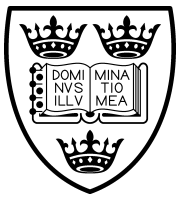
Modelling Nonlinear Dynamical Systems: Chaos, Uncertainty, and Error



David Orrell

Corpus Christi College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2001

For Beatriz

Acknowledgements

Thanks to Tim Palmer and Jan Barkmeijer, for explaining the process of weather forecasting and helping me use the ECMWF models, and to ECMWF and EPSRC for financial assistance. Thanks also to James Hansen, Patrick McSharry, Mark Roulston, and Isla Gilmour for advice and companionship. And thanks finally to my supervisor Leonard Smith, for his guidance, enthusiasm, and support.

Abstract

When nonlinear dynamical models are used to approximate physical systems such as the weather, error arises from one of two causes: the initial condition used by the model, and the model itself. Of these two sources, model error is the less well understood; yet a knowledge of model accuracy is essential for reliable error estimates and model optimisation. This thesis develops a technique for measuring model error in the context of nonlinear systems, and explores the link between model error and the ability of the model to shadow the true system. The methods are tested on a variety of model/system pairs in Chapters 2, 3 and 4. In Chapter 5, issues related to longer term behavior are studied, and connections with short term predictability explored. In Chapter 6, the model error techniques are applied to operational weather forecast models. It is seen that the component of forecast error due to model error tends to grow as the square-root of forecast time, and for the days tested is the dominant source of error out to three days. The results are summarised, and the implications further explored, in Chapter 7.

Contents

1 Introduction **2**

1.1 Outline 4

1.2 Model error vs initial condition error 5

1.3 The Rössler system 7

1.4 Spectral bifurcation diagrams 9

1.5 Other low dimension systems 10

1.5.1 The Lorenz '63 system 10

1.5.2 The Rulkov Circuit 12

1.6 The danger of low dimension systems 15

1.7 Summary 16

2 The Lorenz '96 systems **18**

2.1 The one and two level system equations 18

2.2 Behaviour of the one level system with $n = 4$ 21

2.3 One level systems with dimension 8 and 40 24

2.4 Behaviour of the two level system 29

2.5 Summary 31

3 An exploration of model error using the Lorenz '96 systems **38**

3.1 Model error vs initial condition error (continued) 39

3.1.1 What is initial condition error? 39

3.1.2 What is model error? 41

3.2 Modelling the two level system - the constant model 42

3.2.1 Measuring initial condition error 44

3.2.2 Measuring model divergence 46

3.3 Model error for the constant model 47

3.3.1 Velocity error 47

3.3.2 Forcing error for the constant model 49

3.3.3 Shadowing . . . 51
3.4 An improved model - the linear model . . . 56
3.4.1 Forcing error for the linear model . . . 58
3.4.2 Shadowing for the linear model . . . 58
3.5 The integrated forcing error - a spectral approach . . . 60
3.6 A local linear predictor model . . . 65
3.6.1 Predictor model at low forcing and high coupling . . . 65
3.6.2 Predictor model at high forcing and regular coupling . . . 68
3.7 Summary . . . 69

4 Linearised dynamics and the shadow law **72**
4.1 The linearised dynamics . . . 74
4.2 Model error vs initial condition error (continued) . . . 78
4.3 The shadow estimation technique (SET) . . . 79
4.4 A shadowing law . . . 87
4.5 The model error index . . . 93
4.6 Applications of the shadow test . . . 94
4.6.1 The Lorenz '96 systems . . . 94
4.6.2 The Rössler system . . . 98
4.6.3 The Saltzman system . . . 99
4.6.4 Shadowing and step size . . . 104
4.6.5 The Rulkov circuit . . . 106
4.7 Observed systems . . . 109
4.8 Error due to the projection . . . 112
4.9 Potential problems with the shadow estimation technique . . . 114
4.10 Errors over a subset of variables . . . 118
4.11 Fast techniques to find shadow orbits . . . 121
4.11.1 The linear approximation method . . . 122
4.11.2 The 'pinch' method . . . 124
4.12 Summary . . . 128

5 Climatology **131**
5.1 Introduction . . . 131
5.2 Chaotic in the small, predictable in the large . . . 132
5.3 Modelling the climatology of the two level system . . . 138
5.3.1 Stochastic models . . . 140
5.3.2 Projection on EOF's . . . 141

5.4 Summary 145

6 Operational weather forecasting models 147

6.1 Causes of model error 148
6.2 The perfect model assumption 148
6.3 Multi-model ensembles 152
6.4 Problems with the ensemble approach 153
6.5 Error between models of different resolution 155
6.5.1 The range of ECMWF models 155
6.5.2 The energy metric 156
6.5.3 Forecast errors 158
6.5.4 Calculation of the drift 158
6.5.5 Shadowing 163
6.5.6 Ensemble calculation 164
6.6 The ECMWF operational model 165
6.7 Modelling the model error 174
6.8 Model error in the 500 hPa height 176
6.9 Summary and discussion of results 178

7 Conclusions 182

7.1 Summary of results 182
7.2 Does chaos matter? 184

A Proof of the shadow law 188

B Glossary 191

List of Figures

1.1 Plots of x versus time, y versus x and power spectra versus frequency for various values of c for the Rössler system. 8
1.2 Bifurcation diagram for x as a function of c in the Rössler system. Upper panel is a density plot, middle panel is density of local max/min, and lower panel shows power spectra. 11
1.3 Bifurcation diagram for the Lorenz '63 system, from $F = 25$ to 275. . 13
1.4 Bifurcation diagram for the circuit model. 14
1.5 Zoomed view of bifurcation diagram for the circuit model. 15

2.1 System variables 19
2.2 Time series of x_1 for the one level system given by equation 2.1 with $F = 10$. 20
2.3 Time series of $\tilde{x}_1$ and $\tilde{y}_1$ versus time for the two level system given by equation 2.2 with $F = 10$. 21
2.4 Plots of x_1 versus time, x_2 versus x_1 and log (base 10) power spectra versus frequency for various values of F for the single level system with $n = 4$. 23
2.5 Bifurcation diagram for the one level system with $n = 4$. Upper panel is a density plot of x_1, middle panel is density of local max/min, and lower panel shows the spectral bifurcation diagram, as introduced in the text. 25
2.6 Bifurcation diagrams for the one level system with $n = 8$. Upper panel is the x_1 density plot, middle panel is the density of local max/min, lower panel is the spectral bifurcation diagram. 27
2.7 Period doubled orbit at $F = 3$ for the one level system with $n = 8$. . . 28
2.8 Periodic orbit at $F = 5.235298$ for one level system with $n = 8$. . . . 28

2.9 Close up of spectral bifurcation diagram in region of periodic orbit, for one level system with $n = 8$. Spacing of the harmonics at $F = 5.235298$ is about 0.27, corresponding to a frequency of 36.7 29
2.10 Bifurcation diagrams for one level system with $n = 40$. Upper panel is the x_1 density plot, middle panel is the density of local max/min, lower panel is the spectral bifurcation diagram. 30
2.11 Bifurcation diagram for the 40D two level system, $\tilde{x}$ variables. The ripple effect noticeable in the spectral bifurcation diagram (lower panel) is due to numerical error in the power spectrum routine. 32
2.12 Bifurcation diagrams for 40D two level system, $\tilde{y}$ variables. 33
2.13 Bifurcation diagrams for 40D two level system, $\tilde{y}$ variables, as a function of the coupling coefficient h. 34
2.14 Trajectories of 40D two level system, plotted in $\tilde{x}_1$ and $\tilde{y}_1$ variables, at $F = 4.5$ and 4.6. At the higher forcing, the system is periodic. At the lower forcing, both variables are chaotic, but this is most evident in the $\tilde{y}_1$ orbit. 35
2.15 $\tilde{x}$ and $\tilde{y}$ orbits for true system with $F = 2$ and $c = 1.2$. 36

3.1 Plot showing how model error (dashed line) dominates error due to displacement of initial condition (solid line) for small times, providing the initial error is sufficiently small. Results are RMS errors for an actual model and system (see also Figure 3.4). 43
3.2 Forecast errors for Lorenz model/system, x_1 component. A number of short term model forecasts were initiated at regular intervals along a true trajectory. The errors have been scaled by a factor 20. The total error magnitude over all x_i is also shown, again scaled by a factor 20. Because it contains all components, it is larger than the x_1 error. . . . 44
3.3 Ensemble with initial $\tilde{x}$ displacement standard deviation 0.01 for $F = 10$ true system. The $\tilde{y}$ variables are not perturbed. 45
3.4 Bottom line (solid) shows initial condition displacement error $e_d(t)$ for the true system with $F = 10$. Middle line (long dash) shows $e_d(t)$ for the constant model with $P^c = 9.62$. Top line (short dash) shows model error $e_m(t)$ for the constant model. Note that displacement error grows exponentially at small times, while model error grows linearly. 46
3.5 Initial linear growth of model error for constant model, $P^c = 9.62$, with respect to true system with $F = 10$. 47

3.6 Model error curves for constant model. Order from bottom line to top line is $F = 4, 6, 8, 10$. Error is roughly linear in the early stages. . . . 48
3.7 Initial slope of model error as a function of F. The constant model is near-perfect for $F \leq 1.3$, for which the $\tilde{y}$ variables are periodic. Computational errors are estimated to be about 0.02 in the vertical scale. . . . 48
3.8 Spectral bifurcation diagram of forcing error $\mathbf{F_e}$ for the constant model. Calculations are for one component only; other components are the same by symmetry. As for model variables $\mathbf{x}$, the forcing shows a mix of periodic and chaotic behaviour. . . . 50
3.9 Total power (variance) and standard deviation of forcing error for the constant model. . . . 51
3.10 Schematic diagram showing true orbit (black line starting from x_0) with model shadow orbit (grey line starting from s_0). The model shadows for a time τ. A model trajectory starting at x_0, which shadows a shorter time, is also shown. . . . 53
3.11 Plot of x_1 for a typical constant model shadowing orbit. The shadowing radius is 0.4. The orbit ceases to shadow at $t = 0.5$ (note the radius is over all components of x, not just x_1). . . . 54
3.12 Histogram of shadowing times for the constant model with $F = 10$ over 200 starting points. . . . 55
3.13 Mean shadowing times for constant model at integer values of F, with shadowing radius $r_s = 0.2$ and 0.4 at $F = 10$, and scaled for other values of F. Averaged over twenty initial conditions. . . . 55
3.14 Total shadow forcing errors for the constant model for integer values of F, normalised for shadowing radius (dimensionless). . . . 56
3.15 Scatter plot of local forcing versus local $\tilde{x}$ at $F = 10$. Also shown is the linear interpolation used in the corresponding linear model. . . . 57
3.16 Values of constant α_0 minus F, and slope α_1 for linear model as a function of F. The spike around $F = 1.3$ occurs when the $\tilde{y}$ variables become non-zero. . . . 58
3.17 Model error and displacement error for linear model with $F = 10$. . . 59
3.18 Model error near time zero for constant and linear models for $F = 10$. 59
3.19 Forcing error standard deviation for constant and linear models as a function of F. . . . 60

3.20 Histogram of shadowing times for the linear model with $F = 10$ over 200 starting points. 61
3.21 Shadowing times for linear model at integer values of F. Shadowing radius is 0.2 and 0.4 at $F = 10$, scaled at other values. 61
3.22 Plot of x_1 for a linear model shadowing orbit, with $F = 10$ and shadowing radius $r_s = 0.4$. The orbit cease to shadow at $\tau = 2.49$. Note the shadowing radius is over all components of $\mathbf{x}$, not just x_1. 62
3.23 Weighting function $W(\omega, \tau)$ versus frequency ω for constant model with shadowing time $\tau = 2\pi$. 63
3.24 Total shadow forcing errors for linear model, normalised to shadow radius (dimensionless). Shadowing radius is 0.2 and 0.4 at $F = 10$, scaled for other values. 64
3.25 Power spectra of forcing error for constant and linear models at $F = 10$. 65
3.26 Shadowing orbits for constant, linear and predictor models with $c = 1.2$ and $F = 2$. 66
3.27 Plot of local forcing $\tilde{F}_i$ vs $\tilde{x}_i$ for true system with $c = 1.2$ and $F = 2$. 67
3.28 $\tilde{x}_2$ versus $\tilde{x}_1$ in the learning set. 67
3.29 Plot of local forcing $\tilde{F}_i$ versus $\tilde{x}_i$ in the learning set. 68
3.30 Model error for predictor model in early stages, compared with constant and linear models. 69
3.31 Plot of integrated forcing error versus shadow time for constant, linear and predictor models, and expected limits for shadow radius 0.2 and 0.4 70

4.1 Shadowing orbit for linear model, in polar coordinates with radius the displacement from true orbit, and angle with respect to initial offset. The circle of radius 0.4 represents the shadowing radius. The shadow orbit starts near the extreme radius 0.4 on the right hand side, then exits on the upper left hand side after 2.49 time units. 73

4.2 Forecast errors for Lorenz model/system, x_1 component. As in Figure 3.2, a number of short term model forecasts were initiated at regular intervals along a true trajectory. The errors have here been scaled by a factor 10. The total error magnitude over all x_i is also shown, again scaled by a factor 10. Because it contains all components, it is larger than the x_1 error. The drift, also shown scaled by a factor 10, is the vector sum of the total velocity errors. The almost linear increase up to $t = 0.6$ indicates that the error vectors are accumulating, and are therefore in a similar direction. Above $t = 0.6$, however, the drift begins to reduce because the velocity error has rotated away from its original orientation, and projects negatively onto the drift. 77

4.3 Error growth of perturbed initial conditions for the one level model. In the upper panel, errors are relative to the unperturbed model control. The background contours show error growth for 1000 random perturbations of magnitude 0.2, while the dashed and dotted lines show the perturbations in the positive and negative directions of the leading singular vector. The optimisation time for the singular vector calculation is 0.34; perturbations in these directions give maximum growth at that time. The middle panel shows errors relative to a trajectory of the two level system, and therefore include the effect of model error. The error of the model control is also shown. The lower panel is a zoom of the middle panel near initial time. Note the effect of model error on the ensemble, and the fact that no ensemble member offsets model error over the optimisation time. 80

4.4 Schematic diagram showing how the angle between the positive and negative perturbations in an ensemble can shrink when taken with respect to truth (θ_t) as opposed to the model (θ_m). See also [23], [24] 81

4.5 Normalised dot product (or cosine angle) of the positive and negative singular vector perturbations for the Lorenz model/system, with respect to the true system (dashed line) and the model (solid line). . 82

4.6 Schematic diagram of linearised shadow dynamics, showing the initial ball of radius r about the starting point; the final ellipse at the shadow time with axes $r\sigma_i\mathbf{u}_i$; the drift vector $\mathbf{d}$; and the optimal initial displacement $\mathbf{e}$ which satisfies $\|\mathbf{Me} + \mathbf{d}\| = r$. 84

4.7 Plot of the dissipation coefficient $q(\tau)$ when the singular value multipliers follow a simple power law distribution as described in text. Shown as a function of the leading multiplier. The maximum value is 1.0. As an upper bound on shadow times, the shadow law therefore sets drift equal to twice shadow radius, which equates to a dissipation coefficient of 1.0. 91
4.8 Upper panel shows singular value multipliers, arranged in descending order, for the constant and linear models. Results are averaged over 200 shadow runs at shadow radius 0.4. The + marks the centre point: for either model, more singular value multipliers contract than expand. The centre panel shows the distribution of the dissipation coefficient q over the same shadow runs. The lower panel shows the distribution for the constant model. The constant model is less dissipative than the linear model, largely because it shadows for a shorter time. 92
4.9 Plot of estimated and mean shadow times for constant model, averaged over 20 runs. Shadow radius is 0.2 and 0.4 at $F = 10$, scaled proportionately for other values. 95
4.10 Plot of estimated and actual shadow times for linear model, averaged over 20 runs. Shadow radius is 0.2 and 0.4 at $F = 10$, scaled for other values. 95
4.11 Histogram of 200 shadow times for linear model with $F = 10$, shadow radius 0.4. Upper is real shadow times, lower is estimated times. . . . 96
4.12 Plot of displacements from truth, in first two coordinates, for images under the linear model (full dynamics) and the propagator (linearised dynamics). The accuracy of the linearised dynamics depends on proximity to the true orbit. For the ball of initial conditions at radius 0.4, top panel, the linearised dynamics do not represent the full dynamics, but they are much more successful in the lower panel, where the ball of initial conditions has been shrunk to 0.02. 97
4.13 Shadowing orbit for linear model of two level system with coupling coefficient reduced to 0.5. The reduced effect of the fine-scale $\mathbf{y}$ variables on the large-scale $\mathbf{x}$ variables allows for long shadow times, in this case about 10.1 time units. Shadow radius is 0.4. 98

4.14 Drift calculated over the orbit shadowed in Figure 4.13. The SET fails to predict shadow times in this instance, because the drift remains small and higher order terms, not modelled by the linearised dynamics, eventually dominate. 99
4.15 Real and estimated shadow times for the modified Rössler system. Shadow radius is 0.5 at $F = 10$, and scaled for other values. 100
4.16 Comparison of attractors for the 7-D Saltzman system (upper panels) and the reduced 3-D model (lower panels), for $\lambda = 28$ and 25.1. The model appears to be in better agreement with the system for the lower value of λ. 102
4.17 Analysis of shadowing performance for the 7-D Saltzman system and the reduced 3-D model, for $\lambda = 28$ (left side) and $\lambda = 25.1$ (right side). Shadow radius is 0.1. Upper panels are scatter diagrams of estimated and actual shadow times from 40 initial conditions. Middle panels are histograms of shadow times (note different time scales on left and right). The majority of the estimated shadow times are in good agreement with actual times, though for longer times the SET tends to overestimate the time. Lower panels show the location on the attractor of the longest shadow orbit. 103
4.18 Shadow orbits for the 3-D model of the 7-D Saltzman system, at $\lambda = 28$ and 25.1. Model displacements from truth are projected into a 2-D coordinate system following the true orbit. The results are plotted with time as the third axis. For $\lambda = 28$, the shadow trajectory begins at the left hand side ($t = 0$), oscillates with increasing magnitude around the true orbit, and suddenly fails a short time after switching from one lobe of the attractor to the other. The $\lambda = 25.1$ model shadows a longer time (note the different time scale), apparently because the true system spends more time in one lobe of the attractor. 105
4.19 Diagram showing how too large a step size creates drift. The model, which has twice the step size of the true system, overshoots it. The resulting drift is proportional to the acceleration of the true system. . 106
4.20 Plot of points with local drift ≥ 0.2, and points with shadow times ≤ 0.1, for Lorenz '63. True system has time step 0.01, model has time step 0.02. Errors occur in regions of high acceleration. 107
4.21 Histogram of shadow times for Lorenz '63. True system has time step 0.01, model has time step 0.02. 107

4.22 Distribution of the final points of the model to shadow observations from the Lorenz equations (plus signs). A trajectory of the Lorenz equations (solid line) is also shown for reference. True system has time step 0.01, model has time step 0.02. From [23]. 108
4.23 Circuit errors. From [63]. 109
4.24 Zoom of circuit errors. From [63]. 110
4.25 True system has a stable periodic orbit with radius 1.0, model has periodic orbit with same angular velocity at radius 0.8. 115
4.26 Forecast errors for x component. As in Figure 3.2, a number of short term model forecasts were initiated at regular intervals along a true trajectory (shown unscaled). The total error magnitude over both x and y is also shown, as well as the drift, which accumulates steadily. . 116
4.27 Drift calculated with an integration step of 0.01 and 0.5. The drift reduces with the higher step size. 117
4.28 Forecast error, shadow diameter and drift as a function of prediction time. 117
4.29 Upper panel compares forecast error (solid line), drift (dashed line), and magnitude of the difference between the error and drift vectors (dotted line) for the Lorenz constant model. Lower panel shows the cosine angle between the drift and error vectors. 119
4.30 Forecast errors (solid lines) and drift (dotted lines) compared for each component. 120
4.31 Histogram of shadow times using the Newton method, linear model. . 127
4.32 Histogram of shadow times using the pinch method, linear model. . . 127
4.33 Average ratio of drift to shadow diameter for maximal shadow orbits. Shown are the constant and linear Lorenz models from Figures 3.14 and 3.24 as a function of forcing; the predictor model; the low coupling model; the modified Rössler model from Figure 4.15; and the Saltzman 3D model at $\lambda = 25.1$ and 28. Results are determined by averaging over 20 shadow runs, except for the predictor model and low coupling model where only a single run was tested. All models conform to the shadow law (though these are average rather than RMS results). The low coupling model has very low drift because the extremely long shadow orbits are dominated by nonlinear effects. 130

5.1 Spectral bifurcation diagrams for constant model, linear model, and true system, and model mis-match with true system. 133
5.2 Mean of $\tilde{x}_1$ for the two level system as function of F. Also shown is the mean of x_1 for the constant model (dashed line), with forcing chosen to give the correct ratio $\frac{\langle x_i^2 \rangle}{\langle x_i \rangle}$. The model curve is always below that of the system. 136
5.3 Mean of $\tilde{x}_1^2$ for the two level system as function of F. Also shown is the mean of x_1^2 for the constant model (dashed line), with forcing chosen to give the correct ratio $\frac{\langle x_i^2 \rangle}{\langle x_i \rangle}$. The model curve is always below that of the system. 136
5.4 Ratio of $\langle x_i \rangle$ to $\langle \tilde{x}_i \rangle$ for values of the linear model offset $\alpha_0 - F$ where $F = 10$. The linear model gives $\langle x_i \rangle \approx \langle \tilde{x}_i \rangle$ for offsets near zero. The corresponding slope agrees with that of the linear model used for shadowing purposes. 138
5.5 Plot of $\langle x_i \rangle$ for the linear model, compared to $\langle \tilde{x}_i \rangle$ for the system, over a range of system forcings. Agreement is excellent except for the region of $F = 1.3$ and $F = 7$. 139
5.6 Plot of $\langle x_i^2 \rangle$ for the linear model, compared to $\langle \tilde{x}_i^2 \rangle$ for the system, over a range of system forcings. Agreement is excellent except for the region of $F = 1.3$ and $F = 7$. 139
5.7 EOF's for true system. The horizontal axis is the index i of the variables x_i around the circle. 142
5.8 Histogram of modulus of orbit projected onto EOF's 1 and 2 (vertical axis) and EOF's 3 and 4 (horizontal axis) for true system and models. 143
5.9 Climatology error, expressed by calculating projection of true system and model onto EOF's, then plotting the difference. Vertical axis represents EOF's 1 and 2, horizontal axis EOF's 3 and 4. 144
5.10 Power spectrum of orbit projected onto first EOF for true system and models. 145

6.1 Model errors in the 500 hPa height norm. Left column shows analysis heights for days 1 to 4, middle column shows predicted heights, right column shows errors. Contour interval is 25 for the heights and 2 for the errors. 156
6.2 Errors for temperature T integrated over Europe, for a typical week. . 157

6.3 The upper panel shows T42 (dashed) and T63 (dotted) forecast errors at starting dates 15/10/99,15/11/99,22/12/99,15/1/00,15/2/00. Errors are computed in the energy norm, relative to TL159. The lower panel shows the ratio of the magnitudes of T42 and T63 errors, where each is again taken relative to TL159. Also shown is the ratio of the drift magnitudes. 159
6.4 The upper panel shows a plot of T42 errors with respect to TL159 for starting date 15/10/99. Solid line is the forecast, dashed line is the shadow trajectory which minimised error at 48 hours, dotted line is the drift. The uncorrected drift is in places larger than the forecast error, due to a spin-up error which is probably caused by truncation error. The two estimates for the lower bound on shadow radius, computed using estimates of the drift, are shown by the shaded region. Errors are computed in the energy norm, relative to TL159. The lower panel shows the same for T63. 160
6.5 Comparison of drift for T63 vs TL159, computed with step sizes of 3, 6 and 12 hours. 161
6.6 Upper panel shows plot of errors wrt T42 control for a T42 ensemble generated from leading singular vectors at 1999/10/15, 12 hours GMT. Centre panel shows errors with respect to a TL159 control for the T42 forecast (solid line), the ensemble (dotted) and the ensemble mean (dashed). The lower panel shows the errors which would occur if the error vectors in the upper panel are added to the control error, assuming orthogonality. It can be compared with the centre panel. . . 166
6.7 The left hand panels are as for Figure 6.6, but only the first two ensemble members are shown. The right hand panel shows the corresponding figures for the Lorenz system, from Figure 4.3. Solid line is the control, dashed is the mean, dotted are the two ensemble members. 167
6.8 Plot of TL319 forecast error (solid line) with respect to analysis. Also shown is the drift (dotted). The drift is calculated using a time step of 6 hours for the first day and 24 hours for days 2 and 3. 169
6.9 Upper panel shows the cosine angle for 24 drift vectors at consecutive days over a 100 day period from 15 Oct 1999. Lower panel shows the same for 100 randomly chosen pairs of days from the same period. . . 170

6.10 Magnitude and cosine angle of consecutive 6 and 12 hour drift vectors over a 30 day period from 15 Oct 1999. In the upper left panel, 6 hour drift vectors fall to the left of the dashed line when initiated at times 0 or 12 hours GMT, and to the right for times 6 and 18 hours GMT. The 12 hour drift vectors for the lower panels are formed from two consecutive 6 hour vectors. 171
6.11 Plot of TL319 forecast errors (solid lines) with respect to analysis at four different starting dates, along with the theoretical approximation for model error from 24 hour drifts (dashed) and 6 hour drifts (dotted). The dates used were 1999/10/15, 1999/12/22, 2000/01/15, 2000/02/15, all at 12 GMT. 172
6.12 Plot showing how forecast error is consistent with a combination of model error and displacement error. The + signs shows TL319 RMS forecast errors, with respect to analysis, over five different starting dates. Solid line is the theoretical approximation for model error from equation 6.4. Dotted lines at bottom show series of displacement error curves initiated by the model error after each 6 hour period. Dot-dash line is the sum of the displacement error curves, assuming orthogonality. The dashed line which closely matches the data is the sum of the model error and displacement error curves, again assuming orthogonality. The dates used for the forecast error were 1999/10/15, 1999/11/15, 1999/12/22, 2000/01/15, 2000/02/15, all at 12 GMT. 175
6.13 Plot of TL319 RMS 500 hPa height errors (solid line) with respect to analysis over ten different starting dates, along with the drift (dashed line) and the theoretical approximation for combined model and displacement error (dotted line). The ten forecasts were launched each ten days from 12 GMT on 1999/10/15. 178
6.14 Magnitude and cosine angle for 24 drift vectors at consecutive days over a 100 day period from 15 Oct 1999 in the 500 hPa height norm. 179

7.1 Butterfly. Species *Lorenzus*. First spotted 1963. Incubation period 30 years. 186

A.1 Plot of f and h, which are defined in the text, as a function of the singular value σ_1 over the range for which f is non-negative. 190

What a chimera, then, is man! what a novelty, what a monster, what a chaos, what a subject of contradiction, what a prodigy! A judge of all things, feeble worm of the earth, depositary of the truth, cloaca of uncertainty and error, the glory and the shame of the universe!

Blaise Pascal (1623-1662)

Mankind always sets itself only such problems as it can solve ...

Karl Marx (1859)

Chapter 1

Introduction

Nonlinear dynamical systems are frequently employed to model complicated phenomena, be they an electronic circuit, the trajectory of a spacecraft, or the lifecycle of beetles. One of the most important applications, in terms both of financial expenditure and impact on people's lives and businesses, is the forecasting of the weather. The models in this case contain up to tens of millions of variables, which calculate the future state of the atmosphere from its present measured state to give us our daily forecast.

The models essentially represent an understanding of how the atmosphere operates. Since certain aspects of the weather, such as cloud formation, are too complicated or fine-scale to represent, a parameterisation must be employed. The modeler is also constrained by computer resources; even with the world's most powerful supercomputers, resolution in space and time is still limited. For example, the spacing of the horizontal grid is of the order of 50 kilometres, which is large compared to a thunder storm. As a result, all weather models have one feature in common: they are different from the real weather. In other words, they all contain model error.

Until now, there has been no satisfactory method to calculate the effect of model error on predictability. Techniques which offer partial solutions include measuring the divergence of the model trajectory from observations, and comparing different models. The first approach suffers from the problem of entangling model error with initial condition errors, since as soon as the model diverges slightly from observations, the two are hard to distinguish. This is especially an issue in chaotic systems, where sensitivity to initial condition may be large. The second approach is useful for a comparison of specific models, but fails to give a bound on model error because it doesn't fully sample the space of available models. Models are built up from broadly similar principles, so the fact that a European weather model agrees well with an

American model doesn't imply that they are both accurate - they could both be wrong in similar ways. (A third approach is to randomly perturb model parameters to give an idea of the likely spread, but this is just a variation on the second method, and it isn't clear that any of the perturbed models need be close to the real weather.)

This difficulty in measuring model error is a problem. Measurement of error is fundamental both in science and engineering. Without it, the modeller must rely on intuition or guesswork to improve the models [60]. The problem is especially striking in the context of weather forecasting - a multi-billion dollar industry, yet one, it might be said, without the means to control the quality of its own product; and, as a result, plagued by more than one company which claims to be able to predict the weather out to a year or beyond.

The aim of this thesis is therefore to provide a method for measuring model error in nonlinear dynamical systems and assessing its effect on predictability. To this end, we will address three practical topics, primarily in the context of short to medium range prediction. The first topic is, how do we define model error? For example, how do we decide which of various weather models is the most accurate? Or, if an ensemble of models is being used, with the results in some way averaged over all of them, how do we assign weights to the different models? And how do we compare the magnitude of model error with the likely error in initial condition?

The second topic is, how can we estimate how long a model will shadow (stay close to) a system or set of observations? Many techniques in weather prediction implicitly assume that there exists a model orbit which shadows the true system for some specified time. For small systems it is easy to find actual shadow orbits using optimisation techniques, but for large atmospheric models the computation is difficult, and shadow times are unknown, with expert opinions ranging between several hours and many months. Therefore, how can we cheaply estimate shadow times for a particular model? Can we tell if the model is good enough for ensembles, comprised of a collection of perturbed forecasts, to encapsulate reality?

Finally, given a model with a number of parameters which can be varied, what criterion should be used to optimise the parameters, so that predictability is maximised? How do we know the model is the best that can be achieved, up to a change of structure? And, armed with a knowledge of model error, can forecast accuracy be improved by other means?

1.1 Outline

We will study model error over a range of systems, ranging from simple 3D systems, to the higher dimension Lorenz '96 [40] systems, right up to operational weather models with dimensions of the order 10^7. The first chapter introduces the problem of weather prediction, and the interaction between model error and initial condition error in the context of nonlinear dynamical systems. It also provides some background on nonlinear systems, and introduces the spectral bifurcation diagram. This is a new method of visualising system behaviour through the use of spectral analysis of time series, which will be useful for the higher dimension systems encountered later. Chapter 2 presents the Lorenz '96 systems, which can be viewed as highly idealised atmospheric models, and their behaviour over a range of parameter values is explored using spectral bifurcation diagrams and other tools. The Lorenz systems are particularly suited to the study of model error, and this is a feature exploited in Chapter 3, which develops a formal theory of model error, building up from observed behavior of the Lorenz systems, and arriving at a new measure of model error which is applicable for any model/system pair. In the fourth chapter, a computationally inexpensive technique for estimating shadow times, based on the model error work, is developed, and a shadow law, which underpins the link between model error and shadow times, is presented. The fifth chapter investigates climatological considerations and stochastic models. Links between the optimisation of model climatology and of short term predictability are also explored.

In Chapter 6, the theoretical results developed thus far are applied to weather models at the European Centre for Medium-Range Weather Forecasts (ECMWF). A range of model error and shadowing results are presented for different resolution models, including the operational model. A simple formula for forecast error, which predicts the magnitude of the error for times up to three days, is developed. Finally, in Chapter 7, some future directions of research in this area are proposed, including the possibility of improving forecasts through a greater understanding of model error.

Most of Chapters 3 through 6 represent new work. A summary of the main contributions is as follows:

- A new method of presenting bifurcations using the power spectrum
- A detailed bifurcation analysis of the Lorenz '96 systems
- A method to measure model error independent of initial condition error

- A technique for estimating shadow times
- A simple method to determine an upper bound on shadow times for dissipative models
- Methods for comparing the relative magnitudes of model and initial condition error
- Proofs of fundamental results concerning climatology of the Lorenz systems
- The first estimate of shadow times for weather models
- A formula for predicting the magnitude of forecast error up to three days

1.2 Model error vs initial condition error

A recurrent theme of this thesis will be the relationship between model error and *displacement* error, defined as the error due to the model equations being evaluated at the wrong point. When displacement error occurs at initial time, it is referred to as *initial condition* error. To successfully measure model error, it must somehow be isolated from displacement error.

The influence of the two types of error was cited in the context of weather models by Bjerknes [4], who pointed out in 1911 that the ability to successfully predict the weather requires two things: a sufficiently accurate model, which, given an initial condition, will correctly compute the atmospheric state at the future time; and a knowledge of what that initial condition is. These statements reflect the Laplacian ideal [42] that, if we knew the present state space coordinates of a system, and the forces acting on it, we could predict its future.

A typical atmospheric model can be written as a differential equation of the form

$$\frac{d\mathbf{x}}{dt} = \mathbf{G}(\mathbf{x}), \qquad \mathbf{x}(0) = \mathbf{x_0} \tag{1.1}$$

where the vector $\mathbf{x}$ represents atmospheric variables, the initial conditions at time $t = 0$ are $\mathbf{x_0}$, and the velocity of $\mathbf{x}$ at any time is governed by the function $\mathbf{G}$. Analytical solutions for (1.1) are not generally available, but a numerical solution can be determined by integration. The problem of weather prediction then reduces to knowing the correct initial conditions $\mathbf{x_0}$, and having an appropriate model $\mathbf{G}$.

Now, the equations governing the atmosphere are nonlinear, and therefore capable of showing chaotic behaviour. The hallmark of chaos is sensitivity to initial

conditions, so small variations in $\mathbf{x_0}$ can quickly lead to radically different solutions. This was famously illustrated by Lorenz [37], who encountered it when he discovered that rounding off the initial conditions of his truncated convection model completely changed the solutions. The idea that the atmosphere was a chaotic system soon became enshrined in public lore (somewhat conveniently for forecasters!). What hope is there to know if it will rain on the weekend, if a butterfly somewhere in Brazil could flap its wings and stir up a storm?

It is certainly true that all observations of the weather have a degree of error, and since only a finite number of observations are possible, we never know the exact state of the atmosphere at any given time: the vector $\mathbf{x_0}$ is known only to within a certain tolerance. Therefore, becasue of sensitivity to initial condition, a single run of the model will soon stray from the true path as it is integrated forward in time.

Much effort has gone into addressing the problem of sensitivity to initial condition, and the major weather centres have developed methods of generating ensembles of initial conditions, comprised of perturbations around the observations, all of which are run forward using the model [44, 67, 49]. Statistical statements about the future weather can, in theory, be deduced by examining the ensemble of final states [7].

While chaos makes prediction difficult, it also obscures the effect of model error. As soon as a forecast state diverges from the true weather state, displacement error kicks in. And since there is always some uncertainty about the initial condition, due to observation error and truncation to model resolution, it is hard to separate model error from displacement error even for small times. But that doesn't mean that its contribution can be ignored.

For example, confidence in the ensemble approach would be improved if the model could shadow [23] the true solution, i.e. if there existed some initial condition within the ensemble radius ϵ around $\mathbf{x_0}$ which remained within a tube of radius ϵ of the true solution as it was integrated forward. This would certainly be the case if our model (1.1) was a perfect description of the atmosphere. Suppose, though, the model is flawed (a more likely possibility!), and no such shadowing orbit exists past a time τ. Then no matter what technique we use to generate an ensemble, statistical conclusions drawn by examining the behaviour of the ensemble past that time will be affected, if not made invalid, by model error.

The problems of initial condition error and model error are therefore coupled, and it is impossible to discuss predictability of any system without assessing the effects of model error. We know what happened when Lorenz rounded off his initial conditions,

but how about when he truncated his equations from the full convection model in the first place?

Matters are further complicated by the fact that we are dealing with nonlinear systems. Even simple nonlinear systems are capable of showing highly complex behaviour, and the effect of altering a particular parameter (which is one example of model error) can be hard to analyse. It is therefore necessary to understand some basic properties of nonlinear systems before exploring the topic of model error. Because the systems to be studied are high dimension and fairly complicated, we shall first illustrate some properties of nonlinear dynamical systems, as well as the tools which are used to analyse them, in a simple system due to Rössler.

1.3 The Rössler system

The Rössler system [58] is given by the equations

$$\begin{aligned} \frac{dx}{dt} &= -y - z \\ \frac{dy}{dt} &= x + ay \\ \frac{dz}{dt} &= b + (x - c)z. \end{aligned} \tag{1.2}$$

The constants a and b are here set to 0.1, while c will be treated as a parameter which can be varied. An advantage of this system is that it is very simple - there is only one nonlinear term - and is easier to visualise than the higher dimensional systems which we will come to later. It will also serve as a basis for comparison for those systems, in terms of both similarities and differences.

A standard method of studying the behaviour of dynamical systems is to look for attractors, which can either be a fixed point, a periodic orbit, or a chaotic 'strange attractor' [26]. Some of the attractors for the system (1.2) for various values of c are shown in Figure 1.1. The left hand column is a time series of x, the centre column is y versus x, and the right hand column is a power spectrum of x. Note that not all of the three dimensions are shown: it isn't always necessary to plot the attractor in the full dimension of the space to understand its structure. This will prove useful when we go on to look at 40 dimensional systems! Computations were performed using a fourth order Runge-Kutta scheme [52] with step size 0.01.

For low values of c, such as $c = 3$, the attractor in the xy plane is a periodic orbit consisting of a single loop. The power spectrum consists of a base frequency of

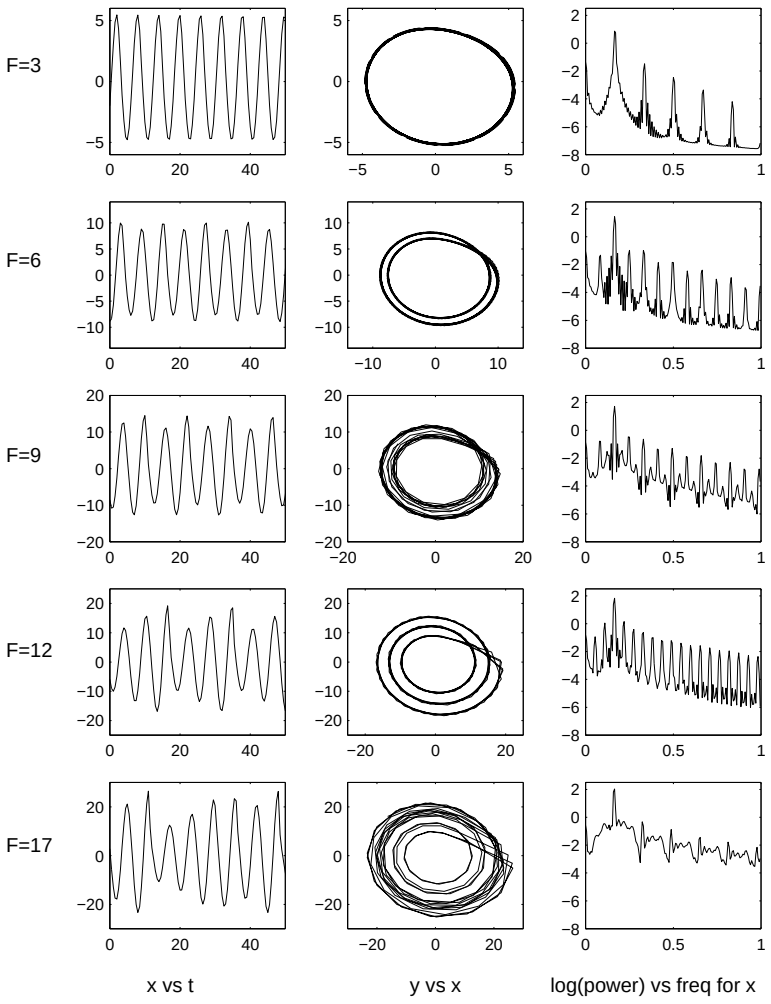


Figure 1.1: Plots of x versus time, y versus x and power spectra versus frequency for various values of c for the Rössler system.

about 0.16, plus its integer multiples (as required for periodicity). As c is increased, the system passes through a bifurcation point, and the single loop becomes a double loop, as shown for $c = 6$. The power spectrum picks up an extra base frequency of about 0.08, and its multiples. Further period doubling bifurcations ensue, at smaller and smaller intervals, until by $c = 9$ the system is chaotic, so there is no periodic orbit. The power spectrum contains a full range of frequencies. This route to chaos, consisting of an infinite number of consecutive period doublings, appears in a wide variety of nonlinear systems [26].

For higher values of c there are occasional windows where the system ceases to be chaotic and reverts to periodic behaviour. For example at $c = 12$ there is a period 3 orbit, and once again there is regular structure in the power spectrum. As c is increased the system again period doubles to chaos. By $c = 17$ the chaotic attractor has grown in size.

Clearly the behaviour of the system depends in a vital way on the parameter c. Rather than examining individual values of c, it is desirable to try to picture how the system changes, and particularly where bifurcations occur, as c is varied continuously. One method to do this is analogous to the bifurcation diagrams of maps such as the logistic map, which simply record the points on the attractor as the bifurcation parameter is increased, either by a scatter plot or a density plot. For example, the top panel of Figure 1.2 shows a density plot of the x variable. For each value of c, it records the density of the x time series, of the sort shown in the left column of Figure 1.1.

While the resulting diagram is interesting and captures much of the behavior, a disadvantage of the method, which doesn't occur with maps, is that because x is a continuous variable, the periodic orbits appear as a continuous band rather than discrete points, and it is hard to distinguish areas of chaos. This is improved in the middle panel, which is again a density plot, but only includes those values of x which are either a local maximum or a local minimum. It is now much easier to distinguish between the areas of chaos, such as $c = 11$, and the periodic window beginning after $c = 12$. A period p orbit produces p separate local maxima, while in a chaotic region, we expect an infinite number of such maxima.

1.4 Spectral bifurcation diagrams

The lower panel of Figure 1.2 is a new kind of bifurcation diagram, dubbed the spectral bifurcation diagram. It was inspired by a technique used to do on-the-fly

measurements of field harmonics in superconducting magnets while the current is being ramped [45]. The diagram is composed by combining the power spectra at different values of c, as shown in the right column of Figure 1.1, into a continuous power histogram. The vertical axis shows frequency, while the greyscale indicates the power at that frequency. For example, at $c = 3$ the heavy line at frequency 0.16 corresponds as before to the periodic orbit with that frequency, and the period double at $c = 6$ is indicated by the appearance of a lower frequency line. Chaotic regions display a smear of frequencies. The periodic window after $c = 12$ appears as a clear band, with lines present only at multiples of the base frequencies. Another, smaller periodic window is also visible just after $c = 10$. The advantages of the spectral bifurcation diagram will become particularly evident in the next chapter.

From the bifurcation diagrams, it is evident that chaotic systems such as the Rössler system are sensitive not just to slight variations in initial conditions, but also to slight variations in parameters. This is one kind of model error, and perhaps the most basic. Below we present two other systems which will be useful in our later investigations into model error.

1.5 Other low dimension systems

1.5.1 The Lorenz '63 system

The story behind this classical system [37] is well known, but worth repeating here. In 1961, Lorenz, a meteorologist with MIT, visited Barry Saltzman of the Travelers Insurance Company Weather Centre in Hartford. Saltzman had been studying the convective motion of a fluid heated from below and cooled from above, a problem first examined by Lord Rayleigh. By considering variations in only two dimensions, Saltzman expanded the solution functions in a Fourier series, substituted this series into the original partial differential equations, and truncated the resulting infinite sum to a set of seven terms (the equations are presented in Chapter 4). In an effort to further simplify the system, and noting Saltzman's comment that under certain conditions all but three of the Fourier coefficients went to zero, Lorenz retained only those terms, and rescaled to obtain the following set of equations:

$$\begin{aligned} \frac{dx}{dt} &= -\sigma x + \sigma y \\ \frac{dy}{dt} &= xz + rx - y \end{aligned}$$

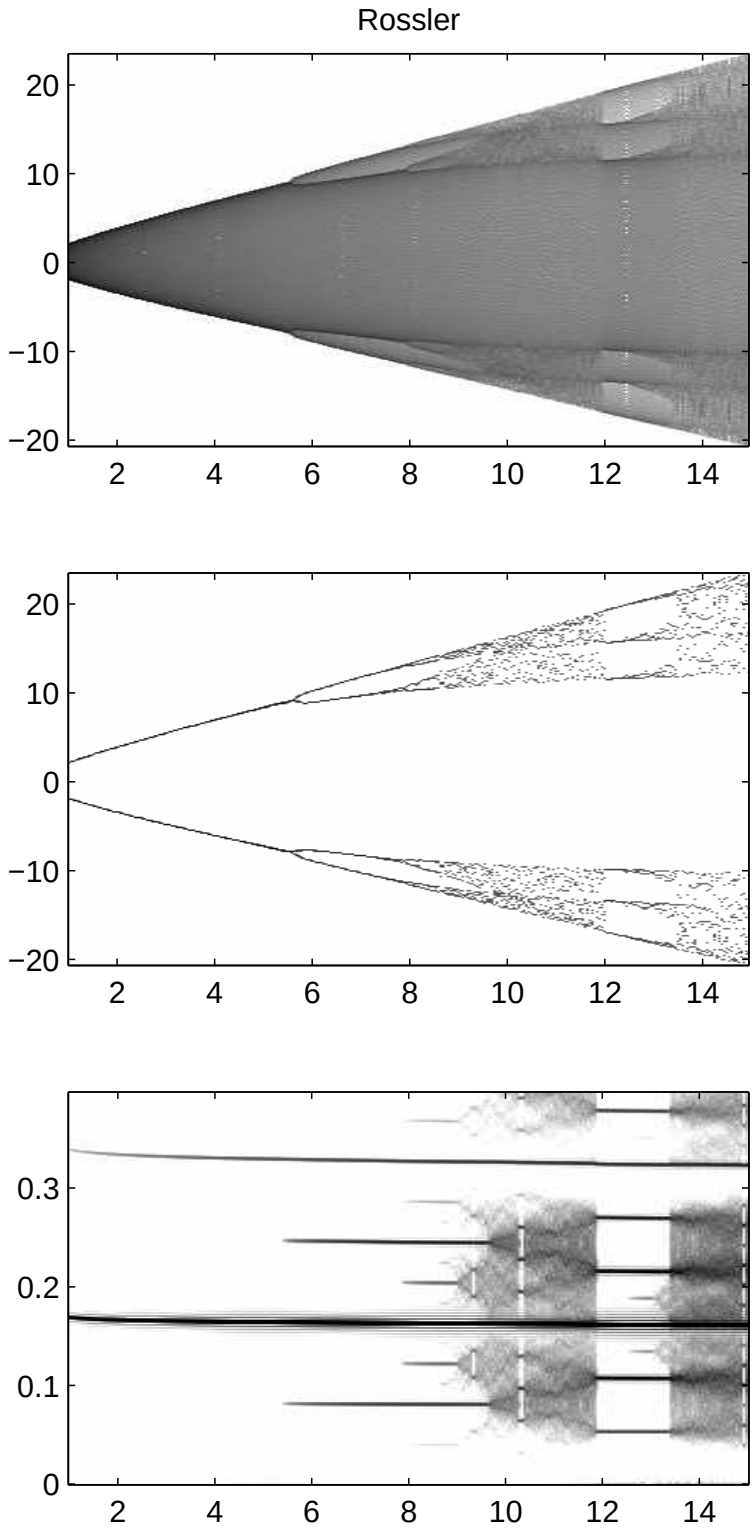


Figure 1.2: Bifurcation diagram for x as a function of c in the Rössler system. Upper panel is a density plot, middle panel is density of local max/min, and lower panel shows power spectra.

$$\frac{dz}{dt} = xy - bz. \tag{1.3}$$

The classical values are $\sigma = 10$, $b = 8/3$, and $r = 28$. The approximation is actually only valid for certain (low) values of r, and the behaviour of these reduced equations no longer says much about convection between plates. However they yield a mathematically very interesting system, whose butterfly-shaped attractor is something of a chaos icon. Staying with the original spirit of ignoring physical reality, we show the bifurcation behaviour for r between 25 and 275 in Figure 1.3 for the z variable. In Chapter 4, we will compare the Lorenz system with the original Saltzman equations.

1.5.2 The Rulkov Circuit

Another system we shall consider later, as an example of a low dimension system which actually approximates a physical system, is given by the Rulkov Circuit equations [59]:

$$\begin{aligned} \frac{dx}{dt} &= y \\ \frac{dy}{dt} &= -x - \delta y + z \\ \frac{dz}{dt} &= \gamma(\alpha f(x) - z) - \sigma y. \end{aligned} \tag{1.4}$$

where $\gamma = 0.2$, $\delta = 0.534$, $\sigma = 1.52$, and α is a parameter to be varied in the range 10 to 30. The function $f(x)$ is given by:

$$f(x) = -\mathrm{sgn}(x)\left(\frac{\sqrt{d(f_1(x) - a)^2 + c} - a}{d}\right) \tag{1.5}$$

and

$$f_1(x) = \|x\| \text{ if } \|x\| \leq a \tag{1.6}$$

$$= -q(\|x\| - p) \text{ if } a < \|x\| \leq b \tag{1.7}$$

$$= -a \text{ if } \|x\| > b \tag{1.8}$$

where $d = \frac{a^2 - c}{a^2}$, $q = \frac{2a}{b-a}$, $p = \frac{a+b}{2}$, $a = 0.5$, $b = 1.8$, and $c = 0.03$.

The bifurcation behaviour of the circuit equations is shown in Figure 1.4 for α in the range 18 to 23, which is where most of the interesting changes occur. Figure 1.5 is a zoomed view of the spectral bifurcation diagram, showing the existence of many small periodic windows. A possible application of the spectral bifurcation diagrams is to search for periodic or quasi-periodic orbits which are difficult to spot using other techniques.

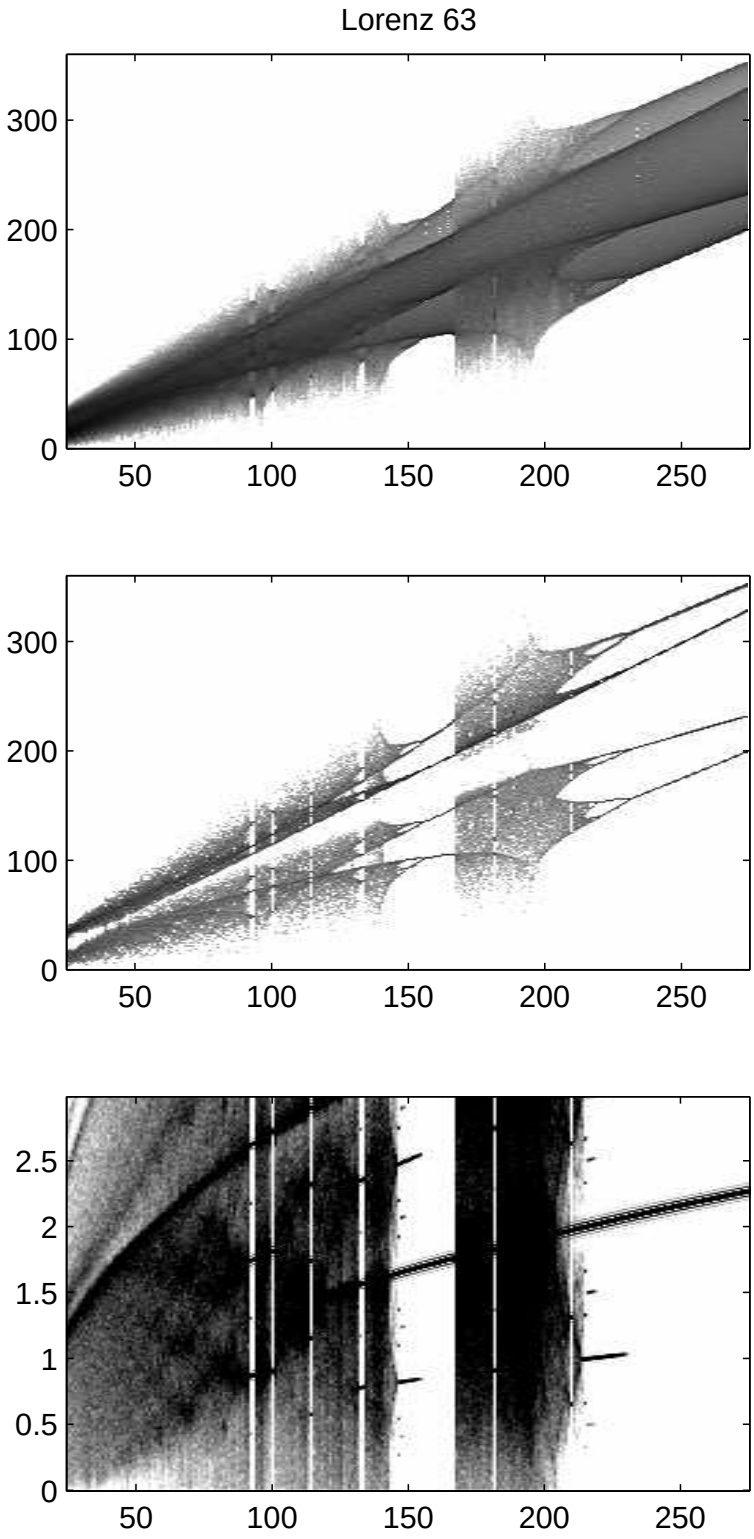


Figure 1.3: Bifurcation diagram for the Lorenz '63 system, from $F = 25$ to 275.

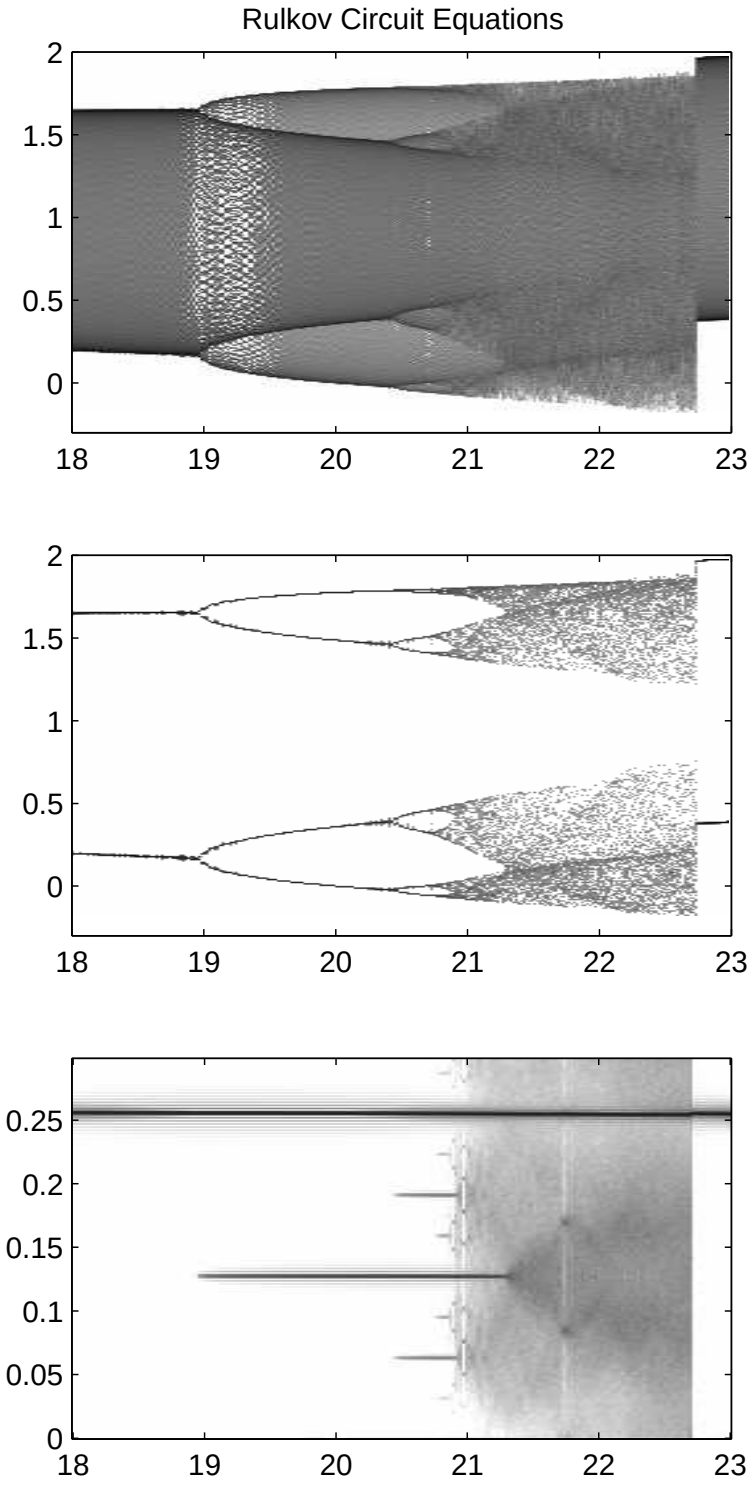


Figure 1.4: Bifurcation diagram for the circuit model.

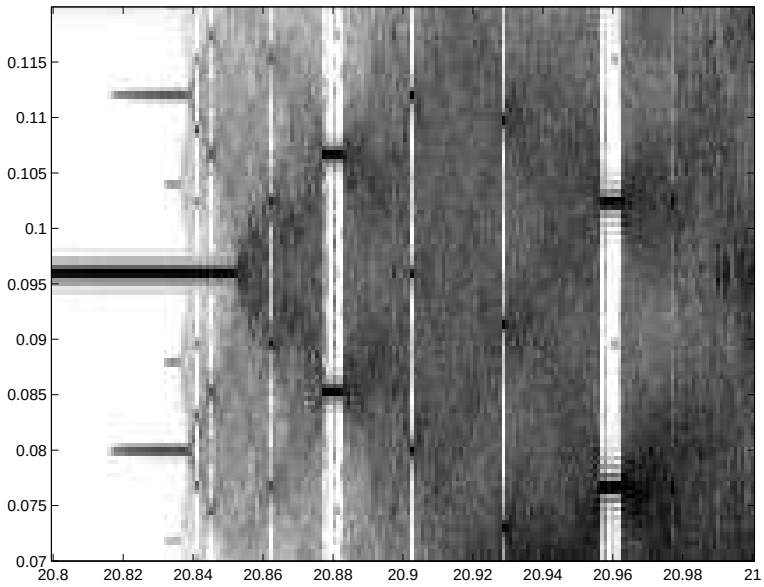


Figure 1.5: Zoomed view of bifurcation diagram for the circuit model.

1.6 The danger of low dimension systems

At this point, one might object that the Rössler, Lorenz or circuit systems have little in common with atmospheric models. The Lorenz '96 systems, presented in the next chapter, are higher dimensional, but in general, if we only do our experiments on 'toy' models, how do we know whether the results will generalise to real atmospheric models?

A similar point was raised at the Newton Institute in Cambridge during its 1996 discussion of atmospheric predictability [64], where it was put as follows:

> Many linear-dynamics-based intuitions are violated in low-dimensional nonlinear systems, like the Lorenz 1963 model; yet these NWP models appear to behave consistently with these intuitions. Is there some principle which indicated that there are pathologies which happen only in low order systems. Do these occur 'Even In, Or Only In' low order systems?

Toy models are at their most effective when used as a kind of thought experiment. An example is James Lovelock's Daisyworld model [41] which proposes an imaginary planet populated with white and dark daisies. As a result of their growth rates, it is seen that they effectively regulate the temperature of the planet. No one thinks this is an accurate model of the real world, but it succeeds in demonstrating a simple

principle which otherwise may not have been evident. It captures the imagination, and its impact is as much rhetorical, like a good analogy, as scientific.

Simple models, however, can easily be misapplied. For example, in many low dimension models, the model error is localised to particular regions of the attractor. If we were to conclude that the atmosphere shares this principle, so that model error varies greatly from day to day, then we may well be mistaken. There is no rule which says that nonlinear dynamical systems have to behave like one another. The problem is exacerbated when results are interpreted graphically; for example, it is easy to argue that the attractor or error pattern of one system might 'look like' that of another, an activity which is more sport than science.

For this reason, the results here are stated wherever possible in such a way that they can be applied to as broad as possible a class of model/system pairs. The new method for measuring model error, for example, is system independent, and we apply it equally to low dimension systems, or, in Chapter 6, operational weather models containing millions of variables.

There is one key difference between low and high dimensional systems, which has an immediately visible effect when comparing the two, and that is related to the concept of orthogonality. Pairs of randomly chosen vectors in a high dimension space have a high probability of being nearly orthogonal. More precisely, the variance of the angle between such vectors in a dimension n space is $\frac{1}{n}$. (To see this, let $\mathbf{x}$ and $\mathbf{y}$ be vectors. Since we are only interested in the angle, we can assume that each vector has magnitude 1. By symmetry, we can also choose $\mathbf{x}$ to be any vector we want. Choose $\mathbf{x}$ to be the vector with first coordinate $x_1 = 1$ and all other coordinates zero. Then $\mathbf{x} \cdot \mathbf{y} = y_1$. But $\sum_{i=1}^{n} y_i^2 = 1$, so the variance of y_1 is $\frac{1}{n}$, which proves the result.) For full weather models, where n is around 10^7, two uncorrelated vectors can therefore be treated as if they are orthogonal. This simple observation accounts for much of the difference in behaviour that we will experience here between weather models and low dimension systems.

1.7 Summary

In this section we have seen that even simple, low dimension systems such as the Rössler, Lorenz or Rulkov Circuit Equations show a rich mix of behaviour, which depends in a highly sensitive fashion on model parameters: slight alterations can throw the system from a periodic orbit into chaos, or vice versa. Since incorrect parameters are just one example of model error (if the model is structurally different

from the true system, then no setting of the parameters need remedy it), in order to understand model error we must be able to visualise model behaviour over a range of parameter values.

Spectral bifurcation diagrams are an excellent tool for this purpose. They clearly indicate when a model is in a chaotic, periodic, or, as we see in the next chapter, a quasi-periodic orbit. The diagrams also give information about the model's attractor, or 'climatology', in terms of the principle frequencies.

Because of the complexity of nonlinear systems, it might seem that comparing two different systems would be a hopelessly complicated task. Fortunately, we will see in chapter 3 that the problem can be made easier by certain emergent properties of the systems. In fact, one unanticipated result for the Lorenz '96 systems is that the question of whether they are chaotic or not has little bearing on the subject of model error.

