Chaos and the Spring Barrier: Aspects of Predictability in the El Niño-Southern Oscillation

by

John Paul Weiss

B. A., Middlebury College, 1991

M. A., University of Colorado, 1994

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

College of Arts and Sciences

1998

This thesis entitled:
Chaos and the Spring Barrier: Aspects of Predictability in the El Niño-Southern Oscillation
written by John Paul Weiss
has been approved for the College of Arts and Sciences

Jeffrey B. Weiss

John Wahr

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Weiss, John Paul (Ph. D., Physics)
Chaos and the Spring Barrier: Aspects of Predictability in the El Niño-Southern Oscillation
Thesis directed by Professor Jeffrey B. Weiss

ENSO, the El Niño-Southern Oscillation, exhibits a seasonal dependence in persistence, known as the "spring barrier," the "predictability barrier," or the "persistence barrier." Persistence is the cyclostationary lag autocorrelation, a form of statistic that accounts for the natural periodicity of a time series. It depends on both a lag and a phase of the year. A mathematical examination of persistence reveals that certain simple time series also generate a "barrier."

Using simple time series consisting of a single sine wave or the sum of two sine waves, it will be shown that frequencies either equal to a biennial cycle or with values that sum to unity produce a barrier-like feature. Time series with continuous but narrow spectral peaks centered on "barrier-producing frequencies" only produce a barrier-like feature if the phase angles vary slowly or coherently across the peaks. Numerical experiments will verify the importance of barrier-producing frequencies and coherency of their phase angles.

In this work, the term "persistence barrier," or simply "barrier," shall denote a feature in persistence similar to that produced by a biennial cycle. A model of this shape is constructed, a model containing four parameters, A, τ, Δp, and p_b. A nonlinear least-squares fit of this model to the persistence of a time series provides a way to measure the properties of the persistence barrier, using these four parameters.

Application of the aforementioned measures to the Southern Oscillation Index and the sea surface temperature in the "NINO3-region" of the equatorial Pacific Ocean shall show that the barrier in these datasets is statistically distinguishable from the persistence of a red-noise process. The measures shall also provide a means of examining interdecadal variability in ENSO. Said examination reveals a weak persistence barrier from 1915 to 1945, a strong barrier during the 1960's and early 1970's, and a weakening barrier after 1975.

Although loss of correlation may provide an indication of a loss of predictability, a better quantification of such a loss are the so-called Lyapunov exponents. Lyapunov exponents describe the exponential growth and decay rates of linearized perturbations to an initial condition. The global Lyapunov exponents, or GLE, describe error growth over

over the whole of the system for all time. A GLE with a positive sign indicates a chaotic system. A more useful quantity, the instantaneous Lyapunov exponents, or ILE, describe error growth "at a point in time" for the system. An average over the ILE for all time produces the GLE.

Examinations of ENSO models demonstrate a quasiperiodic route to chaos in these models, a sign of low-order chaos. These models vary in complexity. This work examines the persistence of a simple system exhibiting the quasiperiodic route to chaos, the damped forced pendulum. It shall be seen that the quasiperiodic route to chaos in the pendulum produces persistence barriers more frequently than other routes to chaos exhibited by the pendulum.

Certain simple periodic time series produce persistence barriers. One may be tempted to assume that persistence barriers are associated with predictable time series. It shall be shown, however, that persistence barriers are also associated with a loss of predictability. Distributions of the ILE, computed for the pendulum on the quasiperiodic route to chaos, show an increase in error growth associated with the value of p_b, the measure of the barrier position.

Dedication

To the Old Souls

Who help us along The Way

Acknowledgements

SOI and NINO3 data from 1950-1996 courtesy Don Garrett, Climate Prediction Center, NOAA. Data from 1896-1949 courtesy Chris Torrence, NCAR.

This work was funded by NOAA grant DOC NA56GP0230.

Thanks to Sue Haupt for her assistance and advice with genetic algorithms, to Chris Torrence for his assistance with ENSO data and discussions, and to Cécile Penland for general discussions.

Additional thanks go to the physics department staff for their assistance during the past seven years.

I would like to offer special thanks to Dr. Jeffrey B. Weiss for his assistance, advice, and guidance during the past four years.

On a personal note, I wish to thank my life-partner, Karl Boggs, for his emotional and moral support during the past five years. I also offer special thanks to Sandy Levy and the staff of the mental health wing at the Mapleton branch of Boulder Community Hospital for assisting me in my struggle with mental illness. Without their professionalism and skill, neither this work nor my life of the past five years would have been possible.

Contents

Chapter

Figures

Figure

Tables

Table

Chapter 1

Introduction

1.1 The El Niño-Southern Oscillation

1.1.1 El Niño: The Phenomenon

On the wall in the hall of one of this university's buildings are a sequence of images, dated from January, 1980 to February, 1998. The images are monthly averages of the vertical temperature profile of the equatorial Pacific Ocean. At both ends of this tableau, the images show warm water extending eastward to replace the normally cooler waters of the eastern Pacific. These are the El Niño events of 1982-83 and 1997-98, the strongest two of this century.

The news media, apparently lacking anything better to do, latched on to "El Niño" in the fall of 1997. El Niño received the blame for a litany of catastrophes: a major blizzard in Colorado, flooding in California, tornadoes in Florida, floods in coastal Peru, and a forest fire in Sumatra so severe, it blackened the skies, leading to a plane crash and closing airports as far away as Singapore. A commentator even made an "El Niño connection" to the financial crisis that hit Southeast Asia in late 1997. Of course, accompanying every mention of El Niño, the media shows a lack of understanding of what El Niño is.[1]

So what is El Niño? The name itself comes from a change in the ocean currents off the coast of Peru. There is an annual decrease in the westward currents off the Peruvian coast during December and January, accompanied by a warming of the ocean's surface temperature (Wyr75). The westward ocean current transports warm surface water to the western Pacific while also causing the upwelling of colder, subsurface waters in the east.

[1] Philander (Phi90) has a description of actual impacts of El Niño, specifically the one in 1982-83.

Riding that upwelling are nutrients that feed the fish population. When the currents slacken, the surface waters warm as upwelling decreases (Phi90). If this change is severe enough, the resulting loss of nutrients causes fish populations to decline. The Peruvian fisherfolk, noticing a connection between their catch, the warming of the waters, and the time of year, named the phenomenon "El Niño," after the Christ-child (Phi90).

At the beginning of this century, Sir Gilbert Walker noticed an interesting connection between the surface atmospheric pressure at various points across the Pacific Ocean (Bje69). Specifically, when the pressure in Darwin, Australia and other points in the western Pacific was high, the pressure in Tahiti and the central and eastern Pacific was low. The converse is also true. He termed this phenomenon the "Southern Oscillation" and related it to changes in rainfall and wind patterns throughout the tropical Pacific and Indian Oceans (Phi90). In the late 1960's, Bjerknes (Bje66, Bje69) made a connection between the Southern Oscillation in the atmosphere and the El Niño phenomenon in the ocean. The two are now known collectively as "ENSO," the El Niño-Southern Oscillation.

1.1.2 The Physics of ENSO

Bjerknes' key contribution to the field of ENSO physics was his proposal of a feedback mechanism connecting the atmosphere and the ocean. Figure 1.1 shows a schematic of that feedback loop. Consider first the presence of warm ocean waters on one side of an ocean basin and colder water on the opposite side. The warmer sea surface temperatures, or SST, generate rising motion. Additionally, warmer temperature decreases the vertical pressure gradient relative to a region of colder temperature. At some upper level, then, there will be a horizontal pressure gradient accompanied by a flow from the warmer end of the ocean to the colder end [Figure 1.1]. Sinking motion over the region of cold SST and another horizontal pressure gradient at the surface complete the circulation pattern. Consequently, there is a surface wind moving from the colder end of the ocean basin to the warmer one. In the Pacific Ocean, these are the easterly trade winds along the equator. Bjerknes termed the entire equatorial cell containing the trade winds the "Walker Circulation" (Bje69).

Let us return now to Figure 1.1 and consider the right side of the diagram to be the eastern end of an ocean basin, in analogy to the Pacific. The westward winds generate a westward surface ocean current, advecting warm water to the western end of the basin while

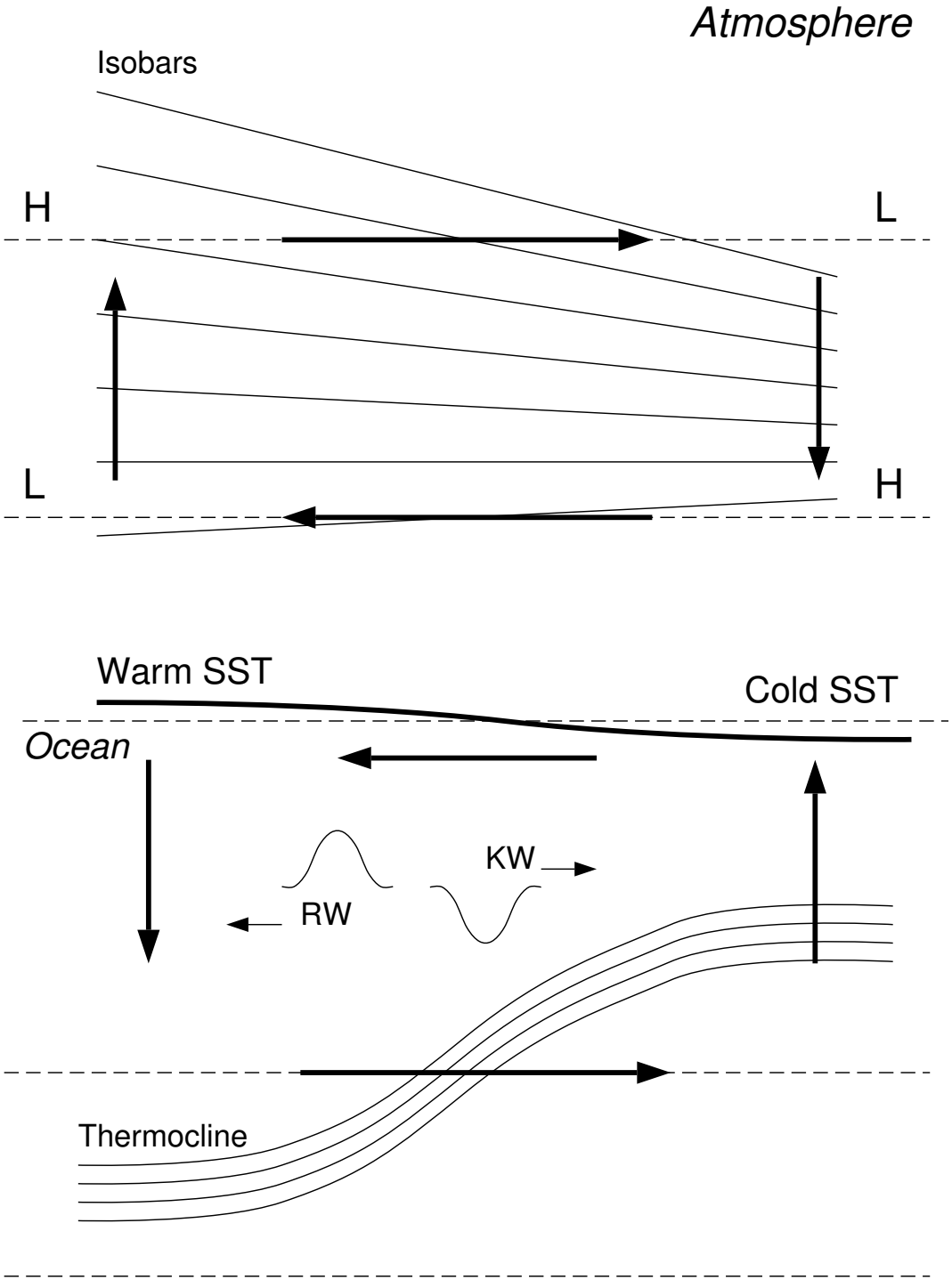

Figure 1.1: Schematic of the basic physics of ENSO. "H" and "L" mark high and low atmospheric pressure, respectively. "RW" and "KW" are abbreviations for "Rossby wave" and "Kelvin wave," respectively.

raising the sea level there. In the western Pacific, this generates the so-called "warm pool" along the equator and raises the sea surface an average of 0.7 meters (Bje66). As noted earlier, the surface current also generates upwelling and downwelling at the eastern and western boundaries of the ocean, respectively, by simple conservation of mass. The circulation is completed by a subsurface current along the thermocline, a steep temperature gradient separating warmer upper-ocean waters from the colder, abyssal waters (Bje69, Phi90). In the Pacific, the thermocline is about 50 meters thick, marking the boundary of an upper layer about 150 meters deep in the west, but surfacing in the east (Phi90). The current thus maintains a sea surface temperature gradient, the same gradient that, ultimately, generates the current. Thus, there is a positive feedback loop between the atmosphere and ocean (Bje66, Bje69).

Wyrtki (Wyr75) notes that a slackening of the surface wind removes the force maintaining the equatorial sea surface slope. He suggests that the resulting propagation of warm water eastward from the warm pool might take place in the form of an equatorial Kelvin wave. Cane and Zebiak (CZ85) comment that the westerly wind bursts common in the Pacific can also excite equatorial Kelvin waves. These Kelvin waves then depress the thermocline in the eastern Pacific, changing the Bjerknes feedback loop's direction. This would lead to an El Niño event. They also remark that ENSO takes place completely in the tropical Pacific. The Pacific Ocean, however, does not experience only El Niño events, but oscillates between two states, the El Niño or warm event, and the corresponding cold event, sometimes termed "La Niña" (Phi90).

In the late 1980's, Graham and White, as well as Schopf and Suarez (GW88,SS88b), published what is now considered the "canonical" description of ENSO physics. Both papers describe El Niño as only one phase of a complete oscillation, just as summer is a single phase of the annual cycle. The cycle begins with some perturbation to the thermocline in the central Pacific, where the system is most sensitive (SS88a, SS88b). The perturbation excites two waves within the thermocline: a westward-propagating equatorial Kelvin wave and an eastward-propagating equatorial Rossby wave. The two waves have opposite signs, i. e. a perturbation produces a downwelling Kelvin wave and an upwelling Rossby wave [or vice versa]. Let us begin with this scenario, indicated, incidentally, in Figure 1.1 by "RW" and "KW." Kelvin waves propagate across the Pacific basin in about 2 to 3 months,

suppressing the thermocline in the east (GW88). This alters the atmospheric circulation, creating more perturbations that generate downwelling Kelvin waves. This leads to the El Niño or warm phase of ENSO. Meanwhile, the corresponding upwelling Rossby waves have been propagating slowly towards the so-called maritime continent on a journey that takes about 9 months (GW88). These Rossby waves reflect off of Indonesia as upwelling Kelvin waves, which propagate rapidly towards South America. The upwelling Kelvin waves raise the thermocline, enhancing the atmospheric circulation which, in turn, generates more upwelling Kelvin waves, along with corresponding downwelling Rossby waves. The system thus enters the La Niña or cold phase of ENSO. The overall period of this oscillation is about 3-5 years (GW88).

In reality, the Kelvin waves reflect from the South American coast as Rossby waves with the same sign. There is, then, an entire train of Kelvin and Rossby waves, each, in sequence, elevating or lowering the thermocline, then reflecting off of the basin boundaries. Naturally, these wave trains dampen out over time, especially upon reflection.

On a final note, there are broader impacts of this oscillation. Due to its effects on atmospheric circulation, the cold phase of ENSO will enhance convection over Indonesia, leading to wetter conditions, while causing drier conditions over coastal South America (RC82). During a warm event, the opposite is true. Precipitation in Indonesia is weaker than average and stronger in coastal South America. During severe warm events, this can manifest itself in severe flooding in Peru and Ecuador.[2] A severe warm event also has connections to weak or absent monsoons in India (SP83, Phi90).

1.1.3 Modeling Efforts

Due to its strong human impact, there is much interest in modeling and forecasting ENSO. To describe ENSO behavior in a compact form, researchers primarily use two indices. The first, the Southern Oscillation Index [hereafter SOI], measures anomalous pressure differences between Darwin, Australia and Tahiti. The second type of index is a mean sea surface temperature [SST] of a predefined region in the equatorial Pacific Ocean, the most commonly used being the SST in the so-called "NINO3" region. This thesis makes heavy

[2] I have seen hypotheses that the ancient Moche civilization of Peru was wiped out by a series of strong El Niño events. Severe flooding wiped out crops, while the cessation of coastal upwelling reduced fish catches.

use of both the SOI and NINO3 SST as representatives of ENSO behavior. I shall return to them in greater detail in chapter 5.

One of the most frequently used models, for forecasts and as a basis for other models, is the two-layer model of Cane and Zebiak (ZC87, hereafter referred to as the CZ model). Cane et. al. (CMZ90, CMZ91) and Bürger et. al. (BZC95a, BZC95b) both analyze the CZ model in detail. Battisti (Bat88) builds upon the CZ model, and later, with Hirst (BH89), linearizes it. They use the linearized version of CZ to derive the simple delay equation model of ENSO first proposed by Suarez and Schopf (SS88b). This model of ENSO as a delay equation, a well-studied class of ordinary differential equations, will return in later discussion.

In addition to the CZ model, several researchers have used a full global circulation model [GCM] to study ENSO (Mee90, BLKR91). GCM's are highly complex, modeling not only the tropical Pacific, but the whole Earth. Consequently, there exist several "intermediate" models, hybrids that do not make some of the simplifying assumptions of the CZ model yet do not attempt to describe the entire globe (Nee90, Kle91, JN93a, JN93b, JN93c). At the other extreme is the model of Vallis (Val88). This highly simplified two-layer model, cast into a form similar to the famous "Lorenz butterfly" model (Lor63), will be used in the work of chapter 6.

1.1.4 Predictability and the "Spring Barrier"

One of the key aspects of forecasting and model analysis is the study of how well a particular model or forecasting technique predicts the behavior of nature. In addition to examining the predictability of models and algorithms, one can also examine the "inherent" predictability of a dataset using lagged autocorrelations. Troup (Tro65) and Wright (Wri79, Wri85) did such an examination, using the SOI and a form of lag autocorrelation called persistence. As defined in their works, persistence is a cyclostationary statistic. Cyclostationary statistics, the appropriate statistics to use when a dataset has a natural periodic cycle (Gla61, HN96), are always a function of an initial — i. e. leading — phase of the natural periodic cycle in the data. Thus, cyclostationary statistics of data with an annual cycle could be a function of seasons, months, days, etc. The persistence, being a lag autocorrelation, is a function of both the initial phase and, of course, a lag.

Figures 1.2 and 1.3 contain "persistence maps" of the SOI and the SST in the NINO3 region of the Pacific ocean, respectively. The maps are similar to ones shown by Wright (Wri79, Wri85) and Torrence and Webster (TW98). Note the rapid decrease in persistence during the boreal spring, at a leading phase of 0.25-0.5 yrs. Webster and Yang (WY92) termed this decrease a "predictability barrier." Torrence and Webster (TW98) suggest that this "barrier" occurs as the the climate system makes a transition from one state to another during the boreal spring. Xue et. al. (XCZB94) additionally show that the barrier is associated with low variance in the data during the spring.

The initial suggestion by Webster and Yang (WY92) that the "barrier" may indicate an inherent limit to ENSO predictability caused some controversy. There has been work suggesting that this decrease in persistence may not necessarily hinder forecasting (CZBC95). Some researchers may have questioned the physicality of this "predictability" barrier. Could it not also be an artifact of the statistics? Before one can answer that question, however, one needs to better understand the properties of cyclostationary statistics in general. The presence of such a shape, hereafter called the "persistence barrier" or simply "barrier," could indicate some interesting underlying behavior.

Although it is not yet clear how the barrier relates to the ability to forecast ENSO, it is nonetheless a statistical property of the system. As such, it warrants further investigation, for deeper understanding of persistence barriers may provide further insight into the causes and dynamics of ENSO. There may even be a connection between persistence barriers and other measures of predictability. Even if there is not such a connection, statistical features of a system can still provide useful benchmarks for comparing models with data. We shall return to these issues more fully in chapters 4-5.

1.2 Chaos Strikes El Niño

1.2.1 About Chaos

The term "chaos" is used to describe any system with aperiodic, irregular behavior that additionally contains no stochastic forcing. Often, the equations of motion for such a system are known exactly, even though the behavior is irregular. Much of the modern research into this field began with the fundamental work of Lorenz (Lor63), who examined

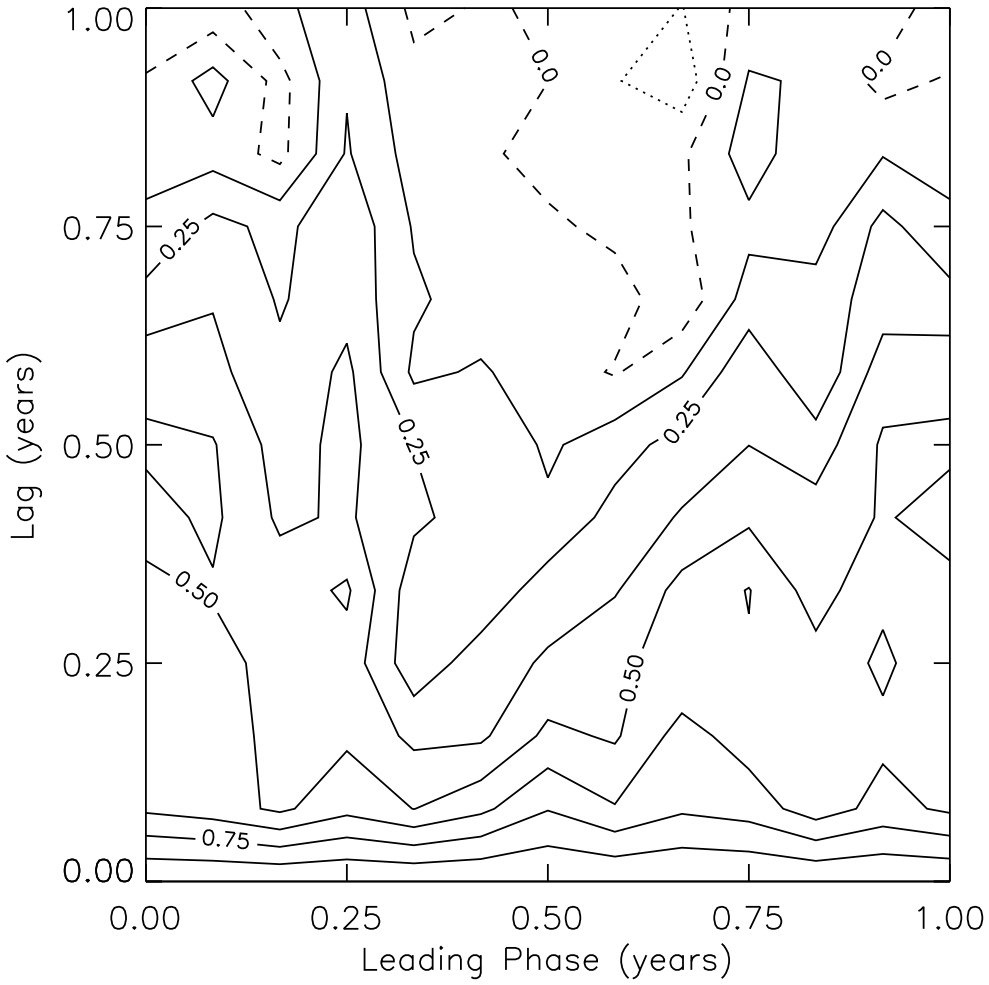

Figure 1.2: Persistence map of the SOI. Contour intervals are in steps of $\frac{1}{8}$. Solid contours denote positive values of persistence, while dotted contours denote negative values. The zero-contour is dashed.

