Tellus (1994), 46A, 651–670
Printed in Belgium – all rights reserved

TELLUS
ISSN 0280–6495

Dynamics of information and uncertainty

By W. C. THACKER*[1] and R. LEWANDOWICZ[2], [1]*Atlantic Oceanographic and Meteorological Laboratory, 4301 Rickenbacker Causeway, Miami, FL 33149, USA*; [2]*Cooperative Institute for Marine and Atmospheric Studies, 4600 Rickenbacker Causeway, Miami, FL 33149, USA*

(Manuscript received 23 September 1993; in final form 23 December 1993)

ABSTRACT

The question of what information might be inferred from a set of data is posed in terms of the uncertainties of model variables determined by a least-squares fit. When a dynamical model is fitted to asynoptic data, the uncertainty can be characterized by a region in the model's phase space surrounding the point associated with the best fit. Changes in the shape and orientation of this region as it evolves indicate how information is redistributed dynamically among the model variables, much as kinetic and potential energy might be redistributed. These ideas are illustrated within the context of single and double oscillator systems. Information about the state of a shallow-water model is shown to depend sensitively on the sampling interval of fictitious altimetric data.

1. Introduction

When computing least-square fits of a linear shallow-water model to simulated satellite altimetric data by adjusting initial conditions, we found remarkable sensitivity to the temporal distribution of the data. Investigating further, we examined the covariance matrix of the errors of the inferred initial state and generally found a broad spectrum of eigenvalues, the smallest corresponding to those aspects of the initial conditions best-determined by the data and the largest corresponding to the worst-determined aspects. We wanted to know what aspects were poorly determined and why.

The errors of the inferred initial state are a consequence of the assumed distribution of errors in the data and of the hypothesis that the shallow-water model with appropriate initial conditions can explain the data. Moreover, because the model is linear, the errors do not depend upon the values of the data. Consequently, it is possible for the errors in the inferred initial state to be small when the model fails to explain the data. Such an inconsistency invalidates the hypothesis of the model's appropriateness. Our data are generated by the model, so we know it is capable of fitting the data. What we don't know is whether the simulated data are sufficient to determine the initial model state completely.

Fitting a straight line to a scatter plot provides a useful analogy (Thacker, 1992): the slope and intercept of the line are like the initial conditions of the model, and posterior estimates of their errors are like those of the inferred initial state. However, the initial state is described by many more parameters than the two required to specify the straight line, and the relationship between the data and the initial conditions is far from intuitive. It is possible that, even when the model explains what appear to be well-sampled data, those data might provide little or no information about some aspects of the initial state. The disparity of the extreme eigenvalues suggests that this is indeed the case.

To proceed, we found it best to restrict our attention to a simple rectangular model requiring only 341 variables to characterize the model's three initial fields (using a C-grid with 11×11 points for the height and 10×11 and 11×10 for the two velocity components) and to examine various configurations of artificial data. Our

* Corresponding author.

model ocean was bounded in latitude by 30°N and 30°S and in longitude by 70°W and the Greenwich meridian, characterizing the tropical Atlantic as a basin with impenetrable zonal and meridional walls. For simplicity, we presumed it would be possible to observe the entire surface at several equally spaced times and asked how the results might differ as the interval between observations were changed. Would sampling every 17 days, for example, provide fairly good information about the model's slower waves but little about the faster waves? Or, conversely, would sampling at every three hours for half a day determine the model's fastest waves much better than its slow waves?

The model's leapfrog time-stepping scheme, with its start-up time step and its need for a device such as a Robert filter for controlling its time-splitting instability, complicates the identification of fast and slow waves and their uncertainties. As indirect indicators of the two branches of the wave spectrum, we chose the model's initial divergence and vorticity fields. Using the error-covariance matrix of the best-fit initial conditions, we computed the error-covariances of the initial divergence and vorticity fields. Fig. 1 shows contours of variance of errors in the initial divergence and vorticity fields inferred from five complete fields of artificial altimetric data spaced 17 days apart. The uncertainties in these two fields, like those of all model variables inferred from the data, stem from the assumed accuracy of the altimetric measurement of the surface elevation and the assumed relationship between the sea-surface elevation and the model's thermocline depth. We have taken the altimetric determination of the surface height to be ± 5 cm, and have assumed a ratio of 200 for thermocline depth to surface elevation (Rebert et al., 1985), but variances can easily be rescaled to accommodate different values. Although the distribution of uncertainty in the divergence and vorticity fields is interesting and not at all obvious, the intent of the figure is only to illustrate that the relative magnitude of the uncertainties is similar for the 2 fields: in regions where they are best determined, both fields are estimated to within $\pm 3 \times 10^{-6}$ sec^{-1}. 17-day sampling does not seem to favor slow waves at the expense of fast waves.

To be somewhat more systematic, we repeated the computation of uncertainties as a function of the sampling interval. In every case there were three fields of hypothetical altimetric data, with the first observations corresponding to the initial time, the second N time steps later, and the third after another N time steps. We computed the error-

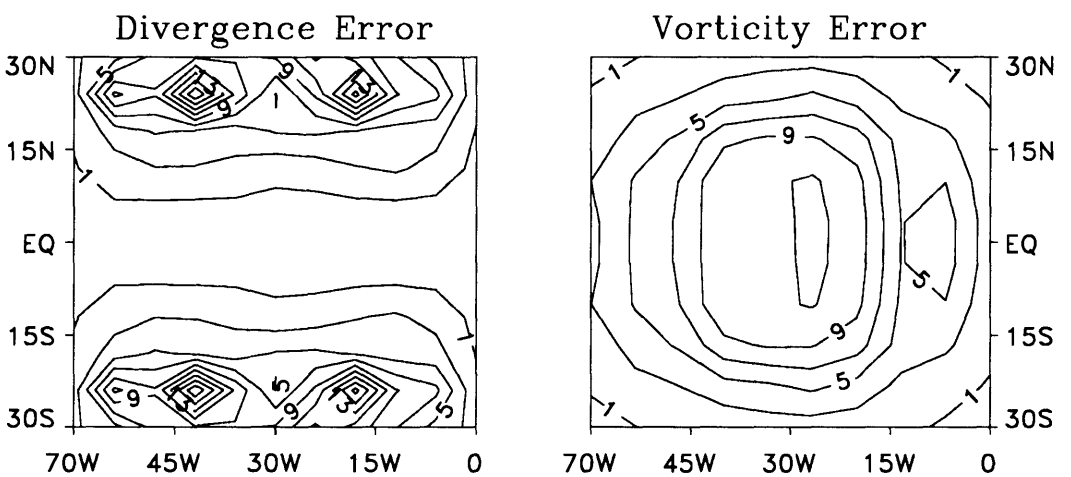

Fig. 1. Variance of errors for best-fit divergence and vorticity fields determined from five fields of hypothetical altimetric data spaced 17 days apart. Units of variance for both fields is 10^{-11} sec^{-2}, which is a consequence of errors in the hypothetical altimeter data being uncorrelated and uniform with assumed variance of 25 cm^2 and of an assumed ratio of 200 for the displacement of the thermocline to that of the surface.

variances of the vorticity and divergence fields for 300 cases (N ranging from 1 to 300), and summarized these fields of variances as averages. Fig. 2 shows the average uncertainties of divergence and vorticity as functions of the sampling interval N. From these results we can conclude that the sampling frequency does not seem to have a simple relationship to the distribution of uncertainty over the wave components of the flow.

The error-covariance matrix of the optimal initial state is the inverse of the Hessian matrix of the quadratic function defining the fit of model to data. Computing the fit, in effect, amounts to inverting the Hessian, so the question of what information the data provide is intimately related to the question of how difficult it is to compute the fit (Thacker, 1989). One measure of computational difficulty is the condition number, i.e., the ratio of the largest to smallest eigenvalues of the Hessian λ_L/λ_S. (Since the 2 matrices share eigenvectors and their eigenvalues are reciprocally related, the condition number is also the ratio of the smallest to largest eigenvalues of the error-covariance matrix.) We found that the condition number was sensitive to the sampling interval. Callies and Eppel (1994) found similar results, also within the context of the shallow-water equations.

The top panel of Fig. 3 shows the condition number as a function of sampling interval for the same sequence of computations giving rise to Fig. 2. Note that the condition number can change by an order of magnitude when the sampling interval is increased by a single time step. The extreme eigenvalues are displayed in the 2nd and 3rd panels of Fig. 3. The trend of eigenvalues decreasing with sampling interval is a consequence of the model's dissipation. When these computations were repeated with less friction, the trend was much less apparent. The general increase in the condition number with sampling period is also due to dissipation via the effect of friction being greater on the smallest eigenvalue. (The tendency toward increasing uncertainty in divergence and vorticity in Fig. 2 is another manifestation of dissipation.) But these trends are not as puzzling as the fluctuations. They indicate that the uncertainty in the best-determined and worst-determined aspects of the initial state are sensitive functions of the sampling period, which suggests that the partition of information among the 341 variables characterizing the model's initial state changes significantly with sampling period. The fourth panel shows the cosine of the angle Θ_S between successive (as sampling increases) eigenvectors associated with the smallest eigenvalues, while the bottom pattern shows the cosine of the angle Θ_L between successive eigenvectors with largest eigenvalues. Clearly, the worst-determined aspect of the initial model state fluctuates much more rapidly than the best-determined aspect. But why does a small change

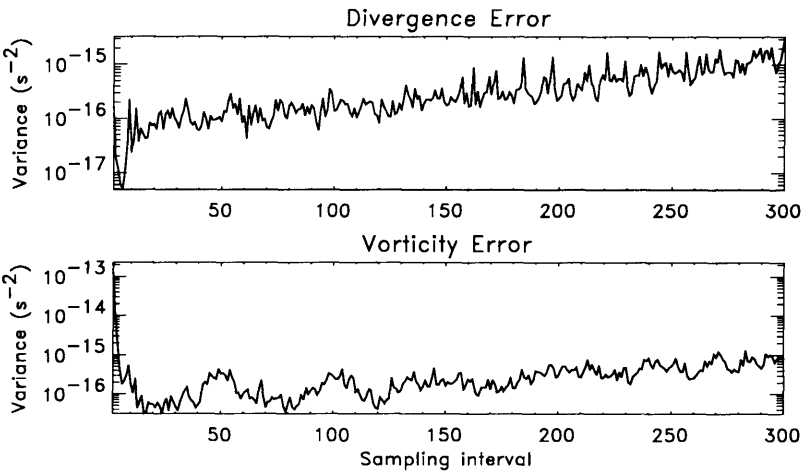

Fig. 2. Average variance of errors for best-fit divergence and vorticity fields determined from three fields of hypothetical altimetric data spaced N time steps apart, where N is the sampling interval and the size of the time step is 3 h.