Chapter 2

The Lorenz '96 systems

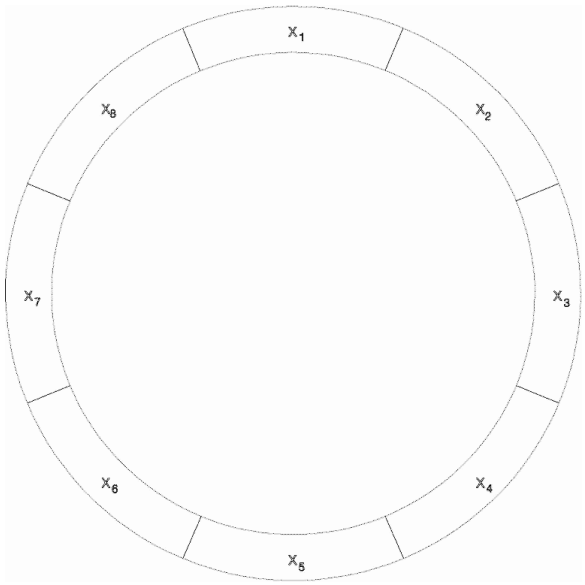
2.1 The one and two level system equations

The Lorenz '96 systems were first introduced as idealised one-dimensional models of the atmosphere [40, 27]. They produce time series which are qualitatively similar to the behavior of variables such as temperature. Another useful property, encountered in Chapter 6, is that, if the parameters are chosen correctly, the model error is seen to resemble that encountered in weather models, in terms of its variation with time and its magnitude compared with displacement error.

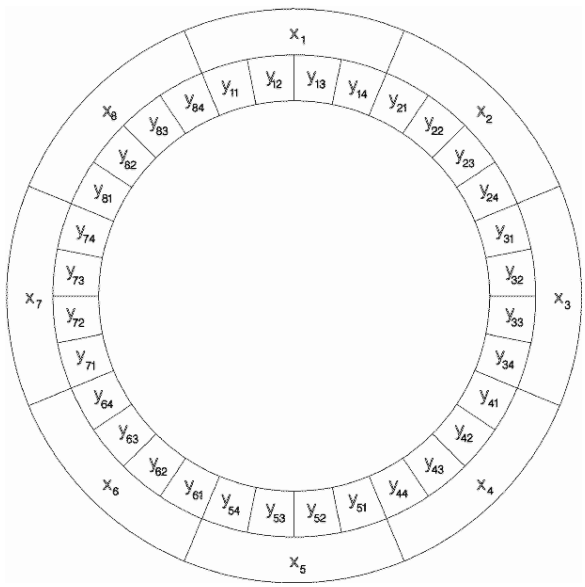
The first system, which we shall refer to as the one level system, contains n variables $x_1, x_2, ..., x_n$, and the equations are

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F, \qquad \text{i=1,...,n} \tag{2.1}$$

where F is a constant, and the index i is cyclic so that $x_{i-n} = x_{i+n} = x_i$. The x_i's can therefore be viewed as variables around a circle, as shown in Figure 2.1(a). In physical terms, they could be values of some atmospheric quantity such as temperature at n equally spaced latitudes around the globe. The constant term F in the equations is external forcing, the linear term is internal damping, and the quadratic terms, which introduce information about the spatial variation of x, represent advection. Of course, the system is only meant to be evocative of atmospheric behaviour, not an accurate model. A typical time series of the system is shown in Figure 2.2, which plots x_1 versus time for $F = 10$, for which the system is chaotic, as seen below. Note the equations are the same for each x_i regardless of index, so each variable has the same statistics.



(a) The one level system variables with $n = 8$.



(b) The two level system variables with $n = 8$ and $m = 4$.

Figure 2.1: System variables

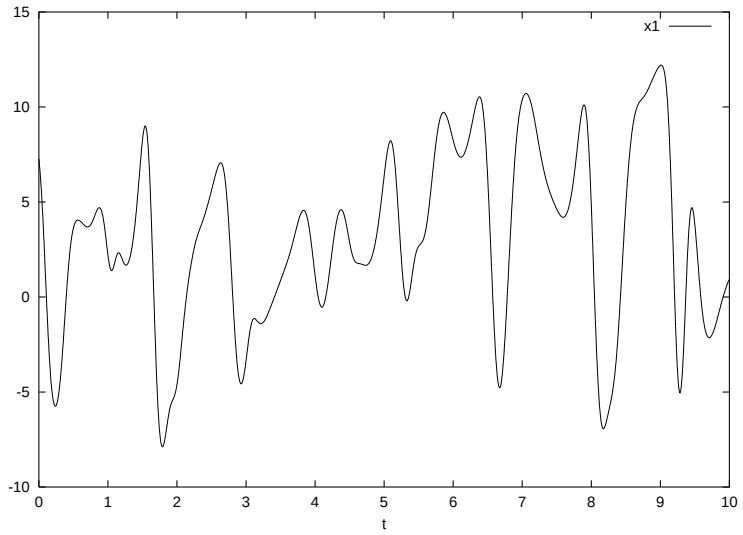


Figure 2.2: Time series of x_1 for the one level system given by equation 2.1 with $F = 10$.

The other system to be discussed will be referred to as the two level system, and incorporates smaller scale motions with shorter time scales. There are n variables $\tilde{x}_i$, together with an additional nm variables $\tilde{y}_{i,j}$ which can be viewed as sitting inside the $\tilde{x}$'s as shown in Figure 2.1(b). The equations [40] are

$$\begin{aligned} \frac{d\tilde{x}_i}{dt} &= \tilde{x}_{i-1}(\tilde{x}_{i+1} - \tilde{x}_{i-2}) - \tilde{x}_i + F - \frac{hc}{b}\sum_{j=1}^{m} \tilde{y}_{i,j} \\ \frac{d\tilde{y}_{i,j}}{dt} &= cb\tilde{y}_{i,j+1}(\tilde{y}_{i,j-1} - \tilde{y}_{i,j+2}) - c\tilde{y}_{i,j} + \frac{hc}{b}\tilde{x}_i \end{aligned} \tag{2.2}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Again the variables are cyclic so that $\tilde{y}_{i+n,j} = \tilde{y}_{i,j}$ and $\tilde{y}_{i,j-m} = \tilde{y}_{i-1,j}$. Following Lorenz [40], we set $b = c = 10$, which has the effect of making the $\tilde{y}$'s fluctuate ten times more rapidly than the $\tilde{x}$'s. The $\tilde{y}$'s can be thought of as convective scale quantities in the atmospheric analogy. The coupling coefficient h is set (except when otherwise specified) to 1. For this thesis we have primarily used $n = 8$ and $m = 4$, though Lorenz originally looked at higher dimensional systems [40].

Figure 2.3 shows a two level system time series of $\tilde{x}_1$ and $\tilde{y}_1$ for $F = 10$ which can be compared with Figure 2.2. It can be seen that a large local value of $\tilde{x}$ tends to excite the $\tilde{y}$ variables, due to the feedback between the two (this is a relationship that will prove useful in Chapter 3).

Our motivation for studying these systems is to examine the effect of model error. Suppose that we consider the two level system to be 'truth', and the one level system

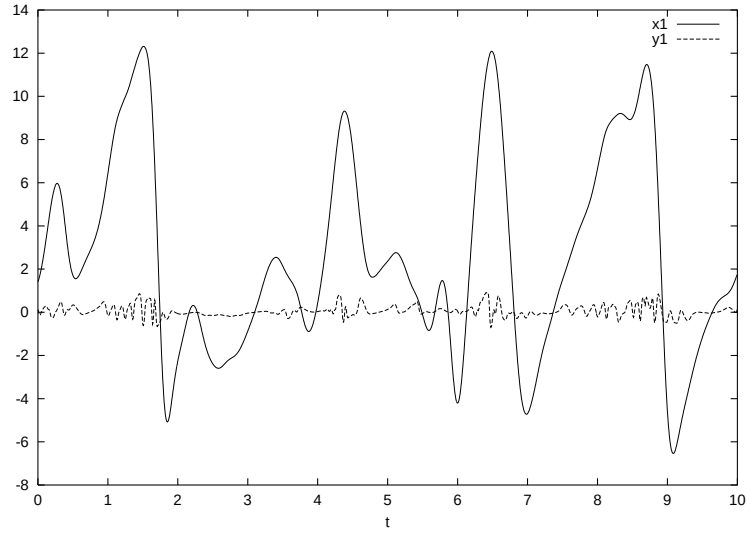


Figure 2.3: Time series of $\tilde{x}_1$ and $\tilde{y}_1$ versus time for the two level system given by equation 2.2 with $F = 10$.

to be a model. If we compare the x equations with the $\tilde{x}$ variables, we see that the model contains a constant forcing, while in the true system the forcing depends on the local $\tilde{y}$ variables. In Section 3, we will consider the constant forcing of the one level system to be a parameterisation of the two level system forcing, just as real atmospheric models incorporate parameterisations of complicated, small scale physical processes. First, though, it is necessary to see what role the forcing has in the structure of the solutions for either system.

2.2 Behaviour of the one level system with $n = 4$

In this section we will consider the one level system with $n = 4$, which is the simplest non-trivial variant. We will derive some of the basic properties of the system, before embarking on a numerical study of the bifurcation behaviour. The equations are:

$$\begin{aligned}
\frac{dx_1}{dt} &= x_4(x_2 - x_3) - x_1 + F \\
\frac{dx_2}{dt} &= x_1(x_3 - x_4) - x_2 + F \\
\frac{dx_3}{dt} &= x_2(x_4 - x_1) - x_3 + F \\
\frac{dx_4}{dt} &= x_3(x_1 - x_2) - x_4 + F
\end{aligned} \tag{2.3}$$

By substituting into the equations, it is easily seen that $x_1 = x_2 = x_3 = x_4 = F$

is a fixed point for all F. The stability of this fixed point can be determined by considering the Jacobean [26], which is:

$$\begin{pmatrix} -1 & x_4 & -x_4 & x_2 - x_3 \\ x_3 - x_4 & -1 & x_1 & -x_1 \\ -x_2 & x_4 - x_1 & -1 & x_2 \\ x_3 & -x_3 & x_1 - x_2 & -1 \end{pmatrix}$$

For $F = 0$, the Jacobean at the solution $x_1 = x_2 = x_3 = x_4 = 0$ is minus the identity, and the only eigenvalue is -1 which ensures stability. For $F = 1$, the Jacobean at the solution $x_1 = x_2 = x_3 = x_4 = 1$ is

$$\begin{pmatrix} -1 & 1 & -1 & 0 \\ 0 & -1 & 1 & -1 \\ -1 & 0 & -1 & 1 \\ 1 & -1 & 0 & -1 \end{pmatrix}$$

An eigenvector of this matrix is (i,-1, $-i$, 1), with associated eigenvalue i. At $F = 1$ an eigenvalue passes through the real axis in the complex plane. This is associated with a Hopf bifurcation [26], where a periodic orbit is produced from a fixed point. Indeed, in Figure 2.4 we see that for F just above 1, the system attractor plotted as x_1 versus x_2 is a near circular stable periodic orbit with period of approximately 2π. The variables x_2, x_3 and x_4 (not shown) also follow periodic orbits but are out of phase with x_1 by $\pi/2$, π and $3\pi/2$ respectively. Viewed as variables on a circle, the solution can then be seen as a wave propagating in a clockwise direction. This direction of propagation is noticeable even when the system is chaotic, and is a consequence of the advection term. The power spectrum of the time series of x_1 shows a peak at frequency $1/2\pi$, as expected, but also reveals a number of higher harmonics at multiples of the base frequency.

As F is increased, the x_1 time series picks up extra local maxima due to the presence of higher harmonics, but there is no sign of period doubling. In the log scale, the power appears to decrease more or less linearly with frequency. This implies that the coefficients in the power spectrum decrease exponentially with frequency. Near $F = 12$, the system becomes chaotic. Around $F = 14.7$ there is a periodic window before becoming chaotic again. The orbit shown in the lower panels has a period of 11.365 time units.

Some of the system behaviour is expressed in the upper two panels of Figure 2.5, which show the density and max/min bifurcations of x_1 (again, it doesn't matter

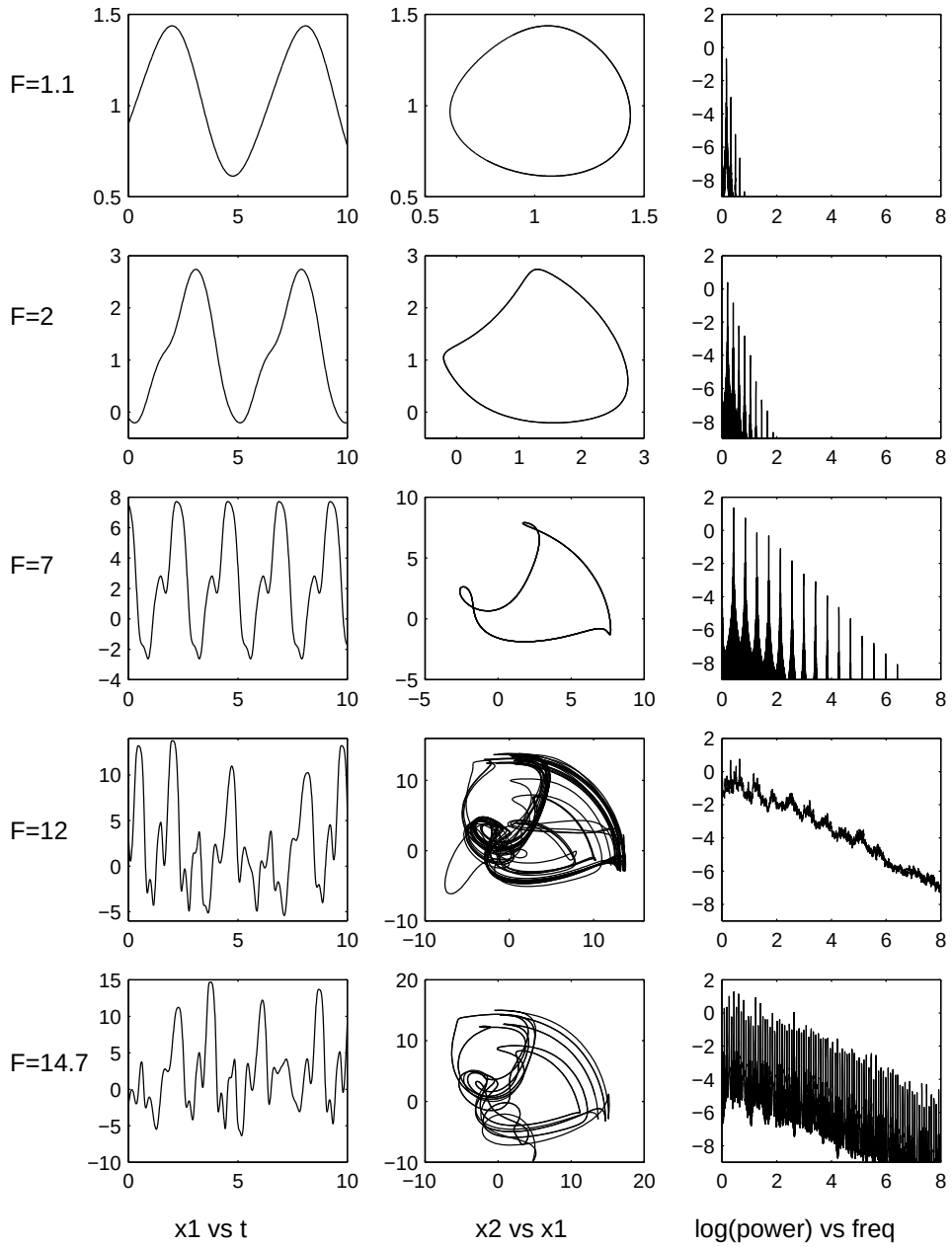


Figure 2.4: Plots of x_1 versus time, x_2 versus x_1 and log (base 10) power spectra versus frequency for various values of F for the single level system with $n = 4$.

which x_i is chosen). It can be compared with Figure 1.2. The line which appears at $F = 1$ represents the nearly linear growth in frequency of the solution as F is increased. The new lines which appear around $F = 3$ and $F = 8$ reflect the higher harmonics which appear for higher F. Unlike in Figure 1.2, however, these lines do not indicate period doubling. The system becomes chaotic around $F = 12$, and the periodic window at $F = 14.7$, as well as a second window just before $F = 16$, can be clearly seen.

While the bifurcation diagrams give useful information, new lines representing the growth of higher harmonics appear out of nowhere, and don't indicate any bifurcation. Because the system picks up progressively higher harmonics as F increases, the more natural approach is the spectral bifurcation diagram in the lower panel. This method also has the advantage of showing which spectra are present in the long-term 'climatology' of the system.

Comparing the spectral bifurcation diagram with Figures 2.4 and the upper two panels, we see that the lines beginning at and after $F = 1$ and continuing to $F = 12$ represent the periodic orbits. These lines are equally spaced in frequency, which means that the orbit for F in this region only contains harmonics which are multiples of its lowest frequency (this ensures periodicity). Around $F = 12$ the chaotic regime begins. The periodic windows, such as the one near $F = 16$, appear as bands of horizontal lines. The period of the orbit at $F = 14.7$ may be estimated from its lowest frequency of about 0.88, which agrees with the observed period 11.365.

2.3 One level systems with dimension 8 and 40

Higher dimension versions of the system display broadly similar behaviour, with some additional complications. One feature of the $n = 8$ system is that it has at least two attractors: a symmetric attractor ($x_5 = x_1, x_6 = x_2, x_7 = x_3, x_8 = x_4$) which is a copy of the $n = 4$ attractor, and a second attractor containing no such points. This symmetric attractor will attract any initial condition which has the required symmetry, while other points are drawn to the other attractor [27]. Therefore periodic orbits corresponding to those in Figure 2.4, even the one at $F = 14.7$ where most orbits are strongly chaotic, all exist in the $n = 8$ system. The analysis below is concerned with the second (asymmetric) attractor.

Figure 2.6 shows bifurcation diagrams for the $n = 8$ case. They are similar to the attractor for the $n = 4$ system, but become chaotic much earlier. Prior to about $F = 2.8$, the attractors for $n = 4$ and $n = 8$ correspond, in the sense that trajectories

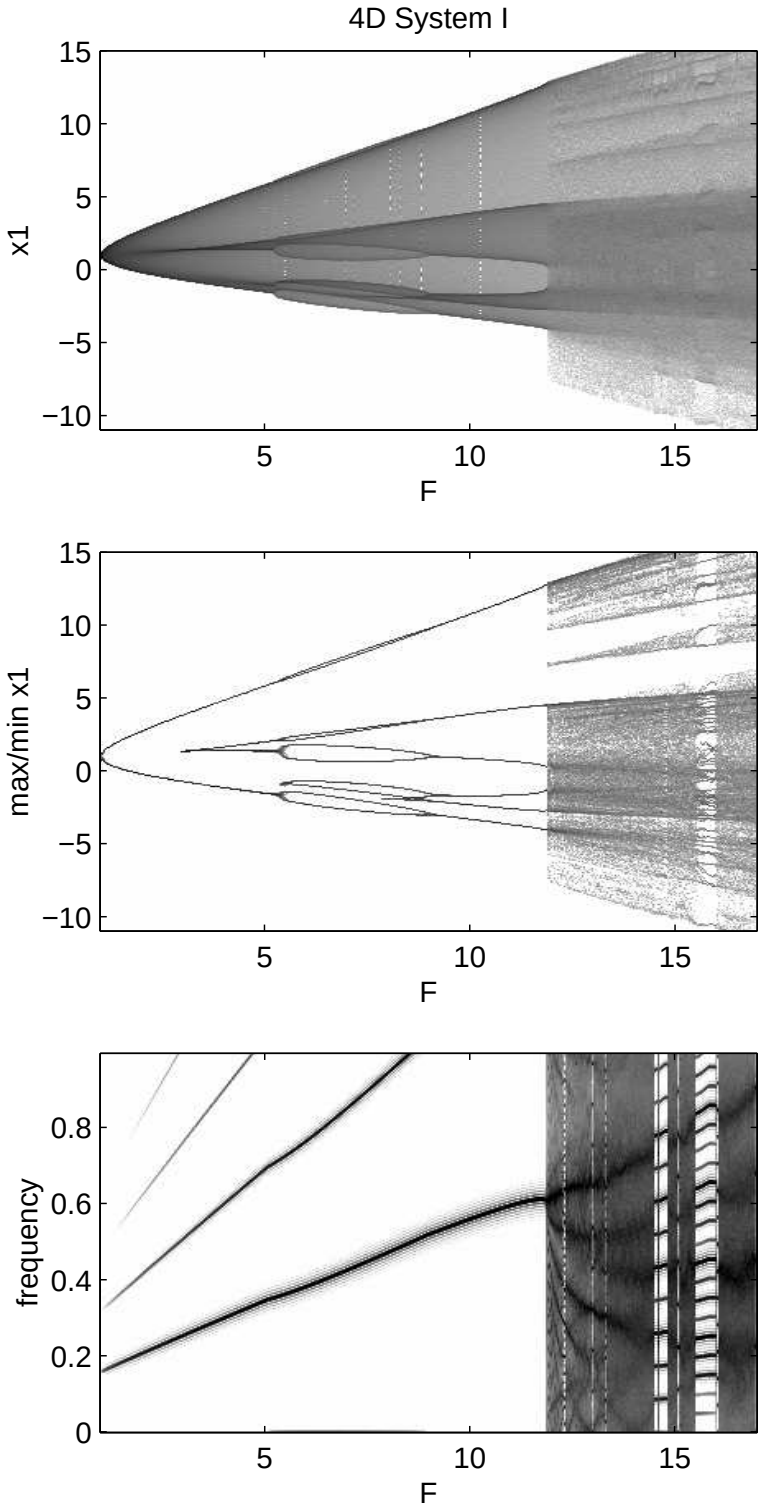


Figure 2.5: Bifurcation diagram for the one level system with $n = 4$. Upper panel is a density plot of x_1, middle panel is density of local max/min, and lower panel shows the spectral bifurcation diagram, as introduced in the text.

in the $n = 8$ system are drawn to the symmetric periodic orbits. However around $F = 2.8$ a period-doubling bifurcation occurs, as shown in Figure 2.7, and we will no longer have $x_1 = x_5$ and so on. By $F = 3.8$, the system appears to be chaotic. At this resolution, it is difficult to pick out periodic windows of any width in the chaotic regime past $F = 4.5$.

The spectral bifurcation diagram in the lower panel of Figure 2.6 reveals completely new features that are not evident from the density and max/min diagrams. The symmetric periodic orbit is indicated by the line beginning at $F = 1$ and frequency 0.16. At $F = 2.8$ a line appears at half the frequency, which corresponds to the period doubling mentioned above. By $F = 4$ we see a broad range of harmonics corresponding to chaos. However from about $F = 4.8$ to $F = 5.6$ there are large windows where the system appears to be non-chaotic (or at least not broad band).

Inspection of the spectral bifurcation diagram reveals that more than one frequency, or its harmonics, are present in these windows. The slopes of the diagonal lines in the range $F = 4.8$ to $F = 5.6$ are different, so the relative balance of the frequencies changes with F. When the frequencies are incommensurate, the result will be a quasi-periodic orbit. In bifurcation diagrams produced either by the maxima method or a Poincaré section method [1], these quasi-periodic orbits appear as a continuous band indistinguishable from chaos.

It is possible to find orbits in the region $F = 4.8$ to $F = 5.6$ which appear to close, as shown in Figure 2.8 for $F = 5.235298$. However the number of decimal places in F attests that this is not an easy task! The period of this orbit is 36.7, which corresponds to a frequency of 0.027. Figure 2.9 is a close-up of the spectral bifurcation diagram. The periodic orbit is located in a region where the spectra are separated by a frequency spacing of 0.027, as expected.

Still another way to view, or experience, the bifurcations is to listen to them. A tape is available which contains a translation of the 8 dimension system into sound. The x_1 and x_5 variables are interpreted as sound waves using MatLab, and played to the left and right speakers respectively. Starting from a periodic orbit at $F = 3.5$, the system is ramped upwards. The periodic orbit increases in speed and sound level, like a motor being accelerated. A distinct change is heard as the system goes chaotic around $F = 3.8$; the sound level drops and becomes irregular, as if the motor is about to stall. Entering the quasi-periodic region around $F = 4.7$, the system once again settles down, though it doesn't quite repeat. Only when held at a value of $F = 5.235298$ is a true rhythm established. It seems that the Lorenz systems are a better model of a car in need of a tune than the atmosphere!

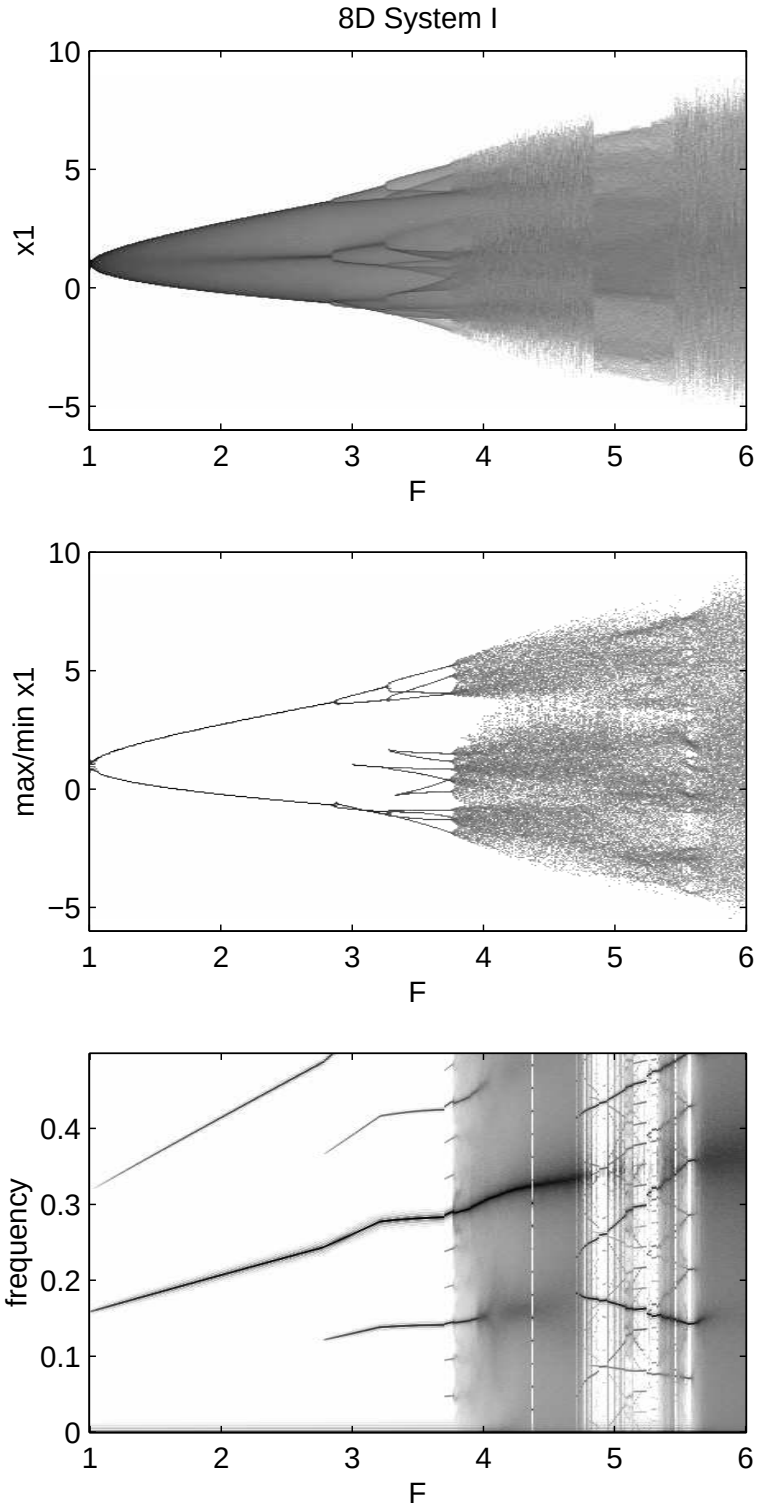


Figure 2.6: Bifurcation diagrams for the one level system with $n = 8$. Upper panel is the x_1 density plot, middle panel is the density of local max/min, lower panel is the spectral bifurcation diagram.

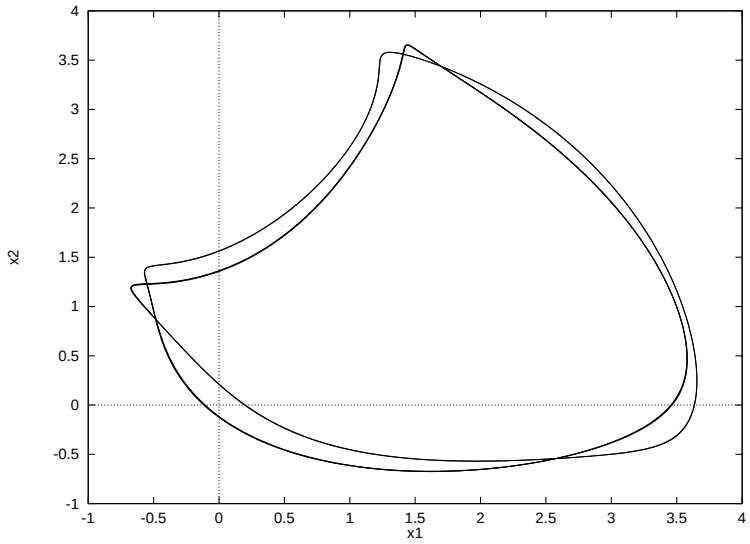


Figure 2.7: Period doubled orbit at $F = 3$ for the one level system with $n = 8$.

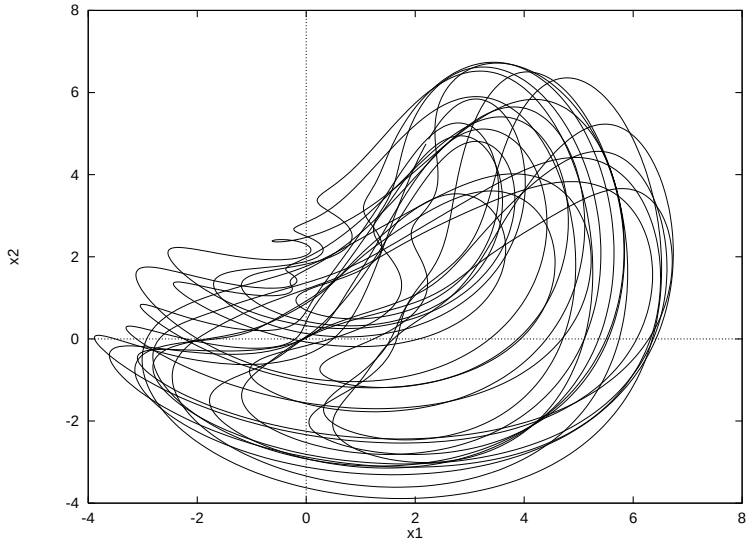


Figure 2.8: Periodic orbit at $F = 5.235298$ for one level system with $n = 8$.

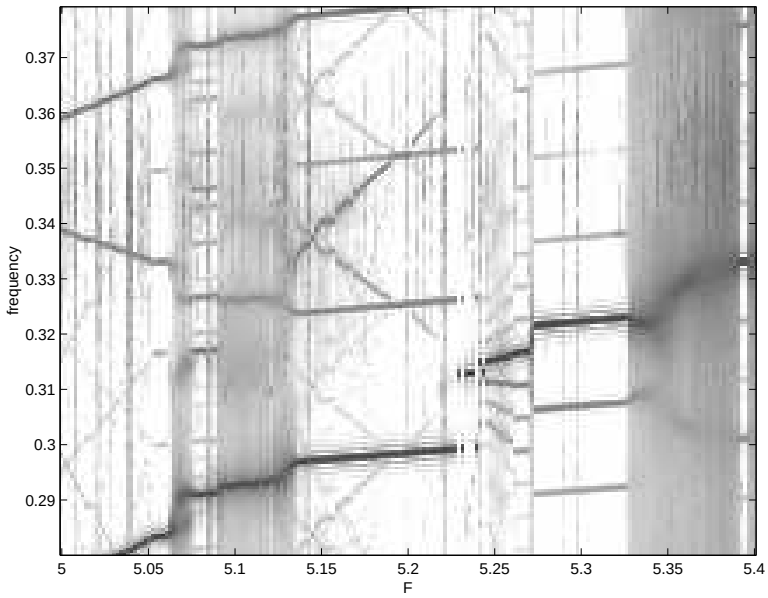


Figure 2.9: Close up of spectral bifurcation diagram in region of periodic orbit, for one level system with $n = 8$. Spacing of the harmonics at $F = 5.235298$ is about 0.27, corresponding to a frequency of 36.7

The dimension of the system can be increased indefinitely, but computations rapidly become expensive. The highest we looked at was $n = 40$. Figure 2.10 shows bifurcations for the 40D one level system. It is again quite similar to the other systems, with the exception that the spectral bifurcation diagram (lower panel) has a somewhat richer appearance in the transition to chaos.

The systems considered so far have all had a constant forcing term F. Other variants are possible; one studied is the case where F depends on the index i. This is analogous to the weather problem where forcing is different over land and over sea [27]. Another possibility is to make the forcing a function of the local value of x_i, or all values of x_i at the current time, or values of x_i at current and previous times, and so on. The two level system may be considered as one such variant, where the forcing depends on small scale $\tilde{y}$ variables which are coupled with the large scale $\tilde{x}$ variables.

2.4 Behaviour of the two level system

The equations for the $\tilde{x}$ variables in the two level system are similar to those of the one level system, with the difference that the constant forcing is replaced by a term which depends on the fast scale $\tilde{y}$ variables. We might therefore expect the $\tilde{x}$ variables

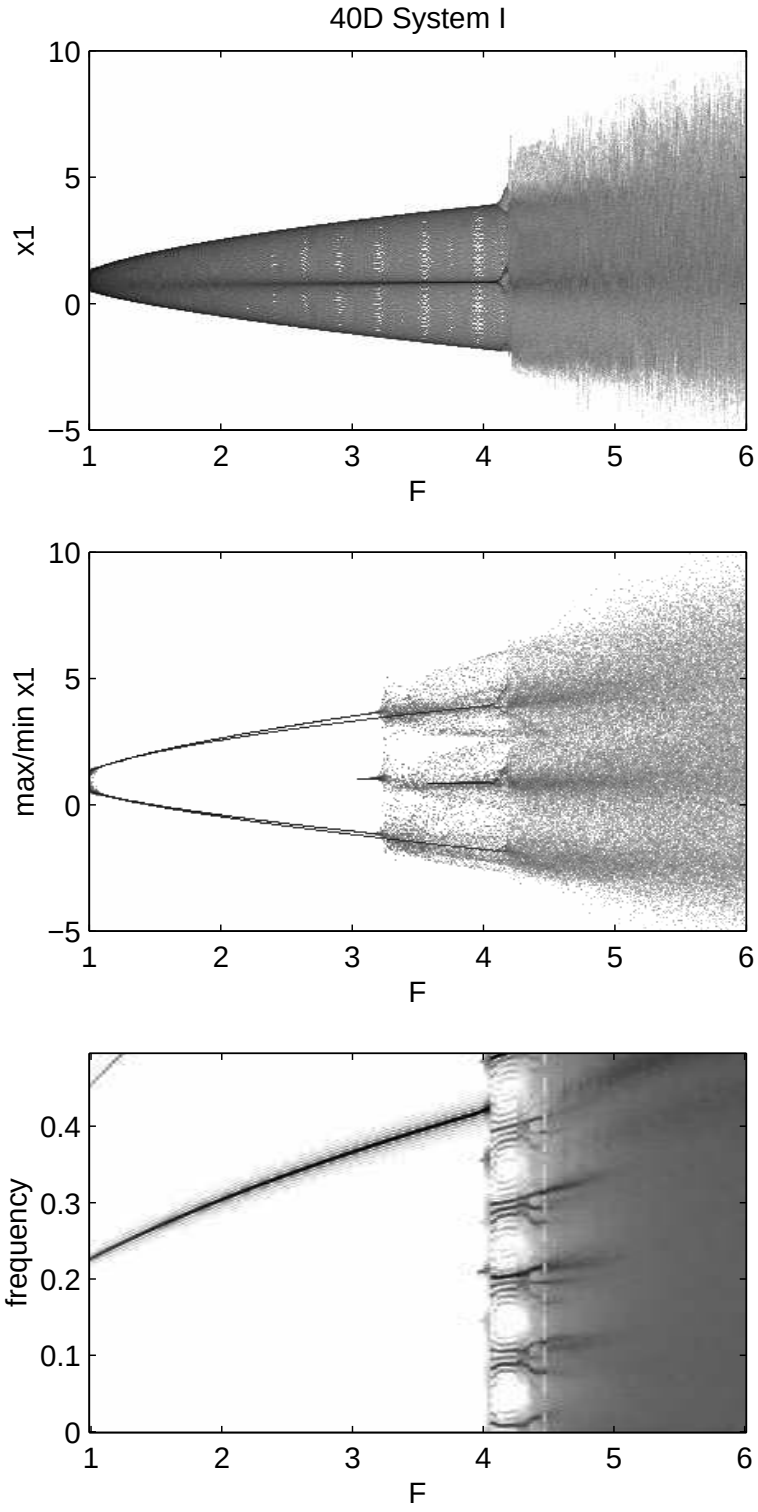


Figure 2.10: Bifurcation diagrams for one level system with $n = 40$. Upper panel is the x_1 density plot, middle panel is the density of local max/min, lower panel is the spectral bifurcation diagram.

to behave like the x's in the one level system, but with an added degree of fuzziness. The density and max/min bifurcation diagrams for the $n = 8, m = 4$ case (dimension 40) in Figure 2.11 bear this out. They are qualitatively quite similar to Figure 2.6. The lines in the region $F = 4$ to $F = 5$ have an added thickness, and correspond to apparently chaotic orbits that are like jostled versions of the periodic orbits seen in the one level system for slightly lower values of F. There is an additional period doubling bifurcation at $F = 1.5$, as the fast scale variables become non-zero. Another noticeable feature is that the $\tilde{y}$ variables tend to decrease the forcing F on average, so the whole diagram is shifted to the right compared to Figure 2.6.

The spectral bifurcation diagram for the two level system in Figure 2.11 can be compared also with that in Figure 2.6. Again it is quite similar to the one level case, with the difference that a full range of spectra, indicating a completely chaotic regime, doesn't occur until around $F = 5.5$ as opposed to $F = 4$. The diagram only shows to $F = 6$, however the system appears to remain chaotic and there aren't any periodic or quasi-periodic windows visible past that point. Figure 2.12 shows bifurcation diagrams for the $\tilde{y}$ variables.

Figure 2.14(a) shows a periodic orbit at $F = 4.6$. Of course, for the system to be periodic the $\tilde{y}$ variables must be periodic as well as the $\tilde{x}$'s, and the path traced out by the $\tilde{y}$'s in 2.14(b) does in fact close. Figures 2.14(c) and (d) is what happens for a slightly smaller value of F. The $\tilde{y}$ variables are chaotic, but the $\tilde{x}$ orbit is nearly periodic.

So far we have only considered bifurcations obtained by varying the parameter F. There are of course other possibilities, such as varying the coefficient h, which controls the coupling between the small scale $\tilde{y}$ variables and the large scale $\tilde{x}$ variables. Figure 2.13 shows bifurcations in the $\tilde{y}$ variables as the coefficient h is varied, while the forcing is held constant at $F = 2$. The spectral bifurcation diagram shows intricate cross-hatching, and a degree of structure that is absent from the other diagrams.

When the coupling coefficient is increased, the two level system is capable of showing quite complicated behaviour even at $F = 2$, where the one level system is periodic. Figure 2.15 shows the $\tilde{x}$ and $\tilde{y}$ orbits. The $\tilde{x}$ variables nearly follow a periodic orbit, while the $\tilde{y}$ variables are clearly quasi-periodic.

2.5 Summary

This concludes the introduction to the Lorenz '96 systems, which have turned out to be interesting in their own right, showing a rich variety of behaviour. As prototype

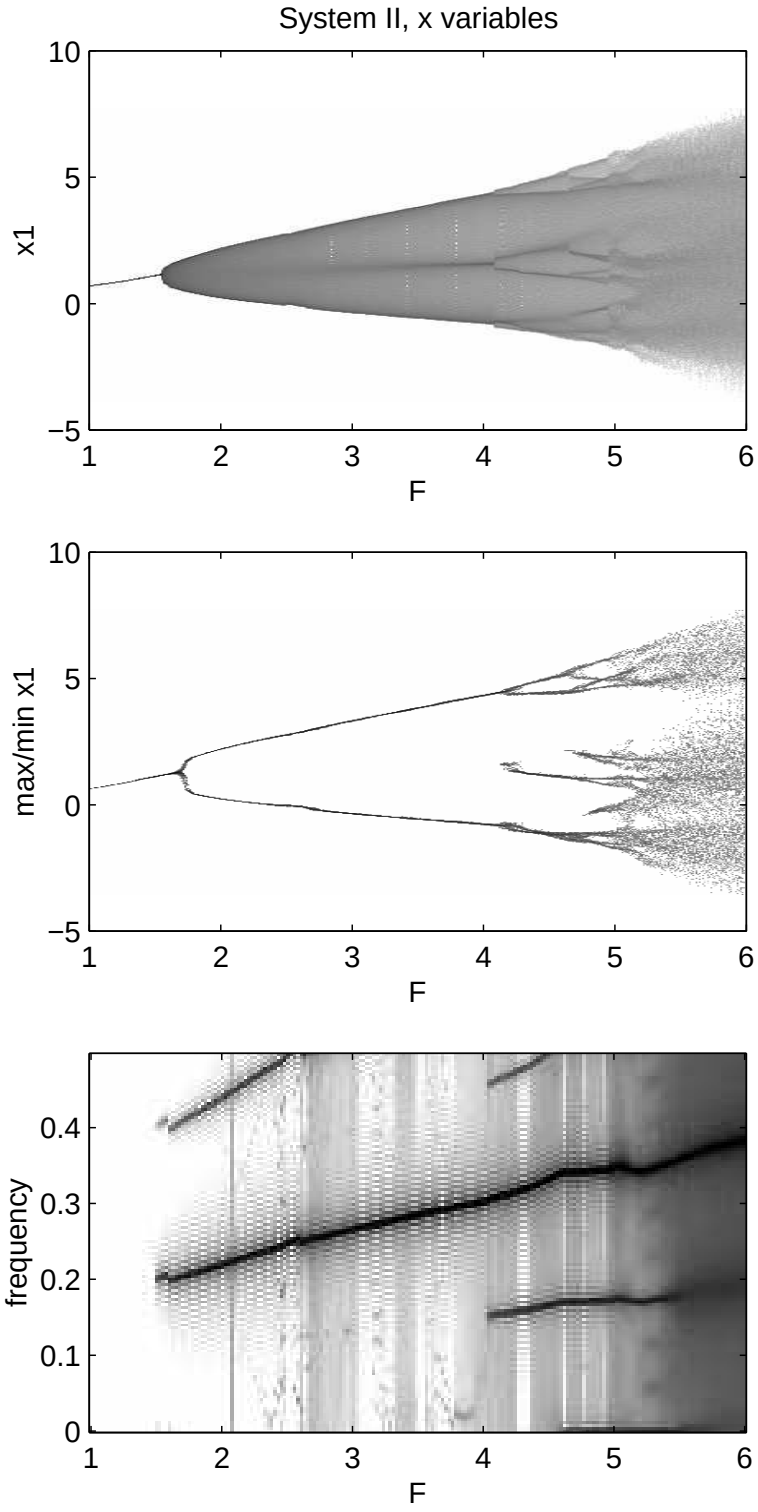


Figure 2.11: Bifurcation diagram for the 40D two level system, $\tilde{x}$ variables. The ripple effect noticeable in the spectral bifurcation diagram (lower panel) is due to numerical error in the power spectrum routine.

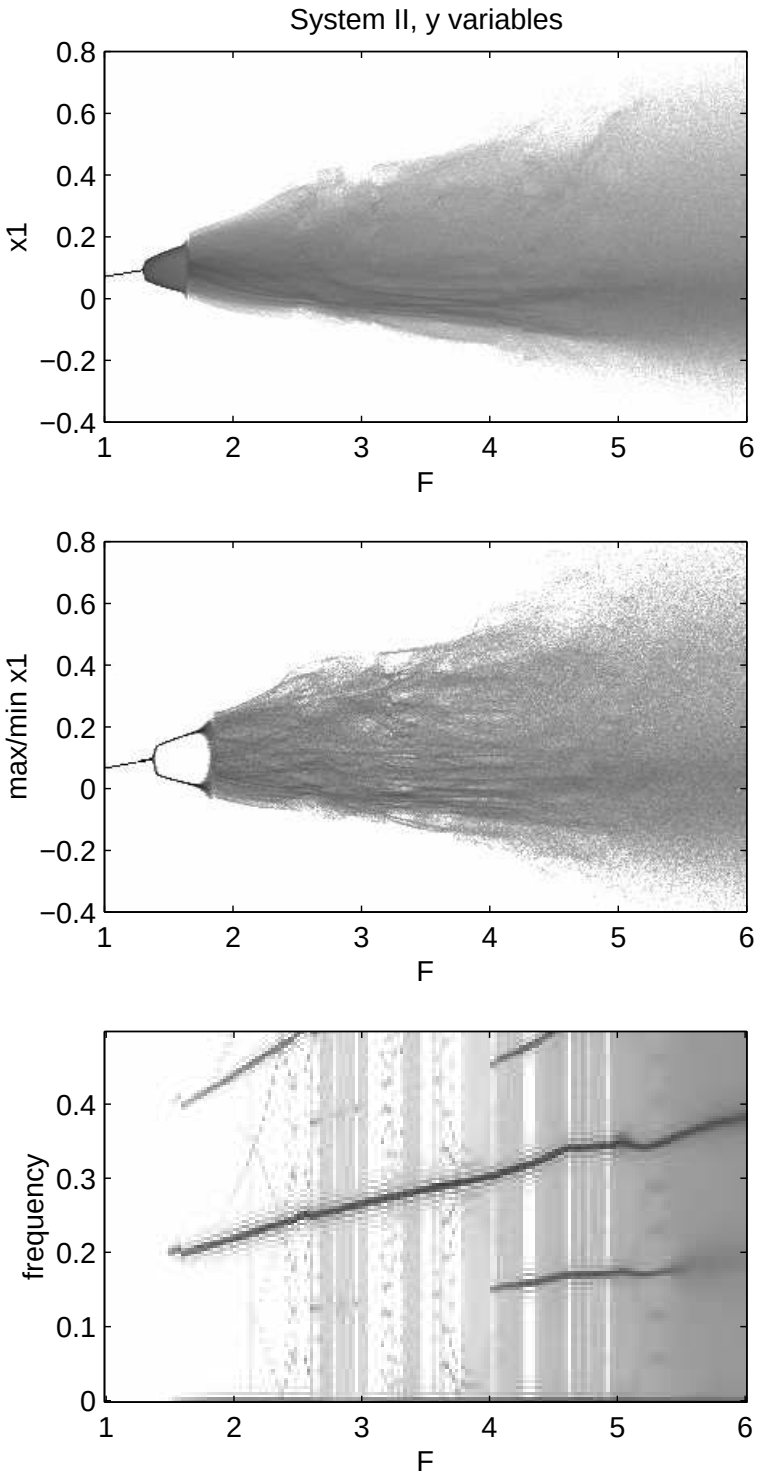


Figure 2.12: Bifurcation diagrams for 40D two level system, $\tilde{y}$ variables.

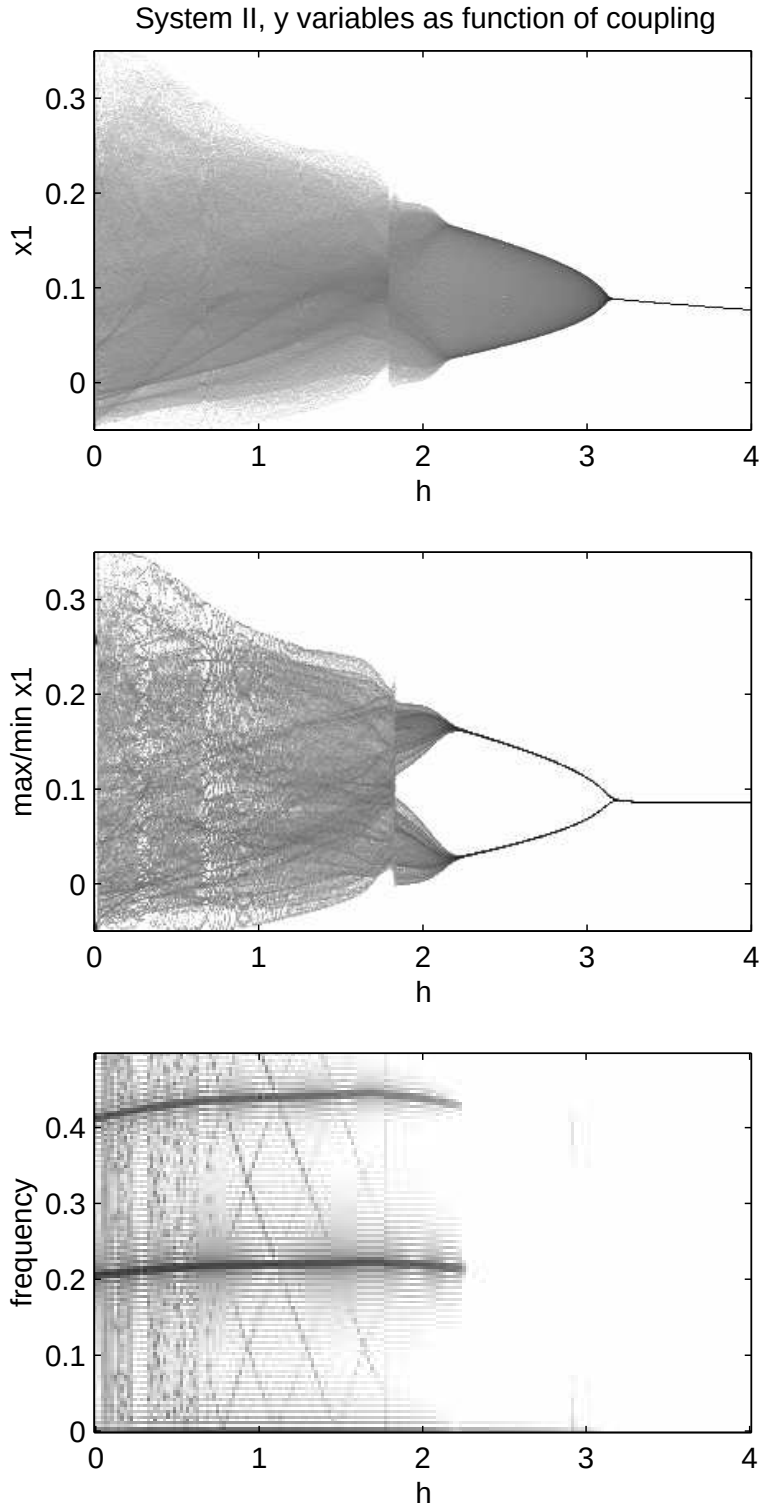
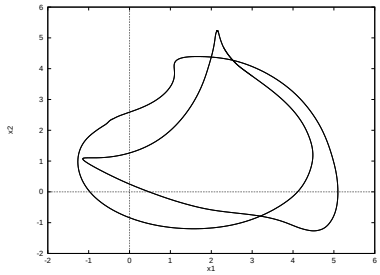
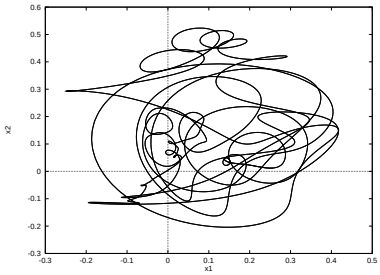


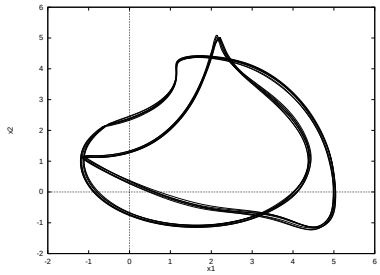
Figure 2.13: Bifurcation diagrams for 40D two level system, $\tilde{y}$ variables, as a function of the coupling coefficient h.



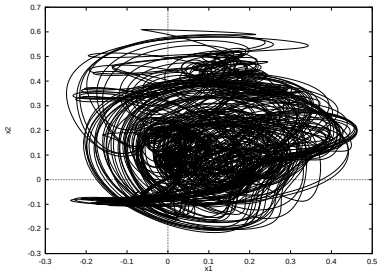
(a) F=4.6, x_2 vs x_1



(b) F=4.6, y_2 vs y_1

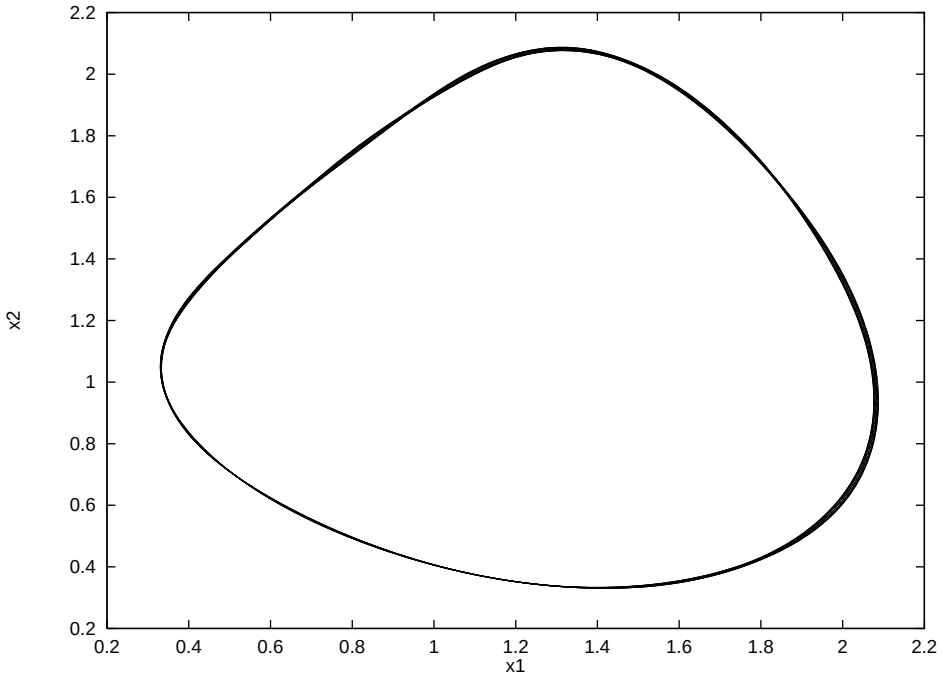


(c) F=4.5, x_2 vs x_1

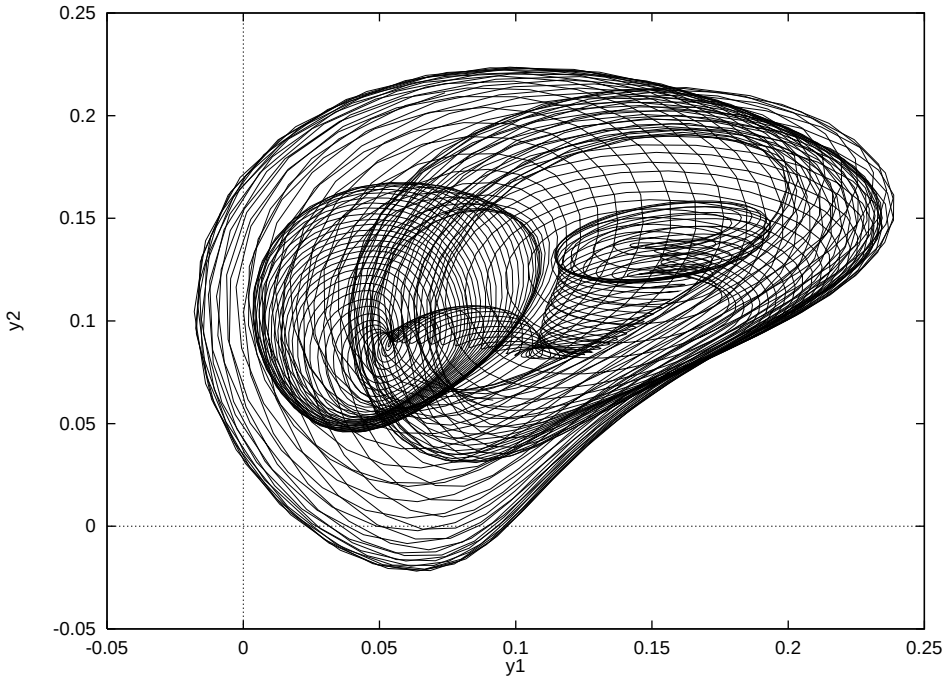


(d) F=4.5, y_2 vs y_1

Figure 2.14: Trajectories of 40D two level system, plotted in $\tilde{x}_1$ and $\tilde{y}_1$ variables, at $F = 4.5$ and 4.6. At the higher forcing, the system is periodic. At the lower forcing, both variables are chaotic, but this is most evident in the $\tilde{y}_1$ orbit.



(a) $\tilde{x}_2$ versus $\tilde{x}_1$



(b) $\tilde{y}_2$ versus $\tilde{y}_1$

Figure 2.15: $\tilde{x}$ and $\tilde{y}$ orbits for true system with $F = 2$ and $c = 1.2$.

models for discussions of atmospheric dynamics, they have the advantage of being higher dimensional and bearing at least a passing resemblance to real atmospheric systems, both in the equations and the resulting dynamics. Spectral bifurcation diagrams are a new and useful tool for analysing such systems, and reveal features such as quasi-periodic orbits which aren't evident in the usual type of bifurcation diagram.

We can now use the systems and the tools developed so far to study the effects of model error. We begin by looking at empirical properties of the error, before making a more formal analysis in Chapter 4. An advantage of the statistical approach is that it allows us to buid up an understanding of model error from observations, while seeing how its behavior is simplified by certain macro-properties of the Lorenz '96 systems.

Chapter 3

An exploration of model error using the Lorenz '96 systems

Accurate measurement of error is a basic plank of dynamical modelling; but as has already been discussed, the measurement of model error in the context of nonlinear dynamical systems is complicated by sensitivity to initial conditions. Even with chaotic systems and models, though, the importance, and ease of measurement, of model error depends on the situation. If the model is enormously wrong, and sensitivity to initial conditions relatively small, then we should have no problem in measuring the model error. On the other hand, if the model is accurate, but highly sensitive to initial condition, it is more difficult to detect what error is due to the model.

The most interesting behaviour occurs when initial condition error and model error vie with each other for importance, as is the case with the Lorenz '96 models which we will study in this chapter. Rather than be deterred by the presence of chaos, we will treat initial condition error and model error as independent entities, to see what can be learnt about their different properties. The approach is primarily experimental, taking the 40D two level system as the 'true system', and attempting to model it with variants of the one level system. The results will motivate the shadow approximation techniques presented in Chapter 4, which incorporate both initial condition and model error in a more complete description.

3.1 Model error vs initial condition error (continued)

Suppose we are comparing a model with a true system, and we wish to assess the effects of model error. In general, the model and true system will have different state spaces [61]. Therefore we require the existence of a projection operator $\mathbf{P}$ going from the true state space to the model-state space. For the time being, in the interest of simplicity, we will assume that the model and system exist in the same state space, and that we know the equations for the true system to be

$$\frac{d\tilde{\mathbf{x}}}{dt} = \tilde{\mathbf{G}}(\tilde{\mathbf{x}}(t)) \qquad \text{true system} \tag{3.1}$$

while the model equations are

$$\frac{d\mathbf{x}}{dt} = \mathbf{G}(\mathbf{x}(t)) \qquad \text{model.} \tag{3.2}$$

The difficulty in measuring model error for such systems is that it is coupled with displacement error. Suppose that trajectories in the model and true systems begin at exactly the same point, so $\mathbf{x}(0) = \tilde{\mathbf{x}}(0)$. At time zero there is no displacement error since the points agree, so the only error is model error. However as soon as a finite time has elapsed and the orbits have diverged, the model trajectory will differ from the true trajectory, and displacement error will come into play.

3.1.1 What is initial condition error?

In chaotic systems, error due to displacement of initial condition is blamed for many woes because it tends to magnify exponentially-on-average [65]. Suppose that the model initial condition is perturbed by a vector $\mathbf{e_d}(0)$. We can then estimate the displacement $\mathbf{e_d}(t)$ at some future time by considering the linearised dynamics around $\mathbf{x}(0)$ [66, 49].

Theorem. Let $\mathbf{x}(t)$ be a solution of the model equation

$$\frac{d\mathbf{x}}{dt} = \mathbf{G}(\mathbf{x}) \tag{3.3}$$

where $\mathbf{G}$ is C^1, and let $\mathbf{x_d}(t) = \mathbf{x}(t) + \mathbf{e_d}(t)$ be a solution with a perturbed initial condition $\mathbf{x_d}(0) = \mathbf{x}(0) + \mathbf{e_d}(0)$. Define the *linear propagator* [66] as

$$\mathbf{M}(t) = e^{\int_0^t \mathbf{J}(\mathbf{x}(t))dt} \tag{3.4}$$

where $\mathbf{J}$ is the Jacobian of $\mathbf{G}$. Then given a reference time $\tau > 0$, and $\epsilon > 0$, there exists a radius r such that, if $\|\mathbf{e_d}(t)\| < r$ for all $t \in [0, \tau]$, then

$$\|\mathbf{e_d}(t) - \mathbf{M}(t)\mathbf{e_d}(0)\| < \epsilon \quad \forall t \in [0, \tau]. \tag{3.5}$$

Proof. The derivation is routine, but will be useful for results developed later. From the system equations,

$$\begin{aligned} \frac{d\mathbf{e_d}(t)}{dt} &= \frac{d(\mathbf{x}(t) + \mathbf{e_d}(t))}{dt} - \frac{d\mathbf{x}(t)}{dt} \\ &= \mathbf{G}(\mathbf{x}(t) + \mathbf{e_d}(t)) - \mathbf{G}(\mathbf{x}(t)). \end{aligned} \tag{3.6}$$

Performing a Taylor expansion of $\mathbf{G}$ around $\mathbf{x}(t)$, and retaining only the first order term, we obtain

$$\frac{d\mathbf{e_d}(t)}{dt} = \mathbf{J}(\mathbf{x}(t))\mathbf{e_d}(t) + \mathbf{R_d}(t) \tag{3.7}$$

where the remainder term $R_d(t)$ is $O(\|\mathbf{e}_d(t)\|^2)$. Therefore, $\exists r_t > 0 \quad \ni \quad \|\mathbf{e_d}(t)\| < r_t \quad \Longrightarrow \quad \|\mathbf{R_d}(t)\| < \epsilon/\tau$. Pick r to be the minimum such r_t (possible since $[0, \tau]$ is a compact set). Integrating from 0 to t for $0 < t \leq \tau$ then gives

$$\mathbf{e_d}(t) = \mathbf{M}(t)\mathbf{e_d}(0) + \int_0^t \mathbf{R_d}(t)dt, \tag{3.8}$$

and

$$\|\mathbf{e_d}(t) - \mathbf{M}(t)\mathbf{e_d}(0)\| = \|\int_0^t \mathbf{R_d}(t)dt\| < \frac{\epsilon}{\tau}t \leq \epsilon \tag{3.9}$$

which proves the result.

The above result implies that the evolution of the error $\mathbf{e_d}(t)$ can be approximated by the linearised dynamics

$$\mathbf{e_d}(t) \approx \mathbf{M}(t)\mathbf{e_d}(0). \tag{3.10}$$

Under the linearised dynamics, a ball of initial conditions therefore evolves into an ellipsoid of final states. The major axes of the ellipsoid and their preimages can be determined by performing a singular value decomposition [25] of $\mathbf{M}(t)$ (which is how ECMWF determines its perturbations aimed for maximum growth [49]). Note the approximation only holds for displaced orbits $\mathbf{x_d}(t)$ which remain within a tolerance r of the reference trajectory $\mathbf{x}(t)$.