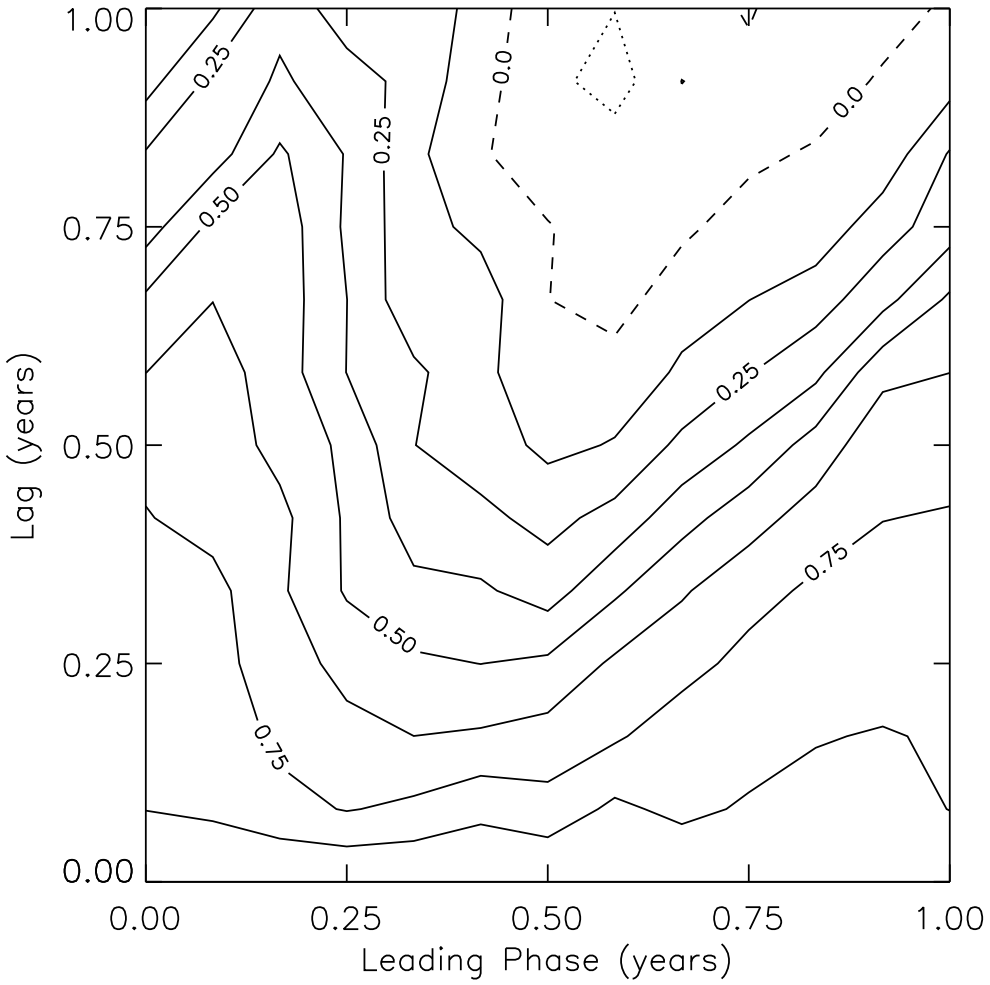

Figure 1.3: Persistence map of the NINO3 SST. The contours are plotted in the same fashion as in Figure 1.2.

the model:

$$\dot{x} = \sigma (y - x), \tag{1.1}$$

$$\dot{y} = -xz + rx - y, \tag{1.2}$$

$$\dot{z} = xy - bz, \tag{1.3}$$

a simplified, truncated model of atmospheric convection. These equations are now well-known as the "Lorenz equation" or the "Lorenz butterfly." Lorenz (Lor63) found that these equations exhibited aperiodic behavior and high sensitivity to initial conditions. That is, even tiny errors in an initial condition can cause drastic changes in the behavior of the system.

Zeng et. al. (ZPE93) published a good summary of the field of "chaos theory," more formally known as nonlinear dynamics, and its applications to the atmospheric sciences. They describe the early work of Poincaré at the beginning of the century, work not examined again until the 1960's. They discuss the definition of a "strange attractor" by Ruelle and Takens, and of the various routes to chaos a dynamical system can take. The general description of the theory of dynamical systems that follows comes from both Zeng et. al. (ZPE93) and Tabor (Tab89).

Consider some physical system, containing many variables that depend on time and, possibly, spatial location. In the theory of dynamical systems, each time-dependent variable of the system constitutes a dimension of a linear vector space, known as "phase space." If the system also depends on location, then the dimensions of phase space are all of the variables evaluated at all possible positions. As the system evolves through time, it traces out a trajectory through phase space. Sometimes, a system's trajectory may stay in one specific region of phase space. That region is termed an "attractor."

It can be very difficult to think about trajectories through phase space with a large number of dimensions. Poincaré thought of a simpler way to envision phase space. Consider a 2-dimensional "slice" of a 3-dimensional phase space, intersecting the trajectory. It may intersect a fixed region of phase space, or it may intersect the trajectory at "regular intervals" in phase space. The intersection points create an image, the so-called "Poincaré Section."[3]

[3]In general, a Poincaré Section is an $(N-1)$-dimensional subspace of an N-dimensional phase space.

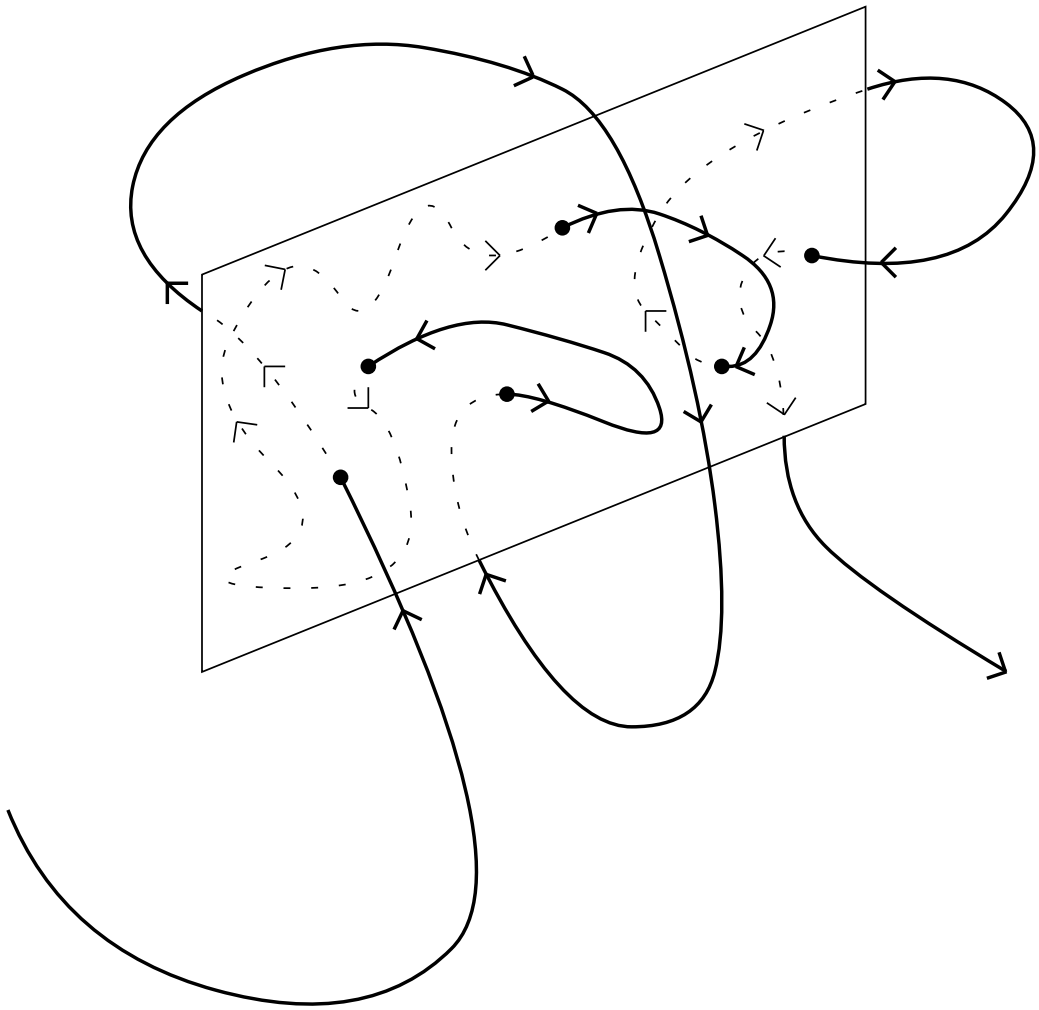

Figure 1.4: Schematic of a Poincaré Section for a 3-dimensional phase space, taken at a fixed location.

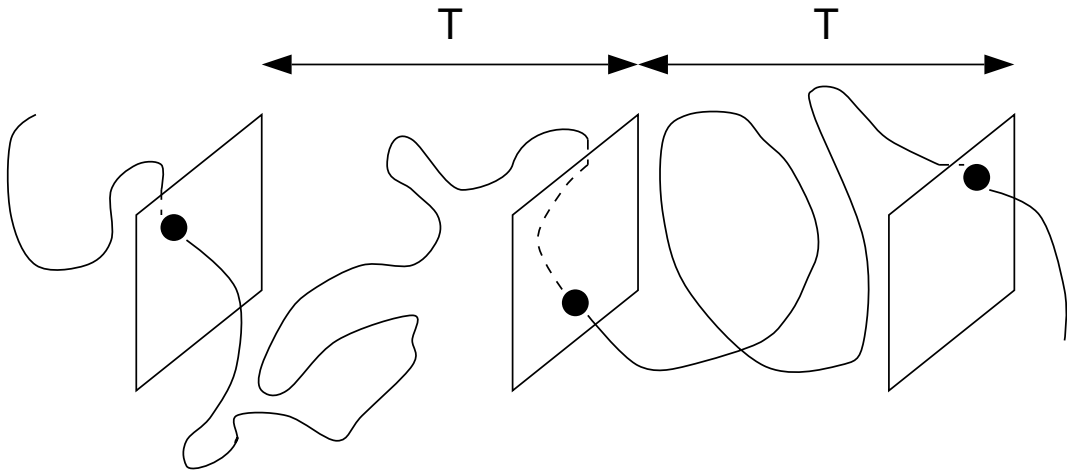

Figure 1.5: Schematic of a Poincaré Section for a 3-dimensional phase space, taken at regular intervals, "T." Note that "T" is not necessarily a time; it may be a distance in phase space.

The Poincaré Section of a periodic system shows an attractor composed of a finite set of points. Other attractors appear as circles or other simple, closed curves. The Poincaré Section of a strange-attractor, however, exhibits a complex, folded structure reminiscent of a fractal.[4]

Portraits of phase space are not the entire story of dynamical systems theory. The equations of motion for a system may contain parameters, independent of time, that describe physical properties of the system, such as damping coefficients, coupling strengths, driving forces, etc. Altering these parameters alters the behavior of the system. One can envision the possible values of every parameter also forming a space, a "parameter space," if you will. Different points in parameter space correspond to different behaviors of the system, i. e. different attractors and trajectories in phase space. It is not uncommon to examine how the attractor in phase space changes as one alters the parameters "along a path" through parameter space. Such a path can describe a "route to chaos."

To elucidate the previous paragraph, consider the following example. Suppose that a system has two different attractors in its phase space. Suppose, too, that one of the system's parameters, say, a damping coefficient, changes the distance between the two attractors. As one alters the damping coefficient, the two attractors eventually overlap. The system now alternates between the two attractors in an aperiodic fashion; the system has become chaotic. The scenario described here is also known as a "crisis" and constitutes a route to chaos.

Another route to chaos, related to the aforementioned "crisis," is the Pommeau-Manneville route. Under this route to chaos, a single attractor splits into two overlapping ones. Both of these types of chaotic systems are known as "intermittent chaos," for the system will exhibit intervals of periodic behavior punctuated by irregular behavior. Another route is the so-called period-doubling bifurcation. As one changes a parameter of the system undergoing periodic motion, the period doubles repeatedly. After going from twice its original period to four times its original period and so on, the system eventually becomes aperiodic, i. e. chaotic. The interested reader can read about these and other routes to chaos in Zeng et. al. (ZPE93) Tabor (Tab89), and the sources they reference.

There is another route to chaos, the quasi-periodic route to chaos. To simplify the discussion, consider a simple system with only two parameters. Suppose that, as one changes

[4]See Tabor (Tab89) for a complete explanation of why this is so.

the parameters, the period of the system remains fixed at a particular value, then "jumps" to a new value, at which it remains fixed. This behavior is known as "mode locking," and a one-dimensional plot of the system's period as a function of one parameter forms a "Devil's Staircase" (JBB84, BBJ84, Bak86). Between any two mode-locked "steps" of the staircase, there are an infinite number of other steps, hence the name "Devil's Staircase." If one were to plot the system's period as a function of all parameters, the mode-locked regions would form "petals," often referred to as "Arnol'd Tongues" or a "Devil's Flower" (Bak86).

Between the tongues, there are regions of non-mode-locked behavior. The system undergoes regular motion, but in a fashion that is "quasiperiodic." Quasiperiodic behavior is not chaotic; it is not irregular, but it is non-repeating. The function $f(t) = \cos(\pi t) + \cos(11t)$ is quasiperiodic. It comes close to repeating its previous values, but never does so exactly. Any function, in fact, that contains the sum of two other functions the ratio of whose frequencies forms an irrational number is quasiperiodic.

Along a particular, "critical" curve through parameter space, the tongues meet. That is, the system experiences mode locking, regardless the parameter values, and the Devil's Staircase is said to be "complete" (JBB84, BBJ84, Bak86). Beyond the critical curve, the tongues "overlap." The system, faced with being in two different mode-locked states simultaneously, alternates between the two in an irregular fashion.

The key point to the previous paragraph is the idea of a "critical curve" — or even a critical surface — in parameter space. The critical curve is simply a specific set of parameter values. Parameters to one side of the critical curve make the system periodic or quasiperiodic. Parameters on the other side produce chaos. At parameter values along the critical curve, the periods of the system collectively form the "complete Devil's Staircase," periodic, mode locked steps between which lie an infinity of other mode locked steps. In fact, this "infinity of steps" has a fractal structure to it (JBB84, BBJ84). The presence of a Devil's Staircase is a hallmark of the quasiperiodic-route to chaos.

1.2.2 El Niño on the Devil's Staircase

Surprisingly, several researchers have found examples of the quasiperiodic route to chaos in models of ENSO. Tziperman et. al. (TSCJ94) used the delay-equation model of Schopf and Suarez (SS88b), adding to it a periodic forcing term. They found transitions from

a stable periodic state to a quasiperiodic regime. Tuning the model's parameters further led to frequency-locked behavior, followed by chaos. Jin et. al. (JNG94) used the more complex Jin-Neelin model (JN93a, JN93b, JN93c), a set of four coupled partial differential equations. Using this model, they constructed a Devil's Staircase using two of the model's parameters.

Because a delay equation is a fairly simple system, one may wonder if the quasiperiodic chaos seen by Tziperman et. al. (TSCJ94) isn't actually a property of the model alone. The work of Jin et. al. (JNG94) would seem to indicate otherwise. Tziperman, however, provides a more convincing example. In a later paper, Tziperman and colleagues (TCZ95) perform computations using the full CZ forecast model (ZC87). Using the equivalent of the SST in the NINO3 region of the model, they perform a phase space reconstruction to examine the frequency behavior of the model's output (TCZ95). First, they see chaos even when the CZ model contains no seasonal forcing. Second, they demonstrate a quasiperiodic route to chaos in the model.

Other researchers have examined examples of chaos in other ENSO models. Chang et. al. (CWLJ94) also use a complicated ENSO model. They find a sequence of period-doubling bifurcations leading to chaos. Wang and Fang do something more interesting (WW96). They derive an ENSO model from "first principles," then simplify it to discern the "essential" dynamical system. Upon this system, they perform a "stability analysis" [see chapter 6 or section B.2 for more information about stability analysis] and compare it to data to verify that it does indeed capture the "basic dynamics" observed in nature. In addition to finding periodic and chaotic behavior, they also find the largest growth of perturbations in the boreal spring, an intriguing connection to the "spring barrier" described in section 1.1.4.

1.2.3 Noisy Niños

A major issue is missing from the previous discussion of chaotic behavior and ENSO. A chaotic system is distinguishable from a stochastic system only when the chaotic behavior is "low-order," i. e. the system's chaotic attractor is an object with few dimensions (Wei). Clearly, the Pacific Ocean and the atmosphere above it do not constitute a low-dimensional system! For this reason, some researchers feel that ENSO is driven by noise. Such works are not the focus of this dissertation and are presented as examples of an alternate viewpoint.

Balmaseda et. al. (BAD94) use an ENSO model with stochastic atmospheric forcing

based on data. Using a linear stochastic model, Xue et. al. (XCZB94) examine issues of predictability, with connections to the persistence barrier of section 1.1.4. Penland and Magorian (PM93) and Penland and Matrosova (PM94) develop an ENSO forecast technique by applying linear inverse modeling to another linear stochastic model of ENSO. With this model, Penland and Sardeshmukh (PS95) find a forecast range of 15 months. The interested reader should also examine the work of Blumenthal (Blu91), who develops some techniques used by these stochastic modeling efforts.

1.3 Stability, Instability, and Predictability

1.3.1 Lyapunov Exponents

There are other ways to describe predictability besides correlations and the forecast skill of models. Lorenz (Lor65) stated three possibilities for the lack of forecast skill in weather predictions:

(1) The system is not deterministic.

(2) Observations used to initialize the models are not accurate enough.

(3) The models and forecast techniques do not describe the system adequately enough.

Correlations can assist the modeler to discern how well his models match the behavior of the actual physical system. The first case, the presence of stochastic forcing, certainly occurs in nature at the quantum level. While predictability behavior at small scales has some effect on larger scale dynamics (Lor69), it is not clear how much, if any, of the quantum-level random forcing propagates to macroscopic scales.

In 1969, Lorenz described the predictability of a system based on the second aforementioned case, namely the effects of errors in initial conditions (Lor69). He formulated three types of predictability. In systems of the first type, initial errors either remain constant or shrink over time. Such systems, e. g. periodic systems, are predictable for all time. Systems of the second and third types exhibit error growth, and are therefore predictable over only a finite length of time. If one can always extend this period of predictability by shrinking the size of the errors, the system fulfills Lorenz' second type of predictability. For the third type of predictability, improving the accuracy of the initial conditions eventually

ceases to have an effect on the predictability of the system. The period of time over which the system is predictable is inherently fixed by the dynamics.

Since Lorenz's descriptions of predictability and its relation to error growth, forecaster often look for "optimal perturbations" in their models (TL95, MK96). These optimal perturbations are directions in the model's phase space, directions in which errors grow the most rapidly. They bear some relation to another quantity, the Lyapunov exponents. Consider a volume of perturbations to a trajectory through phase space, as shown in Figure 1.6. As the system evolves, the volume of perturbations distorts, growing in some directions and contracting in others. The Lyapunov exponents measure the exponential rates of growth and contraction of this volume.

The Lyapunov exponents actually come in three varieties. When the exponents measure growth rates of volumes over the whole of a strange attractor, they are termed "global" (BGGS80a, amongst numerous others). One of the definitions of a "chaotic system" is a system which has at least one positive nonzero Lyapunov exponent. That is, the system experiences exponential error growth.

There is also a version that measures these divergence rates locally over a finite region of the attractor (Nes89). These are of great use in real systems, for which one only needs the error growth behavior during, say, the next week, not until the end of the universe. One can also define "instantaneous" Lyapunov exponents that measure divergence rates "at a point" in phase space (VGR90). I shall return to all of these in great mathematical detail in chapter 2.

1.3.2 Applications to the Atmospheric Sciences

The concept of Lyapunov exponents has numerous applications within the atmospheric sciences. As noted in the previous section, forecasters often use the so-called optimal perturbations to determine the error growth characteristics of their forecast models. These optimal perturbations are, in fact, the so-called "Lyapunov eigenvectors" (MK96) defined later in chapter 2. Additionally, Trevisan and Legnani (TL95) show that measures of error growth based on the optimal perturbations converge to the Lyapunov exponents. Additionally, the optimal perturbations are related to techniques used to reconstruct the phase space of a system from data, specifically the so-called principle-component analysis (ZPE93).

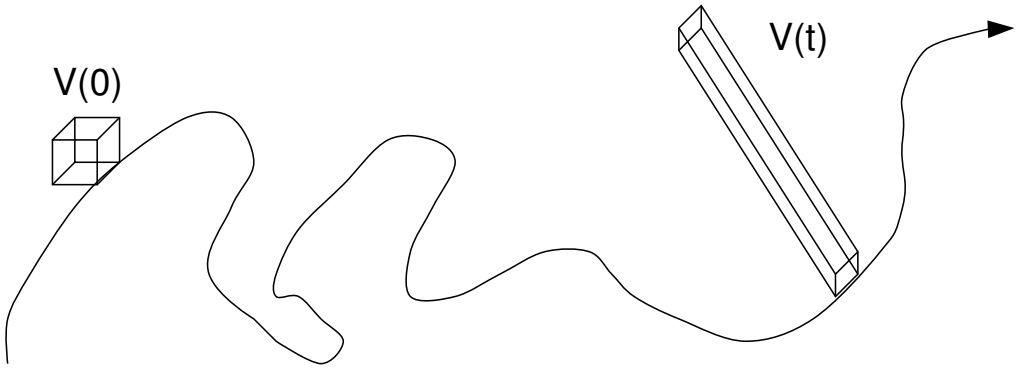

Figure 1.6: Distortion of a volume of perturbations, $V\,(0)$, as it evolves through phase space.

Principle-component analysis is also known as a singular value decomposition or an empirical-orthogonal function analysis. Zeng et. al. (ZPE93) discuss this in greater detail.

Moore and Kleeman (MK96) perform an extensive analysis of the optimal perturbations of the Kleeman model (Kle91). They find a connection between the optimal perturbations and the onset of ENSO events. They also note that the largest error growth for the Kleeman model occurs during the boreal spring, the time of the predictability barrier of section 1.1.4. In a similar vein, Yoden and Nomura (YN93) discuss the relationship between Lyapunov exponents and optimal error growth. They then apply the exponents to such problems in fluid dynamics as zonal flow and Rossby waves.

Vannitsem and Nicolis (VN94) examine the Lyapunov exponents of a thermal convective model. The model is a set of partial differential equations evaluated at over 100 spatial grid points. Consequently, the phase space for this system has a rather large dimension, yielding a correspondingly large number of Lyapunov exponents. That they compute the Lyapunov exponents for such a high-dimensional system is itself rather impressive. Using the exponents, they find that error growth and predictability are due to chaos from several interacting spatial scales. Instabilities at small scales generate chaos while compounding the instabilities at larger scales. Their description of this interaction has self-similar, almost fractal qualities, a fascinating concept.

More closely related to the work of this thesis is the paper by Lepri and colleagues (LGPA93). They examine the Lyapunov exponents of delay equations. Recall from section 1.2.2 that Tziperman et. al. (TSCJ94) used a delay equation model of ENSO to examine manifestations of low-order chaos in ENSO. Lepri et. al. (LGPA93) note that most of the Lyapunov exponents scale as the inverse of the delay time. They also reveal certain special cases in which the largest Lyapunov exponent does not scale. This is an intriguing result: the magnitude of the Lyapunov exponents, the descriptors of error growth for the system, is a function of a time scale of the delay equations, the delay time.

Abarbanel and Lall (AL96) use Lyapunov exponents to examine the predictability timescales of the volume of the Great Salt Lake. The perform a phase space reconstruction on a 150-year data set using the delay method proposed by Takens.[5] The data set contains about 3600 points total. Nevertheless, Abarbanel and Lall are able to perform a reasonable

[5]Zeng et. al. (ZPE93) also discuss this method.

phase space reconstruction. Although some attempts at phase space reconstruction fail unless the dataset contains a vast number of points, Abarbanel and Lall note that the number of data points is not nearly as important as how well the data samples the attractor. In their case, the 3600 measurements of volume changes in the Great Salt Lake create a well-formed attractor (AL96).

The previous paragraph contains a highly important point. There have been numerous attempts to compute measures of chaos from atmospheric and oceanic data, especially fractal dimension (Lor91, ZPE93). Based on these measures, one would get the impression that the entire Earth system was a low-dimensional chaotic system. Lorenz (Lor91) demonstrated why all of these computations of fractal dimensions were vastly underestimated. He examined a system comprised of numerous coupled chaotic oscillators, most of which were weakly coupled to one another. The phase space of this system thus contains a low-dimensional subspace in which coupling is strong, i. e. there is a locally strong fixed point in a region of the attractor. Lorenz integrated this system and estimated the fractal dimension from one of its variables. His results show that some variables produce estimates of dimension much smaller that the true size of the phase space. There are some variables, however, that nevertheless yield a reasonable estimate (Lor91).

The implications for ENSO are immediately clear. Perhaps these manifestations of low-dimensional chaos in the models are only local. That is, in the vast phase space of the Pacific atmosphere-ocean system, there is some low-dimensional region of the attractor containing a strong fixed point. Lorenz feels that this is the case for examples of low-order chaos in the Earth system and urges caution (Lor91). One should not immediately discount such manifestations as nonphysical, but should examine them from a new perspective, as descriptors of local behavior.

1.4 Motivation

Thus far, we have seen how models of the El Niño-Southern Oscillation exhibit features of low-dimensional chaos. As revealed in the previous section, such manifestations of low-dimensional behavior in high-dimensional dynamical system may indicate localized behavior on the system's strange attractor. Despite the numerous papers discussing low-

order chaos in ENSO models, none have yet made any firm connections to physical data. The only connections made are statements of how well a given set of equations model ENSO data. It would certainly be more satisfying if some form of data analysis revealed low-dimensional behavior in the Southern Oscillation Index [SOI] or the sea surface temperature in the "NINO3" region of the equatorial Pacific Ocean [NINO3 SST].