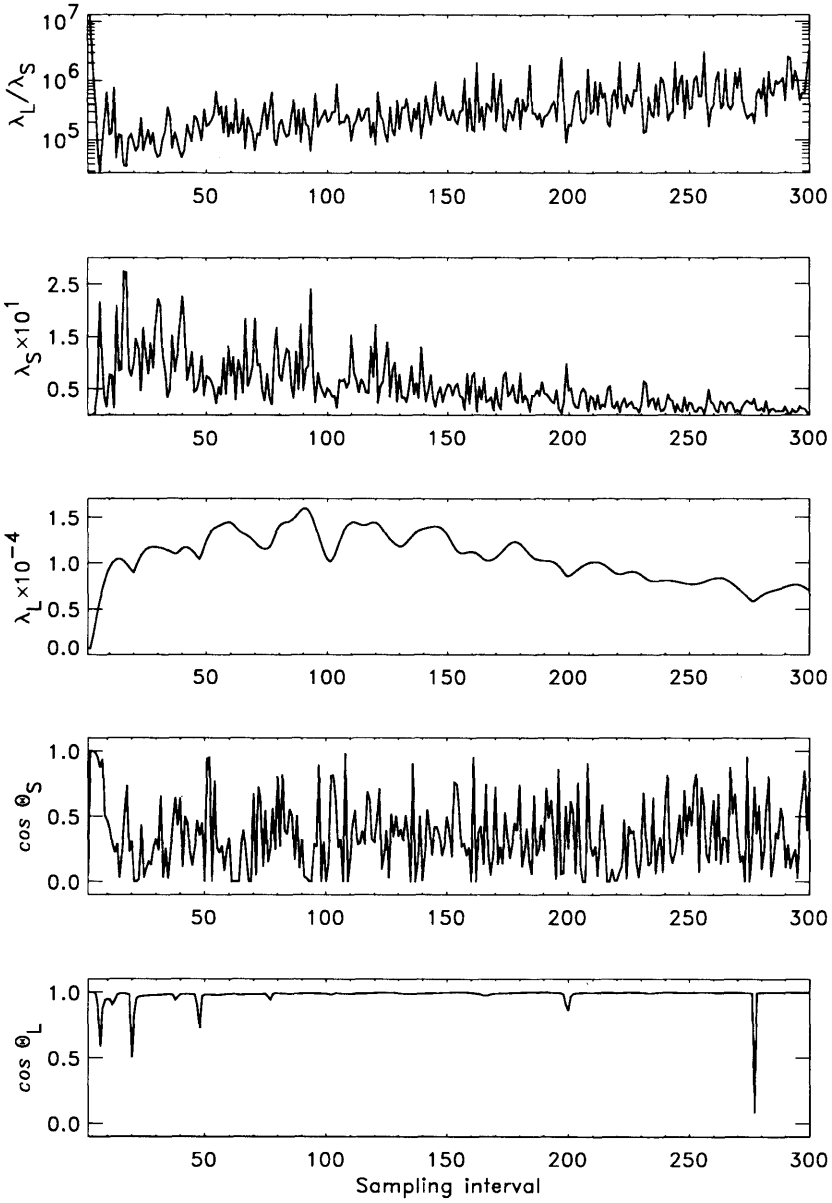

Fig. 3. Condition number, extreme eigenvalues of Hessian, and cosines of angles between successive extreme eigenvectors as functions of sampling interval for three fields of hypothetical altimetric data separated by N computational time steps.

in the sampling interval have such a strong effect on the distribution of information that can be extracted from three fields of altimetric data? And why should it affect the difficulty of the computations so strongly?

The purpose of this paper is to pursue these questions by addressing the question of how information from observations made at one time manifests itself at other times. As information is reciprocally related to uncertainty, the question can be put: how does the uncertainty of the state of a dynamical system evolve? Even though the

linear, shallow-water equations are quite simple, we thought that if we were to understand the sensitivity to sampling time, we should work with a still simpler model. In terms of normal modes, the shallow-water model could be viewed mathematically as a system of oscillators. So we first turned to the double oscillator, because it was the simplest dynamical system with more than one frequency, and we learned that the relationship between sampling frequency and the best-determined aspects of the initial state is far from simple. It was at this point that we realized the answer involved understanding the dynamics of information and uncertainty. We then turned to an even simpler system, the harmonic oscillator, as that provided the possibility of visualizing the evolution of uncertainty and the combination of information from measurements at different times. Below we present our analysis of the single oscillator first, and then discuss the double oscillator.

2. The single oscillator

The focus of this paper is on the question of how measurements made at different times contribute to the specification of the state of a dynamical system at a single time. More specifically, it addresses the way in which observational uncertainties are manifest as uncertainties of the dynamically evolving state. There is no loss of generality in restricting the discussion to a dynamical system with only a few degrees of freedom. Because the simple harmonic oscillator is a well-known textbook example, it provides an ideal example for illustrating the concept of dynamically evolving uncertainty.

The state of the harmonic oscillator is described by the values of two variables, v and x, denoting velocity and displacement from equilibrium position of a unit mass connected to a wall by a spring. These variables evolve in time according to differential equations expressing Newton's second law and the definition of velocity:

$$\frac{dv}{dt} = -\omega^2 x$$
$$\frac{dx}{dt} = v, \qquad (1)$$

where ω^2 is the stiffness of the spring and ω is the frequency of the oscillator. The solution at time t depends on x_0 and v_0, the displacement and velocity at the initial time $t_0 = 0$:

$$x(t) = x_0 \cos \omega t + \frac{v_0}{\omega} \sin \omega t$$
$$v(t) = v_0 \cos \omega t - x_0 \omega \sin \omega t. \qquad (2)$$

Fig. 4 shows the phase-space trajectory for initial conditions $v_0 = 0$ and $x_0 > 0$. Starting on the positive x-axis, the trajectory traces out an ellipse in the clockwise direction with frequency ω. Note that the aspect ratio of the ellipse is a function of the frequency, which depends on the choice of units.

Now suppose the initial conditions were not known precisely, but only that the system was near x_0 with a velocity close to zero at time t_0. The plausible initial states could be represented by a region of phase space, each point of which traces out its own elliptical orbit. Fig. 5a illustrates how an initially circular region evolves. At any instant the region is elliptical. The orientation and aspect ratio vary in time, but the area is constant (i.e., information is conserved). Note how the uncertainty in position is much greater after a quarter cycle than it was initially. This is because knowledge of position has been transformed into knowledge of velocity and vice versa. Fig. 5b, which is the same as 5a, but with units chosen so the phase-space orbits are circular, makes this clear: the elliptical region of plausible states rotates as the mass oscillates. Fig. 5c shows a circular region that has evolved from an initially elliptical

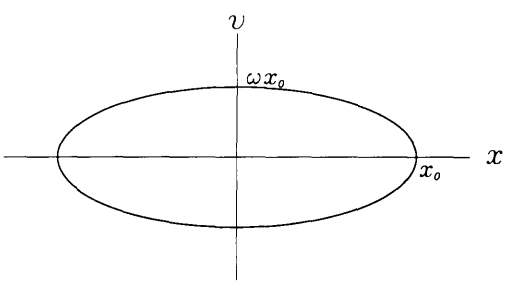

Fig. 4. The trajectory of the state of the harmonic oscillator in phase space is an ellipse. The eccentricity of the ellipse is determined by the frequency of the oscillator, which in turn depends on the choice of units for displacement and velocity.

