Forecasting improvement via optimal choice of sites for observations: a toy model

Sarah Dance

1 Introduction

Oceanic and atmospheric processes can be described as nonlinear dynamical systems. These systems often display a chaotic sensitivity to initial conditions, i.e. trajectories which have close initial conditions in phase space may diverge rapidly in forward time. Since no observable can be measured exactly, it is not possible to know the true state of the system at any given time. Hence, even using a perfect model, it would seem that trying to predict future states accurately is futile. However, it is certainly not true that all directions in phase space are unstable to perturbations. By identifying the most unstable directions and improving the accuracy of measurements at the corresponding locations in physical space, a strategy for forecast improvement can be developed.

Storm tracks are an example of a nonlinear system displaying regions of perturbation growth and damping. In the winter, for example off the East coast of the USA, there are large air temperature gradients between the land and the sea. These give rise to large density gradients and the area becomes unstable to baroclinic instabilities. Further downstream, density gradients are much smaller so perturbations are dissipated. Intuitively, at least some unstable directions in phase space may be expected to correspond to unstable regions in physical space.

Another example of a natural nonlinear system displaying chaotic dynamics is the ocean circulation. The western boundary current and its offshore extension are unstable to small perturbations which grow and form eddies. These eddies affect the flow downstream. Away from the boundaries and strong interior fronts the system is often stable and the eddies dissipate.

In the ocean, observations are very scarce due to the difficulty in making them. Predictability techniques, such as those described later, could be of particular use here in choosing the optimal sites for these measurements. There is also a choice of the kind of measurements to be made. For example, is it better to use lagrangian floats or make a series of observations at a fixed location? Perhaps more sophisticated predictability techniques could be used to answer this question.

The aim of this project was to identify an optimal measurement strategy in a toy model with some qualitatively similar features to a natural system. Unstable directions in phase space were found using different techniques. Simulated forecasts were then carried out based on an approximate system and additional observations of the true system taken at locations determined using information about the most unstable directions. These forecasts were compared

with each other, with the evolution of the true system, and with the evolution of the approximate system unconstrained by additional observations. This work examines a system described by a nonlinear one -dimensional PDE; Lorenz and Emmanuel [Lorenz and Emmanuel, 1998] carried out a similar analysis using coupled ODEs.

2 The Model PDE

The model PDE is given by

$$u_t + \beta u u_x + \lambda u + \alpha u_x + \nu(x)\, u_{xx} + \mu u_{xxx} + \gamma u_{xxxx} = 0 \tag{1}$$

for $0 \leq x < 2\pi$ with periodic boundary conditions and

$$\nu(x) = \nu_0 H(x) \text{ where } H(x) = \begin{cases} 1 & \text{if } 0 \leq x < \frac{2\pi}{5} \\ 0 & \text{if } \frac{2\pi}{5} \leq x < 2\pi \end{cases}$$

This is a modified version of Benney's equation. Here there are two extra terms: a linear advection term and a damping term. A spatial dependence in the coefficient of u_{xx} has been introduced so that there is an area in the domain which is unstable to perturbations and damping elsewhere. Benney's equation can be used to describe motion of thin liquid films and has also arisen in plasma instabilities. It has been shown to admit pulse-like solutions [Balmforth et al., 1997].

Fig. 1 shows a sample trajectory of the system (1) obtained by numerical integration. There are two regimes evident: a vigorous regime (e.g. from $t = 0$ to $t = 15$ or $t = 25$ to $t = 35$) where there are large disturbances in the unstable region, and a quiet regime (e.g. near $t = 20$ or $t = 40$) where the disturbances are much smaller. Nonlinear effects cause some disturbances to move to the left.

An energy equation for (1) is given by

$$\frac{1}{2}\frac{\partial}{\partial t}\int_0^{2\pi} u^2 dx = \nu_0 \int_0^{2\pi/5} (u_x)^2\, dx - \gamma \int_0^{2\pi} (u_{xx})^2\, dx$$

Note that the only positive contributions to the right-hand-side come from the interval $[0, 2\pi/5]$ where $\nu = \nu_0$.