In the special case of a linear system, where the Jacobian J is a constant matrix, then

$$\mathbf{e_d}(t) = e^{\mathbf{J}t}\mathbf{e_d}(0). \tag{3.11}$$

So long as J has positive eigenvalues, trajectories will experience exponential growth. In general, and for the systems considered here, the Jacobian is not constant and the rate of growth can be described as exponential-on-average [65].

3.1.2 What is model error?

Model error can be analysed in much the same way as initial condition error. As before, we initially assume that the system and the model share the same state space (if not, then we require the use of a projection operator from the system state space to that of the model). Let $\tilde{\mathbf{x}}(t)$ be a solution of the system equation

$$\frac{d\tilde{\mathbf{x}}}{dt} = \tilde{\mathbf{G}}(\tilde{\mathbf{x}}) \tag{3.12}$$

where $\tilde{\mathbf{G}}$ is C^1, and let $\mathbf{x_m}(t) = \tilde{\mathbf{x}}(t) + \mathbf{e_m}(t)$ be the solution of the model equation

$$\frac{d\mathbf{x_m}}{dt} = \mathbf{G}(\mathbf{x_m}) \tag{3.13}$$

where $\mathbf{G}$ is C^1 and $\mathbf{e_m}(0) = \mathbf{0}$, so the model orbit begins with zero error relative to the true orbit. Define the *initial velocity error* to be

$$\mathbf{V} = \frac{d\mathbf{e_m}(t)}{dt} \mid_{t=0} \tag{3.14}$$

$$= \frac{d\mathbf{x}(t)}{dt} \mid_{t=0} - \frac{d\tilde{\mathbf{x}}(t)}{dt} \mid_{t=0} . \tag{3.15}$$

Then we have the following simple result.

Theorem. Given $\epsilon > 0$, there exists a time $\tau > 0$ such that

$$\|\mathbf{e_m}(t) - \mathbf{V}t\| < \epsilon \quad \forall t \in [0, \tau]. \tag{3.16}$$

Proof. Performing a Taylor expansion of $\mathbf{e_m}(t)$ around time zero, we have

$$\mathbf{e_m}(t) = \mathbf{e_m}(0) + \mathbf{V}t + R_m = 0 + \mathbf{V}t + R_m \tag{3.17}$$

where the remainder term R_m is $O(\|t\|^2)$. Therefore, there exists a time $\tau > 0$ such that $\|\mathbf{R}_m\| < \epsilon$ for all $t \in [0, \tau]$ which proves the result.

The linearised dynamics of the model error can then be written as

$$\mathbf{e_m}(t) \approx \mathbf{V}t. \tag{3.18}$$

Since the initial velocity error $\mathbf{V}$ is generally non-zero, it follows that, in general, the model error will experience an initial linear growth. This contrasts with the exponential-on-average growth of initial condition error, and implies that, for small times and displacements, model error will dominate initial condition error, as shown in the following corollary.

Corollary. Let the model and system be as above, and assume that the model error $\mathbf{e_m}(\tau)$ at some time $\tau > 0$ is non-zero. Then there exists a radius $r > 0$ such that, if $\|\mathbf{e_d}(0)\| < r$, then $\|\mathbf{e_d}(\tau)\| < \|\mathbf{e_m}(\tau)\|$.

This follows simply from the fact that displacement error can be reduced arbitrarily close to zero by choosing a sufficiently small initial displacement. An illustration of how model error can dominate error due to displacement of the initial condition is provided by Figure 3.1. Model error grows roughly linearly for small times, and soon overwhelms the error due to initial displacement. (Results are for the constant model which we define next - see also Figure 3.4.)

Our definition of velocity error is essentially the same as the definition of tendency error used in [33] in the context of weather models, where the velocity errors were studied in an attempt to isolate their sources in the model, or in [60], which proposed a statistical technique for assessing errors. A similiar term also appears as a residual in the data-fitting technique known as four-dimensional variational assimilation (4D-VAR) [13]. The observation that model error dominates initial condition error for small times is quite trivial, but often overlooked (for example, the comparison of both types of error in [12] for the Lorenz '63 system considered time scales of 50 units, and the initial effects are invisible). It points to an important property of model error which we will exploit in the remainder of the thesis, namely that model error is best measured over small deviations from the true orbit.

3.2 Modelling the two level system - the constant model

The above ideas about model error and displacement error can be demonstrated using the Lorenz '96 systems, with the two level system as truth. Suppose that, in the two level system, only the $\tilde{x}$ variables are known, and the values of the $\tilde{y}$ variables are not known. More formally, we project from the full system state space to the model state space using an operator $\mathbf{T}$ which truncates the vector $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ to $\mathbf{T}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = (\tilde{\mathbf{x}})$. The situation is analogous to real atmospheric systems, where the true system depends on an infinite number of variables that we can only parameterise.

In this case we can write the true system equations as

$$\frac{d\tilde{x}_i}{dt} = \tilde{x}_{i-1}(\tilde{x}_{i+1} - \tilde{x}_{i-2}) - \tilde{x}_i + \tilde{F}_i(t) \qquad \text{true system} \tag{3.19}$$

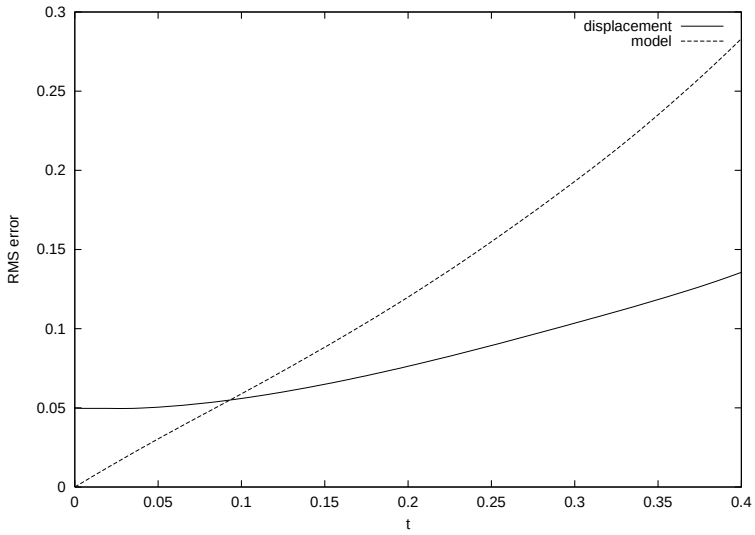


Figure 3.1: Plot showing how model error (dashed line) dominates error due to displacement of initial condition (solid line) for small times, providing the initial error is sufficiently small. Results are RMS errors for an actual model and system (see also Figure 3.4).

for i=1, n, where

$$\tilde{F}_i(t) = F - \frac{hc}{b}\sum_{j=1}^{m}\tilde{y}_{i,j}(t) \tag{3.20}$$

is treated as a forcing which varies in a complicated manner with time. Our goal is to approximate this system using models of the form

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + P_i(t) \tag{3.21}$$

where the n-dimensional vector $\mathbf{P(t)}$ with components $P_i(t)$ is some parameterisation of $\tilde{\mathbf{F}}(\mathbf{t})$.

The simplest parameterisation scheme is to set $P_i(t)$ equal to a constant for all i (this is the same as the one level Lorenz system). A sensible choice of constant would appear to be the mean forcing (we will see in chapter 4 that it is optimal in at least one sense). We therefore define the constant parameterisation as

$$P^c = \langle \tilde{F} \rangle \tag{3.22}$$

where the mean is calculated over a long orbit on the two level system attractor. In general P^c is smaller than F by a small amount, so for example if $F = 10$ the corresponding value of P^c is found to be 9.62. The model is then:

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + P^c \qquad \text{constant model} \tag{3.23}$$

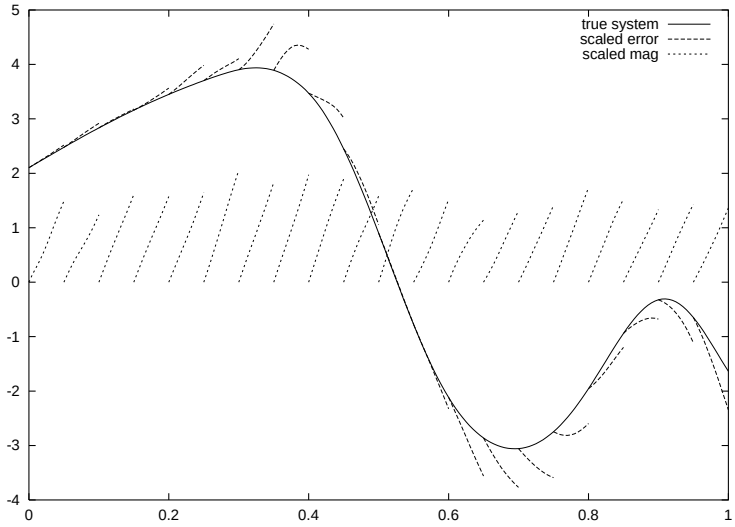


Figure 3.2: Forecast errors for Lorenz model/system, x_1 component. A number of short term model forecasts were initiated at regular intervals along a true trajectory. The errors have been scaled by a factor 20. The total error magnitude over all x_i is also shown, again scaled by a factor 20. Because it contains all components, it is larger than the x_1 error.

Figure 3.2 demonstrates the kind of errors that result when we approximate the true system with the constant model. The solid line is the $\tilde{x}_1$ variable for a trajectory of the true system. At regular times (with spacing of 0.04 on the bottom scale) model trajectories were initiated, starting on the true trajectory. The resulting x_1 trajectories, shown protruding like ribs from the solid line, soon diverge from truth - the errors here have been scaled by a factor 20 to aid visibility. The total error magnitude over all x_i's is shown as the series of diagonal lines, again scaled. Because it contains all components, it is larger (and more constant) than the x_1 component alone. Our aim in this chapter will be to quantify the growth of the model error when averaged over a large number of points.

3.2.1 Measuring initial condition error

One way to quantify an average sensitivity to initial conditions is to measure the root mean square (RMS) error growth over a number of perturbations and a number of starting points on the attractor. Suppose that $\tilde{\mathbf{x}}^{\mathbf{k}}(t)$ is a family of K solutions of the true system starting from different initial conditions on the attractor. For each k=1 to K, we perturb the starting point $\tilde{\mathbf{x}}^{\mathbf{k}}(0)$ by a randomly oriented vector

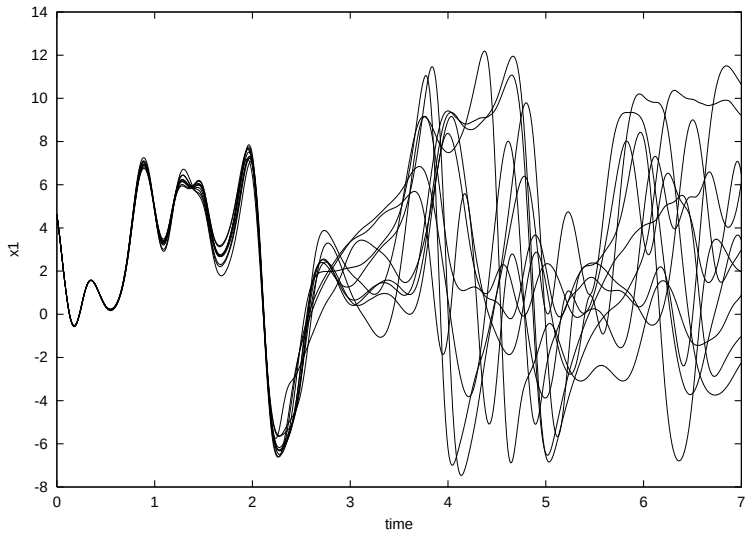


Figure 3.3: Ensemble with initial $\tilde{x}$ displacement standard deviation 0.01 for $F = 10$ true system. The $\tilde{y}$ variables are not perturbed.

whose magnitude has mean 0 and standard deviation ϵ. As an example, Figure 3.3 shows an ensemble of perturbations for the true system (3.20) with $F = 10$ about a single starting point. The standard deviation perturbation size on the $\tilde{x}$ variables is $\epsilon = 0.01$.

The displaced solution $\tilde{\mathbf{x}}_\mathbf{d}^\mathbf{k}(t)$ will diverge from the true orbit by a vector

$$\mathbf{e}_\mathbf{d}^\mathbf{k}(t) = \tilde{\mathbf{x}}_\mathbf{d}^\mathbf{k}(t) - \tilde{\mathbf{x}}^\mathbf{k}(t),$$

with magnitude $e_d^k(t)$. Now, define

$$e_d(t) = \sqrt{\langle e_d^k(t)^2 \rangle} \tag{3.24}$$

where the mean is taken over the K initial conditions. Then $e_d(t)$ is the RMS error growth due to displacement after a time t.

For the particular starting point in Figure 3.3, the trajectories disperse around $t = 3$, then appear to regroup at $t = 5.5$ before diverging again. When averaged over different starting points, however, the behaviour is more uniform. The lower curve in Figure 3.4 shows the RMS error $e_d(t)$ for the true system, while the middle curve is $e_d(t)$ for the constant model. The error growth is characterised by an initial exponential growth, as we would expect from the linearised dynamics, which eventually saturates due to the finite diameter of the attractor.

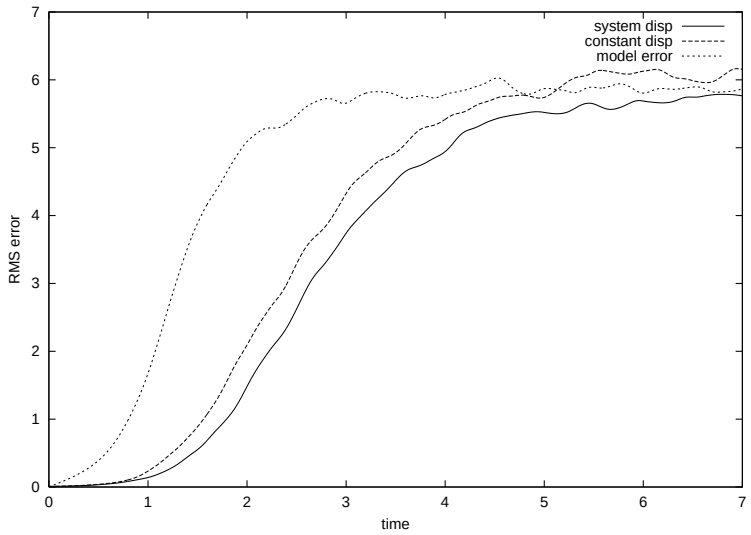


Figure 3.4: Bottom line (solid) shows initial condition displacement error $e_d(t)$ for the true system with $F = 10$. Middle line (long dash) shows $e_d(t)$ for the constant model with $P^c = 9.62$. Top line (short dash) shows model error $e_m(t)$ for the constant model. Note that displacement error grows exponentially at small times, while model error grows linearly.

3.2.2 Measuring model divergence

Divergence of model trajectories can be measured in an analogous manner. Let $\tilde{\mathbf{x}}^{\mathbf{k}}(t)$ be a family of K true system solutions on the attractor as before, and let $\mathbf{x}^{\mathbf{k}}(t)$ be the model solutions with corresponding initial conditions, so for each k we have $\mathbf{x}^{\mathbf{k}}(0) = \tilde{\mathbf{x}}^{\mathbf{k}}(0)$. The model solution will diverge from the true orbit by a vector $\mathbf{e}^{\mathbf{k}}_{\mathbf{m}}(t) = \mathbf{x}^{\mathbf{k}}(t) - \tilde{\mathbf{x}}^{\mathbf{k}}(t)$, with magnitude $e^k_m(t)$. Define

$$e_m(t) = \sqrt{\langle (e^k_m(t))^2 \rangle} \tag{3.25}$$

where the mean is taken over the K initial conditions. Then $e_m(t)$ is the RMS model divergence after a time t.

The divergence of the model from truth follows a somewhat different pattern than displacement error, as shown by the upper curve in Figure 3.4. The error starts at zero, by definition, and increases linearly in the early stages. This is shown more clearly in Figure 3.5, which is a close-up of the initial growth. Between 0.5 and 1.5 time units, the error then enters a phase of exponential growth. Finally, the error saturates as the true solution and the model solution settle on their respective attractors. Therefore, the model error manifests itself at low times as an initial linear

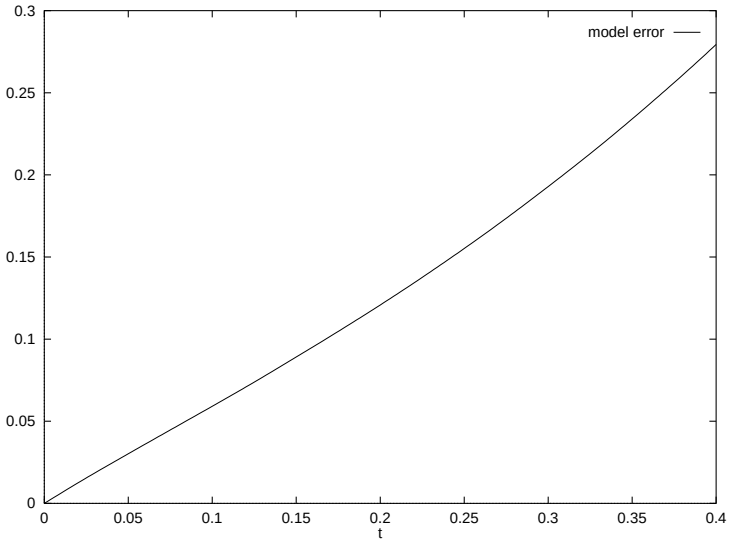


Figure 3.5: Initial linear growth of model error for constant model, $P^c = 9.62$, with respect to true system with $F = 10$.

error, and at longer times as a difference between the attractors of the model and the true system. The latter is a subject we will return to in Chapter 5.

Figure 3.6 shows how the model error varies with different values of F. At each F, the parameterised model forcing P^c is the mean forcing $\langle \tilde{F} \rangle$. For all values of F, initial growth is fairly linear. The extent of the linear growth phase decreases for increasing F.

3.3 Model error for the constant model

3.3.1 Velocity error

In the same way that displacement error is quantified by the initial exponential rate of growth, it seems natural to quantify model error by the slope of the initial linear phase of the model error curve, which gives the velocity error, or rate of model error growth, near time zero. Figure 3.7 shows the initial slope as a function of F. It increases fairly smoothly, and appears to vary, for F above 2, with the square root of F. Note that the constant model is near-perfect for $F \leq 1.3$. [1]

At time zero, the slope of the RMS model error curve can also be determined directly from the model equations, as in the linearised dynamics of the previous

[1] At these values of F the $\tilde{y}$ variables of the true system are periodic, as shown in the bifurcation diagram 2.12.

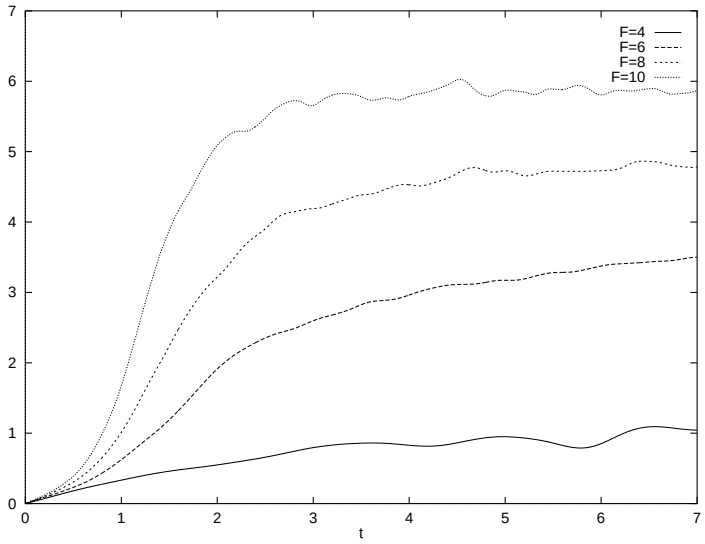


Figure 3.6: Model error curves for constant model. Order from bottom line to top line is $F = 4, 6, 8, 10$. Error is roughly linear in the early stages.

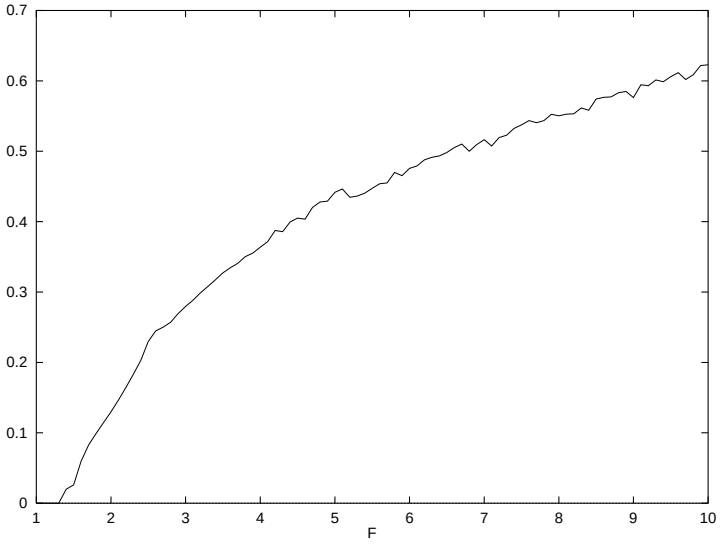


Figure 3.7: Initial slope of model error as a function of F. The constant model is near-perfect for $F \leq 1.3$, for which the $\tilde{y}$ variables are periodic. Computational errors are estimated to be about 0.02 in the vertical scale.

section. To make this explicit, suppose $\tilde{\mathbf{x}}^{\mathbf{k}}(t)$ for k=1 to K is a family of solutions for the true system with initial conditions on the attractor, and $\mathbf{x}^{\mathbf{k}}(t)$ are the model system solutions with the same initial conditions. The model error vectors are then $\mathbf{e}^{\mathbf{k}}_{\mathbf{m}}(t) = \mathbf{x}^{\mathbf{k}}(t) - \tilde{\mathbf{x}}^{\mathbf{k}}(t)$. Since $\mathbf{x}^{\mathbf{k}}(0) = \tilde{\mathbf{x}}^{\mathbf{k}}(0)$, subtraction of the model equations (3.21) from the true system equations (3.20) with $t = 0$ give

$$\frac{d\mathbf{e}^{\mathbf{k}}_{\mathbf{m}}(t)}{dt} \mid_{t=0} = \frac{d\mathbf{x}^{\mathbf{k}}(t)}{dt} \mid_{t=0} - \frac{d\tilde{\mathbf{x}}^{\mathbf{k}}(t)}{dt} \mid_{t=0} = \mathbf{F}^{\mathbf{k}}_{\mathbf{e}} \tag{3.26}$$

where

$$\mathbf{F}^{\mathbf{k}}_{\mathbf{e}} = \mathbf{P}^{\mathbf{k}}(0) - \tilde{\mathbf{F}}^{\mathbf{k}}(0) \tag{3.27}$$

is an n dimensional vector of the forcing difference between the model and the true system, as calculated at the initial condition $\tilde{\mathbf{x}}^{\mathbf{k}}(0)$. For the constant model, each component of the vector $\mathbf{P}^k$ is equal to P^c independent of k. The vector $\frac{d\mathbf{e}^{\mathbf{k}}_{\mathbf{m}}(t)}{dt} \mid_{t=0}$ is the difference between the model velocity and the true velocity on the true attractor, which is just the velocity error. With the Lorenz system, velocity error is caused by, and equal to, the forcing error $\mathbf{F}_{\mathbf{e}}$, since the other terms are identical.

Now, referring to Figure 3.5 and equation (3.25), the slope of the RMS error curve at time zero is

$$\lim_{t \to 0} \frac{1}{t} \sqrt{\langle (e^k_m(t))^2 \rangle} = \lim_{t \to 0} \sqrt{\langle (\frac{e^k_m(t)}{t})^2 \rangle} \tag{3.28}$$

where $e^k_m(t) = \|\mathbf{e}^{\mathbf{k}}_{\mathbf{m}}(t)\|$. By (3.26) this is just

$$\sqrt{\langle (F^k_e)^2 \rangle} \tag{3.29}$$

where $F^k_e = \|\mathbf{F}^{\mathbf{k}}_{\mathbf{e}}\|$, and the average is taken over the K initial conditions on the true attractor. Since P^c was chosen to be the mean of $\tilde{F}(0)$ over the attractor, the initial rate of RMS model error growth for the constant model is equal to the standard deviation of the forcing error, and initial model error can be determined directly through the properties of the forcing error. More generally, the initial slope of the error curve is given by the velocity error on the true attractor.

3.3.2 Forcing error for the constant model

The forcing $\tilde{F}$ in the true system shows a mix of periodic, quasi-periodic and chaotic behaviour for different values of F, as we might expect from Section 2. Figure 3.8 is a spectral bifurcation diagram for the difference in forcing $\mathbf{F}_{\mathbf{e}}$ (equation 3.27) between the model and the true system. Again, the analysis is for one component only of $\mathbf{F}_{\mathbf{e}}$; behaviour of the other components is the same by symmetry. Because the model

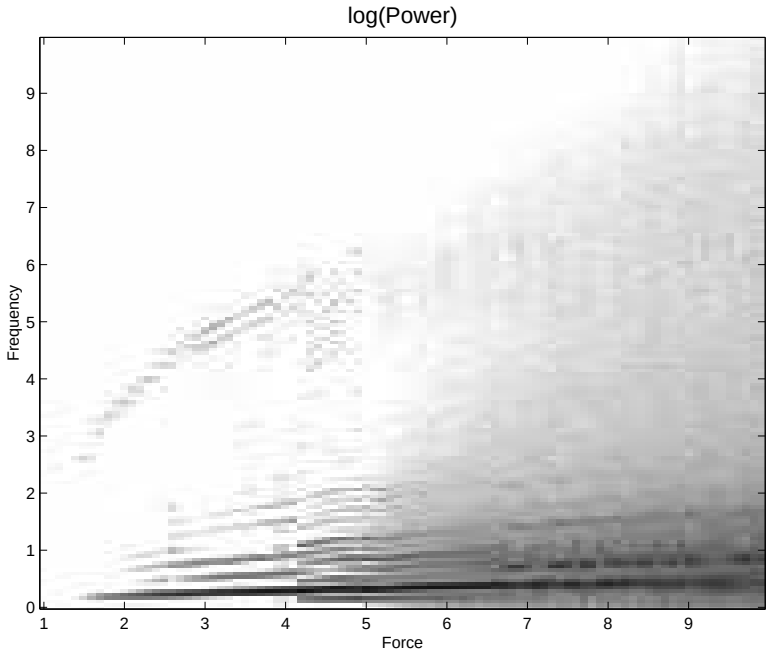


Figure 3.8: Spectral bifurcation diagram of forcing error $\mathbf{F_e}$ for the constant model. Calculations are for one component only; other components are the same by symmetry. As for model variables $\mathbf{x}$, the forcing shows a mix of periodic and chaotic behaviour.

forcing is the mean forcing in the true system, the average forcing error will vanish, and the constant term in the Fourier expansion is zero. Other components, however, remain unchanged.

The forcing error again shows a mix of behaviour, and one might conclude that model error will depend in an irregular way on the parameter F. Actually, this isn't the case; for what interests us is not whether the forcing error is chaotic, but merely, from 3.29, its standard deviation. This can of course be calculated directly, but it is also illustrative to note that, by definition of the power spectrum [53], the sum of the powers over all frequencies is just the variance. Therefore the sum of the spectra for a particular F gives the forcing error variance at that F. When this calculation is performed, the complexity of the bifurcation diagram disappears, revealing a simple pattern: the almost straight line in Figure 3.9.

Comparing Figures 3.9 and 3.7, we see that the standard deviation of the slope agrees with the standard deviation of the forcing, as expected. The standard deviation (square root of the variance) is also plotted in Figure 3.9. Because the variance of the forcing grows approximately linearly with F, for $F > 2$, we can say that, for the

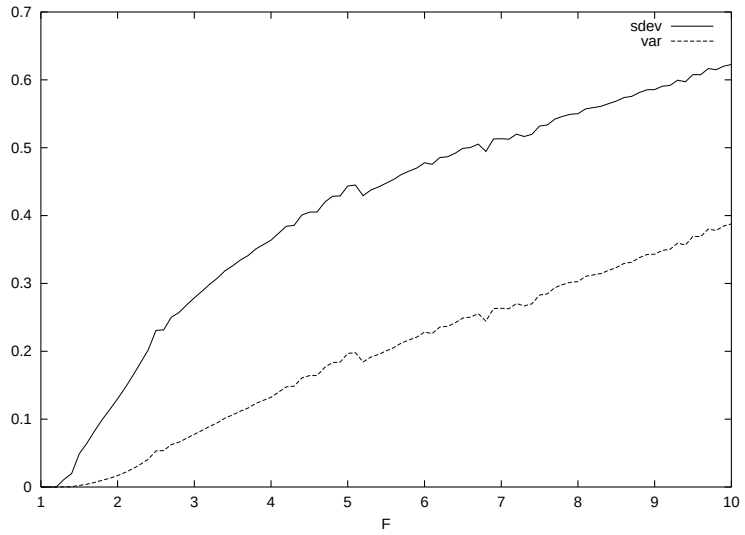


Figure 3.9: Total power (variance) and standard deviation of forcing error for the constant model.

constant model, the velocity error as expressed by the slope of the error curve near time zero tends to increase with the square root of F. This is a result that wouldn't be obvious by perusal of the forcing error spectral bifurcation diagram, Figure 3.8.

3.3.3 Shadowing

An important test of a model is that it be capable of shadowing the true system for a specified prediction period. Various closely related definitions of shadowing exist in the literature. The main one is ϵ-shadowing from the *Anosov-Bowen shadowing lemma* [2, 5]. A concern in modelling chaotic systems is that, due to sensitivity of initial conditions, a numerical solution may rapidly diverge from the true system trajectory. The shadowing lemma proves that, given certain restrictions on the system such as uniform hyperbolicity and invertibility, there at least exists a true solution, starting from a slightly displaced initial condition, which ϵ-shadows the numerical trajectory. To take just the definition of ϵ-shadowing from the lemma, it is as follows: given $\epsilon \geq 0$, a series of discrete times t_i with $a \leq t_i \leq b$, and a numerically generated model orbit $\mathbf{p}(t_i)$, then a system trajectory $\mathbf{s}(t_i)$ ϵ-shadows $\mathbf{p}(t_i)$ if

$$\|\mathbf{s}(t_i) - \mathbf{p}(t_i)\| \leq \epsilon \tag{3.30}$$

for every $a \leq t_i \leq b$.

Note that the shadowing lemma itself addresses a different problem to the one that we wish to solve: it states that a true trajectory shadows a numerical model

trajectory, while we wish to do the opposite, namely determine whether model trajectories shadow the true orbit. The lemma also assumes that we know the true system trajectory, which isn't the case when the system is the atmosphere.

A distinct shadowing problem was phrased in [23] to address the more practical question of modelling a system which is only known through a series of observations, each with observational uncertainty. This led to the definition of ι-shadowing: a model is said to *ι-shadow* the observed system for a time τ at radius r if it stays within a radius r of the observed time series over that time. Further, the model is said to be *consistent* with the observations if it ι-shadows with the shadow tolerance ϵ equal to the observational uncertainty (this can also be phrased using a Gaussian observational uncertainty).

The definition we will use in this thesis is the same as ι-shadowing, with the sole difference that we treat the shadow tolerance as a variable that is set independently, rather than derived from some error distribution (we discuss observation error separately). We therefore have the following definitions.

Definition: Given a true orbit $\tilde{\mathbf{x}}$, and shadow radius r_s, we say that a model orbit $\mathbf{x}$ *shadows* $\tilde{\mathbf{x}}$ for a time τ, if

$$\tau = \sup(t_s : \|\mathbf{x}(t) - \tilde{\mathbf{x}}(t)\| \leq r_s \quad \forall\ 0 \leq t \leq t_s). \tag{3.31}$$

Such a model orbit is called a *shadow orbit*. There is also a corresponding definition for discrete time series t_i.

Definition: Given a true trajectory, starting from a specified initial condition, and a shadow radius r_s, the *shadow time* of the model for that initial condition is the maximal time τ for which a model trajectory shadows the true trajectory within the specified radius. More loosely, when shadowing is simulated numerically, the shadow time is the longest time found by the numerical technique. Which of these definitions apply will be clear from the context.

Figure 3.10 is a schematic diagram which illustrates the definition of a shadow orbit. The true trajectory is shown as a solid line starting at x_0. The tube of radius r_s is shown as a shaded region. The model shadow orbit, the grey line starting from s_0, stays within the tube for a time τ. The model trajectory starting at x_0, however, shadows a shorter time.

The shadowing process depends on an interplay between model and displacement error, and for this reason is a necessary, rather than sufficient, condition for a good model. For example, if the model is sufficiently sensitive to displacement, it may be possible to find a perturbed initial condition which offsets a large model error for a

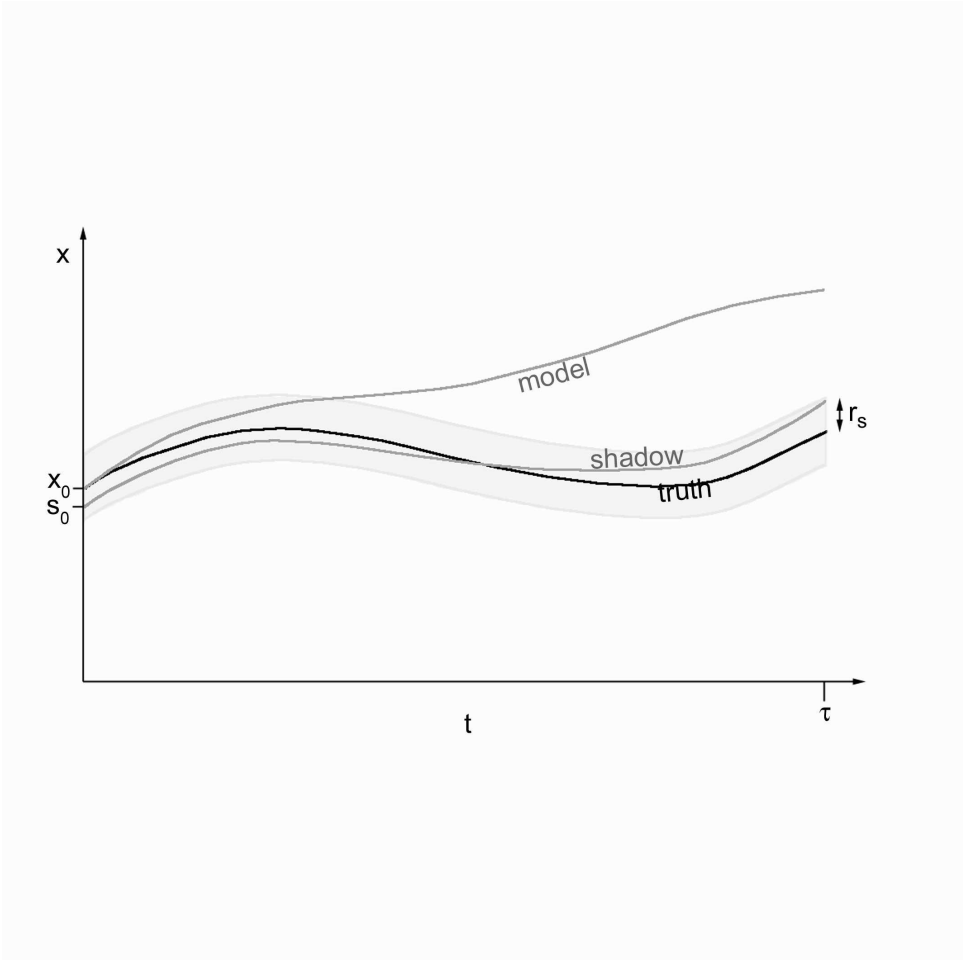


Figure 3.10: Schematic diagram showing true orbit (black line starting from x_0) with model shadow orbit (grey line starting from s_0). The model shadows for a time τ. A model trajectory starting at x_0, which shadows a shorter time, is also shown.

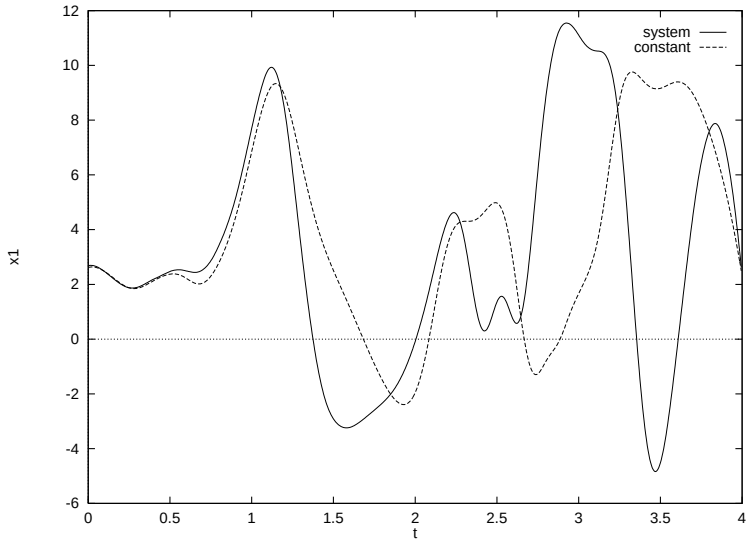


Figure 3.11: Plot of x_1 for a typical constant model shadowing orbit. The shadowing radius is 0.4. The orbit ceases to shadow at $t = 0.5$ (note the radius is over all components of x, not just x_1).

certain time; however the majority of initial conditions will most likely diverge quickly from truth (a badly controlled shot gun may hit its target, but only by spraying pellets in every direction). Therefore it would be unsafe to conclude that the model is a good approximation to the real system - it may be, or it may not.

An example of a shadowing orbit for the constant model is shown in Figure 3.11 (a longer one, for an improved model, is Figure 3.22). It was computed by an optimisation routine which searches over the possible perturbed initial conditions within a radius r_s, here 0.4, for the one with the longest shadowing time. The shadowing capabilities of a model depend on the particular initial condition. A histogram of the results over 200 points is shown in 3.12.

Figure 3.13 shows average shadowing times for the constant model, evaluated at integer values of F. The shadowing radius r_s has been scaled with F, in order that it stays in proportion with the size of the attractor. Two sets of results are shown, with the scaled radius $r_s = 0.2$ for $F = 10$, and $r_s = 0.4$ for $F = 10$. From the difference between the two radii, there appears to be a roughly linear relationship between shadowing radius and mean shadowing times.

One might also expect a relationship between initial state space velocity error, as computed from the forcing error, and shadowing times. Figure 3.14 shows the total RMS forcing error accumulated over the average shadow time, normalised to shadowing radius. It is calculated by taking the standard deviation of the forcing

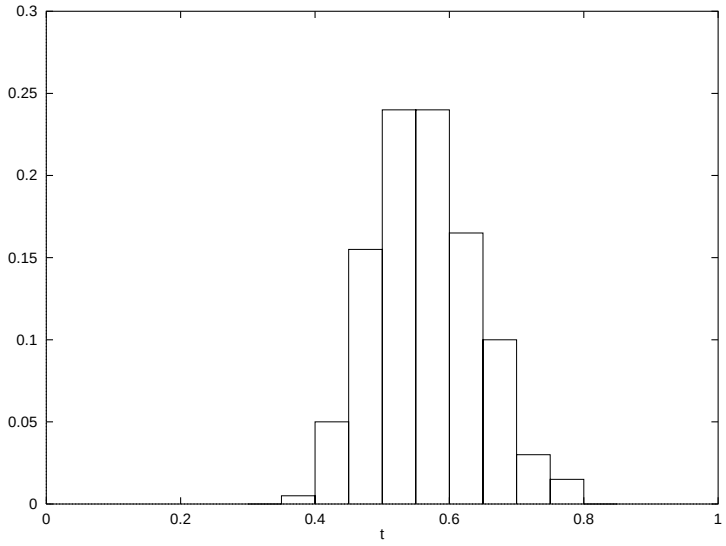


Figure 3.12: Histogram of shadowing times for the constant model with $F = 10$ over 200 starting points.

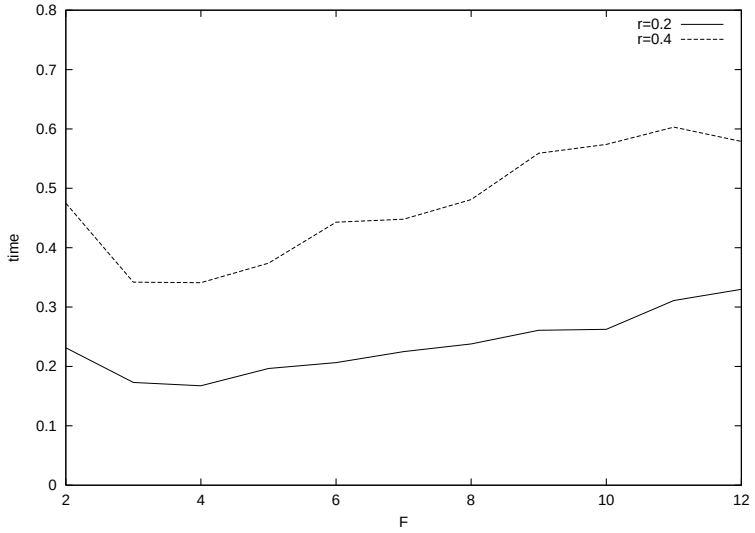


Figure 3.13: Mean shadowing times for constant model at integer values of F, with shadowing radius $r_s = 0.2$ and 0.4 at $F = 10$, and scaled for other values of F. Averaged over twenty initial conditions.

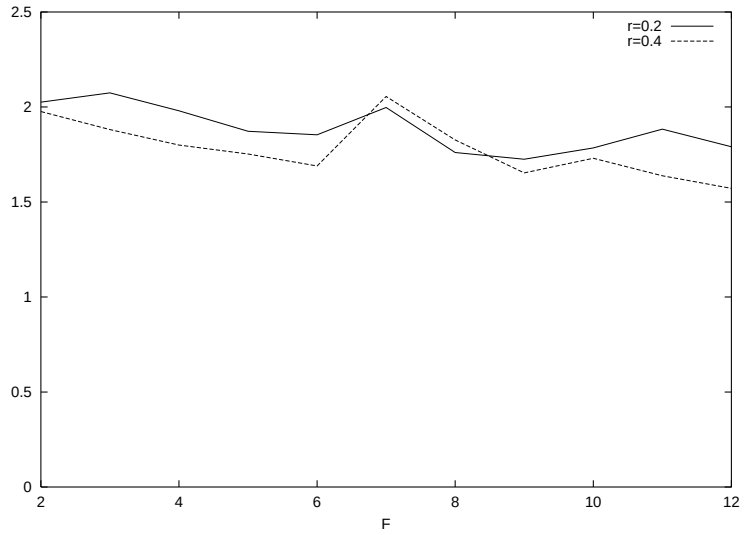


Figure 3.14: Total shadow forcing errors for the constant model for integer values of F, normalised for shadowing radius (dimensionless).

error, multiplying by the shadow time, and dividing by the shadowing radius r_s. The result is a dimensionless number. For this model, the numbers are fairly constant at about 2 over a range of forcing. It therefore appears that, for the constant model, shadow times vary inversely with initial model error and linearly with shadowing radius. Since initial model error depends on the square root of F for this model, we can also say that shadowing times vary inversely with the square root of F.

Of course, these simple, empirically deduced relationships between shadow time, shadow radius and initial model error won't necessarily generalise to other model/system pairs. In the next section we will go on to look at more sophisticated models which will reduce forcing error, and therefore lead to improved shadowing times. We will also refine our technique of calculating total forcing error so that it holds for these more general cases.

3.4 An improved model - the linear model

There are many different ways that one could go about refining the constant model so that it better approximates the true system, but in this chapter we shall look at just two further models: a simple model that makes the forcing a function of the local x_i variables, and a more sophisticated approach which utilises the fact that the system exists on a low-dimensional attractor. In Chapter 5 we will also look at two models designed to reproduce the general 'climatological' behaviour.

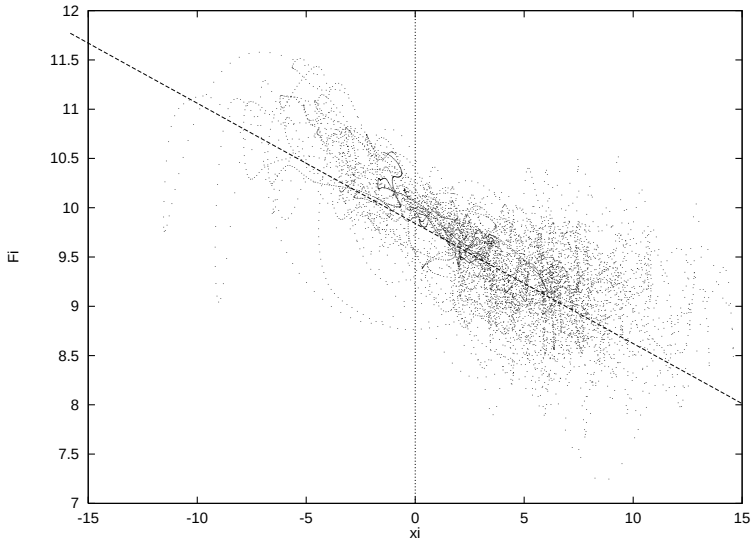


Figure 3.15: Scatter plot of local forcing versus local $\tilde{x}$ at $F = 10$. Also shown is the linear interpolation used in the corresponding linear model.

One of the first things to be noted about the true system in Section 2 was that the magnitude of the $\tilde{y}_{i,j}$ variables tends to be greater in regions where the local $\tilde{x}_i$ variable is also larger (see for example Figure 2.3). It is therefore reasonable to look for a connection between the local $\tilde{x}_i$ component and the forcing $\tilde{F}_i$ experienced by that component, which depends on the local $\tilde{y}_{i,j}$'s. This is shown more clearly by Figure 3.15, which is a plot of pairs $(\tilde{x}_i, \tilde{F}_i)$ where $\tilde{x}_i$ is the value of a particular $\tilde{x}$ component and

$$\tilde{F}_i = F - \frac{hc}{b} \sum_{j=1}^{m} \tilde{y}_{i,j} \tag{3.32}$$

is the local forcing. There is a definite linear tendency to the data, which can be fit using linear regression, leading to a formula

$$\mathbf{P_i^l}(\tilde{x}_i) = \alpha_0 + \alpha_1 \tilde{x}_i. \tag{3.33}$$

We can then define a new model

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + P_i^l(x_i) \qquad \text{linear model} \tag{3.34}$$

and apply our various tests of model error. The constant α_0 and slope α_1 must be calculated for each value of F. Figure 3.16 shows how they vary as a function of F. There is a peak around $F = 1.3$ when the $\tilde{y}$ variables become non-zero, but apart from that they are fairly constant.

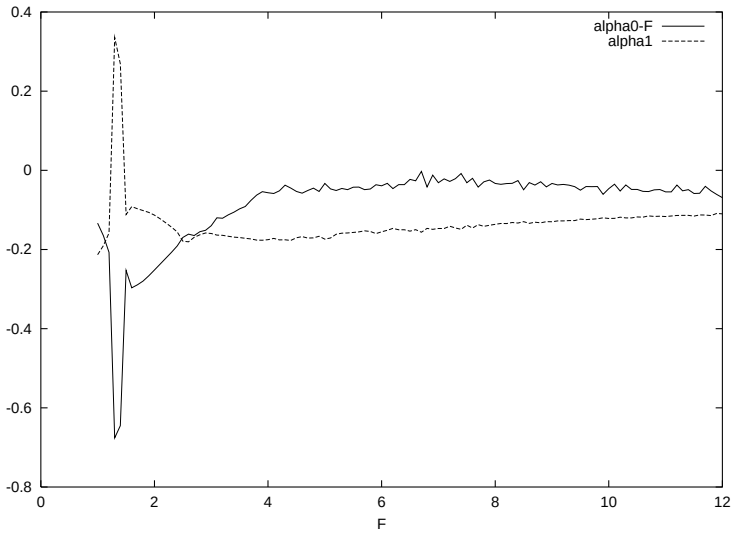


Figure 3.16: Values of constant α_0 minus F, and slope α_1 for linear model as a function of F. The spike around $F = 1.3$ occurs when the $\tilde{y}$ variables become non-zero.

3.4.1 Forcing error for the linear model

Figure 3.17 shows model error for the linear model for $F = 10$, along with initial condition displacement error. The errors are calculated in the same RMS manner as was used for the constant model errors in Figure 3.4. Figure 3.18 is a close-up of initial model error near time zero, compared with model error for the constant model. The model error slope is lower for the new model, and error doesn't grow at as constant a rate (the growth curve has negative curvature). The standard deviation of forcing error, which gives the error slope at time zero, is shown in Figure 3.19 as a function of F. The graph is obtained as follows: at each value of F, the true system forcing is determined, the correct values of constant α_0 and slope α_1 are determined by linear regression, the linear system forcing is subtracted from the true forcing, and the standard deviation of the result is then calculated by sampling over the attractor. When compared with the constant model forcing errors, we see an improvement of slightly more than 50 percent.

3.4.2 Shadowing for the linear model

If forcing error is a good measure of model quality, then one might expect that a 50 percent improvement in forcing error would translate into a similiar improvement in shadowing times. In fact, the improvement is considerably greater. Figure 3.21 shows the shadowing results where the shadowing radius is 0.2 and 0.4 at $F = 10$,

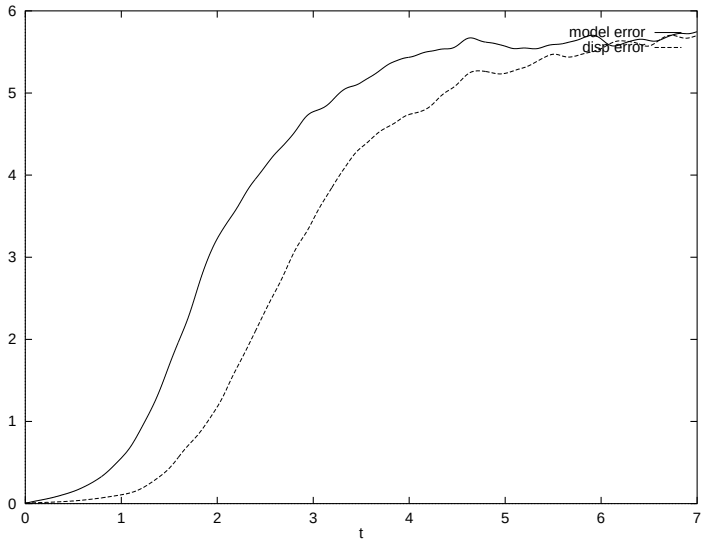


Figure 3.17: Model error and displacement error for linear model with $F = 10$.

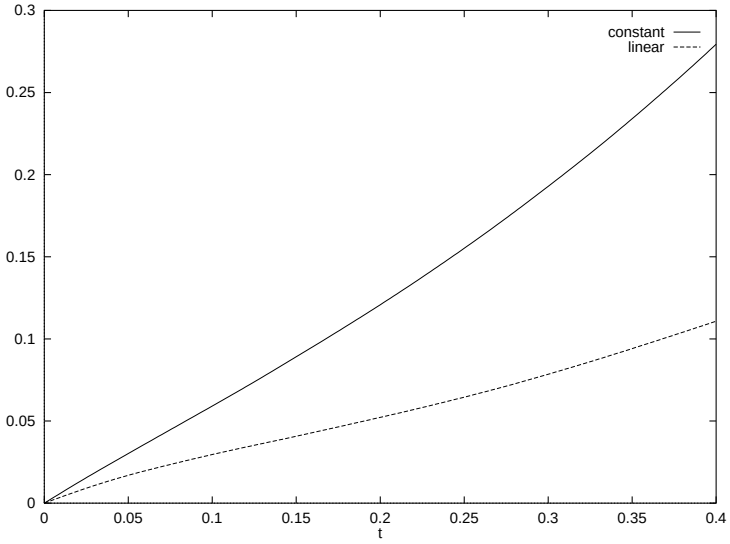


Figure 3.18: Model error near time zero for constant and linear models for $F = 10$.

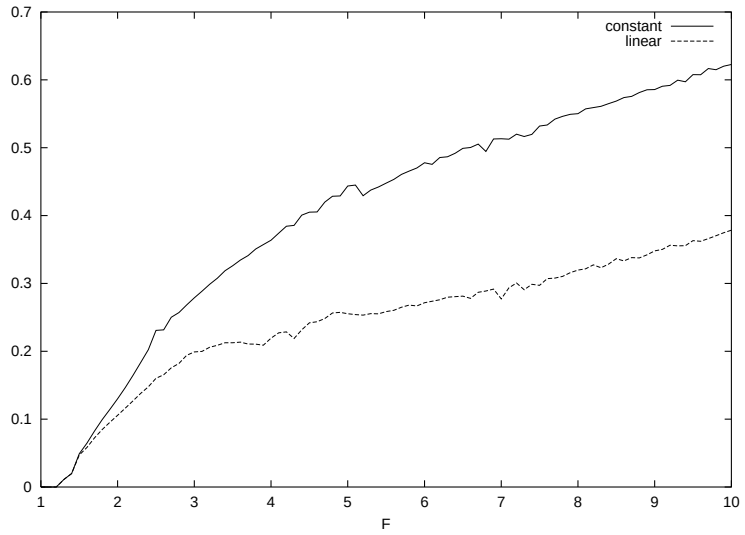


Figure 3.19: Forcing error standard deviation for constant and linear models as a function of F.

and scaled for other values of F as before. At $F = 10$, for example, the 0.4 radius shadowing times have been improved by a factor of more than 4. Figure 3.22 is a shadow orbit for $F = 10$, which compares with Figure 3.11. The particular shadow time here is 2.49 as opposed to 0.5 for the constant model case. A histogram of the results over 200 points is shown in Figure 3.20.

The reason for this dramatic increase in shadowing ability is that the forcing error gives the initial slope of the model error, but for the linear model the slope of the error curve decreases with time. In order to get a better measure of forcing error over the shadow period, we must take into account the fact that it increases nonlinearly with time.

3.5 The integrated forcing error - a spectral approach

The forcing error contains power over a range of frequencies. However, the contribution to error $e_m(t)$ of the model relative to the system over a certain time τ is frequency dependent, since higher frequencies will tend to cancel themselves out. The definition of which frequencies are high and low will depend on the reference time τ. In terms of the linearised model error dynamics, equation 3.18, we could say that the model error after a time τ depends on the vector integral of the velocity error, not

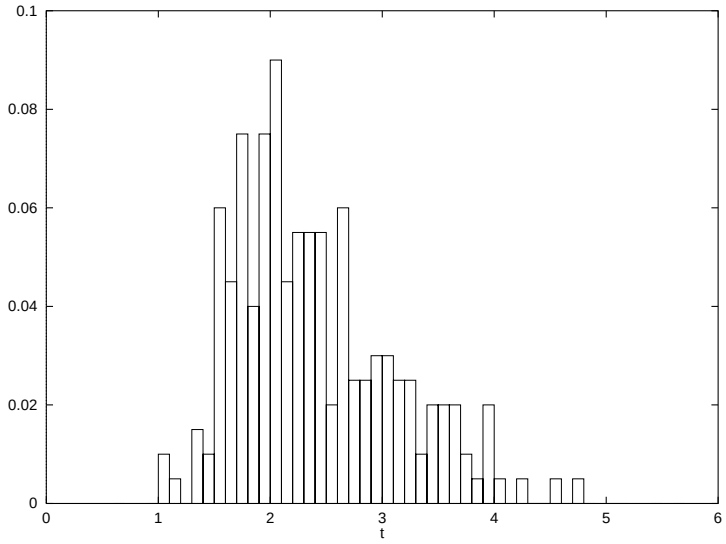


Figure 3.20: Histogram of shadowing times for the linear model with $F = 10$ over 200 starting points.

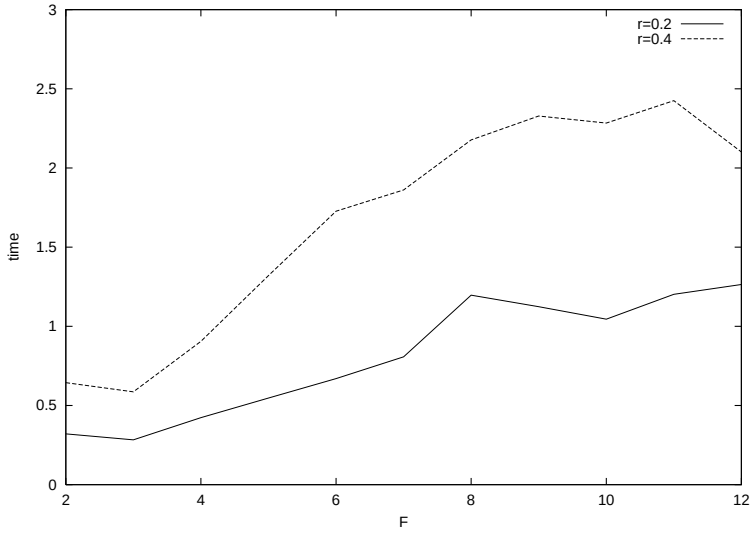


Figure 3.21: Shadowing times for linear model at integer values of F. Shadowing radius is 0.2 and 0.4 at $F = 10$, scaled at other values.

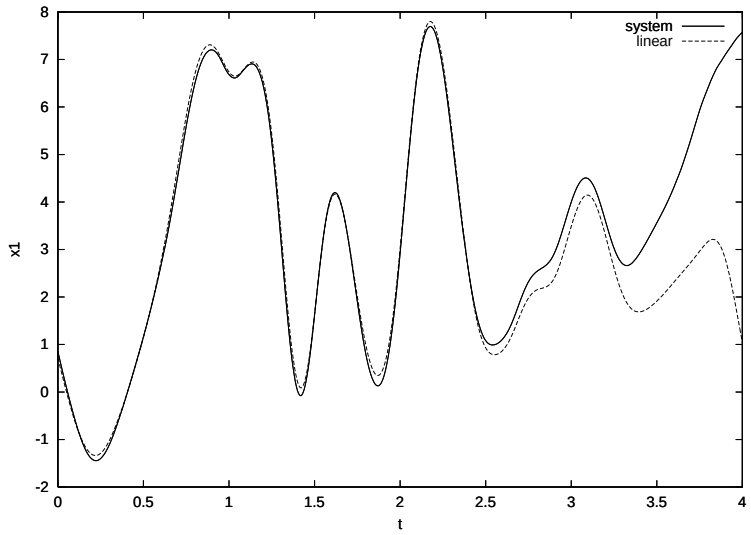


Figure 3.22: Plot of x_1 for a linear model shadowing orbit, with $F = 10$ and shadowing radius $r_s = 0.4$. The orbit cease to shadow at $\tau = 2.49$. Note the shadowing radius is over all components of $\mathbf{x}$, not just x_1.

just its initial value.

We therefore need a way to compute the total forcing error as a function of time. One way to do this, given a specific time τ, is to calculate the integral of the forcing error along a number of segments of the true attractor, all of which have length τ, and then derive the standard deviation of the resulting integral. This will be the standard deviation of the total forcing error experienced over that time, and is the approach we will adopt in the next chapter.

Another method, which is instructive and aids interpretation of general classes of error, is to use a power spectrum approach. For the constant model, we summed the terms of the forcing error power spectrum to get the variance. Since the forcing error increased linearly with time, multiplying by a time τ gave the variance of the integrated forcing error over that time (see Figure 3.14). We can do something similar in the nonlinear case, by correctly weighting each term of the spectrum to reflect its contribution to the integral. This will allow us to obtain the integrated forcing error from the forcing error spectrum, but, more importantly, will show which terms in the spectrum contribute most to model error.

The correct weighting for each term in the power spectrum is determined by considering the contribution of the corresponding sine wave to the total forcing error integral. The constant term p_0 will integrate over a time τ to $p_0\tau$. However the power p_ω at frequency ω corresponds to a cosine wave of the form $cos(\omega t + \theta)$, where θ is a

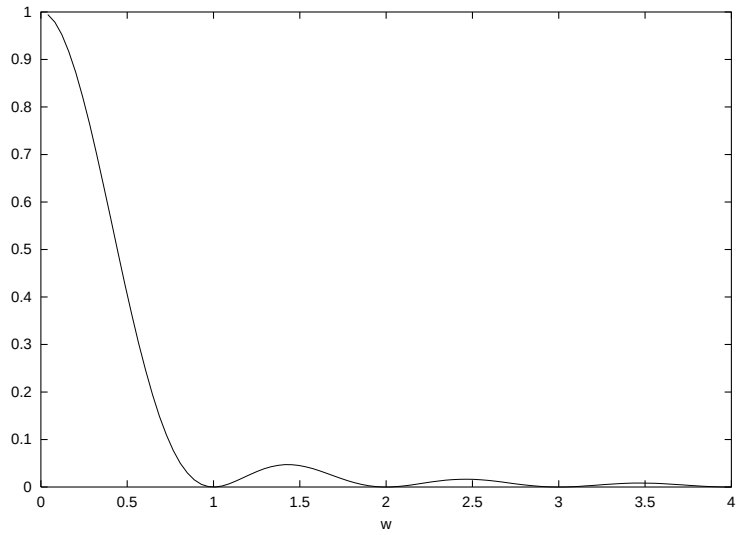


Figure 3.23: Weighting function $W(\omega, \tau)$ versus frequency ω for constant model with shadowing time $\tau = 2\pi$.

particular angle at time zero. The integral of the wave over the shadowing time is

$$\int_0^\tau cos(\omega t + \theta)dt = \frac{1}{\omega}(sin(\omega\tau + \theta) - sin(\theta)) \tag{3.35}$$

The correct weighting for this term of the power expansion, which is the square of the Fourier coefficient, is therefore

$$\frac{1}{\tau}\frac{1}{\omega^2}\langle(sin(\omega\tau + \theta) - sin(\theta))^2\rangle \tag{3.36}$$

where the expectation is over all initial angles θ. The resulting weight is

$$W(\omega, \tau) = \frac{4}{(\omega\tau)^2}sin^2(\frac{\omega\tau}{2}). \tag{3.37}$$

Figure 3.23 shows a plot of the weighting function W for $\tau = 2\pi$. It is 1.0 for frequency zero, and effectively cuts off powers with corresponding periods smaller than $2\pi/\tau$. The implication is that only forcing error frequencies with periods greater than $2\pi/\tau$ will contribute significantly to total forcing error, and therefore to expected shadowing times (subject to the caveat that the magnitudes of high frequency spectra are small compared to the shadowing radius r_s: even high frequencies will prevent the system from shadowing if the resulting oscillations are larger than the shadow radius).