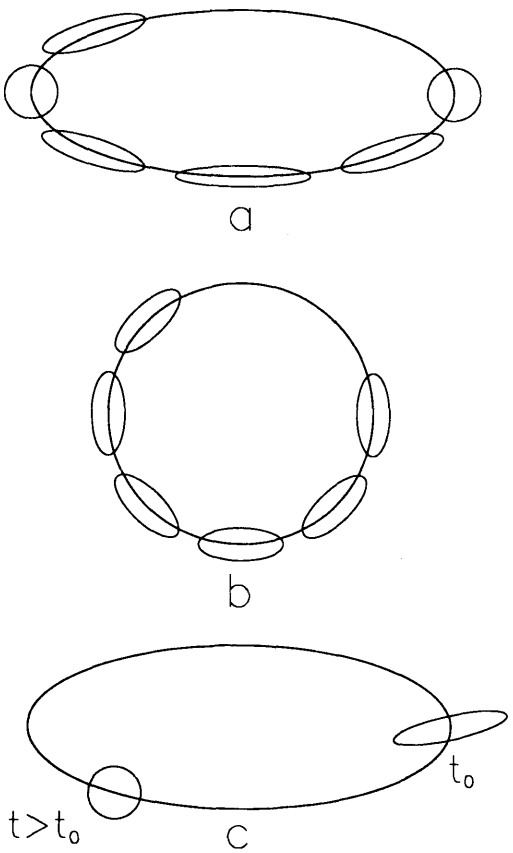

Fig. 5. Dynamics cause information to be redistributed among the state variables as the system evolves. (a) Units of displacement and velocity are chosen to be the respective observational uncertainties, and, for a spring with frequency less than one in these units, equal uncertainty of the two state variables starting from maximum extension results in increased (decreased) accuracy of the velocity (displacement) estimate at the time of maximum velocity. (b) The same with units chosen so the trajectory is circular and the dynamic operator is skew-symmetric. (c) An estimate of the state variables at $t > t_0$ having equal and uncorrelated errors (circular contour) corresponds to initial state variables with unequal and correlated errors (elliptical contour).

region, or, equivalently, a region of plausible initial conditions consistent with measurements made at a later time. The forward evolution of uncertainty is essential to Kalman filtering, while backward evolution is characteristic of least-square fits obtained by adjusting initial conditions.

The best-fit initial conditions, x_{best} and v_{best}, are determined by finding the minimum of a function $J(x_0, v_0)$, measuring model-data misfit in the least-squares sense (see, for example, Le Dimet and Talagrand, 1986, or Thacker and Long, 1988):

$$J = \frac{1}{2} \sum_i \left(\frac{m_i - d_i}{\sigma_i} \right)^2. \tag{3}$$

If the model counterparts m_i of the data d_i (with standard deviations σ_i) are linearly related to x_0 and v_0, which is the case for measurements of position and velocity at any time t, then J is a quadratic function of x_0 and v_0. In matrix notation the quadratic function can be written:

$$J = \tfrac{1}{2}(\xi - \xi_{\text{best}})^{\mathrm{T}} H (\xi - \xi_{\text{best}}) + J_{\min}. \tag{4}$$

where $\xi^{\mathrm{T}} = (x_0, v_0)$ and H is the Hessian matrix of second derivatives of J.

How good are the estimates x_{best} and v_{best}? The objective function J can be thought of as being the logarithm of a likelihood function. As the data are modeled by linear functions of the initial conditions, the likelihood function, when normalized, takes the form of a bivariate normal probability-density function, $p(x_0, v_0)$:

$$p = \frac{1}{2\pi \sqrt{\det \Sigma_{\text{best}}}} \exp((\xi - \xi_{\text{best}})^{\mathrm{T}} \times H(\xi - \xi_{\text{best}})), \tag{5}$$

where the covariance matrix Σ_{best} is the inverse of H and where $\det \Sigma_{\text{best}}$ is its determinant. Contours of p are ellipses in the (x_0, v_0)-plane centered on the optimal initial conditions. The larger the ellipse, the greater the uncertainty in the estimate of the initial state. The contour corresponding to values of (x_0, v_0) for which the argument of the exponential is one-half, i.e., for which $J - J_{\min} = \frac{1}{2}$, is the standard-deviation ellipse (Fig. 6). Its size, shape, and orientation are determined by the types of observations used in determining the fit, by the uncertainties in these observations, by the number of observations, and by the observation times relative to t_0. If the principal axes of the standard-deviation ellipse were aligned with the (x_0, v_0) axes, errors in the best-fit estimates of x_0 and v_0 would be uncorrelated, and its eccentricity would reflect the relative accuracy of x_{best} and v_{best} *as expressed in some chosen units.* If the units were

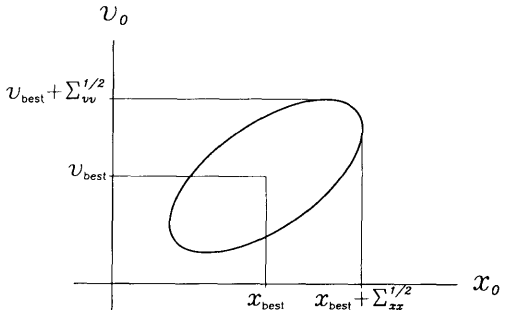

Fig. 6. The uncertainty of the state of the harmonic oscillator can be characterized by the standard-deviation contour of the normal probability density. The best-fit value lies at the center of the ellipse, and the variance of error in the best-fit estimate of the displacement (velocity) is the square of the distance between the center and the line tangent to the ellipse and perpendicular to the x-axis (v-axis). If error-covariances were zero, the ellipse would be a circle. The short (long) axis of the ellipse corresponds to the linear combinations of x and v that has the least (most) uncertainty.

chosen to be the standard deviations of the observational errors, the ellipse would be a circle.

Fig. 5 illustrates the evolution of the collective uncertainty in the state of the harmonic oscillator resulting from some unspecified measurements, but it says nothing about the information contributed by individual measurements d_i via the least-squares fit. How do the data determine the shape of the initial region of uncertainty? Two distinct concepts are involved: (1) the propagation of information (or, equivalently, uncertainty) from the various measurements to a common time and (2) the combination of information from the measurements into a best estimate for the initial state. Both are complicated by the fact that a single measurement can determine only one aspect of the dynamic state; e.g., measuring the position of the mass provides no information about its velocity. We have already discussed the propagation of uncertainty of the complete state, and it is simple enough to visualize incomplete information as corresponding to an infinitely elongated ellipse, so we turn now to the question of combining information.

Estimating the state vector using several independent estimates of the entire state is analogous to estimating a single random variable from several independent measurements. In the single-variable case, the estimate x_{best} is an average of the measurements weighted according to their information content, i.e., inversely proportional to their error variance:

$$x_{best} = \sigma_{best}^2 \left(\frac{\hat{x}_1}{\hat{\sigma}_1^2} + \cdots + \frac{\hat{x}_n}{\hat{\sigma}_n^2} \right), \tag{6}$$

and the variances of the measurements combine like resisters in parallel to give the variance of the estimate:

$$\frac{1}{\sigma_{best}^2} = \frac{1}{\hat{\sigma}_1^2} + \cdots + \frac{1}{\hat{\sigma}_n^2}. \tag{7}$$

In the multivariate case, the estimate is again a weighted average of the vector measurements, where the weights are matrices:

$$\xi_{best} = \Sigma_{best}(\hat{\Sigma}_1^{-1}\hat{\xi}_1 + \cdots + \hat{\Sigma}_n^{-1}\hat{\xi}_n), \tag{8}$$

and the error-covariance matrices again combine like resisters in parallel:

$$\Sigma_{best}^{-1} = \hat{\Sigma}_1^{-1} + \cdots + \hat{\Sigma}_n^{-1}, \tag{9}$$

Fig. 7 illustrates how the information in two pairs of measurements combine. The first panel shows a circular confidence region of the first pair (compare with Fig. 5a), an elliptical confidence

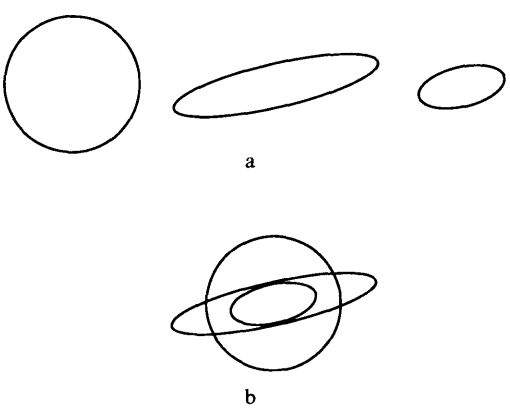

Fig. 7. Two estimates of the model state combine in such a way that the uncertainty of the combination is smaller than that of either estimate. In panel (a) the right-most ellipse represents the uncertainty based on the combined information characterizing the initial states of Figs. 5a, c. In (b) centers are superposed, showing the confidence interval of the combination lies within that of each estimate.

region for the second measurement pair transformed to the initial time (compare with Fig. 5c), and a smaller elliptical confidence region for the estimate based on both pairs. The second panel shows the ellipses with their centers superimposed to illustrate that the confidence region of the combined estimate falls within the intersection of the individual confidence regions.

As we mentioned above, we can't expect all variables constituting the state of a dynamical system to be measured simultaneously. For example, we might only have equipment for measuring the position of the oscillating mass but nothing for measuring its instantaneous velocity. If we have only one measurement of position, the state is not fully determined. The confidence ellipse becomes an infinite region parallel to the v-axis. Fig. 8a shows such an infinite region for a measurement at $t > t_0$ and the region from which it evolved. Fig. 8b shows the confidence region for initial conditions resulting from two measurements of position, one corresponding to that of the top panel and the other made at t_0 rather than t. (Note that, even though the covariance matrices corresponding to these infinite regions are infinite, their inverses are finite, and the sum of the inverses is not singular, so the covariance matrix for the estimated initial conditions is finite.)