The parameter values used throughout the rest of this discussion are: $\lambda = 6.4 \times 10^{-2},\ \beta = 2.0,\ \nu_0 = 1.7 \times 10^{-2}, \alpha = 5.0 \times 10^{-2},\ \mu = 1.0 \times 10^{-5},$
$\gamma = 8.3 \times 10^{-5}$

These parameters were chosen so that growth was significant in the unstable region of the domain and was damped as the disturbances moved into the stable region. The damping was sufficiently small to allow perturbations to grow again on re-entering the unstable region.

Simple estimates of the intrinsic time-scales of the system may be obtained by considering a linearisation about the zero state. Assuming $\nu = \nu_0$ over the whole domain, trying a solution of the form $u \propto e^{ikx+(\sigma - i\omega)t}$ and neglecting nonlinear terms gives:

$$\sigma = -\lambda + \nu k^2 - \gamma k^4,\ \omega = \alpha k - \mu k^3$$

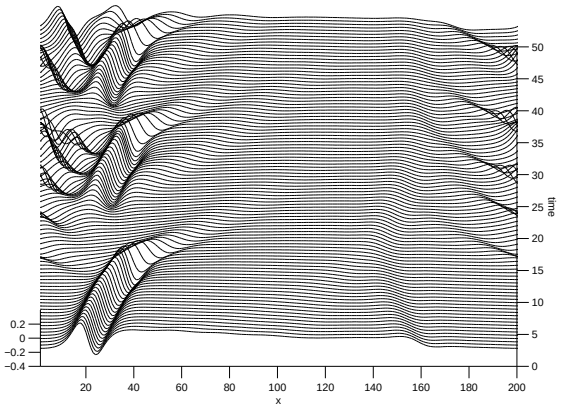

Figure 1: A segment of a trajectory of equation (1) The equation was solved numerically using a 200 point grid. Here the x-axis is labelled by grid point. Each curve represents the state of the system at a particular instant in time. Time progresses into the paper, with each curve separated by 0.5 time units. The $u(x)$ scale is on the left.

For the chosen parameters $\sigma(k)$ has a maximum at $k = \sqrt{102} \simeq 10$. So in the growing region of the domain, for a wavenumber $k = 10$, estimates of the growth rate and phase speed are $\sigma = 0.8$ and $c = 0.05$ respectively.

3 Predictability Techniques

In this section the techniques used to find the unstable directions are described. These methods include an empirical approach (Ensemble vector analysis) and more quantitative techniques based on a linearised system (Singular vector and Lyapunov vector analysis). The discussion takes place in a general setting with the system described by

$$\dot{X} = f(X)$$

with $X \in U^{\text{open}} \subseteq \mathbb{R}^n$, $f : U \to \mathbb{R}^n$

3.1 Ensemble Vectors

Ensemble vectors (EV) provide one way of estimating the most unstable directions along a given trajectory as given by an empirical approach. Note that this name is not in common use in the predictability literature. Fig. 2 shows a chosen fiducial (reference) trajectory in phase space. Let this trajectory be given by $X_{ref}(t)$. To find an ensemble vector take a ball of initial conditions of a given (small) radius δ around the reference initial condition, $X_{ref}(0)$. Pick an ensemble of points randomly distributed on this ball and flow them forward under the full nonlinear PDE (1) for time t_{opt}, the optimisation time. (t_{opt} is generally taken to be the same as the forecast period.) The hope is that the trajectory the furthest away from

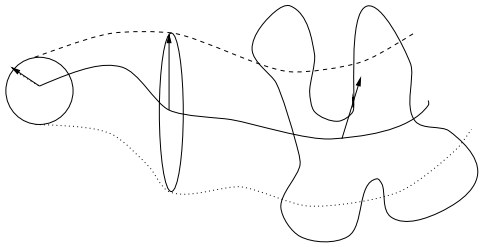

Figure 2: A fiducial trajectory is shown as a solid line. The evolution of a ball of initial conditions about the reference initial condition is depicted, with the arrows representing the instantaneous maximum growth directions. Two other trajectories with initial conditions a distance δ from the fiducial initial condition are shown as dotted and dashed lines.

the reference trajectory at optimisation time has experienced growth aligned with the most unstable direction. Let this trajectory be given by $\tilde{X}(t)$. The ensemble vector is defined by

$$EV = \kappa \left(\tilde{X} \left(t_{opt} \right) - X_{ref} \left(t_{opt} \right) \right)$$

where κ is a normalising factor, chosen so that $\|EV\| = 1$.