Section 1.1.4 discussed the so-called "predictability barrier" seen in the persistence of both the SOI and NINO3 SST. However, there is as yet no connection between persistence and predictability as defined in a dynamical systems sense, i. e. the Lyapunov exponents. In fact, no one appears to have examined the properties of the statistic termed "persistence." A barrier-like feature in the persistence of a time series, a feature such as those seen in Figures 1.2 and 1.3, may indicate interesting dynamics. Furthermore, those papers that attempt to prove or disprove the presence or absence of a barrier (XCZB94,CZBC95,BDA95, TW98) use only visual comparisons of persistence. A more quantitative, less subjective measure of barrier-like features in persistence would be preferable.

These ideas raise a question: Could the "barriers" in persistence somehow be connected to low-order chaotic behavior? Such a connection would neatly connect the "predictability barrier" to dynamical systems behavior while simultaneously providing a data-based manifestation of low-dimensional chaos. Making some progress towards answering this question is the focus of this research.

I shall begin this research with two background chapters. Chapter 2 examines the Lyapunov exponents, including the global, finite-time, and instantaneous exponent spectra, as well as describing the Lyapunov eigenvectors. A secondary purpose of this chapter is a unified summary of the numerous, nontrivial papers that originally defined these quantities. In chapter 3, we will examine issues surrounding the fitting of data to nonlinear models. Section 3.2 will discuss genetic algorithms and their uses with other nonlinear optimization techniques. Section 3.3 provides a definition of a correlation-like quantity and confidence intervals as applied to the problem of nonlinear least-squares fitting.

Chapter 4 begins the primary focus of the thesis with a detailed analysis of persistence. Section 4.2 presents a formal definition of cyclostationary statistics, the class of statistics to which persistence belongs. Sections 4.2.2-4.2.5 then calculate the persistence for a progression of simple but increasingly complex time series. Not all time series have a

persistence that resembles the "persistence barrier" seen in ENSO data. Using the results of these sections, I shall describe a set of criteria that a timeseries must fulfill in order to produce persistence "barrier." Section 4.3 defines a model containing the key features of the shape known as the "barrier." Henceforth, I shall use the term "persistence barrier" or "barrier" to describe this shape. A nonlinear least-squares fit to this model provides a quantitative measure of the properties of persistence barriers in data.

In chapter 5, I will apply the results of chapter 4 to the El Niño-Southern Oscillation.[6] Section 5.2.1 uses the aforementioned measure to quantify the barriers in the NINO3 SST and the SOI. Section 5.2.2 numerically examines the conditions for producing a barrier, revealing a connection between persistence barriers and phase-locking to the annual cycle. Finally, in section 5.3, I use the measure of barrier properties to study the interdecadal variability of ENSO.

The connection between persistence barriers and phase-locking to the annual cycle is tantalizingly similar to the quasiperiodic route to chaos in systems such as the damped forced pendulum (JBB84,BBJ84). Because a damped forced pendulum is the simplest physical system exhibiting the quasiperiodic route to chaos, chapter 6 begins with a detailed description of this system. We shall learn that the pendulum contains other routes to chaos, such as period-doubling and intermittency. As noted in section 1.2.2, ENSO models also exhibit such routes to chaos. I shall also examine the ENSO model of Vallis (Val88), a version of the Lorenz equations [Eqs. (1.1-1.3)] with periodic forcing. The chapter concludes with an attempt to connect the instantaneous Lyapunov exponents to persistence using these two dynamical systems.

[6]The material in chapters 4 and 5 has been submitted to the Journal of the Atmospheric Sciences (WW97a, WW97b). The papers were co-authored with Dr. Jeffrey B. Weiss.

Chapter 2

Lyapunov Exponents

2.1 Concepts of Stability and Predictability

2.1.1 Introduction

The study of stability is a fundamental feature of dynamical systems analysis. In cases where one cannot integrate the equations of motion, stability analysis can provide an approximate picture of the system's behavior. Stability analysis is concerned with the behavior of linearized perturbations to certain "special" initial conditions of the system. It is a topic that chapter 6 returns to. A related concept is the effect of linearized perturbations, or "errors," to any initial condition of a system under scrutiny. One typically categorizes error growth behavior into three basic types: exponential growth, exponential contraction, and some other form of growth or contraction. The **Lyapunov exponents** are simply these exponential growth rates. Positive exponents indicate an instability, for perturbations grow at an exponential rate. In contrast, the exponential contraction of perturbations implies a stability.

There are two types of Lyapunov exponents. One type measures the global stability of an attractor of the system. A second, local version measures the stability in a specific region of an attractor of the system. The local exponents are useful for, say, examining how well a model will make a forecast, or which variables of the model exhibit the greatest instability. As for the global exponents, one definition of "chaos" in a dynamical system is that the system have at least one positive global Lyapunov exponent.

For either case, the Lyapunov exponents only describe **linearized** stability, the evolution of a perturbation under approximate equations of the motion. Two initial condi-

tions evolving under the full, nonlinear dynamics of the system cannot diverge exponentially forever. While the initial divergence of these two initial conditions may be exponential, they can never be further apart than the size of the attractor of the system. Therefore, Lyapunov exponents are best used locally in phase space when examining predictability, the goal of this work.

The use of local Lyapunov exponents for examining the predictability of a system is only one reason for the present chapter. The second reason has to do with the literature on Lyapunov exponents. In addition to almost every author using their own notation, several papers independently define the same quantity in different forms. There are some fundamental papers, such as those by Oseledec (Ose68) and Benettin et. al. (BGGS80a,BGGS80b). However, these works are heavily mathematical and understanding them is not trivial. Therefore, this chapter shall also serve as a survey of the field and its techniques.

The chapter begins with a description of notation in sections 2.1.2 and 2.2.1. Section 2.2 presents formal definitions of the Lyapunov exponents, including the global Lyapunov exponent spectrum. In sections 2.2.4-2.2.5, we will examine the effect of reorthonormalization and lay the theoretical groundwork for both computing the Lyapunov exponents and defining local Lyapunov exponents. Building upon this groundwork, section 2.3 defines two kinds of local Lyapunov exponent spectra. The chapter concludes with a description of issues surrounding the numerical computation of the local Lyapunov exponents.

2.1.2 Dynamical Systems: Notation

The basis for the work of this chapter, as well as parts of chapter 6 and appendix A, is a notation commonly used in the analysis of dynamical systems (Tab89,NL92,BG90). One takes the equations of motion, the differential equations describing the system, and casts them into the form:

$$\dot{\vec{x}}\left(t\right) = \frac{d\vec{x}}{dt}\left(t\right) = \vec{F}\left(\vec{x}\left(t\right)\right), \tag{2.1}$$

where $\vec{x}\left(t\right)$ is a vector containing the various variables underlying the dynamics of the system, and $\vec{F}\left(\vec{x}\left(t\right)\right)$ describes the differential equations governing them.

A concept commonly used in relation to Eq. (2.1) is that of phase space (Tab89, NL92,BG90). Let us denote an N-dimensional phase space with the symbol $\mathcal{H}^{[N]}$, using the

brackets to distinguish dimensions from exponents. $\mathcal{H}^{[N]}$ is simply the vector space defining all mathematically possible values of $\vec{x}\,(t)$. Individual $\vec{x}\,(t)$ are thus points in the phase space, $\mathcal{H}^{[N]}$. Contained in $\mathcal{H}^{[N]}$ is a manifold — an N-dimensional volume — defining all of the physical values of $\vec{x}\,(t)$, the possible values determined by the dynamics. It is customary to refer to a path from $\vec{x}\,(t_0)$ to $\vec{x}\,(t_1)$ along this manifold as a trajectory through phase space.

Writing a dynamical system in the form of Eq. (2.1) permits one to examine the effect of perturbations, $\delta\vec{x}\,(t)$. Using a Taylor expansion of (2.1), one can show (Tab89,NL92, BG90):

$$\delta\dot{\vec{x}}\,(t) = \mathbf{J}\left(\vec{x}\,(t)\right)\cdot\delta\vec{x}\,(t)\,, \tag{2.2}$$

where

$$\begin{aligned} \mathbf{J}\left(\vec{x}'\right) &= \frac{\partial\vec{F}}{\partial\vec{x}}\left(\vec{x}'\right), \\ \Longleftrightarrow \quad J_{ij}\left(\vec{x}'\right) &= \frac{\partial F_i}{\partial x_j}\left(\vec{x}'\right), \end{aligned} \tag{2.3}$$

is known as the Jacobian matrix. One can use the eigenvalues of the Jacobian to describe the stability of the system in a linear sense (Tab89, NL92, BG90). A second use comes from integrating (2.2):

$$\delta\vec{x}\left(T+t_0;\,\vec{x}\,(t_0)\right) = \mathbf{M}\left(T;\,\vec{x}\,(t_0)\right)\cdot\delta\vec{x}\,(t_0)\,, \tag{2.4}$$

where $\mathbf{M}\left(T;\,\vec{x}\,(t_0)\right)$ is the linear operator that evolves an initial perturbation, $\delta\vec{x}\,(t_0)$, associated with a point in phase space, $\vec{x}\,(t_0)$, to a new time, $T+t_0$. It is easy to show that

$$\mathbf{M}\left(T;\,\vec{x}\,(t_0)\right) = \exp\left(\int_{t_0}^{T+t_0}\mathbf{J}\left(\vec{x}\,(t')\right)\,dt'\right). \tag{2.5}$$

Using Eqs. (2.4) and (2.5), one can derive the following useful property of $\mathbf{M}$:

$$\mathbf{M}\left(T+S;\,\vec{x}\,(t_0)\right) = \mathbf{M}\left(T;\,\vec{x}\,(S+t_0)\right)\cdot\mathbf{M}\left(S;\,\vec{x}\,(t_0)\right). \tag{2.6}$$

Upon seeing (2.5), one might argue that the notation "$\mathbf{M}\left(T;\,t_0\right)$" is better, being more compact. I do not like this convention. It implies a dependence solely on time. In Equation (2.2), there are two initial conditions, $\delta\vec{x}\,(t_0)$, the initial perturbation, **and** $\vec{x}\,(t_0)$, the initial conditions on the true dynamics. $\mathbf{M}$ thus contains an implicit dependence on

$\vec{x}(t_0)$. One might then ask, "Why not use $\vec{x}(T+t_0)$ in the notation for the arguments of $\mathbf{M}$ instead of T? Isn't there also an implicit dependence on $\vec{x}(T+t_0)$ in $\mathbf{M}$?" The answer to the latter question is, "Not necessarily."

Equation (2.2) describes the **linearized** dynamics of perturbations, not their true dynamics. While $\delta\vec{x}(t_0)$ may be the distance between two points in phase space, $\vec{x}_1(t_0)$ and $\vec{x}_2(t_0)$, evolving $\delta\vec{x}(t_0)$ using (2.2) or (2.4) may remove this connection. In other words, even though $\delta\vec{x}(t_0) = \vec{x}_2(t_0) - \vec{x}_1(t_0)$, $\delta\vec{x}(T+t_0;\, \vec{x}(t_0))$ from Eq. (2.4) won't necessarily equal $\vec{x}_2(T+t_0) - \vec{x}_1(T+t_0)$, especially not for $T \gg 0$. Hence, the notation "$\mathbf{M}(T;\, \vec{x}(t_0))$" indicates that the new perturbation depends on the initial point, $\vec{x}(t_0)$, of the true dynamics and the evolution timespan, T. This is also the reason for using the notation "$\delta\vec{x}(T+t_0;\, \vec{x}(t_0))$" to indicate the new perturbation generated by evolving $\delta\vec{x}(t_0)$ with $\mathbf{M}$. Lorenz and others (Lor65, Lor84, YN93) refer to $\mathbf{M}$ as the "error matrix," while Goldhirsch et. al. (GSO87) call it the "stability matrix." Physically, $\mathbf{M}$ is the linearized perturbation evolution operator, while the Jacobian is the matrix most associated with stability. I shall therefore avoid the more confusing terms as "error matrix" and "stability matrix."

All of the work in this chapter is for "flows." That is, time is a continuous variable. There is another way to express a dynamical system, namely as a mapping (Tab89, NL92, BG90, BBJC84, and numerous others). For mappings, time is discrete, and Equations (2.1), (2.2) and (2.4) become

$$\vec{x}_{n+1} = \vec{F}(\vec{x}_n), \tag{2.7}$$

and

$$\delta\vec{x}_{n+1} = \mathbf{J}(\vec{x}_n) \cdot \delta\vec{x}_n, \tag{2.8}$$

where the Jacobian matrix is still defined as in Eq. (2.3). Notice that there is no longer any need for an integrated form of the Jacobian. Although unused in this chapter, later chapters will refer to these equations for mappings.

2.1.3 Predictability

Before delving into the esoteric intricacies of Lyapunov exponents, let us briefly return to the discussion of predictability in section 1.3.1. Lorenz (Lor69) defines three types of predictability in terms of the behavior of errors. In the first type [Figure 2.1],

errors remain constant or shrink over time. There is an obvious connection to negative Lyapunov exponents, especially if the rate of contraction is exponential. Thus, a system with only negative Lyapunov exponents describe behavior that is predictable over all time. The presence of a null-exponent may indicate that errors remain constant, in which case the system is still predictable over all time.

The second type of predictability defined by Lorenz describes errors growing over time [Figure 2.2]. The amount by which they grow depends on the size of the initial error. Again, there is a clear connection to Lyapunov exponents. When the errors grow exponentially, the system will have at least one positive Lyapunov exponent. A null-exponent may also indicate error growth, but in a non-exponential fashion. Thus, a system with at least one positive Lyapunov exponent will be predictable over only a finite length of time, a length of time that scales with the size of the errors.

The third type of predictability is a variant of the previous one. Errors still grow over time, but they do not scale with the size of the initial error, as show in Figure 2.3. After some length of time, the error growth ceases to depend on the initial size of the error. There are several possible connections to Lyapunov exponents. A system with a global Lyapunov exponent equal to $+\infty$ will exhibit error growth that does not scale with the size of the error. A stochastic system will exhibit this property. Finally, if the linearized error growth exceeds the size of the attractor, the true, nonlinear error behavior will cease to resemble the linearized, exponential error growth.

In summary, a system whose global Lyapunov exponents are all negative is predictable over all time. Perturbations will contract. On the other hand, at least one positive global Lyapunov exponent indicates growing perturbations. Such a system will only be predictable over a finite length of time. One can always lengthen that time by reducing the size of the initial perturbation. Lyapunov exponents of zero indicate a lack of **exponential** growth or contraction of perturbations. Perturbations may still grow or contract in some non-exponential way, e. g. geometrically.

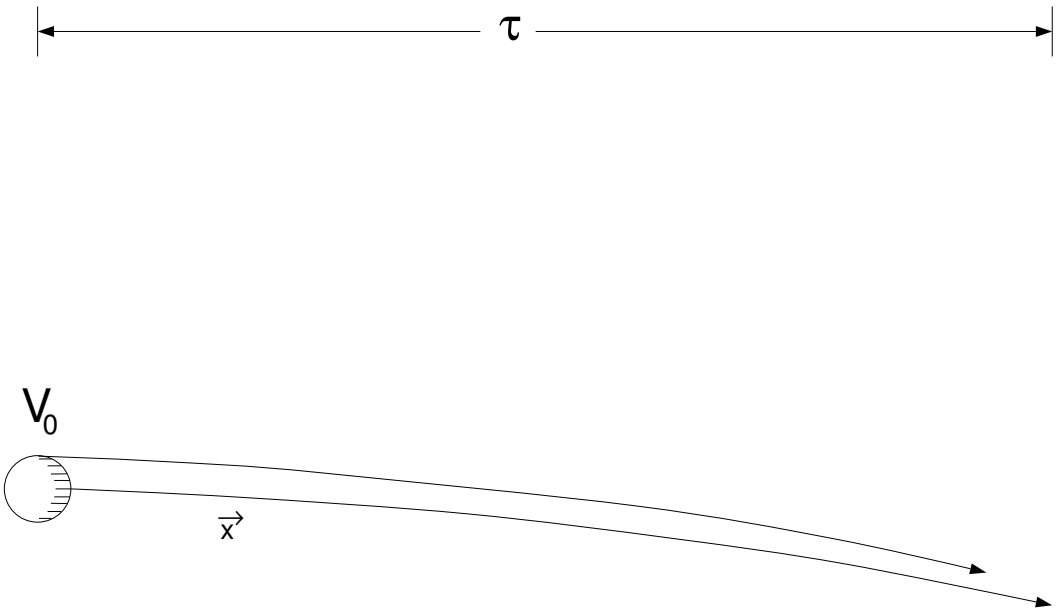

Figure 2.1: Evolution of a sphere of perturbations, V_0, over a length of time, τ. The perturbations converge.

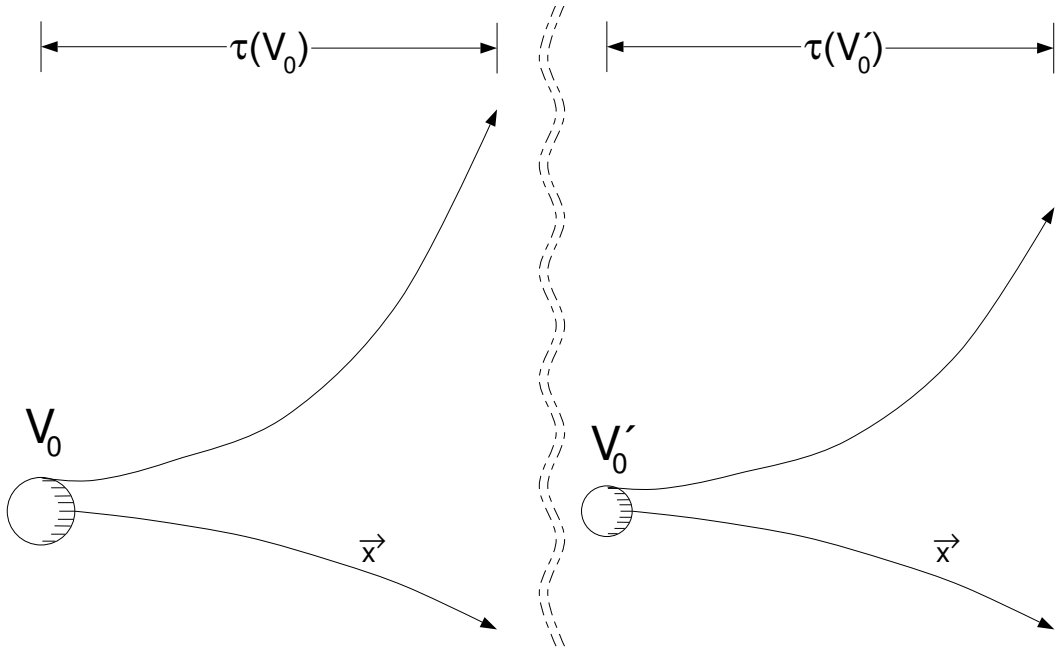

Figure 2.2: Evolution of two spheres of perturbations, V_0 and $V_0' < V_0$, over a length of time, τ. The perturbations diverge, but scale as a function of their initial size.

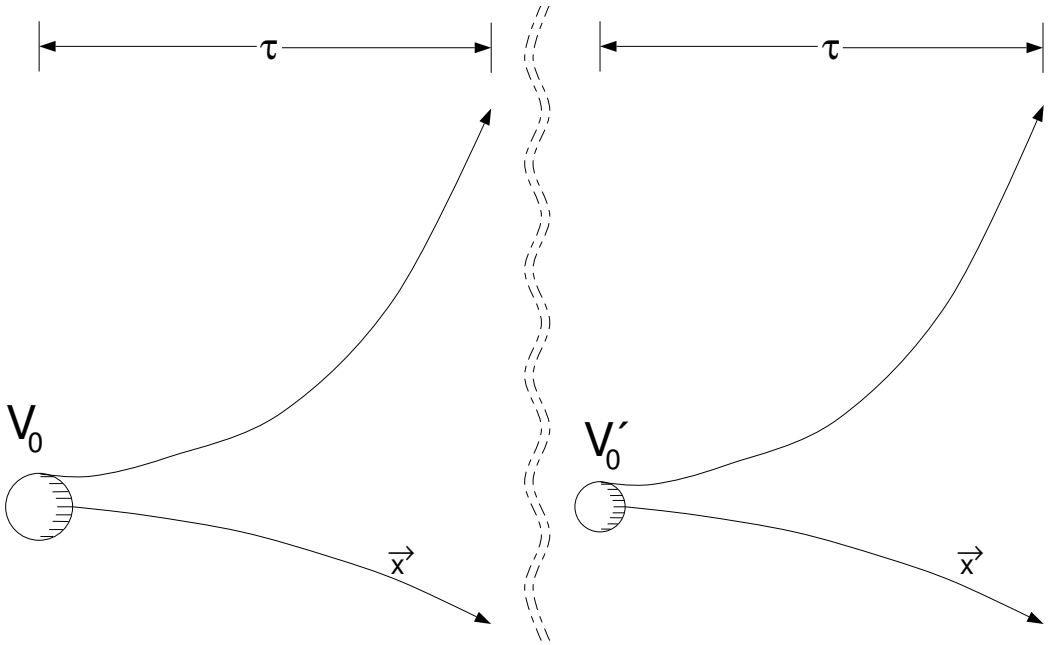

Figure 2.3: Evolution of two spheres of perturbations , V_0, and $V_0' < V_0$, over a length of time, τ. The perturbations diverge and do not scale.

2.2 Lyapunov Exponents: Definitions, Derivations, and Development

2.2.1 Some Notation

The step from the qualitative description of Lyapunov exponents in the last section to the mathematical definitions of the exponents is a large one. The details are often far from obvious and, as noted in section 2.1.2, the notation varies greatly from paper to paper. With the lack of a uniform notation in mind, let us now define one for our own uses throughout the remainder of this work.

As in section 2.1, let $\mathcal{H}^{[N]}$ denote the phase space of a dynamical system, a system whose perturbations evolve via Equations. (2.2-2.5). The perturbation evolution operator, $\mathbf{M}$, is a linear operator in $\mathcal{H}^{[N]}$. Let us use a notation similar to the "bra" and "ket" notation to indicate the action of operators on vectors, as well as to denote the dot product defined on the space $\mathcal{H}^{[N]}$.[1] For those unfamiliar with this notation, it has the following form:

- The "bra," $|\vec{v}\rangle$, denotes a column vector in $\mathcal{H}^{[N]}$.
- The "ket," $\langle\vec{v}|$, denotes a row vector in $\mathcal{H}^{[N]}$.
- The dot-product defined on $\mathcal{H}^{[N]}$ actually has a few different forms:
  (1) Between a matrix operator, $\boldsymbol{\mathcal{L}}$, and a vector, the dot product is $\boldsymbol{\mathcal{L}}\,|\vec{v}\rangle$ or $\langle\vec{v}|\,\boldsymbol{\mathcal{L}}$.
  (2) The "braket"[2] denotes a scalar product, or inner-product. It is $\langle\vec{v}_1\,|\,\vec{v}_2\rangle$ for two vectors and $\langle\vec{v}_1\,|\,\boldsymbol{\mathcal{L}}\,|\,\vec{v}_2\rangle$ for contraction of an operator, $\boldsymbol{\mathcal{L}}$, by two vectors.
  (3) The outer-product, $|\vec{v}_1\rangle\,\langle\vec{v}_2|$, is another form of dot product; however, it generates a matrix operator from the two vectors instead of a scalar.[3]
- "$|\vec{v}\rangle^\dagger$" denotes the adjoint operator, mathematically equivalent to performing a vector-transpose followed by a complex-conjugate, or vice-versa.[4] By definition, $|\vec{v}\rangle^\dagger = \langle\vec{v}|$ and $(\boldsymbol{\mathcal{L}}\,|\vec{v}\rangle)^\dagger = \langle\vec{v}|\,\boldsymbol{\mathcal{L}}^\dagger$.

[1] The choice of this notation is not arbitrary. While the "bra" and "ket" notation is most commonly used in quantum mechanics, it applies equally well to any linear vector space. It is also the notation used by Goldhirsch et. al. (GSO87) in their work on Lyapunov exponents.

[2] or perhaps "bracket," though the latter could also mean the parentheses "[...]"

[3] If $\boldsymbol{\mathcal{L}} = |\vec{v}_1\rangle\,\langle\vec{v}_2|$, then $\mathcal{L}_{ij} = v_{1_i} v_{2_j}$.

[4] For real-valued vectors, as is the case for a dynamical system, the adjoint operator reduces to a vector-transpose.

The magnitude of a vector in this notation is simply:

$$\| \vec{v} \| \equiv \sqrt{\langle \vec{v} \,|\, \vec{v} \rangle}. \tag{2.9}$$

Because the definition of Lyapunov exponents uses volumes, we will need to define a projection operator:

$$\boldsymbol{\mathcal{P}}_k (\hat{e}) \equiv \begin{cases} \sum_{i=1}^{k-1} |\hat{e}_i\rangle \langle \hat{e}_i| & k > 1, \\ \mathbf{0} & k = 1, \end{cases} \tag{2.10}$$

where $\mathbf{0}$ is the zero-matrix and the set of vectors, $\{\hat{e}_k\}$ is an orthonormal basis of the space $\mathcal{H}^{[N]}$. Under the definition presented in Eq. (2.10), one of the basis vectors in $\{\hat{e}_k\}$ must be aligned with at least one of the vectors undergoing projection. Thus, $\boldsymbol{\mathcal{P}}_n (\hat{e}) |\vec{v}_n\rangle$ implies that $\hat{e}_1 = \frac{|\vec{v}_1\rangle}{\| |\vec{v}_1\rangle \|}$. The purpose of this particular definition of a projection operator will become clear shortly.