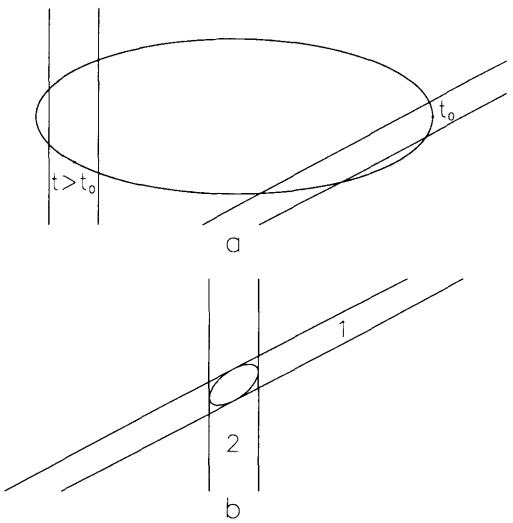

Fig. 8. (a) Uncertainty in the state of oscillator from a single measurement of displacement at $t > t_0$ corresponds to infinite uncertainty in a linear combination of the state variables at the initial time. (b) When combined with a measurement of displacement at t_0, uncertainty is finite for all aspects of the model state.

The algebra of information in eqs. (8) and (9) can be generalized to accommodate measurements providing only partial information about the state at any given time. Suppose, for example, that only one aspect of the state is measured, providing an estimate d_i for that aspect (at the time of the measurement) with standard deviation σ_i. Then there will be a term in the least-squares objective function corresponding to this measurement:

$$J_i = \frac{1}{2\sigma_i^2}(m_i - d_i)^2, \tag{10}$$

where m_i is assumed to be a linear function of the model's initial state ξ:

$$m_i = \alpha_i^{\mathrm{T}} \xi. \tag{11}$$

Using the coefficient vector α, we can represent the scalar measurement as a measurement of the state vector containing information about only one aspect of the state:

$$J_i = \frac{1}{2\sigma_i^2}\left(\xi - \frac{d_i}{\alpha_i^{\mathrm{T}}\alpha_i}\alpha_i\right)^{\mathrm{T}} \times \alpha_i\alpha_i^{\mathrm{T}}\left(\xi - \frac{d_i}{\alpha_i^{\mathrm{T}}\alpha_i}\alpha_i\right), \tag{12}$$

which can be written:

$$J_i = \tfrac{1}{2}(\xi - \hat{\xi}_i)^{\mathrm{T}}\,\hat{\Sigma}_i^{-1}(\xi - \hat{\xi}_i). \tag{13}$$

The data vector $\hat{\xi}_i = \alpha_i d_i/(\alpha_i^{\mathrm{T}}\alpha_i)$ points in the direction in which variance is finite and equal to σ_i^2, and the rank-one weight matrix $\hat{\Sigma}_i^{-1} = \alpha_i\alpha_i^{\mathrm{T}}/\sigma_i^2$ is the inverse of an infinite error-covariance matrix. Several such measurements can be combined according to equations (8) and (9).

Suppose the two measurements of position were separated in time by exactly half of the oscillator period. It is clear that when the infinite confidence region of the second measurement is propagated to the time of the first measurement, it will be parallel to the first measurement's infinite confidence region. When combined, the two measurements provide no information about the initial velocity. The same would be true for an arbitrary number of position measurements, as long as they are separated by multiples of the half period. We like

to call this stroboscopic sampling. If the sampling period is almost, but not quite, stroboscopic, the infinite regions are almost, but not quite, parallel, and the uncertainty in the initial velocity is large. Our first result is that stroboscopic sampling should be avoided if at all possible.

Before proceeding with the two-oscillator example, where we shall examine the complications caused by more than one dynamical frequency, it is best to continue with the simpler example and see what roles friction and nonlinearity might play.

The equations for a damped oscillator are:

$$\frac{dv}{dt} = -\omega_0^2 x - rv$$
$$\frac{dx}{dt} = v, \tag{14}$$

where r is the friction coefficient and ω_0 is the frequency in the limit of no friction. When the friction is relatively weak, i.e., when $r^2 < 4\omega_0^2$, the oscillatory frequency is $(\omega_0^2 - r^2/4)^{1/2}$ and the damping rate is $r/2$. The phase-space trajectories spiral in toward the origin, which is the stable rest state of the system. When the friction is relatively strong, the motion is overdamped; there is no oscillation and there are two damping rates, $r/2 \pm (\omega_0^2 - r^2/4)^{1/2}$. The faster damping corresponds to initial conditions that project the mass away from its rest position. Ultimately the mass turns, and the slower damping dominates as it approaches the origin.

Fig. 9a illustrates the evolution of an initial circular confidence region for an underdamped case. Note how its area shrinks as it spirals toward the origin, indicating that information is *increasing*. This makes sense, as we know with certainty that the mass will ultimately occupy its rest position. Conversely, when estimating the initial conditions, the most recent measurements are the *least* reliable. When estimating initial conditions from similar measurements separated by large intervals, as in Fig. 3, the data collected at the initial time provide the most information, and the relative contributions from the subsequent data decrease with increasing sampling interval. The overdamped case (Fig. 9b) is similar, the principal difference being that there is no spiral in the approach to the origin. To visualize an unstable situation, reverse the direction of time along these trajectories: 9a would correspond to a growing wave, while 9b would correspond to growth without oscillation.

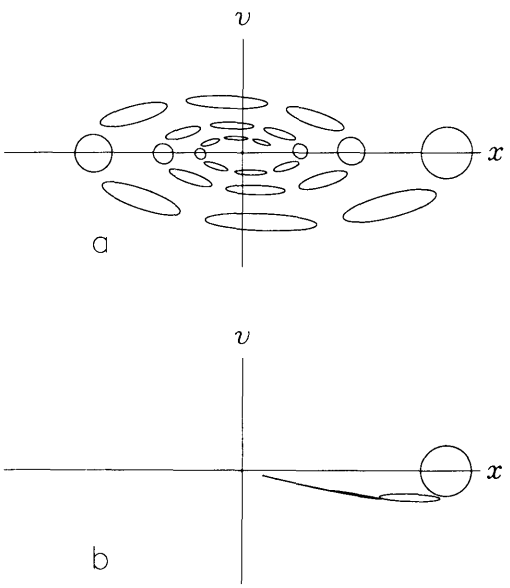

Fig. 9. Trajectory of uncertainty for (a) damped and (b) overdamped oscillator.

Blumenthal (1991) has attributed dynamical growth of uncertainty to the fact that the eigenvectors of the dynamical evolution operator are not orthogonal. In the overdamped case, the oscillator's evolution operator has two real non-orthogonal eigenvectors, and it is easy to see the initial growth of uncertainty in the position. But, for the critically damped case, there is only one eigenvector (the matrix is defective), and there is still initial growth of uncertainty in the position. When there is no friction and units are chosen so that phase-space orbits are circular, the dynamical matrix is skew-symmetric and has orthogonal eigenvectors, but there is still initial growth in uncertainty of the position corresponding to the rotation of the standard-deviation ellipse (compare with Fig. 5). More important than non-orthogonality is the transformation of information and uncertainty by the dynamics. For conservative systems, uncertainty is redistributed among the variables, and, in the case of friction or instability, there is interchange of uncertainty with a background source or sink.

The equations for a nonlinear oscillator are:

$$\frac{dv}{dt} = -\omega_0^2(1+\beta x^2)\,x$$
$$\frac{dx}{dt} = v, \tag{15}$$

where β is the increased-stiffness parameter. If $\beta > 0$, the spring gets stiffer as it is stretched or compressed from its relaxed length, causing phase-space orbits to be more rectangular. The solution can be written in terms of elliptic functions, and the expression for the period involves elliptical integrals of the first kind (Cunningham, 1958). The most significant aspect of the motion is the dependence of frequency on amplitude: the farther the spring is initially stretched before it is released, the faster it will oscillate. Consequently, an initially elliptical region of plausible states does not map onto some other elliptical region at a later time. Equivalently, if the uncertainty is normally distributed at one time, it will not be at later times, and if the nonlinearity is strong, the inverse of the Hessian of the misfit function may not be a good approximation to the covariance matrix for posterior errors.

Fig. 10 illustrates how a region of initial states consistent with a single observation of the displacement is transformed after the time corresponding to the period of the trajectory through the midpoint of the region. (Transforming backward in time can be visualized by flipping the picture about the x-axis.) This example corresponds to a strong nonlinearity, i.e., to $\beta x_{max}^2 = \frac{1}{2}$, where x_{max} is the center of the observed range of displacement; the restoring force is 50% greater than for infinitesimal displacements. And the larger the magnitude of the unmeasured velocity, the more nonlinear the dynamics. Points within the plausible region corresponding to very fast motion are on trajectories with very short periods, so during the elapsed interval, they might orbit the origin several times. (The picture is cropped; the transformed region should spiral about the origin, but only the portion of the region with modest initial velocities is shown.)

Note that two measurements of displacement spaced at an interval equal to the period of the illustrated orbit *do* provide some information about the initial velocity, the velocity is most likely

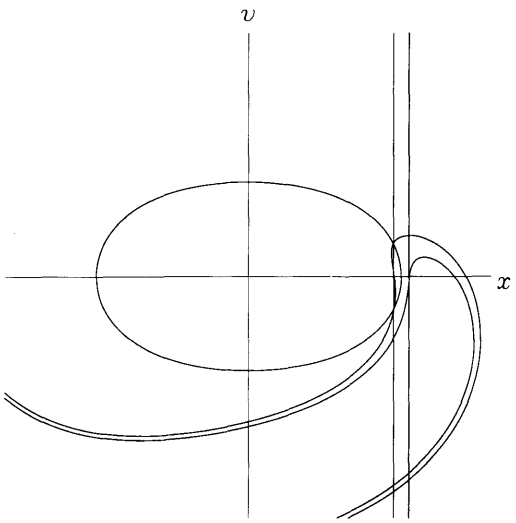

Fig. 10. Uncertainty resulting from a single observation of position of an undamped nonlinear oscillator with $\beta x_{max}^2 = \frac{1}{2}$. The state is likely to be in the region between the parallel lines at the time of the measurement and in the region between the curved lines at a later time corresponding to the period of the ellipsoidal orbit.

confined to the intersection of the two regions. (See Fig. 10.) The intersection consists of an infinite number of disjoint regions, corresponding to a multi-modal probability density that is far from normal. If large values of velocity can be excluded on grounds other than the two observations, perhaps only a single region need be considered. And if the nonlinearity is sufficiently weak, this region should resemble the intersection of two ellipses with almost parallel axes, and the uncertainty after the observations are combined should be approximately described by normal statistics.