Note that for the purposes of this project the Euclidean norm was used to measure distances in phase space. Any other norm could have been used. It is clear that the unstable directions calculated are norm dependent. There is some debate over which norm is the best measure of trajectory divergence, see [Smith, 1995].

3.2 Singular Vectors and Lyapunov Vectors

These techniques are based on linearising the system about a reference trajectory.

Recall

$$\dot{X} = f(X)$$

with $X \in U^{\text{open}} \subseteq \mathbb{R}^n, f : U \to \mathbb{R}^n$.

Consider a small perturbation x of the state X. For sufficiently small times t the evolution of the perturbation can be described by the linear approximation

$$\dot{x} = Df(X(t))x \tag{3}$$

This equation can be written in integral form

$$x(t) = L(t, t_0)x(t_0).$$

Given an inner product $< \cdot, \cdot >$, define the perturbation norm at time t by

$$\|x(t)\|^2 = \langle x(t), x(t) \rangle = \langle x(t_0), L^* L x(t_0) \rangle$$

where L^* is the adjoint of L. In the case that L is a real matrix, $L^* = L^T$, the transpose of L.

L^*L is often referred to as the Oseledec operator. It is clearly a symmetric operator. It follows that all its eigenvalues σ_i^2 are real and there exists an orthonormal basis of eigenfunctions v_i which satisfy

$$L^*Lv_i\left(t_0\right) = \sigma_i^2 v_i\left(t_0\right)$$

Note that the eigenvalues may be repeated and the basis of eigenfunctions may not be unique.

At a future time t, these eigenfunctions evolve to $v_i\left(t\right) = Lv_i\left(t_0\right)$ and

$$LL^*v_i\left(t\right) = \sigma_i^2 v_i\left(t\right)$$

Assuming that the basis of eigenfunctions is complete (in the tangent space of perturbations), any perturbation can be written as a linear combination of eigenvectors.

$$x(t) = \sum_i c_i v_i(t)$$

It follows that

$$\max_{x(t_0)\neq 0}\left(\frac{\left\|x(t)\right\|^2}{\left\|x(t_0)\right\|^2}\right) = \sigma_1^2$$

where σ_1^2 is the largest eigenvalue of L^*L. (For simplicity, we have assumed that the eigenspace corresponding to σ_1 is one dimensional.)

Maximum energy growth over the interval (t_0, t) is associated with the largest eigenvalue and the corresponding eigenvectors of L^*L and LL^*: $v_1(t_0)$ at initial time and $v_1(t)$ at optimisation time.

It is useful to have a name for these eigenvectors. In the predictability literature there are many conflicting definitions in use. In order to avoid any confusion here, define

***Singular vectors* (SV)** as the eigenvectors of L^*L e.g. $v_1(t_0)$ is the initial perturbation which will have grown the most at optimisation time.

***Lyapunov vectors*(LV)** as the eigenvectors of LL^* e.g. $v_1(t) = Lv_1(t_0)$ represents the orientation at time t, of the vector which experiences maximum growth over the chosen optimisation time t.

3.2.1 Relationship to 'Normal' Mode Analysis

Normalised eigenmodes ξ_i of $Df(X(t))$ with eigenvalues λ_i give rise to solutions $\xi_i e^{\lambda_i(t-t_0)}$ of the linear perturbation equation (3). Operator L is not usually *normal* so that in general $\langle \xi_i, \xi_j \rangle \neq \delta_{ij}$ If the initial disturbance is a linear combination of the eigenmodes then the fastest growing eigenmode will eventually dominate the solution, so for sufficiently large t

$$LV = eigenmode \times phase\ factor.$$

It is possible to have decaying eigenmodes, but transient growth in the Lyapunov vectors. This is due to the lack of orthogonality in the eigenmodes. An example of this phenomenon is given in [Smith et al., 1997].