All projection operators have the following useful properties:

$$[\boldsymbol{\mathcal{P}}_k (\hat{e})]^2 = \boldsymbol{\mathcal{P}}_k (\hat{e}), \tag{2.11}$$

$$[\boldsymbol{\mathcal{P}}_k (\hat{e})]^\dagger = \boldsymbol{\mathcal{P}}_k (\hat{e}). \tag{2.12}$$

Thus, the operator, $[\mathbf{I} - \boldsymbol{\mathcal{P}}_k (\hat{e})]$, where $\mathbf{I}$ is the identity matrix, is useful for defining orthogonalizations. From Eqs. (2.11-2.12), it has the properties $[\mathbf{I} - \boldsymbol{\mathcal{P}}_k (\hat{e})]^2 = [\mathbf{I} - \boldsymbol{\mathcal{P}}_k (\hat{e})]$ and $[\mathbf{I} - \boldsymbol{\mathcal{P}}_k (\hat{e})]^\dagger = [\mathbf{I} - \boldsymbol{\mathcal{P}}_k (\hat{e})]$.

Using the projection operator of Eq. (2.10), one can now define the d-dimensional volume $\mathrm{Vol}^{[d]} \{\vec{v}_k\}$ generated by the vectors $|\vec{v}_1\rangle, \ldots, |\vec{v}_d\rangle$. Let $\mathcal{A}^{[d]} \subseteq \mathcal{H}^{[N]}$ be the d-dimensional subspace of phase space spanned by the $|\vec{v}_1\rangle, \ldots, |\vec{v}_d\rangle$. That is, one can construct an orthonormal basis for $\mathcal{A}^{[d]}$ using the set of $|\vec{v}_k\rangle$. The vectors $|\vec{v}_1\rangle, \ldots, |\vec{v}_d\rangle$, however, need not be orthogonal, much less orthonormal, to generate $\mathrm{Vol}^{[d]} \{\vec{v}_k\}$.

Next, let $\{\hat{\eta}_k\}$ be the basis of $\mathcal{A}^{[d]}$ with one basis vector aligned with $|\vec{v}_1\rangle$, i. e. $\hat{\eta}_1 = \frac{|\vec{v}_1\rangle}{\| |\vec{v}_1\rangle \|}$. The operator $[\mathbf{I} - \boldsymbol{\mathcal{P}}_j (\hat{\eta})]$ will therefore project each vector $|\vec{v}_k\rangle$ onto one of the basis vectors. The product of the projected components forms the volume. One can therefore

define $\mathrm{Vol}^{[d]}\{\vec{v}_k\}$ in one of three ways:

$$\mathrm{Vol}^{[d]}\{\vec{v}_k\} = \prod_{i=1}^{d} \| \left[\mathbf{I} - \boldsymbol{\mathcal{P}}_i(\hat{\eta})\right] |\vec{v}_i\rangle \|, \tag{2.13}$$

$$= \prod_{i=1}^{d} \sqrt{\langle \vec{v}_i \,|\, \left[\mathbf{I} - \boldsymbol{\mathcal{P}}_i(\hat{\eta})\right] \,|\, \vec{v}_i\rangle}, \tag{2.14}$$

$$= \prod_{i=1}^{d} \| \vec{v}_i \| \qquad \text{if } \langle \vec{v}_i \,|\, \vec{v}_k \rangle = 0 \quad \forall\, i \neq k, \tag{2.15}$$

i. e. Equation (2.15) is valid for vectors which are already orthogonal. Again, the $|\vec{v}_k\rangle$ need not be orthogonal.

2.2.2 "Generic" Global Exponents

Though the reader may be familiar with the so-called "global exponent spectrum" from other sources (Tab89, NL92, BG90), there is an older, more general definition of a Lyapunov exponent. Let $\mathcal{A}^{[d]} \subseteq \mathcal{H}^{[N]}$ be a d-dimensional subspace of phase space and $\mathcal{V}_0^{[d]} \in \mathcal{A}^{[d]}$ be any volume in that subspace. We shall again use the perturbation evolution operator of Equation (2.4). One can then define the ***d*th-order global Lyapunov exponent** as (BGGS80a):

$$\lambda^{[d]}\left(\mathbf{M}, \mathcal{A}^{[d]}\right) \equiv \lim_{t' \to \infty} \frac{1}{t'} \ln\left(\frac{\mathcal{V}^{[d]}(t')}{\mathcal{V}_0^{[d]}}\right), \tag{2.16}$$

where $t' = t + t_0$ and

$$\mathcal{V}^{[d]}(t') = \mathrm{Vol}^{[d]}\left\{\mathbf{M}(t;\, \vec{x}_0)\;;\, \mathcal{V}_0^{[d]}\right\}, \tag{2.17}$$

denotes the volume formed by evolving $\mathcal{V}_0^{[d]}$ with $\mathbf{M}(t;\, \vec{x}_0)$, while $\vec{x}_0$ is shorthand for $\vec{x}(t_0)$. The $\lambda^{[d]}$ therefore describe the rate of exponential growth or contraction of $\mathcal{V}_0^{[d]}$, a volume of perturbations to $\vec{x}_0$, some initial point in phase space. Oseledec (Ose68) proved that (2.16) exists for all orders as long as $\mathbf{M}(t;\, \vec{x}_0)$ is "sufficiently well-behaved." Namely, $\frac{1}{t} \ln |\det\left[\mathbf{M}(t;\, \vec{x}_0)\right]|$ must be finite in the limit $t \to \infty$ (Ose68) .

The exponents defined in (2.16) depend solely on the dynamics, $\mathbf{M}(t;\, \vec{x}_0)$, and the particular subspace, $\mathcal{A}^{[d]}$, containing $\mathcal{V}_0^{[d]}$ and $\mathcal{V}^{[d]}(t')$ (BGGS80a). They do not depend on the choice of initial volume, $\mathcal{V}_0^{[d]}$. They may, however, depend on the choice of $\vec{x}_0$ if the dynamical system in question has multiple attractors.

These d^{th}-order Lyapunov exponents are, in part, historical. They were, at one point, the only definition of a Lyapunov exponent. Consequently, there was an exponent at every dimension for every subspace of phase space. The definition of the more familiar Lyapunov exponent spectrum was a breakthrough that simplified this collection of d^{th}-order exponents

The d^{th}-order exponents defined in Equation (2.16) have a very useful property. In their ground-breaking work on Lyapunov exponents, Benettin et. al. (BGGS80a, BGGS80b) demonstrate the following property:

$$\lambda_{\max}^{[d]}\left(\mathbf{M}\right) = \sum_{i=1}^{d} \lambda_i^{[1]}\left(\mathbf{M}\right), \tag{2.18}$$

where $\lambda_{\max}^{[d]}$ is the largest d^{th}-order exponent for all d-dimensional subspaces $\mathcal{A}^{[d]}$. The order-1 exponents, $\lambda_i^{[1]}$, are typically numbered in order of decreasing magnitude, i. e. $\lambda_1^{[1]} \geq \lambda_2^{[1]} \geq \ldots \geq \lambda_N^{[1]}$. Use of this convention is assumed in Equation (2.18). Therefore, using the order-1 exponents, one can compute the largest d^{th}-order exponents.

Additionally, Benettin et. al. (BGGS80a) note that one can readily obtain $\lambda_{\max}^{[d]}$ by simply choosing a subspace, $\mathcal{A}^{[d]}$, at random. To see why, consider the one-dimensional case, where the possible $\mathcal{A}^{[1]}$ are simply the directions defined by any set of basis vectors. Any initial vector, $\mathcal{V}_0^{[1]} = |\vec{v}_0\rangle$, chosen at random, will likely contain a component in the maximally expanding direction. After evolving $|\vec{v}_0\rangle$ long enough, the component in the maximally expanding direction will dominate the resulting vector. Only if one were unlucky and, say, picked a $|\vec{v}_0\rangle$ which pointed in a contracting direction would the previous statements be false.

On a final note, some readers may be used to definitions of Lyapunov exponents in which $\mathcal{V}_0^{[d]} \to 0$ as $t \to \infty$. This is only necessary for $\mathcal{V}_0^{[d]}$ evolving under the full nonlinear dynamics.[5] Benettin et. al. (BGGS80a,BGGS80b) note that, for a linear operator, $\mathbf{M}\left(t;\, \vec{x}_0\right)$, and any two volumes, $\mathcal{V}^{[d]}$, $\mathcal{U}^{[d]} \in \mathcal{A}^{[d]}$,

$$\frac{\text{Vol}^{[d]}\left\{\mathbf{M}\left(t;\, \vec{x}_0\right)\,;\, \mathcal{V}^{[d]}\right\}}{\mathcal{V}^{[d]}} = \frac{\text{Vol}^{[d]}\left\{\mathbf{M}\left(t;\, \vec{x}_0\right)\,;\, \mathcal{U}^{[d]}\right\}}{\mathcal{U}^{[d]}}, \tag{2.19}$$

[5] When evolving an ensemble of initial condition [i. e. the volume $\mathcal{V}_0^{[d]}$] under the full dynamics of the system, one must take $\mathcal{V}_0^{[d]} \to 0$ as $t \to \infty$ to eliminate any effect from the size of the initial volume.

as long as $\mathbf{M}(t;\vec{x}_0)$ is invertible. This relation holds for any subspace $\mathcal{A}^{[d]} \in \mathcal{H}^{[N]}$. Thus the choice of initial volume, $\mathcal{V}_0^{[d]}$, is indeed irrelevant, since the perturbation evolution operator, $\mathbf{M}$, is linear. All initial volumes, therefore, will have the equivalent effect on Eq. (2.16). The theorem stated in (2.19) will be useful in section 2.2.4.

2.2.3 The Global Exponent Spectrum

With the definition of order-d Lyapunov exponents in Equation (2.16) and the property described by Equation (2.18), one can now define the familiar **global Lyapunov exponent spectrum** [GLE]:

$$\lambda_n \equiv \lambda_n^{[1]} = \lambda_{\max}^{[n]}(\mathbf{M}) - \sum_{i=1}^{n-1} \lambda_i, \tag{2.20}$$

for $1 \leq n \leq N$. Again, one chooses a numbering scheme such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$. The N global exponents define a set of expanding and/or contracting directions in phase space. Equation (2.20), in combination with Equation (2.16), is the fundamental definition of the GLE's. All other expressions for the λ_n originate from these two equations.

There is another interesting relation one can derive from Eqs. (2.16) and (2.18). Consider the case of the N^{th}-order Lyapunov exponent. From Eq. (2.16), it is clear that $\delta\mathcal{V}^{[N]} = \lambda^{[N]}\mathcal{V}^{[N]}\delta t$, where $\delta\mathcal{V}^{[N]}$ is the change in volume over some small time interval, δt. From Liouville's theorem, however, $\delta\mathcal{V}^{[N]} = \delta t \int d\mathcal{V}\,\mathrm{tr}(\mathbf{J})$, where $\mathrm{tr}(\mathbf{J})$ is the trace of the Jacobian matrix from Eqs. (2.2-2.3) and the integral is over the volume occupied by the attractor (BG90). Using these facts, one can show that:

$$\sum_{i=1}^{N} \lambda_i = \lim_{T\to\infty} \frac{1}{T} \int_{t_0}^{T} dt' \,\mathrm{tr}\left(\mathbf{J}\left(\vec{x}\left(t'\right)\right)\right), \tag{2.21}$$

relating the GLE spectrum to the dissipation of the system.

Before going any further, let us take time to interpret Eq. (2.20). Suppose that the orthonormal basis $\left\{\hat{f}_n\right\} \in \mathcal{H}^{[N]}$describes the directions of exponential expansion and contraction, numbered in the same fashion as the GLE spectrum. Thus $\left|\hat{f}_1\right\rangle$ indicates the maximally expanding direction, $\left|\hat{f}_2\right\rangle$ the second most rapidly expanding direction, and so on. As noted in section 2.2.2, any randomly selected vector, $|\vec{v}_0\rangle$, will likely have components in

all N directions of the $\left\{\hat{f}_n\right\}$ basis. After evolving that vector to some new vector, $|\vec{v}\,(t+t_0)\rangle$, using the evolution operator, $\mathbf{M}$, $\left|\hat{f}_1\right\rangle$ will be the dominant component of $|\vec{v}\,(t+t_0)\rangle$ as long as t is large. Now consider the same situation in the 2-dimensional case. One evolves some initial volume, $\mathcal{V}_0^{[2]}$, to a final volume, $\mathcal{V}^{[2]}\,(t+t_0)$, over a long time interval, t. $\mathcal{V}^{[2]}\,(t+t_0)$ will then be aligned with the 2-dimensional subspace of $\mathcal{H}^{[N]}$ defined by $\left|\hat{f}_1\right\rangle$ and $\left|\hat{f}_2\right\rangle$. By subtracting λ_1 from $\lambda_{\max}^{[2]}$, one is essentially removing the effect of the maximally expanding direction, leaving only the rate of growth or contraction due to $\left|\hat{f}_2\right\rangle$. This is exactly the logic Benettin et. al. (BGGS80a, BGGS80b) used to define the full GLE spectrum.

Let us now examine some of the alternate forms of Eq. (2.20), alternate definitions of the global Lyapunov exponent spectrum. Let $|\vec{v}_1\rangle, \ldots, |\vec{v}_d\rangle$ be the set of vectors that generates $\mathcal{V}_0^{[d]}$. One can then apply Eq. (2.13) to Eq. (2.17) to get:

$$\mathcal{V}^{[d]}\,(t') = \prod_{i=1}^{d} \parallel \left[\mathbf{I} - \boldsymbol{\mathcal{P}}_i\,(\hat{\eta})\right] \cdot \mathbf{M}\,(t;\,\vec{x}_0)\,|\vec{v}_i\rangle \parallel, \tag{2.22}$$

where $t' = t + t_0$ and the orthonormal basis $\{\hat{\eta}_i\}$ fulfills the requirement for Eq. (2.13), namely that $\hat{\eta}_1 = \frac{|\vec{v}_1\rangle}{\||\vec{v}_1\rangle\|}$. Using (2.22) in (2.20) to perform a proof by induction, one finds the following forms of the GLE spectrum:

$$\lambda_n = \lim_{t\to\infty} \frac{1}{t} \ln\left(\frac{\parallel \left[\mathbf{I} - \boldsymbol{\mathcal{P}}_i\,(\hat{\eta})\right] \cdot \mathbf{M}\,(t;\,\vec{x}_0)\,|\vec{v}_n\rangle \parallel}{\parallel \vec{v}_n \parallel}\right), \tag{2.23}$$

$$= \lim_{t\to\infty} \frac{1}{t} \ln\left(\left\| \left[\mathbf{I} - \boldsymbol{\mathcal{P}}_i\,(\hat{\eta})\right] \cdot \mathbf{M}\,(t;\,\vec{x}_0) \left|\vec{\delta}_n\right\rangle \right\|\right). \tag{2.24}$$

The second form, Equation (2.24), comes from using Eq. (2.19) to replace the set of $\{\vec{v}_n\}$ with the set of unit vectors, $\left\{\vec{\delta}_n\right\}$. One defines the $\vec{\delta}_n$ with the Kronecker-delta function, namely $\left[\vec{\delta}_n\right]_i = \delta_{ni}$. Equation (2.24) is one of the more common forms of the GLE spectrum, since a unit volume is a convenient initial condition to use.

2.2.4 The Effect of Orthonormalization

One of the major innovations in the work by Benettin et. al. (BGGS80b) is demonstrating the effect of orthonormalization on the global exponent spectrum. Specifically, one can regularly "interrupt" the evolution of the initial volume, $\mathcal{V}_0^{[d]}$, and orthonormalize the volume before continuing, yet the GLE will remain unaffected. Such a process aids the com-

putation of the GLE spectrum, since $\mathcal{V}^{[d]}(t+t_0)$ is growing exponentially and will eventually become mathematically unwieldy for a computer.

Consider the following redefinition of Equation (2.17):

$$\mathcal{V}_k^{[d]} = \mathrm{Vol}^{[d]}\left\{\mathbf{M}\left(\Delta t_k;\, \vec{x}\left(t_{k-1}\right)\right) ;\, \boldsymbol{\mathcal{G}}\left\{\mathcal{V}_{k-1}^{[d]}\right\}\right\}, \tag{2.25}$$

where the time intervals, $\Delta t_k = t_k - t_{k-1}$, $k \geq 1$, need not all be the same length, and the operator,

$$\boldsymbol{\mathcal{G}}\left\{\mathrm{Vol}^{[d]}\left\{\vec{v}_n\right\}\right\}: \quad \mathrm{Vol}^{[d]}\left\{\vec{v}_n\right\} \longmapsto \mathrm{Vol}^{[d]}\left\{\hat{\eta}_n\right\},$$

maps the volume defined by the set of d vectors, $\{\vec{v}_n\}$, to the volume defined by $\{\hat{\eta}_n\}$, the set of basis vectors generated by orthonormalizing the $\{\vec{v}_n\}$. The resulting volume is, naturally, a unit volume. The type of orthonormalization performed by $\boldsymbol{\mathcal{G}}$ can, in principle, be anything.

At this point, one can apply the theorem of Eq. (2.19) to (2.25):

$$\frac{\mathrm{Vol}^{[d]}\left\{\mathbf{M}\left(\Delta t_k;\, \vec{x}\left(t_{k-1}\right)\right) ;\, \boldsymbol{\mathcal{G}}\left\{\mathcal{V}_{k-1}^{[d]}\right\}\right\}}{\boldsymbol{\mathcal{G}}\left\{\mathcal{V}_{k-1}^{[d]}\right\}} = \frac{\mathrm{Vol}^{[d]}\left\{\mathbf{M}\left(\Delta t_k;\, \vec{x}\left(t_{k-1}\right)\right) ;\, \mathcal{V}_{k-1}^{[d]}\right\}}{\mathcal{V}_{k-1}^{[d]}}. \tag{2.26}$$

Note that the denominator on the left-hand side of (2.26) is just 1. One can now use (2.26) with (2.25) to find:

$$\mathcal{V}_k^{[d]}\mathcal{V}_{k-1}^{[d]} = \mathrm{Vol}^{[d]}\left\{\mathbf{M}\left(\Delta t_k;\, \vec{x}\left(t_{k-1}\right)\right) ;\, \mathcal{V}_{k-1}^{[d]}\right\}. \tag{2.27}$$

At this point, one uses Eq. (2.27) in a proof by induction, regrouping all of the $\mathbf{M}\left(\Delta t_k;\, \vec{x}\left(t_{k-1}\right)\right)$, to show that

$$\prod_{k=1}^{\kappa} \mathcal{V}_k^{[d]} = \mathrm{Vol}^{[d]}\left\{\prod_{k=1}^{\kappa} \mathbf{M}\left(\Delta t_k;\, \vec{x}\left(t_{k-1}\right)\right) ;\, \mathcal{V}_0^{[d]}\right\}. \tag{2.28}$$

Therefore, the product of the volumes $\mathcal{V}_1^{[d]}, \ldots, \mathcal{V}_\kappa^{[d]}$ is invariant under the effect of reorthonormalizations at times $t_1, \ldots, t_k$. Furthermore, recall Equation (2.6) from earlier in this chapter. It should now be obvious that the right-hand side of Eq. (2.28) is equal to $\mathcal{V}^{[d]}(t_\kappa)$. Thus, the product of the volumes $\mathcal{V}_1^{[d]}, \ldots, \mathcal{V}_\kappa^{[d]}$ is also equal to the volume one would have arrived at had one simply evolved $\mathcal{V}_0^{[d]}$ from time t_0 to time t_κ.

Benettin et. al. (BGGS80b) note the utility of (2.28) for computing the GLE spectrum. I shall use a slightly different notation than they do but sketch out the same proof. First, they make the convenient replacement of $\mathcal{V}_0^{[d]} = \mathrm{Vol}^{[d]}\left\{\vec{\delta}_j\right\}$, where $\left|\vec{\delta}_j\right\rangle$ are the unit-vectors defined in the previous section. Next, they let the operator, $\boldsymbol{\mathcal{G}}$, be Gram-Schmidt orthonormalization. Let us denote the set of vectors defining the reorthonormalized volumes by $\{\hat{\eta}_j\left(t_k;\,\vec{x}_k\right)\}$. Then:

$$\left|\hat{\eta}_n\left(t_k;\,\vec{x}_k\right)\right\rangle = \frac{\left[\mathbf{I}-\boldsymbol{\mathcal{P}}_n\left(\hat{\eta}\left(t_k;\,\vec{x}_k\right)\right)\right]\cdot\left|\vec{v}_n\left(t_k;\,\vec{x}_k\right)\right\rangle}{\sqrt{\left\langle\vec{v}_n\left(t_k;\,\vec{x}_k\right)\,\middle|\,\left[\mathbf{I}-\boldsymbol{\mathcal{P}}_n\left(\hat{\eta}\left(t_k;\,\vec{x}_k\right)\right)\right]\,\middle|\,\vec{v}_n\left(t_k;\,\vec{x}_k\right)\right\rangle}}, \tag{2.29}$$

where $\left|\vec{v}_n\left(t_k;\,\vec{x}_k\right)\right\rangle = \mathbf{M}\left(\Delta t_k;\,\vec{x}\left(t_{k-1}\right)\right)\cdot\left|\hat{\eta}_n\left(t_{k-1};\,\vec{x}_{k-1}\right)\right\rangle$ and $\vec{x}_k \equiv \vec{x}\left(t_k\right)$. One can now use Equations (2.29), (2.28), (2.25), (2.20), and (2.16) in a proof by induction to show that

$$\lambda_n = \lim_{\substack{t_\kappa\to\infty\\ \kappa\to\infty}} \frac{1}{t_\kappa}\sum_{k=1}^{\kappa}\ln\left(\left\|\left[\mathbf{I}-\boldsymbol{\mathcal{P}}_n\left(\hat{\eta}\left(t_k;\,\vec{x}_k\right)\right)\right]\cdot\mathbf{M}\left(\Delta t_k;\,\vec{x}\left(t_{k-1}\right)\right)\left|\hat{\eta}_n\left(t_{k-1};\,\vec{x}_{k-1}\right)\right\rangle\right\|\right). \tag{2.30}$$

This is yet another form for the definition of the GLE spectrum (BGGS80b). It is also the one used by Wolf et. al. (WSSV85) to craft a computer algorithm for numerically calculating the full GLE spectrum. The dual limits in (2.30) express that t_κ can increase, the number of time intervals, κ, can increase, or both. They are equivalent; the choice of both the size and number of the t_k is flexible. In addition to this flexibility, the definition of the GLE spectrum in Eq. (2.30) has one other benefit. Specifically, the use of repeated orthonormalizations generates the exponents in the canonical order. Thus the λ_n in Eq. (2.30) have the property $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$ without any reordering of the indices.[6]

2.2.5 Of Eigenvalues and Eigenvectors

Another oft-referenced work of the field is that by Goldhirsch et. al. (GSO87, referred to in this section as GSO87). This section concerns itself with their results, a relationship between the perturbation evolution operator, $\mathbf{M}\left(t;\,\vec{x}\left(T\right)\right)$, and the operator defined as $\left[\mathbf{M}^\dagger\mathbf{M}\right]\left(t;\,\vec{x}\left(T\right)\right) = \left[\mathbf{M}\left(t;\,\vec{x}\left(T\right)\right)\right]^\dagger\cdot\mathbf{M}\left(t;\,\vec{x}\left(T\right)\right)$. The former is related to the linear stability of the system, while the latter describes Lyapunov stability (GSO87). Abarbanel

[6] It is easy to prove this by using the heuristic description in section 2.2.3 for the generation of the full GLE spectrum.

et. al. (ABK91b, ABK91a, ABK92) call $\mathbf{M}^{\dagger}\mathbf{M}$ the "Oseledec matrix."[7]

I would now like to introduce a minor change in notation. Because so many quantities are functions of $\vec{x}(t)$, let us move the time-dependence of $\vec{x}$ into a subscript. Thus, $\vec{x}_t = \vec{x}(t)$, a far more compact notation.