3. The double oscillator

Our original motivation was to understand the connection between sampling frequency and dynamic frequencies when fitting a model to data. The single oscillator illustrated (1) how uncertainty and information are transformed dynamically and (2) that problems can be expected when the sampling period is commensurate with the period of oscillation. To see what additional complications arise when the dynamics support more

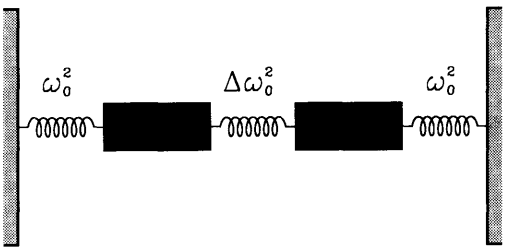

Fig. 11. Configuration of double oscillator described by eqs. (16). The masses are equal and the springs connected to the wall have the same stiffness.

than one frequency, we consider the simplest system with more than one frequency, a double oscillator. We have chosen the mass-spring configuration illustrated in Fig. 11, because its normal modes of oscillation are easily visualized. The masses are equal, as are the springs that connect them to the walls, so one normal mode has the center of mass at rest and the other has no relative motion.

For springs with stiffness ω_0^2 and $\Delta\omega^2$, the motion is governed by the equations:

$$\begin{aligned}
\frac{dv_1}{dt} &= -\omega_0^2 x_1 - \Delta\omega^2(x_1 - x_2),\\
\frac{dv_2}{dt} &= -\omega_0^2 x_2 - \Delta\omega^2(x_2 - x_1),\\
\frac{dx_1}{dt} &= v_1,\\
\frac{dx_2}{dt} &= v_2,
\end{aligned} \tag{16}$$

where x_1, x_2, v_1, and v_2, are the displacements from equilibrium and velocities of two masses. In the limit $\Delta\omega = 0$, the masses are uncoupled and the double oscillator becomes two identical single oscillators with frequency ω_0. The characteristic frequency of the normal mode with no relative motion between the two masses is $\omega_s = \omega_0$ and the characteristic frequency of the mode, with the center of mass at rest is $\omega_f = (\omega_0^2 + 2\,\Delta\omega^2)^{1/2}$. The solution can be expressed in terms of the initial conditions x_{10}, x_{20}, v_{10}, v_{20}:

$$\begin{aligned}
x_1(t) &= \frac{1}{2}(x_{10} + x_{20})\cos\omega_s t\\
&\quad + \frac{1}{2}(x_{10} - x_{20})\cos\omega_f t\\
&\quad + \left(\frac{v_{10} + v_{20}}{2\omega_s}\right)\sin\omega_s t\\
&\quad + \left(\frac{v_{10} - v_{20}}{2\omega_f}\right)\sin\omega_f t,\\
x_2(t) &= \frac{1}{2}(x_{10} + x_{20})\cos\omega_s t\\
&\quad - \frac{1}{2}(x_{10} - x_{20})\cos\omega_f t\\
&\quad + \left(\frac{v_{10} + v_{20}}{2\omega_s}\right)\sin\omega_s t\\
&\quad - \left(\frac{v_{10} - v_{20}}{2\omega_f}\right)\sin\omega_f t,\\
v_1(t) &= \frac{1}{2}(v_{10} + v_{20})\cos\omega_s t\\
&\quad + \frac{1}{2}(v_{10} - v_{20})\cos\omega_f t\\
&\quad + \frac{\omega_s}{2}(x_{10} + x_{20})\sin\omega_s t\\
&\quad + \frac{\omega_f}{2}(x_{10} - x_{20})\sin\omega_f t,\\
v_2(t) &= \frac{1}{2}(v_{10} + v_{20})\cos\omega_s t\\
&\quad - \frac{1}{2}(v_{10} - v_{20})\cos\omega_f t\\
&\quad + \frac{\omega_s}{2}(x_{10} + x_{20})\sin\omega_s t\\
&\quad - \frac{\omega_f}{2}(x_{10} - x_{20})\sin\omega_f t.
\end{aligned} \tag{17}$$

The trajectory of the center of mass is an ellipse in the $(x_1 + x_2, v_1 + v_2)$-plane with eccentricity ω_s, and that of the relative separation is an ellipse in the $(x_1 - x_2, v_1 - v_2)$-plane with eccentricity ω_f. In general, both normal modes are present. Because energy is conserved, we know that phase-space

Tellus A/46/5-8

trajectories lie on the surface of a four-dimensional ellipsoid:

$$\frac{\omega_s^2}{2}(x_1+x_2)^2+\frac{1}{2}(v_1+v_2)^2+\frac{\omega_f^2}{2}(x_1-x_2)^2 + \frac{1}{2}(v_1-v_2)^2=E, \tag{18}$$

where E is the initial energy of the system. The trajectories are difficult to visualize, but some idea of the nature of the four-dimensional trajectories can be gleaned from two-dimensional projections. Fig. 12 shows the trajectory of a system with $\omega_s=2\pi/1.9$ and $\omega_f=2\pi/0.4$ and initial conditions $x_1=1$ and $x_2=v_1=v_2=0$ projected on various

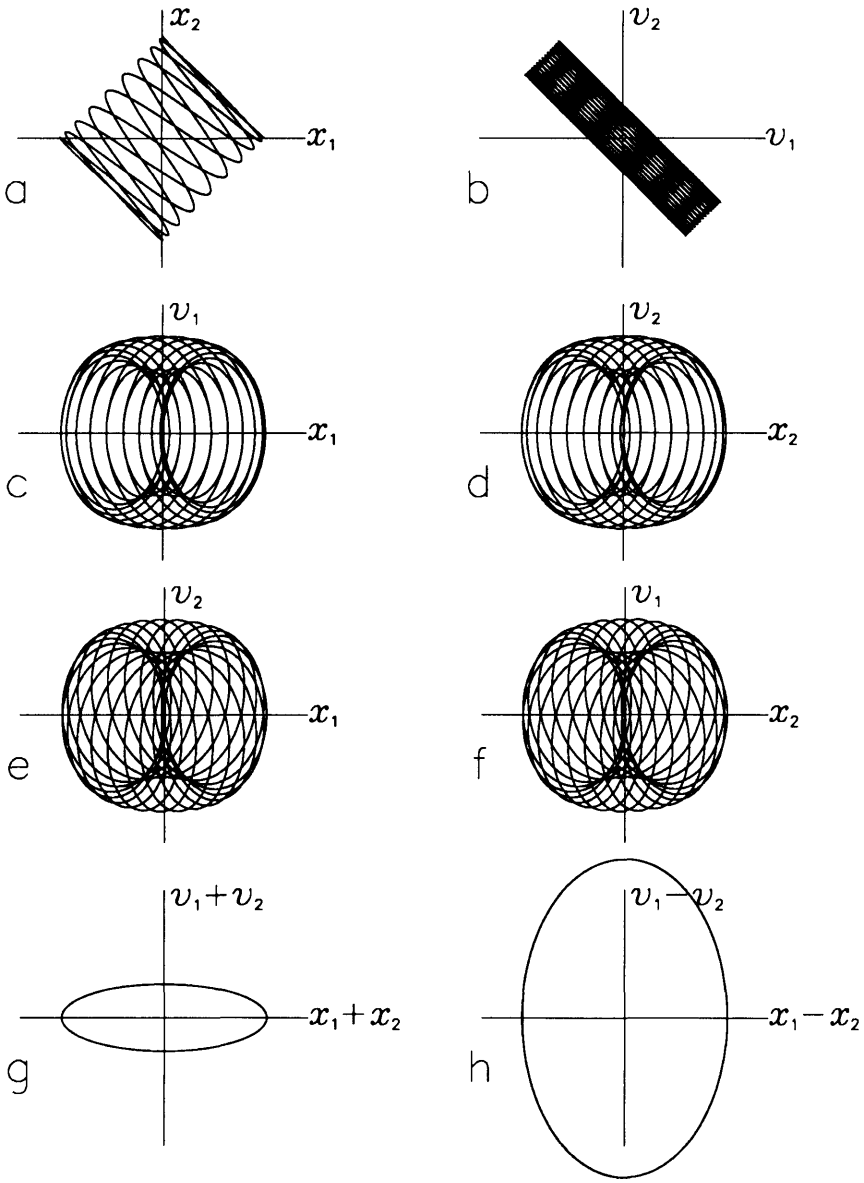

Fig. 12. Trajectory of the double oscillator with normal-mode periods of 0.4 and 1.9 projected onto various 2-dimensional planes. The elliptical orbits of the slow and fast modes are evident in (g) and (h), respectively.

two-dimensional planes. The fact that, other than the ellipses corresponding to the normal modes, these Lissajous figures are not very easy to interpret is an indication of the difficulty in comprehending the phase-space trajectory of a dynamical system as simple as the double oscillator.