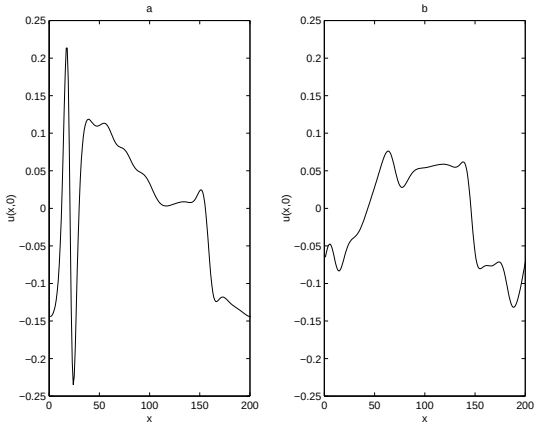

Figure 3: Panel (a) shows the initial condition (I) from the vigorous regime. Panel (b) shows the initial condition (II) from the quiet regime.

4 Results – unstable directions

Equation (1) was solved numerically using fourth-order Runge-Kutta time-stepping and centered differences in space on a 200 point grid. The linear forward tangent propagator L was calculated by linearising the numerical scheme about a fiducial trajectory and computing its effect on basis vectors. The SV and LV were calculated using matlab.

An alternative calculation for systems with a large number of variables is to code the adjoint model (see [Talagrand and Courtier, 1987]) and then use a power method to calculate the eigenvectors and eigenvalues. This was attempted here, but the results were found to be very bad due to the slow convergence of the power method.

Two sets of initial conditions were used to generate the fiducial trajectories (see Fig. 3): I) a representative from the vigorous regime of the equation and II) a representative from the quiet regime.

A plot of first singular exponents i.e. $\log(\sigma_1/t)$, is given in Fig. 4. Panel (a) (initial condition I) appears to indicate that the rescaled first singular value is asymptoting to some constant, although this is not so clear from Panel (b) (initial condition II). This constant is less than the simple scale estimate in Section 2, mainly because here the unstable region occupies only part of the domain. The differences between Panel (a) and Panel (b) indicate that the transient growth rate is highly dependent on the region of phase space from which the initial condition is taken. The following theoretical asymptotic relation links the first singular value, σ_1 to the first Lyapunov exponent, l_1 (see [Palmer, 1996] for further details).

$$l_1 = \lim_{t_{opt}\to\infty} \left[\frac{1}{t_{opt}} \log \sigma_1 \right]$$

The SV and LV are given in Figs. 5, 6, 7 and 8. Note that the unstable directions are localised to the unstable region of the domain. For the initial condition coming from the more

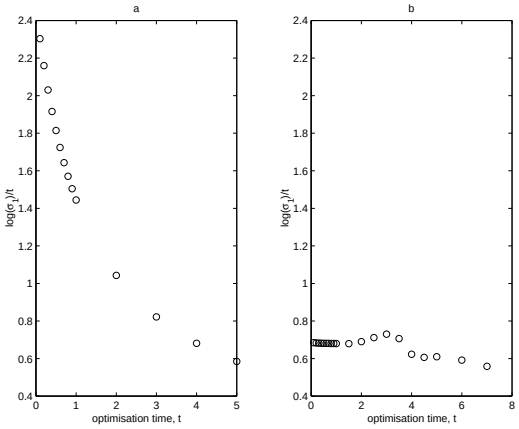

Figure 4: Rescaled singular values for (a) initial condition I, (b) initial condition II. These values provide an estimate of the fastest growth rates.

vigorous regime (I) the basic forms of the SV and LV do not change over time. The peaks broaden, but remain in the same location in physical space. For the initial condition coming from the quiet regime (II) the SV and LV show more variation with time. This may be due to faster movement of smaller disturbances through the unstable region, or to differences in evolution of the distinct steep gradients initially present in and near the unstable region.