In Figure 3.24 the total forcing error experienced by a typical shadow orbit for the linear model is shown. Shadow radius r_s is 0.2 and 0.4 at $F = 10$, and scaled at other values. The results have again been normalised by dividing by the shadowing

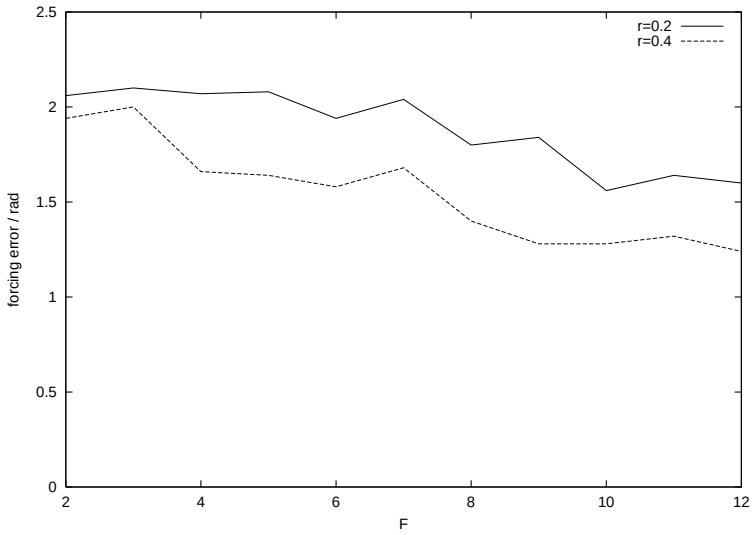


Figure 3.24: Total shadow forcing errors for linear model, normalised to shadow radius (dimensionless). Shadowing radius is 0.2 and 0.4 at $F = 10$, scaled for other values.

radius r_s, and the numbers are again near 2, except at higher values of forcing. The relationship is less exact than it was for the constant model, but it seems fair to say that the main reason the linear model shadows so much better than the constant model is because it reduces the low frequency forcing error.

This is indicated more clearly in Figure 3.25, which is a plot of the forcing error for the two systems. Forcing error for the constant model is equal to the true system forcing, minus the constant term at zero frequency (the constant model forcing doesn't vary with time). The linear model, however, reduces the low frequency spectra up to a frequency of about 3.0, while leaving the higher frequencies unchanged. We conclude that, at least in this example, the key to improving shadowing is to reduce the low frequency forcing error. The improvement in shadowing can be large, even if the reduction in total forcing error is modest.

If the forcing error has a white noise spectrum, so there is equal power at each frequency, then the integral of the forcing error will increase with the square root of time, as for a random walk [11]. Another way to interpret the success of the linear model over the constant model is therefore to say that its power spectrum is less 'red'. This will be of interest in Chapter 6 when we come to look at weather models, where errors have in the past been modelled as white or red noise.

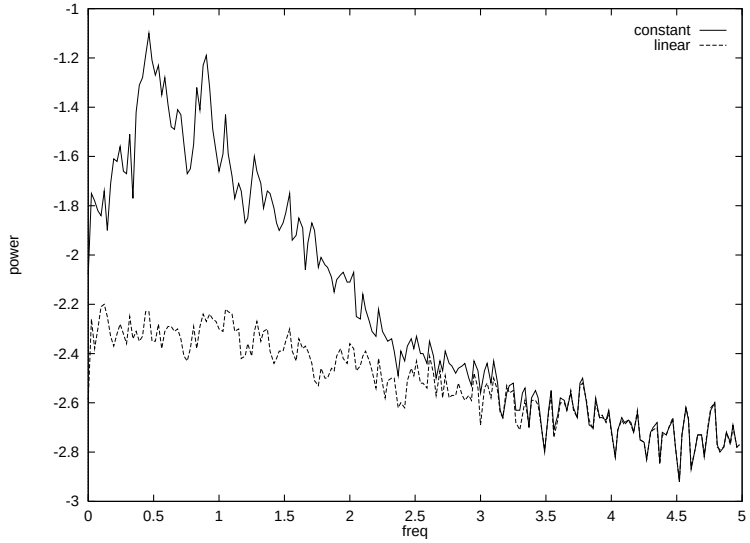


Figure 3.25: Power spectra of forcing error for constant and linear models at $F = 10$.

3.6 A local linear predictor model

A more sophisticated approach to modelling the true system is to take advantage of the fact that the true system forcing may exist on a low dimensional attractor. Two cases were examined, the normal system with $F = 10$, and a modified system where the coefficient c which determines the coupling between the $\tilde{x}$ and $\tilde{y}$ variables is set to a higher value of 1.2 and F is reduced to 2, as was done in Figure 2.15.

3.6.1 Predictor model at low forcing and high coupling

Referring to the bifurcation diagram Figure 2.13 and the system trajectories in Figure 2.15 for coupling coefficient 1.2 and forcing 2, we see that the two-level system with these parameter values is in a quasi-periodic state. If we attempt to model this system with the constant or linear models, we run into the problem that at the low level of forcing these systems are periodic and have either the wrong amplitude or the wrong frequency (or both). Figure 3.26 shows the best matches that we could find by varying the parameters. The constant model has $P^c = 1.5$, while the linear model has constant term $\alpha_0 = 1.8$ and slope $\alpha_1 = -0.2$. Neither of them shadow for the initial condition shown as long as 3 time units.

Figure 3.27 is a plot of local forcing $\tilde{F}_i$ vs $\tilde{x}_i$ for the true system, and can be compared with Figure 3.15. The linear model uses the information in the figure by drawing a straight line through the data and deriving a relationship between $\tilde{x}_i$ and the local forcing $\tilde{F}_i$. However for this system it is possible to do much better, because

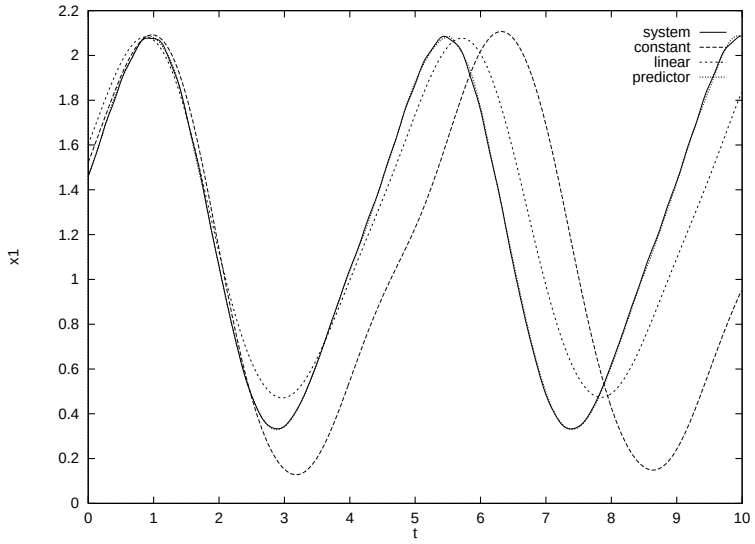


Figure 3.26: Shadowing orbits for constant, linear and predictor models with $c = 1.2$ and $F = 2$.

the forcing exists on a low-dimensional attractor: in this case, a quasi-periodic orbit. If we specify all the $\tilde{x}_i$'s then we can determine the forcing to arbitrary accuracy.

The technique used to do this is local linear prediction [61]. The first step is to construct a learning set, consisting of points $\tilde{\mathbf{x}}$ on the attractor of the true system (projected into model space) and corresponding local forcing values. The learning set is built incrementally. For each new $\tilde{\mathbf{x}}$, the existing learning set is used to predict the forcing. If the prediction fails to land within a prescribed tolerance of the true forcing, the point is added to the learning set. Predictions are made using the local linear method: nearest neighbours to the point in question are selected and a linear interpolation performed to estimate the corresponding forcing. The number of nearest neighbours used is variable, but here was set to 16. Should the forcing exist on a low dimensional attractor, the process will almost certainly converge so that the learning set effectively spans the attractor in an efficient way.

Figure 3.28 is a plot of the learning set as $\tilde{x}_2$ versus $\tilde{x}_1$, which can be compared with the orbit in Figure 2.15(a). Figure 3.29 is $(\tilde{x}_i, \tilde{F}_i)$ pairs, which compares with Figure 3.27. We see that the learning set just consists of points distributed fairly regularly over the quasi-periodic attractor.

Once a learning set has been constructed, we can define a predictor function $P_i^p(\mathbf{x})$ as follows: given a vector $\mathbf{x}$, look up the nearest neighbours in the learning set. Each of those points has an associated forcing. Perform a linear interpolation to give the

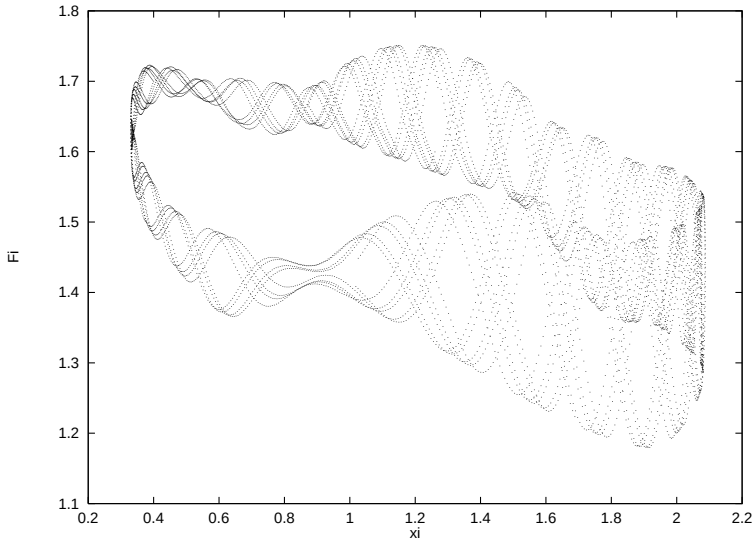


Figure 3.27: Plot of local forcing $\tilde{F}_i$ vs $\tilde{x}_i$ for true system with $c = 1.2$ and $F = 2$.

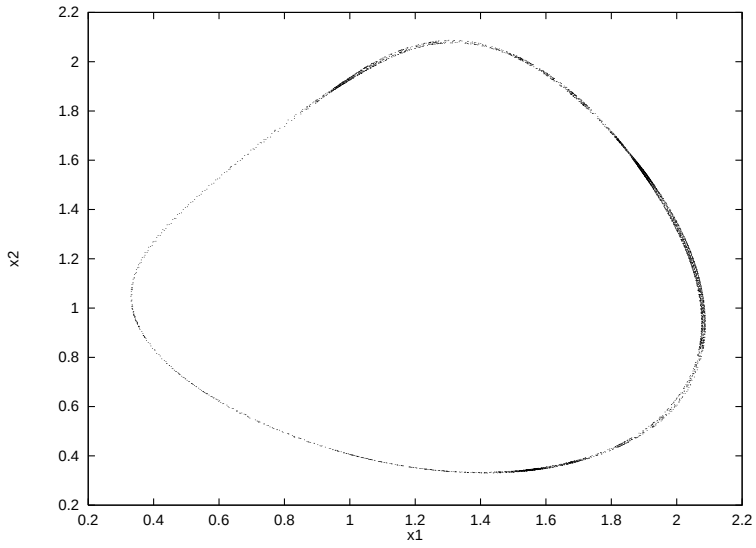


Figure 3.28: $\tilde{x}_2$ versus $\tilde{x}_1$ in the learning set.

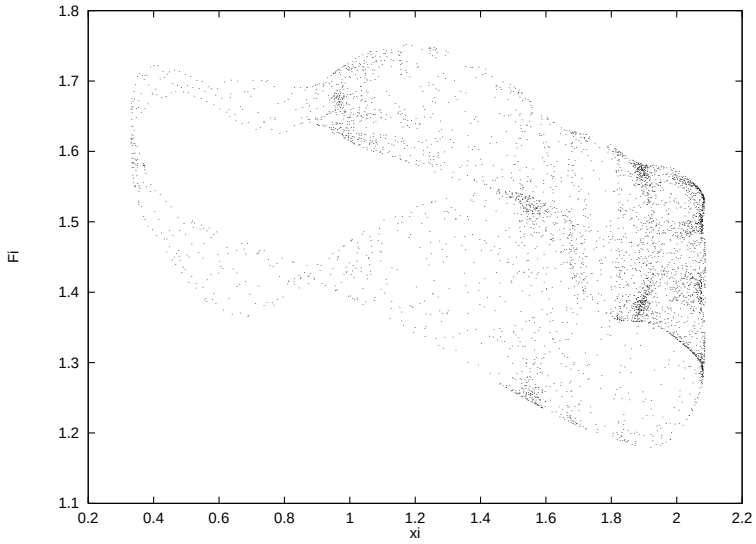


Figure 3.29: Plot of local forcing $\tilde{F}_i$ versus $\tilde{x}_i$ in the learning set.

predicted forcing for the new point. We thus arrive at the model

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + P_i^p(\mathbf{x}) \qquad \text{predictor model.} \tag{3.38}$$

Shadowing results for this system are excellent, as we would expect. The shadowing orbit in Figure 3.26 for the predictor model is nearly indistinguishable from the true system, and appears to follow the true quasi-periodic orbit indefinitely.

3.6.2 Predictor model at high forcing and regular coupling

The local linear predictor method was applied to model the true system with the normal coupling coefficient of 1.0, and $F = 10$. The prediction ability of the local linear method doesn't converge as well as for the previous case, so with a learning set of 4096 points the variance of the forcing error is about 0.6, which is the same achieved by the constant model. Much of the variance, however, is in higher frequencies, which have reduced effect on shadowing times. Figure 3.30 shows the model error curve compared with the other systems. The slope at time zero is the same as for the constant model, due to the high frequency variance, but the curve soon flattens out and the slope becomes closer to that of the linear model. We would therefore expect shadow results to be somewhere between the results for the other two models, i.e. in a range [0.23,0.8] at a shadowing radius of 0.2, and [0.5,1.9] for shadowing radius 0.4.

It is possible to improve this estimate by plotting the integrated forcing error as a function of time, which is done in Figure 3.31 for all three systems. As expected, the

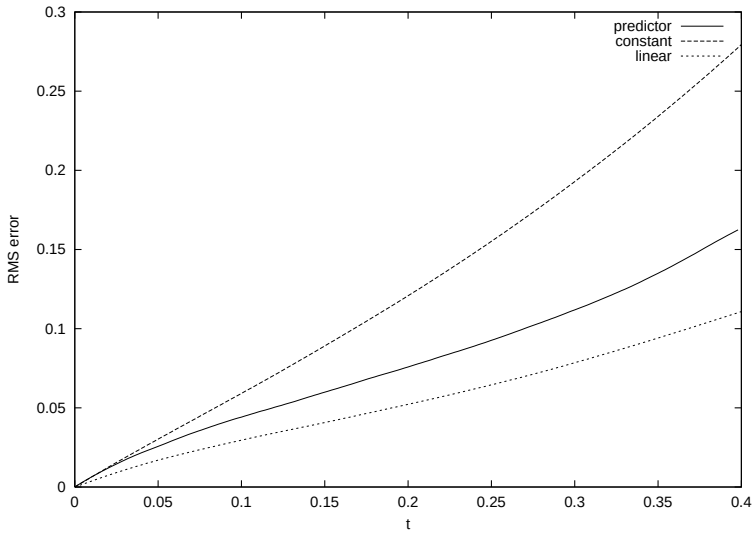


Figure 3.30: Model error for predictor model in early stages, compared with constant and linear models.

predictor model accumulates forcing error over a shadowing orbit at a rate somewhere between the two other models. Also shown are lines joining the points where the constant and linear models fail to shadow at shadowing radius 0.2 and 0.4. From this graph we would expect the predictor model to shadow for about 0.55 time units at shadow radius 0.2 and 1.2 time units at shadow radius 0.4.

Actual shadowing calculations give shadow times of 0.65 time units at radius 0.2 and 1.3 time units at radius 0.4. These are in the right range, and show that the forcing error curves give a good indication of shadowing times, even though the models being compared are quite different. Forcing error has the advantage of being much faster to calculate than explicit shadow orbits, which can be a factor for complex models such as real atmospheric models.

3.7 Summary

In this chapter we have investigated model error from a mostly empirical point of view, using the Lorenz system and its various models to explore issues such as velocity error, displacement error, and shadowing. The principal finding has been that short to medium range predictability, as measured by shadowing times, depends to a large extent on integrated forcing error, which in turn depends on the forcing error frequency spectrum. For the constant model, the integrated forcing error increases roughly linearly with time, at a rate determined by the square root of the forcing. The

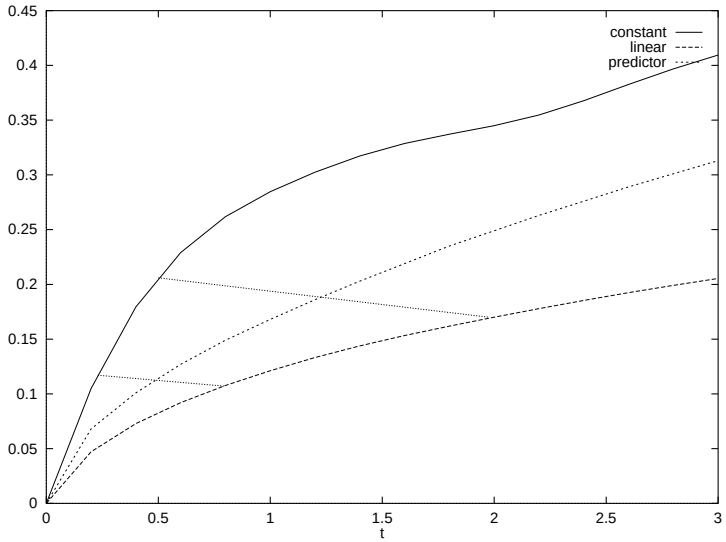


Figure 3.31: Plot of integrated forcing error versus shadow time for constant, linear and predictor models, and expected limits for shadow radius 0.2 and 0.4

fact that this holds despite the complexity of the forcing error's spectral bifurcation diagram is an unanticipated emergent property of the model.

One way to view the forcing error is like the tiller of a boat. If the tiller is held first too far to one side, then too far to the other, but with frequent corrections - as in high frequency forcing error - then the boat will tend to stay on the right path. If, however, the tiller is held too long to one side before correcting - as is the case with low frequency forcing error - then the boat will drift far off course. This is the case even if the average error over all time is zero. The constant Lorenz model has zero average forcing error, but still gives significant short to medium range prediction errors. It therefore seems possible that a weather model which also has zero average forcing error, and produces long term forecasts in balance with the climatology, may still fail in the short term.

Because of the important role of low frequency velocity error, any change to the model which addresses this will improve performance. It was shown with the linear model that relatively minor improvements in parameterisation, which succeed in reducing low frequency velocity error, can have an amplified effect on shadowing times.

While shadowing times for the constant model varied almost linearly with shadow radius and velocity error, the relationship was less clear cut for the linear model. It appears that velocity error in itself is not sufficient to predict shadow times. Since

shadowing is the result of an interplay between initial condition error and model error, this isn't surprising.

In the next chapter, we will change our approach from an experimental one - looking from the outside in - to a more detailed one, where we analyse the combined dynamics of model and initial condition error. In doing so, we develop a more sophisticated way to estimate shadowing times, which will be applicable to any dynamical model of any system.

Chapter 4

Linearised dynamics and the shadow law

For the Lorenz '96 systems, it was found that shadowing times depended largely on the velocity error of the model relative to the true system. In general, though, shadowing performance will also depend on other characteristics of the model. In effect, there is a trade off between displacement error and model error, and a shadow orbit can loosely be viewed as one which succeeds in offsetting the effect of model error by a good choice of initial displacement.

In this chapter we will further investigate these two types of error, and study how they interact. Our first aim is to develop a robust measure of model error, motivated by the results of the previous chapter. By considering the linearised dynamics, we develop a hierarchy of techniques for estimating shadow times, without the need to produce explicit shadow orbits. A shadow law, which gives a lower bound on shadow radius in terms of the model error, is derived. The methods are tested on a variety of systems, as preparation for the application to weather models in Chapter 6. Finally, we use the insights gained to propose fast methods of producing shadow orbits.

One of the main goals of this thesis is to quantify the effect of model error on shadow times. When we consider the complexity of a typical shadowing orbit, it might seem unlikely that shadow times can easily be predicted just from some measure of model error, without actually searching for shadow orbits as was done in Chapter 3. For example, Figure 4.1 is another view of the shadow orbit for the linear model shown previously in Figure 3.22. It is an attempt to picture what is going on in 8 dimensional space. The displacement vectors have first been projected onto the hyper-plane normal to the true orbit. The radius of each point is then calculated as

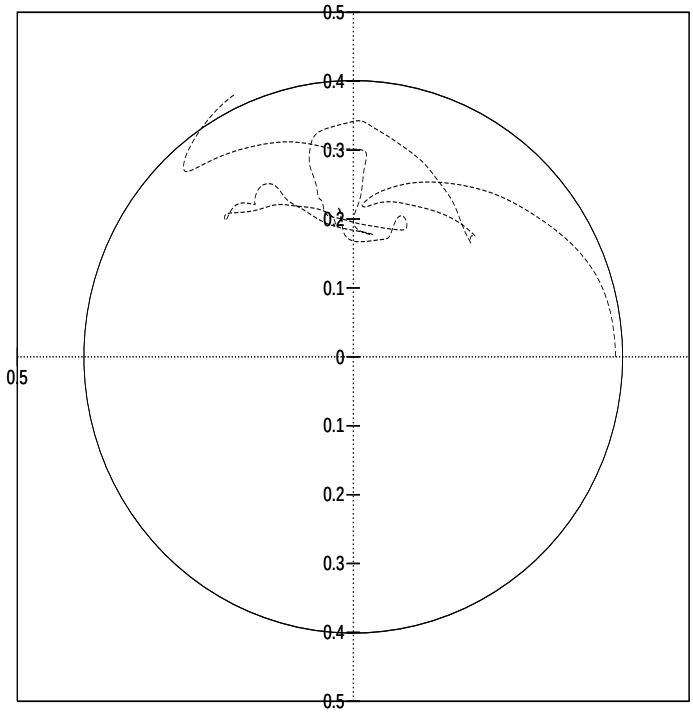


Figure 4.1: Shadowing orbit for linear model, in polar coordinates with radius the displacement from true orbit, and angle with respect to initial offset. The circle of radius 0.4 represents the shadowing radius. The shadow orbit starts near the extreme radius 0.4 on the right hand side, then exits on the upper left hand side after 2.49 time units.

the displacement of the shadow orbit from truth, while the angle is the angle of the displacement at that time with the original displacement. The shadow orbit starts near the extreme radius 0.4 on the right hand side, then exits on the left hand side after 2.49 time units. (As an aside, if we could search for the longest shadowing orbit over all possible starting points, it would always start near the outer radius. This is because, if the longest orbit started at some other radius, then we could run time backwards until the orbit exited. Using the exit point as an initial condition would then produce a longer orbit, contradicting our assumption that the original orbit was longest.)

It is clear from Figure 4.1 that the ability of a model to shadow will depend not only on model error, but on a complex interaction between model error and displacement error. A model with large sensitivity to initial condition, for example, may stand a better chance of producing an orbit that shadows, simply because nearby orbits tend to diverge out in all directions (the machine gun analogy). A model's

sensitivity to initial conditions need not be a bad thing if it is shared by the real system, but the point is that an estimate of shadow times based on model error alone couldn't take this effect into account.

A theme of this thesis (and other studies of complex systems), though, is that the systems are sometimes more complicated in the details than in their overall behaviour. For that reason, as well as the results of Chapter 3, there are grounds for optimism that even complicated shadow trajectories can, at least in an average sense, show certain predictable properties. We begin the search for such properties, though, by considering, not macroscopic behaviour, but what one might consider the opposite: the linearised dynamics of error evolution.

4.1 The linearised dynamics

It was found in the previous chapter that a useful indicator of shadowing ability was the integrated velocity error, which was calculated by linearising the model error and integrating. Motivated by that result, we now apply the same technique to the interaction between displacement error and model error, by linearising both types of error around the true orbit. The following theorem states that, even in the presence of model error, the evolution of errors can be approximated by the linearised dynamics.

Theorem. Let $\tilde{\mathbf{x}}(t)$ be a solution of the system equation

$$\frac{d\tilde{\mathbf{x}}}{dt} = \tilde{\mathbf{G}}(\tilde{\mathbf{x}}), \tag{4.1}$$

and let $\mathbf{x}(t)$ be a solution of the model equation

$$\frac{d\mathbf{x}}{dt} = \mathbf{G}(\mathbf{x}), \tag{4.2}$$

where $\tilde{\mathbf{G}}$ and $\mathbf{G}$ are C^1. If the true system and the model system exist in separate spaces, then as before we implicitly assume the existence of a projection operator taking the true system into the model system space, but omit it from the equations below for clarity. Define $\mathbf{e}(t) = \mathbf{x}(t) - \tilde{\mathbf{x}}(t)$, $\mathbf{G_e}(\tilde{\mathbf{x}}(t)) = \mathbf{G}(\tilde{\mathbf{x}}(t)) - \tilde{\mathbf{G}}(\tilde{\mathbf{x}}(t))$. The linear propagator of the model around the true orbit is

$$\mathbf{M}(t) = e^{\int_0^t \mathbf{J}(\tilde{\mathbf{x}}(t))dt}, \tag{4.3}$$

where $\mathbf{J}$ is the Jacobian of $\mathbf{G}$. We also define the propagator from time s to time t as

$$\mathbf{M_s}(t) = e^{\int_s^t \mathbf{J}(\tilde{\mathbf{x}}(t))dt}. \tag{4.4}$$

Then given a reference time $\tau > 0$, and $\epsilon > 0$, there exists a radius $r > 0$ such that, if $\|\mathbf{e}(t)\| < r$ for all $t \in [0, \tau]$, then

$$\|\mathbf{e}(t) - \mathbf{M}(t)\mathbf{e}(0) - \int_0^t \mathbf{M_s}(t)\mathbf{G_e}(\tilde{\mathbf{x}}(t))dt\| < \epsilon \quad \forall t \in [0, \tau]. \tag{4.5}$$

Proof. From the system equations, we can write

$$\begin{aligned} \frac{d\mathbf{e}}{dt} &= \frac{d\mathbf{x}}{dt} - \frac{d\tilde{\mathbf{x}}}{dt} \\ &= \mathbf{G}(\mathbf{x}(t)) - \tilde{\mathbf{G}}(\tilde{\mathbf{x}}(t)) \\ &= \mathbf{G}(\mathbf{x}(t)) - \mathbf{G}(\tilde{\mathbf{x}}(t)) + \mathbf{G}_e(\tilde{\mathbf{x}}(t)) \\ &= \mathbf{G}(\tilde{\mathbf{x}}(t) + \mathbf{e}(t)) - \mathbf{G}(\tilde{\mathbf{x}}(t)) + \mathbf{G}_e(\tilde{\mathbf{x}}(t)). \end{aligned} \tag{4.6}$$

Performing a Taylor expansion of $\mathbf{G}$ around $\tilde{\mathbf{x}}(t)$, and retaining only the zero and first order term, we obtain

$$\frac{d\mathbf{e}}{dt} = \mathbf{J}(\tilde{\mathbf{x}}(t)) \cdot \mathbf{e}(t) + \mathbf{G}_e(\tilde{\mathbf{x}}(t)) + \mathbf{R}(t) \tag{4.7}$$

where the remainder term $\mathbf{R}(t)$ is $O(\|\mathbf{e}(t)\|^2)$. Therefore, $\exists r_t > 0 \quad \ni \quad \|\mathbf{e}(t)\| < r_t \implies \|\mathbf{R_d}(t)\| < \frac{\epsilon}{\tau}$. Pick r to be the minimum such r_t (possible since $[0, \tau]$ is a compact set). Integrating from 0 to t for $0 < t \leq \tau$ then gives [48]

$$\mathbf{e}(t) = \mathbf{M}(t)\mathbf{e}(0) + \int_0^t \mathbf{M_s}(t)\mathbf{G_e}(\tilde{\mathbf{x}}(t))dt + \int_0^t \mathbf{R}(t)dt, \tag{4.8}$$

and

$$\|\mathbf{e}(t) - \mathbf{M}(t)\mathbf{e}(0) - \int_0^t \mathbf{M_s}(t)\mathbf{G_e}(\tilde{\mathbf{x}}(t))dt\| = \|\int_0^t \mathbf{R}(t)dt\| < \frac{\epsilon}{\tau}t \leq \epsilon \tag{4.9}$$

which proves the result.

Now

$$\int_0^t \mathbf{M_s}(t)\mathbf{G_e}(\tilde{\mathbf{x}}(t))dt = \int_0^t \mathbf{G_e}(\tilde{\mathbf{x}}(t))dt + \int_0^t (\mathbf{I} - \mathbf{M_s}(t))\mathbf{G_e}(\tilde{\mathbf{x}}(t))dt. \tag{4.10}$$

Let

$$\mathbf{d}(t) = \int_0^t \mathbf{G}_e(\tilde{\mathbf{x}}(t))dt. \tag{4.11}$$

Since the linear propagator approaches the identity matrix as time (or shadow radius) goes to zero, it is easily seen that the term

$$\int_0^t (\mathbf{I} - \mathbf{M_s}(t))\mathbf{G_e}(\tilde{\mathbf{x}}(t))dt \tag{4.12}$$

is $O(r^2)$, and acts as a relatively small perturbation on $\mathbf{d}(t)$. We will therefore neglect this higher order term, but return to estimate it for weather models in Chapter 6.

Therefore the evolution of the error $\mathbf{e}(t)$ can be approximated by the linearised dynamics

$$\mathbf{e}(t) \approx \mathbf{y}(t) + \mathbf{d}(t). \tag{4.13}$$

where

$$\mathbf{y}(t) = \mathbf{M}(t)\mathbf{e}(0). \tag{4.14}$$

Note the approximation only holds for model orbits $\mathbf{x}(t)$ which remain within the tolerance r of the reference trajectory $\tilde{\mathbf{x}}(t)$. In other words, it only holds for orbits which shadow at that radius. The size of the radius will depend on the system, the model, and the allowed error ϵ.

Note also that the linear propagator $\mathbf{M}(t)$ is now calculated along the true trajectory $\tilde{\mathbf{x}}(t)$, rather than the model trajectory. It can determined by directly integrating the model Jacobian along the true orbit, but this is a lengthy procedure. A commonly used alternative is to estimate the linear propagator by computing the trajectories of $(n+1)$ (or more) slightly displaced orbits. In this case, though, the technique cannot be applied directly since the orbits will be incorrect. A way round this problem is to consider the modified system

$$\frac{d\mathbf{x}}{dt} = \mathbf{G}(\mathbf{x}(t)) - \mathbf{G}_e(\tilde{\mathbf{x}}(t)) \tag{4.15}$$

where $\tilde{\mathbf{x}}(t)$ is the true orbit which we are trying to shadow. The function $\mathbf{G}_e(\tilde{\mathbf{x}}(t))$ is a function only of t, and the Jacobian of $\mathbf{G} - \mathbf{G}_e$ is exactly the same as the Jacobian $\mathbf{J}$ of $\mathbf{G}$. However the modified system 4.15 also has the property that $\tilde{\mathbf{x}}$ is a solution. Therefore the linear propagator for this system, evaluated on the $\tilde{\mathbf{x}}$ orbit, is the integral of the Jacobian over that orbit, as required.

Since $\mathbf{G}$ and $\tilde{\mathbf{G}}$ specify the state space velocities of the model and true system respectively, the function $\mathbf{G}_e$ is just the velocity error. The vector $\mathbf{d}(t)$ in 4.11 is therefore the integral of the velocity error along the true orbit, and has the dimension of distance. Setting $\mathbf{e}(0)$ equal to the zero vector in (4.13), we have

$$\mathbf{e}(t) \approx \mathbf{d}(t) \tag{4.16}$$

so $\mathbf{d}(t)$ is the approximate displacement of a model solution, which is started on the true initial condition, from the true solution after a time t. Define the *local model drift* after a time t to be

$$d(t) = \|\mathbf{d}(t)\| \tag{4.17}$$

(the drift is of course also dependent on the true orbit $\tilde{\mathbf{x}}$ which is taken as set). The model drift is then a good indicator of the model's predictive capacity: the smaller

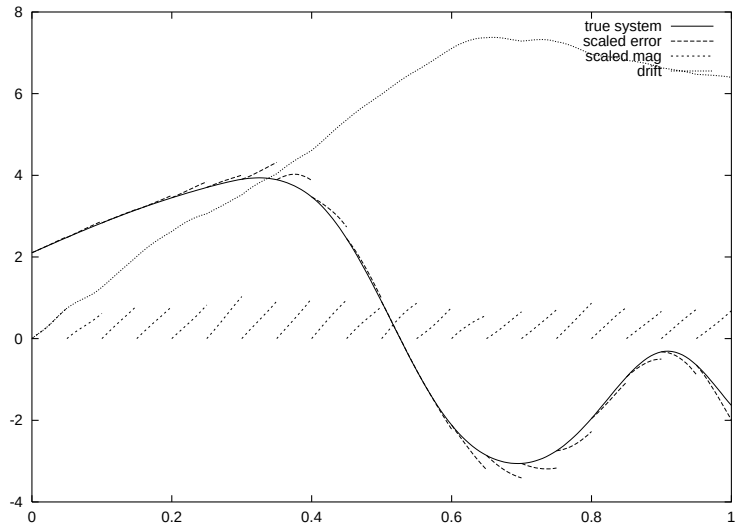


Figure 4.2: Forecast errors for Lorenz model/system, x_1 component. As in Figure 3.2, a number of short term model forecasts were initiated at regular intervals along a true trajectory. The errors have here been scaled by a factor 10. The total error magnitude over all x_i is also shown, again scaled by a factor 10. Because it contains all components, it is larger than the x_1 error. The drift, also shown scaled by a factor 10, is the vector sum of the total velocity errors. The almost linear increase up to $t = 0.6$ indicates that the error vectors are accumulating, and are therefore in a similar direction. Above $t = 0.6$, however, the drift begins to reduce because the velocity error has rotated away from its original orientation, and projects negatively onto the drift.

the drift $d(t)$, the better the prediction of the true system's position after the elapsed time t. In the Lorenz '96 systems, velocity error on the true orbit is due entirely to forcing error, so for those systems the drift is the same as integrated forcing error.

Figure 3.2 in Chapter 3 showed the velocity errors for the Lorenz system. Figure 4.2 is similar, but the scaling of the errors has been reduced from 20 to 10, and the drift, which is the vector integral of the velocity errors, is also shown. Because the drift is state dependent, it is not necessarily systematic over long periods, but it may tend to accumulate in the short to medium term. In the figure, the drift increases steadily at first, which one would expect if the velocity errors were in roughly the same direction. Above $t = 0.6$, however, the drift begins to reduce, implying that the velocity error has now rotated so that its dot product with the drift is negative.

4.2 Model error vs initial condition error (continued)

The linearised dynamics provide a conceptual framework which helps to separate out the respective roles in the shadow process of model error and initial condition error, and in the next section this framework is used to develop a technique for estimating shadow times. First, though, it is worth clarifying the fact that these two forms of error, due to the model and displacement, need bear no special relation to each other: the latter is a property of the model alone, while the former depends also on the true system. For example, suppose that, after a given time τ, we calculate the leading singular vectors of the model's linear propagator, which define the direction of displacement which gives largest growth at time τ. These directions will depend purely on the model, regardless of the true system. The drift vector $\mathbf{d}(t)$, however, is a measure of the difference between the model and the true system. There is no reason for it to be aligned, or not, with the leading singular vectors. As a consequence, an ensemble of initial conditions, formed by perturbing in the directions of the leading singular vectors (as is done at ECMWF), need not offset model error.

For example, Figure 4.3 shows errors of the constant model relative to the two level Lorenz system (compare also Figure 3.4). In the upper panel, perturbations of size 0.2 are added to the model initial condition in the positive and negative directions of the leading singular vector, to form a two-member ensemble. Relative to the model, these perturbations have grown at time 0.34 by about a factor 5.0. Also shown in the background is the density of errors found by randomly perturbing the initial condition by an amount 0.2 and taking a histogram of the resulting errors over 1000 runs. The singular vector perturbations give maximum displacement for $t = 0.34$, as expected by construction, but not for higher times.

In the lower panels, where errors are shown relative to truth, the situation is very different. The errors are larger than for the previous case, so if ensemble spread is measured relative to truth it will be larger than if measured relative to the model control. Also, neither singular vector perturbation effectively offsets model error, compared to the random displacements. The negative perturbation nearly captures the maximum error, but other random displacements do slightly better: the 'worst-case' displacement now depends on the drift vector.

This figure is a graphic illustration of two facts which must be taken into account when model error is significant. Firstly, spread will appear smaller if measured with

respect to the model than to truth. Secondly, no member of the ensemble need succeed in counteracting the effect of model error.

Figure 4.4 is a schematic diagram showing how model error can distort an ensemble and affect the spread. The initial perturbations $\mathbf{v}^+$ and $\mathbf{v}^-$, aligned with the leading singular vector, evolve to $\mathbf{u}_\mathbf{m}^+$ and $\mathbf{u}_\mathbf{m}^-$ under the model dynamics, but to $\mathbf{u}_\mathbf{t}^+$ and $\mathbf{u}_\mathbf{t}^-$ under the true dynamics. The angle θ_t therefore shrinks, and the vectors $\mathbf{u}_\mathbf{t}^+$ and $\mathbf{u}_\mathbf{t}^-$ are no longer approximately anti-parallel. Since the evolved perturbations with respect to truth are larger than those with respect to the model, the model spread is expected to be smaller than the true spread.

A similar effect can be seen in 4.5, which shows the cosine of the angle enclosed by the positive and negative perturbations for the constant model. At initial time the cosine angle is -1, indicating that the perturbations are anti-parallel. For times up to about $t = 0.6$ the perturbations taken with respect to the model remain nearly anti-parallel, but with respect to truth the cosine angle actually become positive. This implies that both perturbations have effectively crossed over to the same side of the true orbit, not a desirable property if the ensemble is supposed to encompass truth [23].

A good 'sanity test' for any ensemble, therefore, is to take dot products of perturbations in this manner, and follow their evolution with time. This was done in detail for the ECMWF models by Gilmour [23], [24], with the difference that perturbations were measured relative to the model control itself, as oppposed to the observed weather, so model error wasn't a factor. It was found that the test fails anyway due to nonlinearity of the model - a separate problem.

4.3 The shadow estimation technique (SET)

Because the linearised dynamics model the evolution of small errors around a true trajectory, they can be used to model the shadowing process. In this section, we develop a hierarchy of tests which allow the determination of approximate shadow times for a given model/system pair.

Referring to equation 4.14 of the linearised dynamics, we can write the linear propagator matrix in its singular value decomposition (SVD) form [25] as

$$\mathbf{M}(t) = \mathbf{U}(t)\mathbf{\Sigma}(t)\mathbf{V}^T(t). \tag{4.18}$$

If $\mathbf{M}$ is an n by n matrix, then $\mathbf{U}$ and $\mathbf{V}$ are matrices of the same dimension with orthonormal columns, while $\mathbf{\Sigma}$ is a diagonal matrix with positive diagonal entries. The

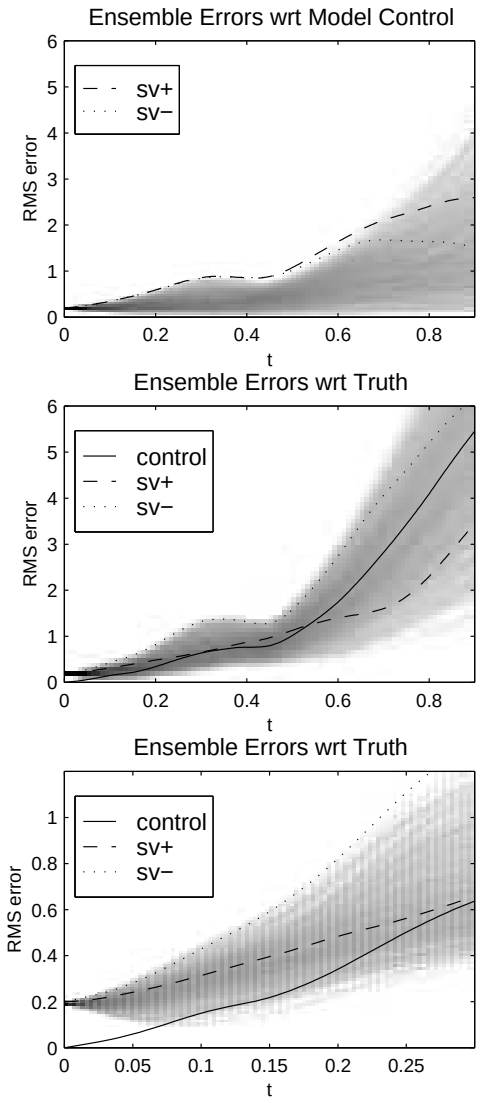


Figure 4.3: Error growth of perturbed initial conditions for the one level model. In the upper panel, errors are relative to the unperturbed model control. The background contours show error growth for 1000 random perturbations of magnitude 0.2, while the dashed and dotted lines show the perturbations in the positive and negative directions of the leading singular vector. The optimisation time for the singular vector calculation is 0.34; perturbations in these directions give maximum growth at that time. The middle panel shows errors relative to a trajectory of the two level system, and therefore include the effect of model error. The error of the model control is also shown. The lower panel is a zoom of the middle panel near initial time. Note the effect of model error on the ensemble, and the fact that no ensemble member offsets model error over the optimisation time.

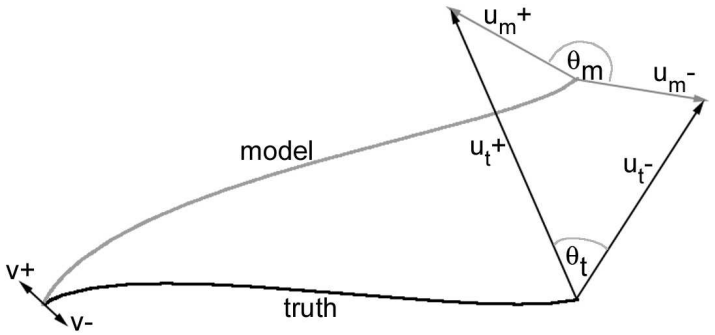


Figure 4.4: Schematic diagram showing how the angle between the positive and negative perturbations in an ensemble can shrink when taken with respect to truth (θ_t) as opposed to the model (θ_m). See also [23], [24]

.

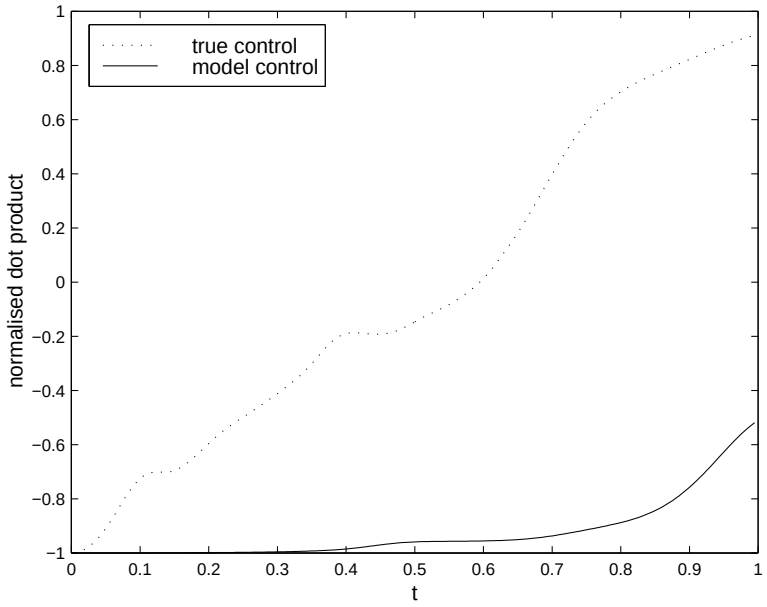


Figure 4.5: Normalised dot product (or cosine angle) of the positive and negative singular vector perturbations for the Lorenz model/system, with respect to the true system (dashed line) and the model (solid line).

linear propagator maps the i'th column $\mathbf{v}_i$ of $\mathbf{V}$ to the i'th column $\mathbf{u}_i$ of $\mathbf{U}$ multiplied by the i'th diagonal entry σ_i of $\mathbf{\Sigma}$. Hyperspheres in the space of $\mathbf{V}$ columns are therefore mapped to ellipsoids in the space of $\mathbf{U}$ columns, where each axis is multiplied by a factor σ_i.

Referring to equation (4.14), it therefore follows that if the initial condition lies within a ball of radius r_s about $\tilde{\mathbf{x}}(0)$, then after an elapsed time t the point $\mathbf{y}(t)$ will satisfy the ellipsoid equation

$$\sum_{i=1}^{n} \frac{(\mathbf{y}(t) \cdot \mathbf{u_i}(t))^2}{\sigma_i(t)^2} < r_s^2 \tag{4.19}$$

where $\mathbf{y}(t) \cdot \mathbf{u_i}(t)$ is the projection of $\mathbf{y}(t)$ onto the i'th basis vector given by the columns of $\mathbf{U}(t)$. From equation (4.11), this is the same as

$$\sum_{i=1}^{n} \frac{((\mathbf{e}(t) - \mathbf{d}(t)) \cdot \mathbf{u_i}(t))^2}{\sigma_i(t)^2} < r_s^2. \tag{4.20}$$

Now, an initial condition displaced by the vector $\mathbf{e}(0)$ will shadow until a time t if $\|\mathbf{e}(t)\| < r_s$ where r_s is the shadow radius. Because $\mathbf{e}(t)$ is in the ellipsoid given by (4.20), this is the same as saying that the distance between the origin and the offset ellipsoid should be smaller than the shadow radius r_s.

The direct way to solve this problem is to find the initial displacement which has the smallest final displacement under the linearised dynamics. Before doing so, we first note that, to a good approximation, the desired result will be true if the zero vector lies within an enlarged ellipsoid, where all the axes have been increased by an amount r_s (the agreement is exact at the poles, and very close elsewhere). This is shown schematically in Figure 4.6. The enlarged ellipse is all images

$$\mathbf{U}(t)(\mathbf{\Sigma}(t)+\mathbf{I})\mathbf{V}^T(t)\cdot\mathbf{e}(0)+\mathbf{d}(t), \tag{4.21}$$

where the identity $\mathbf{I}$ has been added to $\mathbf{\Sigma}(t)$ to stretch each axis an amount r_s. We therefore obtain a simple shadowing condition. The model will shadow the true system for a time τ if τ is the smallest positive time such that

$$\sum_{i=1}^{n}\frac{(\mathbf{d}(\tau)\cdot\mathbf{u_i}(\tau))^2}{(1+\sigma_i(\tau))^2}=r_s^2. \tag{4.22}$$

Since the geometric argument of expanding each axis of the ellipse was only approximate, the above condition will give shadow times which are slightly incorrect. In fact, it is easy to see that it will tend to slightly underestimate shadow times of the linearised dynamics. Suppose that condition (4.22) is satisfied. Then the ellipse of images contains the zero vector, and there is an initial displacement $\mathbf{e}$ which satisfies

$$\mathbf{U}(t)(\mathbf{\Sigma}(t)+\mathbf{I})\mathbf{V}^T(t)\cdot\mathbf{e}+\mathbf{d}(t)=\mathbf{0}. \tag{4.23}$$

Rearrangement gives

$$\mathbf{U}(t)\mathbf{\Sigma}(t)\mathbf{V}^T(t)\cdot\mathbf{e}+\mathbf{d}(t)=-\mathbf{U}(t)\mathbf{V}^T(t)\mathbf{e}. \tag{4.24}$$

Now $\|\mathbf{U}(t)\mathbf{V}^T(t)\mathbf{e}\|=\|\mathbf{e}\|$ since the matrices $\mathbf{U}(t)$ and $\mathbf{V}^T(t)$ are orthonormal. Also, by assumption, $\|\mathbf{e}\|\leq r_s$, so it follows that

$$\|\mathbf{U}(t)\mathbf{\Sigma}(t)\mathbf{V}^T(t)\cdot\mathbf{e}+\mathbf{d}(t)\|\leq r_s \tag{4.25}$$

and the point $\mathbf{e}$ shadows under the linearized dynamics. The approximation therefore will tend to underestimate shadow times.

A more accurate estimate of shadow times using the linearised dynamics can be attained by directly solving the following problem:

$$\begin{aligned}\text{minimise } C(\mathbf{e}) &= \|\mathbf{U}(t)\mathbf{\Sigma}(t)\mathbf{V}^T(t)\cdot\mathbf{e}(t)+\mathbf{d}(t)\|\\ \text{subject to } \|\mathbf{e}(0)\| &\leq r_s.\end{aligned} \tag{4.26}$$

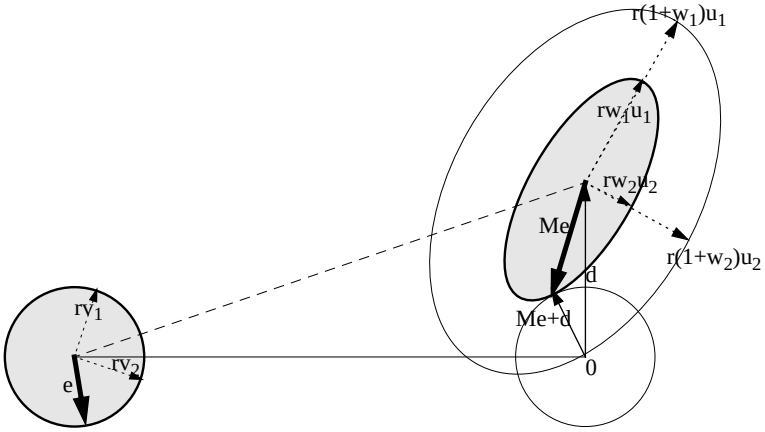


Figure 4.6: Schematic diagram of linearised shadow dynamics, showing the initial ball of radius r about the starting point; the final ellipse at the shadow time with axes $r\sigma_i \mathbf{u}_i$; the drift vector $\mathbf{d}$; and the optimal initial displacement $\mathbf{e}$ which satisfies $\|\mathbf{Me} + \mathbf{d}\| = r$.

In other words, find the initial displacement $\mathbf{e}(0)$ which has the smallest final displacement (we neglect for now the checking of displacements at intermediate times). This optimisation problem can be solved using the Lagrange method. Dropping the dependence on time for clarity, we seek stationary points of

$$(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{e}+\mathbf{d})^T(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{e}+\mathbf{d})+\lambda\mathbf{e}^T\mathbf{e} \tag{4.27}$$

where λ is a scalar Lagrange multiplier. Setting the variation equal to zero gives

$$\mathbf{V}\mathbf{\Sigma}\mathbf{U}^T(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{e}+\mathbf{d})+\lambda\mathbf{e}=0. \tag{4.28}$$

Solving for e, and using the fact that $\mathbf{U}^T\mathbf{U}=\mathbf{V}^T\mathbf{V}=\mathbf{I}$, we find

$$\mathbf{e}=-(\mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T+\lambda\mathbf{I})^{-1}\mathbf{V}\mathbf{\Sigma}\mathbf{U}^T\mathbf{d}. \tag{4.29}$$

Now it is easily seen by direct substitution that

$$(\mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T+\lambda\mathbf{I})^{-1}=\mathbf{V}\mathbf{A}(\lambda)\mathbf{V}^T \tag{4.30}$$

where $\mathbf{A}(\lambda)$ is the diagonal matrix with i'th diagonal entry

$$a_i(\lambda)=(\sigma_i^2+\lambda)^{-1}. \tag{4.31}$$

Substituting into the expression for e gives

$$\mathbf{e}=-\mathbf{V}\mathbf{A}(\lambda)\mathbf{\Sigma}\mathbf{p} \tag{4.32}$$

where $\mathbf{p}=\mathbf{U}^T\mathbf{d}$.

The next step is to solve for λ. We first note that the linearised dynamics map points $\mathbf{e}$ on the outside of the shadow radius ball to image points on the outside of the ellipse. The maximum shadow time under the linearised dynamics (not necessarily the real dynamics) will occur when there is just one point that shadows, so we can assume that the image is on the perimeter of the ellipse, and therefore that $\mathbf{e}$ satisfies $\|\mathbf{e}\|=r_s$. Referring to equation (4.32), this implies

$$\mathbf{e}^T\mathbf{e}=\mathbf{p}^T(\mathbf{\Sigma}\mathbf{A}(\lambda))^2\mathbf{p}=r_s^2, \tag{4.33}$$

which is equivalent to

$$\sum_{i=1}^{n}(p_i\sigma_i a_i(\lambda))^2-r_s^2=0. \tag{4.34}$$

The multiplier λ can therefore be found by using Newton's method to find zeros of the above expression. Once λ has been determined, we solve for $\mathbf{e}$ using (4.32).

The image point $\mathbf{U\Sigma V}^T\mathbf{e}+\mathbf{d}$ is then seen to be $\lambda\mathbf{UA}(\lambda)\mathbf{p}$, with magnitude equal to the square root of $\lambda^2\sum_{i=1}^{n}(p_i a_i(\lambda))^2$.

We thus arrive at a new shadowing condition, which is in two stages. Given a specified time t, first calculate the initial displacement which yields the minimum displacement at that time. Then check to see whether the magnitude of the image point is smaller than the shadow radius. If it is, then the model shadows until time t under the linearised dynamics (again neglecting what happens at intermediate times, which need also to be checked).

The shadowing condition involves more computation than (4.22), but is easy to implement. Also, since (4.22) will tend to underestimate the shadow time, the detailed test need only be carried out when the simpler test fails. For the models studied here, it usually adds about a percent or less, as measured in terms of the allowable drift over a shadow orbit, and the difference goes to zero as the shadow radius decreases.

Finally, since the multipliers σ_i are all positive, we can write

$$\sum_{i=1}^{n}\frac{(\mathbf{d}(\tau)\cdot\mathbf{u_i}(\tau))^2}{(1+\sigma_i(\tau))^2}\leq\sum_{i=1}^{n}(\mathbf{d}(t)\cdot\mathbf{u_i}(t))^2=\sum_{i=1}^{n}\mathbf{d}_i(t)^2=d(t)^2 \tag{4.35}$$

where we have also used the fact that the vectors u form an orthonormal basis. It follows from equation (4.22) that if the drift $d(\tau)$ at time τ is smaller than r_s, the model should shadow at least until that time.

We therefore can apply a hierarchy of shadow tests, each of increasing complexity. For increasing times t we first test equation (4.35), to see if the drift is smaller than the shadow radius. If this fails, we test equation (4.22), to check if the enlarged ellipse contains the zero vector. When that fails, we can do a full solution of the eigenvalue problem (if desired, though its effect is small). We shall refer to this procedure as the *shadow estimation technique*, or SET. The SET depends only on the model drift, the modified linear propagator, and the shadowing radius, and is applicable whenever the modified linear propagator is a good approximation to the system dynamics at distances smaller than the shadow radius from the true attractor.

We can now quantify our observation that a model which tends to scatter orbits in all directions given small displacements may shadow quite well (a shot gun may hit its target better than a well-aimed rifle). Such a model will have large multipliers σ_i, so that a ball of initial conditions blows up into a large ellipse. This will reduce the left hand side of equation (4.22). If we compare between models where these multipliers are similar, then the dominant factors are model drift and shadowing radius. In the next section, we derive a law that applies to any model which is locally dissipative, i.e. more like a rifle than a shot gun.

4.4 A shadowing law

Under certain circumstances, shadow times can be estimated by considering model drift only, without recourse to the linear propagator. Suppose first that shadow times are relatively short, so that the singular value multipliers (which tend to unity as t goes to zero) are close to 1.0. Then from equation 4.22 we have

$$\begin{aligned} r_s^2 &= \sum_{i=1}^{n} \frac{(\mathbf{d}(\tau) \cdot \mathbf{u_i}(\tau))^2}{(1+\sigma_i(\tau))^2} && (4.36) \\ &\approx \sum_{i=1}^{n} \frac{(\mathbf{d}(t) \cdot \mathbf{u_i}(t))^2}{4} && (4.37) \\ &= \frac{\sum_{i=1}^{n} \mathbf{d_i}(t)^2}{4} && (4.38) \\ &= \frac{d(\tau)^2}{4}. && (4.39) \end{aligned}$$

The shadowing time τ then satisfies

$$d(\tau) = \|\mathbf{d}(\tau)\| \approx 2r_s \tag{4.40}$$

and so is the time at which the model drift exceeds the shadow diameter.

In such cases, the ratio of drift (or equivalently, for the Lorenz systems, integrated forcing error) to shadow radius for a typical shadow orbit should be approximately 2. This is exactly what was found in Figure 3.14 for the constant model. The number 2 actually appears to be an upper bound, for all but low values of F.

The property clearly holds when displacement error is effectively zero, since, for a reference time τ and drift vector $\mathbf{d}(\tau)$, the model trajectory can simply begin at a displacement of $-0.5\mathbf{d}(\tau)$ and end at a displacement of $-0.5\mathbf{d}(\tau)$. It will also tend to hold in a statistical sense, though, whenever

$$\langle \sum_{i=1}^{n} \frac{(\mathbf{d}(\tau) \cdot \mathbf{u_i}(\tau))^2}{(1+\sigma_i(\tau))^2} \rangle \approx \frac{d(\tau)^2}{4} \tag{4.41}$$

which is a much weaker condition.

Suppose that the model exists in a high dimension state space, and the components of the drift vector are uncorrelated either with each other or with the direction of the singular vectors. In this case,

$$\langle (\mathbf{d}(\tau) \cdot \mathbf{u_i}(\tau))^2 \rangle = \langle d_i^2(\tau) \rangle = \frac{1}{n} \langle d^2(\tau) \rangle. \tag{4.42}$$

Suppose now that we fix the magnitude of the drift vector, and the multipliers σ_i, and take expected values of both sides of equation 4.22 over all possible orientations of the singular vectors. Then, from equations 4.22 and 4.42,

$$\langle r_s^2 \rangle = \langle \sum_{i=1}^{n} \frac{(\mathbf{d}(\tau) \cdot \mathbf{u_i}(\tau))^2}{(1+\sigma_i(\tau))^2} \rangle \tag{4.43}$$

$$= \frac{d(\tau)^2}{n} \langle \sum_{i=1}^{n} \frac{1}{(1+\sigma_i(\tau))^2} \rangle \tag{4.44}$$

where the shadow radius r_s is now a function of the orientation of the singular vectors.

If the model is locally dissipative in the sense that it contracts volumes in state space [47] over the finite time τ, then we claim that the sum on the right hand side has a minimum value of $\frac{n}{4}$ when all $\sigma_i = 1$. To see this, consider first the case where the model exactly preserves volume, which will occur if

$$\prod_{i=1}^{n} \sigma_i = 1 \tag{4.45}$$

where we have dropped the dependence on τ. Writing the minimisation problem as a Lagrangian, we seek minima of

$$\sum_{i=1}^{n} \frac{1}{(1+\sigma_i)^2} + \lambda(\prod_{i=1}^{n} \sigma_i - 1) \tag{4.46}$$

where λ is a constant multiplier. Taking partial derivatives with respect to σ_j, and setting to zero, gives

$$0 = \frac{-2}{(1+\sigma_j)^3} + \lambda \prod_{i \neq j} \sigma_i = \frac{-2}{(1+\sigma_j)^3} + \frac{\lambda}{\sigma_j} \tag{4.47}$$

where we have used the fact that $\prod_{i=1}^{n} \sigma_i = 1$. Therefore

$$\sigma_j = \frac{2}{\lambda}(\sigma_j + 1)^3. \tag{4.48}$$

This equation represents the intersection between a straight line and a cubic in σ_j, and has two solutions for $\lambda > 16$, and a single solution when $\lambda = 16$ and all $\sigma_j = 1$. Since λ is the same for all j, the multipliers σ_j can only take on one of a maximum two values.

We claim that the solution $\sigma_j = 1$ for all j represents a global minimum. The full proof is complicated, and is given in the Appendix. For the 2-D case, a geometrical argument is also possible. We wish to show that

$$\frac{1}{(1+\sigma_1)^2} + \frac{1}{(1+\sigma_2)^2} \geq \frac{1}{2} \tag{4.49}$$

which is the same as saying that the point $x_1 = 1$, $x_2 = 1$ is inside the ellipse

$$\frac{x_1}{(1+\sigma_1)^2} + \frac{x_1}{(1+\sigma_2)^2} \geq \frac{1}{2}. \tag{4.50}$$

When $\sigma_1 = 1$ and $\sigma_2 = 1$, the point $(1, 1)$ is at the boundary of the ellipse. Any other volume retaining ellipse, which is contracted along one axis and expanded along the other, will not contain this point. It therefore follows that

$$\frac{1}{(1+\sigma_1)^2} + \frac{1}{(1+\sigma_2)^2} > \frac{1}{2} \tag{4.51}$$

if σ_1 (and by implication σ_2) doesn't equal 1.

Since the critical point with $\sigma_i = 1$ for all i represents a global minimum, it follows, from equation 4.44, that

$$\begin{aligned} \langle r_s^2 \rangle &= \frac{d(\tau)^2}{n} \langle \sum_{i=1}^{n} \frac{1}{(1+\sigma_i(\tau))^2} \rangle & (4.52) \\ &\geq \frac{d(\tau)^2}{n} \frac{n}{4} & (4.53) \\ &= \frac{d(\tau)^2}{4}, & (4.54) \end{aligned}$$

or

$$\sqrt{\langle r_s^2 \rangle} \geq \frac{d(\tau)}{2}. \tag{4.55}$$

Therefore the allowable shadow radius, in an RMS sense, for a given magnitude of drift, is greater than or equal to half the drift. If the model is locally strictly dissipative, rather than volume preserving, the inequality is replaced by a strict inequality. This is a powerful result, since it applies to an extremely broad class of chaotic models, including, typically, those which have an attractor [47].

Note that we are treating the shadow radius r_s as a function of the shadow time τ, while in the shadow calculations of the Lorenz '96 and other systems we solved for the shadow time as a function of the shadow radius. Since $\tau(r_s)$ is a monotonically increasing function, it is possible to invert the problem in this way. We will see in Chapter 6 that for weather model it is usually more convenient to solve for the shadow radius as a function of shadow time.

Define the *dissipation coefficient* $q(\tau)$ as

$$q(\tau) = \sqrt{\frac{n}{4\langle \sum_{i=1}^{n} \frac{1}{(1+\sigma_i(\tau))^2} \rangle}}. \tag{4.56}$$

Then equation 4.44 can be written

$$d(\tau) = 2q(\tau)\sqrt{\langle r_s^2(\tau) \rangle}. \tag{4.57}$$

The dissipation coefficient $q(\tau)$ is a measure of local model dissipation in state space. A model where all the singular value multipliers equal 1 has a dissipation coefficient of exactly 1. If model error is high, or the shadow radius is small, then shadow times will be short and the dissipation coefficient will be near 1, so RMS drift will approximately equal the shadow diameter. The dissipation coefficient can either be calculated directly, or estimated from some idea of the likely distribution.

As an illustration of a volume preserving model, suppose that the magnitudes of the n singular vector multipliers $\sigma_i(\tau)$, when arranged in descending order, follow a power law distribution, so that

$$\sigma_i(\tau) = \sigma_1^{1-\frac{2i}{n}}. \tag{4.58}$$

The largest singular vector multiplier is therefore σ_1, and the smallest is $\sigma_n = \sigma_1^{-1}$. An equal number of directions contract as expand in phase space, and because the product of the multipliers is 1, such a model would preserve state space volume.