Fig. 13 illustrates the transformation of uncertainty as reflected by a projection onto the (x_1, x_2)-plane for the same trajectory as shown in Fig. 12. The four-dimensional uncertainty ellipsoid is initially spherical, corresponding to uncorrelated errors in units determined by the observational accuracies. To generate the projections of the evolving uncertainties, we used the fact that the projection of an ellipsoid corresponding to a particular covariance matrix is the ellipse corresponding to the sub-matrix associated with the axes of the projection plane. The fact that the area of the projected ellipse decreases and then increases means that our knowledge about x_1 and x_2 increases at the expense of v_1 and v_2 and then decreases again as v_1 and v_2 regain their original precision. The volume of the 4-dimensional standard-deviation ellipsoid is constant in time, indicating that our knowledge of the state of the double-oscillator system as an entity is always equivalent to our knowledge of its initial state.

Suppose we could measure only the position of the first mass. Clearly, determination of the state of the double oscillator requires at least four measurements. To explore how the sampling frequency impacts the uncertainty of the initial state, we considered a series of experiments in which the four-dimensional trajectory of a double oscillator with characteristic frequencies 0.4 and 1.9 is fitted to ten (equally certain) measurements $x(t_0)$, $x(t_0 + \Delta t)$, ..., $x(t_0 + 9\,\Delta t)$, where Δt varied from experiment to experiment. For each fit, the determinant of the Hessian matrix was computed as a measure of the information about the system extracted from the ten measurements. This determinant, which is inversely proportional to the volume of the 4-dimensional standard-deviation ellipsoid, vanishes when some aspect of the double oscillator's state is not determined by the 10 measurements.

The top panel of Fig. 14 shows the determinant of the Hessian matrix plotted as a function of Δt. The fluctuations are reminiscent of those in Figs. 2, 3. However, the determinant of the Hessian is a better measure of quantity of information than are the variables plotted in those figures. Condition number, for example, is more an indicator of computational difficulty, as it can increase when information is added about the best-determined aspect or when information is reduced for the worst-determined aspect. Furthermore, whereas individual eigenvalues are not dimensionally homogeneous, the determinant has recognizable dimensions, so it can be considered to be a

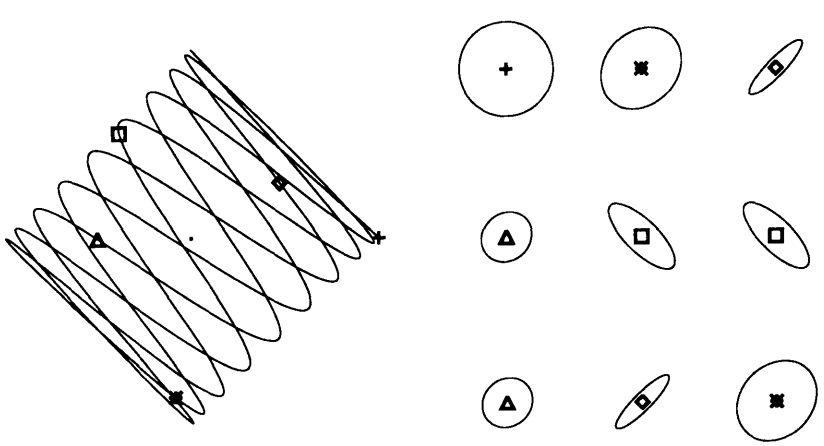

Fig. 13. The ellipses represent uncertainty in the state of a double oscillator having normal-mode periods 0.4 and 1.9 projected onto the (x_1, x_2)-plane at intervals of $\frac{1}{9}$ the return period, starting at the initial time on the top left and ending on the bottom right. The symbols cross-reference the ellipses with points on the trajectory of their centers.

meaningful quantity independent of a particular system of units. And, as mentioned above, the square root of the determinant is proportional to a volume in phase space; for non-dissipative dynamics, this volume remains constant, even though its shape may change as information is redistributed dynamically among the state variables.

Values of 10^{-6} in 14 can be interpreted as zero. The first thing to notice is that the Hessian becomes singular as $\Delta t \to 0$. This reflects the fact that uncertainty of estimates of velocity approaches $2\,\Delta x_1/\Delta t$ as $\Delta t \to 0$, where Δx_1 is the uncertainty of the x_1 measurements. An even more conspicuous result is that the Hessian is very often singular. If the singularities occurred at only the stroboscopic frequencies, there would be only *one* zero on the plot, because the repeat period of the double oscillator is 7.6! Although the top panel of Fig. 14 shows results for sets of ten equally spaced measurements, the picture would be much the same if only the first four measurements were used. In particular, there would be no change in the distribution of zeros. The last 6 observations shed

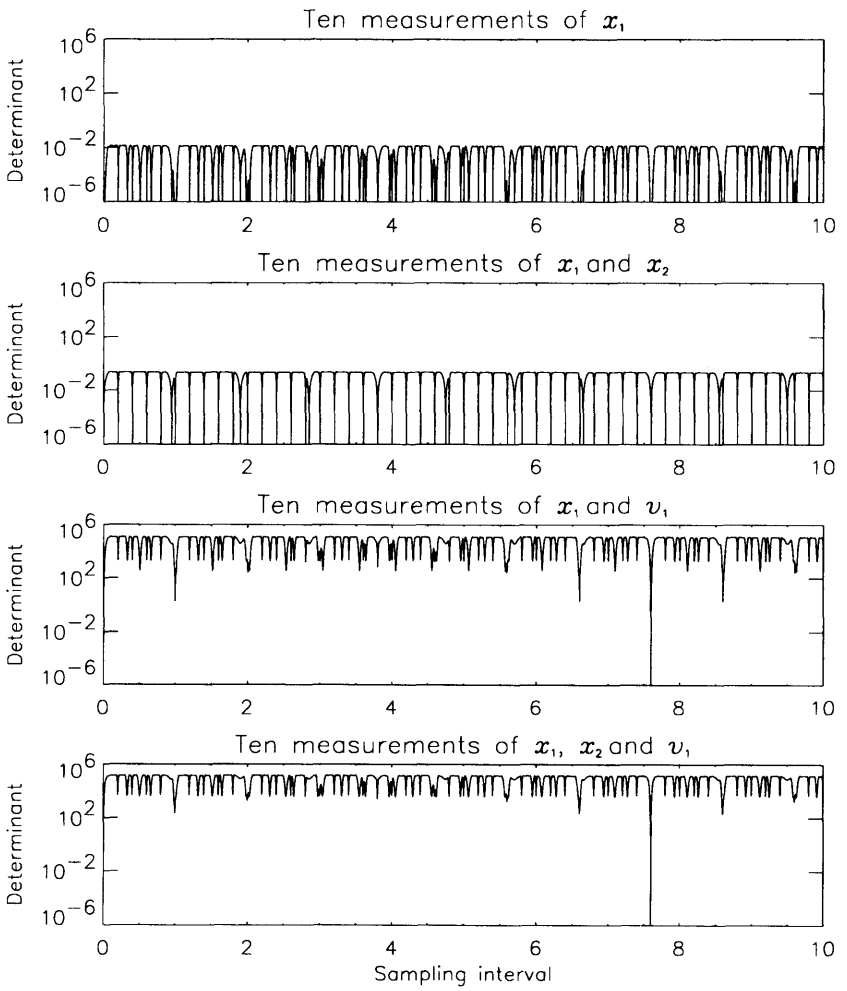

Fig. 14. Determinant of Hessian for 10 equally spaced measurements of x_1 (top panel) x_1 and x_2 (second panel), x_1 and v_1 (third panel), and x_1, x_2, and v_1 (bottom panel) as a function of the sampling interval for a system with normal-mode periods of 0.4 and 1.9.

no light on the aspects of the double oscillator's state that the first four have failed to sample.

To investigate the large number of unfavorable sampling frequencies, we asked under what mathematical circumstances the Hessian should be singular. Because we are solving a linear least-squares problem, the Hessian is singular whenever the rank of the model matrix is not the same as the number of variables to be determined. The model matrix is the matrix associated with the equations relating the initial state to the model counterparts of the measurements. In this case, the 10 model equations are of the form:

$$\begin{pmatrix} m(t_0) \\ m(t_1) \\ m(t_1) \\ \vdots \\ m(t_9) \end{pmatrix} = \begin{pmatrix} e^{i\omega_s t_0} & e^{-i\omega_s t_0} & e^{i\omega_f t_0} & e^{-i\omega_f t_0} \\ e^{i\omega_s t_1} & e^{-i\omega_s t_1} & e^{i\omega_f t_1} & e^{-i\omega_f t_1} \\ e^{i\omega_s t_2} & e^{-i\omega_s t_2} & e^{i\omega_f t_2} & e^{-i\omega_f t_2} \\ & \ddots & & \\ e^{i\omega_s t_9} & e^{-i\omega_s t_9} & e^{i\omega_f t_9} & e^{-i\omega_f t_9} \end{pmatrix} \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix}. \tag{19}$$

The left-hand side represents the measurements and the right-hand side involves the (real) unknown initial conditions x_{10}, x_{20}, v_{10}, and v_{20} via (complex) unknowns A, B, C, and D, the exact relationship depending on the quantity measured. For example, when $m = x_1$, as in the computational example:

$$\begin{aligned} A &= \frac{1}{4}(x_{10} + x_{20}) - \frac{i}{4\omega_s}(v_{10} + v_{20}), \\ B &= \frac{1}{4}(x_{10} + x_{20}) + \frac{i}{4\omega_s}(v_{10} + v_{20}), \\ C &= \frac{1}{4}(x_{10} - x_{20}) - \frac{i}{4\omega_f}(v_{10} - v_{20}), \\ D &= \frac{1}{4}(x_{10} - x_{20}) + \frac{i}{4\omega_f}(v_{10} - v_{20}). \end{aligned} \tag{20}$$

The 1st and 2nd columns of the model matrix are the same when the sampling interval $\Delta t = t_{i+1} - t_i$ is a multiple of the half period of the slow mode, and the 3rd and 4th when it is a multiple of the half period of the fast mode. More generally, when the exponents for two columns differ by a multiple of 2π, those columns will be identical:

$$\begin{aligned} &(\omega_f \pm \omega_s)\,\Delta t = 2l\pi \\ \text{or} \quad &\omega_s\,\Delta t = m\pi \\ \text{or} \quad &\omega_f\,\Delta t = n\pi, \end{aligned} \tag{21}$$

where l, m, and n are integers. Under these conditions, the rank of the model matrix is less than four, and the Hessian is singular. These results do not depend on the number of measurements or on the variable being measured; they do require the measurements to be of a single variable and to be uniformly spaced in time.