Fig. 9 shows EV for both initial conditions, chosen from ensembles of six randomly distributed initial conditions a distance δ from the fiducial trajectory. Unlike the SV and LV the EV are not localised. Since the ensembles contained only six members there is no guarantee that the EV are in fact aligned with the most unstable directions. A larger sample and hence more computational work would be needed in order to find the EV with a greater probability of alignment with the true maximal growth direction.

5 Forecasting

Starting from an exact initial condition for a fiducial trajectory, an initial condition for a forecast trajectory is obtained by perturbing the value at each grid point with white noise of a specified maximum amplitude. Grid points are chosen at which to make 'additional measurements' using information obtained via the predictability techniques described earlier. The exact details of this choice are explained in Fig. 10. At the chosen grid points the perturbed initial condition is reset to the exact values of the fiducial initial condition. The system is then integrated forward to obtain a forecast trajectory. A forecast time of 5.0 time units was chosen since this is approximately the time taken for a large peak to grow to maximum amplitude in the growing region of the domain.

As a comparison, another strategy was also used to reset points. This was done by picking points randomly from the unstable region of the domain.

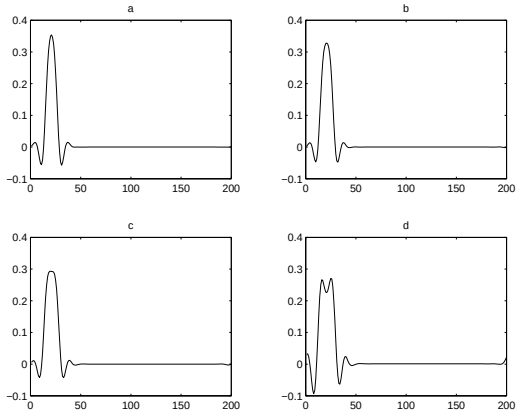

Figure 5: Singular vectors for initial condition I a) optimisation time 0.3, b) optimisation time 0.5, c) optimisation time 1.0, d) optimisation time 5.0

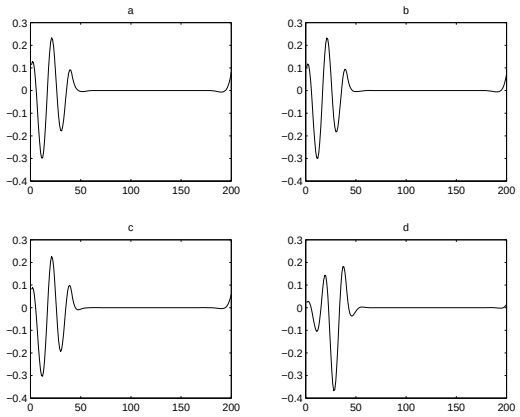

Figure 6: Singular vectors for initial condition II a) optimisation time 0.3, b) optimisation time 0.5, c) optimisation time 1.0, d) optimisation time 5.0

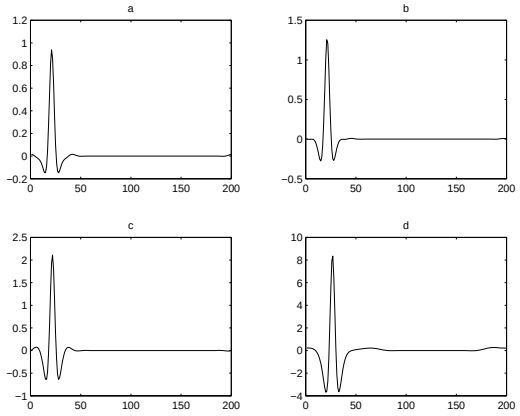

Figure 7: Lyapunov vectors for initial condition I a) optimisation time 0.3, b) optimisation time 0.5, c) optimisation time 1.0, d) optimisation time 5.0