Given the ideal power law distribution, and assuming the dimension n is large, we can approximate the sum by an integral, so

$$\sum_{i=1}^{n} \frac{1}{(1+\sigma_i(\tau))^2} = \int_{-n/2}^{n/2} \frac{1}{(1+\sigma_1^{1-\frac{2s}{n}})^2} ds \tag{4.59}$$

$$= \frac{n}{2} - \frac{n}{2log(\sigma_1)} \frac{\sigma_1 - 1}{\sigma_1 + 1}. \tag{4.60}$$

The dissipation coefficient is therefore

$$q(\tau) = \frac{1}{2\sqrt{1 - \frac{1}{log(\sigma_1)} \frac{\sigma_1 - 1}{\sigma_1 + 1}}}. \tag{4.61}$$

Values of q are plotted as a function of σ_1 in Figure 4.7. The maximum value of 1.0 occurs for $\sigma_1 = 1.0$, as expected.

Real models often show a similiar, roughly power law distribution. Figure 4.8 plots the distribution of singular value multipliers, and dissipation coefficient q (where the expectation operator in equation 4.57 is supressed), for the constant and linear models. Results are averaged over 200 shadow runs at shadow radius 0.4. The + marks the centre point: for either model, more singular value multipliers contract than expand. The models are therefore more dissipative than the power law distribution described above. The constant model has a lower q than the linear model, and so is more dissipative. This is largely because q is evaluated over maximal shadow orbits, and the constant model shadows for shorter times than the linear model.

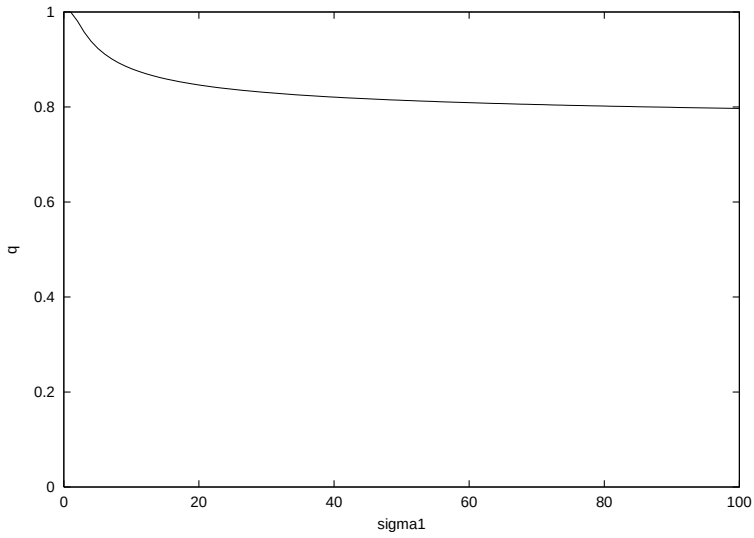


Figure 4.7: Plot of the dissipation coefficient $q(\tau)$ when the singular value multipliers follow a simple power law distribution as described in text. Shown as a function of the leading multiplier. The maximum value is 1.0. As an upper bound on shadow times, the shadow law therefore sets drift equal to twice shadow radius, which equates to a dissipation coefficient of 1.0.

We summarise these results with the following formal statement of the *shadow law*.

The Shadow Law. Suppose that the model of a true system is locally dissipative in the sense that the singular value multipliers σ_i of the linear propagator, evaluated along the true orbit over a time τ, are volume contracting, i.e. $\prod_{i=1}^{n} \sigma_i \leq 1$. Assume also that the drift vectors are uncorrelated with the singular vectors. Then, as a function of drift $d(\tau)$, an approximate lower bound on shadow radius is given by

$$\sqrt{\langle r_s^2(\tau) \rangle} = \frac{1}{2} d(\tau). \tag{4.62}$$

When model error is high, or shadow times are short, then the shadow radius will approach this bound, so

$$r_s(\tau) \approx \frac{1}{2} d(\tau). \tag{4.63}$$

The shadow law therefore provides a lower bound on shadow radius, in terms of drift, whenever the model is dissipative over the time tested. The shadow rule will prove indispensable for the weather models encountered in Chapter 6, for which it is impossible to calculate all the singular vectors, and the full SET cannot be invoked.

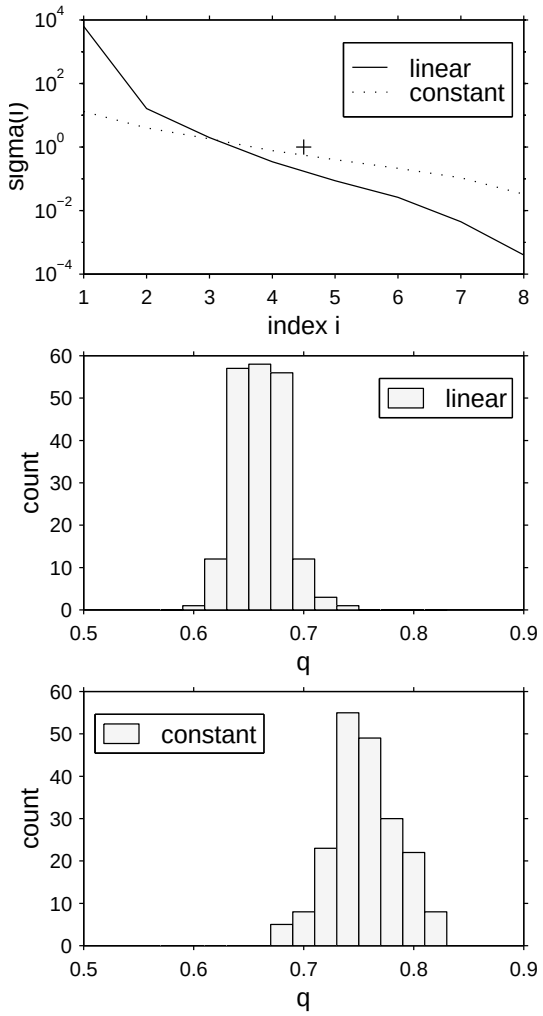


Figure 4.8: Upper panel shows singular value multipliers, arranged in descending order, for the constant and linear models. Results are averaged over 200 shadow runs at shadow radius 0.4. The + marks the centre point: for either model, more singular value multipliers contract than expand. The centre panel shows the distribution of the dissipation coefficient q over the same shadow runs. The lower panel shows the distribution for the constant model. The constant model is less dissipative than the linear model, largely because it shadows for a shorter time.

4.5 The model error index

While the measurement of model error may be complicated by sensitivity to initial conditions, the degree to which this holds true depends on the strength of model error relative to displacement error. For example, if a model is only moderately sensitive to initial conditions, but generates huge model error, then it should be relatively easy to measure model error. If, on the other hand, the model is excellent, but both it and the system are highly sensitive to initial condition, then measuring model error will be more difficult.

The relative strength of model and displacement error also determines which places limits on shadow times. When model error dominates, shadow times will be determined primarily by the drift, while if model error is small the model dissipation must be taken into account.

We propose two different measures for the comparison of model error with displacement error. The first compares the forces of drift with those of dissipation. Using the definition of the dissipation coefficient $q(\tau)$, the linearised dynamics in RMS form were written in equation 4.57 as

$$d(\tau) = 2q(\tau)\sqrt{\langle r_s^2(\tau)\rangle}. \tag{4.64}$$

We define the first model error index $M1(\tau)$ to be

$$M1(\tau) = \langle \frac{d(\tau)}{2r_s q(\tau)} \rangle. \tag{4.65}$$

$M1(\tau)$ provides a measure of the relative strength of model error, as measured by the drift $d(\tau)$, compared to the dissipation, as measured by $q(\tau)$. If it is the case that $M1(\tau) > 1$, then model error dominates dissipation, and the model won't be expected to shadow for the time τ at radius r_s.

For large models, it may be impossible to evaluate the dissipation coefficient $q(\tau)$. For dissipative models, and any of the models studied in this thesis, it holds that $q(\tau) < 1$. Therefore we have

$$M1(\tau) < \frac{d(\tau)}{2r_s}. \tag{4.66}$$

Another measure of model error relative to displacement error is to use, instead of the dissipation, the growth of the leading singular vector. Such a measure would be useful when judging the likely impact of model error on ensembles created by perturbing in the direction of the leading singular vectors. We therefore define

$$M2(\tau) = \frac{d(\tau)}{2r_s\sigma_{max}(\tau)} \tag{4.67}$$

where $\sigma_{max}(\tau)$ is the maximum singular value multiplier at the shadow time, $d(\tau)$ is the drift at the shadow time, and r_s is the shadow radius. If model error is high, then drift is about equal to the shadow diameter, so the expression simplifies to

$$M2(\tau) \approx \frac{1}{\sigma_{max}(\tau)}. \tag{4.68}$$

For the constant model typical values of $M2(\tau)$ are about 0.10, and for the linear model about 0.017. Its precise value will vary around the attractor, especially for systems like Saltzman where shadow times vary enormously.

Either one of these measures can be used to compare different model/system pairs. In Chapter 6, we use $M2$ to compare the likely effect of error on ensembles in weather models, with the corresponding effect on ensembles of the Lorenz '96 constant model.

4.6 Applications of the shadow test

4.6.1 The Lorenz '96 systems

The SET is a general method for estimating shadow times, using only the drift and singular vector multipliers, that can be applied to any model/system pair. In this section we test the method for a number of cases, by comparing the estimated times with actual shadow times.

The first comparison is with the Lorenz '96 models already studied in some detail. Figures 4.9 and 4.10 show the results when applied to the constant and linear models, where the truth is the two level system. Shadow times have been averaged over twenty runs. The shadow radius has again been scaled, and takes the values 0.2 and 0.4 at $F = 10$. Agreement is quite good. A more detailed view of the results for the linear model at $F = 10$ is given by Figure 4.11, which compares a histogram of shadow times over 200 starting points.

The shadow law states that the maximum drift tolerated over a shadow orbit is bounded by the shadow diameter. For the constant model the bound is close to actual shadow times, as was found in Figure 3.14 where the ratio of drift to shadow radius for the constant model was about 2 (it actually exceeds 2 by a small amount for lower values of forcings where the model is only weakly dissipative). For longer shadow times, however, the shadow law can over-estimate substantially. For example, with the linear model at $F = 10$ and shadow radius 0.4, an estimate of an upper bound using the shadow law would give 4.05. The SET gives an answer of 2.36, while the actual shadow time is 2.28. Using the shadow law as a guide to shadow

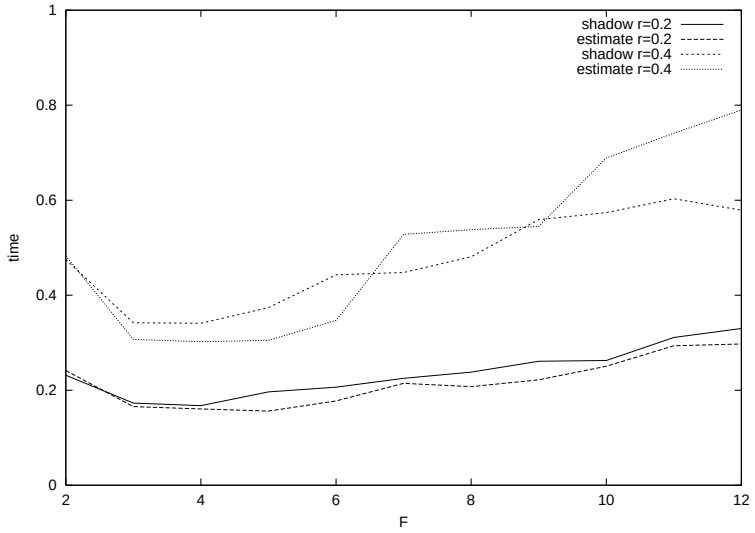


Figure 4.9: Plot of estimated and mean shadow times for constant model, averaged over 20 runs. Shadow radius is 0.2 and 0.4 at $F = 10$, scaled proportionately for other values.

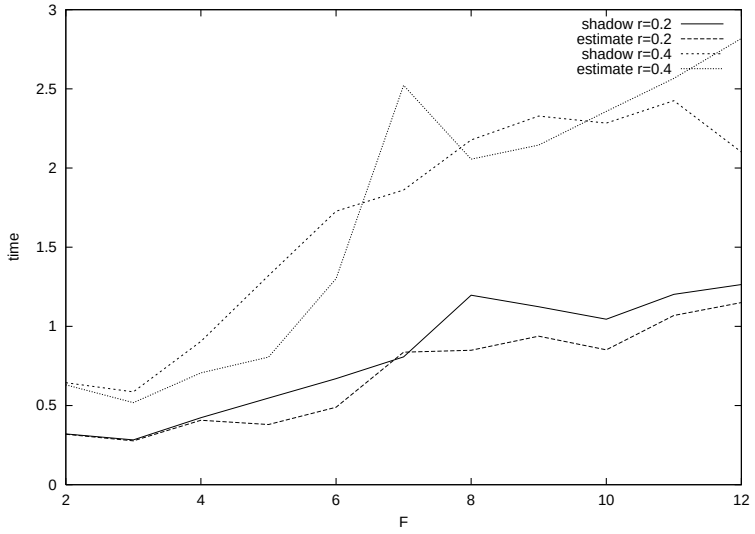


Figure 4.10: Plot of estimated and actual shadow times for linear model, averaged over 20 runs. Shadow radius is 0.2 and 0.4 at $F = 10$, scaled for other values.

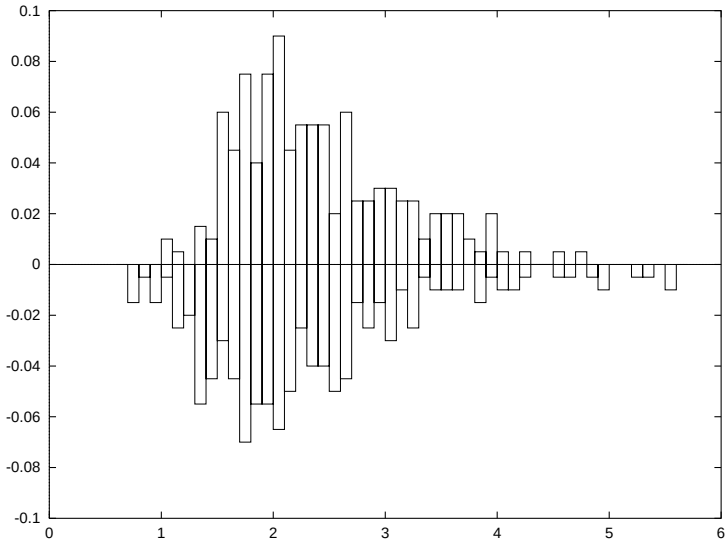
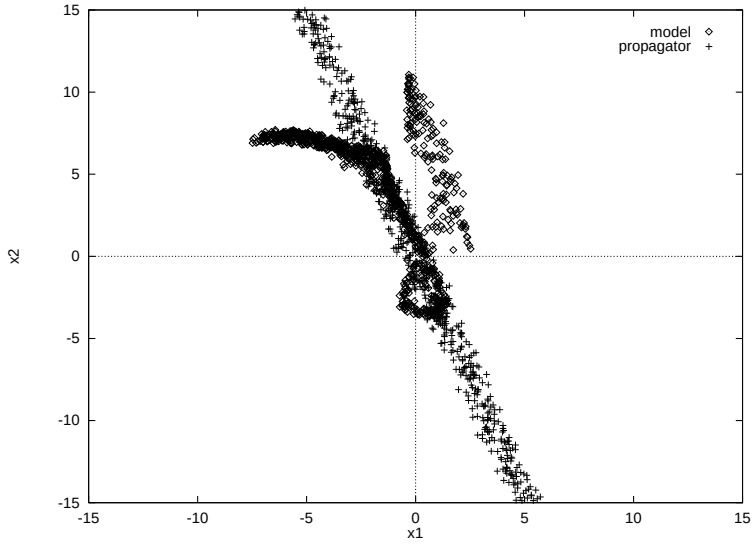


Figure 4.11: Histogram of 200 shadow times for linear model with $F = 10$, shadow radius 0.4. Upper is real shadow times, lower is estimated times.

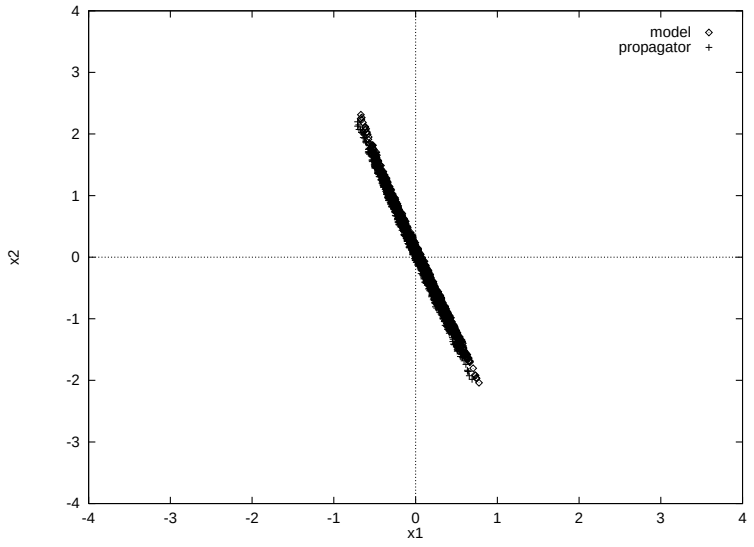
times, rather than an upper bound, works best in situations where model error is large relative to displacement error, simply because under those circumstances the dissipation coefficient $q(t)$ will be near 1.

It might seem strange that the SET can work so well at relatively long times, when one would expect that the linear propagator will no longer be accurate. This fear would appear to be validated by Figure 4.12(a), which shows the images of a 0.4 radius ball of initial displacements under both the full and the linearised propagator dynamics. The linearised dynamics produce an ellipse, as expected, while the full dynamics produces a somewhat contorted distribution. For most points the two completely disagree, but for points near the origin, which correspond to shadow trajectories, the match is in fact quite good. Figure 4.12(b) shows the images of a ball of points of radius 0.02 around the shadow point. Both the full dynamics and the linearised dynamics produce an ellipse close to the origin. The essence of the SET is that it only tries to model trajectories which actually shadow, and for these orbits, so long as the shadow radius is sufficiently small, the linearised dynamics are valid.

Of course, the validity of the linear approximation depends on time as well as shadow radius, and if a model shadows for extremely long times then the SET will no longer be reliable. An example of a very long shadow orbit is shown in Figure 4.13. It was obtained by changing the coupling of the two level system from 1.0 to 0.5, which reduces the effect of the fine-scale $\tilde{y}$ variables on the large-scale $\tilde{x}$ variables. The resulting system was then modelled with the linear model. When the SET was



(a) Displacements of a 0.4 radius ball around starting point.



(b) Displacements of a 0.02 radius ball around shadow point. Note the reduced scale.

Figure 4.12: Plot of displacements from truth, in first two coordinates, for images under the linear model (full dynamics) and the propagator (linearised dynamics). The accuracy of the linearised dynamics depends on proximity to the true orbit. For the ball of initial conditions at radius 0.4, top panel, the linearised dynamics do not represent the full dynamics, but they are much more successful in the lower panel, where the ball of initial conditions has been shrunk to 0.02.

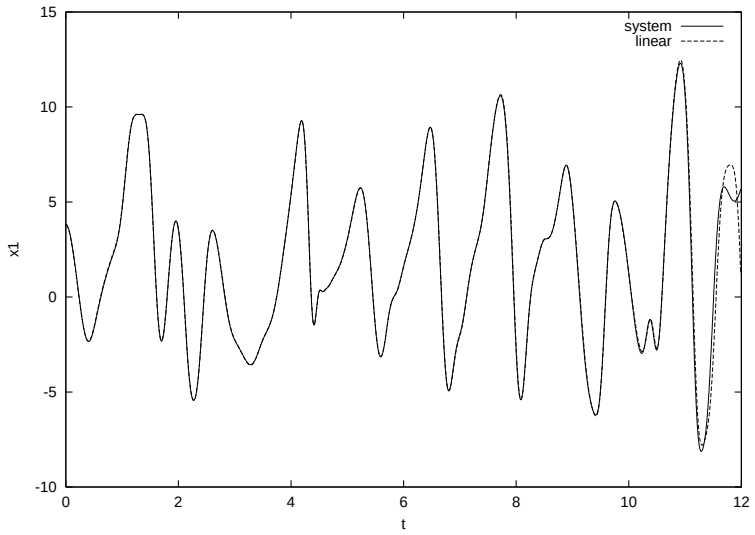


Figure 4.13: Shadowing orbit for linear model of two level system with coupling coefficient reduced to 0.5. The reduced effect of the fine-scale **y** variables on the large-scale **x** variables allows for long shadow times, in this case about 10.1 time units. Shadow radius is 0.4.

tested with this model, it failed to converge. The reason is that, over time periods such as this, higher order terms, which are not taken into account by the linearised dynamics, will eventually dominate. Figure 4.14 shows the drift accumulated over the shadow orbit. The shadow law gives an upper bound on total drift over the shadow orbit of about twice the shadow radius, or 0.8, which is well above the actual value of 0.12.

4.6.2 The Rössler system

The Lorenz '96 systems have a certain symmetry in that the equation for each variable **x** is the same, and it is possible that for some reason this symmetry might make shadow behaviour easier to predict. As a check against this, the method was used to predict shadowing times where the system is a modified version of the Rössler system,

$$\begin{aligned} \frac{dx}{dt} &= -y - z + \alpha xy/c \\ \frac{dy}{dt} &= x + 0.1y \\ \frac{dz}{dt} &= 0.1 + (x - c)z, \end{aligned} \tag{4.69}$$

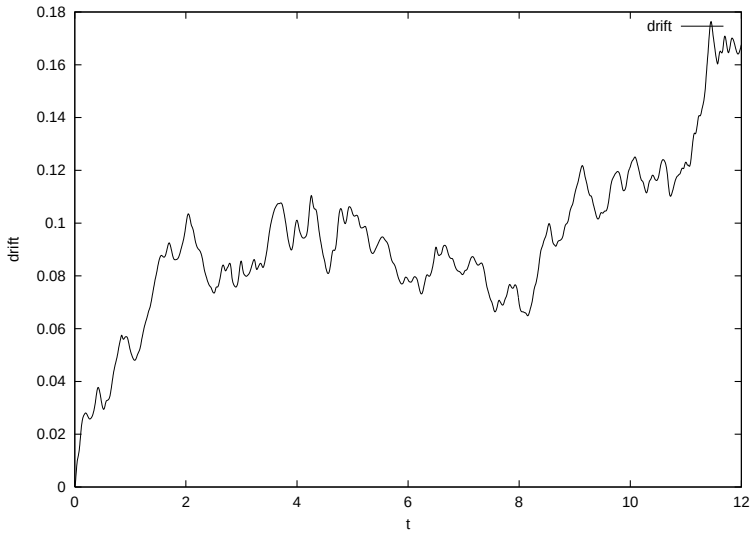


Figure 4.14: Drift calculated over the orbit shadowed in Figure 4.13. The SET fails to predict shadow times in this instance, because the drift remains small and higher order terms, not modelled by the linearised dynamics, eventually dominate.

and the model is the regular system of equation (1.2). The difference between the two is therefore the term $\alpha xy/c$ in the first equation, which was chosen as just one example of a nonlinear, asymmetric error term. Figure 4.15 shows the real and estimated shadow times for $\alpha = 0.3$. The SET is equally effective for this system.

4.6.3 The Saltzman system

With any such concocted example, though, we can be accused of choosing the system to prove the point (just about anything can be demonstrated using simple systems, since there are so many of them). A more convincing example might be one taken from the literature. It was mentioned in the introduction that we are familiar with the effect of Lorenz truncating the initial conditions of his convection model, but less familiar with the effects of truncating the equations of the model to three dimensions in the first place. Because of the historical importance of this 3-D model, we present a detailed investigation of its shadowing relative to the 7-D Saltzman system from which it was derived.

The equations for Saltzman's 7-D system [37] are:

$$\begin{aligned} \frac{dA}{dt} &= 23.521BC - 1.500D - 148.046A \\ \frac{dB}{dt} &= -22.030AC - 1.589E - 186.429B \end{aligned}$$

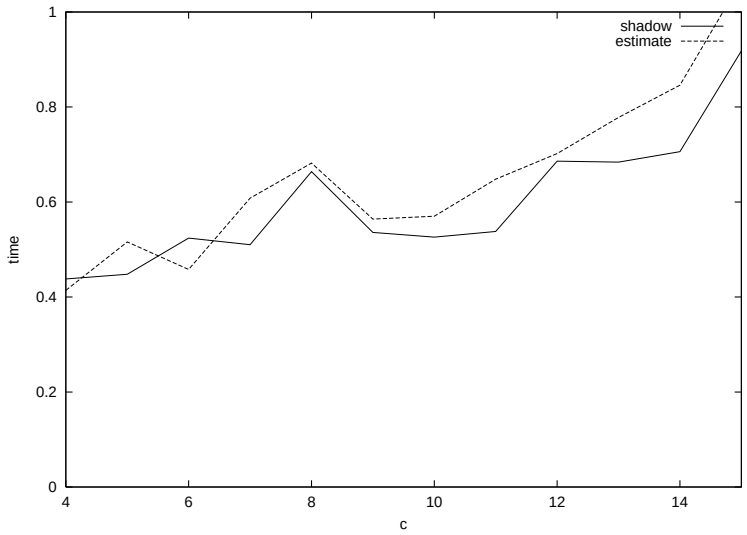


Figure 4.15: Real and estimated shadow times for the modified Rössler system. Shadow radius is 0.5 at $F = 10$, and scaled for other values.

$$\begin{aligned}
\frac{dC}{dt} &= 1.561AB - 0.185F - 400.276C \\
\frac{dD}{dt} &= -16.284CE - 16.284BF - 13.958AG - 1460.631\lambda A - 14.805D \\
\frac{dE}{dt} &= 16.284CD - 16.284AF - 18.610BG - 1947.508\lambda B - 18.643E \\
\frac{dF}{dt} &= 16.284AE + 16.284BD - 486.877\lambda C - 40.028F \\
\frac{dG}{dt} &= 27.916AD + 37.220BE - 39.479G
\end{aligned} \tag{4.70}$$

where λ is the Reynold's number. Lorenz noted that in certain circumstances the variables apart from A, D and G tended to zero. Therefore he built his system using only those three variables. He also rescaled so that each had a similar magnitude.

In order to check how his approximation compares to the full Saltzman system, we considered the reduced model

$$\begin{aligned}
\frac{dA}{dt} &= -1.500D - 148.046A \\
\frac{dD}{dt} &= -13.958AG - 1460.631\lambda A - 14.805D \\
\frac{dG}{dt} &= 27.916AD - 39.479G.
\end{aligned} \tag{4.71}$$

We then set $X = A, Y = D/100$ and $Z = G/100$ in both model and system (the rescaling is similar to that which Lorenz used).

If λ is such that the other parameters do indeed tend to zero, then the 3-D model and the 7-D system will obviously agree, and shadowing times will be infinite. We therefore seek regimes where this is not the case. It was found that a critical point exists between $\lambda = 25$ and $\lambda = 25.1$, so for $\lambda > 25.1$ the other parameters do not go to zero. We then compared the model and sytem for $\lambda = 28$, the typical value used in the Lorenz system, and $\lambda = 25.1$.

Figure 4.16 shows orbits on the attractor for the system and model at each value of λ. The attractors are inverted from those of the Lorenz '63 system because of a sign change. For $\lambda = 28$, the familiar butterfly wings of the model have, in the case of the full system, grown a body as well. The attractors at $\lambda = 25.1$ appear in closer agreement. Of course, these are observations of the climatologies, rather than model error on a true orbit, but the shape of the attractors might lead one to expect better shadowing performance for the lower setting of λ than for the higher.

This suspicion is borne out by Figure 4.17, which shows a detailed analysis of the shadowing performance. The upper panels are scatter diagrams of estimated and actual shadow times from 40 initial conditions, for shadow radius 0.1. The middle panels are histograms of shadow times. Note the different scale of the $\lambda = 28$ (left side) and $\lambda = 25.1$ (right side) results. For $\lambda = 28$, the longest shadow orbit is about 2 time units (the units differ from those of the Lorenz system). For $\lambda = 25.1$, many points shadow for longer than that. In either case the majority of the estimated shadow times are in good agreement with actual times, though for longer times the SET tends to overestimate the time.

The lower panels show the location on the attractor of the longest shadow orbit. The orbit at $\lambda = 28$ is shorter than the $\lambda = 25.1$ orbit on the right, even though it manages the transition from one lobe of the attractor to the other.

The Saltzman system is interesting because shadow times vary enormously depending on the position on the true attractor. Despite this variation, the SET does a good job of predicting shadow times for the majority of points (the top-left scatter diagram shows two points which fail for times under 0.5, but this only represents 5 percent of the total number tested). The 3-D model appears to be a better approximation to the full system at the lower setting of $\lambda = 25.1$, which is just above the threshold where other variables go to zero.

Figure 4.18 is another way to view shadow orbits. The model displacements from truth have been projected into a 2-D coordinate system (x_p, y_p) following the true orbit, and plotted with time t as the third axis. The graph, viewed from left to right, represents the perturbations that one would experience up and down (y_p) and from

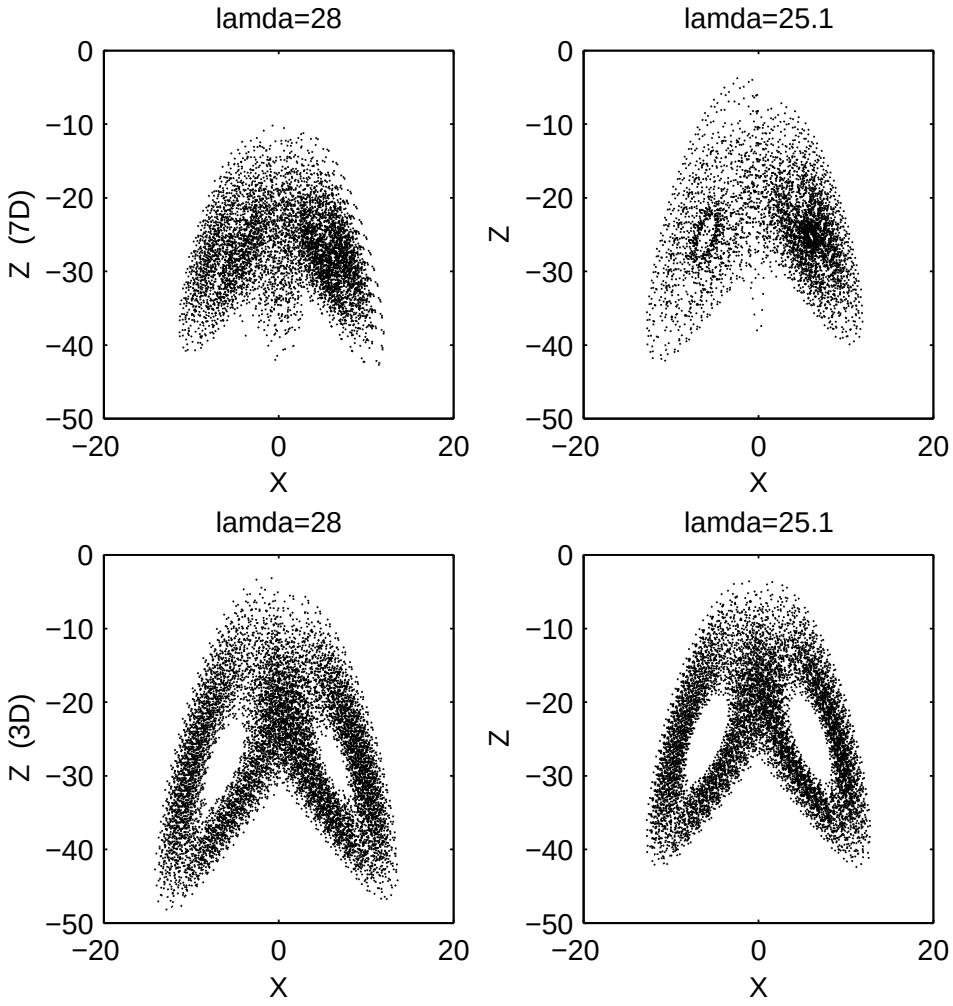


Figure 4.16: Comparison of attractors for the 7-D Saltzman system (upper panels) and the reduced 3-D model (lower panels), for $\lambda = 28$ and 25.1. The model appears to be in better agreement with the system for the lower value of λ.

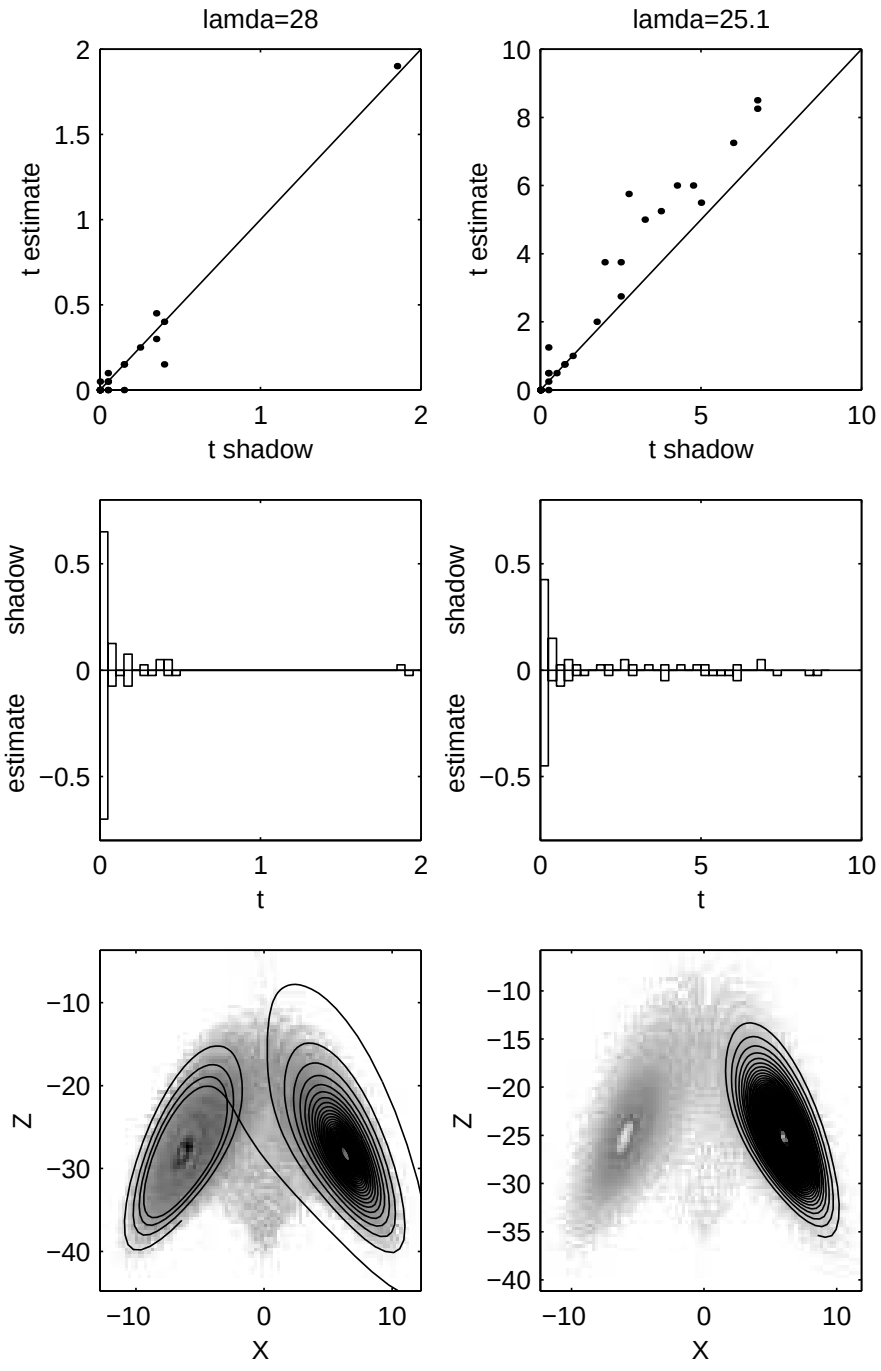


Figure 4.17: Analysis of shadowing performance for the 7-D Saltzman system and the reduced 3-D model, for $\lambda = 28$ (left side) and $\lambda = 25.1$ (right side). Shadow radius is 0.1. Upper panels are scatter diagrams of estimated and actual shadow times from 40 initial conditions. Middle panels are histograms of shadow times (note different time scales on left and right). The majority of the estimated shadow times are in good agreement with actual times, though for longer times the SET tends to overestimate the time. Lower panels show the location on the attractor of the longest shadow orbit.

side to side (x_p) while attempting, figuratively speaking, to follow the true path in a model car. For either value of λ, the displacements follow a regular pattern that slowly grows in magnitude. It appears that the reason the $\lambda = 25.1$ model shadows longer is because the system spends longer in each lobe of the attractor, thus presenting an easier path to follow.

4.6.4 Shadowing and step size

One cause of model error is an insufficiently small step size used during integration of the model's differential equations. Selection of an appropriate step size and integration scheme is of course a field in itself; our aim is merely to rephrase the problem in terms of model error, and illustrate the techniques developed so far. Figure 4.19 is a schematic diagram showing how a large step integration will create a drift error, as compared with an integration performed using two smaller steps. The drift after one step of Δ is just the difference between the trajectories at time 2Δ, which can be written

$$\mathbf{d}(\Delta) = (\frac{d\mathbf{x}(\Delta)}{dt} - \frac{d\mathbf{x}(0)}{dt})\Delta. \tag{4.72}$$

The term in brackets equals the change in velocity, so the drift can be written in terms of the acceleration:

$$\mathbf{d}(\Delta) \approx \frac{d^2\mathbf{x}(0)}{dt^2}\Delta^2. \tag{4.73}$$

This quantity can easily be calculated for each point.

As an example, suppose that we wish to determine whether a step size of 0.02 is sufficiently small for integration of the Lorenz '63 system. As a test, we could check how well the model at that time step shadows the system with a reduced step size of 0.01.

Figure 4.21 is a histogram of shadow times for shadow radius 0.01, where the true system has a step size of 0.01, and the model has a step size of 0.02. A small number of points, about 2.4 percent, fail to shadow for longer than a single time step. These points are marked by circles in Figure 4.20.

From the shadow law that maximum drift is twice the shadow radius (shadow times approach the upper bound when model error is high), we would expect the points which fail to shadow with the points where drift exceeds a value of 0.2. These points have also been marked in Figure 4.20. The correspondence is almost exact. They are simply the points which experience high acceleration. (They also coincide with areas of state space contraction [65].)

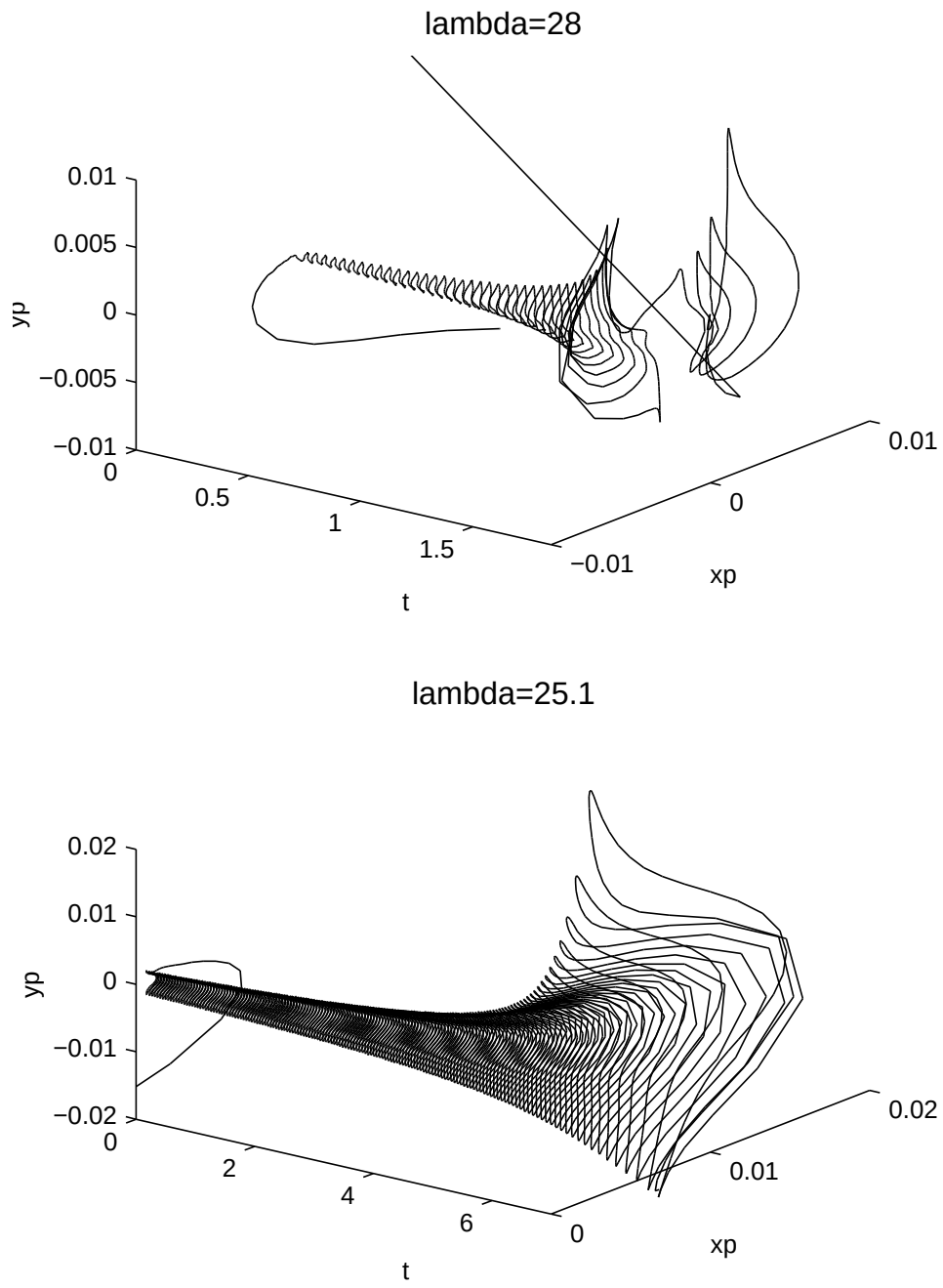


Figure 4.18: Shadow orbits for the 3-D model of the 7-D Saltzman system, at $\lambda = 28$ and 25.1. Model displacements from truth are projected into a 2-D coordinate system following the true orbit. The results are plotted with time as the third axis. For $\lambda = 28$, the shadow trajectory begins at the left hand side ($t = 0$), oscillates with increasing magnitude around the true orbit, and suddenly fails a short time after switching from one lobe of the attractor to the other. The $\lambda = 25.1$ model shadows a longer time (note the different time scale), apparently because the true system spends more time in one lobe of the attractor.

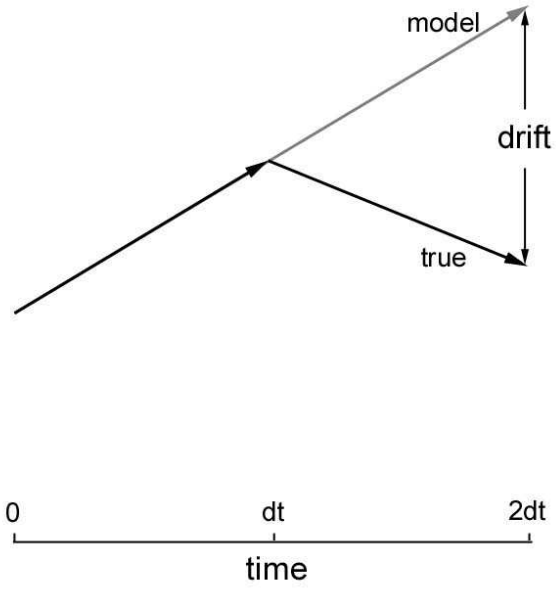


Figure 4.19: Diagram showing how too large a step size creates drift. The model, which has twice the step size of the true system, overshoots it. The resulting drift is proportional to the acceleration of the true system.

This contrasts with Figure 4.22 from Gilmour [23], which shows points where shadow orbits fail for a larger observational tolerance. She demonstrated a connection between the points where shadowing fails, and points with fastest error doubling times [62], which both tend to occur in the transition region between zones of the attractor. The figures serve again to highlight the distinction between initial condition error and model error.

4.6.5 The Rulkov circuit

The systems considered so far have only existed inside a computer. However, the techniques of studying model error apply equally well to observations of real systems, for which the true equations are unknown (or don't exist). In the next section, we formalise the treatment of observed systems. As a prelude, we here consider a model of an actual electrical circuit, which will further clarify the distinction between model and initial condition error.

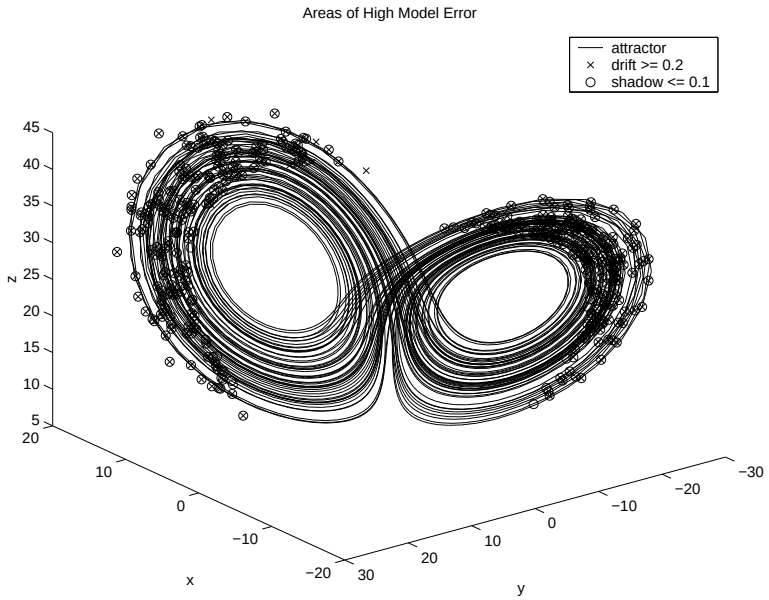


Figure 4.20: Plot of points with local drift ≥ 0.2, and points with shadow times ≤ 0.1, for Lorenz '63. True system has time step 0.01, model has time step 0.02. Errors occur in regions of high acceleration.

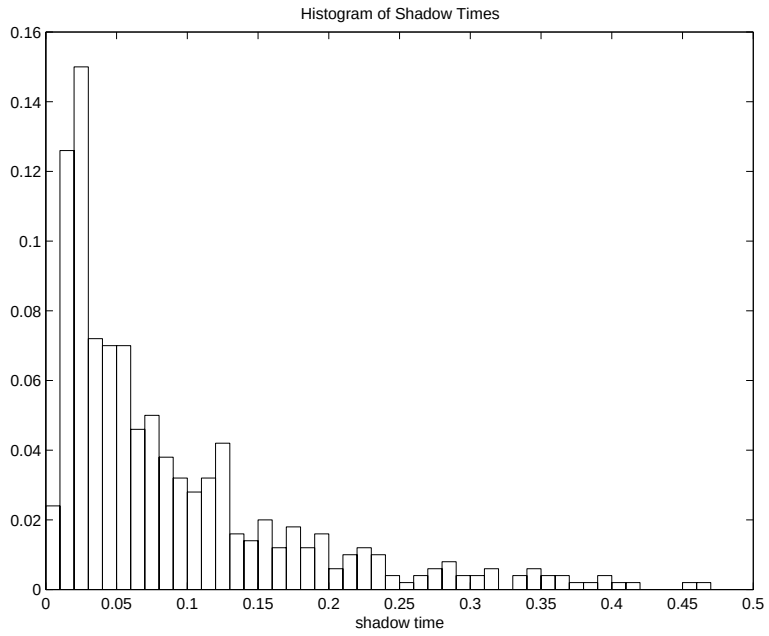


Figure 4.21: Histogram of shadow times for Lorenz '63. True system has time step 0.01, model has time step 0.02.

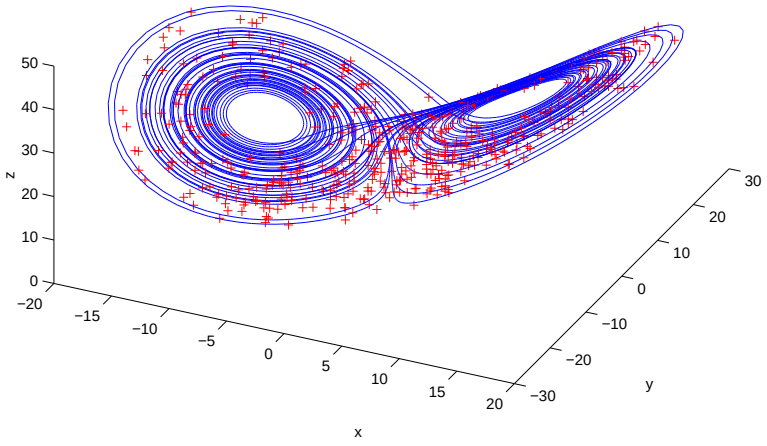


Figure 4.22: Distribution of the final points of the model to shadow observations from the Lorenz equations (plus signs). A trajectory of the Lorenz equations (solid line) is also shown for reference. True system has time step 0.01, model has time step 0.02. From [23].

The Rulkov circuit [59] was presented in Chapter 2 as an example of a nonlinear dynamical system. The equations model the behaviour of a real electrical circuit, but there will never be an exact correspondence between the two: even simple circuits don't follow neat mathematical laws, when examined in detail. We therefore take the physical circuit, projected into model space, to be truth, and the mathematical approximation as the model.

Figure 4.23 shows predicted and observed points for a number of trajectories starting from neighbouring points. Each of the initial conditions used was a point on an orbit of the true system, projected into model space. It was possible to find a number of initial conditions in close proximity because the circuit is recurrent: i.e. if the system is run for sufficiently long times it experiences a near return, within a specified tolerance, to the initial condition. As a result, the ensemble of model trajectories can be verified against an ensemble of true trajectories, which makes model error much easier to detect [63].

Referring to the figure, around time 7830 both the observed and predicted trajectories suddenly diverge, so at this point both the model and the system itself have high sensitivity to initial condition. Therefore to say that a model has high sensitivity to initial condition is not to say that it is wrong, or that this is an undesirable feature,

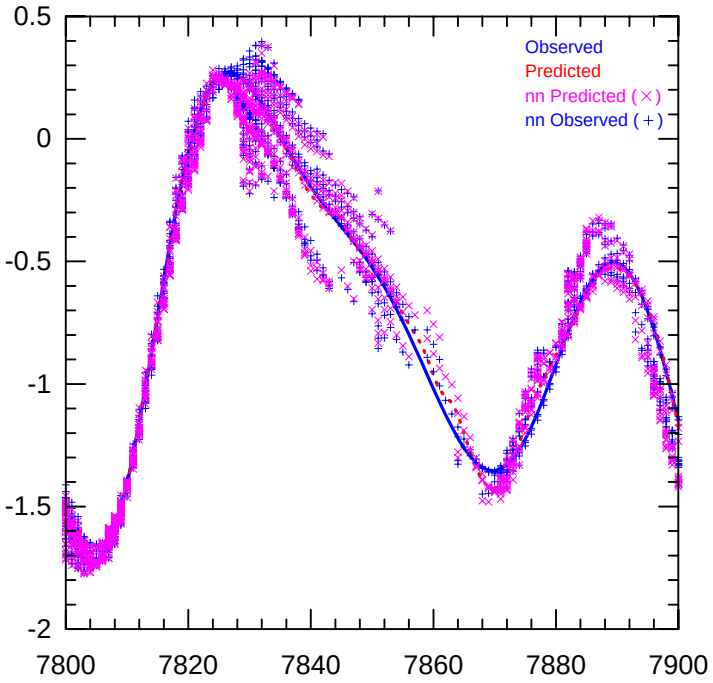


Figure 4.23: Circuit errors. From [63].

since the system itself may have the same behaviour. Rather, it means that the initial condition must be chosen very exactly to get an accurate prediction. In this case, the model seems to track the observations quite well for the majority of points.

In the zoomed view, Figure 4.24, however, it is seen that, near the local maximum at time 7825, the forecasts are systematically lower than the observations. This is a result of local model error.

Recurrence makes model error at a particular point easier to detect, because an ensemble of forecasts can be compared with an ensemble of true trajectories. Since the atmosphere is unlikely to repeat itself even once before it eventually boils away into space, recurrence isn't a feature that we can exploit to generate initial conditions for weather models. However, this needn't be a limitation; model error and shadow times are determined principally by the linearised dynamics, which don't distinguish whether the initial conditions lie on an attractor or not.

4.7 Observed systems

The development of the linearised dynamics assumed that we know the underlying equations of the true system. For most systems of practical interest, such as the

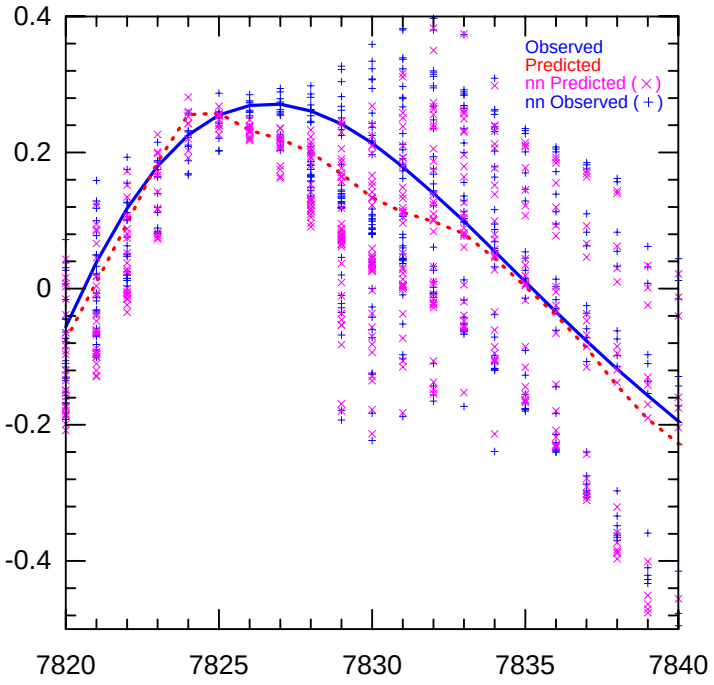


Figure 4.24: Zoom of circuit errors. From [63].

weather, this won't be the case. The best that we can do is interpolate through a series of observations, each of which will be corrupted to some extent by observation error (in weather forecasting the arrived at interpolation is known as the *analysis* [17]). Fortunately, because the linearised dynamics refer only to the specific trajectory of the true system which we are trying to shadow, it isn't necessary to know the equations that underlie it. Suppose that the true system is expressed, as with the weather, from an analysis, so that

$$\tilde{\mathbf{x}} = \mathbf{x_a} + \mathbf{h_a} \tag{4.74}$$

where $\mathbf{x_a}$ is the analysed solution and $\mathbf{h_a}$ is an error term due to imperfect observations and analysis. We further assume that $\mathbf{x_a}$ is continuous and piecewise differentiable. If the analysis is only known at discrete points, then we can use some smooth interpolation for intermediate points, or alternatively the equations below could be written in discrete form (we prefer the continuous form for clarity and consistency with previous results).

Linearising the equations as before, we can write

$$\begin{aligned} \frac{d\mathbf{e}}{dt} &= \frac{d\mathbf{x}}{dt} - \frac{d\tilde{\mathbf{x}}}{dt} \\ &= J(\tilde{\mathbf{x}}(t)) \cdot \mathbf{e}(t)) + \mathbf{G}(\tilde{\mathbf{x}}(t)) - \frac{d\tilde{\mathbf{x}}}{dt} + \mathbf{R_e}(t) \end{aligned}$$

$$= J(\mathbf{x_a}(t) + \mathbf{h_a}(t)) \cdot \mathbf{e}(t)) + \mathbf{G}(\mathbf{x_a}(t) + \mathbf{h_a}(t)) - (\frac{d\mathbf{x_a}}{dt} + \frac{d\mathbf{h_a}}{dt}) + \mathbf{R_e}(t)$$

$$= J(\mathbf{x_a}(t)) \cdot \mathbf{e}(t)) + \mathbf{G}(\mathbf{x_a}(t)) + J(\mathbf{x_a}(t)) \cdot \mathbf{h_a}(t) + \mathbf{R_e}(t) + \mathbf{R_h}(t) - (\frac{d\mathbf{x_a}}{dt} + \frac{d\mathbf{h_a}}{dt})$$

where the remainder term $R_e(t)$ is $O(\|\mathbf{e}(t)\|^2)$, and $R_h(t)$ is $O(\|\mathbf{h_a}(t)\|^2)$.

If we neglect the second order remainder terms, and integrate, we then obtain

$$\mathbf{e}(\tau) \approx \mathbf{M}_a(\tau) \cdot \mathbf{e}(0) + \int_0^\tau \mathbf{G}(\mathbf{x_a}(t))dt - \mathbf{x_a}(\tau) + \mathbf{x_a}(0) + \mathbf{E}(\tau) \tag{4.75}$$

where the error term

$$\mathbf{E}(\tau) = \int_0^\tau \mathbf{J}(\mathbf{x_a}(t)) \cdot \mathbf{h_a}(t)dt - \mathbf{h_a}(\tau) + \mathbf{h_a}(0) \tag{4.76}$$

depends only on the analysis error and the model. If the likely analysis error distribution is known, then the error term $\mathbf{E}$ can be estimated without needing further information about the true system. Note that in the above formulation it is not necessary to calculate the velocity of the true system, since this term has been integrated out. Only the start and finish analysis points $\mathbf{x_a}(0)$ and $\mathbf{x_a}(\tau)$ need be known.

The linearised equations can be used to obtain model drift and estimate shadowing times as for the case without error. The only difference is that there will now be an additional error term. Consider for example the calculation of the analysed model drift
$d_a(\tau) = \|\mathbf{d_a}(\tau)\|$, where

$$\mathbf{d_a}(\tau) = \int_0^\tau \mathbf{G}(\mathbf{x_a}(t))dt - \mathbf{x_a}(\tau) + \mathbf{x_a}(0) + \mathbf{E}(\tau). \tag{4.77}$$

If we assume that observation errors are uncorrelated, then averaging the results over many integrations should give a good measure of average model drift, and hence model quality.

As an example, suppose that the vectors $\mathbf{h_a}(t)$, sampled every Δ time units, follow a white noise distribution with variance $K = \langle \mathbf{h_a^2}(t) \rangle$, and suppose the vectors $\mathbf{J}(\mathbf{x_a}(t)) \cdot \mathbf{h_a}(t)$ follow a similiar distribution but with variance ρK. Then the integral in the expression for $\mathbf{E}(\tau)$ is just a random walk, and we can write

$$\langle (\int_0^\tau \mathbf{J}(\mathbf{x_a}(t)) \cdot \mathbf{h_a}(t)dt)^2 \rangle = \rho K \tau. \tag{4.78}$$

Therefore

$$\langle \mathbf{E}^2(\tau) \rangle = \rho K \tau + 2K \tag{4.79}$$

and the expected value of the error in the drift calculation due to observation error can be explicitly calculated.

The conclusion is that model drift is a robust measure of model error, which doesn't necessitate a direct calculation of the true system velocity. Of course, if observation error is greater than model error it will be difficult to separate the two. When we look at weather models in Chapter 6, we will concern ourselves with shadowing the analysis, which is the best approximation to the weather, rather than the observations themselves (if the model can't shadow the analysis within the analysis uncertainty, then it can't shadow the real weather either).

4.8 Error due to the projection

Until now, we have ignored the role of the projection operator $\mathbf{P}$ which maps the true system state space into that of the model. However, the projection can introduce significant errors. In the case of the weather, for example, the projection operator maps the real weather onto the model grid, using an assimilation process which is dependent on observations but also to a large extent on the interpolation scheme, which itself is a functon of the model [17]. In data-poor regions, the assimilation process will be particularly prone to error.

To see how such errors affect the drift calculation, suppose that there exists a particular projection $\mathbf{P_T}$ and a corresponding model

$$\frac{d\mathbf{x}}{dt} = \mathbf{G_T}(\mathbf{x}(t)) \tag{4.80}$$

such that

$$\mathbf{x}(t) = \mathbf{P_T}(\tilde{\mathbf{x}}(t)) \tag{4.81}$$

provided $\mathbf{x}(0) = \mathbf{P_T}(\tilde{\mathbf{x}}_\mathbf{0})$, where $\tilde{\mathbf{x}}_\mathbf{0} = \tilde{\mathbf{x}}(0)$. In other words, the model is perfect given the particular projection $\mathbf{P_T}$. Taking derivatives with respect to time at initial time, we have

$$\mathbf{G_T}(\mathbf{P_T}(\tilde{\mathbf{x}}_\mathbf{0})) = \frac{d}{dt}\mathbf{P_T}(\tilde{\mathbf{x}}_\mathbf{0}) = \frac{\partial \mathbf{P_T}(\tilde{\mathbf{x}}_\mathbf{0})}{\partial \tilde{\mathbf{x}}}\tilde{\mathbf{G}}(\tilde{\mathbf{x}}_\mathbf{0}). \tag{4.82}$$

Now, suppose that the actual projection is given by the function $\mathbf{P} = \mathbf{P_T} + \mathbf{P_E}$, where $\mathbf{P_E}$ is an error term, and the actual model is given by $\mathbf{G} = \mathbf{G_T} + \mathbf{G_E}$. Then the error will be

$$\mathbf{e}(t) = \mathbf{x}(t) - \mathbf{P}(\tilde{\mathbf{x}}(t)) \tag{4.83}$$

with initial velocity

$$\frac{d\mathbf{e}(0)}{dt} = \mathbf{G}(\mathbf{P}(\tilde{\mathbf{x}}_\mathbf{0})) - \frac{\partial \mathbf{P}}{\partial \tilde{\mathbf{x}}}\tilde{\mathbf{G}}(\tilde{\mathbf{x}}_\mathbf{0}) \tag{4.84}$$

$$= \mathbf{G}(\mathbf{P}(\tilde{\mathbf{x}}_\mathbf{0})) - \frac{\partial \mathbf{P_T}}{\partial \tilde{\mathbf{x}}}\tilde{\mathbf{G}}(\tilde{\mathbf{x}}_\mathbf{0}) - \frac{\partial \mathbf{P_E}}{\partial \tilde{\mathbf{x}}}\tilde{\mathbf{G}}(\tilde{\mathbf{x}}_\mathbf{0}) \tag{4.85}$$

$$= \mathbf{G}(\mathbf{P}(\tilde{\mathbf{x}}_\mathbf{0})) - \mathbf{G_T}(\mathbf{P}(\tilde{\mathbf{x}}_\mathbf{0})) - \frac{\partial \mathbf{P_E}}{\partial \tilde{\mathbf{x}}}\tilde{\mathbf{G}}(\tilde{\mathbf{x}}_\mathbf{0}) \tag{4.86}$$

$$= \mathbf{G}(\mathbf{P_T}(\tilde{\mathbf{x}}_\mathbf{0})) - \mathbf{G_T}(\mathbf{P_T}(\tilde{\mathbf{x}}_\mathbf{0})) + \mathbf{G}(\mathbf{P}(\tilde{\mathbf{x}}_\mathbf{0})) - \mathbf{G}(\mathbf{P_T}(\tilde{\mathbf{x}}_\mathbf{0})) - \frac{\partial \mathbf{P_E}}{\partial \tilde{\mathbf{x}}}\tilde{\mathbf{G}}(\tilde{\mathbf{x}}_\mathbf{0}) \tag{4.87}$$

where the first two terms reflect error in the model relative to the perfect model, and the last three terms reflect error in the projection operator. In the perfect model case where $\mathbf{G} = \mathbf{G_T}$, we have

$$\frac{d\mathbf{e}(0)}{dt} = -\frac{\partial \mathbf{P_E}}{\partial \tilde{\mathbf{x}}}\tilde{\mathbf{G}}(\tilde{\mathbf{x}}_\mathbf{0}), \tag{4.88}$$

so even with a perfect model, there will be a velocity error term caused by the projection operator.