The bad sampling intervals in the computational example are multiples of four fundamental intervals: 0.2000, 0.3304, 0.5067, and 0.95. The smallest is half the period of the fast normal mode; multiples of this interval are easily identified in the top panel of Fig. 14, the first being the first nonzero singularity. The 2nd corresponds to the sum $\omega_f + \omega_s$; see the 2nd nonzero singularity in the top panel. The 3rd corresponds to the difference $\omega_f - \omega_s$; see the 4th nonzero singularity. The largest is half the period of the slow mode; its multiples can also be recognized.

Although the mathematics of the singularities is easy to understand, it is not obvious what information is missing from badly spaced measurements. The reason for this difficulty stems from the fact that the conditions for singularities are derived from the columns of the model matrix, while the measurements are associated with its rows. Careful analysis shows that when the displacement of the first mass is sampled at multiples of the shorter normal-mode half period, the information that is missing is the initial relative velocity of the two masses. This is reasonable, as the relative motion is associated with the higher frequency. Similarly, sampling at multiples of the longer half period provides no information about the initial velocity of the center of mass. We made no effort to discern the shortcomings of the other bad sampling schedules.

The ocean is characterized by many frequencies, some of which are quite small. If, in analogy with this example, we were able to measure *only one* variable, it is clear that the situation would be practically impossible. It would be difficult to avoid bad sampling intervals. But, recalling the motivating problem of fitting a shallow-water model to satellite altimetric data, it is clear that a more reasonable situation would be where there are several, possibly many, variables measured simultaneously, and perhaps different variables measured at different times. To explore something closer to this sort of situation within the context of the double oscillator, we repeated the experiment using data for the displacements of both masses, the simplest analogue of the synoptic altimetric coverage of the entire ocean. Results of the sampling-interval analysis are shown in the second panel of Fig. 14. The situation is somewhat better as only multiples of the fundamental half periods are present, but the high normal-mode frequency still causes difficulty.

Rationalizing these results mathematically proceeds along the same lines as before. In this case there are 20 measurements, and the model matrix reflects the differences in the types of measurements:

$$\begin{pmatrix} x_1(t_0) \\ x_1(t_1) \\ \vdots \\ x_1(t_9) \\ x_2(t_0) \\ x_2(t_1) \\ \vdots \\ x_2(t_9) \end{pmatrix} = \begin{pmatrix} e^{i\omega_s t_0} & e^{-i\omega_s t_0} & e^{i\omega_f t_0} & e^{-i\omega_f t_0} \\ e^{i\omega_s t_1} & e^{-i\omega_s t_1} & e^{i\omega_f t_1} & e^{-i\omega_f t_1} \\ & & \ddots & \\ e^{i\omega_s t_9} & e^{-i\omega_s t_9} & e^{i\omega_f t_9} & e^{-i\omega_f t_9} \\ e^{i\omega_s t_0} & e^{-i\omega_s t_0} & -e^{i\omega_f t_0} & -e^{-i\omega_f t_0} \\ e^{i\omega_s t_1} & e^{-i\omega_s t_1} & -e^{i\omega_f t_1} & -e^{-i\omega_f t_1} \\ & & \ddots & \\ e^{i\omega_s t_9} & e^{-i\omega_s t_9} & -e^{i\omega_f t_9} & -e^{-i\omega_f t_9} \end{pmatrix} \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix}. \tag{22}$$

The minus signs in the 3rd and 4th columns prevent singularities related to sums and differences of the normal-mode frequencies.

The 3rd panel of Fig. 14 shows results for the case where the displacement and velocity of the first mass are measured. The only singularities are multiples of the return period for the double oscillator, i.e., true stroboscopic sampling. And the bottom panel shows the same singularities for measurements of both displacements and one velocity. Note that, even though the Hessian is not singular for nearly as many sampling frequencies, some sampling frequencies are much worse than others. For example, a sampling interval of 1.0 leads to a small determinant in the lower two panels, indicating that less information is obtained than might be obtained with a slightly different interval. Looking more closely, you can see that the minima of the determinant of the Hessian in the lower two panels align precisely with the singularities in the top panel! We have made no attempt to analyze why this is the case, but we can conclude that sampling intervals that are related to the characteristic frequencies as in eqs. (21) should be avoided.

The singularities associated with observing displacement-velocity pairs (lower 2 panels of Fig. 14) require a well-defined return period, i.e., they require the ratio of normal-mode frequencies to be a rational number. If the ratio is irrational, there will be no exact return period, but there will be times when the system is close to its initial state. At those times it is reasonable to expect that the Hessian is almost singular. Fig. 15 illustrates what happens when the ratio of frequencies cannot be constructed from small integers, i.e., when the exact return period is large, 7596 and 75996 respectively. In both cases, the singularity at $\Delta t = 7.6$, which appears in 14a through 14d, shows up as an almost singular Hessian in the neighborhood of 7.6. (In the lower panel the determinant is so small that it can't be distinguished form a singularity on the plot.) Note that the bad sampling intervals near 1.0, 6.6, and 8.6 also persist under small perturbations of the ratio of normal-mode frequencies. In fact, both panels of Fig. 15 are remarkably like 14c in almost every detail. This suggests that, for a system with many natural frequencies, the quality of the sampling schedule is stable under small perturbations (uncertainties) of the natural frequencies of the model.

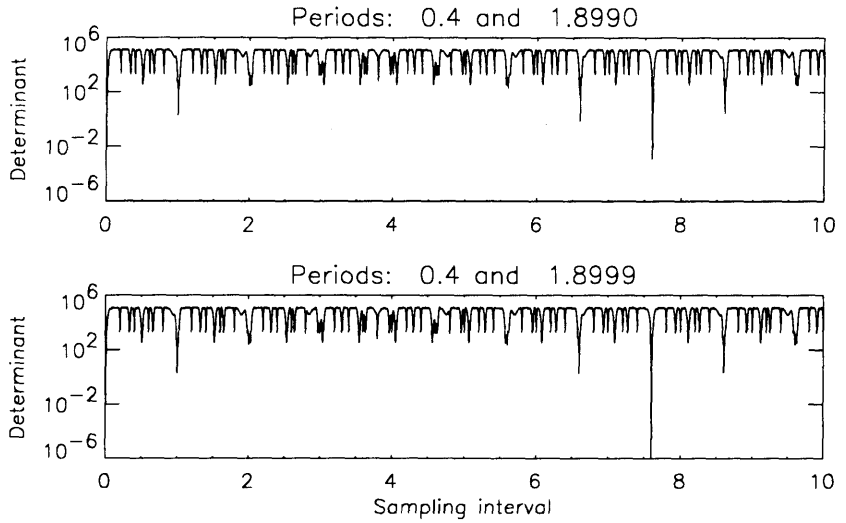

Fig. 15. Determinant of Hessian for 10 equally spaced measurements of x_1 and v_1 as a function of the sampling interval for fast normal mode having period 0.4 and slow normal mode with period 1.8990 and 1.8999.

4. Shallow-water model

Although it is not feasible to analyze an oceanic model as completely as the single and double oscillators, it is possible to discuss it qualitatively. To do so, we start by considering the continuous-time form of the model, where the spatial domain remains discretized but complications of the time-stepping scheme are avoided. The general form for such a linear model is:

$$\frac{df}{dt} = Mf, \tag{23}$$

where the model state f, e.g., velocity and elevation fields, can be represented as a vector and the temporal evolution operator M can be represented as a matrix. The characteristic frequencies of the model are found by solving the eigenvalue problem:

$$Me_i = \lambda_i e_i. \tag{24}$$

Because M is real, its eigenvalues λ_i are either real or complex conjugate pairs. An eigenvalue's imaginary part is like a normal-mode frequency, while its real part is a dissipation rate (or growth rate, in the case of an unstable model). Real eigenvalues correspond to non-oscillatory motion (recall the overdamped oscillator), but they need not come in pairs. Eigenvalues that are zero correspond to conserved linear properties.