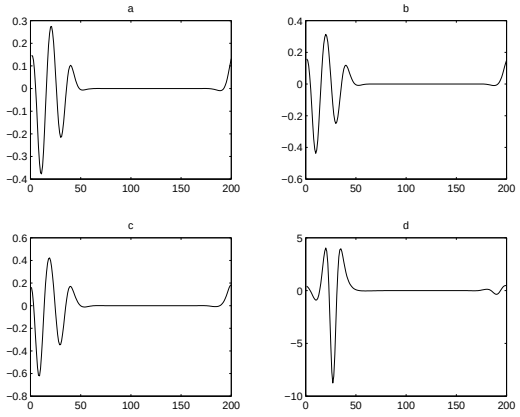

Figure 8: Lyapunov vectors for initial condition II a) optimisation time 0.3, b) optimisation time 0.5, c) optimisation time 1.0, d) optimisation time 5.0

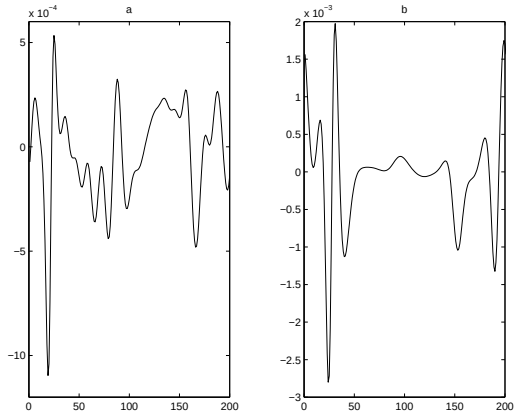

Figure 9: Ensemble vectors with optimisation time 0.3 a) initial condition I, b) initial condition II

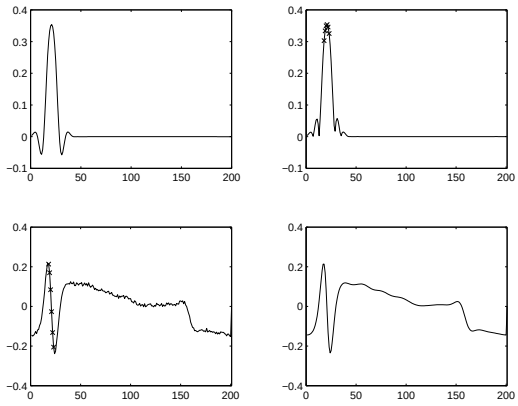

Figure 10: The plot on the top left shows an example singular vector. The absolute values of the components of this vector were taken (top right) and sorted by size. The grid points corresponding to the largest were selected (the first six are marked as x). At these grid points the values of the perturbed initial condition were reset to the exact values (bottom left) as given by the true initial condition (bottom right).

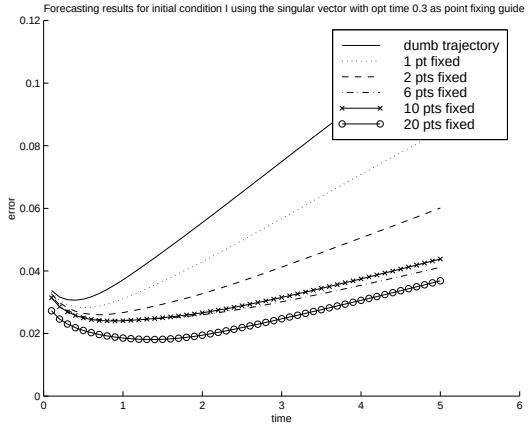

Figure 11: Variation of error with number of points reset. Note that the error is measured in Euclidean norm so that the true trajectory would be represented by the zero line on this plot. The 'dumb' trajectory corresponds to the perturbed initial condition with no points reset. Maximum initial pointwise perturbation size = 1.0×10^{-2}

A brief experiment was carried out to determine the number of points to fix in order to achieve a reasonable improvement over the forecast from the 'dumb' trajectory (the perturbed trajectory with no points reset to their exact values). The results for initial condition I with maximum initial pointwise perturbation 1.0×10^{-2} using the singular vector with optimisation time 0.3 as point resetting guide are given in Fig. 11.