As a simple example, consider the 2-D case where

$$\mathbf{P}(x_1, x_2) = (x_1, x_2 + f(x_1, x_2)) \tag{4.89}$$

$$\mathbf{P_T}(x_1, x_2) = (x_1, x_2) \tag{4.90}$$

$$\mathbf{P_E}(x_1, x_2) = (0, f(x_1, x_2)) \tag{4.91}$$

where f is some C^1 function. The x_1 variable could correspond to a well-observed area, while the x_2 variable corresponds to a poorly observed area. Then we find

$$\frac{d\mathbf{e}(0)}{dt} = -(\frac{\partial f}{\partial x_1}\frac{dx_2}{dt}, \frac{\partial f}{\partial x_2}\frac{dx_2}{dt}). \tag{4.92}$$

Therefore the error due to inaccurate observation of x_2 also creates errors in x_1.

From our calculation of the drift, therefore, we cannot tell if the error is due to the model parameters being incorrect, or the projection operator being wrong. This is because the definition of the model implicitly assumed a certain projection operator when the initial condition was picked. This does not mean, however, that errors which appear to be due to the model are in fact due to sensitivity to initial condition. All we have done in the above treatment is decompose the model error into two parts: that due to model equations, and that due to the projection. The shadow law still states that if model error is large, no orbit can be found by taking small perturbations around the initial condition that will shadow for a long time. For weather models, projection is likely to occur over data-poor regions where the interpolation scheme, which involves the forecast model, fails to give an accurate estimate of the real weather.

4.9 Potential problems with the shadow estimation technique

While the SET has worked quite well with a number of models, and should hold in general providing the shadow radius is sufficiently small, there will be situations where it fails to work correctly. This will be the case, for example, if the shadow radius is so large that the model error inside the shadow tube varies significantly from its value on the true orbit. The drift, which is calculated on the true orbit, may then give a misleading indication of the real model error experienced by a shadow orbit.

Situations where the model error either dramatically increases or decreases away from the true orbit are easy to produce, and may occur with real weather models; it is sometimes said of a weather model, for example, that it experiences an initial spin-up error because it is in some way out of balance (similiar to the way that a low dimension model experiences a transient orbit before settling on its attractor). An interpretation of model error might then be that the model is out of balance at start but soon moves back towards balance. The model error may then decrease as the model trajectory moves away from the true trajectory, even while remaining within the specified shadow radius. In this case the calculation of the drift could give an artificially high value.

The linearised dynamics can still be applied, but in such situations it is preferable to linearise around a model trajectory that starts from truth, rather than the true trajectory itself. The reason for this is that model error near the true trajectory doesn't reflect model error at points off that trajectory, but still within the shadow radius. (The main reason we linearised around the true orbit was to allow calculation of model error over a range of prediction times, particularly in cases where the model can shadow for long periods.) Equation (4.13) will remain the same, with the difference that the linear propagator $\mathbf{M}$ is calculated around the model orbit rather than around the true orbit, and the drift vector $\mathbf{d}$ becomes the forecast error. The linearisation will only be valid, as before, until the time at which the model orbit leaves the shadow radius. The SET can also be used, with the same limitation, and the shadow law remains valid.

An example of a model which is 'out of balance' with truth is shown in Figure 4.25. The equations of the true system are

$$\frac{dx}{dt} = x - y - x(x^2 + y^2)/a^2$$

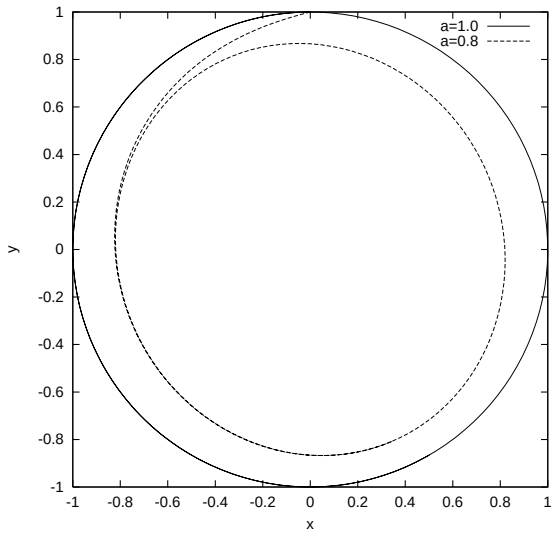


Figure 4.25: True system has a stable periodic orbit with radius 1.0, model has periodic orbit with same angular velocity at radius 0.8.

$$\frac{dy}{dt} = x + y - y(x^2 + y^2)/a^2 \tag{4.93}$$

with $a = 1.0$, which has a stable periodic orbit at radius 1.0, while the model has the same equations but with $a = 0.8$, so the periodic orbit is at 0.8. When started from the point $(0, 1)$, as in the figure, the model quickly moves away from truth to the smaller radius.

Figure 4.26 shows how drift accumulates along the true orbit. The velocity error, shown in the x direction, continually pulls the orbit towards the smaller radius, and the drift accumulates steadily. For a prediction period of 0.5, the drift is about 0.1. If model error is high, we could therefore expect a shadow radius of half the drift, or 0.05. The multipliers of the linear propagator over that prediction period are 0.249 and 0.816, which are both smaller than one so the state space is contracting (i.e. the model is locally dissipative). For a displacement of 0.05 the maximum error due to initial condition is 0.816 times 0.05, which is 0.041, much smaller than the drift. Because the model is highly dissipative, it is reasonable to estimate shadow times from the drift alone, so that shadow diameter should approximately equal the drift over the prediction time.

A feature of this model, though, is that the drift will depend critically on the time step used for the integration. Figure 4.27 shows the drift calculated with time step 0.01 (equal to the integration step for the system) and 0.5. The drift for the

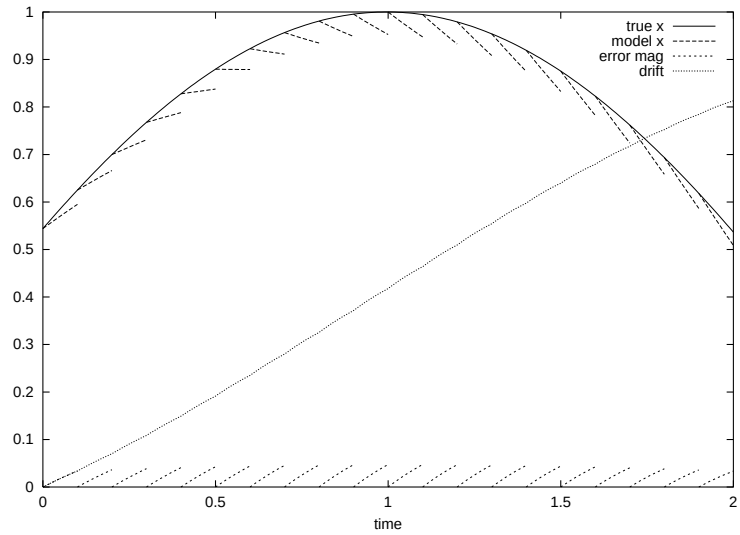


Figure 4.26: Forecast errors for x component. As in Figure 3.2, a number of short term model forecasts were initiated at regular intervals along a true trajectory (shown unscaled). The total error magnitude over both x and y is also shown, as well as the drift, which accumulates steadily.

longer time step is considerably lower. This isn't due, however, to the effect of displacement error dampening out model error; rather, it is because the model error reduces as the model trajectory approaches its attractor. Model error therefore depends on the shadow radius. If the shadow radius were specified as 0.2, then the model could shadow indefinitely, since the periodic orbit at radius 0.8 is within the shadow tolerance of the true orbit at 1.0.

This effect can be seen in Figure 4.28, which shows the shadow diameter as a function of prediction time, along with the drift and the forecast error starting from the point (0,1). Shadow times approach infinity for a shadow diameter of 0.4. The shadow law, which states, in the case of high model error, that drift approximately equals shadow diameter, doesn't hold beyond a shadow diameter of about 0.2. For larger shadow diameters, linearising the dynamics about the true orbit is no longer valid, so it is better to linearise about the model trajectory. The shadow law then states that shadow diameter is about equal to the forecast error. From the figure, we see that this relationship holds up to times of about 0.6.

To summarise, the SET should work for any model/system pair, sufficing the shadow radius is chosen sufficiently small. If the shadow radius is chosen too large, then model error may vary with distance from the true orbit, and the drift calculation may be sensitive to integration step. In such cases, it may be preferable to use forecast

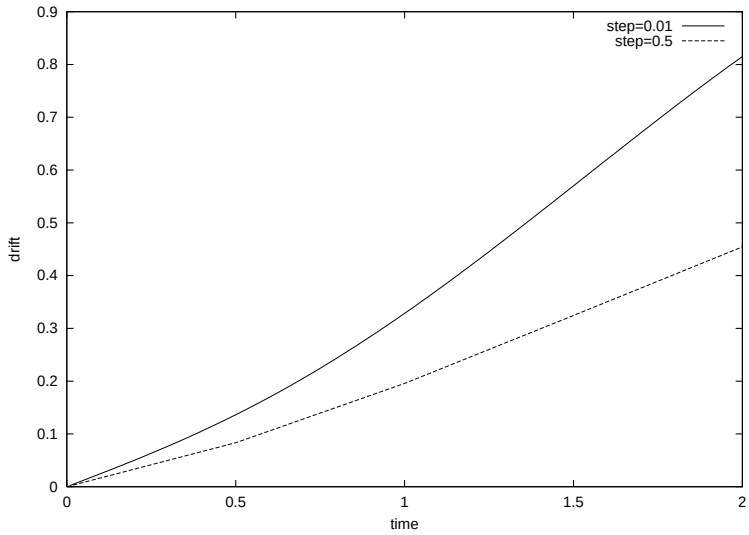


Figure 4.27: Drift calculated with an integration step of 0.01 and 0.5. The drift reduces with the higher step size.

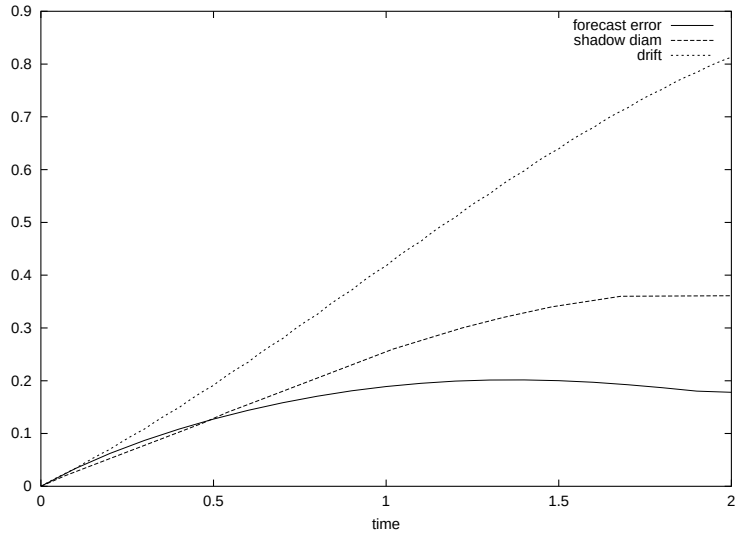


Figure 4.28: Forecast error, shadow diameter and drift as a function of prediction time.

error as a guide to shadow times rather than the drift.

Of course, the best way to determine shadow times is to test for actual shadow orbits. We discuss below ways in which this can be achieved, even for extremely large problems such as weather models. First, though, we examine another kind of difficulty that arises in the calculation of model error, when errors are measured, not over all variables, but only an unrepresentative subset.

4.10 Errors over a subset of variables

The treatment so far, and the development of the linearised dynamics, has assumed that errors are measured using a norm which takes into account all variables of model space. In this section, we consider the likely problems that can occur if this requirement is not satisfied, and the norm measures only a subset of variables. For example, weather models are often verified against analysis, mostly for historical reasons, using the 500 hPa geopotential height. Since the variables in the model typically include surface pressure, two horizontal wind components, temperature, moisture and geopotential height [43], knowing the last of these alone, and at only one level, won't be a very complete indication of the atmospheric state. For the Lorenz system, it is the equivalent of measuring only x_1, and trying to determine the quality of the model based on this alone.

The linearised dynamics assume that the model equations and initial condition are completely known, so restricting error measurements to a subset of variables will affect their accuracy. For example, equation 4.13 states that the error at any time is approximately given by the sum of the drift, and the initial displacement multiplied by the linear propagator. For small times and zero initial displacement, the linear term vanishes and the error is about equal to the drift. This is shown in Figure 4.29 for the Lorenz constant model. The drift closely approximates the error up to about $t = 0.2$.

It isn't the case, however, that errors over individual components e_i are equally well approximated by components d_i of the drift. Figure 4.30 compares the error and the drift for each component. The first component e_1, for example, has departed from its drift equivalent by about $t = 0.1$. It seems that drift is better at approximating the magnitude of the error than its direction. This is evident also from the top panel of Figure 4.29, where the magnitude of the difference between the drift and the error is shown to grow in an exponential manner, and the lower panel of the same figure, which plots the cosine angle between the drift and the error vector.

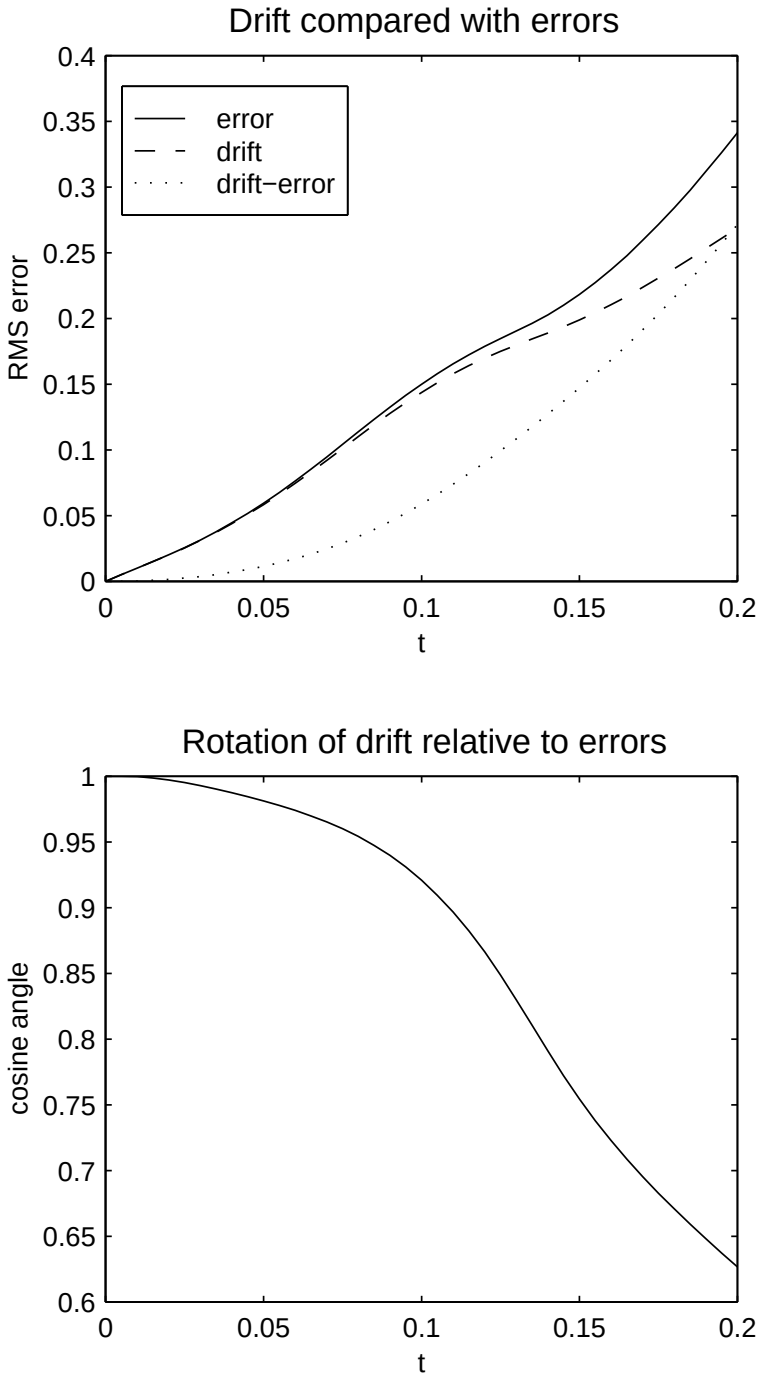


Figure 4.29: Upper panel compares forecast error (solid line), drift (dashed line), and magnitude of the difference between the error and drift vectors (dotted line) for the Lorenz constant model. Lower panel shows the cosine angle between the drift and error vectors.

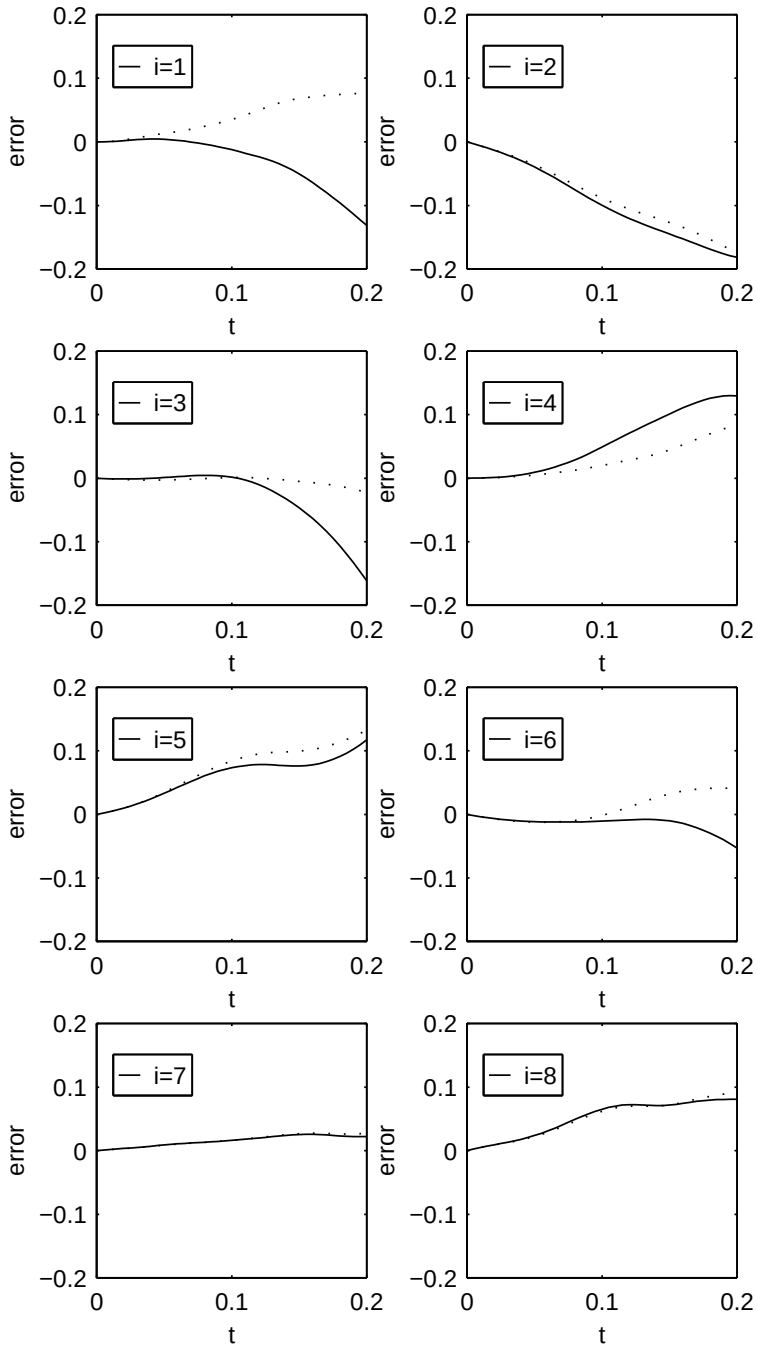


Figure 4.30: Forecast errors (solid lines) and drift (dotted lines) compared for each component.

The reason for this behaviour is that the linear propagator acts on displacements in two ways: by rotating, and by stretching (for example, it rotates the first singular vector $\mathbf{v}_1$ into $\mathbf{u}_1$, and stretches by a factor σ_1). Rotations don't affect the magnitude of the displacement, but they will transfer error in one component into error in another. It follows that individual components will be more sensitive to the effects of the linear propagator than the total error. It is also much harder to separate out the effects of model and initial condition error, since displacements may be large in those components not measured, and create exaggerated errors.

Shadowing times will also be influenced if errors are measured over only a subset of variables. For example, it is easy to imagine that the Lorenz constant model might shadow the two level system for longer times if only x_1 was taken into account, since enormous distortions could be accommodated in the other variables. For weather models, this would be like having a model which shadows indefinitely at 500 hPa, but is completely wrong at ground level. Model error is best understood by considering all variables, and for this reason the work with weather models is performed using an energy metric which represents the energy in the atmosphere over all levels.

4.11 Fast techniques to find shadow orbits

The SET allows one to assess the shadowing capability of a model without actually having to produce the shadowing orbit, and therefore takes much less time. However the linearised dynamics can also be exploited to produce actual shadow orbits.

Shadow orbits to this point have been found by a multi-dimensional optimisation technique known as the simplex method [53]. The simplex method is a somewhat brute force approach since it doesn't use any derivative information and requires many points to be tested. It is therefore impractical for larger systems such as weather models, for which derivative information can be extracted by using the adjoint, as we see in Chapter 6.

In this section we briefly outline two techniques to produce actual shadow orbits, using derivative information, in a more efficient way. The first method will require computation of singular vectors, while the second method has been designed to work as closely as possible with existing weather model code at ECMWF. The methods are illustrated using the Lorenz '96 system. Since we limit ourselves to a single starting point, the shadow times will generally be shorter than those found by the simplex method; in some situations there will be more than one local maximum, so the brute

force approach is necessary to find them all. However computation times are orders of magnitude smaller.

4.11.1 The linear approximation method

To find a shadow orbit is to seek a point in the ball of initial conditions, with radius equal to the shadow radius, which will remain within the shadow radius of the true system for the longest amount of time. It is therefore possible to pose the shadowing problem as a general optimisation problem. Beginning with an initial prediction time t, we solve

$$\begin{aligned} \text{minimise } C(\mathbf{e}) &= \|\mathbf{e}(t)\mathbf{e}^T(t)\| \\ \text{subject to } \|\mathbf{e}(0)\| &\leq r_s \end{aligned} \tag{4.94}$$

where $\mathbf{e}$ is the error term. The optimal solution will minimise the offset from truth at time t over all initial conditions within the shadow radius. If the final error also satisfies

$$\|\mathbf{e}(t)\| \leq r_s \tag{4.95}$$

then the prediction time t can be increased, and the process repeated.

Strictly speaking, we should demand that the error $\mathbf{e}(t)$ remain within the shadow radius for all intermediate times between 0 and t; however we shall relax that condition for the time being, because for the systems considered here it will usually hold so long as the trajectory is within the shadow tube at the initial and final points.

Were the system perfectly linear, the minimisation problem could be solved in a single step just by solving the Lagrangian problem of equation 4.27. With nonlinear systems, a one-step approach isn't feasible, since the system is sensitive to small displacements, and if the step is too large it is likely to miss the optimum completely. The standard optimisation approach in such cases is to begin with a starting point, and iterate slowly from it, using a rule to determine the step direction, until further steps no longer produce an improvement [22].

The problem then becomes how to choose the step direction. One method which has been proposed is to search in the space of the leading singular vectors, the rationale being that these are the perturbations which give the maximum final displacement and are therefore most likely to offset model error. A topic we have often returned to, however, is that model error and displacement error are not the same thing, and are generally in different directions.

In fact, it is easy to imagine situations where the largest step must be taken in the direction of the trailing, rather than leading, singular vector. Referring to the schematic diagram of the linearised dynamics, Figure 4.6, suppose that the model error is aligned with the evolved trailing singular vector u_2. Then the initial perturbation to offset it must be in the direction of u_1, and, if the multiplier σ_2 is much smaller than unity, then the initial displacement must be much larger than would be the case if the model error were aligned with the leading singular vector u_1.

In the case of weather models, searching for an optimal displacement in the subspace of leading singular vectors, which typically has a dimension of about 25 in an overall space of millions, would be limiting one's self to a rather impoverished set of step directions. (One advantage of singular vectors, though, is that they give the maximum final displacement for a given initial displacement, so if no other information is available they may be a good place to start.)

The approach adopted here is again based on the linearised dynamics, which has been shown to hold to a good approximation for those trajectories which shadow (the ones we are interested in). We first transform the minimisation problem to an approximate linear one:

$$\begin{aligned} \text{minimise } C(\mathbf{e}) &= \|\mathbf{U}(t)\mathbf{\Sigma}(t)\mathbf{V}^T(t)\cdot\mathbf{e}(t)+\mathbf{d}(t)\| \\ \text{subject to } \|\mathbf{e}(0)\| &\leq r_s. \end{aligned} \tag{4.96}$$

This is a constrained optimisation problem, which is more complicated than an unconstrained one, since the optimal direction will depend on whether the boundary on the initial condition $\mathbf{e}(0)$ is active or not, i.e. if $\|\mathbf{e}(0)\| = r_s$. The constraint is important, because solving the unconstrained problem won't give the same solution, even if the process is stopped when the initial condition exceeds the boundary condition [22].

Various techniques exist to solve this problem, but the schematic diagram of the linearised dynamics, Figure 4.6, suggests a simple and novel approach. It was already seen in the development of the shadow test that the ball of initial conditions will contain one point that shadows if the enlarged ellipse, where each axis is increased by unity, contains the zero vector. Therefore the constrained problem can approximately be solved by finding an initial condition $\mathbf{e}(0)$ which satisfies

$$\mathbf{U}(t)(\mathbf{\Sigma}(t)+\mathbf{I})\mathbf{V}^T(t)\cdot\mathbf{e}(0)+\mathbf{d}(t)=\mathbf{0}. \tag{4.97}$$

The vector $\mathbf{e}(0)$ can be solved for directly:

$$\mathbf{e}(0) = -\mathbf{V}^T(t)(\mathbf{\Sigma}(t)+\mathbf{I})^{-1}\mathbf{U}^T(t)\mathbf{d}(t). \tag{4.98}$$

Note that $\mathbf{\Sigma}(t) + \mathbf{I}$ is positive diagonal, hence invertible. The actual step would be taken in this direction, but with a reduced magnitude determined by the optimisation routine.

A scheme was implemented which proceeded as follows. The initial condition was chosen to coincide with the true orbit. The initial prediction time was chosen (typically 0.05 units), and the first step taken in the direction suggested by the zero of the enlarged ellipse. The step size was chosen to be a factor (typically 0.5) of the difference between the initial condition's magnitude and the shadow radius. Thus the steps would never exceed the shadow radius. The new initial condition was then run forward under the full dynamics until it ceased to shadow. This time became the new prediction time. The process was then repeated until it ceased to improve shadow times.

The method was tested by calculating shadow times of the Lorenz '96 linear model relative to the two level system. With shadow radius 0.2, the technique gave a median shadow time of 0.95, as opposed to 0.97 for the brute-force simplex method. Considering that the new method only begins from a single starting point, and is orders of magnitude faster, this is an excellent result. For shadow radius 0.4, the median shadow time is 1.80, as opposed to 2.35 for the simplex method. For the increased shadow radius, there is a greater possibility of multiple local minima, which can only be found by using a number of starting points. Figure 4.31 shows a histogram of shadow times for the linear model using the new optimisation method. It can be compared with Figure 4.11.

4.11.2 The 'pinch' method

While the linear approximation method is much faster than the simplex method, it requires the computation of singular vectors. For weather models it is possible to calculate the linear propagator matrix $\mathbf{M}$, using the adjoint (see Chapter 6), but the dimension of the matrix prohibits computation of all the singular vectors. We therefore need an optimisation scheme which can work with the linear propagator in its raw form. Also, optimisation methods are currently used with ECMWF models for purposes such as 4D-Var data assimilation, so the method should be capable of exploiting existing code for those models.

The linearised optimisation problem, written now without the singular value decomposition, is

$$\text{minimise}\, C(\mathbf{e}) \quad = \quad \|\mathbf{M}(t) \cdot \mathbf{e}(t) + \mathbf{d}(t)\|$$

$$\text{subject to } \|\mathbf{e}(0)\| \quad \leq \quad r_s \tag{4.99}$$

which again is a constrained problem. One approach to such a problem is the penalty method, which transforms the constrained problem into an unconstrained one by adding a penalty term:

$$\text{minimise } C(\mathbf{e}) = \|\mathbf{M}(t) \cdot \mathbf{e}(t) + \mathbf{d}(t)\| + \lambda(\|\mathbf{e}(0)\| - r_s^2) \tag{4.100}$$

where λ is some suitably large constant. The above formulation will force the initial condition $\mathbf{e}(0)$ to have radius r_s; alternatively, the penalty function could switch on only if the radius exceeds the shadow radius.

There is a symmetry to the shadow problem, however, which doesn't distinguish between the initial and final displacements; we could equally well minimise the initial displacement subject to the final displacement being within the shadow radius, i.e.

$$\begin{aligned} \text{minimise } C(\mathbf{e}) \quad &= \quad \|\mathbf{e}(0)\| \\ \text{subject to } \|\mathbf{M}(t) \cdot \mathbf{e}(t) + \mathbf{d}(t)\| \quad &\leq \quad r_s. \end{aligned} \tag{4.101}$$

This would mean that the final displacement became the penalty term, instead of the initial displacement.

A balanced approach, then, is to minimise the sum of the initial and final displacements

$$\text{minimise } C(\mathbf{e}) = \|\mathbf{M}(t) \cdot \mathbf{e} + \mathbf{d}(t)\|^2 + \|\mathbf{e}(0)\|^2 \tag{4.102}$$

without specifying what the shadow radius r_s should be. For a particular prediction time t, this method should produce the orbit which minimises the initial and final displacements. The shadow radius r_s can then be taken as the maximum of these two values. We assume that intermediate values will remain within bounds; this is easily checked for.

One method to determine the step direction would be to find the gradient of the cost function $C(\mathbf{e})$, which is given by

$$2\mathbf{M}^T(t)(\mathbf{M}(t) \cdot \mathbf{e} + \mathbf{d}(t)) + 2\mathbf{e}(0) \tag{4.103}$$

and step along the negative of the gradient (the so-called steepest descent method). This approach was used in early versions of the ECMWF sensitivity code [55]. A disadvantage of the method is that it tends not to converge well if the gradient matrix is ill-conditioned [22].

A more robust technique, commonly used in 4D-Var [16], is to take a step in the direction determined by Newton's method. Because of the quadratic form of the cost function, the Hessian of the cost function can be calculated as

$$2\mathbf{M}^T(t)\mathbf{M}(t) + 2\mathbf{I}. \tag{4.104}$$

By using the Hessian information, Newton's method can achieve quadratic convergence. (Implementing it with low dimension models is more straightforward than with 4D-Var, which is a subject in its own right [36, 15].)

The optimisation scheme is then as follows: for an initial prediction time, determine an initial condition $\mathbf{e}(0)$ which minimises the sum of displacements, by taking a sequence of steps in the Newton direction. The shadow radius is taken to be the largest of the initial and final displacements. Then increase the prediction time by an increment, and repeat the process, using the previous shadow point as the new starting point. The process is repeated until the specified shadow time is exceeded (or a curve of shadow radius versus shadow time stored, and the time for the specified radius read off by interpolation).

This 'pinch' method, which finds the shadow orbit by simultaneously minimising the initial and final displacements, gives results which are less accurate than the previous method: at a radius of 0.2, the average shadow time is 0.94 (the brute-force method gives 0.97), while for shadow radius 0.4 the average shadow time is 1.62, a substantial reduction from the 2.28 of the brute-force method. For good models with long shadow times and complicated shadow orbits as seen in 4.1, the 'pinch' method may not be adequate.

A useful feature of this technique, however, is that it can be implemented in the ECMWF code with a fairly minimal degree of work; and it is this which has motivated its development here. Its efficiency will depend on how long the ECMWF models can shadow (if shadowing times are such that the model ceases to shadow before it becomes nonlinear, then just about any optimisation routine will suffice).

In the next chapter, we will discuss the longer term characteristics of the Lorenz system and its models. In other words, we will turn our attention from the short to medium range 'weather' to the longer term 'climate'. Before doing so, we briefly summarise the progress made so far in the understanding of model error.

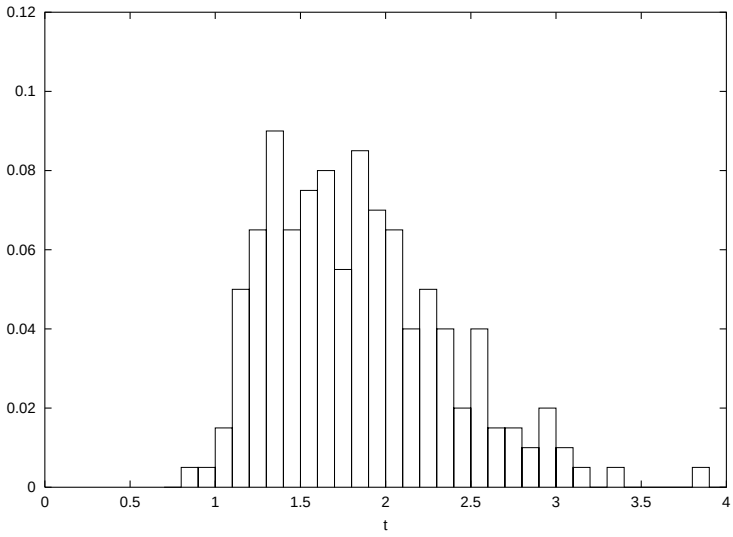


Figure 4.31: Histogram of shadow times using the Newton method, linear model.

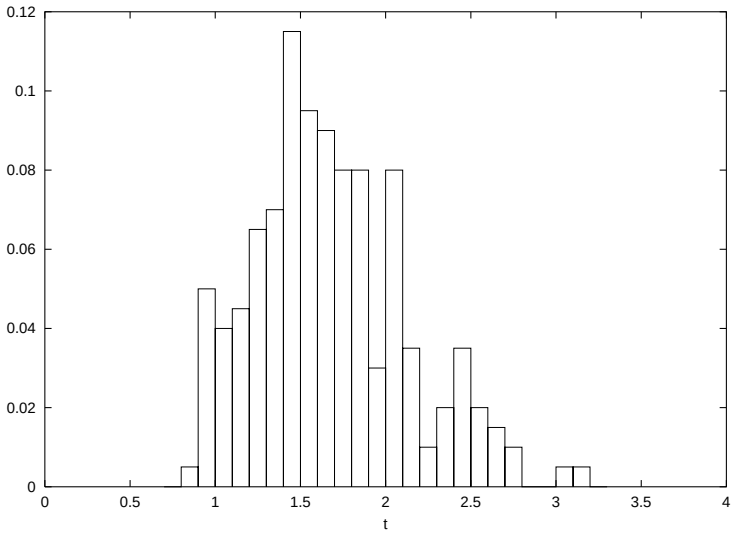


Figure 4.32: Histogram of shadow times using the pinch method, linear model.

4.12 Summary

In this chapter, results from experiments on variants of Lorenz '96 and other model/system pairs have been translated into a number of insights into model error and shadowing. Equation 4.22, which describes the linearised dynamics near a true orbit for any model/system pair, summarises the relationship between the two. Shadow times can be estimated using the SET, which is based on the linearised dynamics, and includes the drift and a linear propagator term. For any locally dissipative model, the shadow law gives an approximate upper bound on shadow times in terms of the drift alone; and when model error is high, the allowable drift over a shadow orbit is about equal to twice the shadow radius.

The methods have been tested over a range of model/system pairs. The SET was found to work well for both the constant and linear Lorenz models, though the technique fails when coupling in the true system is reduced to half its normal value, because shadow times become excessively long (the shadow law still holds) A modified version of the Rössler system showed that the equations need not be symmetric, while comparisons of the full 7-D Saltzman system with its 3-D model showed that the SET can work even when shadow times vary greatly over the attractor.

As a method to measure model error, it is worth distinguishing between local model drift, and other measures such as forecast error after a certain period, or the tendency error at a particular time. Forecast errors convolute initial condition and model error, while tendency error doesn't allow for the fact that model error can be non-additive over the prediction period.

In the introduction, three questions were raised, asking how do we measure model error, how do we estimate shadow times, and how do we optimise a model's parameters. From the above discussion, we are now in a position to address these points. The research into model error indicates:

- Model drift, as defined in terms of integrated velocity error, is a useful measure of model error, and the primary determinant of predictability
- Shadow times can be estimated for any model/system pair using the model drift and (for longer shadow times) a modified version of the model's linear propagator matrix, without the need to produce an actual shadowing orbit
- For locally dissipative models, shadow times are bounded above by the shadow law, which states that RMS drift over shadow orbits must be smaller than twice the shadow radius

- Predictability is best optimised by minimising the model drift (which is determined by the model's low frequency velocity error)

As an example of the last point, the constant model was chosen to have its forcing equal to the average true forcing, which minimises the RMS velocity error. This is equivalent to minimising the drift in the limit as the integration time goes to zero. In general, a model can be optimised by minimising its drift over a specified prediction time. For example, if the goal is to predict over a five day period, the 5 day drift can be calculated at various points on the true attractor, and the model parameters chosen to minimise it. Alternatively, if the velocity error power spectrum is calculated on the attractor, the drift can be minimised over a range of prediction times simply by adjusting the weighting of the power spectra to calculate the expected drift at each time.

The key result from this chapter is the shadow law, which provides an easily computable upper bound on shadow times. It is a mathematically demonstrable and easily verified statement which applies across a broad range of dispersive, chaotic models. Figure 4.33 is a graphic illustration of the shadowing law: for the more than 60 experiments conducted with a number of model/system pairs, the ratio of drift to shadow diameter over a shadow orbit is near or below 1. We will later use this simple result to address the question of model error in weather forecasting.

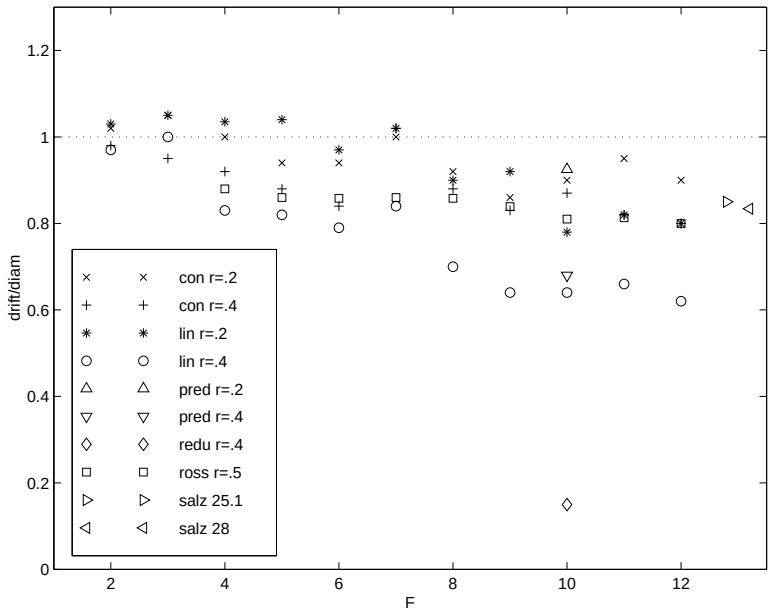


Figure 4.33: Average ratio of drift to shadow diameter for maximal shadow orbits. Shown are the constant and linear Lorenz models from Figures 3.14 and 3.24 as a function of forcing; the predictor model; the low coupling model; the modified Rössler model from Figure 4.15; and the Saltzman 3D model at $\lambda = 25.1$ and 28. Results are determined by averaging over 20 shadow runs, except for the predictor model and low coupling model where only a single run was tested. All models conform to the shadow law (though these are average rather than RMS results). The low coupling model has very low drift because the extremely long shadow orbits are dominated by nonlinear effects.

Chapter 5

Climatology

5.1 Introduction

Prediction problems have been described by Lorenz [39] as falling into two categories. Problems which depend on the initial condition, such as short to medium range weather forecasting, or El Nino, are described as 'predictions of the first kind'. Longer term problems, such as effects on the Earth's climate of volcanic emissions or carbon dioxide levels, are referred to as predictions of the second kind.

In general, modelling the climatology seems to be a somewhat easier problem than modelling short term behaviour. For example, numerous models have been constructed which do a reasonably good job of modelling certain aspects of financial time series, yet predicting the next stockmarket crash is still an elusive goal. The converse also holds: it is easy to construct a model of the Lorenz '96 system which predicts short term, but, due to a small damping term, eventually trends to zero.

It should also be noted that, while model climatology is affected by model error, it does not seem possible to measure model error in a meaningful manner by analysing the climatology alone. In general the true system is only known through observations of a true orbit. Therefore model error is strictly speaking only defined on projections of true orbits into model space (for how can we measure model error in a region of state space that the true system never enters?).

Despite these caveats, there appear to be some links between short and long range predictability. For example, referring to Figure 4.16, it was noticed that the attractor of the Saltzman 7-D system was closer to that of the reduced 3-D model at a parameter value of $\lambda = 25.1$ than at the higher value of 28. It also turned out that shadowing was much improved at the lower parameter setting. Intuitively, it seems reasonable

that there be a connection between the two kinds of predictability. Also, shadowing was seen in Chapter 3 to depend primarily on low frequency errors - precisely the type that one might expect to influence long-term climatology.

In this chapter, we turn our attention to issues related to climatology and predictions of the second kind, concentrating on the Lorenz '96 system and its models. We will examine how climatology of the model or system depends on forcing; what properties hold over a range of forcings; and how the system climatology might be modelled at a particular forcing.

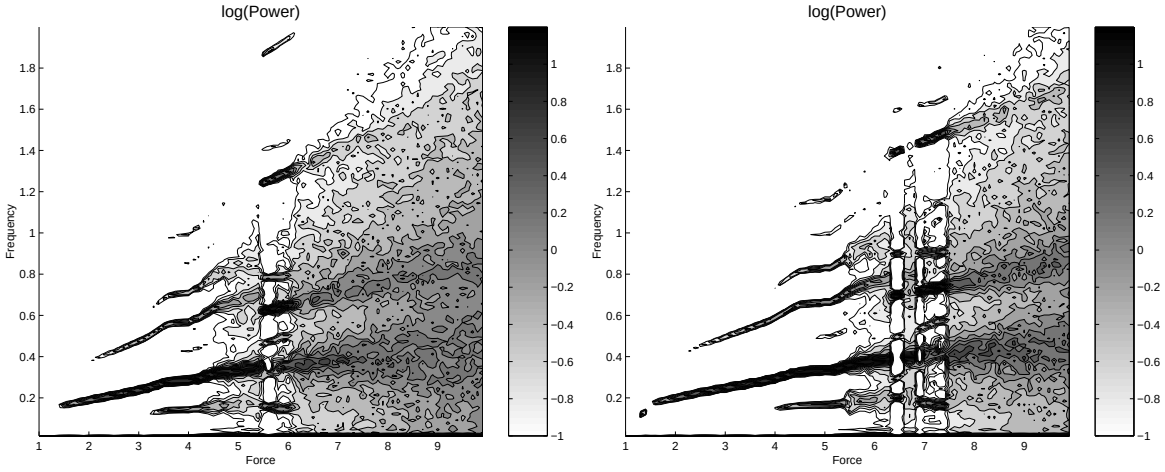
One question, of special relevance in high dimension systems, is how to judge when one attractor is similar to another. A possible measure of a system's climatology is the amount of power contained at different frequencies. Spectral bifurcation diagrams express this information over a range of forcings, and therefore provide a snapshot of climatological variation.

Figure 5.1 shows such diagrams for the true system and constant model. Also shown is the difference between the two, i.e. the mismatch between the attractors as expressed in terms of power spectra. For the constant model there is clearly a difference around $F = 6$ where the true system is chaotic but the constant model is periodic or quasi-periodic. Also around $F = 2.5$ there is a mismatch in the frequencies of the periodic orbits, which appears as a split in the lines. The linear model shows a general improvement of fit over the constant model.

These figures encapsulate a great deal of detailed information, but it is hard to draw any general conclusions from them - especially if we are more interested in general behaviour rather than whether the model is chaotic or periodic. Another, somewhat simpler, measure of climatology is to consider the first and second order moments, i.e. $\langle x_i \rangle$ and $\langle x_i^2 \rangle$. We might then ask whether optimising the model for these macroscopic quantities is the same as optimising for short term predictability. In the next section we prove that this depends on the model; in one case the two aims are at odds, while in another they appear to agree.

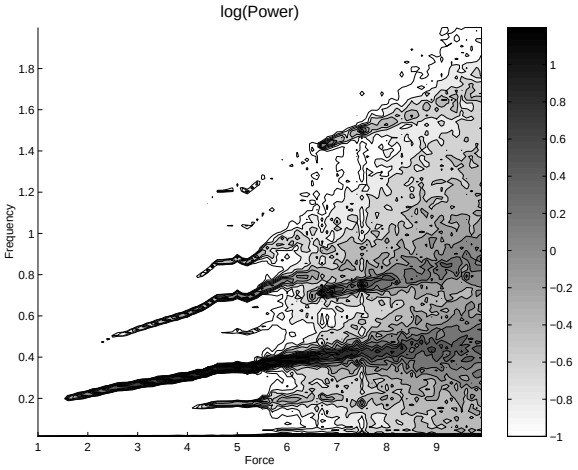
5.2 Chaotic in the small, predictable in the large

The Lorenz '96 systems undergo complex changes in behaviour as forcing is increased. Nevertheless, it was seen in Chapter 3 that quantities such as forcing error vary in a simple manner as a function of forcing. Similar relationships can be deduced for $\langle x_i \rangle$ and $\langle x_i^2 \rangle$ by averaging the model equations over long time periods. These

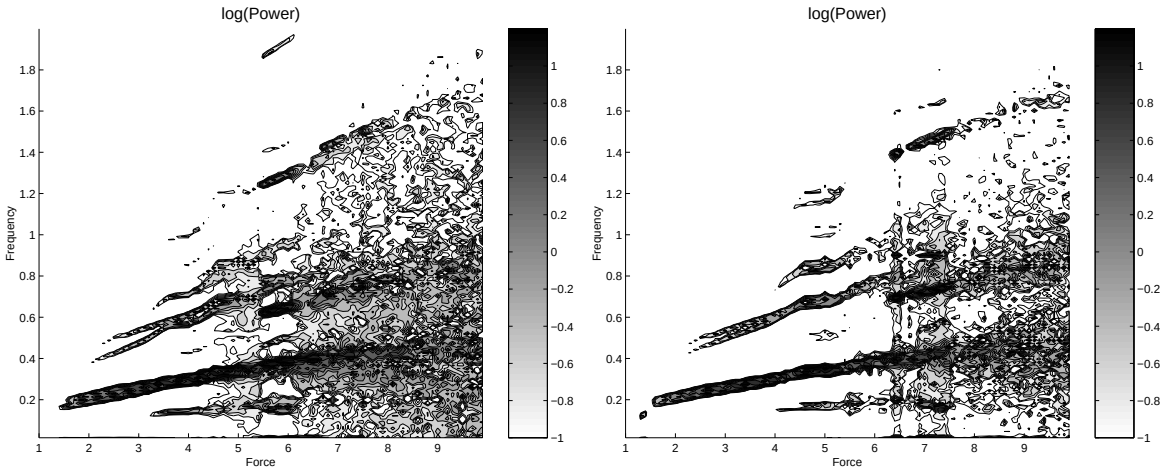


(a) Constant model

(b) Linear model



(c) True system



(d) Constant model mismatch

(e) Linear model mismatch

Figure 5.1: Spectral bifurcation diagrams for constant model, linear model, and true system, and model mis-match with true system.

relationships can then be exploited to determine how best to model the system's macroscopic behaviour with the constant or linear model.

The one level system equation 2.1 is:

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F. \tag{5.1}$$

Multiplying each side of this equation by x_i, we obtain

$$x_i \frac{dx_i}{dt} = \frac{1}{2}\frac{dx_i^2}{dt} = x_1 x_{i-1}(x_{i+1} - x_{i-2}) - x_i^2 + F x_i. \tag{5.2}$$

Summing over all i, the advection terms cancel out, leaving

$$\frac{1}{2}\sum_{i=1}^{n} \frac{dx_i^2}{dt} = -\sum_{i=1}^{n} x_i^2 + F\sum_{i=1}^{n} x_i. \tag{5.3}$$

Let L be a real number. Then

$$\frac{1}{L}\int_0^L \frac{1}{2}\sum_{i=1}^{n} \frac{dx_i^2}{dt} dt = -\frac{1}{L}\int_0^L \sum_{i=1}^{n} x_i^2 dt + \frac{1}{L} F \int_0^L \sum_{i=1}^{n} x_i dt. \tag{5.4}$$

Now, taking the limit as L goes to infinity, the left hand side is just

$$\lim_{L\to\infty} \frac{1}{L}\frac{1}{2}\sum_{i=1}^{n} x_i^2. \tag{5.5}$$

It is easily seen, for example by the Trapping Region Lemma [1], that x_i is bounded, and so the above term goes to zero in the limit. The first term on the right hand side, meanwhile, converges to $n\langle x_i^2 \rangle$, where the average is over the attractor, and the second term is $n\langle x_i \rangle$. Therefore we obtain the result that the mean of x_i^2 is equal to the forcing times the mean of x_i:

$$\langle x_i^2 \rangle = F\langle x_i \rangle. \tag{5.6}$$

A similar technique can be applied to the two level system. Equation 2.2 for the large scale variables is

$$\frac{d\tilde{x}_i}{dt} = \tilde{x}_{i-1}(\tilde{x}_{i+1} - \tilde{x}_{i-2}) - \tilde{x}_i + F - \frac{hc}{b}\sum_{j=1}^{m} \tilde{y}_{i,j}. \tag{5.7}$$

The same procedure as that followed above gives a similar result, but now there is an additional term due to the $\tilde{y}_{i,j}$ variables:

$$\langle \tilde{x}_i^2 \rangle = F\langle \tilde{x}_i \rangle - \frac{mhc}{b}\langle \tilde{x}_i \tilde{y}_{i,j} \rangle. \tag{5.8}$$

The system equation for the $\tilde{y}_{i,j}$ variables is

$$\frac{d\tilde{y}_{i,j}}{dt} = cb\tilde{y}_{i,j+1}(\tilde{y}_{i,j-1} - \tilde{y}_{i,j+2}) - c\tilde{y}_{i,j} + \frac{hc}{b}\tilde{x}_i. \tag{5.9}$$

Multiplying now by $\tilde{y}_{i,j}$ and proceeding as above gives

$$\langle \tilde{y}_{i,j}^2 \rangle = \frac{mh}{b}\langle \tilde{x}_i \tilde{y}_{i,j} \rangle. \tag{5.10}$$

Combining these equations yields

$$\langle \tilde{y}_{i,j}^2 \rangle = \frac{1}{mc}(F\langle \tilde{x}_i \rangle - \langle \tilde{x}_i^2 \rangle). \tag{5.11}$$

This result means that information about the fine-scale $\tilde{y}_{i,j}$ variables can be deduced by observing only the large-scale $\tilde{x}$ variables.

Suppose now that we wish to model the macroscopic behaviour of the two level system using the constant model with forcing P^c. For the model, we have

$$\frac{\langle x_i^2 \rangle}{\langle x_i \rangle} = P^c \tag{5.12}$$

while for the system we have

$$\frac{\langle \tilde{x}_i^2 \rangle}{\langle \tilde{x}_i \rangle} = F - \frac{mc\langle \tilde{y}_{i,j}^2 \rangle}{\langle \tilde{x}_i \rangle}. \tag{5.13}$$

If we demand that the ratio of the first and second moments agree, so

$$\frac{\langle x_i^2 \rangle}{\langle x_i \rangle} = \frac{\langle \tilde{x}_i^2 \rangle}{\langle \tilde{x}_i \rangle}, \tag{5.14}$$

then it follows that

$$P^c = \frac{\langle \tilde{x}_i^2 \rangle}{\langle \tilde{x}_i \rangle}. \tag{5.15}$$

The value of P^c arrived at is not the same as the value used in Chapter 3 for shadowing purposes. At $F = 10$, for example, the optimal forcing is 8.87 as opposed to 9.63 for shadowing. Also, the chosen value of P^c gives the correct ratio of $\langle x_i^2 \rangle$ to $\langle x_i \rangle$, but never the correct value of either term. This is seen in Figures 5.2 and 5.3, which show $\langle \tilde{x}_i \rangle$ and $\langle \tilde{x}_i^2 \rangle$ respectively, along with the corresponding values of $\langle x_i \rangle$ and $\langle x_i^2 \rangle$ for the constant model with forcing P^c. In either graph, the curve for the model is below the curve for the system, so it is impossible to arrive at a constant model which has both $\langle x_i^2 \rangle$ and $\langle x_i \rangle$ correct simultaneously.

An interesting feature of the graphs is that the quantities vary in a regular manner with forcing. The mean $\langle \tilde{x}_i \rangle$ goes approximately with the square root, and $\langle \tilde{x}_i^2 \rangle$ with

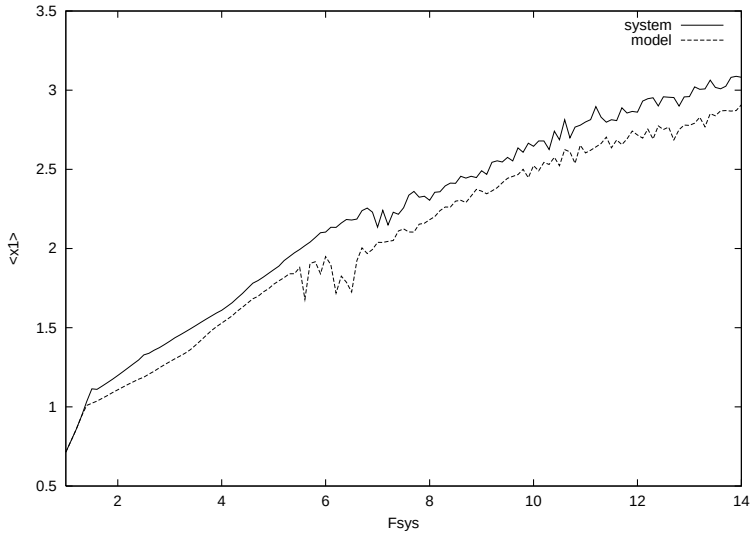


Figure 5.2: Mean of $\tilde{x}_1$ for the two level system as function of F. Also shown is the mean of x_1 for the constant model (dashed line), with forcing chosen to give the correct ratio $\frac{\langle x_i^2 \rangle}{\langle x_i \rangle}$. The model curve is always below that of the system.

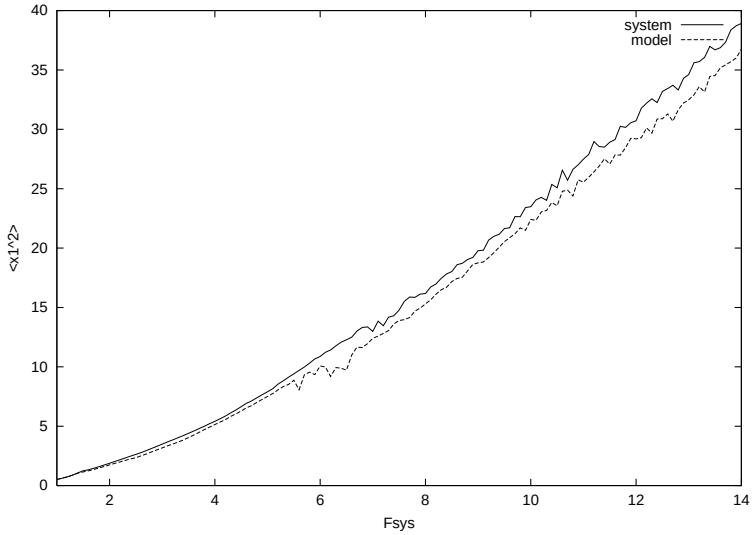


Figure 5.3: Mean of $\tilde{x}_1^2$ for the two level system as function of F. Also shown is the mean of x_1^2 for the constant model (dashed line), with forcing chosen to give the correct ratio $\frac{\langle x_i^2 \rangle}{\langle x_i \rangle}$. The model curve is always below that of the system.

the square of forcing. It seems reasonable that model error should also vary in a simple way with forcing.

One might expect to do a better job of modelling the two level system with the linear model, which has two parameters to adjust. The relationship between $\langle x_i^2 \rangle$ and $\langle x_i \rangle$ can be computed for the linear model just as for the constant model. Recall that the linear model has a forcing term $\mathbf{P^l}$ with components given by

$$\mathbf{P_i^l}(\tilde{x}_i) = \alpha_0 + \alpha_1 \tilde{x}_i. \tag{5.16}$$

Following the procedure above, we calculate that

$$\frac{\langle x_i^2 \rangle}{\langle x_i \rangle} = \frac{\alpha_0}{1 - \alpha_1} \tag{5.17}$$

with the additional linear term in the parameterisation introducing a factor $1 - \alpha_1$ in the denominator. Thus, to preserve the ratio $\frac{\langle x_i^2 \rangle}{\langle x_i \rangle}$ of the true system, we require

$$\frac{\alpha_0}{1 - \alpha_1} = \frac{\langle \tilde{x}_i^2 \rangle}{\langle \tilde{x}_i \rangle} \tag{5.18}$$

which solved for α_1 gives

$$\alpha_1 = 1 - \alpha_1 \frac{\langle \tilde{x}_i^2 \rangle}{\langle \tilde{x}_i \rangle}. \tag{5.19}$$

Therefore, given a value of α_0, the corresponding value of α_1 can be found.

Figure 5.4 shows how the ratio of $\langle x_i \rangle$ to $\langle \tilde{x}_i \rangle$ changes with the forcing offset $\alpha_0 - F$. A graph of the ratio of $\langle x_i^2 \rangle$ to $\langle \tilde{x}_i^2 \rangle$ is indistinguishable. The ratios is approximately 1.0 when the offset is zero, or $\alpha_0 = F$. The corresponding value of α_1 is then

$$\alpha_1 = 1 - F \frac{\langle \tilde{x}_i \rangle}{\langle \tilde{x}_i^2 \rangle}. \tag{5.20}$$

For $F = 10$, the resulting slope is $\alpha_1 = -0.127079$. To the margin of error, these coefficients are indistinguishable from the values $\alpha_0 = 10 - 0.046$ and $\alpha_1 = -0.122$ used in the linear model for shadowing purposes.

In fact, the linear model, as derived for shadowing, turns out do a fine job of reproducing the true system's macroscopic behaviour over a range of forcings. Figure 5.5 shows $\langle x_i \rangle$ for the linear model compared to $\langle \tilde{x}_i \rangle$ for the system as a function of system forcings, while Figure 5.5 shows $\langle x_i^2 \rangle$ compared to $\langle \tilde{x}_i^2 \rangle$. Agreement is excellent except in the regions near $F = 1.3$ and $F = 7$. The area around $F = 1.3$ was found in Chapter 3 to be a problem for both the constant and linear models, since this is the point where the fine-scale variables become non-zero in the true system. Referring to

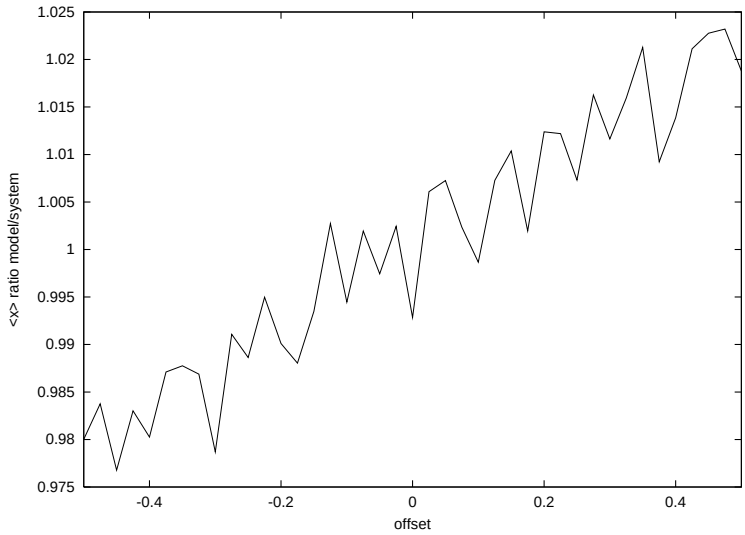


Figure 5.4: Ratio of $\langle x_i \rangle$ to $\langle \tilde{x}_i \rangle$ for values of the linear model offset $\alpha_0 - F$ where $F = 10$. The linear model gives $\langle x_i \rangle \approx \langle \tilde{x}_i \rangle$ for offsets near zero. The corresponding slope agrees with that of the linear model used for shadowing purposes.

the spectral bifurcation graphs in Figure 5.1, the problem around $F = 7$ appears to be that the system is already chaotic, while the model is in a quasi-periodic region. Apart from these areas, correspondence is almost exact. The linear model is certainly the simplest model which successfully reproduces the macroscopic behaviour of the true system.

Beyond $\langle x_i \rangle$ and $\langle x_i^2 \rangle$, one might ask what other aspects of the climatology can be modelled. In the next section we look at the $F = 10$ two level system climatology in greater detail, and consider other ways of approximating it.

5.3 Modelling the climatology of the two level system

The linear model may be the simplest model to capture the mean and variance of the two level system, but, as seen by the spectral bifurcation diagram Figure 5.1, it is still not perfect at modelling the power spectrum. In this section we try other approaches to find models which produce a similar climatology to that of the true system, for the specific forcing $F = 10$, where the definition of 'similar' is broadened to include phase space plots and the power spectrum.