Let

$$\left[\mathbf{M}^{\dagger}\mathbf{M}\right](S;\vec{x}_T)\cdot\left|\hat{f}_i(\vec{x}_T)\right\rangle = \rho_i(S;\vec{x}_T)\left|\hat{f}_i(\vec{x}_T)\right\rangle, \tag{2.31}$$

where S denotes another time interval, define the eigenvalues and eigenvectors of the operator $\mathbf{M}^{\dagger}\mathbf{M}$. Naturally, the eigenvectors, $\left|\hat{f}_i(\vec{x}_T)\right\rangle$, define a basis of the phase space, $\mathcal{H}^{[N]}$. GSO87 prove that, for $S \gg 1$, the eigenvalues are $\rho_i(S;\vec{x}_T) \approx e^{\lambda_i S}$. Additionally, they show that the eigenvectors, $\left|\hat{f}_i(\vec{x}_T)\right\rangle$, approximate the expanding and contracting directions of the system — **at** $\vec{x}_T$. This is one of the key points of the GSO87 paper: the eigenvectors of $\left[\mathbf{M}^{\dagger}\mathbf{M}\right](S;\vec{x}_T)$ depend **only** on location in phase space, $\vec{x}_T$, **not** on the evolution time, S. The latter appears in the eigenvalues. GSO87 name $\rho_i(S;\vec{x}_T)$and $\left|\hat{f}_i(\vec{x}_T)\right\rangle$ the **Lyapunov eigenvalues** and **eigenvectors**, respectively.

Lorenz (Lor84) also discusses the eigenvalues and eigenvectors of $\mathbf{M}^{\dagger}\mathbf{M}$, though in the context of a singular-value decomposition of $\mathbf{M}$. He notes that $\rho_i(S;\vec{x}_T)^{-\frac{1}{2}}$ are the singular values of $\mathbf{M}(S;\vec{x}_T)$, relating them to the lengths of the semiaxes of an ellipsoid of perturbations. This is another common conceptualization of $\mathcal{V}^{[d]}(t)$ and is equivalent to the parallelpiped view presented in this work. Lastly, Lorenz denotes his version of the $\left|\hat{f}_i(\vec{x}_T)\right\rangle$ as Lyapunov vectors. Both this and the terminology of GSO87 are used interchangeably in the literature. They are also referred to as "singular vectors," since the $\left|\hat{f}_i(\vec{x}_T)\right\rangle$ are by definition the singular vectors of the matrix $\mathbf{M}(S;\vec{x}_T)$ (MK96). In this work, let us herafter use the term "Lyapunov eigenvector" for the sake of clarity.

Sections 2.2.3-2.2.4 described the evolution of a basis of unit vectors, $\left\{\vec{\delta}_i\right\}$. The longer one evolves the $\left\{\vec{\delta}_i\right\}$ using $\mathbf{M}$, with either periodic reorthonormalization or a single final reorthonormalization, the better the resulting basis vectors approximate the expanding and contracting directions of the system. It should therefore come as no surprise that the $|\hat{\eta}_j(t_\kappa;\vec{x}_{t_\kappa})\rangle \rightarrow \left|\hat{f}_j(\vec{x}_{t_\kappa})\right\rangle$ for sufficiently large κ, t_κ, or both. This is useful computationally.

[7]Note, however, that they define it for the dynamics of a mapping, not a flow as is done here.

To simplify the mathematics, let $|\vec{v}_n(T;\,\vec{x}_0)\rangle = \mathbf{M}(T-t_0;\,\vec{x}_0)\cdot\left|\vec{\delta}_n\right\rangle$. Thus,

$$\left|\hat{f}_n(\vec{x}_T)\right\rangle \cong \frac{\left[\mathbf{I}-\boldsymbol{\mathcal{P}}_n\left(\hat{f}(\vec{x}_T)\right)\right]\cdot|\vec{v}_n(T;\,\vec{x}_0)\rangle}{\sqrt{\left\langle\vec{v}_n(T;\,\vec{x}_0)\,\middle|\,\left[\mathbf{I}-\boldsymbol{\mathcal{P}}_n\left(\hat{f}(\vec{x}_T)\right)\right]\,\middle|\,\vec{v}_n(T;\,\vec{x}_0)\right\rangle}}, \tag{2.32}$$

for sufficiently large time, $T \gg t_0$. Equation (2.32) therefore describes a means of approximating the Lyapunov eigenvectors.

Once one has evolved some initial set of perturbations, using (2.32), to generate an initial set of Lyapunov eigenvectors, $\left|\hat{f}_i(\vec{x}_T)\right\rangle$, further computations become quite easy. Consider the following interesting theorem: if $|\vec{g}_j(\vec{x}_{S+T})\rangle \equiv \left[\mathbf{I}-\boldsymbol{\mathcal{P}}_j\left(\hat{f}(\vec{x}_{S+T})\right)\right]\cdot\mathbf{M}(S;\,\vec{x}_T)\cdot\left|\hat{f}_j(\vec{x}_T)\right\rangle$, and $\left|\hat{f}_j(\vec{x}_{S+T})\right\rangle \equiv \frac{|\vec{g}_j(\vec{x}_{S+T})\rangle}{\|\,\vec{g}_j(\vec{x}_{S+T})\,\|}$, then:

$$\left|\hat{f}_j(\vec{x}_{S+T})\right\rangle = \rho_j(S;\,\vec{x}_T)^{-\frac{1}{2}}\,\mathbf{M}(S;\,\vec{x}_T)\cdot\left|\hat{f}_j(\vec{x}_T)\right\rangle, \tag{2.33}$$

where

$$\rho_i(S;\,\vec{x}) = \left\|\,\mathbf{M}(S;\,\vec{x})\left|\hat{f}_i(\vec{x})\right\rangle\right\|^2. \tag{2.34}$$

To prove Equation (2.33), assume it is valid for the $(j-1)^{\text{th}}$ Lyapunov eigenvector. A proof by induction follows easily. The operator, $\mathbf{M}(S;\,\vec{x}_T)$, therefore merely stretches and rotates the $\left\{\hat{f}_i(\vec{x}_T)\right\}$ basis, as noted in GSO87.

One can now use the results of this section to formulate a fourth form of the GLE spectrum. Applying Eq. (2.32-2.34) to Eq. (2.30), one can show:

$$\lambda_n = \lim_{S\to\infty}\frac{1}{S}\ln\left(\left\|\,\mathbf{M}(S;\,\vec{x}(T))\left|\hat{f}_n(\vec{x}(T))\right\rangle\right\|\right). \tag{2.35}$$

This is the form of the GLE spectrum used by GSO87. Notice that repeated reorthonormalizations have no effect on the evolution of the Lyapunov eigenvectors due to Eq. (2.33). Not only is (2.35) convenient for computing the GLE spectrum, it also leads to the definition of local versions of the Lyapunov exponent spectrum.

2.3 Local or "Instantaneous" Lyapunov Exponents

2.3.1 Definition

2.3.1.1 Two Versions: FTLE and ILE

The results presented in section 2.2.5 lay the foundation for a definition of a more local description of perturbation behavior. The GLE spectrum describes the behavior of linear perturbations in terms of exponential divergence rates over the whole of the system's attractor. [Let us ignore the case of multiple attractors for simplicity.] For real physical systems, one often cannot examine the whole of the attractor, nor does one wish to. In such circumstances, it is convenient to define a local version of the Lyapunov exponent spectrum, one which describes the exponential divergence rates of linear perturbations in a specific region of the attractor.

There are two common versions of a "local Lyapunov exponent" in the literature. I shall first present their definitions, along with a brief description of each. I shall wait, however, until section 2.3.1.2 to present references for these definitions. There are several of them, and I do not wish to distract the reader with line upon line of citations.

The first type of local exponent spectrum examines the divergence rate along a trajectory on the attractor. The definition follows easily from Eq. (2.35):

$$\lambda_i\left(S;\,\vec{x}\left(T\right)\right) \equiv \frac{1}{S}\ln\left(\left\|\,\mathbf{M}\left(S;\,\vec{x}\left(T\right)\right)\left|\hat{f}_i\left(\vec{x}\left(T\right)\right)\right\rangle\,\right\|\right), \tag{2.36}$$

where $\left|\hat{f}_i\left(\vec{x}\left(T\right)\right)\right\rangle$ is the i^{th} Lyapunov eigenvector, computed as described by Equations (2.32-2.33). Clearly, $\lambda_i = \lim_{S\to\infty}\lambda_i\left(S;\,\vec{x}\left(T\right)\right)$, a property which any definition of a local Lyapunov exponent must satisfy. I prefer to call this particular form of local exponent a **finite-time Lyapunov exponent**, or FTLE. The FTLE spectrum describes the stability of the attractor along the trajectory from $\vec{x}\left(T\right)$ to $\vec{x}\left(S+T\right)$. They are therefore a function of the starting point in phase space, $\vec{x}\left(T\right)$, and a time interval, S.

The second type of local exponent spectrum is an interesting generalization of the FTLE spectrum:

$$\lambda_i\left(\vec{x}\left(T\right)\right) \equiv \lim_{S\to 0}\lambda_i\left(S;\,\vec{x}\left(T\right)\right), \tag{2.37}$$

$$= \lim_{S\to 0} \frac{1}{S} \ln\left(\left\| \mathbf{M}\left(S;\, \vec{x}\,(T)\right) \left|\hat{f}_i\left(\vec{x}\,(T)\right)\right\rangle \right\|\right), \tag{2.38}$$

$$= \frac{d}{dS} \ln\left(\left\| \mathbf{M}\left(S;\, \vec{x}\,(T)\right) \left|\hat{f}_i\left(\vec{x}\,(T)\right)\right\rangle \right\|\right)\Bigg|_{S=0}. \tag{2.39}$$

Equation (2.37) is the primary definition, and, in some sense, describes the exponential divergence rate of linear perturbations "at a point" on the attractor, $\vec{x}\,(T)$. Equations (2.38-2.39) follow logically from Eq. (2.37). The best description of this type of local exponent is by Vavriv et. al. (VGR90), who call it the **instantaneous Lyapunov exponent** spectrum, or ILE for short.

Because the ILE is a differential quantity, one can readily describe the GLE or the FTLE for any time interval in terms of the ILE. The following three relations formulate this property of the ILE spectrum mathematically. They are stated without proof:

$$\lambda_i = \lim_{t\to\infty} \frac{1}{t} \int_0^t \lambda_i\left(\vec{x}\,(t')\right)\, dt', \tag{2.40}$$

$$\lambda_i\left(t_1;\, \vec{x}\,(t_0)\right) = \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} \lambda_i\left(\vec{x}\,(t')\right)\, dt', \tag{2.41}$$

$$\lambda_i\left(t_\kappa;\, \vec{x}\,(t_0)\right) = \frac{1}{t_\kappa - t_0} \sum_{k=1}^{\kappa} \left(t_k - t_{k-1}\right)\, \lambda_i\left(t_k;\, \vec{x}\,(t_{k-1})\right). \tag{2.42}$$

The work of section 2.3.2 will make these easier to prove.

2.3.1.2 In the Literature

There are, unfortunately, no one or two fundamental works which define Equations (2.36-2.42). Instead, there is a broad collection of papers in a variety of journals, each presenting some aspect of the FTLE or ILE spectrum. To make matters worse, many researchers call their particular version the "local Lyapunov exponents," making no distinction between the FTLE or ILE.

Following work by Nese (Nes89) on local divergence rates, Abarbanel et. al. (ABK91b) define a spectrum of "local" Lyapunov exponents. Their definition is, in fact, a form of the FTLE spectrum of Eq. (2.36), defined for mappings instead of flows. This is useful for

estimating an FTLE spectrum from discrete data (ABK91a, ABK92). Later, Yoden and Nomura (YN93) presented Eq. (2.36) as a FTLE spectrum based on a GLE spectrum, one defined using Eq. (2.35). It is therefore their definition which I use in this work.

For definitions of the ILE, I have found no less than six sources, each of which appears to have independently formulated the ILE. Four of the six provide a definition for only $\lambda_1(\vec{x}(T))$, the ILE associated with the maximal GLE. Eckhardt and Yao (EY93) and Gupte and Amritkar (GA96) all define $\lambda_1(\vec{x}(T))$ using a limit. The form of their definitions is akin to Eq. (2.38). In contrast, Trevisan and Legnani (TL95) and Yamaguchi (Yam96) define $\lambda_1(\vec{x}(T))$ with a differential, yielding a version of Eq. (2.39). Both of these papers also define a finite-time version of the exponent by averaging $\lambda_1(\vec{x}(T))$ over a time interval, a relationship noted here in Eq. (2.41).

Jolly et. al. (JTX95) present a most strange definition of the ILE. Their "instantaneous" Lyapunov exponent is actually a form of the FTLE defined by Abarbanel et. al. (ABK91b), using a time interval corresponding to a single iteration of a mapping. For a mapping, where time is discrete, this is somewhat akin to the ILE. For a flow, however, it is simply the FTLE with an arbitrary, small time interval. Furthermore, their exponents have the property that the computation of an exponent at $\vec{x}(t_k)$ influences the value of that exponent at later points $\vec{x}(t_{k+1})$, $\vec{x}(t_{k+2})$, ... This is simply wrong; it means that different computations of the $\lambda_i(\vec{x}(T))$ at the same $\vec{x}(T)$ will produce utterly different values. Granted, any computation of the ILE can only be an estimate of their true value. Thus any changes in the ILE from computation to computation should only reflect changes in the error of the estimation algorithm. Hopefully, this is indeed what Jolly et. al. (JTX95) mean.

Predating all of the aforementioned papers is one from the former Soviet Union. Not only did Vavriv et. al. (VGR90) "beat everyone to the punch," so to speak, they define a full ILE spectrum, not just $\lambda_1(\vec{x}(T))$.[8] Their definition is in the form shown in Eq. (2.39). They also mention some of the properties of the ILE spectrum presented here in Eqs. (2.40) and (2.41). If their definition of the ILE spectrum is, in turn, based on an earlier work, I am not aware of it.

[8] As an aside: this is not an uncommon phenomenon in nonlinear dynamics. Western researchers seem often to spend several years working on a problem that the Russians already solved and presented in a single publication. It is therefore well worth the effort of drudging through poorly-translated Russian papers to make certain one is not "reinventing the wheel."

2.3.2 A New Version

My own minor contribution to the field of ILE was inspired by Eckhardt and Yao (EY93). In their work, they presented a definition of $\lambda_1(\vec{x}(T))$, the ILE associated with the maximal GLE, using a form similar to Eq. (2.38). However, they then state, without proof, a second version of $\lambda_1(\vec{x}(T))$ in terms of the Lyapunov eigenvectors, $\left|\hat{f}_i(\vec{x}(t))\right\rangle$, and the Jacobian matrix, $\mathbf{J}$. In this section, I present, with proof, a similar definition of the full ILE spectrum.

For the purpose of making the proof readable, I will again use the notation $\vec{x}_t = \vec{x}(t)$. Recall the definition of the ILE spectrum presented in Equation (2.39):

$$\lambda_i(\vec{x}(T)) = \frac{d}{dS} \ln\left(\left\| \mathbf{M}(S;\vec{x}(T)) \left|\hat{f}_i(\vec{x}(T))\right\rangle \right\|\right)\Bigg|_{S=0}.$$

The argument of the logarithm, by Eq. (2.34), is simply $\sqrt{\rho_i(S+T;\vec{x}_T)}$, the root of the Lyapunov eigenvectors. I shall now show that the Lyapunov eigenvectors have the property:

$$\frac{d}{dS} \ln\left[\sqrt{\rho_i(S;\vec{x}_T)}\right] = \left\langle \hat{f}_j(\vec{x}_{S+T}) \,\middle|\, \Re\left[\mathbf{J}(\vec{x}_{S+T})\right] \,\middle|\, \hat{f}_j(\vec{x}_{S+T}) \right\rangle, \tag{2.43}$$

where $\rho_i(S;\vec{x}_T)$ is as defined in Eq. (2.34) and $\Re[z] = \frac{1}{2}\left(z^\dagger + z\right)$ denotes the real part of a complex number or operator.

The proof of Eq. (2.43) is as follows: consider the derivative of $\mathbf{M}(S;\vec{x}_T)$ with respect to S. Using Eq. (2.5) and denoting the time derivative using a dot, one can show that $\dot{\mathbf{M}}(S;\vec{x}_T) = \mathbf{J}(\vec{x}_{S+T}) \cdot \mathbf{M}(S;\vec{x}_T)$. Applying this to the time derivative of Equation (2.34):

$$\dot{\rho}_i(S;\vec{x}_T) = \left\langle \hat{f}_i(\vec{x}_T) \left[\dot{\mathbf{M}}^\dagger(S;\vec{x}_T)\cdot\mathbf{M}(S;\vec{x}_T) + \mathbf{M}^\dagger(S;\vec{x}_T)\cdot\dot{\mathbf{M}}(S;\vec{x}_T)\right] \hat{f}_i(\vec{x}_T)\right\rangle, \tag{2.44}$$

one can show that:

$$\begin{aligned}
\dot{\rho}_i(S;\vec{x}_T) &= \left\langle \hat{f}_i(\vec{x}_T)\right| \left[\mathbf{M}^\dagger(S;\vec{x}_T)\cdot\mathbf{J}^\dagger(\vec{x}_{S+T})\cdot\mathbf{M}(S;\vec{x}_T)\right. \\
&\quad + \left.\mathbf{M}^\dagger(S;\vec{x}_T)\cdot\mathbf{J}(\vec{x}_{S+T})\cdot\mathbf{M}(S;\vec{x}_T)\right] \left|\hat{f}_i(\vec{x}_T)\right\rangle, \qquad (2.45) \\
&= 2\left\langle \hat{f}_i(\vec{x}_T) \,\middle|\, \mathbf{M}^\dagger(S;\vec{x}_T)\cdot\Re\left[\mathbf{J}(\vec{x}_{S+T})\right]\cdot\mathbf{M}(S;\vec{x}_T) \,\middle|\, \hat{f}_i(\vec{x}_T)\right\rangle. \qquad (2.46)
\end{aligned}$$

One can now use Eq. (2.33) to evolve the $\left|\hat{f}_i\left(\vec{x}_T\right)\right\rangle$ in (2.46), finding:

$$\dot{\rho}_i\left(S;\,\vec{x}_T\right) = 2\rho_i\left(S;\,\vec{x}_T\right)\left\langle \hat{f}_i\left(\vec{x}_{S+T}\right)\,\middle|\,\Re\left[\mathbf{J}\left(\vec{x}_{S+T}\right)\right]\,\middle|\,\hat{f}_i\left(\vec{x}_{S+T}\right)\right\rangle. \tag{2.47}$$

The remainder of the proof now becomes straightforward.

Using Eq. (2.43) in Eq. (2.39), we arrive at the following form of the ILE spectrum:

$$\lambda_i\left(\vec{x}\left(T\right)\right) = \left\langle \hat{f}_i\left(\vec{x}\left(T\right)\right)\,\middle|\,\Re\left[\mathbf{J}\left(\vec{x}\left(T\right)\right)\right]\,\middle|\,\hat{f}_i\left(\vec{x}\left(T\right)\right)\right\rangle. \tag{2.48}$$

When the elements of the Jacobian are not complex numbers, the ILE are a contraction of the Jacobian by the Lyapunov eigenvectors, all of which are evaluated at the same point in phase space. Whether the Jacobian is complex or real-valued, Eq. (2.48) makes deriving Eqs. (2.40-2.42) trivial. Additionally one can use Eq. (2.40) and (2.21) to show:

$$\sum_{i=1}^{N}\lambda_i\left(\vec{x}\left(t\right)\right) = \mathrm{tr}\left(\mathbf{J}\left(\vec{x}\left(t\right)\right)\right), \tag{2.49}$$

where "$\mathrm{tr}\left(\ldots\right)$" is the trace of a matrix. Eckhardt and Yao (EY93) make mention of this relationship of the local Lyapunov exponents to local measures of the dissipation of the system.

2.3.3 Computation, Uses, and Units

2.3.3.1 Producing Lyapunov Eigenvectors

Computing the ILE spectrum using Eq. (2.48) requires the Lyapunov eigenvectors, $\left|\hat{f}_i\left(\vec{x}\left(t\right)\right)\right\rangle$. Given the equations of motion, one can always calculate the Jacobian matrix, $\mathbf{J}\left(\vec{x}\left(t\right)\right)$. Computing $\mathbf{M}\left(S;\,\vec{x}\left(t\right)\right)$, however, is not always possible or practical. There is, however, a way to estimate the $\left|\hat{f}_i\left(\vec{x}\left(t\right)\right)\right\rangle$ numerically: Equation (2.32). By starting with any set of unit vectors, $\vec{\delta}_i$, one need only numerically integrate the entire set $\vec{\delta}_i$ of using Eq. (2.2) for a "long enough" time interval, T. [Of course, one must also simultaneously integrate the equations of motion.] Once this integration is complete, a simple Gram-Schmidt orthonormalization of the results yields an adequate approximation of the $\left|\hat{f}_i\left(\vec{x}\left(T\right)\right)\right\rangle$.

Section 2.2.5 showed that there is a modification to the aforementioned technique.

Instead of a single, long integration and Gram-Schmidt orthonormalization of the $\left\{\vec{\delta}_i\right\}$ for some length of time, T, one could perform numerous short integration/orthonormalization steps equivalent to T. After enough such steps, one again has an adequate approximation of the $\left|\hat{f}_i\left(\vec{x}\left(T\right)\right)\right\rangle$. Thus, the algorithm by Wolf et. al. (WSSV85) will generate a set of approximate Lyapunov eigenvectors after some initial number of steps to integrate away transients. Once this initial alignment integration is complete, any successive integration and Gram-Schmidt orthonormalization produces the next set of Lyapunov eigenvectors, $\left|\hat{f}_i\left(\vec{x}\left(T+\Delta t\right)\right)\right\rangle$. This is guaranteed by Eq. (2.33) and the other results of section 2.2.5.

One may wonder how small one must make the steps, Δt, between successive computations of the $\left|\hat{f}_i\left(\vec{x}\left(T+k\Delta t\right)\right)\right\rangle$, where k is an integer. [The use of a fixed Δt is, of course, a convenience.] A related question is how the ILE depend on Δt. In both cases, the size of Δt is irrelevant. The definition of the ILE spectrum in Eq. (2.48) depends only on the Lyapunov eigenvectors and the Jacobian. The latter certainly does not depend on orthonormalizations of perturbations. As for the former, the $\left|\hat{f}_i\left(\vec{x}\left(T+k\Delta t\right)\right)\right\rangle$ will remain good approximations to their true values as long as Δt is not "too large." Computers have a finite numerical precision. If one evolves the set of Lyapunov eigenvectors for too long, the separation between the maximally growing and maximally contracting directions will reach machine precision, and the Gram-Schmidt orthonormalization will fail. Even if the computation has not reached machine precision, roundoff errors may still appear in the orthonormalization. Thus, there is an upper limit on the size of Δt, the time between successive computations of the $\left|\hat{f}_i\left(\vec{x}\left(T+k\Delta t\right)\right)\right\rangle$. In practice, this interval need only be very small if one is also estimating the GLE spectrum from the ILE's.

There are two ways to perform the alignment integration and obtain the first set of $\left|\hat{f}_i\left(\vec{x}\left(T\right)\right)\right\rangle$. If $\vec{x}\left(t_0\right)$ is unimportant, i. e. one does not need the values of the ILE spectrum at $\vec{x}\left(t_0\right)$, one can just perform the alignment integration from t_0 to T. The actual value of T will differ from problem to problem. If one needs the ILE spectrum at $\vec{x}\left(t_0\right)$, one can use a technique described by Eckhardt and Yao (EY93). First, integrate $\vec{x}\left(t_0\right)$ backwards to find $\vec{x}\left(t_0-T\right)$. Now perform the alignment integration from t_0-T to t_0 to find $\left|\hat{f}_i\left(\vec{x}\left(t_0\right)\right)\right\rangle$, the Lyapunov eigenvectors at t_0.

There is one warning about the method described in this section: it will **only** work if one uses Gram-Schmidt to perform orthonormalization. Sections 2.2.4-2.2.5 con-

tain the reasons why. Using QR-factorization or singular-value decomposition to perform the orthonormalization may or may not work.[9] Regardless, Gram-Schmidt orthonormalization offers an advantage over other orthonormalization techniques. By the very definition of Gram-Schmidt orthonormalization, the first Lyapunov eigenvector, $\left|\hat{f}_1\left(\vec{x}\left(t\right)\right)\right\rangle$, will always be aligned with the maximally expanding direction at $\vec{x}\left(t\right)$. Note that I assume that $\left|\hat{f}_1\left(\vec{x}\left(t\right)\right)\right\rangle$ is the first vector generated by Gram-Schmidt orthonormalization, $\left|\hat{f}_2\left(\vec{x}\left(t\right)\right)\right\rangle$ the second, and so on. As a result, the first ILE, $\lambda_1\left(\vec{x}\left(t\right)\right)$, always generates the largest GLE under Eq. (2.40). The same holds for other exponents. If one uses some other technique, say, one involving QR-factorization, the aforementioned relationship will most certainly not hold, and one will need to sort one's GLE spectrum to place it in the canonical order.

2.3.3.2 Bits-per-Second

Like many other aspects of Lyapunov exponents, there is no one standard set of units used in the literature. Consequently, a researcher comparing his numerical results to a test case he has chosen from the literature could spend days looking for an error in his code when the discrepancy is actually due to units. The obvious units for Lyapunov exponents of any variety are $\frac{1}{\text{time}}$. All of the definitions in this chapter use these units.