For the initial condition from the vigorous regime it was found that six points were necessary to give a good result (Fig. 12). Note that since the SV did not change much with optimisation time, the first six points selected for resetting were identical for experiments using the SV associated with different optimisation times. The same is true for the LV. All the strategies employed offered an improvement over the dumb trajectory. The strategy employing the SV gave the best results.

With a larger noise amplitude of 1.0×10^{-1} the results from resetting points using SV were still good initially. However, Fig. 13 shows that eventually the errors become comparable with those for the dumb trajectory. This is after 5 or 6 e-folding times (approximated using the growth rates given in Fig. 4 so the perturbed trajectory will be so far away from the fiducial trajectory in phase space that the linear approximation is no longer relevant. Here SV are no longer valid tools.

For the initial condition from the quiet regime (II) it was more difficult to get good results. Fig. 14 shows results from resetting twelve points at initial time. One explanation for this disparity of results between the two initial conditions is that it is easier to follow the movement of a peak along a trajectory than it is to predict which small perturbation will grow in a transition between the quiet and the vigorous regimes of the equation.

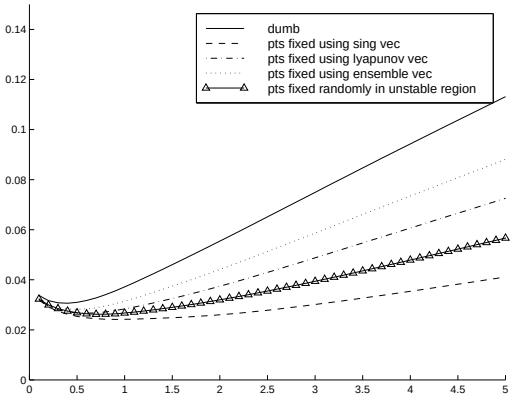

Figure 12: Forecasting results with initial condition I, time against error measured in Euclidean norm. Maximum pointwise initial perturbation size = 1.0×10^{-2}

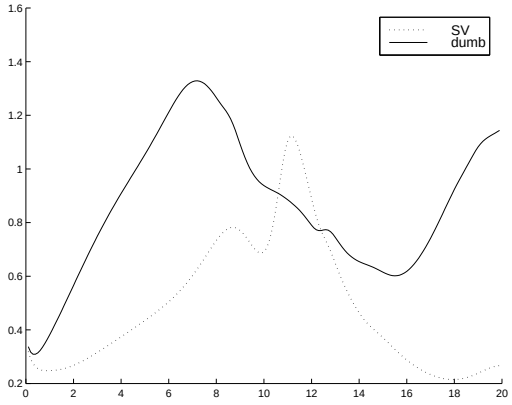

Figure 13: Forecasting results with initial condition I, time against error measured in Euclidean norm. Maximum pointwise initial perturbation size = 1.0×10^{-1}

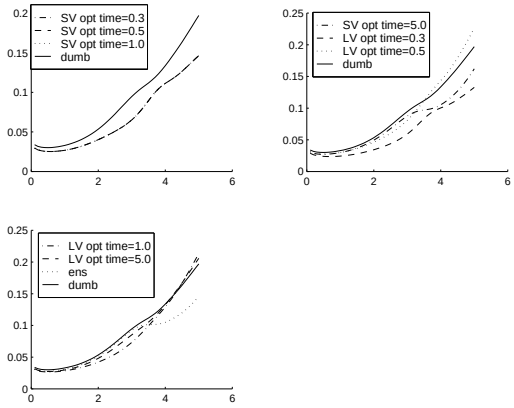

Figure 14: Forecasting results with initial condition II, time against error measured in Euclidean norm. Maximum pointwise initial perturbation size = 1.0×10^{-2}

5.1 Climatology

Suppose there is no data from the present time. A climatology can be constructed by averaging all past data. The climatology and results obtained using the climatology as the initial trajectory are given in Fig. 15. The results are fairly poor. One contributing factor is that the climatology initial condition lies in a different region of phase space to the other initial conditions studied here.

