with *ΔF=25*, *S$_y$=0.2*, and the Lorenz '96 system with *ΔF=15*, *S$_y$=0.5*. Observations are each *Δt=0.025*; a smaller time step of *0.005* was used for the Runge-Kutta integration scheme. Error growth in either case is similar. Right panel shows the drift squared: the difference between the two cases is now clearly visible. The linear fit gives an estimate for the Lorenz '63 case of *ΔF=23.6*, *S$_y$=0.18*, and for the Lorenz '96 system *ΔF=15.2*, *S$_y$=0.47*, which are close to the correct values. Results are RMS over 500 initial conditions.

Figure 2. Plot of drift squared for S_y=0.1 and 0.2 for the Lorenz '63 and '96 models. Model and system are the same, so all error is due to the Gaussian observational noise. The slope can be used to estimate β, and in either case gives β =11.5. Results are RMS over 500 initial conditions.

Figure 3. Left panel shows measurement of β in the Chua circuit, for two values of the parameter c_3. At $c_3=50$ the circuit has a periodic attractor, while at $c_3=25$ the circuit is chaotic. Perturbation size is $S_y=0.03$ for the former case and $S_y=0.05$ for the latter. The values of β determined from the slope are 30 and 16 respectively. Right panel shows the drift squared for the cases shown. Observations are each $\Delta t=0.01$; a smaller time step of *0.001* was used for the Runge-Kutta integration scheme. Predicted errors are in agreement with the actual terms. Results are RMS over 500 initial conditions.

Figure 4. Top left panel shows plot of drift in the Lorenz '63 system, due to a one percent error in the Reynold's parameter *R*, and an observational error of S_y=0.2. Observations are each *Δt*=0.025. The expected *y*-intercept is shown by the solid square symbol. Its value was estimated by a linear interpolation from the first two points, giving an estimate of 0.19. The drift due to the model only, shown by the dashed line, can then be estimated by subtracting off the observational component, as discussed in the text. Top right panel shows forecast error and drift at time 0.2 for various values of R_m; the optimum value is 28 which is the same as the system. The model drift provides the most distinct minimum of the three curves. Lower panels show same calculation for the Chua system with c_3=25, where the model error includes both a stochastic component and a component due to a two percent error in the parameter c_3. Stochastic component of drift can be estimated from the minimum model drift. Results are RMS over 500 initial conditions.
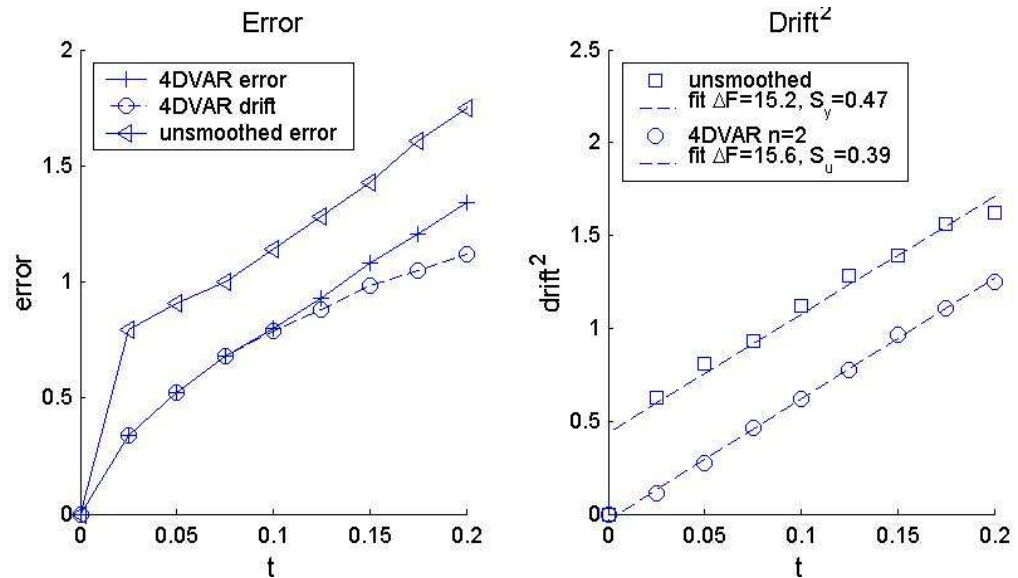
Figure 5. Left panel shows zoomed view of forecast errors for the Lorenz '96 system, with *ΔF=15*, *S_y=0.5*, and *ΔT=0.025*, for the untreated and 4DVAR case with assimilation window *n=2*. The 4DVAR drift (circles) takes on a square-root shape. Right panel shows the drift squared. The linear fit gives an estimate for the 4DVAR case of *ΔF=15.6*. The value *S_u=0.39* was measured from the actual variance; the exact value has little impact on the estimate of *ΔF*. Results are RMS over 500 initial conditions.