There is a second set of units for Lyapunov exponents used in the literature: $\frac{\text{bits}}{\text{time}}$. There is a branch of statistical mechanics which has related the entropy of a system to a loss of information about that system. There are similar measures of "entropy," or loss of information, for nonlinear systems (Tab89, BG90, NL92). The Lyapunov exponents are related to such measures (BGS76). It is therefore customary to consider the Lyapunov exponents as the divergence rates on powers of 2, hence units of $\frac{\text{bits}}{\text{time}}$. To perform the conversion to these units, one must replace every $\ln\left(\ldots\right)$ in the expressions of this chapter by $\log_2\left(\ldots\right)$, the logarithm at base-2. Considering that $\log_2\left(y\right) = \frac{\ln(y)}{\ln(2)}$, one can just as easily divide by $\ln\left(2\right)$ to convert from $\frac{1}{\text{time}}$ to $\frac{\text{bits}}{\text{time}}$. The reason for this warning should now be clear: it is not easy to detect a consistently missing factor of $\frac{1}{\ln(2)}$!

[9] I have not investigated this possibility.

2.3.3.3 Probability Distribution Functions

Once one computes the ILE at numerous points on an attractor, a natural question arises: aside from computing the GLE spectrum, what can one do with all of these ILE's? The answer lies in a paper by Abarbanel et. al. (ABK91b). They state that all moments of the probability distribution functions, or PDF's, of an ILE exist for each exponent in the spectrum. They additionally note that the mean of the various PDF's converge to the GLE spectrum as the number of ILE's in the distribution increases. The PDF's of the ILE's therefore describe properties of the system under examination (AB93). Several works use these PDF's in their analyses (AB93, JTX95, TL95, AL96, GA96, Yam96). Yamaguchi (Yam96) looks at the PDF's of the FTLE's, examining their behavior as a function of the energy of the system. This idea is readily extensible: one can look at the PDF's of a local Lyapunov exponent versus any quantity of interest to learn something about the stability behavior of the system. This is exactly what is done in chapter 6.

2.3.3.4 A Word about Mappings and Final Thoughts

Section 2.1.2 noted that the work of this chapter has focussed on flows, dynamical systems in which time is continuous. Datasets, however, are discrete. Thus, the concept of a mapping, described in section 2.1.2, is more appropriate for datasets. The work of Abarbanel et. al. (ABK91b, ABK91a, ABK92) estimates Lyapunov exponents from datasets. The basic technique involves the QR-decomposition method proposed by Eckman et. al. (EKRC86). The QR-decomposition method also works with flows and is an alternative to the techniques described by this chapter.

There is one thing the reader should be aware of. Because Abarbanel et. al. (ABK91b, ABK91a, ABK92) are working with datasets and therefore with a mapping involving discrete time, the definitions of the Lyapunov exponents change. Specifically, the perturbation evolution operator, $\mathbf{M}(t;\vec{x})$, must be replaced by the Jacobian, $\mathbf{J}(\vec{x}_n)$. The properties and relations described for $\mathbf{M}(t;\vec{x})$ **do not** carry over to the Jacobian of a mapping! Notably, the Lyapunov eigenvectors as defined in this chapter simply do not apply to a mapping. The same is true for the ILE spectrum. Instead, other quantities fulfill the same roles. If one needs to estimate the GLE of a timeseries, see the works by Abarbanel

et. al. (ABK91b, ABK91a, ABK92). These papers also describe versions of a local Lyapunov exponent spectrum appropriate for mappings.

The work of this chapter has focussed not on mappings, but on flows, dynamical systems whose equations of motion are continuous functions of time. Sections 2.1.2 and 2.2.1 provided the necessary notational framework. Each subsequent section build upon the previous ones. Section 2.2.2 defined the order-d Lyapunov exponents, the exponential growth rates of d-dimensional volumes in the phase space of the system. From these, section 2.2.3 built the global Lyapunov exponent spectrum, the fundamental quantities describing the divergence behavior of linearized perturbations to a trajectory through phase space. Sections 2.2.4 and 2.2.5 concerned theory useful in the actual computation of the global exponent spectrum, while laying the groundwork for the definition of local Lyapunov exponents in section 2.3. One version of local exponents, the "instantaneous" Lyapunov exponents, will be used heavily in chapter 6.

Chapter 3

Nonlinear Regression and Optimization

3.1 Basic Concepts and Notation

3.1.1 Introduction

The term "optimization" generally refers to seeking the maximum or minimum of some function. Of particular interest are error functions, such as

$$\chi^2(\vec{a}) = \sum_{\text{all } \vec{x}} \left[T(\vec{x};\, \vec{a}) - y(\vec{x})\right]^2, \tag{3.1}$$

where $y(\vec{x})$ is a point in a dataset, taken at time or location $\vec{x}$, and $T(\vec{x};\, \vec{a})$ is some model with a set of parameters, $\vec{a}$, to which one wishes to fit the dataset y. The process of minimizing $\chi^2(\vec{a})$, commonly known as regression or least-squares fitting, is itself a subfield of optimization, one with a wide range of applications. For example, if $T(\vec{x};\, \vec{a})$ were some forecast model, one could use (3.1) to determine how well T describes the data one is forecasting. This example reveals an additional concern of regression beyond mere minimization of $\chi^2(\vec{a})$: "goodness of fit." Goodness of fit itself comes in two flavors, how well does the model describe the data at the value(s) of $\vec{a}$ which minimize $\chi^2(\vec{a})$, and how well does that "minimizing value" of $\vec{a}$ actually describe a minimum of χ^2.

When most researchers think of fitting data to a model, the model they have in mind is typically some linear function, or at best a low-order polynomial, with only one or two parameters which are simply multiplicative coefficients. While linear models have certain advantages, simplicity among them, they restrict the modeler's ability to best describe the data or glean the most information from the model. Consider the case where one uses

regression for pattern recognition. A linear fit of data to a target pattern or set of target patterns — say, uppercase letters of the alphabet — may fail to distinguish two different cases apart. The letters "E" and "F" may have the same linear fit to a given piece of input data, for example.

This research itself uses regression for pattern recognition, as seen in chapters 4-6. The example described above is exactly the problem experienced during initial attempts to use simple linear regression to perform the pattern recognition scheme needed by this research. Yet, despite some of the obvious advantages of nonlinear regression, there is appallingly little said about it in standard textbooks. Most lack any mention of nonlinear regression techniques. Of those that do, almost all give it only brief mention and, perhaps, a reference to an article or two in a mathematical journal. It is this situation which inspired the present chapter about nonlinear regression.

In addition to serving as some introduction to nonlinear regression for other researchers, this chapter documents the basis of techniques used by the remainder of this work. With that in mind, in the remainder of section 3.1, I begin with a description of notation and the basic methods of nonlinear regression. Section 3.2 describes genetic algorithms and their uses in nonlinear regression, especially their application to the problem of multiple minima in χ^2. In section 3.3, the chapter concludes with correlation coefficients, confidence intervals, and when one can use them.

3.1.2 Notation

Both linear and nonlinear regression have features in common. For the purpose of future discussion, let us adopt the matrix notation of Ryan (Rya97). The model and the data thus become $\vec{T}(\vec{a}) \equiv T(\vec{x};\, \vec{a})$ and $\vec{y} \equiv y(\vec{x})$, respectively. Equation (3.1) then becomes:

$$\chi^2(\vec{a}) = \left\| \vec{\mathcal{E}}(\vec{a}) \right\|^2 \qquad \text{where } \vec{\mathcal{E}}(\vec{a}) \equiv \vec{T}(\vec{a}) - \vec{y}. \tag{3.2}$$

Thus $\vec{\mathcal{E}}(\vec{a})$ describes the errors between the model and the data at all evaluation points, $\vec{x}$. The data, $\vec{y}$, is more commonly known as the dependent variable, where $\vec{x}$ is the independent variable. In fact, $\vec{x}$ may represent more than one independent variable, e. g. time, latitude, and longitude. In this case, $\vec{x}$ will be a vector of all **unique combinations** of all values

of the independent variables. The use of vector notation for the parameters of the model is for personal convenience; $\vec{a}$ is far more compact a notation than a_1, a_2, ... , a_k. The actual parameters need not have any dimensionality to them.

The reader may be familiar with another definition for the error vector, $\vec{\mathcal{E}}(\vec{a})$. Specifically,

$$\mathcal{E}_i(\vec{a}) = \frac{T_i(\vec{a}) - y_i}{\sigma_i} \tag{3.3}$$

where $\vec{\sigma}$ are systematic errors, possibly due to measurement, in the the data points, $\vec{y}$. The purpose of doing this is twofold. First, (3.3) describes a weighting of the fit errors. If a particular data point has a high measurement error associated with it, then it makes little sense to give its deviation from the model much credence. Second, (3.3) is a normalization, one which permits some convenient assumptions. We shall return to this latter point in greater detail in section 3.3. Of course, the use of $\vec{\sigma}$ becomes meaningless if one doesn't have the measurement errors in $\vec{y}$, if $\vec{y}$ is a synthetic dataset, for which the concept of measurement error is moot, or if the very idea of systematic errors in $\vec{y}$ makes no sense. Let us therefore assume either definition of $\vec{\mathcal{E}}(\vec{a})$ in future discussion.

3.1.3 Nonlinear Regression

The primary differences between linear and nonlinear regression lie with the parameters, $\vec{a}$. The nonlinearity in nonlinear regression is not due to functions of the independent variables, $\vec{x}$, but due to the parameters. As an example, consider the functions $T_i = \sum_{j=0}^{n} a_j x_i^j$, $T_i = a_0 \sin(x_i + e^{x_i}) + a_1 \arcsin(\sinh(x_i))$, and $T_i = \frac{1}{a\sqrt{2\pi}} e^{-(x_i)^2/2a^2}$. The first function is an n^{th}-order polynomial. Its parameters, $a_0 \ldots a_n$, are all linear coefficients. Hence this model is linear in its parameters. The second model, too is linear in its two parameters, even though the model itself is a highly nonlinear function of $\vec{x}$. In both of these cases, some form of linear regression will do nicely.[1] The third function, however, is nonlinear in both $\vec{x}$ and its sole parameter, a. Additionally, it cannot be transformed to render the problem a linear one, as discussed in section 3.2.4. For this function, one has no choice but to use nonlinear regression.

The astute reader may be wondering if the methods for linear regression are simply

[1] The reader interested in regression of such problems should see (PTVF92, Rya97, Spi95) for specific details and techniques.

a subset of those for nonlinear regression. This is not the case. Linear regression can use techniques to solve directly for the value of $\vec{a}$ that minimizes χ^2 (PTVF92, Rya97, Spi95). Clearly, this is impossible to do in the nonlinear case, where some of the parameters may be analytically uninvertible. The standard technique for nonlinear regression involves assuming that $\chi^2(\vec{a})$ can be approximated quadratically near a minimum using Taylor expansion (PTVF92, Rya97):

$$\chi^2(\vec{a}_{\rm new}) \cong \chi^2(\vec{a}) + \vec{\nabla}_{\vec{a}}^{\rm T}\chi^2(\vec{a}) \cdot \delta\vec{a} + \frac{1}{2}\delta\vec{a}^{\rm T} \cdot \left(\vec{\nabla}_{\vec{a}} \cdot \vec{\nabla}_{\vec{a}}\right)\chi^2(\vec{a}) \cdot \delta\vec{a}, \tag{3.4}$$

where $\vec{\nabla}_{\vec{a}}$ denotes the vector of partial derivatives with respect to each of the parameters, the superscript "T" denotes the transpose,[2] and $\delta\vec{a} = \vec{a}_{\rm new} - \vec{a}$. Additionally, let $\vec{a}_{\rm min}$ denote a minimum of $\chi^2(\vec{a})$. Note that either $\vec{a}_{\rm new}$ **or** $\vec{a}$ in Eq. (3.4) could be $\vec{a}_{\rm min}$. The assumption is that both $\vec{a}_{\rm new}$ and $\vec{a}$ are sufficiently close enough to the minimum for (3.4) to hold.

There is an important equation one can derive from (3.4), an equation which is the foundation of a technique called the Gauss-Newton method (PTVF92, Rya97). Let

$$\mathbf{D}(\vec{a}) \equiv \frac{1}{2}\left(\vec{\nabla}_{\vec{a}} \cdot \vec{\nabla}_{\vec{a}}\right)\chi^2(\vec{a}) \tag{3.5}$$

define the so-called curvature matrix. Then

$$\mathbf{D}(\vec{a}) \cdot \delta\vec{a} = -\vec{\nabla}_{\vec{a}}^{\rm T}\chi^2(\vec{a}_{\rm new}) \tag{3.6}$$

describes the set of equations, termed "normal equations," which one can use to minimize $\chi^2(\vec{a})$ (PTVF92,Rya97). The reason for assuming that both $\vec{a}_{\rm new}$ and $\vec{a}$ are "sufficiently close enough to $\vec{a}_{\rm min}$" should now be clear: the right-hand side of (3.6) would be zero if $\vec{a}$ were the minimum! The act of using Eq. (3.6) to locate a minimum is actually an application of the method of steepest descent (PTVF92, HH98), and involves some techniques beyond simply solving (3.6), which itself must be done iteratively. A detailed description of these issues is beyond the scope of this work. The interested reader should see Ryan (Rya97), Press

[2]The definition of the dot product is such that the dot product of two column vectors [or two row vectors] produces a matrix, not a scalar. Many texts examined by the author, however, "play fast and loose" with this notation, making it unclear how one performs the dot product in several cases. Let us be more rigorous and make copious use of the transpose.

et. al. (PTVF92), or even the MINPACK library routine `lmder.f`, available from NETLIB (MGH80).

3.2 Multiple Minima and Genetic Algorithms

3.2.1 The Problem of Multiple Minima

As noted in the previous section, the normal equations used in minimizing $\chi^2(\vec{a})$ assume that one is "sufficiently close enough to the minimum." How close is "sufficiently close" is rather problem-dependent. Quadratic minimization techniques are reasonably good at finding **a** minimum for most problems. Regardless, one needs to specify a starting point to the quadratic minimization algorithm. If $\chi^2(\vec{a})$ has only one minimum, one's work is relatively easy. If, however, $\chi^2(\vec{a})$ has more than one minimum, then one must somehow locate the smallest of these — the global minimum — in order to truly optimize χ^2 and find the best-fit parameters, $\vec{a}_{\min}$. If one is lucky, then the minimization problem at hand is simple enough so that one can use knowledge of the problem and some clever tricks to pick a point near the global minimum. Typically, one is not so lucky.[3]

One might be tempted to simply pick the initial parameter value for the quadratic minimizer at random. Generalizing this idea, one could randomly pick an ensemble of initial $\vec{a}$, perform the minimization via Eq. (3.6) for each one, then find which $\vec{a}$ in the ensemble produced the smallest χ^2. Such a scheme will always work given a large enough ensemble; however, it is computationally expensive in the extreme. A related scheme that requires less computer resources simply computes $\chi^2(\vec{a})$ at each of the values of $\vec{a}$ in the ensemble of random parameters. The value of $\vec{a}$ producing the smallest χ^2 then becomes the initial parameter value for the quadratic minimizer. This second scheme will also work if the ensemble is large enough.

Using a large ensemble of random $\vec{a}$ to find which produces the smallest χ^2 is still an inefficient scheme for locating a point near the global minimum. There are many techniques that use random parameter values in combination with some of the properties of the problem to find the global minimum more efficiently. Two described here are simulated annealing and

[3] This is especially true in the case of optimization for pattern recognition, where one performs a fit to the same model using multiple datasets. It therefore becomes difficult to know anything about $\chi^2(\vec{a})$ in general and design a clever means of always picking a point near the global minimum.

genetic algorithms. Both are general optimization schemes which can be used for locating maxima and minima. The following section discusses genetic algorithms in detail.

The method of simulated annealing mimics the process of slowly cooling a liquid to form a single crystal. The technique involves randomly choosing each of the parameters from a particular range. These ranges grow progressively smaller according to a schedule, with the center of the range remaining near the optimal value from the previous selection cycle. One can find a full description of simulated annealing in the papers by Kirkpatrick (KGV83, Kir84).

3.2.2 Genetic Algorithms: An Overview

Genetic algorithms, or GA's, use biological evolution as the paradigm for randomly seeking a minimum. A full description of genetic algorithms can be found in the excellent text by Goldberg (Gol89), the source for this sketch of how GA's work. In a GA, each set of parameters, $\vec{a}$, is analogous to a biological entity, with $\chi^2(\vec{a})$, the function to minimize, a descriptor of the "fitness" of that entity. A particular ensemble of $\vec{a}$ is analogous to a population of animals. Just as in nature, where the most fit animals mate to produce offspring and the least fit die, the "most-fit"[4] $\vec{a}$ combine their traits via "crossover" to produce new parameter values, while one removes the "least fit" $\vec{a}$ from the ensemble. One also "mutates" the parameters to prevent "inbreeding," providing a continual random exploration of the space of possible $\vec{a}$. The entire process looks something like:

(1) Select an initial population.

(2) Evaluate the fitness of each member [using the function to be minimized or maximized].

(3) Discard the $N_{\rm pop} - N_{\rm breed}$ members with the lowest fitness [$N_{\rm pop}$ is the size of the population].

(4) Use crossover to combine the "genes" from the $N_{\rm breed}$ members with the highest fitness and produce $N_{\rm pop} - N_{\rm breed}$ new members.

(5) Randomly select $\mu\%$ of the population and randomized a "gene" in each.

[4] or, in this case, best fit

(6) Repeat steps 2-5 until some convergence criteria is met, or until the desired number of iterations.

The question remains: what constitutes a "gene" for a set of numerical values, and how does one perform crossover with them? Goldberg describes encoding the $\vec{a}$ as a binary number (Gol89). The "genes" are each bit, and crossover becomes a process of selecting which bits from each parent to pass to the offspring. Mutation, similarly, involves selecting a gene and changing its value. One might be wondering how one should select the genes for crossover and mutation. This is where genetic algorithms change from science to art. There is no set method; one should choose a particular scheme for crossover and mutation which best suits the problem (Gol89, HH98).

In addition to there being an art to the crossover and mutation schemes, the encoding scheme is itself subject to modification. For the purpose of regression, Haupt and Haupt (HH98) describe a real-value encoding, in which the elements of a given $\vec{a}$ are the genes. Under this encoding, crossover is a linear combination:

$$\vec{a}_1' = w \cdot \vec{a}_1 + (1-w) \cdot \vec{a}_2, \tag{3.7}$$

$$\vec{a}_2' = (1-w) \cdot \vec{a}_1 + w \cdot \vec{a}_2, \tag{3.8}$$

where $w \in [0,1]$ is a uniform random number. Mutation is performed much the same, except that the gene to mutate is replaced with a new random value for that parameter. Haupt and Haupt also advocate using an larger initial population and mutation with "elitism" (HH98). Selecting some $N_{\mathrm{pop}_0} > N_{\mathrm{pop}}$, one performs an initial evaluation of fitness between steps 1 and 2 of the GA and keeps the N_{pop} fittest members for the initial population of the GA. Under elitism, the N_{elite} fittest members of the population never undergo mutation. As with other aspects of genetic algorithms, the choice of N_{elite}, N_{pop_0}, and N_{pop}, as well as N_{breed} and μ, is something of an art.

3.2.3 Using Genetic Algorithms with Standard Nonlinear Optimization Techniques

The second most important criteria for choosing N_{elite}, N_{pop_0}, N_{pop}, N_{breed} and μ, is convergence. One wishes the GA to converge to the correct answer as quickly as feasible

without sacrificing accuracy. When a GA is simply the "front-end" for a traditional quadratic minimization algorithm, the concern for accuracy is less strict. One need only find a point in the vicinity of the global minimum, not the value of the minimum itself. One can therefore run the GA for a fixed number of iterations, N_{iters}, tuning this number and N_{pop} to perform as few evaluations of $\chi^2(\vec{a})$ as possible. Values of N_{iters} and the other numbers used by the GA in this work are listed in appendix A.

Some of the properties of the custom genetic algorithm used in this work may be of general use. Because $\chi^2(\vec{a})$ is a minimized quantity, using it to rank population members requires a minor conversion. Letting $R = \sum_{\text{all } \vec{a}} \frac{1}{\chi^2(\vec{a})}$, the fitness of a given individual $\vec{a}$ is $\frac{1}{R\chi^2(\vec{a})}$. The GA also performs crossover using $0 \leq w \leq 1.49$. The purpose of this form of crossover is to prevent "inbreeding" (HH98). Additionally, the choice of $0 \leq w \leq 1$ assumes that the "parent" parameters bracket $\vec{a}_{\text{min}}$. However, it is quite possible that $\vec{a}_{\text{min}}$ lies slightly outside of the bounds formed by the parents. Haupt and Haupt state that one should use a combination of $0 \leq w \leq 1.5$ and $0 \leq w \leq 1.0$ to produce three offspring, keeping only those that do not generate invalid values of $\vec{a}$ (HH98). In the case of this work, invalid parameter values are not an issue,[5] so using $0 \leq w \leq 1.49$ in Eqs. (3.7) and (3.8) simplifies the algorithm.

3.2.4 Doing Things the Hard Way

While the techniques described in the last section are useful, one should be careful before using an intricate nonlinear optimization technique on the first try. As noted earlier, one may not need to use a GA if the problem does not involve multiple minima or if it is easy to locate a neighborhood around the global minimum. Additionally, some preprocessing of the data, $y(\vec{x})$, may be useful. It takes far less time to check a 3x3x3 tensor for uniformity than it would take a nonlinear optimizer to come to the same conclusion. Similarly, if the model never exceeds a value of 2.5, one should do something about those outliers in the data with a value of 10,000 before performing the optimization.[6]

Additionally, the design of $T(\vec{x}; \vec{a})$ can be as important as the details of the regres-

[5] Actually, the unscaled parameters do have bounds. However, the actual equation for $T(\vec{x}; \vec{a})$ rescales those parameters into an unbound form. See appendix A for details.

[6] There are special techniques for handling data containing outliers. See Ryan (Rya97) or Press et. al. (PTVF92) for details.

sion algorithm. For example, the function $T(\vec{x};\vec{a}) = a_0 + \frac{1}{a_1\vec{x}} + a_2\vec{x}$ is "falsely nonlinear," since one can use $a_1' = \frac{1}{a_1}$ to linearize the problem. Similarly, one can convert $\vec{T}(\vec{a}) = e^{a_0+a_1\vec{x}}$ to a linear problem by fitting $a_0 + a_1\vec{x}$ to $\vec{y}' = \ln(\vec{y})$. Finally, one can rewrite the function $f(\vec{x}) = \frac{\exp(a_0+a_1\vec{x})}{1+\exp(a_0+a_1\vec{x})}$ as $\frac{f(\vec{x})}{1-f(\vec{x})} = \exp(a_0 + a_1\vec{x})$ and convert this to a linear problem. This is called logistic regression, and has some special considerations to it (Rya97). Even if one must use a model that is a nonlinear function of the parameters, one can make the problem far easier to compute using a clever transformation of the parameters. Consider the example $\tanh\left(\frac{\vec{x}}{a}\right) = \tanh(a'\vec{x})$. As seen in Eq. (3.6), nonlinear quadratic regression algorithms require derivatives of $\chi^2(\vec{a})$. Derivatives with respect to the parameters of a term such as $\tanh\left(\frac{\vec{x}}{a}\right)$ are more complicated and require more computation than the derivative of $\tanh(a'\vec{x})$. Beware of making the problem more difficult than it is.

3.3 Goodness of Fit and Confidence Intervals

3.3.1 Correlation Coefficients

When performing regression, one cannot merely optimize the error, $\chi^2(\vec{a})$, by finding $\vec{a}_{\min}$. One must provide some quantitative description of how precisely $\vec{a}_{\min}$ describes the minimum and how well the model fits the data. The latter must be independent of the details of any particular model or data set, otherwise it will be useless for comparison purposes. The value of $\chi^2(\vec{a}_{\min})$, for example, depends heavily upon both the model and the data used in the regression. A better quantity for measuring the goodness of fit is the correlation coefficient.