Along with the constant and linear models, we consider also two other models

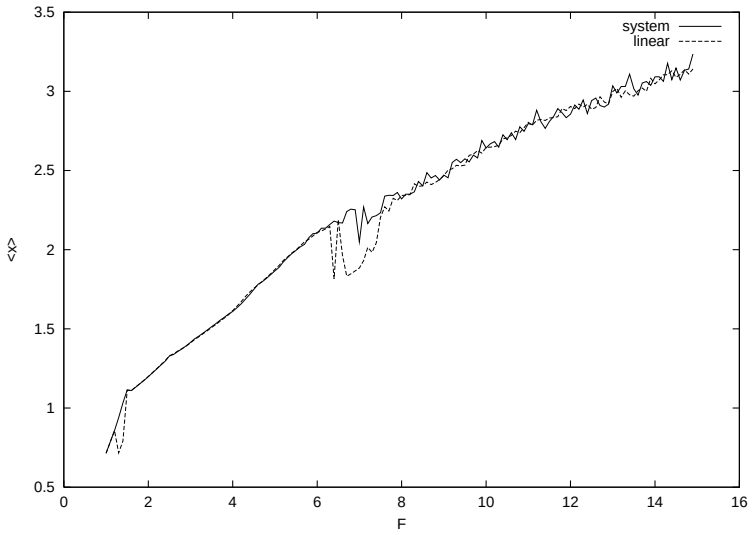


Figure 5.5: Plot of $\langle x_i \rangle$ for the linear model, compared to $\langle \tilde{x}_i \rangle$ for the system, over a range of system forcings. Agreement is excellent except for the region of $F = 1.3$ and $F = 7$.

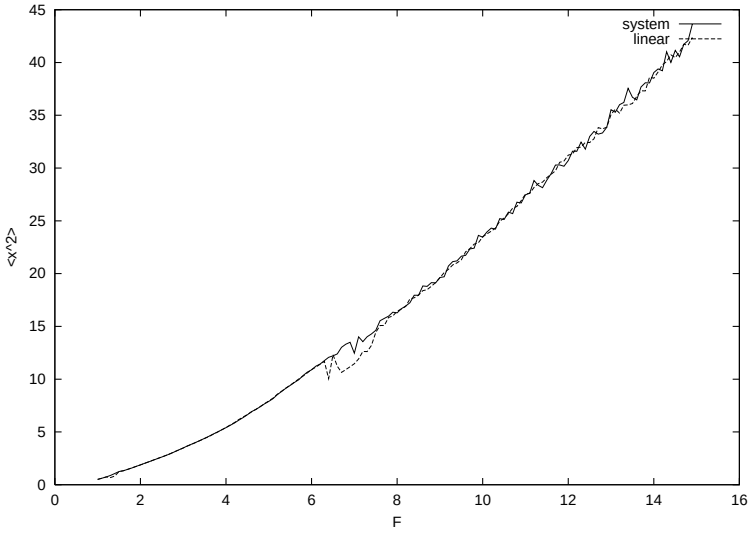


Figure 5.6: Plot of $\langle x_i^2 \rangle$ for the linear model, compared to $\langle \tilde{x}_i^2 \rangle$ for the system, over a range of system forcings. Agreement is excellent except for the region of $F = 1.3$ and $F = 7$.

constructed using stochastic methods. The motivation here for stochastic models is to see if the climatology can be improved by adding random terms which make the model statistically similar to the true system.

5.3.1 Stochastic models

The first stochastic model we will consider draws the forcing at each point from a distribution of observed forcings. Suppose we observe the true system forcing $\tilde{\mathbf{F}}(t)$ at K points on the attractor, where the points are chosen so their distribution reflects the natural measure. We then define the model

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + P_i^r \qquad \text{random model} \tag{5.21}$$

where P_i^r is chosen at random from the distribution at each time step. In practice, the size of the distribution was 10,000 points, taken from an orbit at intervals of 0.185 time units, which is the decorrelation time for $\tilde{\mathbf{F}}(t)$.

The second stochastic model attempts to better model the data by using an AR(1) fit [11] to generate a time series of the form

$$P^{AR}(n) = \langle \tilde{F} \rangle + a_1 P^{AR}(n-1) + a_0. \tag{5.22}$$

The covariance term a_1 is given by e^{-1/n_d}, where n_d is the (non-integer) number of time steps corresponding to the decorrelation time for $\tilde{\mathbf{F}}(t)$. For this model, the resulting covariance was $a_1 = 0.97$. The term a_0 is a random term, with zero mean and variance 0.165 chosen to make the AR(1) series variance match the true variance. The model is then

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + P_i^{AR} \qquad \text{AR(1) model.} \tag{5.23}$$

Adding stochastic terms to a model seems unlikely to improve shadowing performance, since random perturbations will only add to the forcing error variance, which was seen in Chapter 3 to limit shadowing times. In fact, we have to be careful about how we define shadowing times for these systems. In the case of the random model, for example, there will be one series of random choices of the forcing which will be exactly the same as for the true system, and therefore shadow indefinitely. What we can ask instead is whether adding the random terms on average increases or decreases the time that the model will track the true system. As expected, the answer is that it decreases tracking times. The constant model shadows at $F = 10$ and shadow radius

0.4 for about 0.6 time units, while the random model tracks on average 0.46 time units, and the AR(1) model an average 0.32 time units.

Note also that, if reducing the drift is the goal, then, because the drift measures the integral of the velocity error over a fixed time, it follows that a parameter varied stochastically with time will give the same drift as one where the same parameter is held constant at some intermediate value over the prediction period. Therefore stochastic models offer no real advantage over non-stochastic models in improving short term predictability. The question is then whether they affect the long term behaviour.

5.3.2 Projection on EOF's

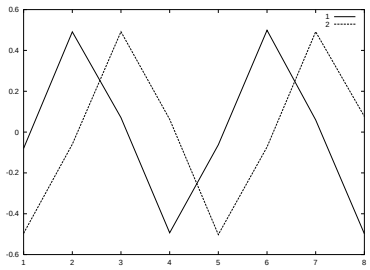
One technique used by meteorologists to analyse the climatology is to look at the projections of the system onto the empirical orthogonal functions, or EOF's. The EOF's are defined as the eigenvectors of the matrix $\mathbf{O^T O}$, where $\mathbf{O}$ is a K by n matrix containing K points distributed on the attractor, and n is the dimension of the model (in this case 8). The eigenvalues indicate the degree of variance attributable to each EOF. Therefore the EOF with highest eigenvalue will have the highest variance.

The first four EOF's for the true system are shown in Figure 5.7. Model EOF's are similar. The first two pairs of EOF's can be viewed as pairs of standing waves around the circle, which are out of phase by a quarter period. Because the indices are cyclic, the starting point is arbitrary, and only the phase and the relative phase difference is important.

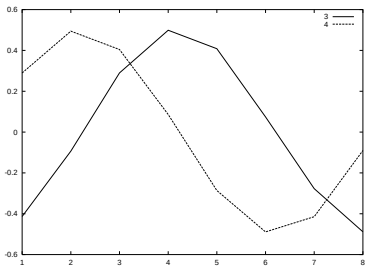
There is a small but significant difference between models in the degree of variance for which each EOF is responsible. The results are summarised below, and are accurate to about 0.1 percent. The linear model is in good agreement, while the stochastic models and the constant model are all out by the same amount, which is about 1 percent for the first two EOF's and 0.5 percent for the next two.

Table of percentage of variance in EOF's.

model	1 and 2	3 and 4
truth	22.6	14.3
constant	21.5	13.8
linear	22.7	14.1
random	21.5	13.8
AR(1)	21.5	13.9



(a) EOF's 1 and 2



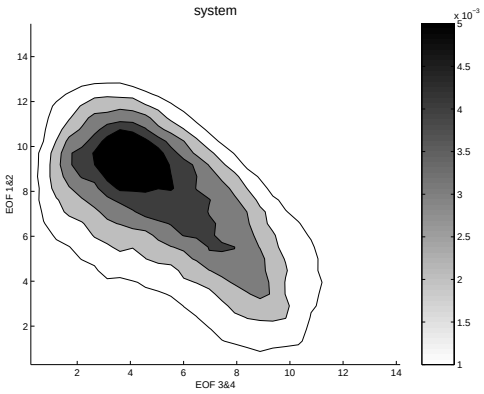
(b) EOF's 3 and 4

Figure 5.7: EOF's for true system. The horizontal axis is the index i of the variables x_i around the circle.

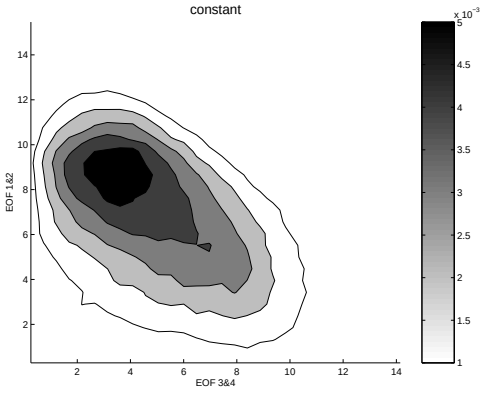
In Chapter 2, we studied the system by looking at orbits of x_1 versus x_2. We now refine this approach by first projecting the orbit onto the higher variance EOF's. This will capture the high variance aspects of the system behaviour in an efficient way. For example, the first two EOF's are responsible for about 45 percent of the variance, while the next two are responsible for another 27 percent. A further improvement is to do a contour plot of the probability density in the EOF's, rather than a simple trace of the orbit.

Because of the rotational symmetry in the systems, the first two EOF's are phase shifted versions of one another, as are the next two. One approach is to project onto EOF's 1 and 3. Another method, which gives slightly clearer figures, is to calculate the projection onto the first two EOF's, get the modulus, project onto the next two EOF's, get the modulus, and plot a histogram of these two numbers. This has been done in Figure 5.8. The difference between the true system and the models is shown in Figure 5.9. Again, results for the stochastic models are similar to the constant model, while the linear model gives the best results. The histograms were generated by calculating 250,000 points, sampled once every 0.2 time units from a long orbit. A test was also performed with only 50,000 points. Results are similar to the long orbit, suggesting that the difference between the true system and the models is a real one, and not a numerical artefact.

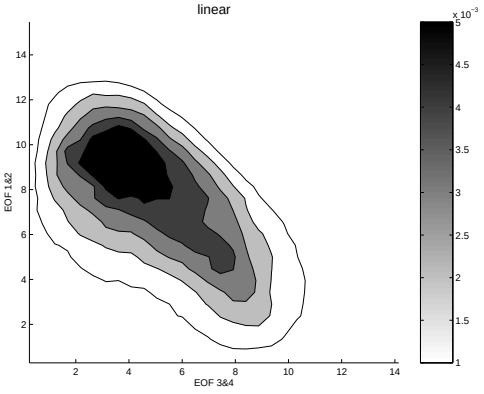
The final method considered for viewing the climatologies was to look at the power spectrum of an orbit's projection onto the first EOF. Figure 5.10 shows the spectra for each model compared with the full system. The linear model again has the best



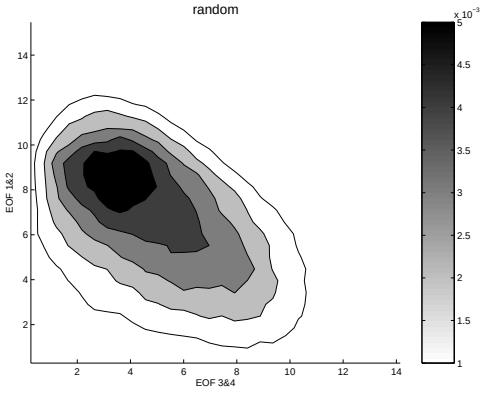
(a) true system



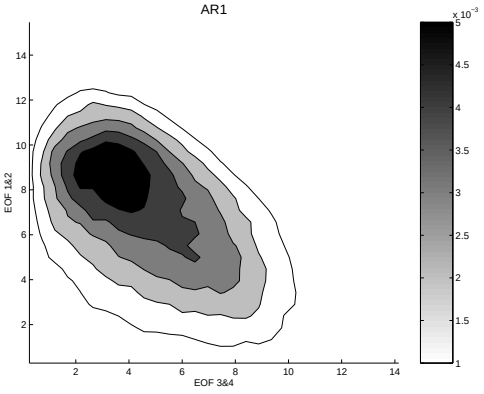
(b) constant model



(c) linear model

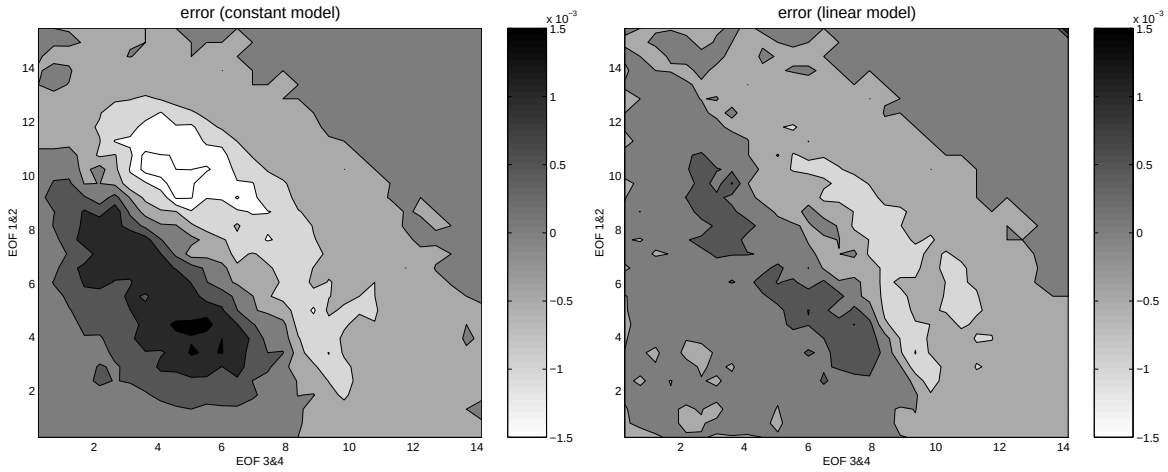


(d) random model



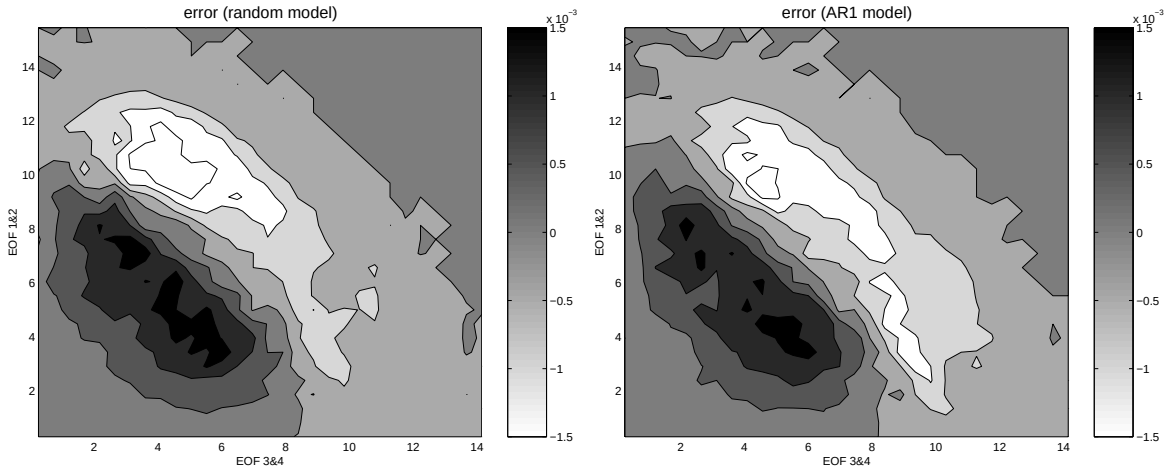
(e) AR(1) model

Figure 5.8: Histogram of modulus of orbit projected onto EOF's 1 and 2 (vertical axis) and EOF's 3 and 4 (horizontal axis) for true system and models.



(a) constant model

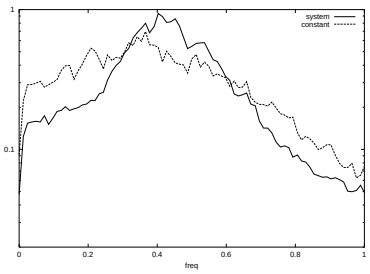
(b) linear model



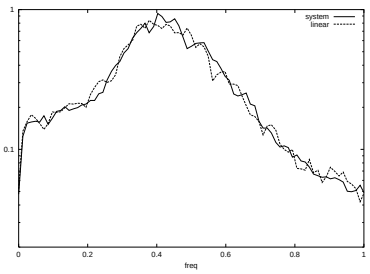
(c) random model

(d) AR(1) model

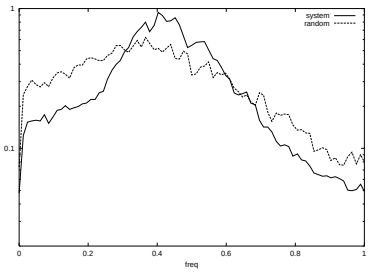
Figure 5.9: Climatology error, expressed by calculating projection of true system and model onto EOF's, then plotting the difference. Vertical axis represents EOF's 1 and 2, horizontal axis EOF's 3 and 4.



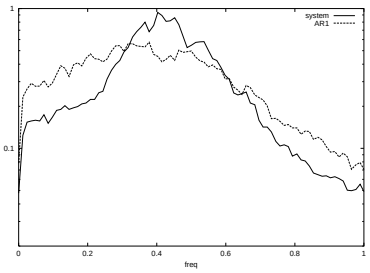
(a) constant model



(b) linear model



(c) random model



(d) AR(1) model

Figure 5.10: Power spectrum of orbit projected onto first EOF for true system and models.

fit.

5.4 Summary

In this chapter we looked first at analytically derived properties of the Lorenz system climatologies. Properties relating the mean to the variance were found, which do not depend on whether the system is in a chaotic or periodic state. It was shown that optimising the mean and variance of the constant model result in a constant forcing which is different from that used for short term prediction, while for the linear model the parameters closely matched those used for shadowing. Therefore the optimisation of short term predictability may, or may not, be the same as optimising for climatology, depending on the particular model/system pair.

Attempts were then made to model the two level climatology, using variants of the one level system. The conclusion appears to be that the linear model, which is best for shadowing, is also best at reproducing the climatology of the full system. This may be related to the fact, seen in Chapter 3, that the linear model reduces low

frequency (and therefore long term) model error. Adding stochastic terms does little to improve the constant model, even if the term is an AR(1) fit to the real errors. This result seems unsurprising, since, in general, we would expect the climatology to be influenced by the average forcing, and less by short term random fluctuations. The random models are actually worse at shadowing than the constant model, which is consistent with the hypothesis that model error is dominated by the forcing error variance.

Chapter 6

Operational weather forecasting models

Up until this point the theory related to model error has been applied to low and medium dimension systems. Since the key results have been determined from the linearised dynamics, which are valid for any model/system pair provided the shadow radius is sufficiently small, the methods are equally applicable to models with very high dimension, such as weather models.

Referring to the linearised dynamics equation (4.13), a fair amount is known about the linear propagator $\mathbf{M}$ for such models, because of the investigations into singular vectors and directions of fastest growth for perturbations in initial conditions. The neglected part of the equation is the drift $\mathbf{d}$, about which very little is known [30].

In this chapter we begin to rectify that imbalance by studying the dynamics of a number of different resolution models in use at ECMWF. In the same way as for the one-layer Lorenz model versus the two-layer system, we will calculate model drift and shadow times, first between the different models, and then between the operational model and the analysis (the closest thing available to the real weather). Finally we discuss methods to improve the forecasts by using information about the likely error.

Before going on to examine the models in detail, though, we first describe some of the chief characteristics of weather models to understand how they work and where error can arise. A summary of results from previous investigations into model error will also hopefully cast some light on how we arrived at the odd situation of knowing more about the error's first order term - the linear propagator - than its zero order term - the drift.

6.1 Causes of model error

Global weather prediction models of the type used at ECMWF and other national or international meteorological centres are extremely complex models, with the order of 10^7 variables. The models are formulated using Galerkin truncations of the fluid dynamic partial differential equations, which describe the evolution of mass, energy, momentum and composition, including terms representing sources and sinks [49, 69]. The models are integrated on some of the fastest computers in the world, with typical speeds of 10^{11} floating point operations per second.

One problem with such huge models is that they are too complex: there are many things that can go wrong, and the size of the models makes it difficult to analyse the errors and determine the cause. Another problem is that they are not complex enough. A typical spatial resolution is about 50 km horizontally and 1 km vertically. The limit to the resolution is determined, not by some scientific choice, but by the capacity of the computer. Therefore any fine-scale processes must be parameterised, in the same way that the forcing in the Lorenz one-level model was used to parameterise the two-level system.

Apart from the finite resolution, there are many other possible causes of model error. The Earth's atmosphere must be one of the hardest modelling tasks that mankind has ever attempted. Anyone who has built a finite element model of a mechanical structure is aware of the potential for unforeseen error (the author's own experience in this regard is with superconducting magnets, where accuracies of parts in 10^4 or better are attainable in theory, but less often in practice [46]). For example, the most important constituent in the atmosphere for the fluid dynamics is water (in its various phases). Unfortunately it is also one of the most difficult to model, and processes to do with the formation and dissipation of clouds need to be modelled parametrically. Other potential sources of error are the interaction between the weather and the earth, such as surface heat fluxes or momentum transfer through tomography; incorrect assessment of radiation due to poor cloud forecasts; and inaccurate model interpolation over data-poor regions, which leads to projection errors. The models currently in use are therefore definitely wrong; the question is, how wrong are they?

6.2 The perfect model assumption

While model error certainly has a role to play in weather forecasting, most investigations into error, at least over the last decade, have concentrated on the initial

condition, and led to the development of ensemble methods. The reason for this emphasis on initial condition isn't clear, though it may have been due in part to the interest in chaotic systems engendered by Lorenz's discovery in 1963 [37] that the atmosphere is chaotic. They say that a scientific revolution takes thirty years to be absorbed: a key paper exactly thirty years on was that by Toth and Kalnay which introduced the breeding vector method for producing ensemble perturbations (a method similar to the singular vector method, but choosing directions which have grown most quickly in the recent past). Here is a quote from that 1993 paper [67]:

> The replacement of single operational forecasts by an ensemble of initial forecasts reflect explicitly the recognition that the atmosphere is a chaotic system. As pointed out by Lorenz (1963), even an infinitesimally small perturbation (as would be produced, for example, by the 'wings of a butterfly') introduced into the state of an atmosphere at a given time will result in an increasingly large change of the evolution of the atmosphere with time, so that after about two or three weeks the trajectories of the perturbed and the original atmosphere would be completely different.
>
> Lorenz's discovery led to ... the realization that many apparently deterministic systems, like the atmosphere and its numerical models, are also chaotic: arbritrarily small perturbations evolve into large differences with time.
>
> If we are willing to run an ensemble of forecasts from slightly perturbed initial conditions, then averaging the ensemble can filter out some of the unpredictable components of the forecast, and the spread among the forecasts should provide some guidance on the reliability of the forecasts.

The stated aims of ensemble forecasting, therefore, are to provide a more accurate forecast, from the mean, and a confidence level, from the spread. The technique will obviously work best when model error is small, and it was felt that models had improved enough over those of the 1960's and 70's that model error had become almost irrelevant. From *Toth et al* [68]:

> In the early years of NWP, forecast errors due to simplified model formulations dominated the total error growth. The traditional perception that forecast errors are primarily due to model errors date back to those early years. By now, however, models have become much more sophisticated and it is the errors that arise due to instabilities in the atmosphere

(even in case of small initial errors) that dominate forecast errors. The recognition of this situation requires a major shift in the perception of NWP.

For the purposes of the calculations, then, the model was assumed to be perfect:

> In this paper we will assume that our numerical model is essentially perfect ... As *Reynolds et al.* (1993) have showed, the forecast error in the extra-tropics is dominated by the error originating from the unstable growth of initial errors, and not by model deficiencies.

A similar assumption was made for the ECMWF ensemble prediction scheme (EPS) in *Buizza et al* [7]:

> From its inception, the EPS has been based on the premise that medium-range forecast errors are predominately associated with uncertainties in initial conditions.

These are statements of what is known as the 'perfect model assumption', and it underlies most of the development of ensemble techniques based on perturbations of the initial condition (other techniques perturb the model as well, and we will come to them below). The assumption appears in different forms whenever such techniques are discussed. Usually it is posed only as a working assumption, but sometimes it is expressed almost as a statement of fact. From *Buizza et al* [7]:

> ... the hypothesis of the dominant role of initial uncertainties is certainly valid in the early forecast range ...

The same paper goes on to say that the perfect model assumption doesn't always hold: in fact,

> ... model errors can become as important as initial condition uncertainties in the medium forecast range.

The belief that model error is only important for longer forecast times could be dubbed the 'nearly perfect model' assumption. It is effectively saying that model error is initially small, causing a perturbation which is then amplified by 'flow-dependent instabilities of the chaotic climate attractor' [49]. It actually refers, not to model error itself, but to the displacement error which is initiated by a small perturbation. The drift for such a nearly perfect model would still be small, and the model error index low.

Papers quoted in support of the nearly perfect model assumption include *Downton et al* [21] and *Richardson et al* [57]. The first paper noted that different models often gave different results, and set out to discover whether this was due to the models themselves or the fact that they were initiated from different analyses. It examined in detail six cases during the autumn/winter of 1985/86 where the UK Meteorological Office (UKMO) operational forecast disagreed significantly with the ECMWF forecast. The approach used was to run the ECMWF forecast from the interpolated UKMO analysis, and vice versa. In most instances, it seemed that the models produced similar forecasts providing they were initiated with the same analysis. 'Similar' here was not so much in terms of RMS fields, but in various qualitative properties of the 500 hPa heights, such as development of lows, highs, troughs, ridges and so on. The emphasis was on errors after five days.

The second paper studied the relative effects of using different analyses and different models for 25 cases in the winter/spring 1996/97 period. In each case, a perturbation was made to the ECMWF analysis approximating the difference between it and the UKMO analysis. Forecasts with the ECMWF model from this analysis were compared with the EPS control forecast, to determine analysis differences, and with the UKMO model to determine model differences. It was found that the effect of using a different analysis was 'substantially greater' than that of using different models, as measured by RMS errors in the 500 hPa height. At day 5, model differences were found to account for only 25 percent in the Northern and 15 percent in the Southern hemispheres, though this was considered an upper bound since it contained also errors in the representation of the UKMO analysis.

Another paper taking a similar approach was *Harrison et al* [28]. It noted that the 'the weight of evidence appears to suggest that analysis differences are the more critical in controlling forecast divergence', but also that 'the overall contribution of model and analysis dependencies to the divergence of forecasts have not been fully elucidated and further evaluation is desirable'. The paper went on to examine two case studies comparing the ECMWF T63 model with the UKMO Unified Model at comparable resolution. Four ensembles, each with 33 members, were constructed using all permutations of models and analyses. The initial perturbations for the ensemble were generated from ECMWF singular vectors. It was found that 'significant differences between all four ensemble sets were found in each case-study', where the emphasis was again on the medium range (5 days). The writers concluded that it might be preferable to include both models, so that the ensemble contained, not only different initial conditions, but different models: a multi-model ensemble.

6.3 Multi-model ensembles

The use of ensembles comprised of different models is an extension of the technique of initial condition ensembles. If different models give different results, then incorporating all the models in the ensemble should take that effect into account.

The concept of using a combination of models to provide a forecast is actually not very new [28]. Meteorologists have access to products from all the major weather centres, and have always used their experience of model performance and atmospheric behaviour to choose the model which seems most applicable. (The final forecast, though, was usually based on a single model.) This tradition is carried on by companies like Risk Management Solutions, which, when predicting hurricane tracks for insurance companies, use a combination of available models and historical data, or by the Fleet Numerical Meteorological Centre which regularly produces a 72 hour forecast which is the mean of the forecast from several centres [30]. It has been found that the difference between forecasts can be a good predictor of forecast skill [75].

Multi-model ensembles come in two flavours. The first is to use a combination of models from different centres. The second approach is to perturb the parameters of a single model, analogous to the randomly perturbed Lorenz models of Chapter 5. This can be viewed as an attempt to add a perturbation to the model which captures the likely nature and extent of model error.

In the past, attempts were made to account for model error by adding random perturbations to the entire model, rather than particular parameters. *Philips* [51] suggested using a white noise description for the model error. *Bennet and Budgell* [3] claimed that the tail of the spectrum should be constrained, so as to be consistent with regularity of model solutions. Such a model error description was used by *Cohn and Parrish* [14], who adjusted the length scale of the model error to the length scale used in the National Centers for Environmental Prediction regional analysis system. *Dee* [19] investigated the estimation of model error parameters using an analysis of innovations.

As *Houtekamer et al.* [30] pointed out, it wasn't clear whether such an idealised model error had the same characteristics as the real error, or whether their addition would aid an ensemble system. The whole principle behind ensembles, to put things rather bluntly, is that we add garbage to the solution in the hope that the ensemble of perturbed solutions will give an improved picture of where truth lies; but we at least want to add the right kind of garbage.

A more sophisticated scheme is to actually perturb those physical parameters in the model which are felt to have a degree of uncertainty. When combined with initial condition errors, this means that every uncertain variable is perturbed - a method dubbed the system simulation experiment, or SSE [73, 74, 31, 50]. *Houtekamer et al.* [30] incorporated into their SSE different parametrizations of a number of model areas. Particular attention was focused on the areas of orography and deep convection, which were thought to be particularly deficient [56], and the treatment was also extended to horizontal diffusion, radiation, and gravity wave drag.

A similar approach was taken by *Buizza et al.* [8] at ECMWF, focussing on the parametrization of the diabatic tendency. The diabatic forcing term for each grid point was chosen randomly from a prescribed range. The forcing term also varied with time: it was noted that 'even if the parametrized and actual diabatic heating fields agree on average (i.e. over many time steps) at the chosen grid point, there must inevitably be some standard deviation in the time-step by time-step difference between observed and modelled heating'. The scheme was therefore similar in principle to that in Chapter 5 where random perturbations were assigned to the forcing of the one-level system.

The SSE approach doesn't make the perfect, or nearly perfect, model assumption; but it does assume that the models can be corrected, or at least substantially improved, by varying the parameters. This is what we might call the 'structurally perfect assumption'.

6.4 Problems with the ensemble approach

The use of ensembles has become quite broadly accepted in the meteorological community, and ensemble calculations have been executed routinely at ECMWF since 1992. They seem well adapted to the problem of addressing initial condition error, because the error in that case is in an unknown direction, but is (probably) within a certain magnitude. It is also possible to choose the perturbations which grow the fastest, and therefore estimate the likely spread of forecasts.

Referring to Figure 4.3, though, the usefulness of an ensemble forecast, in terms of the mean and, to a lesser degree, the spread, will depend on the model error. If model error is high, then the ensemble mean may be no more accurate than a single control forecast. As stated in [68]:

> The ensemble strategy will work only if the models are good enough that model-related errors do not dominate the final error fields.

Therefore we are brought back to the perfect, or at least the near-perfect, model assumption. Unfortunately, evidence for the near-perfect model assumption is mostly circumstantial, and is based on the observation that models from different weather centres produce similar results. In fact, this is hardly surprising, because of the process by which the models are built: the meteorologists all read the same books and attend the same conferences, so when an advance is made in one area it is adopted fairly quickly by the others. It is notable that the one paper which found a distinct difference between forecasts [29] atributed it to a problem with one of the models, which was eventually corrected. In practice, it has been found that ensembles consistently underestimate the spread, and that the mean is no better than the control [6, 70]; characteristics which are both compatible with high model error.

The use of ensemble techniques to understand model error is even more problematic than its use for initial condition error, though for different reasons. It has been stressed in this thesis that model error and initial condition error are different entitities; therefore they demand different approaches. Ensemble methods are at least theoretically suited to initial condition error, since the true initial condition is assumed to lie within some ball of radius corresponding to the analysis error. Model error, in contrast, is more difficult to address. It could be simply impossible to construct a suitable set of equations [63]. Perturbing model coefficients won't help if the model is structurally deficient. With initial conditions, we know the type, if not the direction, of the garbage that we want to add; with the model, we can make educated guesses about uncertainty of certain parameters, but have no guarantee that we have addressed the real source of error.

The most important difference between ensembles of initial conditions and of models, though, is that we can choose those initial conditions which, out of all possible perturbations, will create the largest error, but we can never do the same for models. An ensemble of models from different centres is a very poor sample of model space; and a stochastically perturbed model will not represent the real errors if the model is not structurally perfect. Indeed, there may be no accessible set of equations that perfectly mimic the dynamics of the system [63].

The focus here will therefore be, not on creating ensembles, but on measuring model error and determining its characteristics. (After all, the constant model was improved, not by taking an ensemble of models with different constant forcings, but by looking at how a simple parameterization could reduce error: the linear model.) This is not to say that the ensemble approach isn't adaptable to model error; rather that, as for initial conditions, if we intend to perturb our model by adding garbage to

it, we need to have a very good idea which kind of garbage we should add. We begin that investigation by comparing ECMWF models of different resolution.

6.5 Error between models of different resolution

6.5.1 The range of ECMWF models

The forecast models at ECMWF have undergone a number of changes in resolution since operational forecasting began over 20 years ago. Here is a brief summary of their historical development.

In April 1983, a 15-level finite difference model, based on a regular longitude/latitude grid, was replaced by a T63 16-level spectra model, with the extra level in the planetary boundary layer. Spectra models exploit the spherical geometry of the globe by using a truncated series of spherical harmonics (products of sinusoidal functions in the zonal directions and Legendre functions in the meredinial direction). A T63 model truncates the series above order 63.

Further improvements followed. In May 1985, the horizontal resolution increased to T106. A year later, vertical resolution became 19 levels, with the three extra levels in the stratosphere. In September 1991, horizontal resolution became 213 and vertical 31, with layer spacing reduced by a factor of about two. In April 1998, spectral resolution became T319, but used a 'linear-grid' option in which the computational grid remained the same (about 60 km) as for the old T213. In March 1999, vertical layers increased to 50, with a layer spacing of about 1.5 km over most of the stratosphere.

The EPS scheme, meanwhile, was initiated in 1992 with a T63L19 model (the 19 refers to the vertical levels, the 'L' refers to the linear grid option). In December 1996 the resolution was increased to T159L31. The singular vectors, which are expensive to compute, are based on a lower resolution T42L31 model. There also exists a tangent linear version of T42L31, which linearises the adiabatic component of the model so that an adjoint can be constructed [10].

There therefore exists a fairly extensive suite of models from which to choose. The resolution experiments in this thesis were based on the lower resolution T42L31 and T63L31 models, using T159L31 as 'truth'. These models have the advantage of keeping the same number of vertical levels, so interpolation isn't required over that scale. Also an adjoint exists for T42L31, which is required for the computation of shadow orbits.

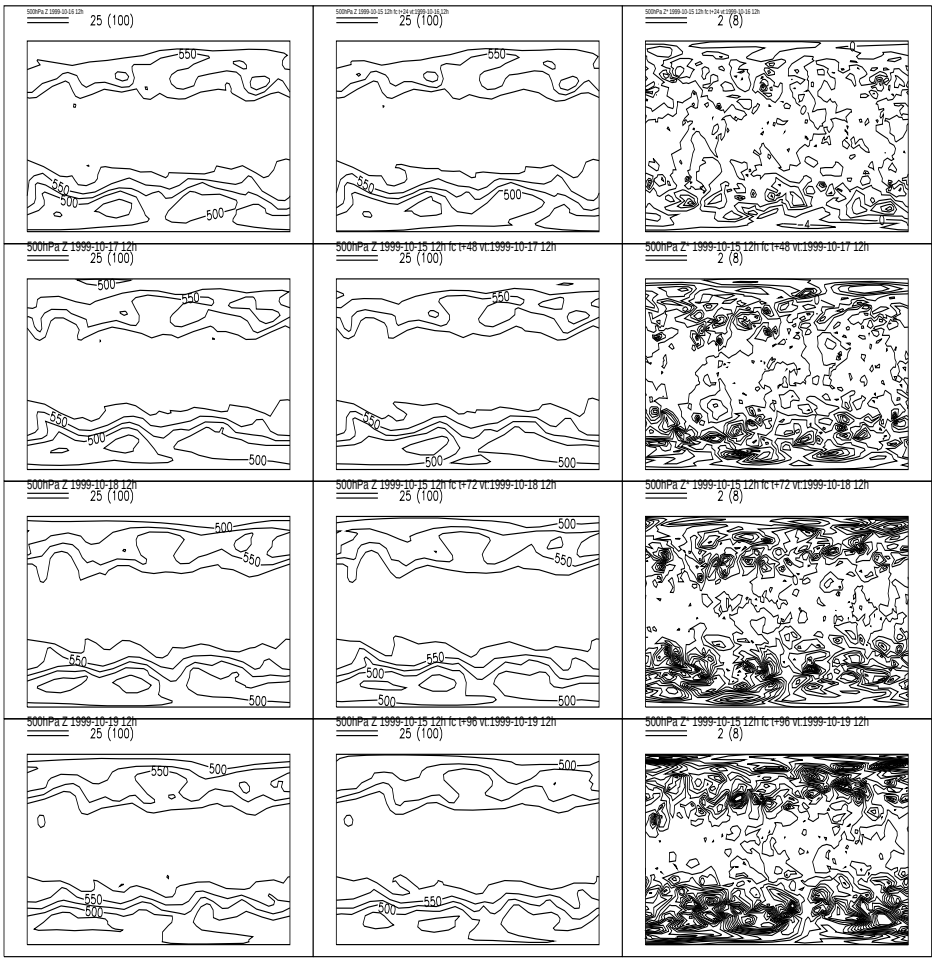


Figure 6.1: Model errors in the 500 hPa height norm. Left column shows analysis heights for days 1 to 4, middle column shows predicted heights, right column shows errors. Contour interval is 25 for the heights and 2 for the errors.

6.5.2 The energy metric

In order to calculate RMS errors, it is first necessary to choose a metric. One possible choice, which is used commonly by meteorologists, is the 500 hPa height. Figure 6.1 shows how error grows in this metric over a typical four day forecast. After a couple of days the difference between the analysed heights (left column) and predicted heights (middle column) has become noticeable. The error (right column) appears to have a finer structure than the height fields themselves.

The 500 hPa height metric may be useful for meteorological interpretation of the weather, but it is less suitable for shadow calculations since it only takes into account a limited set of atmospheric variables, namely the geopotential at one level. The situation would be the same as doing shadow computations for the 8D Lorenz system

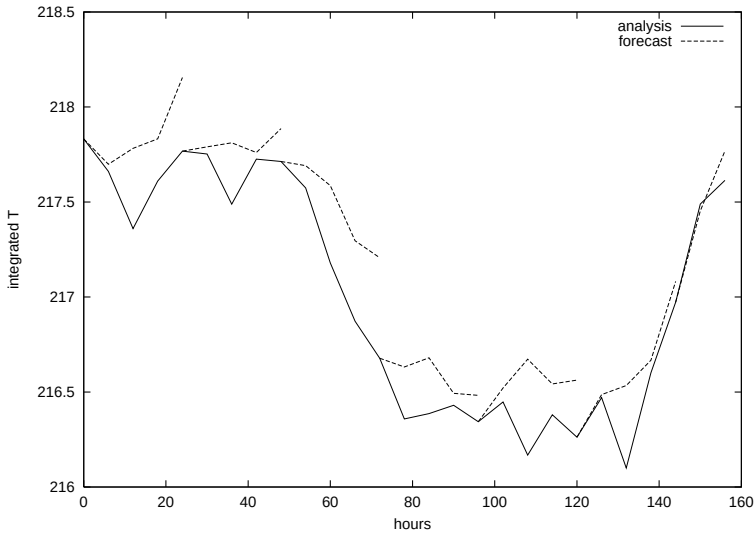


Figure 6.2: Errors for temperature T integrated over Europe, for a typical week.

with a single variable; as mentioned in Chapter 4, it may be possible to shadow for example x_1, but only by introducing huge distortions into the other variables. In weather terms, the pressure might be fine in the middle atmosphere, but completely wrong at ground level.

A more complete measure of the atmospheric state is given by a simplified version of the total energy metric. The vector used to describe the atmospheric state $\mathbf{x}$ at a particular time is

$$\mathbf{x} = (u, v, T) \tag{6.1}$$

where u and v are the zonal and meridional wind components, and T is the temperature. Figure 6.2 shows errors for one of these variables, the temperature T, integrated over Europe. We define the energy norm to be

$$\langle \mathbf{x}, \mathbf{x} \rangle = 1/2 \int_0^1 \int \int_\Sigma (u^2 + v^2 + (C_p/T_r)T^2) d\Sigma (\partial p_r/\partial \eta) d\eta. \tag{6.2}$$

The energy norm equals the sum of the kinetic energy of the wind error and the potential energy stored in the temperature error, and is the same as the total energy norm but with the relatively small surface pressure component omitted. T_r is a reference temperature, p_r a reference pressure, and C_p the specific heat at constant pressure for dry air. Σ is the horizontal domain, taken here to be northwards of 30 degrees, and η the vertical coordinate. Details are in [9].

The energy norm appears more complicated than the standard Euclidean norm, but it can be viewed numerically as a weighted sum of squares of (u, v, T) errors over a

finite element grid. Quantities such as singular vectors can be calculated in this norm just as they were for the Lorenz models in the Euclidean norm, with the difference that the matrix transpose of the linear propagator becomes an adjoint model [20], calculated with respect to the total energy inner product.

6.5.3 Forecast errors

The upper panel of Figure 6.3 shows RMS errors in the energy metric at five different five-day forecasts starting at different dates. For comparison, the typical analysis variance, which is used to determine perturbation size in ensemble forecasts, is about 45 units on this scale. The model trajectories diverge from the true system (TL159) at a fairly constant rate, with T63 consistently performing better than T42 as one would expect. The lower panel shows the ratio of T42 errors to T63 errors. What is surprising is the uniformity of the results; there is little evidence of fickle sensitivity to initial conditions for these five starting dates. Nor does growth appear to be exponential in shape, which is the typical characteristic of initial condition error. Rather, the curvature is negative, so rate of growth actually decreases with time.

Interpretation of the forecast results is complicated by the ambiguity in the starting points. The forecast errors are not all zero at time zero because of the truncation operator which translates TL159 fields to T42 or T63 fields. For the forecasts considered here, the mismatch is about 40 energy units for T42 and 25 for T63. This still allows the possibility that truncation error is responsible for the divergence of forecasts: a small initial error is magnified by the nonlinear dynamics, and the problem would not then be of model error, but of sensitivity to initial conditions. In that case it would be possible to shadow for extremely long times, since the negligible model error could be counteracted by an appropriate choice of initial displacement.

6.5.4 Calculation of the drift

From the forecast alone, we can't separate out the effects of model error and initial condition error, since as soon as the model diverges from the true orbit initial condition error begins to grow. We therefore calculate the drift. A number of short, twelve hour forecasts were made with T42 and T63, starting at twelve hour intervals along the TL159 forecast, and the results integrated numerically to give the drift. Figure 6.4 shows how the drift accumulates with time for T42 and T63. The ratio of the drifts is also shown in the lower panel of Figure 6.3; as for the forecasts, it is nearly constant at 1.4 over the forecast time.

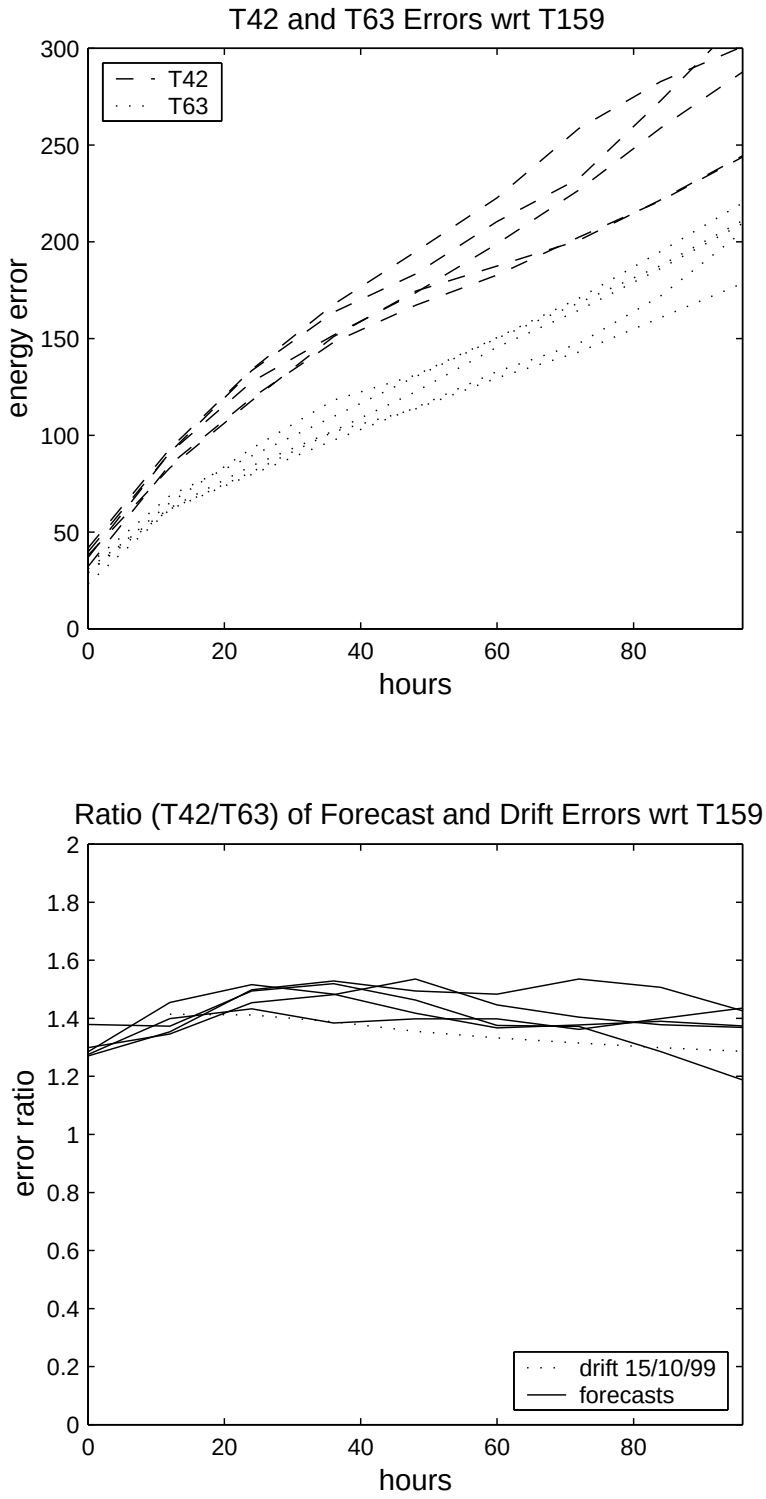


Figure 6.3: The upper panel shows T42 (dashed) and T63 (dotted) forecast errors at starting dates 15/10/99,15/11/99,22/12/99,15/1/00,15/2/00. Errors are computed in the energy norm, relative to TL159. The lower panel shows the ratio of the magnitudes of T42 and T63 errors, where each is again taken relative to TL159. Also shown is the ratio of the drift magnitudes.

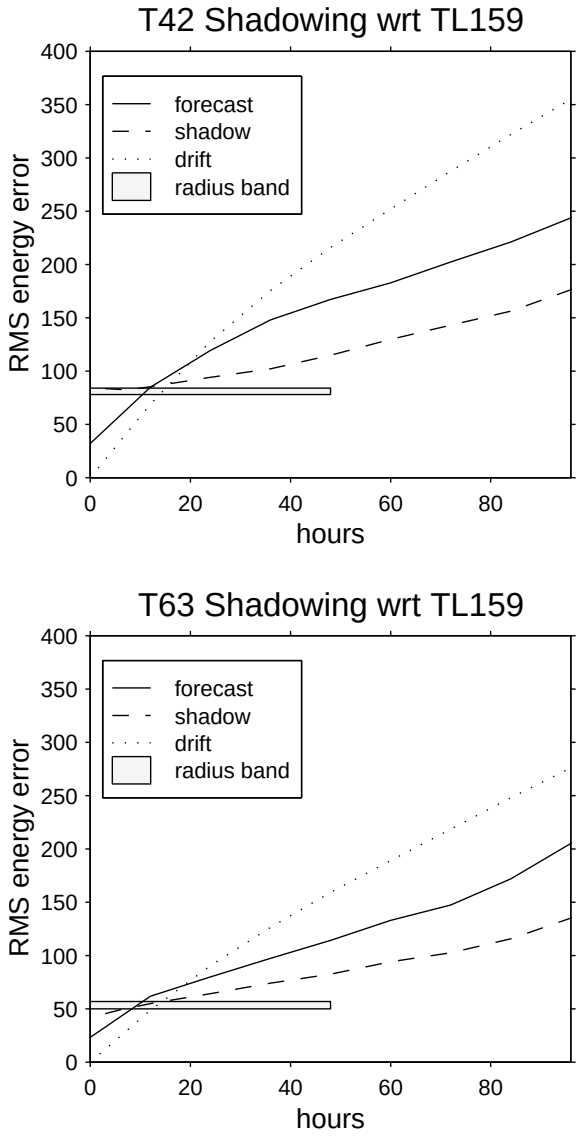


Figure 6.4: The upper panel shows a plot of T42 errors with respect to TL159 for starting date 15/10/99. Solid line is the forecast, dashed line is the shadow trajectory which minimised error at 48 hours, dotted line is the drift. The uncorrected drift is in places larger than the forecast error, due to a spin-up error which is probably caused by truncation error. The two estimates for the lower bound on shadow radius, computed using estimates of the drift, are shown by the shaded region. Errors are computed in the energy norm, relative to TL159. The lower panel shows the same for T63.

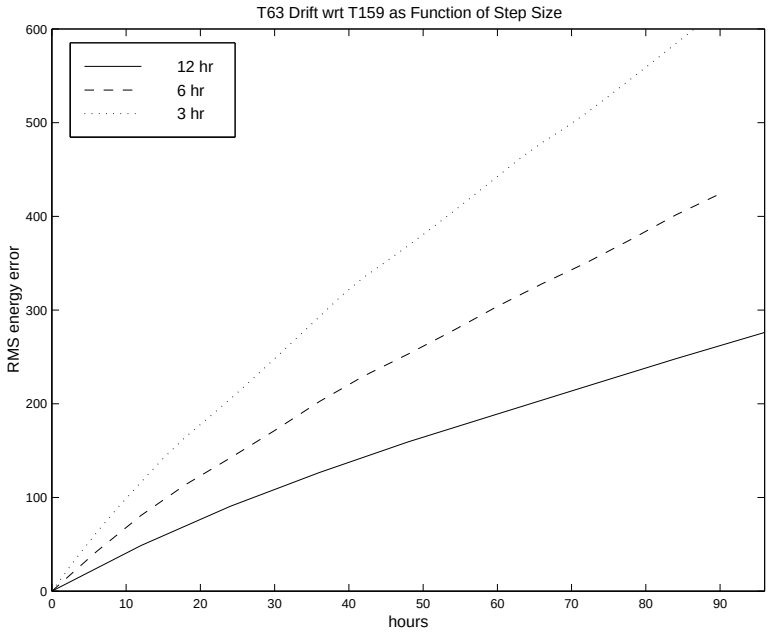


Figure 6.5: Comparison of drift for T63 vs TL159, computed with step sizes of 3, 6 and 12 hours.

If model error were truly negligible, then we would expect the drift to be smaller than the forecast error, since the tendency difference is always calculated on the TL159 orbit where displacement error is minimal. In fact, the magnitude of the drift is close to the magnitude of the forecast error. At times it is even larger: the reason appears to be that there is an initial spin-up error associated with each short forecast, which may be an artefact due to the initial truncation error. Tests with different time steps show that the drift calculation is dependent on step size. For example, Figure 6.5 shows the drift for T63 versus T159 for step lengths of 3, 6 and 12 hours. The results show a marked discrepancy between the different step sizes, with the 6 hour step length giving a drift about 50 percent higher than the 12 hour number, and the 3 hour step drift higher by the same amount again.

This spin-up error, whose signature is a lack of scale invariance in the drift calculation, appears to be an unavoidable feature of the inter-model comparisons. It means that a portion of the drift is due to spin-up effects, and the calculated drift is artificially high. We will therefore attempt to deal with it using two methods, and note that the same problem does not occur in the next section, where the operational forecast is compared with the analysis, and the drift calculation is seen to scale with time step.

The first approach is to reduce the drift by the errors incurred during each small

forecast. For example, the drift calculation over 48 hours involves four short forecasts, so compared to a normal forecast there are three additional 20 unit errors. We could therefore correct the 48 hour drifts by 60 units, giving a drift of 156 units for T42 and 99 units for T63.

The second approach is similar to that adopted for the system of Chapter 4 with periodic orbits on a circle, shown in Figure 4.25. In that example we circumvented the problem of spin-up error by linearising the error around the model control rather than truth. The control error therefore serves as a proxy for drift. This method gives a result for 48 hour drift of 168 units for T42, and 114 units for T63.

Given an estimate of the drift, we can determine its effect on shadow times from the shadow law. The law has been shown to work for a range of low and medium dimension models, but does it apply to full weather models, in all their complexity? The conditions for the law to hold are that the model must be locally dissipative, and the shadow times must be sufficiently short so that the linearised dynamics are valid for shadow orbits. The first condition surely holds. For the second condition, it is known that the model becomes nonlinear within a day or so [23]. However, if the linearisation is done about truth, and only shadow orbits are considered, then the linearisation will hold for longer times because the displacement is limited by the shadow radius (the error is $O(\|r^2\|)$). It therefore seems reasonable to expect that the shadow law will apply.

The shadow law states, in an RMS sense, that the minimum shadow radius for a set drift should be equal to half the drift. As mentioned above, there are two methods for estimating the drift given the large truncation errors. If we correct the drift by subtracting the initial errors, we obtain an expected minimum shadow radius of about 78 units for T42, and 50 units for T63. If we use instead the control error as a proxy for drift, we find a radius of 84 units for T42, and 57 units for T63. The results of the two different methods are shown by the shaded region in Figure 6.4.

To summarise, it appears that both T42 and T63 have significant model error relative to TL159. Drift varies with step size, but is highest for the shorter step, implying that it is not due to initial condition, and even when spin-up effects are subtracted it still accounts for most of the total forecast error. Estimating the minimum achievable shadow radius at a specified time of 48 hours from the drift gives for T63 a radius of around 50-57 units, close to the analysis variance, and for the T42 model around 78-84 units.

6.5.5 Shadowing

The above results are approximate, and hampered somewhat both by the initial truncation error and the (perhaps related) fact that drift increases with lower step size. It might still be possible to construct some scenario where the drift is due, not to the model, but to the truncation error being consistently in the same, rapidly growing direction (for example, the direction of the leading singular vector). Were this the case, and model error was in fact small, then it should be possible to find shadow orbits which shadow for two days with a much smaller shadow radius than given above. The only way to test this is to look for actual shadow orbits.

An ECMWF algorithm, originally designed to find optimal perturbations to offset forecast errors [55], was employed to search for such orbits. The method, based on that used in 4DVAR data assimilation [36, 15], uses a Newton step minimisation procedure, and is similar to the 'pinch' method described in Chapter 4, with the difference that only the final energy error at 48 hours is minimised rather than the sum of the initial and final. The method is therefore not perfect, but since the optimisation time is quite short and the model reasonably linear over that period [34, 54, 71], it should produce satisfactory results. The gradient of the cost function is determined by use of the T42 adjoint, which will only be an approximation to the true adjoint for T63. A total of 50 iterations were performed.

Figure 6.4 shows the orbits which the program found for T42 and T63. At time two days, the minimised error of the T42 forecast is 114 units, while for T63 it is 82 units.

The optimisation procedure gradually increases the initial error while it decreases the final error, and since for both T42 and T63 the initial condition error is still smaller than the final error it appears that the process isn't quite complete. Convergence was limited by computer time and the efficiency of the algorithm, but it seems reasonable that trajectories could be found which had the same initial and final displacements equal to the average of the two. For T42, the average of initial and final is about 99 units, while for T63 it is about 63.5. These are still above the lower bound estimate shown in Figure 6.4.

Orbits with longer shadowing times may exist, as the method used isn't optimal; however, this concern would be more of an issue for longer shadow times where the model was less linear. The fact that shadow behaviour is consistent with results from the drift implies that the drift does not overestimate the contribution of model error, and confirms that both T42 and T63 have significant model error relative to T159. A typical shadow tolerance for operational purposes would be the same as the analysis

variance, i.e. 45 units. It therefore appears that both models fail to shadow at that radius while they are still in what would be considered a linear regime.

6.5.6 Ensemble calculation

If model error is significant compared to displacement error, then the behaviour of initial condition ensembles will be affected, as discussed in Chapter 4 for the Lorenz systems. A suitable guide to the possible impact is the model error index $M2(\tau)$ given in equation 4.68, which compares the drift with the growth of the leading singular vector.

Leading singular vectors normally grow in the energy norm by a factor of about 20 over 48 hours (the energy itself may increase by 400, so RMS errors will be the square root of that). From equation 4.68, the model error index for T42 will be

$$M2(\tau) = \frac{1}{\sigma_1} = 0.05. \tag{6.3}$$

To compare with the Lorenz systems, this is closer to the constant than the linear model. Since model error had a significant effect on ensembles for the Lorenz constant model, we can expect it to do the same here.

To test the effect, an ensemble was formed for the T42 model by adding scaled displacements, equal in magnitude to the analysis variance, in the subspace of the leading 25 singular vectors. A total of 50 initial conditions were generated from the positive and negative perturbations. Figure 6.6 shows the resulting errors with repect to both the T42 control (upper panel) and the T159 control (middle panel). The lower panel was generated by summing the errors in the upper panel with the control errors, with the assumption that they are orthogonal. If the upper panel errors are caused by initial condition error, while the control errors are primarily caused by the model, then, because of the high dimension of the space, it is safe to assume they are orthogonal. The agreement between the centre and lower panels confirms this.

Figure 6.7 compares the two only leading singular vector perturbations of the weather model with those of the Lorenz system (see also Figure 4.3). It illustrates some of the key similarities and differences between the weather models and the Lorenz systems, and between high and low dimension systems in general. The upper panels follow quite similar curves for either system. In the lower left panel, the T42 control has a negative curvature, unlike the Lorenz system. We will see later that this curvature is characteristic of model error in high dimension systems. The biggest difference is that in the lower dimension Lorenz system the model error significantly

affects the error for the singular vector perturbations, increasing it in one case and decreasing it in the other, while in the weather model the errors appear largely unaffected. Also the lower panel for the Lorenz system, which assumed the errors were orthogonal, is a less good fit to the real errors than for the weather model. The explanation lies in the dimension of the space: in a low dimension system the initial condition and model errors have a much higher probability of interacting than in a high dimension space, where we can assume they are orthogonal.

The conclusion is that, for weather models, we may not only assume that initial condition error and model error are uncorrelated, as demanded by the shadow law, but also that they are nearly orthogonal. As a result, no ensemble member manages to reduce model error. The ensemble mean, also shown in the centre panel, closely tracks the perturbed forecasts; this is not surprising, since, if the model is in a linear regime, the positive and negative perturbations will tend to cancel in the average. The implication, at least for this particular day, is that running an ensemble of T42 forecasts wouldn't be much more informative than a single deterministic forecast.

6.6 The ECMWF operational model

Of course, T42 and T63 haven't been used operationally for some time; the current standard at ECMWF is TL319. Also, we want to shadow the real weather, not TL159. Our real interest is therefore to compare TL319 with the analysis (our closest approximation to the real weather).

Previous calculations of model error in this thesis have primarily been with respect to a true system which is described by differential equations. For the Lorenz model, the true system was the two-level equations. For the experiments above, the true system was TL159. As mentioned in Chapter 4, though, the same model error techniques can be applied equally well to comparisons between a forecast and an interpolated set of observations or analysed trajectory.

To further illustrate this point, one goal of measuring model error is to estimate shadow times. Shadow orbits can be found explicitly, as in the previous section, by a code which minimises the RMS error at a specified future time (here 48 hours). However the program doesn't distinguish whether the target trajectory (specifically, the desired value at 48 hours) comes from an analysis or a model; it can be used with either. In the same way, our estimates of shadow times, derived from the drift vector, are applicable whether the drift is calculated relative to a model trajectory or an analysed trajectory.

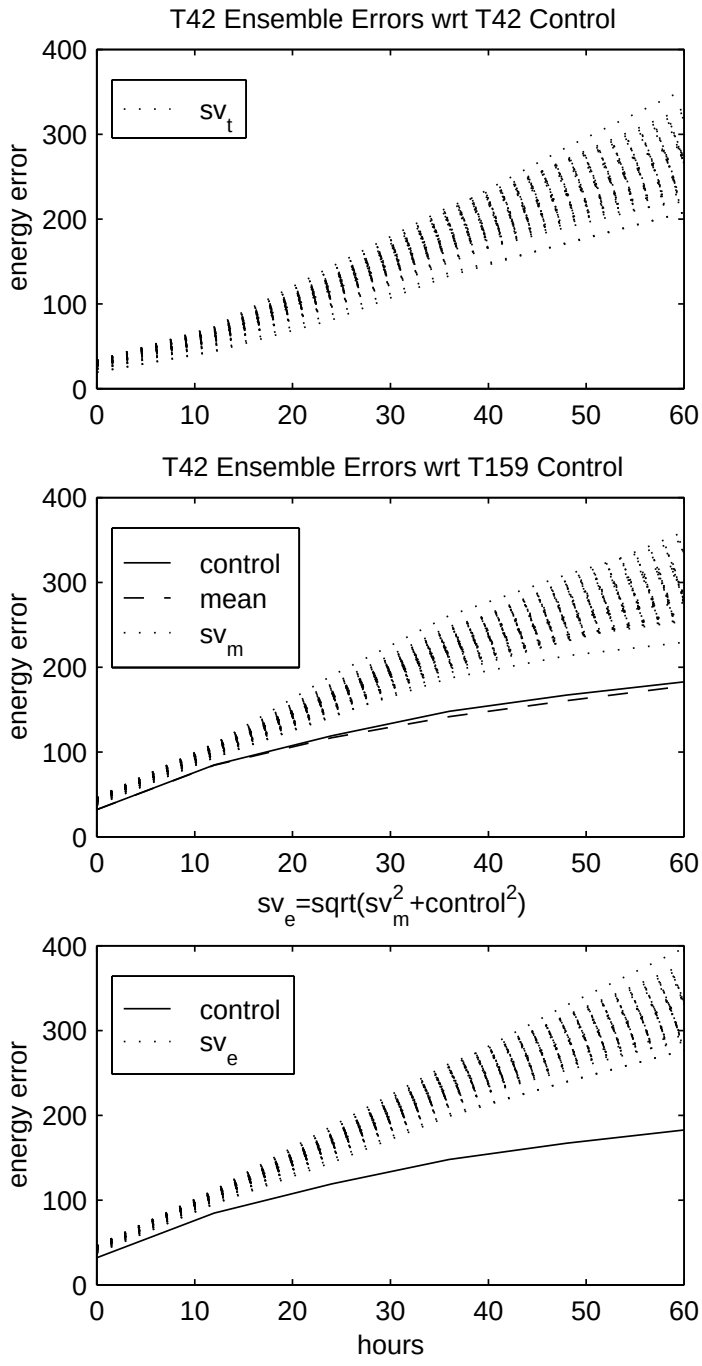


Figure 6.6: Upper panel shows plot of errors wrt T42 control for a T42 ensemble generated from leading singular vectors at 1999/10/15, 12 hours GMT. Centre panel shows errors with respect to a TL159 control for the T42 forecast (solid line), the ensemble (dotted) and the ensemble mean (dashed). The lower panel shows the errors which would occur if the error vectors in the upper panel are added to the control error, assuming orthogonality. It can be compared with the centre panel.