There are 341 eigenvalues for the continuous-time counterpart to the simple shallow-water model discussed in the introduction. One is zero, because the model conserves mass. (It also conserves energy, but that is not a linear property of the system and doesn't give rise to a zero eigenvalue.) The remaining are 170 complex-conjugate pairs. In other words, the model resembles a chain of 170 masses coupled by springs. If the system is initially located within a spherical region of phase space, this region transforms into a rotating, pulsating, 341-dimensional ellipsoid, the motion of which is characterized by 170 frequencies. We had difficulty visualizing and comprehending the transformation of uncertainty for the double oscillator where only two frequencies were involved, but we could see that information passed from one dynamical variable to the others at the characteristic frequencies of the system. Because of this, information obtained by sampling at intervals related to sums or differences of the natural frequencies resulted in redundant information about some variables at the expense of others. For a model having 170 frequencies, the redistribution of information is qualitatively similar, but more complex. But having many more frequencies, which give rise to extremely many sums and differences, there are very many sampling intervals that can fail to provide information about some aspect of the model state.

The actual numerical model has a leap-frog time-stepping scheme, which computes the next model state from the states at the present and previous time levels. As such it is not consistent with the continuous-time form of the model (eq. (23)); the information from the 2 time levels is like the 2 sets of initial conditions needed for 2nd-order differential equations. This scheme has a well-known time-splitting instability that is due to what are referred to as computational modes, which can be suppressed by devices like the Robert filter. The point to be made here is that this scheme gives rise to twice as many frequencies as you would expect on the basis of the continuous-time model. So there are even more combinations of frequencies determining unfavorable sampling intervals than there would be for a two-level time-stepping scheme.

The details of the finite-difference model can influence its fit to data in other ways. For example, to start a leap-frog scheme from a single time level, a 2-level scheme is needed for the first step. We used a semi-implicit scheme, where a forward step was used to compute the velocity field, and then this velocity field was used in a forward step in computing the elevation field. Then the leap-frog scheme was used to compute all fields at the 3rd level. We found that, when fitting to fields of simulated altimetric data at the first 3 time levels, the Hessian was singular. In fact, it had 121 zero eigenvalues! Even though there were ostensibly enough data to determine all 3 fields, a complete field of information was missing. We were able to verify that this was due to the semi-implicit start-up step in 2 ways. First, we were able to show analytically, when the earth's curvature was neglected, that this scheme indeed should lead to a singular Hessian. And then we replaced the scheme with a fully explicit scheme, where all fields at the second time level were computed directly from the initial fields, and we found that the Hessian was no longer singular. The three fields of data were sufficient, just as intuition would have it.

Fig. 16 illustrates the variation of the determinant of the Hessian matrix with sampling interval when the shallow-water model is being fit to three equally-spaced fields of altimetric data. This figure differs from Figs. 14 and 15 in 2 ways. First, the range of the values, which span 400 orders of magnitude rather than 4, so fluctuations of several orders of magnitude are not visible. Second, the sampling interval is restricted to multiples of the model's three-hour time step, so the most extreme fluctuations, singularities in particular, might not have been encountered. (We have analyzed the discrete-time oscillator and found that, while it is possible to have time steps consistent with stroboscopic sampling, such steps are not likely to be chosen at random.) Still, the information about the system that can be extracted from the data doesn't appear as sensitive to the sampling interval as is the case for the double oscillator. The dramatic fluctuations in Figs. 2, 3 appear more to be the result of information being passed rapidly between different aspects of the model and thus being partitioned differently over the 341 variables as the sampling interval changes than to be the result of significantly less information about the system as a whole. The trend of decreasing information with increasing sampling interval is clear; it is due to dissipation. The best sampling interval appears to be 17 time steps, i.e., about 2 days. Comparing with Fig. 3, you can see that for this interval the smallest eigenvalue is as large as it ever gets, the largest is at a relative minimum, and the

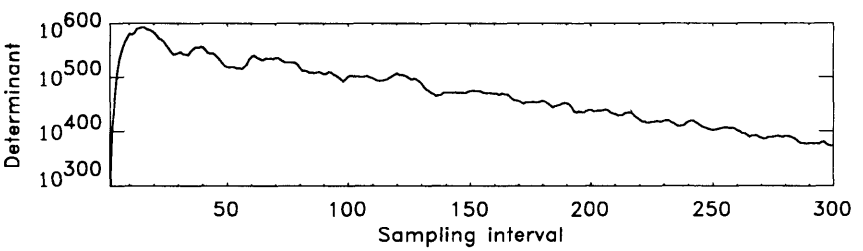

Fig. 16. Determinant of the Hessian matrix as a function of sampling interval separating in time three fields of fictitious altimetric data to which the shallow-water equations are fitted.

condition number has its second best value. The computations are relatively well-conditioned when the sampling provides the most information.

5. Conclusions

The 1st point that we would like to emphasize is that fitting a model to a time series by adjusting its initial conditions is quite different from analyzing the time series by Fourier methods. In particular, the sampling interval does not determine the highest frequencies that are detectable, nor does the length of the series determine the lowest. The frequencies that make up the series are intrinsic to the model. The distribution of energy over these frequencies depends on the initial conditions, which play no role in time-series analysis. Given the initial conditions, without projecting them on the model's normal modes, it is difficult to say how the energy is distributed, e.g., how much of the flow can be attributed to Rossby wave and how much to gravity waves, but that distribution governs the subsequent evolution. In numerical weather prediction, sophisticated initialization procedures have been developed to remove unwanted gravity-wave energy. Here we have shown that the relationship between the sampling interval and distribution of energy over normal modes as determined by a least-squares fit is far from intuitive.

Secondly, we want to emphasize the result that some aspects of the model state might not be determined by the available observations, even when the number of available observations greatly exceeds the number of initial conditions, if the sampling interval corresponds to sums or differences of the model's normal-mode frequencies. The examples of the single and double oscillators illustrated how such sampling provides corroborative information about already observed aspects of the system while leaving other aspects unobserved. A complete analysis to identify which aspects are unobserved is in general quite difficult even for a simple shallow-water-equation model, but it is important to recognize the possibility of such unfortuitous sampling. Not only does it leave some aspects of the model state undetermined, but it presents a computational problem that is singular or ill-conditioned. Furthermore, as oceanic and atmospheric models typically involve large numbers of characteristic frequencies with even larger numbers of sums and differences, a given sampling schedule is likely to provide little or no information about some aspects of the state. And without a computationally expensive analysis of posterior errors, it would be hard to know what information is missing.

Perhaps the most important conclusion that can be drawn from this study is that care must be taken to formulate the least-square fit in such a way that unfortuitous sampling is less problematic. This can be done by incorporating prior estimates for all variables along with the actual measurements. This additional information plays exactly the same role as the so-called background or first guess in optimal interpolation. (See, for example, Lorenc, 1986, or Thacker, 1992). It limits the uncertainty of those aspects of the solution for which the data provide no information. By limiting uncertainty, it controls ill-conditioning due to the smallest eigenvalues of the Hessian. And if the prior information dominates the large eigenvalues through its demand that the best-fit field be smooth, the prior error-covariance matrix might serve as an effective preconditioner.

We chose the title "Dynamics of information and uncertainty", because the paper addresses the question of what information might be extracted from asynoptic data through a least-squares model fit. We have shown how the least-squares formalism can be thought of as providing two services: one is the dynamical transformation of information to a common time, and the other is the combination of the information into an optimal estimate for the model state. And we have shown how both aspects might be visualized: dynamics of information can be thought of in terms of transformations on regions of phase space, and the combination of information satisfies an algebra that guarantees the region of uncertainty is smaller after measurements are combined. Each individual measurement can be represented geometrically as an infinite region in phase space bounded only in one direction. The uncertainty of all measurements, taken collectively, can be visualized roughly as the intersection of infinite regions of uncertainty. Clearly, the directions of finite uncertainty of all the individual measurements must span the entire phase space. Moreover, the directions must be sufficiently orthogonal, so the overlap of the regions does not extend too far in

any direction. Visualizing in detail the collective information of all the measurements determining the fit is difficult because of the high dimensionality of the model's phase space, the very large number of measurements, and the complicated transformation of information among the variables by the dynamics. Nevertheless, it is possible to imagine what sort of situations might occur. For example, some measurements may be almost inconsistent and others providing corroborating information; in neither case would they restrict the possible states along new directions. But as long as the sampling is adequate, i.e., there is information restricting the region in all phase-space directions, redundant information simply serves to improve the estimate of some combinations of variables, while inconsistencies can provide grounds for rejecting the model.

REFERENCES

Blumenthal, M. B. 1991. Predictability of a coupled ocean-atmosphere model. *J. Climate* **4**, 766–784.

Callies, U. and Eppel, D. P. 1994. On the sensitivity of a least-squares fit of discretized linear hyperbolic equations to data. *Tellus*, in press.

Cunningham, W. J. 1958. *Introduction to nonlinear analysis.* McGraw-Hill, New York.

Le Dimet, F. X. and Talagrand, O. 1986. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus* **38A**, 97–110.

Lorenc, A. E. 1986. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.* **112**, 1177–1194.

Rebert, J. P., Donguy, J. R., Eldin, G. and Wyrtki, K. 1985. Relation between sea level, thermocline depth, heat content, and dynamic height in the tropical Pacific Ocean. *J. Geophys. Res.* **90**, 11719–11725.

Thacker, W. C. and Long, R. B. 1988. Fitting dynamics to data. *J. Geophys. Res.* **93**, 1227–1240.

Thacker, W. C. 1989. The role of the Hessian matrix in fitting models to measurements. *J. Geophys. Res.* **94**, 6177–6196.

Thacker, W. C. 1992. Oceanographic inverse problems. *Physica D* **60**, 16–37.