The familiar correlation coefficient,

$$r = \frac{\overline{(\vec{y}-\overline{\vec{y}})(\vec{z}-\overline{\vec{z}})}}{\sigma_{\vec{y}}\sigma_{\vec{z}}}, \tag{3.9}$$

is actually a measure of how well two data sets, $\vec{y}$ and $\vec{z}$, are connected linearly. In the above equation, $\sigma_{\vec{y}}$ and $\sigma_{\vec{z}}$ denote the standard deviation and the bar denotes the mean over all data points. When $\vec{z}$ is a least-squares fit of a line to $\vec{y}$, r is a normalized criteria of the goodness of fit. There is a more general definition of the correlation coefficient, a definition

that applies to other models. It has the form:

$$R^2 = \frac{\overline{\left(\vec{T}\left(\vec{a}_{\min}\right) - \overline{\vec{y}}\right)^2}}{\sigma_{\vec{y}}^2}, \tag{3.10}$$

where $\overline{\left(\vec{T}\left(\vec{a}_{\min}\right) - \overline{\vec{y}}\right)^2} = \sigma_{\text{expl}}^2$ is termed the "explained variance" in the data and $\sigma_{\vec{y}}^2$ is the total variance (Spi95). Note the use of an uppercase letter to distinguish this "correlation coefficient" from the linear correlation coefficient. It is easy to show that

$$\sigma_{\vec{y}}^2 = \frac{1}{N_{\text{data}}}\chi^2 + \sigma_{\text{expl}}^2 + 2\sigma_{\text{cross}}$$

where $\sigma_{\text{cross}} = \overline{\left(\vec{y} - \vec{T}\right)\left(\vec{T} - \overline{\vec{y}}\right)}$ and N_{data} denotes the number of data points. Clearly, $\frac{1}{N_{\text{data}}}\chi^2$ describes variations in the data that are not explained by the model.[7] Thus, if $\sigma_{\text{cross}} = 0$, then σ_{expl}^2 does indeed describe variations in the data that are explained by the model. Under this case, where the cross-term is zero, Eq. (3.10) simplifies to:

$$R^2 = 1 - \frac{\chi^2}{\overline{(y - \overline{y})^2}}, \tag{3.11}$$

thus $R^2 \leq 1$. It is easy to show that any polynomial model, $T_i = \sum_{j=0}^{n} a_j x_i^j$, satisfies the criteria $\sigma_{\text{cross}} = 0$.

Unfortunately, $\sigma_{\text{cross}} \neq 0$ in general. This by itself is no great trouble, for as long as the cross-term is a positive number, the correlation coefficient defined by (3.10) will be between 0 and 1 inclusive. Alas, there is no way to prove that $\sigma_{\text{cross}} > 0$ in general, opening the possibility of $R^2 >$ 1in Eq. (3.10)! However, if one were to use a general form of correlation coefficient, defined qualitatively as $R^2 = \frac{\text{explained variance}}{\text{total variance}}$, then one could use Eq. (3.11), which fulfills this qualitative definition. The only time Eq. (3.11) fails as a definition of a general correlation coefficient is if $\frac{1}{N_{\text{data}}}\chi^2 > \sigma_{\vec{y}}^2$, in which case $R^2 < 0$.

Kvålseth (Kvå85) discusses the issues surrounding several forms of R^2, including Eqs. (3.10), (3.11), and the square of the correlation coefficient, Eq. (3.9), between $\vec{y}$ and $\vec{T}\left(\vec{a}_{\min}\right)$. He recommends the use of Eq. (3.11) in most cases, since those cases where it

[7] Ryan (Rya97) indicates that it might be better to use $\frac{1}{N_{\text{data}} - N_{\text{params}}}\chi^2$ in some cases, since this is an unbiased statistic.

returns a value of $R^2 < 0$ are due either to a complete lack of fit between the data and the model, or to outliers in the data. Data outliers were not an issue for this research; however, the reader should note that data with outliers requires special handling not discussed here. If one's data contains outliers, whether one uses linear or nonlinear regression, one should consult a text such as Ryan (Rya97) for techniques of robust regression and Kvålseth for a definition of R^2 which is insensitive to outliers.

3.3.2 Confidence Intervals: Concept

Regardless what form of goodness of fit or correlation coefficient one decides to use, one can always compute a quantitative description of how precisely $\vec{a}_{\min}$ describes the minimum. This is the confidence interval, denoted here by $\delta\vec{a}\,(q)$. Statistically, a confidence interval describes the probability, q, that the computed $\vec{a}_{\min}$ is within $\pm\delta\vec{a}$ of the "true" solution, assuming that any errors in the regression are due to random systematic errors (Spi95).

As stated earlier, one can always compute $\delta a_k\,(q)$, the confidence interval on the k^{th} parameter at q-percent. One randomly varies the k^{th} parameter about the minimum, computing $\chi^2\left(\vec{a}_{\min} + \delta a_k \hat{k}\right)$ for each perturbation (PTVF92). This Monte-Carlo technique generates a distribution of values of χ^2 about the computed minimum, $\chi^2\,(\vec{a}_{\min})$. The values of $\vec{a}_{\min} + \delta a_k \hat{k}$ which bound q-percent of this distribution is the confidence interval. One can then repeat this process with all of the parameters. Further details of this Monte-Carlo method are in Press et. al. (PTVF92).

3.3.3 Estimates of Confidence

The aforementioned Monte-Carlo method, though guaranteed to work in all cases, is nevertheless computationally intensive. There are, however, ways to analytically estimate the confidence intervals on the parameters. Consider Eq. (3.4), evaluated at $\vec{a}_{\min}$:

$$\Delta\chi^2 = \delta\vec{a}^{\,\mathrm{T}} \cdot \mathbf{D}\,(\vec{a}_{\min}) \cdot \delta\vec{a} \tag{3.12}$$

where $\mathbf{D}\,(\vec{a}_{\min})$ is the curvature matrix, Eq. (3.5), evaluated at the computed minimum, and $\Delta\chi^2 = \chi^2\,(\vec{a}_{\min} + \delta\vec{a}) - \chi^2\,(\vec{a}_{\min})$. Note that this equation defines an ellipsoid. In analogy

to the Monte-Carlo technique of the previous section, let us consider the confidence interval about the minimum for the k^{th} parameter, where $\delta\vec{a} = \delta a_k \hat{k}$. If one knew the probability distribution governing $\Delta\chi^2$, one could then replace $\Delta\chi^2$ by the value of that distribution at the q^{th} percentile.

At this point, Press et. al. (PTVF92) note that one can use the normal equations defined in Eq. (3.6) in combination with (3.12) to compute δa_k. They also introduce several assumptions in order to perform the computation. Let us not make the same assumptions at this point, but simply denote the probability distribution governing $\Delta\chi^2$ as $\mathcal{P}_{\Delta\chi^2}$. Furthermore, let $\mathcal{P}_{\Delta\chi^2}(q)$ denote the q^{th}-percentile value of $\mathcal{P}_{\Delta\chi^2}$. For variations, δa_k, about the minimum, the normal equations, (3.6), become

$$\mathbf{D}(\vec{a}_{\min}) \cdot \delta a_k = -\frac{\partial\left[\chi^2\right]}{\partial a_k}\left(\vec{a}_{\min} + \delta a_k \hat{k}\right). \tag{3.13}$$

One can now use Eqs. (3.6) and (3.13) to solve for δa_k by replacing $\Delta\chi^2$ in (3.6) by $\mathcal{P}_{\Delta\chi^2}(q)$ (PTVF92):

$$\delta a_k(q) = \pm\sqrt{\mathcal{P}_{\Delta\chi^2}(q)} \cdot \sqrt{\left[\mathbf{D}(\vec{a}_{\min})\right]^{-1}_{kk}}, \tag{3.14}$$

where $\left[\mathbf{D}(\vec{a}_{\min})\right]^{-1}$ denotes the inverse of the curvature matrix evaluated at the minimum. $\delta a_k(q)$ is thus the confidence interval for the k^{th} parameter at the q^{th} percentile.

There are only two lingering questions to answer: is $\mathbf{D}(\vec{a}_{\min})$ invertible and what is $\mathcal{P}_{\Delta\chi^2}$? The answer to the first question is "yes" if the determinate of $\mathbf{D}(\vec{a}_{\min})$ is nonzero. Now, at a true minimum, $\Delta\chi^2 > 0$ for any perturbation away from the minimum. By Eq. (3.12), this implies that $\vec{b}^{\mathrm{T}} \cdot \mathbf{D}(\vec{a}_{\min}) \cdot \vec{b} > 0 \quad \forall\, \vec{b}$. One can then use this condition to prove that the curvature matrix has the following properties at any true minimum:

(1) $\operatorname{tr}(\mathbf{D}) > 0$,

(2) The eigenvalues of $\mathbf{D}$ are all positive real numbers,

(3) $\det|\mathbf{D}| > 0 \qquad \Rightarrow$ $\mathbf{D}$ is invertible,

(4) $[\mathbf{D}]^{-1}_{ii} > 0 \quad \forall\, i$,

(5) $|D_{ik}| < \sqrt{D_{ii}\, D_{kk}} \qquad i \neq k$.

The first property is straightforward to prove. Properties (2)-(4) are easy to prove using a similarity transform to diagonalize $\mathbf{D}$. To prove the last property, use the condition on $\mathbf{D}$ with the vector $b_j = \delta_{ij} + c\delta_{kj}$, where δ_{ij} denotes the Kronecker delta function, to solve for an interval bounding c. One then uses the midpoint of that interval to compute the inequality of property (5).

Leaving $\mathcal{P}_{\Delta\chi^2}$ unspecified gives the researcher some freedom in the choice of a hypothesis for the data. Recall that a confidence interval uses the basic hypothesis that all deviations of the model from the data, $\vec{\mathcal{E}}(\vec{a}_{\min})$, are due to random systematic errors. As of yet, we have made no assumptions about what kind of probability distribution, $\mathcal{P}_{\mathcal{E}}$, might govern the errors $\mathcal{E}_i(\vec{a}_{\min})$. $\mathcal{P}_{\mathcal{E}}$ then determines the probability distribution governing $\chi^2(\vec{a})$ and hence $\mathcal{P}_{\Delta\chi^2}$. Though it is both commonplace and reasonable to assume that $\mathcal{P}_{\mathcal{E}}(x) = \frac{1}{\sigma_{\mathcal{E}}\sqrt{2\pi}} e^{-x^2/2\sigma_{\mathcal{E}}^2}$, the normal distribution, Eq. (3.14) does not require that the errors be normally distributed. If, say, $\mathcal{P}_{\mathcal{E}}$ were some strange probability distribution that could not be approximated by a normal distribution, one could still use (3.14) to compute confidence intervals. One need only derive a form for $\mathcal{P}_{\Delta\chi^2}$ from $\mathcal{P}_{\mathcal{E}}$.

Let us now use the assumption that $\mathcal{P}_{\mathcal{E}}$ is a normal distribution with zero mean and a variance, $\sigma_{\mathcal{E}}^2$. If one uses data with associated systematic errors, then this assumption applies to $\vec{\mathcal{E}}$ as defined in Eq. (3.3). The square of a normally distributed variable with unit variance is a χ^2-distributed variable with one degree of freedom (Spi95). Furthermore, the χ_N^2 distribution, where N denoted the number of degrees of freedom, has the property $\chi_M^2 \pm \chi_N^2 = \chi_{M\pm N}^2$ (Spi95). Although $\mathcal{P}_{\mathcal{E}}$ does not have unit variance, the probability distribution of the quantity $\frac{\vec{\mathcal{E}}}{\sigma_{\mathcal{E}}}$ does. Thus, for a randomly selected value of $\vec{a}$, $\frac{\chi^2(\vec{a})}{\sigma_{\mathcal{E}}^2} \equiv \hat{\chi}^2(\vec{a})$ is a χ^2-distributed variable with N_{data} degrees of freedom. This is not the case, however, for values of $\vec{a}$ computed from the data.

Whenever one uses data to compute a single quantity, any statistic of the data which also involves the derived quantity "loses" a degree of freedom (Spi95,PTVF92). Hence $\hat{\chi}^2(\vec{a}_{\min})$ has $N_{\text{data}} - N_{\text{params}}$ degrees of freedom, and $\hat{\chi}^2\left(\vec{a}_{\min} + \delta a_k \hat{k}\right)$, which permits the k^{th} parameter to vary, has $N_{\text{data}} - N_{\text{params}} + 1$ degrees of freedom [N_{params} denotes the number of parameters]. Thus $\frac{\Delta\chi^2}{\sigma_{\mathcal{E}}^2}$ is a χ^2-distributed variable with one degree of freedom. It is then straightforward to show that $\mathcal{P}_{\Delta\chi^2} = \chi_1^2 \cdot \sigma_{\mathcal{E}}^2$, where χ_1^2 denotes a one degree of

freedom chi-squared distribution. Therefore:

$$\delta a_k(q) = \pm\sigma_{\mathcal{E}} \sqrt{\chi_1^2(q) \cdot [\mathbf{D}(\vec{a}_{\min})]^{-1}_{kk}}, \tag{3.15}$$

where $\chi_1^2(q)$ is the q^{th}-percentile value of the χ_1^2 distribution.

The previous expression is similar to the form of the confidence interval derived by Press et. al. (PTVF92). Ryan (Rya97) derives a quantity similar to (3.15), except that he finds that $\mathcal{P}_{\Delta\chi^2} = \left[t_{N_{\text{data}}-N_{\text{params}}}\left(\frac{1+q}{2}\right)\right]^2 \cdot \sigma_{\mathcal{E}}^2$, where t_M is a Student-t distribution with M-degrees of freedom. Using properties of this distribution, and properties of a few others, one can show that $t_M\left(\frac{1+q}{2}\right) = \sqrt{\chi_1^2(q)}$ for $M \geq 100$. Ryan further notes that the expectation value of $\sigma_{\mathcal{E}}^2$ is simply $\frac{1}{N_{\text{data}}}\chi^2(\vec{a}_{\min})$ (Rya97). It is fairly straightforward to show this algebraically. Therefore, the final form for the algebraic estimate of the confidence interval is:

$$\delta a_k(q) = \pm\sqrt{\chi_1^2(q) \cdot [\mathbf{D}(\vec{a}_{\min})]^{-1}_{kk} \cdot \frac{1}{N_{\text{data}}}\chi^2(\vec{a}_{\min})}. \tag{3.16}$$

Lastly, because we have assumed that the $\vec{\mathcal{E}}$ are small and have zero mean, we can approximate the curvature matrix using $\mathbf{D}(\vec{a}_{\min}) \cong \frac{1}{2}\left(\vec{\nabla}_{\vec{a}}\,\chi^2\right) \cdot \left(\vec{\nabla}_{\vec{a}}\,\chi^2\right)\Big|_{\vec{a}_{\min}}$ (PTVF92). Note that this approximate $\mathbf{D}(\vec{a}_{\min})$ is a symmetric matrix, making computation and inversion much easier.

3.3.4 Final Cautions

This chapter has focussed on the task of fitting a model to data. It began with a basic definition of least-squares fitting in section 3.1 with Equations (3.1-3.3) and a basic description of regression using models which are nonlinear functions of their parameters. It then proceeded to the method of genetic algorithms in section 3.2, including how one could use a genetic algorithm to compute an initial parameter value for a "traditional" nonlinear regression method. Having defined a confidence interval for the parameters with Equation (3.16), the chapter now concludes with some remarks about possible pitfalls.

Sections 3.2.4 and 3.3.1 both remarked on situations that might need special handling. Specifically, one must be certain that the model does indeed describe the data within reason, especially when using Eq. (3.11) as a measure of goodness of fit. If one does use

Eq. (3.11) as a measure of goodness of fit, one must be sure to compute it using the actual data and model, not linearized versions. Consider once again the model $\vec{T}(\vec{a}) = e^{a_0 + a_1 \vec{x}}$, fit using the linearization $\vec{T}' = a_0 + a_1 \vec{x}$ and $\vec{y}' = \ln(\vec{y})$. To compute R^2, one must use a χ^2 calculated with $\vec{T}$ and $\vec{y}$, **not** $\vec{T}'$ and $\vec{y}'$, which will likely produce a different value of χ^2.

The mention of linearizing the model raises the issue of types of regression. One should use a model and method appropriate to the problem. There are some cases where it makes sense to linearize or perform some other form of transformation on the model and avoid using a full nonlinear regression method. There are other cases where a nonlinear model would describe the data best. When one's data has outliers, one should use some form of robust regression method, such as those described in Ryan (Rya97).

Lastly, one should design the model itself so that it, and, if necessary, its derivatives are easy to compute and have the "most compact" form possible. Successive derivatives of a nonlinear model with a parameter of $\frac{1}{a_i}$ will be far less complicated if one transforms this parameter to $a_i' = \frac{1}{a_i}$. Such a rescaling has no effect on the value of χ^2. It may, however, have an effect on the regression algorithm's convergence time. It behooves the researcher to transform the parameters in such a way that there are no "flat regions," i. e. a range of values in a_i where large changes in a_i produce only minute changes in χ^2.

For a rescaled model, i. e. $\vec{T}(\vec{a}) \rightarrow \vec{T}(f(\vec{a}))$, one should compute $\mathbf{D}$, the curvature matrix, using the unscaled version. Otherwise, the confidence intervals defined by Eq. (3.16) will describe the transformed version of the parameters, $f(\vec{a})$, not the parameters themselves. One could transform these "$\delta f(a_k)(q)$" back into the true confidence intervals, $\delta a_k(q)$, though this may be more work than it is worth.

There are two other concerns involving the confidence intervals defined by Eq. (3.16). Suppose that 9 of your 12 parameters, $\vec{a}$, become undefined when $a_3 = 0$. It makes little sense, then, to compute all 144 elements of $\mathbf{D}(\vec{a}_{\min})$ and then invert it. In fact, it is quite possible that only nine of those 144 elements will be nonzero, the nine elements of $\mathbf{D}(\vec{a}_{\min})$ corresponding to a_3 and the two parameters which are still defined. The full matrix may even be ill-conditioned or uninvertible. In such a scenario, where the value of one parameter renders one or more other parameters undefined, one should only invert the submatrix of $\mathbf{D}(\vec{a}_{\min})$ corresponding to the derivatives of the defined parameters. The inverted submatrix will provide rough estimates of the confidence intervals on the defined parameters. One can

then set the confidence intervals of the parameters with undefined values to some appropriate number.

Lastly, consider how one might compute $\mathbf{D}(\vec{a}_{\min})$. It may be possible to derive $\mathbf{D}(\vec{a}_{\min})$ analytically, yet this may be more work than is necessary. A curvature matrix calculated from finite-difference derivatives of χ^2 may work just as well. In one particular case, encountered in this research, it will work better. Appendix A contains further details on this matter.

Chapter 4

Persistence Barriers: Theory

4.1 Introduction

The persistence barrier, shown in Figures 1.2 and 1.3, has raised a speculation (CZBC95): could this "barrier" be an artifact of the statistics? To answer this question, one should examine persistence in detail. Such an examination is the focus of this chapter.[1] Persistence, a form of lag autocorrelation, is part of a class of statistics known as cyclostationary statistics (HN96). Cyclostationary statistics are the appropriate class of statistics for time series with an external periodicity, such as an annual cycle, and depend on some "phase" of that cycle, such as months of the year.

Section 4.2 presents a formal definition of cyclostationary statistics, and then calculates the persistence for a progression of simple but increasingly complex time series. Some time series have a persistence which does not depend on the phase of the year, others have phase-dependence which resemble the persistence barrier seen in ENSO data, and others have a phase-dependence which looks completely different from that seen in ENSO data. Our focus here will be the persistence barrier, those phase-dependent patterns which match the ENSO data. Section 4.2 shows that a persistence barrier is due to the presence of harmonics in a time series, harmonics satisfying certain conditions. Finally, section 4.3 develops a method to quantitatively measure the properties of a persistence barrier, using the nonlinear optimization techniques of chapter 3.

[1] The material in this chapter has been submitted to the Journal of the Atmospheric Sciences and was originally co-authored with Dr. Jeffrey B. Weiss (WW97a).

4.2 The Mathematics of Persistence

4.2.1 Definition

Common forms of statistical analysis assume that the data is stationary in time. Most geophysical systems, however, are characterized by periodic cycles, such as the annual cycle or the diurnal cycle. It is thus appropriate to include this external periodicity in the statistical analysis; this results in cyclostationary statistics (HN96). Cyclostationary statistics differ from stationary statistics in that the various quantities, such as the mean, variance, etc., are periodic functions of the phase with respect to the external periodic cycle.

In this paper, the relevant external periodicity is the annual cycle, so let us measure time in units of years. For the following definitions, let $x(t)$ be a time series, and denote the phase of the year by p, $0 \leq p < 1$. The cyclostationary mean is defined as

$$\langle x(t) \rangle (p) = \frac{1}{N} \sum_{y=0}^{N-1} x(y+p) \,. \tag{4.1}$$

Thus, $\langle \cdots \rangle (p)$ represents an average over N years at fixed phase p. In the case of the ENSO data, Eq. (4.1) is merely the mean annual cycle. It will be convenient to also write the cyclostationary mean as $\langle x \rangle (t)$ where the dependence on t is understood to mean the phase of t. The cyclostationary variance is defined in terms of the departure from the annual cycle, i. e. the anomaly, $x'(t) = x(t) - \langle x \rangle (t)$,

$$\sigma_y^2(p) = \frac{1}{N-1} \sum_{y=0}^{N-1} \left[x'(y+p) \right]^2 . \tag{4.2}$$

The cyclostationary lag autocorrelation, also called persistence, is then

$$C(p,l) = \frac{\langle x'(t) x'(t-l) \rangle (p)}{\sigma_y(p) \sigma_y(p-l)} \,. \tag{4.3}$$

Let us refer to p as the leading phase, or just phase if there is no confusion. I will hereafter focus on lags of one year or less, $l \leq 1$, and will refer to the resulting field as a persistence map.

4.2.2 Persistence of a Sine Wave

To gain insight into the behavior of persistence, let us first consider simple harmonic time series for which one can calculate the statistics in closed form. Considering simple harmonic functions additionally lays the groundwork for interpreting the persistence of more complex time series in terms of their Fourier transforms.

First, consider a single sinusoidal periodic oscillation,

$$x(t) = a\sin(2\pi\nu t + \phi), \tag{4.4}$$

where $\nu > 0$ [negative frequencies are equivalent to shifting ϕ by π]. An important identity needed to analytically calculate the persistence of (4.4) is:

$$\left\langle e^{\pm i2\pi\nu t}\right\rangle(p) = \begin{cases} 0 & \nu \neq n, \\ e^{\pm i2\pi\nu p} & \nu = n, \end{cases} \tag{4.5}$$

where n is an integer. The persistence, Eq. (4.3), divides into three cases depending on whether: 1) ν is an integer n; 2) ν is a half-integer $(2n+1)/2$; or 3) neither:

$$C(p,l) = \begin{cases} \text{undefined}, & \nu = n; \quad (4.6) \\ \dfrac{\sin\left(2\pi\nu(p-\beta)\right)\sin\left(2\pi\nu(p-\beta-l)\right)}{\left|\sin\left(2\pi\nu(p-\beta)\right)\sin\left(2\pi\nu(p-\beta-l)\right)\right|}, & \nu = \frac{2n+1}{2}; \quad (4.7) \\ \cos(2\pi\nu l), & \nu \neq \frac{n}{2}, \quad (4.8) \end{cases}$$

where $\beta \equiv -\phi/2\pi\nu$. A time series with $\nu = n$, Eq. (4.6), corresponds to an annual cycle or a harmonic of an annual cycle. Since the persistence is calculated from the departure from the annual cycle, it is undefined in this case. If ν is neither an integer nor a half-integer [Eq. (4.8)] then the persistence does not depend on leading phase.

The remaining case, where ν is a half-integer [Eq. (4.7)] corresponds to a biennial cycle or one of its harmonics. In this case, $C(p,l)$ can be written as

$$C(p,l) = \frac{x(p)x(p-l)}{|x(p)x(p-l)|}, \tag{4.9}$$

a quantity divided by its absolute value, and thus is either 1 or -1 indicating perfect cor-

relation or anticorrelation. By defining a periodic step function (i. e. a square wave) as $\Theta(x) = (-1)^{\mathrm{Int}(x)}$, where $\mathrm{Int}(x)$ is the integer part of x, one can write C as

$$C(p,l) = \Theta\left(2\pi\nu\left(p-\beta\right)\right)\Theta\left(2\pi\nu\left(p-\beta-l\right)\right). \tag{4.10}$$

Thus, in this case, C depends on the leading phase as well as the lag. A contour plot of Eq. (4.10) for $\nu = \frac{1}{2}$ appears in Figure 4.1. Note the similarities to the persistence maps of the SOI and NINO3 sea surface temperature, Figures 1.2 and 1.3. The parameter β determines the position of the transition in the map from positive to negative persistence. From Eq. (4.4), one sees that β is also the time at which $x(t)$ crosses zero.

4.2.3 Persistence of Two Sine Waves

Let us now turn to a time series composed of two sine waves:

$$x(t) = a_1 \sin(2\pi\nu_1 t + \phi_1) + a_2 \sin(2\pi\nu_2 t + \phi_2), \tag{4.11}$$

where ν_1 and ν_2 are positive numbers. The calculations are similar to those of section 4.2.2. Again, the persistence breaks into three cases: 1) ν_1 or ν_2 are integers; 2) neither ν_1 nor ν_2 are integers and either ν_1 or ν_2 is a half-integer; and 3) ν_1 and ν_2 are neither integers nor half-integer.