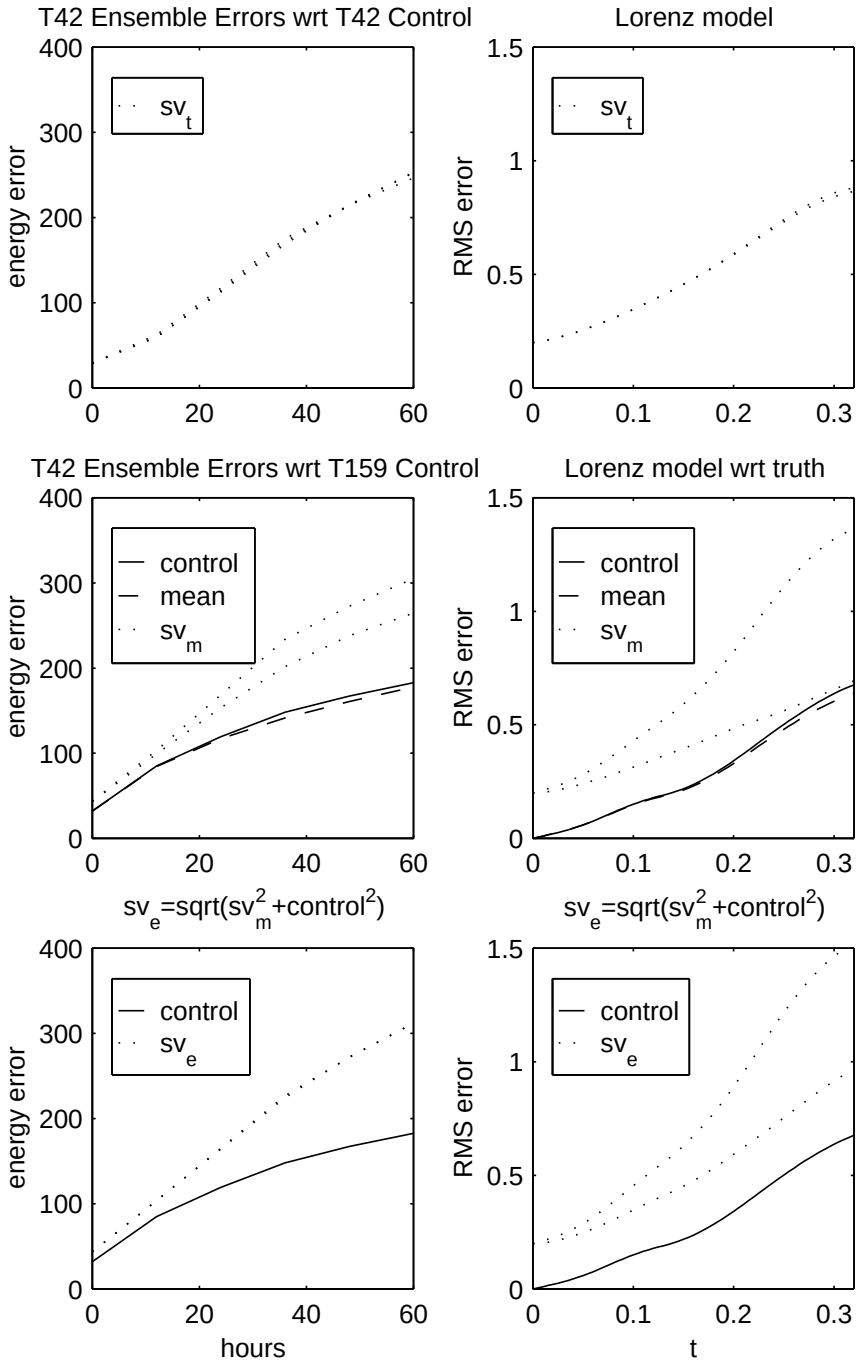


Figure 6.7: The left hand panels are as for Figure 6.6, but only the first two ensemble members are shown. The right hand panel shows the corresponding figures for the Lorenz system, from Figure 4.3. Solid line is the control, dashed is the mean, dotted are the two ensemble members.

We will therefore use the analysis, not as a proxy for truth, but as the true orbit $\tilde{\mathbf{x}}(\mathbf{t})$ itself (in, of course, a discrete form). An advantage when comparing the operational forecast with the analysis is that there is no truncation error, since both are at the same resolution. And because the forecast is always initiated at the analysis for that day, the two have the same initial condition. Therefore we don't have the problem, encountered in Figure 6.4, that truncation error contributes to forecast error.

It also appears that the spin-up error which occurred with the inter-model comparisons is not an issue when the truncation error is removed. Figure 6.8 shows error growth for a single forecast, along with the drift. The drift was calculated by summing short 6 hour forecasts for the first day to capture the fast initial growth, followed by 24 hour forecasts for days 2 and 3. Unlike with the inter-model comparisons, the calculation is not sensitive to step size, so summing 6, 12 or 24 hour forecasts give similar results. This can be seen, for example, by the fact that the drift over one day, calculated by summing four 6 hour forecasts, agrees closely with the forecast error at 24 hours, which would be the value of the drift if a 24 hour step were used. Spin-up errors, whose signature is a strong time-step dependence in the drift calculation, are not present to a noticeable extent.

The drift closely tracks the forecast error out to three days, in a manner compatible with high model error, and the initial slope is about three times greater than that for T63 versus TL159. The most striking feature of the curve, though, is its pronounced negative curvature, which is hard to reconcile with the exponential-on-average [65] growth expected from displacement error.

The shape of the curve makes more sense when we examine the nature of the short forecast errors which make up the drift calculation. Figure 6.9 shows histograms of the cosine of the enclosed angle of the 24 hour drift vectors, for consecutive and randomly chosen days over a hundred day period. The mean for the consecutive days is 0.081, which is a significant correlation considering the dimension of the space. Note that the distribution for consecutive days in the upper panel is shifted significantly to the right of the distribution for random pairs of days, implying that drift is persistent on a timescale of one day. (We return to discuss the fact that neither are mean zero below.)

One might expect that this correlation would increase for shorter times, but the reality is less straightforward. The left two panels of Figure 6.10 show the magnitudes and cosine angles for drift vectors calculated every 6 hours instead of daily. The magnitudes fall into two camps: those to the left of the dashed line are initiated at 0 or 12 hours GMT, while those to the right are initiated at 6 or 18 hours GMT.

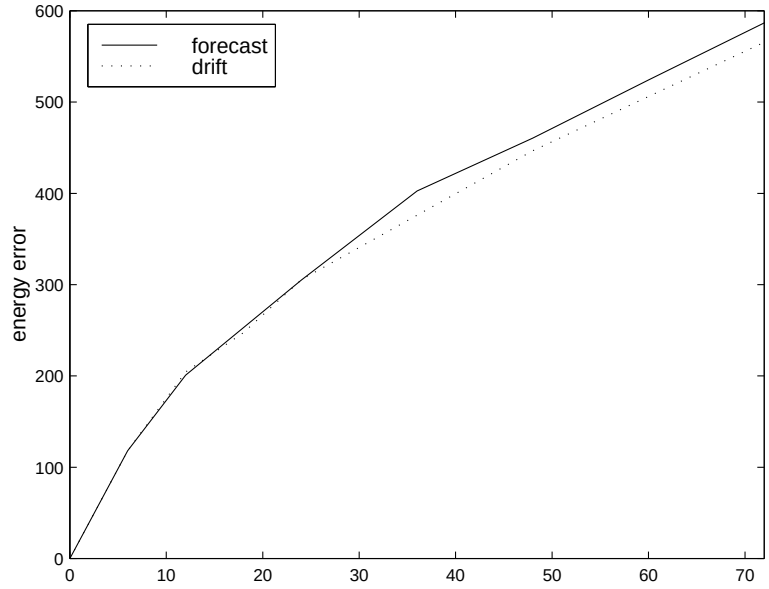


Figure 6.8: Plot of TL319 forecast error (solid line) with respect to analysis. Also shown is the drift (dotted). The drift is calculated using a time step of 6 hours for the first day and 24 hours for days 2 and 3.

The variation is probably related to the amount of data available to construct the analysis at each time. The mean cosine angle for consecutive vectors is 0.084, which is little higher than for 24 hour vectors. In order to smooth out some of the variation, consecutive drift vectors were combined to give 12 hour drift vectors, shown in the right hand panels. Both the magnitude and the cosine angle are more tightly focussed, with a mean cosine angle of 0.125.

We can use information about the mean magnitude and cosine angle of drift vectors to build a theoretical equation for model error. Suppose the drift over T hours has average magnitude d_m, and the cosine angle for consecutive days has mean c_m (we assume that correlations become negligible for periods of over one day). Then the drift is given by

$$d(t) = d_m \sqrt{\frac{t}{T}(1 + 2c_m) - 2c_m} \tag{6.4}$$

where $t > T$ is the time in hours.

This modified square-root curve has negative curvature, as did the drift of the Lorenz '96 models (see for example Figure 3.31). If the correlation c_m equals zero, then the drift is a square-root curve. This would be the case if the velocity error was equivalent to white noise.

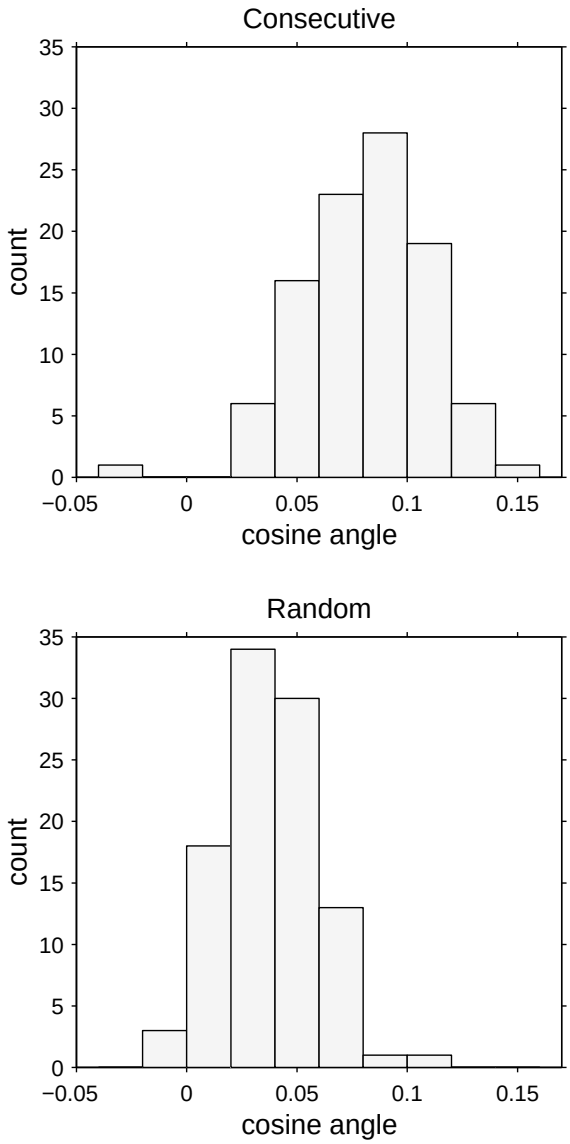


Figure 6.9: Upper panel shows the cosine angle for 24 drift vectors at consecutive days over a 100 day period from 15 Oct 1999. Lower panel shows the same for 100 randomly chosen pairs of days from the same period.

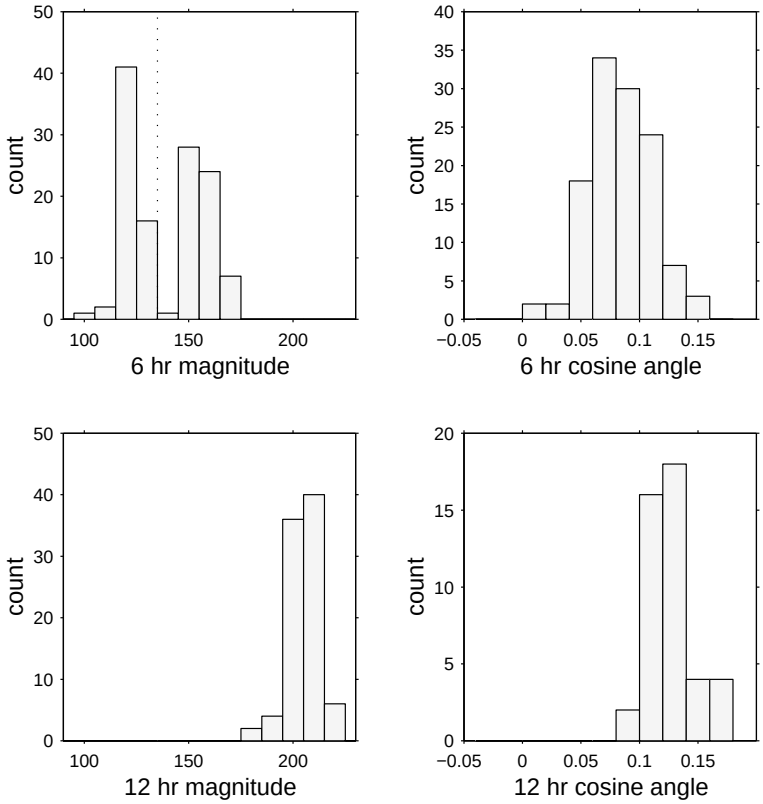


Figure 6.10: Magnitude and cosine angle of consecutive 6 and 12 hour drift vectors over a 30 day period from 15 Oct 1999. In the upper left panel, 6 hour drift vectors fall to the left of the dashed line when initiated at times 0 or 12 hours GMT, and to the right for times 6 and 18 hours GMT. The 12 hour drift vectors for the lower panels are formed from two consecutive 6 hour vectors.

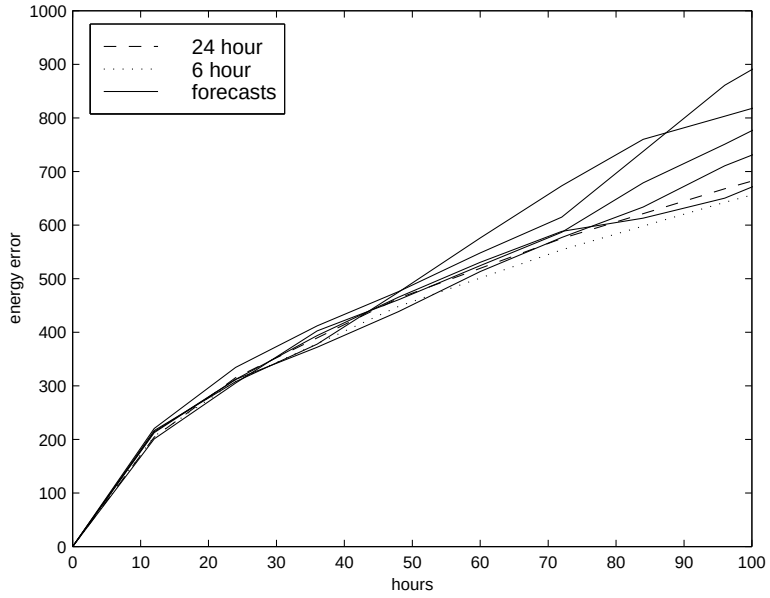


Figure 6.11: Plot of TL319 forecast errors (solid lines) with respect to analysis at four different starting dates, along with the theoretical approximation for model error from 24 hour drifts (dashed) and 6 hour drifts (dotted). The dates used were 1999/10/15, 1999/12/22, 2000/01/15, 2000/02/15, all at 12 GMT.

Figure 6.11 shows forecast errors for five starting dates, along with the model error as estimated using equation 6.4 with parameters calculated from 24 hour and 6 hour forecasts. The 24 hour curve has $T = 24$, $d_m = 315$ and $c_m = 0.081$. For the 6 hour curve, vectors were combined as above to form 12 hour drifts, in order to filter out some of the short-term variability, and reduce correlations between non-consecutive drift vectors. The values used were then $T = 12$, $d_m = 205$ and $c_m = 0.125$. In either case the theoretical curves closely match the forecast errors up to a time of three days.

It seems remarkable that forecast errors for five different days in five different months can be modelled using such a small amount of information, namely the mean magnitude and cosine angle of consecutive drift vectors. The weather itself may be chaotic, but our degree of ignorance of its future state is extremely reliable. Note that the technique will work less well in low dimension spaces, where correlations between random vectors is higher.

Although model error appears to dominate forecast error, displacement error will of course play a role, if only as a by-product of model error, as soon as the model departs from truth. The convolution of model error and displacement error will be

complex, but a rough picture can be obtained by assuming that, on top of the model error, an additional displacement error term, which is initiated by the model error, is added to the theoretical curve. For example, the drift calculation is performed by summing a series of short forecasts. We then suppose that each of these errors creates a displacement which magnifies exponentially. Therefore the model error over the first six hours leads to a displacement, which then grows at an exponential rate from that time on. The next six hours brings another displacement which also will magnify at the same rate. Each of these displacement error curves, initiated every 6 hours, are shown at the bottom of Figure 6.12. We assume a doubling time of 2.5 days, in accordance with the estimate given in [38]. Each displacement curve starts with magnitude zero, because it represents the additional error after the original displacement. After 2.5 days the error has doubled, so each curve has a magnitude equal to that of the original displacement, which is the same as the 6 hour drift. Summing each of these separate curves, and assuming orthogonality, which is justified given the dimension of the space, gives the total displacement error curve shown as the dot-dash line. When this displacement error is added to the drift, again assuming orthogonality, we arrive at the upper dashed line, which is an excellent fit to the RMS forecast errors.

Of course, the plot isn't meant to be an accurate representation of how model error and displacement error convolute. Nor does it confirm that error doubling times are 2.5 days; indeed, the displacement error is assumed to be orthogonal to the original error, which differs from the usual definition of doubling times. The graph's aim is merely to show that observed forecast errors are consistent with a combination of a large model error term, and a secondary displacement error term. It also seems reasonable, though, that forecast error, being a mix of model error and displacement error, could be loosely viewed as the sum of square root and exponential growth curves. The resulting curve has an initial negative curvature phase, followed by a nearly linear growth phase in the middle term, before eventually saturating. Interpolating Figure 6.12 forward, the model error and displacement error portions become roughly comparable in magnitude after about five days, though saturation effects will also come into play by that time.

The displacement error curve, being a sum of lagged exponential terms, isn't quite an exponential itself. It can be calculated explicitly by adding each of the separate error terms exactly as described above. If the e-folding time is $1/a$ (so the doubling time, again assuming that error growth is orthogonal to the original displacement, is

$log(2)/a$), and the individual drift vectors initiated each t_r time units have magnitude d_r, then the total displacement error $p(t)$ is seen to be

$$p(t) = d_r \sqrt{\frac{e^{at}-1}{e^{at_r}-1}\left(\frac{e^{at}+1}{e^{at_r}+1}-2\right)+\frac{t}{t_r}}. \tag{6.5}$$

In Figure 6.12, for example, t_r is 6 hours and d_r is the drift at 6 hours. Taking the limit as t_r goes to zero, we have

$$p(t) = s\sqrt{\frac{1}{2a}(e^{at}-1)(e^{at}-3)+t} \tag{6.6}$$

where $s = \lim_{t_r \to 0} \frac{d_r}{\sqrt{t_r}}$, which can be estimated for example from the 6 hour drift (the limit will exist if the drift varies with the square root of time).

The displacement error $p(t)$ created by the drift corresponds to the term in Eq. 4.12 which was omitted from the linearised dynamics. We see that for weather models, it is a relatively small effect. At 24 hours, it is about 10 percent of the drift, and at 12 hours it is only five percent.

The effect of the drift on shadow times can be estimated by using the shadow law to determine the likely drift for a certain shadow radius. The mean 24 hour drift over the days tested was 315, while the mean 6 hour drift for the days tested is 138. An upper estimate of shadow time from the 6 hour drift for a radius of 45 units is then about three to four hours. This result is definitely on the low end of what has been considered the likely range, and is a rather surprising result. It means that the dominant term in equation 4.13 is $\mathbf{d}$ rather than $\mathbf{M}$; model error rather than displacement error, drift rather than chaos. Weather models may be sensitive to initial conditions, but according to these results, they fail to shadow after just a few hours, and well before chaotic nonlinear growth becomes an issue.

6.7 Modelling the model error

While the conclusion that model error is responsible for the majority of forecast error over times of three days may seem less than encouraging, one positive note is that, because the drift vectors show a degree of coherence with time, it might be possible to develop techniques which compensate for it. For example, suppose we are making a 24 hour forecast, and we know the drift $\mathbf{d_p}$ from the preceding 24 hours. If the 24 hour forecast gives $\mathbf{d}$, and we assume that the cosine angle with $\mathbf{d_p}$ is c_m, then vector algebra shows that using $\mathbf{d} - c_m \mathbf{d_p}$ as the corrected forecast yields a fractional

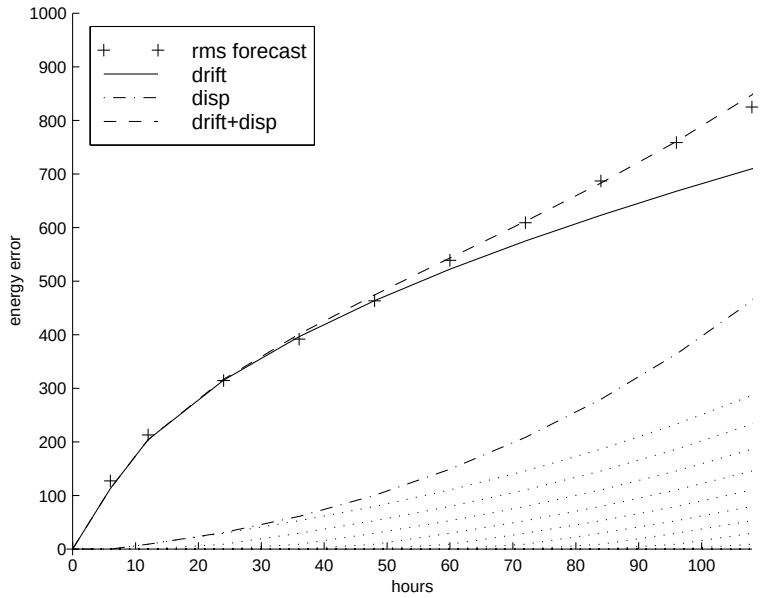


Figure 6.12: Plot showing how forecast error is consistent with a combination of model error and displacement error. The + signs shows TL319 RMS forecast errors, with respect to analysis, over five different starting dates. Solid line is the theoretical approximation for model error from equation 6.4. Dotted lines at bottom show series of displacement error curves initiated by the model error after each 6 hour period. Dot-dash line is the sum of the displacement error curves, assuming orthogonality. The dashed line which closely matches the data is the sum of the model error and displacement error curves, again assuming orthogonality. The dates used for the forecast error were 1999/10/15, 1999/11/15, 1999/12/22, 2000/01/15, 2000/02/15, all at 12 GMT.

improvement of $1 - \sqrt{1 - c_m^2}$. For $c_m = 0.081$ the correction is 0.33 percent. The correction increases to 0.78 percent if the 12 hour drift vectors, with mean correlation 0.125, are used instead. These are small improvements, but don't cost anything, and can probably be improved by considering more sophisticated schemes.

Such techniques are similar to the statistical approach for correcting forecast errors of [60] or [35], or the method proposed in [18]. Model parameters could also be tuned to reduce the forecast error, as suggested in [72].

An unexpected result was that the 24 drifts taken at random days in Figure 6.9 also shows a degree of coherence, with an average of 0.038. This implies that the model drift will not tend to zero over at least seasonal time scales. A first step to improve the model would therefore be to tune out this constant drift.

The square root shape is of course reminiscent of the integrated errors caused by a white noise spectrum, as discussed in Chapter 3. Perhaps a white or red noise model similiar to that in [51] is appropriate as a description of model error. It seems more likely, though, that the model error is not entirely random, and probably exists in a subspace of smaller dimension than the full space. What is certain is that adding white noise to the model won't make it more realistic, any more than stochastically varying the forcing improved the Lorenz model's performance in Chapter 5.

One topic that we haven't investigated is the spatial structure, or precise cause, of the model error, which is a topic of future work. Nor have we attempted to determine what component of the model error is due to model formulation and what is due to the projection. One method may be to examine data-rich and data-poor areas separately. The approach must be used with care, though, because the mathematics behind the linearised dynamics assumes that the model is well described by the equations and by the initial condition. If the model is limited to a small region, this condition will be violated, since the behaviour of the model in the specified region will be influenced by events in the other regions, and measurement of the drift vector will to some extent be affected.

6.8 Model error in the 500 hPa height

We have used the total energy norm because it provides a fairly complete description of the atmospheric state. If the weather is viewed as a flow of energy, then the total energy gives the amount of that energy associated with error. It also has the benefit of describing the sensible components of the atmospheric state, namely heat and wind. Meteorologists, though, often prefer to use the 500 hPa height to describe the

atmospheric state and record errors. The reason appears to be more historical than corresponding to any special property of this particular height [35].

As mentioned earlier, the 500 hPa height isn't an appropriate variable for our investigations into model error, since it is too incomplete a description of the atmosphere. However there is no reason why we shouldn't look at 500 hPa errors, and try to interpret them using the tools developed so far.

Figure 6.13 shows a plot of 500 hPa RMS height errors from ten separate forecasts. The errors no longer follow a square root curve, but grow quite linearly. One might think that the drift vectors are highly correlated, so that they add almost linearly as for the lower dimension Lorenz system, but Figure 6.14 shows this is not the case: the cosine angles have mean 0.13, which is higher than the total energy angles but still small.

Figure 6.13 also shows the approximate proportion of error due to displacement error, calculated using equation 6.5. A doubling time of 1.8 days was used in order to fit the data, which is faster than the 2.5 days used for the total energy error. The faster time makes sense when we compare the situation with the Lorenz system in Figures 4.29 and 4.30. When an incomplete set of parameters is used as a metric, the errors will be susceptible to the effect of rotations which preserve error magnitude but rotate error from one component to the other. Therefore one reason for the difference between the total energy errors and the 500 hPa errors is that 500 hPa model error is more sensitive to rotational displacement error, and therefore more likely to feed into a rapidly growing mode.

Another possible reason for the difference is that weather models may be better at predicting 500 hPa heights than they are at predicting other variables such as wind or temperature throughout the atmosphere, for example near the Earth's surface. (The fact, though, that the 500 hPa height was not a strong *source* of model error would not imply that it is *unaffected* by model error, which can advect in from other locales, or enter from other parameters through the primitive equations.)

In general, model errors are best analysed using as global a measure as possible. Just as attempting to predict the future direction of a highly complex stock market using only a single index is a risky (but popular) endeavour, so it may be misleading to interpret model error by its effect on height level.

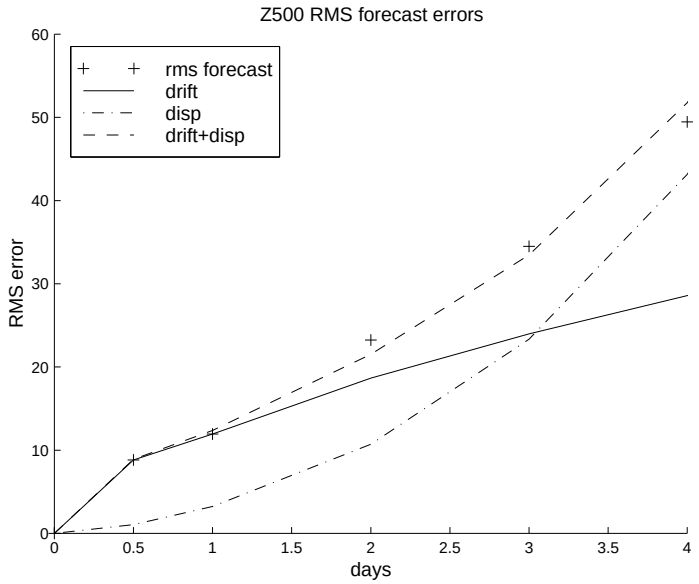


Figure 6.13: Plot of TL319 RMS 500 hPa height errors (solid line) with respect to analysis over ten different starting dates, along with the drift (dashed line) and the theoretical approximation for combined model and displacement error (dotted line). The ten forecasts were launched each ten days from 12 GMT on 1999/10/15.

6.9 Summary and discussion of results

In this chapter we have applied the methods for measuring model error to operational forecast models. Inter-model comparisons show that the T42 and T63 models have significant error relative to the TL159 control. Calculations of drift are hampered by spin-up effects, but indicate that model error is significant. Estimates of minimum shadow radius using the shadow law are in accordance with shadow orbits obtained using a sensitivity algorithm, and ensemble behavior is also consistent with high model error.

When the techniques were applied to the TL319 model relative to the analysis, it was found, unsurprisingly, that model error was higher than for the inter-model comparisons. The forecast error was found to be dominated by model error out to three days, and could be represented as a sum of a square-root model error curve, together with exponential displacement error curves. Estimated shadow times at the observational tolerance are in the region of only 3-4 hours.

We did not attempt to locate the cause of the model error, nor determine what proportion of the error is due to model structure and what to projection error over data-poor areas. We have concerned ourselves only with the magnitude of the error.

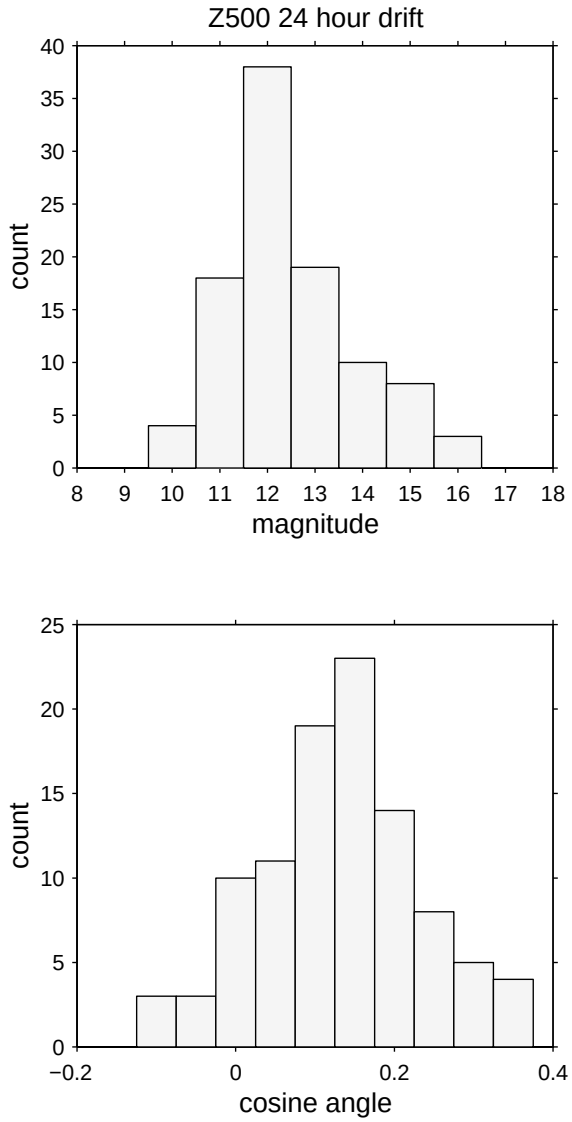


Figure 6.14: Magnitude and cosine angle for 24 drift vectors at consecutive days over a 100 day period from 15 Oct 1999 in the 500 hPa height norm.

What is clear is that making small displacements around the initial condition will not offset the effect of the error or, from the shadow law, produce a shadow orbit.

The above results indicate that model error has serious repercussions for many aspects of forecasting. For example, the technique for generating an analysis is essentially an attempt to force a model solution into fitting a set of observations. It appears, however, that the TL319 forecast is already generating errors of the same magnitude as the observation error after just a few hours. This complicates the analysis procedure and makes convergence unlikely.

Ensemble techniques which generate multiple initial conditions will also be affected. The usefulness of an ensemble is directly related to the model error: if model error is negligible, then the ensemble tells us everything we need to know about error distribution; but if model error is much larger than initial condition error, then the ensemble technique is just an expensive way of producing many wrong forecasts instead of one. For weather forecasts, we are in an intermediate position, so ensemble techniques contain information about some fraction of the error, but neglect an important component.

Ensemble techniques were designed to tackle the problem of sensitivity to initial conditions; however their use as a method to similarly tackle model error appears to be less justified. Singular vectors give a precise measure of initial condition error, in terms of the maximum error growth after a certain time, which can be used to generate initial conditions for an ensemble. No such method exists to produce the model perturbations which give maximum growth. Taking a collection of existing models and lumping them together in an ensemble may be an effective way of screening out particularly dud forecasts, but doesn't really address the problem of model error. Models which incorporate stochastic perturbations suffer from the same problem; there is no way to tell which are the correct perturbations to make.

Model error has usually been treated as some inherently unmeasurable quantity; but the fact is that model error, a zero order term, is *easier* to measure than initial condition error, a first order effect that requires calculation of singular vectors or similar. Only when model error is so small that it is dwarfed by initial condition error will it become difficult to measure. The assumption that atmospheric models have reached that state appears to be optimistic. Estimates of model error for T42 and T63 with respect to TL159, arrived at by drift, shadow and ensemble calculations, all indicate that error, even between these forecast models, is dominated by the model rather than initial condition. It is unsurprising, therefore, that drift calculations

indicate an even greater error between the operational forecast and the analysed weather.

Weather models are extremely complicated entities which have to incorporate all kinds of interactions between air, water, the ground and so on. They are also looking for a fairly small signal - weather fluctuations - on a large background field - the climate. Any finite element modeller or other person with experience of modelling physical systems must view the weather as one of the most complex problems imaginable, and regard with awe the progress that meteorologists have made. At the same time, such a person would find the perfect or nearly-perfect model hypotheses puzzling. Models may have improved substantially in recent years, but, as analysis of forecast errors has shown, they are certainly not perfect. The best way to proceed must be to measure the error, optimise the model to reduce it, and then do whatever is possible to predict and possibly offset the residue error.

Chapter 7

Conclusions

7.1 Summary of results

In this thesis we have studied model error over a range of systems, from the smallest 3D to the largest operational weather model. Of particular interest were the medium-dimension Lorenz '96 systems of Chapter 2: small enough to compute rapidly, but large enough to produce behaviour qualitatively similar to atmospheric variables. The intricate beauty of these systems, with their interplay of periodic, quasi-periodic and chaotic orbits, was revealed, for the first time, with spectral bifurcation diagrams.

In Chapter 3, it was shown that a fundamental difference between model error and initial condition error is that model error has a non-zero initial slope. This slope was termed the velocity error. The complexity of the Lorenz system behaviour as the forcing parameter F was varied led one to suspect that velocity error would be equally complex; yet it was found that emergent properties of the systems made velocity error surprisingly smooth as a function of the forcing. For the constant model, the model error, both in terms of initial velocity error and shadowing times, simply varied with the square root of F.

The linear model, which employed a simple linear parameterisation of the forcing error, gave dramatically improved shadowing behaviour, even though the reduction in initial velocity error was relatively modest. The reason was found to be that low frequency forcing error was the primary determinant of shadow times, and since the linear model had a less 'red' forcing error power spectrum than the constant model, its performance was improved.

This result led to the detailed investigation of the shadowing process in Chapter 4. By linearising around the true attractor, the displacement of a shadow trajectory from

truth was approximated by a linear equation, involving a zero order drift term and the first order propagator term. The linearised dynamics were used to develop a shadow estimation technique (SET), which was seen to accurately predict shadow behaviour even over prediction times when the system itself had ceased to be linear. This was because the linearised dynamics applied only to shadow orbits, which remained within the shadow radius of the true orbit.

The linearised dynamics, when applied to locally dissipative models, led to the discovery of a shadow law, which states that, in an RMS sense, minimum shadow radius is bounded below by half the drift. When model error is high, the minimum shadow radius approaches the bound, and drift is approximately equal to the shadow diameter. Even with an estimate of shadow times, it is always desirable to compute actual shadow orbits for verification; therefore optimization schemes to find shadow orbits, even for large models with long computation times, were proposed and tested.

The usefulness of ensemble techniques, in terms for example of the accuracy of computed spread, was found to depend critically on the model error. If the model error dominates displacement error, then it is natural that ensembles should give less information about the likely correct forecasts. It was noted that in high dimension spaces, model error and displacement error are expected to be nearly orthogonal, making it unlikely that model error could be offset by searching in the space of singular vectors.

Chapter 5 turned attention away from the short and medium range, and looked at the question of predictability of the second kind; long term climate. Analytic results concerning the mean and variance of the Lorenz '96 systems were derived, and it was found that the linear model, something of a champion among simple models, was capable of matching both quantities over a large range of forcings, and suggested a link between short term and long term predictability. Other models, which invoked various stochastic schemes to simulate the properties of the true forcing, were also tested. For the models studied, the stochastic approach appeared to have no benefit over the non-stochastic models.

Finally, Chapter 6 applied the techniques developed for the lower dimensional systems to full weather models. First, the T42 and T63 models were compared with TL159. An upper bound on shadow times was estimated using the drift, and compared with results using an optimisation program. The two methods gave compatible results.

The fact that model error outweighed displacement error was confirmed when an ensemble of T42 initial conditions was run. As expected, no ensemble member reduced

error, and the ensemble mean offered no improvement over the control forecast. Any 'return to skill' by an ensemble member would be due to chance, since, if it isn't better after a day, there's no reason to expect that it should be better after a week. The initial condition errors were also seen to be nearly orthogonal to the model error, in agreement with theory.

The operational TL319 forecast was then compared with the analysis to estimate how long the forecast could shadow the true weather. The drift calculations indicated that model error was higher than for the inter-model comparisons, and shadow times were estimated to be in the region of 3 to 4 hours, which would have severe impact on analysis and ensemble techniques. Another square root law emerged: model error, and indeed forecast error, varied with a simple square root formula up to about three days. It meant that forecast error could be predicted just from a knowledge of the mean drift magnitude and cosine angle between consecutive drift vectors, both of which are fairly stable quantities. Despite the immense complexity of the weather, and of the model, the difference between the two after a certain time is remarkably constant from day to day or month to month. To globally conserved quantities such as momentum or mass, we may now add a new one: error.

Knowing the size of the error is one thing; knowing its direction is another. (It may be interesting that all forecasts are wrong by the same amount, but it isn't very useful.) The fact that model error was so large and consistent was at least seen to offer a potential solution. If the error had predictable features, then it could be possible to effectively model the model error, and thus correct the forecast. Similar techniques, based on predictor methods, had been used to some effect with the Lorenz systems. The model could also be improved by tuning parameters to minimise drift. Without such a measure of model error, though, it would be impossible to make much progress in improving the model, since to do so would be like working in the dark.

7.2 Does chaos matter?

As Bjerknes first said, forecast error is due to a combination of model error and initial condition error. The former is mostly a question of physics or engineering; how well can we model the complicated physical laws governing the atmosphere with a set of differential equations truncated to a finite grid? The latter effect is related to chaos theory; how sensitive is the atmosphere (or the model) to small perturbations in the initial condition?

Which of these sources of error is more important will depend on the particular system and model. In the case of the atmosphere, the answer to this question hasn't been known. Good estimates of initial condition error are available, from singular vector and other methods, but there hasn't been a corresponding measure of model error; at the same time, and contradictorily, it has been widely assumed to be small.

In fact, considering the enormous number of articles written for meteorological journals, a search reveals that relatively few refer to model error in the abstract (most of those have been referenced in this thesis). It is slightly puzzling why more attention hasn't been paid to the subject of model error. As mentioned earlier, it is easier to measure model error, a zero order effect, than it is to measure initial condition error, which depends on complicated first order derivative estimates based on a large number of forecasts and requiring the use of an adjoint model. Also, an examination of typical RMS forecast error plots reveals not the quasi-exponential growth characteristic of displacement error, but the square root curve that is characteristic of model error. So why was it assumed that forecast error was primarily due to initial condition error, and not the model? Why jump to the more complicated first order explanation before eliminating the simple zero order cause?

Some reasons suggest themselves. Initial condition error shifts the blame for bad forecasts away from the forecast centres towards the inherent unpredictability of the weather. Model error, meanwhile, isn't as culturally important in meteorology as it is in other fields (say bridge design) which also employ sophisticated and complicated modelling techniques. Any engineer is familiar with the sense of anxiety that mingles with anticipation as a project nears completion: the computer models are replaced by a physical object which will follow not quite the same rules, and at the same time any mistake or omission in the calculation will become very evident (consider the Millennium bridge). Meteorologists, in contrast, aren't responsible for the weather, and therefore by implication their responsibility is dimmed when, as in any case always happens, it does something other than predicted. And meteorologists don't get sued (though people have tried [32]).

Part of the reason, though, must also be due to the entity pictured in Figure 7.1. Butterflies, it seems, can do more than stir up storms by flapping their wings. They can also deflect the course of entire branches of science. This particular example was spotted to his great credit by Ed Lorenz in 1963, but, like other revolutions of its type, took 30 years to fully develop. It was then that the technique of ensemble forecasting was introduced: a net in which to catch the unruly but attractive butterfly of chaos.

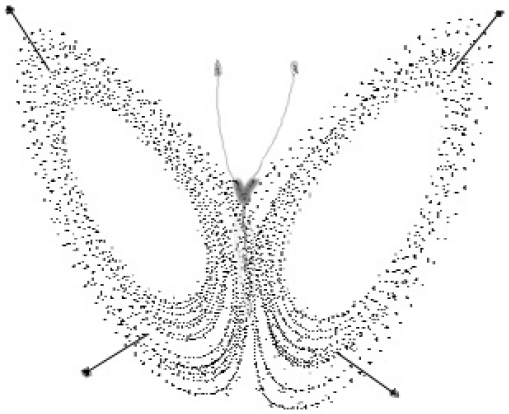
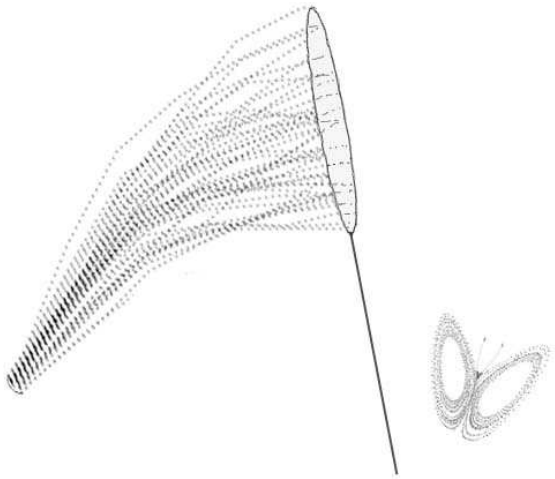


Figure 7.1: Butterfly. Species *Lorenzus*. First spotted 1963. Incubation period 30 years.

Chaos may well place an upper limit on prediction times of a couple of weeks. It seems premature, though, to worry too much about that if the models are already introducing significant errors after just a few hours. The ensemble net has missed its target almost before it is thrown, and the butterfly escapes.

Weather forecasting ranks as one of science's greatest and most prodigous modelling endeavours; but, like most human pursuits, it has yet to banish the effects of error, uncertainty, or chaos. Ensemble techniques have played their part in improving our understanding of the latter. In seeking methods to further improve forecasts, though, it would be preferable to devote additional resources to analysing model error, and using the information thus gained to develop the model. For it is here, rather than in the effects of chaos or the flapping of an insect's wings, that the primary cause of near and medium range forecast error lies.



Appendix A

Proof of the shadow law

Referring to equation 4.44, the proof of the shadow law rests on showing that, for a dissipative model, the sum

$$\sum_{i=1}^{n} \frac{1}{(1+\sigma_i)^2} \tag{A.1}$$

has a minimum value of $\frac{n}{4}$. Suppose first that the model is volume preserving, so

$$\prod_{i=1}^{n} \sigma_i = 1. \tag{A.2}$$

Writing the minimisation problem as a Lagrangian, we seek minima of

$$\sum_{i=1}^{n} \frac{1}{(1+\sigma_i)^2} + \lambda(\prod_{i=1}^{n} \sigma_i - 1) \tag{A.3}$$

where λ is a constant multiplier. Taking partial derivatives with respect to σ_j, and setting to zero, gives

$$\sigma_j = \frac{2}{\lambda}(\sigma_j + 1)^3 \tag{A.4}$$

which has two solutions for $\lambda > 16$, and a single solution when $\lambda = 16$ and all $\sigma_j = 1$. Since λ is the same for all j, the multipliers σ_j can only take on one of a maximum two values apart from 1, and they must also satisfy equation A.2.

We claim that the solution $\sigma_j = 1$ for all j, for which the sum in equation A.1 is equal to $\frac{n}{4}$, represents a global minimum. We do this by examining the other critical points. Suppose that some other arrangement of σ_j's satisfies the criticality requirement. For a particular value of λ, there are only three possibilities for each σ_j: σ_1, σ_2, or 1, where $\sigma_1 < 1$ and $\sigma_2 > 1$ are roots of equation A.4. We set n_1 equal to the number of occurences of σ_1, n_2 the number of σ_2's, n_3 the number of 1's, and $n_s = n_1 + n_2$, so $n_s + n_3 = n$.

Let $\beta = \frac{2}{\lambda}$. Then from equation A.4,

$$\sigma_1 = \beta(\sigma_1 + 1)^3 \tag{A.5}$$

and

$$\sigma_2 = \beta(\sigma_2 + 1)^3. \tag{A.6}$$

From equation A.2,

$$1 = \prod_{i=1}^{n} \sigma_i = \sigma_1^{n_1}\sigma_2^{n_2} \tag{A.7}$$

so, taking the n_s'th root of each side,

$$\sigma_1^a\sigma_2^b = 1 \tag{A.8}$$

where $a = \frac{n_1}{n_s}$ and $b = \frac{n_2}{n_s}$. Substituting the expressions from equations A.5 and A.6 into A.8, we can solve for β as

$$\beta = (1 + \sigma_1)^{-3a}(1 + \sigma_2)^{-3b} \tag{A.9}$$

so equation A.5 then becomes

$$\sigma_1 = \beta(\sigma_1 + 1)^3 = (1 + \sigma_1)^{3-3a}(1 + \sigma_2)^{-3b}. \tag{A.10}$$

Solving for σ_2, we obtain

$$\sigma_2 = \sigma_1^{-\frac{1}{3b}}(1 + \sigma_1) - 1. \tag{A.11}$$

Substituting into equation A.8 gives $g(\sigma_1, a) = 1$, where the function g is defined as

$$g(\sigma_1, a) = \sigma_1^a(\sigma_1^{-\frac{1}{3(1-a)}}(1 + \sigma_1) - 1)^{1-a}. \tag{A.12}$$

For a given value of σ_1, the requirement that $g(\sigma_1, a) = 1$ can be used to solve for $a = f(\sigma_1)$. It is easily seen that f is monotonically decreasing from 0 to 0.2364 and negative for $0.2364 < \sigma_1 < 1$ (σ_1 is less than 1 by assumption). Since we require $a > 0$, it follows that σ_1 is in the range 0 to 0.2364. The function f, for σ_1 in that range, is shown plotted in Figure A.1.

We now show that the sum A.1, evaluated at such a critical point, has a value greater than $\frac{n}{4}$. We can write

$$\sum_{i=1}^{n} \frac{1}{(1 + \sigma_i)^2} = \frac{n_1}{(1 + \sigma_1)^2} + \frac{n_2}{(1 + \sigma_2)^2} + \frac{n_3}{4} \tag{A.13}$$

$$= n_s\left(\frac{a}{(1 + \sigma_1)^2} + \frac{b}{(1 + \sigma_2)^2}\right) + \frac{n_3}{4} \tag{A.14}$$

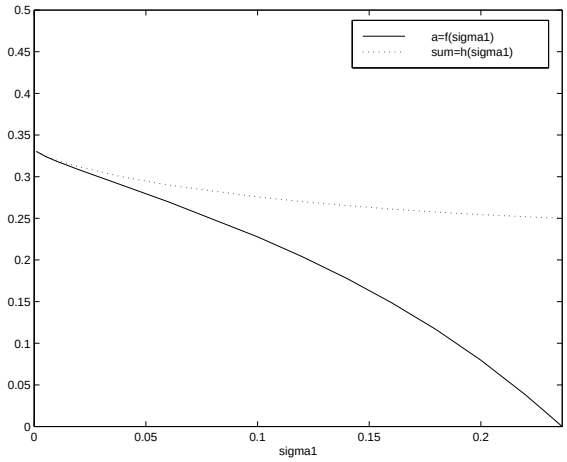


Figure A.1: Plot of f and h, which are defined in the text, as a function of the singular value σ_1 over the range for which f is non-negative.

$$= n_s\left(\frac{a}{(1+\sigma_1)^2} + \frac{b}{\sigma_1^{-\frac{2}{3b}}(1+\sigma_1)^2}\right) + \frac{n_3}{4} \tag{A.15}$$

$$= n_s\left(\frac{f(\sigma_1)}{(1+\sigma_1)^2} + \frac{1-f(\sigma_1)}{\sigma_1^{-\frac{2}{3(1-f(\sigma_1))}}(1+\sigma_1)^2}\right) + \frac{n_3}{4} \tag{A.16}$$

$$= h(\sigma_1) + \frac{n_3}{4} \tag{A.17}$$

where

$$h(\sigma_1) = \frac{f(\sigma_1)}{(1+\sigma_1)^2} + \frac{1-f(\sigma_1)}{\sigma_1^{-\frac{2}{3(1-f(\sigma_1))}}(1+\sigma_1)^2}. \tag{A.18}$$

The function $h(\sigma_1)$, shown in Figure A.1, is also monotonically decreasing over the range 0 to 0.2364, and has a minimum value of $\frac{1}{4}$ at $\sigma_1 = 0.2364$, for which $a = 0$. In the nontrivial case that $a > 0$, we have $h(\sigma_1) > 0.25$, so

$$\sum_{i=1}^{n} \frac{1}{(1+\sigma_i)^2} = n_s h(\sigma_1) + \frac{n_3}{4} > \frac{n_s}{4} + \frac{n_3}{4} = \frac{n}{4}. \tag{A.19}$$

It thus follows that the critical point with $\sigma_i = 1$ for all i represents a global minimum, as desired. The case where the model is strictly dissipative, so that $\prod_{i=1}^{n} \sigma_i < 1$, follows easily.

Appendix B

Glossary

Analysis. Meteorological term denoting the best approximation to the real weather, as expressed in terms of model variables. It is obtained by interpolating a combination of observed data and model predictions.

Displacement error. Error due to model equations being evaluated at a point other than the true point projected into model space.

Drift. Magnitude of integrated velocity error, evaluated over the projection into model space of a segment of a true orbit. Used as a measure of model error, and to estimate shadow times, via the SET, or bound them via the shadow law.

ECMWF. The European Centre for Medium-Range Weather Forecasts, located in Reading, UK.

Forcing. Refers to a term in a system or model ode, usually to represent some external input to a physical system, such as solar heat in the case of the weather, or forcing of a pendulum.

Forcing error. Velocity error due to error in forcing term.

Four-dimensional variational assimilation (4D-VAR). A technique which determines the analysis by combining observed data with a model forecast initiated usually 6-12 hours earlier.

Initial condition error. Displacement error at initial time, caused by incorrect initial condition. May be large for chaotic systems due to sensitivity to initial condition.

Initial velocity error. The velocity error measured at initial time.

Integrated forcing error. Forcing error integrated over the projection into model space of a segment of a true orbit.

Integrated velocity error. The velocity error integrated over the projection into model space of a segment of a true orbit. Has dimension distance. See drift.

Model error. Refers to error due to the difference between model equations and true system, as measured on the projection of a true orbit into model space.

Shadow estimation technique. A procedure for estimating shadow times for a given initial condition and shadow radius, without explicitly producing a shadow orbit. Referred to as SET.

Shadow law. A law which states that, for any dissipative model, and in an RMS sense, an approximate lower bound on shadow radius is given by half the drift.

Shadow orbit. Given a specific radius r and true orbit, a shadow orbit is a model trajectory which stays within the radius r of the true orbit, as measured in model state space.

Shadow radius. The radius used in shadowing calculations.

Shadow time. The time for which a shadow orbit stays within the shadow radius of the true orbit.

Velocity. The rate of change of a system or model variable. In the case of model variables the velocity can be calculated using the ode.

Velocity error. The difference between the system velocity at a particular point, measured in model space, and the model velocity at the projection of that point into model state space.

Bibliography

[1] K.T. Alligood, T.D. Sauer, and J.A. Yorke. *Chaos: An Introduction to Nonlinear Dynamical Systems.* Springer-Verlag, New York, 1996.

[2] D.V. Anosov. Geodesic flows and closed Riemannian manifolds with negative curvature. *Proc. Steklov Inst. Math.*, **90**, 1967.

[3] A.I. Bennett and W.P. Budgell. Ocean data assimilation and the Kalman filter: Spatial regularity. *J. Phys. Oceanogr.*, **17**:1583–1601, 1987.

[4] V. Bjerknes. *Dynamic Meteorology and Hydrography, Part II. Kinematics.* Gibson Bros., Carnegie Institute, New York, 1911.

[5] R. Bowen. ω-limit sets for axiom A diffeomorphisms. *J. Diff. Eqns.*, **18**:333–339, 1975.

[6] R. Buizza. Linearity of the optimal perturbations time evolution and sensitivity of ensemble forecasts to perturbation analysis. *Quart. J. Roy. Meteor. Soc.*, 1995.

[7] R. Buizza, J. Barkmeijer, T.N. Palmer, and D.S. Richardson. Current status and future developments of the ECMWF ensemble prediction system. *Meteorol. Applic.*, **7**:163–175, 2000.

[8] R. Buizza, M. Miller, and T.N. Palmer. Stochastic simulation of model uncertainties in the ECMWF ensemble prediction scheme. In T. Palmer, editor, *Predictability*, Shinfield Park, Reading UK, 1997. European Centre for Medium-Range Weather Forecasting.

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

[10] R. Buizza, T. Petroliagis, T.N. Palmer, J. Barkmeijer, M. Hamrud, A. Hollingsworth, A. Simmons, and N. Wedi. Impact of model resolution

and ensemble size on the performance of an ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **124(550B)**:1935–1960, 1995.

[11] C. Chatfield. *The Analysis of Time Series.* Chapman and Hall, New York, 1989.

[12] P.C. Chu. Two kinds of predictability in the lorenz system. *J. Atmos. Sci.*, **56**:1427–1432, 1999.

[13] S.E. Cohn. An introduction to estimation theory. *J. Meteorol. Soc. Japan*, **75**:257–288, 1997.

[14] S.E. Cohn and D.F. Parrish. The behavior of forecast error covariances for a Kalman filter in two dimensions. *Mon. Wea. Rev.*, **119**:1757–1785, 1991.

[15] P. Courtier and O. Talagrand. Variational assimilation of meteorological observations with the adjoint vorticity equation. Part 2: Numerical results. *Q.J.R. Meteorol. Soc.*, **113**:1329–1347, 1994.

[16] P. Courtier, J.-N. Thepaut, and A. Hollingsworth. A strategy for operational implementation of 4d-var, using an incremental approach. *Q.J.R. Meteorol. Soc.*, **120**:1367–1388, 1994.

[17] R. Daley. *Atmospheric Data Analysis.* Cambridge University Press, Cambridge, 1991.

[18] F. D'Andrea and R. Vautard. Reducing systematic errors by empirically correcting model errors. *Tellus*, **52A**:21–41, 2000.

[19] D.P. Dee. On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**:1128–1145, 1995.

[20] J.F. Le Dimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations. *Tellus*, **37A**:97–110, 1988.

[21] R.A. Downton and R.S. Bell. The impact of analysis differences on a medium-range forecast. *Meteorol. Mag.*, **117**:279–285, 1988.

[22] P.E. Gill, W. Murray, and M. Wright. *Practical Optimization.* Academic Press, New York, 1981.

[23] I. Gilmour. Nonlinear model evaluation: ι-shadowing, probabilistic prediction and weather forecasting. *D. Phil. Thesis, Oxford University*, 1998.

[24] I. Gilmour, L.A. Smith, and R. Buizza. On the duration of the linear regime: Is 24 hours a long time in weather forecasting? *J. Atmos. Sci.*, **58**:3525–3539, 2001.

[25] G.H. Golub and C.F. Van Loan. *Matrix Computations*. The John Hopkins University Press, 1989.

[26] J. Guckenheimer and P. Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springer-Verlag, New York, 1983.

[27] J. A. Hansen. Adaptive observations in spatially-extended, nonlinear dynamical systems. *Ph.D. Thesis, Oxford University*, 1998.

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

[29] A. Hollingsworth, A.C. Lorenc, M.S. Tracton, K. Arpe, G. Cats, S. Uppala, and P. Kallberg. The response of numerical weather prediction systems to fgge level iib data. part ii: Forecast verifications and implications for predictability. *Q.J.R. Meteorol. Soc.*, **111**:67–101, 1985.

[30] P.L. Houtekamer, L. Lefaivre, J. Derome, H. Ritchie, and H. Mitchell. A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**:1225–1242, 1996.

[31] C.P. Arnold Jr. and C.H. Dey. Observing-systems simulation experiments: Past, present and future. *Bull. Amer. Meteor. Soc.*, **67**:687–695, 1986.

[32] S. Junger. *The Perfect Storm*. Fourth Estate, London, 1997.

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

[34] J.F. Lacarra and O. Talagrand. Short-range evolution of small perturbations in a barotropic model. *Tellus A*, **40**:81–95, 1988.

[35] C.E. Leith. Objective methods of weather prediction. *Ann. Rev. Fluid Mech.*, **10**:107–128, 1978.

[36] J.M. Lewis and J.C. Derber. The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus*, **37A**:309–322, 1985.

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

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

[39] E. N. Lorenz. Climate predictability. the physical basis of climate modelling. *WMO, GARP Publication Series*, **16**:132–136, 1975.

[40] E.N. Lorenz. Predictability - a problem partly solved. In T. Palmer, editor, *Predictability*, Shinfield Park, Reading UK, 1996. European Centre for Medium-Range Weather Forecasting.

[41] J.E. Lovelock. *Gaia. A New Look at Life on Earth.* Oxford University Press, Oxford, 1979.

[42] Pierre-Simon Marquis de Laplace. *Thèorie Analytique des Probabilitiés.* Paris, 1820. Reproduced in the Ouvres complètes de Laplace, Paris, Volume 11, 1886.

[43] K. McGuffie and A. Henderson-Sellers. *A Climate Modelling Primer.* John Wiley and Sons, Chichester, 1996.

[44] F. Molteni, R. Buizza, T.N. Palmer, and T. Petroliagis. The ECMWF ensemble prediction system: Methodology and validation. *Q.J.R. Meteorol. Soc.*, **122**:73–119, 1996.

[45] D. Orrell. A method for harmonic field measurements during magnet ramp. *SSCL Technical Report, Superconducting Super Collider Laboratory, 2550 Beckleymeade Avenue, Dallas*, 1995.

[46] D. Orrell, J. Turner, P. Giavonnoni, F. Kircher, C. Lyraud, J. Perot, JM. Rifflet, and P. Vedrine. Mechanical, thermal and electromagnetic design for a superconducting quadrupole for the ssc. *IEEE Transactions on Magnetics*, **30**(44):1954–1957, 1994.

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

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

[49] T.N. Palmer. Predicting uncertainty in forecasts of weather and climate. *Reports on Progress in Physics*, **63**:71–116, 2000.

[50] T.N. Palmer and D.L.T. Anderson. The prospects for seasonal forecasting - a review paper. *Quart. J. Roy. Meteor. Soc.*, **114**:755–793, 1994.

[51] N.A. Philips. The spatial statistics of random geostrophic modes and first-guess error. *Tellus*, **38A**:314–322, 1986.

[52] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes.* Cambridge University Press, Cambridge, 1986.

[53] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, Cambridge, 1993.

[54] F. Rabier and P. Courtier. Four-dimensional assimilation in the presence of baroclinic instability. *Q.J.R. Meteorol. Soc.*, **118**:649–672, 1992.

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

[56] C. Reynolds, P.J. Webster, and E. Kalnay. Random error growth in the nmc medium range forecasting model. *Mon. Wea. Rev.*, 1993.

[57] D. Richardson. The relative effect of model and analysis differences on ECMWF and ukmo operational forecasts. In T. Palmer, editor, *Predictability*, Shinfield Park, Reading UK, 1997. European Centre for Medium-Range Weather Forecasting.

[58] O. E. Rössler. An equation for continuous chaos. *Phys. Lett. A*, **57A(5)**:397–398, 1976.

[59] N. Rulkov, L. Tsimring, and H. Abarbanel. Tracking unstable orbits in chaos using dissipative feedback controls. *Phys. Rev. E*, **50**:314–324, 1994.

[60] S. Schubert and Y. Schang. An objective method for inferring sources of model error. *Monthly Weather Review*, **124**:325–340, 1996.

[61] L. A. Smith. The maintenance of uncertainty. In G. Cini Castagnoli and A. Provenzale, editors, *Past and Present Variability in the Solar-Terrestial System: Measurement, Data Analysis and Theoretical Models*, **CXXXIII** of *International School of Physics "Enrico Fermi"*, pages 177–246, Bologna, 1997. Il Nuovo Cimento.

[62] L.A. Smith. Local optimal prediction: exploiting strangeness and the variation of sensitivity to initial condition. *Phil. Trans. R. Soc. Lond. A*, **348**:371–381, 1994.

[63] L.A. Smith. Disentangling uncertainty and error: On the predictability of nonlinear systems. In A. I. Mees, editor, *Nonlinear Dynamics and Statistics*, pages 31–64, Boston, 2000. Birkhauser.

[64] L.A. Smith, I. Gilmour, and P.L. Read. A brief overview of the seminar on atmospheric predictability. *OCIAM Technical Report, OCIAM, 24-29 St. Giles, Oxford, OX1 3LB, U.K.*, 1996.

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

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

[67] Z. Toth and E. Kalnay. Ensemble forecasting at nmc: the generation of perturbations. *Bull. Am. Meteorol. Soc.*, **74**:2317–2330, 1993.

[68] Z. Toth, E. Kalnay, S. Tracton, S. Wobus, and J. Irwin. A synoptic evaluation of the ncep ensemble. In T. Palmer, editor, *Proc. 5th Workshop on Meteorological Operating Systems*, Shinfield Park, Reading UK, 1996. European Centre for Medium-Range Weather Forecasting.

[69] K.E. Trenberth. *Climate System Modelling.* Cambridge University Press, Cambridge, 1992.

[70] H.M. Vandendool and L. Rukhovets. Current status and future developments of the ECMWF ensemble prediction system. *Wea. Forecasting.*, **9**:457–465, 1994.

[71] T. Vukicevic. Nonlinear and linear evolution of initial forecast errors. *Mon. Wea. Rev.*, **119**:1603–1611, 1991.

[72] W. Wergen. The effect of model errors in variational assimilation. *Tellus*, **44A**:297–313, 1992.

[73] D.L. Williamson and A. Kasahara. Adaptation of meteorological variables forced by updating. *J. Atmos. Sci.*, **28**:1313–1324, 1971.

[74] D.L. Williamson and A. Kasahara. Evaluation of tropical wind and reference pressure experiments for observing systems. *Tellus*, **24**:100–115, 1972.

[75] R. Wobus and E. Kalnay. Two years of operational prediction of forecast skill at nmc. In *Proc. 10th Conf. on Numerical Weather Prediction*, Portland, OR, 1994. Amer. Meteor. Soc.