6 Conclusions

The form of the SV and LV depends on

1. the initial conditions
2. the optimisation time

Further study needs to be done in order to make this dependence more precise. For example, it has been suggested that the peak of the singular vector might correspond to the largest gradient in the initial condition. This hypothesis has not yet been tested.

In a more realistic scenario it would be necessary to calculate the most unstable directions from a noisy set of data rather than from the true initial conditions. The noisy data is likely to be less smooth than the true initial condition. If the unstable directions do have a strong dependence on the gradients in the initial condition then smoothing algorithms will have to be employed.

In general, singular vectors produced the best forecasting results. The common sense approach of picking random points in the unstable region also seemed to work quite well. The quality of the forecasting results is dependent on the fiducial trajectory – whether the initial condition lay in the vigorous or the quiet regime. The results indicate that in the quiet regime

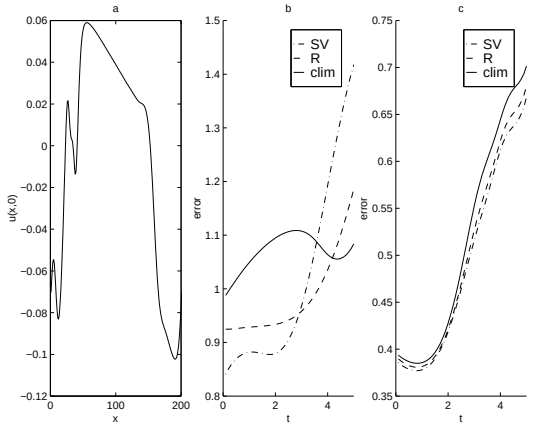

Figure 15: a) shows the initial condition for the climatology, b) and c) show forecasting results with six points reset using SV and random selection in the unstable region (R) as point picking guide. b) shows the results with initial condition I, c) shows the results with initial condition II

the most unstable direction changes rapidly along the fiducial trajectory, thus making it much harder to predict which perturbations will grow the most in the transition to the vigorous regime. This would suggest that having more accurate measurements for more than just the initial time would also be an interesting variant of the technique to study.

For practical applications it may be the case that a more accurate forecast is required only at certain locations in physical space (for example an oil rig in the middle of the ocean). For such instances forecast improvement may be offered by using a weighted norm to calculate the unstable directions (these will be different from those calculated using the Euclidean norm).

The SV and LV are only a valid tool for time-scales when the perturbed trajectories are still sufficiently close to the fiducial trajectory. The development of techniques which can follow the evolution of the most unstable directions over longer time-scales would be very valuable.

7 Acknowledgments

My thanks go to Steve Meacham for all his help, encouragement and never ending patience throughout the summer. I am also grateful to Drew Poje and Lenny Smith for their assistance.

References

[Balmforth et al., 1997] Balmforth, N. J., Ierley, G., and Worthing, R. (1997). Pulse dynamics in an unstable medium. *SIAM J. Appl. Math.*, 57(1):205–251.

[Lorenz and Emmanuel, 1998] Lorenz, E. N. and Emmanuel, K. A. (1998). Optimal sites for supplementary weather observation: Simulation with a small model. *J. Atmos. Sci.*, 55:399–414.

[Palmer, 1996] Palmer, T. N. (1996). Predictability of the atmosphere and oceans: from days to decades. *NATO ASI Series*, 1(44).

[Smith, 1995] Smith, L. A. (1995). Accountability and error in ensemble forecasting. *ECMWF Proceedings of Seminar on Predictability.*

[Smith et al., 1997] Smith, L. A., Ziehmann, C., and Fraedrich, K. (1997). Uncertainty dynamics and predictability in chaotic systems. *Quart. J. Roy. Meteorol. Soc.*, 123:1–30.

[Talagrand and Courtier, 1987] Talagrand, O. and Courtier, P. (1987). Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Quart. J. Roy. Meteorol. Soc.*, 113:1311–1328.