If either ν_1 or ν_2 is an integer, then that Fourier mode of the time series is removed by computing the anomaly. The persistence is then determined from Eqs. (4.6)-(4.8) applied to the remaining mode. When neither frequency is an integer, the persistence can still be found in closed form. To facilitate the discussion of these cases, note that one can express persistence in the form:

$$C(p,l) = \frac{V(p,l)}{\sqrt{V(p,0)V(p-l,0)}}, \tag{4.12}$$

where the function $V(p,l)$ is, apart from a constant factor, the cyclostationary covariance, $\langle x(t)\, x(t-l)\rangle\,(p)$. For the case where one of the frequencies is an odd half-integer and the other is not, specifically $\nu_1 = n/2$ and $\nu_2 \neq n/2$, $V(p,l)$ has the form:

$$V(p,l) = \widehat{C} + \alpha^2 \cos\left(2\pi n\left[p-\beta-l/2\right]\right), \tag{4.13}$$

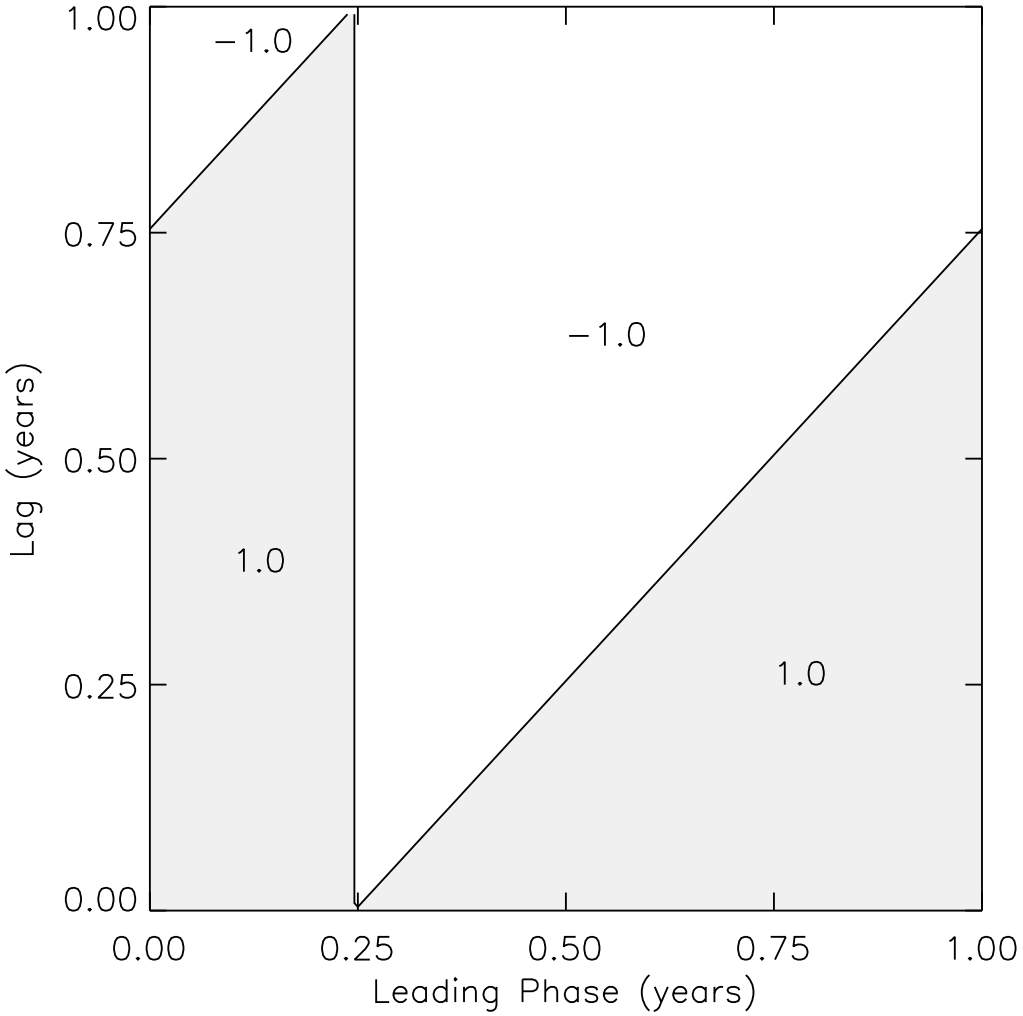

Figure 4.1: Persistence map of a single sine wave, Eq. (4.10), with $\beta = 0.25$ and $\nu_1 = \frac{1}{2}$. The shaded area denotes a persistence of $+1$, while the unshaded area is a persistence of -1.

where $\alpha = a_1/a_2$ is the ratio of amplitudes, $\widehat{C} = \alpha^2 \cos(2\pi\nu_1 l) + \cos(2\pi\nu_2 l)$, and $\beta = -\phi_1/\pi n$. When both frequencies are half-integers, then C can be written in the same form as the single frequency, biennial case, Eq. (4.9). It cannot, however, be reduced to Eq. (4.10). C has the values ± 1, and the transitions occur at phases where $x(t)$ crosses zero. Persistence maps for both of these cases are shown in Figure 4.3. Notice that phase-dependence can appear in forms other than that found in the SOI and NINO3 SST. I shall use the term "barrier" to denote those phase-dependent patterns in the persistence map which are qualitatively similar to the pattern seen in the SOI and NINO3 SST. Note also that in the case where only one of the two modes is a biennial cycle, the drop in correlation is continuous, as for the ENSO data.

I have additionally explored the behavior of $C(p, l)$ for values of a_1, a_2 , ϕ_1, and ϕ_2 different than those used in Figure 4.3. The phases ϕ_1 and ϕ_2 determine the position of any leading-phase-dependent features. When $C(p, l)$ is continuous, as in the case defined by Eq. (4.13), the values of a_1 and a_2 control the strength of the gradients of the continuous features. For example, decreasing $|a_1 - a_2|$ would narrow the transition region between positive and negative persistence. Furthermore, if $a_1 \gg a_2$, then $x(t) \cong a_1 \cos(2\pi\nu_1 t + \phi_1)$ and the persistence changes accordingly.

The final case, where ν_1 and ν_2 are neither integers nor half-integers, falls into two subcases:

$$V(p,l) = \begin{cases} \widehat{C}, & \nu_1 \pm \nu_2 \neq m; \quad (4.14) \\ \widehat{C} + 2\alpha \cos(\pi\hat{\nu} l) \cos(2\pi m(p - \beta - l/2)), & \nu_1 \pm \nu_2 = m; \quad (4.15) \end{cases}$$

where m is an integer, α and $\widehat{C}$ are defined as in Eq. (4.13), $\hat{\nu} = \nu_1 \mp \nu_2$, and $\beta = -(\phi_1 \pm \phi_2)/2\pi m$. Figure 4.5 contains plots of the persistence for the $\nu_1 \pm \nu_2 = m$ case [Eq. (4.15)] for a variety of frequencies. The values of α and $\hat{\nu}$ determine how strong the gradients are in the persistence map. Specifically, as α nears 1, and $\hat{\nu}$ nears 0, the gradients of the continuous features grow stronger, becoming infinite at $\alpha = 1$. As might be expected, if $\alpha \gg 1$ or $\alpha \ll 1$, $C(p, l)$ loses phase-dependence. When the amplitudes are equal, i. e. $\alpha = 1$, the persistence

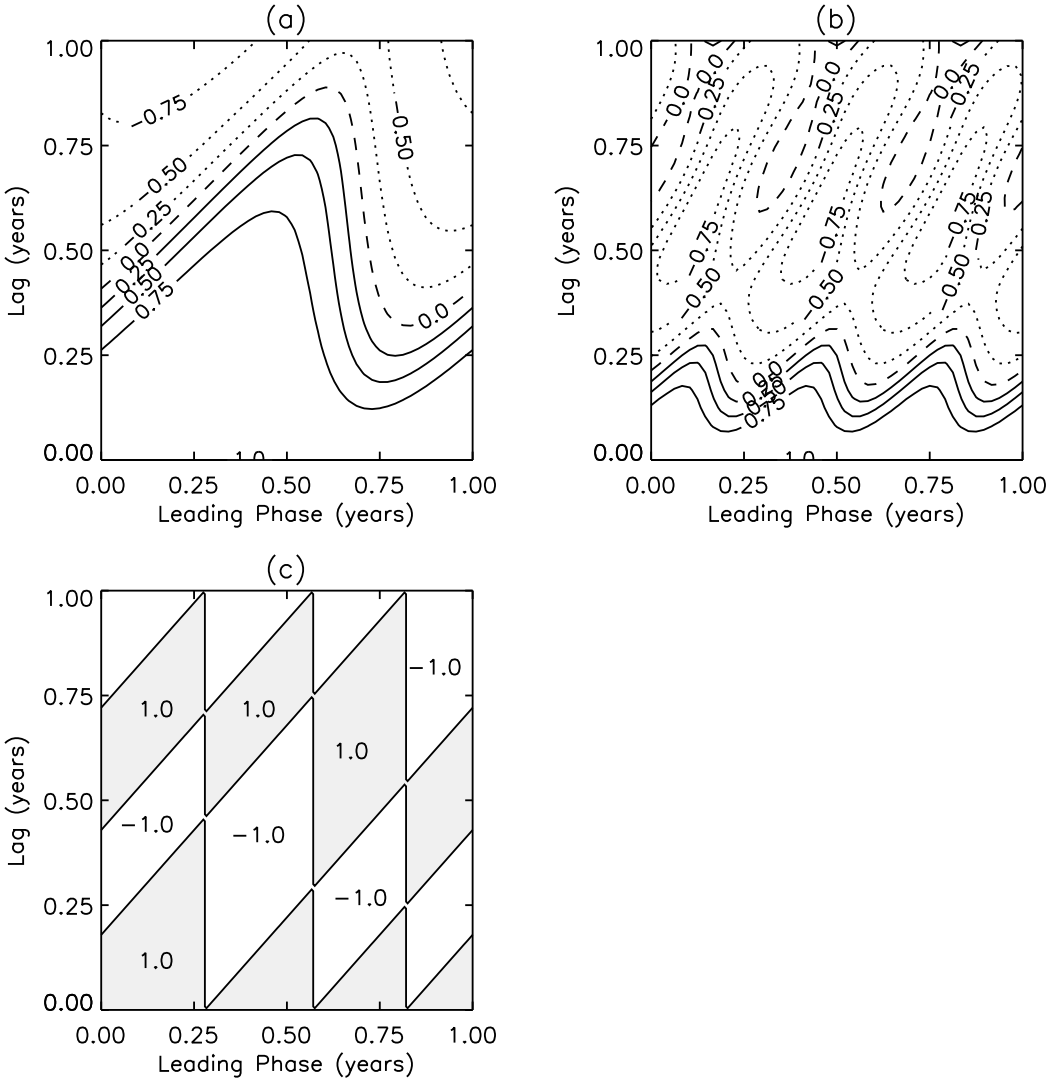

Figure 4.3: Persistence maps of time series composed of two harmonics, Eq. (4.11), for four different pairs of frequencies. In all cases, $a_1 = 1.0$, $a_2 = 0.75$, and $\phi_1 = \frac{\pi}{3}$, $\phi_2 = \frac{\pi}{2}$. All four cases represent time series with a biennial cycle present. Shading is as in Figure 4.1. Contour intervals are in steps of $\frac{1}{4}$, with the contour lines drawn as described in Figure 1.2. (a) $\nu_1 = \frac{1}{2}$ and $\nu_2 = \frac{3}{10}$. (b) $\nu_1 = \frac{4}{5}$ and $\nu_2 = \frac{3}{2}$. (c) A biennial cycle with $\nu_1 = \frac{3}{2}$ and $\nu_2 = \frac{1}{2}$.

defined by Eq. (4.15) simplifies to

$$C(p,l) = \cos(\pi\hat{\nu}l)\Theta\left(m(p-\beta)\right)\Theta\left(m(p-\beta-l)\right), \tag{4.16}$$

which is the barrier of a single biennial harmonic modified by a phase-independent factor which oscillates with lag. When the two frequencies are equal, $\nu_1 = \nu_2 = m/2$, then $\hat{\nu} = 0$, and the biennial result, Eq. (4.10), is exactly recovered.

Thus, when $\nu_1 \pm \nu_2 = m$, the persistence is similar to that of a single biennial cycle. Denote any pair of harmonics with frequencies ν_i and ν_j which satisfy $\nu_i \pm \nu_j = m$ as **$\mathbf{m}^{\pm}$-complementary**. From Figure 4.5, one sees that $\nu_i + \nu_j = 1$ results in a barrier, while other $\mathrm{m}^{\pm}$-complementary harmonics produce other forms of phase-dependence. Let us refer to harmonics with this relationship, $\nu_i + \nu_j = 1$, as unit-complementary. For unit-complementary harmonics, the smaller the separation of the frequencies and the more equal their amplitudes, the stronger the barrier will be. Note that both frequencies, ν_i and ν_j, in a unit-complementary pair must be less than 1, since both frequencies are also positive.

Summarizing the analytic results of the previous two sections, Table 4.1 describes those cases where the persistence contains some form of phase-dependence. Persistence takes the form of a barrier for a single biennial model, the sum of a biennial and non-biennial mode, or for the sum of a unit-complementary pair of harmonics. Other forms of phase-dependence occur for a biennial harmonic or other $\mathrm{m}^{\pm}$-complementary pairs.

4.2.4 Persistence of a Stochastic Timeseries

Let us next examine the persistence of a random time series. Consider the following first order Markov process:

$$\dot{x} = -\frac{1}{\tau}x + \xi(t), \tag{4.17}$$

where $\xi(t)$ is Gaussian white noise with zero mean and unit variance. This stochastic system describes a diffusion process, and is also known as the Ornstein-Uhlenbeck process.

Because (4.17) is a stochastic timeseries, its ensemble average persistence map is the appropriate quantity to calculate. The derivation involves standard techniques of stochastic calculus (Gar90). Once one shows that the cyclostationary variance [Eq. (4.2)] of (4.17)

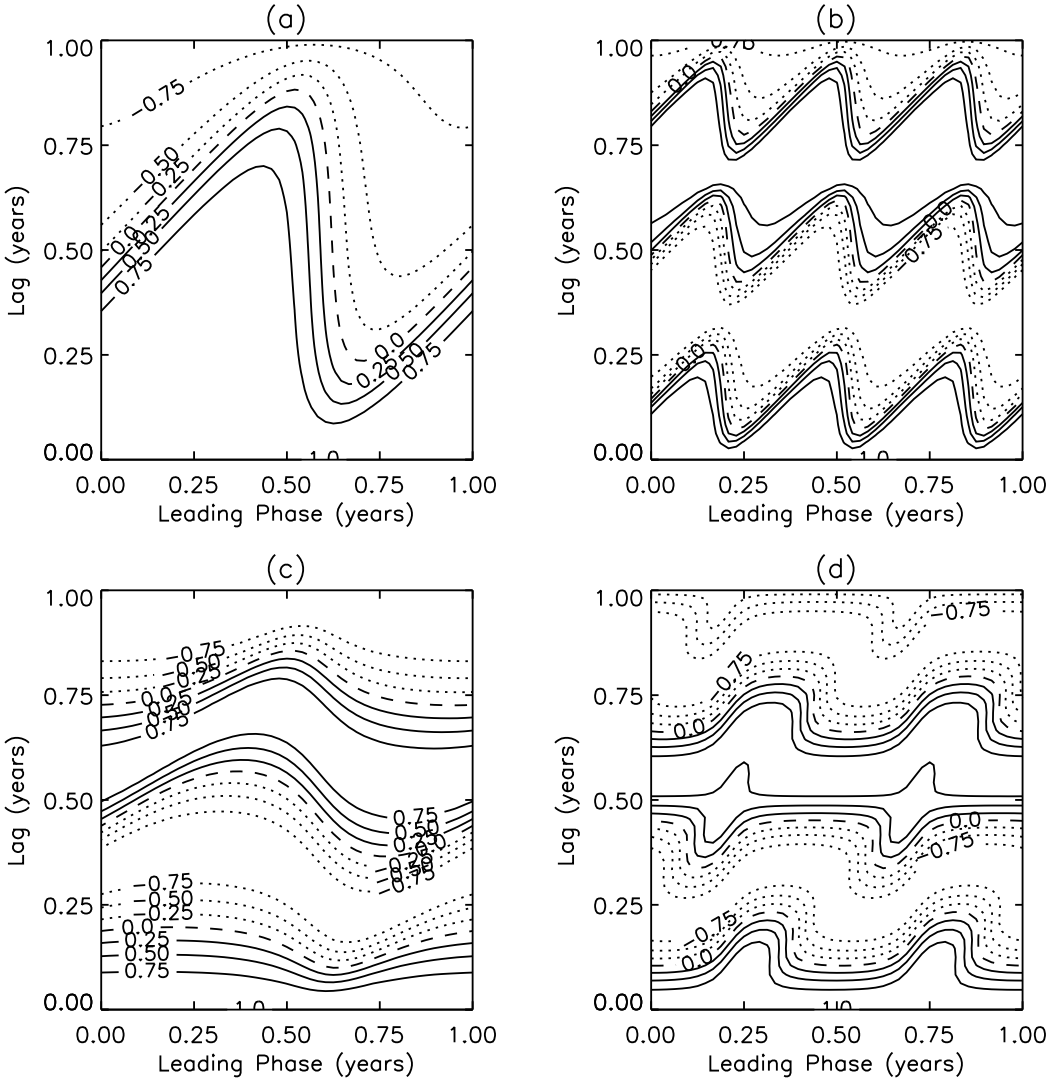

Figure 4.5: Persistence map of Eq. (4.15) with $a_1 = 1.0$, $a_2 = 0.5$, $\phi_1 = \frac{\pi}{3}$, $\phi_2 = \frac{\pi}{2}$, and $\nu_1 \pm \nu_2 = m \in \mathcal{Z}$. Contours are drawn as described in Figure 4.3. (a) $\nu_1 = \frac{2}{5}$, $\nu_2 = \frac{3}{5}$, $\nu_1 + \nu_2 = 1$. (b) $\nu_1 = \frac{7}{5}$, $\nu_2 = \frac{8}{5}$, $\nu_1 + \nu_2 = 3$. (c) $\nu_1 = \frac{8}{5}$, $\nu_2 = \frac{3}{5}$, $\nu_1 - \nu_2 = 1$. (d) $\nu_1 = \sqrt{3}$, $\nu_2 = 2 + \sqrt{3}$, $\nu_1 - \nu_2 = -2$.

Table 4.1: Summary of phase-dependence in persistence maps of time series from a single sine wave with frequency ν [Eq. (4.4)] and from the sum of two sine waves with frequencies ν_1 and ν_2 [Eq. (4.11)].

persistence pattern	ν	ν_1, ν_2
barrier	biennial [Eq. (4.10)]	biennial + non-biennial [Eq. (4.13)]
		unit-complementary, $\nu_1+\nu_2=1$ [Eq. (4.15)]
other phase dependence	biennial harmonic [Eq. (4.10)]	biennial + any biennial harmonic [Eq. (4.9)]
		two biennial harmonics [Eq. (4.9)]
		$m^{\pm}$-complementary with $\nu_1+\nu_2\neq 1$ [Eq. (4.15)]

is a constant, the remainder of the derivation is straightforward. The ensemble average persistence is simply:

$$\langle C(p,l)\rangle_{\text{ensemble}} = e^{-l/\tau}. \tag{4.18}$$

The parameter τ thus signifies both a diffusion time for the Ornstein-Uhlenbeck process and a decorrelation time for the persistence. Notice that the persistence map, Eq. (4.18), depends only on the lag.

4.2.5 Persistence of a General Time Series

A general time series is described by a continuous Fourier transform,

$$x(t) = \int_0^\infty d\nu\, a(\nu)\sin(2\pi\nu t + \phi(\nu)), \tag{4.19}$$

and the amplitudes and phase angles are now functions of the frequency. For the general case, calculating the persistence in closed analytic form appears intractable.

There is a special case where the persistence can be easily calculated. If $a(\nu) = 0$ for the biennial cycle, for all of its harmonics, $\nu = n/2$, and for all $\mathrm{m}^{\pm}$-complementary frequencies, then the persistence is the sum of the persistence of the individual harmonics given by the result for a single sine wave, Eq. (4.8). In general, however, such superposition fails.

If the spectrum of x consists of a single biennial peak or two $\mathrm{m}^{\pm}$-complementary peaks, then the persistence can be calculated in certain limits. Consider a spectrum with peaks at ν_1 and ν_2, $a(\nu) = a_1\eta(\nu-\nu_1) + a_2\eta(\nu-\nu_2)$, where $\eta(x)$ is a symmetric, positive-definite function with its maximum at zero and a width determined by a parameter, Δ. If ν_1 and ν_2 are unit-complementary, $\nu_1+\nu_2 = 1$, there is now a continuum of unit-complementary harmonics contained in the two peaks. Similarly, in the case of a single peak at $\nu_1 = \nu_2 = \frac{1}{2}$, there is a continuum of unit-complementary harmonics surrounding this one peak.

Assume that the peaks are narrow, $\Delta \ll 1$, and that their shape is Gaussian, $\eta(\nu) = \frac{1}{\sqrt{2\pi\Delta^2}} e^{-\nu^2/2\Delta^2}$. The latter assumption merely simplifies the form of the final result; any narrow distribution will suffice. Let us also assume that $\nu_1 < \nu_2$, $\nu_1 > \Delta$, $\nu_2 < 1-\Delta$, and $\nu_1+\Delta < \nu_2-\Delta$. In other words, the two peaks neither overlap each other nor extend beyond

the frequencies 0 and 1. Using these assumptions, one can show that the cyclostationary mean of Eq. (4.19) is approximately zero.

Deriving the cyclostationary covariance, $\langle x(t)x(t-l)\rangle_y(p)$ is lengthy and not straightforward. I therefore wish to describe some of its more pertinent points. The covariance contains a product of two integrals. One commutes these integrals with the cyclostationary average and perform the cyclostationary mean on the integrand first. Using Eq. (4.5) and trigonometric identities, one can then reduce one of the two integrals to a discrete sum. One then makes use of the symmetry and narrowness of the peaks to approximate the integral of the sum to a single integral from $-\infty$ to ∞. After some additional reductions, one can show that:

$$\begin{aligned} V(p,l) &\cong \widehat{C}e^{-2\pi^2 l^2 \Delta^2} \\ &\quad -\alpha \int_{-\infty}^{\infty} \mathrm{d}\nu \; e^{-(\nu/\Delta)^2} \left[\cos\left(p_1+\phi_{12}\right)+\cos\left(p_2+\phi_{21}\right)\right], \end{aligned} \tag{4.20}$$

where $V(p,l)$ is defined by Eq. (4.12), α and $\widehat{C}$ are the same as in Eq. (4.14), $p_i = 2\pi\left(p - l\left(\nu_i - \nu\right)\right)$, and $\phi_{ij} = \phi\left(\nu_i+\nu\right)+\phi\left(\nu_j-\nu\right)$. This equation can also be used for the case of a single biennial peak, $\nu_1 = \nu_2 = \frac{1}{2}$, by setting $a_1 = a_2$.

The persistence can be further simplified in two different limits. If the phase angle, $\phi(\nu)$, is rapidly varying over the width of a peak, then $\cos(p_i+\phi_{ij})$ acts as a random variable and the integral in Eq. (4.20) vanishes. The persistence then becomes

$$C(p,l) \cong e^{-2\pi^2 l^2 \Delta^2} \frac{\widehat{C}}{\alpha^2+1}. \tag{4.21}$$

This is the persistence of the two-mode non-complementary frequency case, Eq. (4.14), multiplied by an exponential decay in lag, a decay whose rate is governed by the width of the peak, Δ. This result can be generalized to multiple sets of unit-complementary harmonics: if the phase varies rapidly over all of the peaks, then the resulting persistence map is independent of the phase of year.

One can also simplify the persistence in the limit of slowly varying phase angles. Assuming $\phi''\left(\nu_i\right)\Delta^2 \ll 1$, and $\phi'(\nu_1)-\phi'(\nu_2) \ll 1$, where prime denotes differentiation with respect to ν, one Taylor expands the cosine terms in Eq. (4.20) around $\nu = 0$. If one keeps

terms up to $\mathcal{O}\left(\Delta^2\right)$, Eq. (4.20) reduces further to

$$\begin{aligned} V(p,l) &\cong e^{-2\pi^2 l^2 \Delta^2} \left\{ \widehat{C} - 2\alpha \cos\left(\pi \hat{\nu} l\right) \Big[\cos\left(2\pi \left[p - \beta - l/2\right]\right) \right. \\ &\quad \left. - \epsilon_2 \sin\left(2\pi \left[p - \beta - l/2\right]\right) \Big] \right\}, \end{aligned} \tag{4.22}$$

where $\hat{\nu} = \nu_1 - \nu_2$, $\epsilon_2 = \left(\phi''(\nu_1) + \phi''(\nu_2)\right) \Delta^2/4$, and $\beta = -(\phi(\nu_1) + \phi(\nu_2))/2\pi$. This expression is similar to the two-mode unit-complementary case, Eq. (4.15), and reduces to it in the limit $\Delta \to 0$. Therefore, when ϕ is a slowly-varying function, the persistence of a time series with two narrow unit-complementary spectral peaks is similar to that of a time series composed of two unit-complementary sine waves. A plot of the persistence map resulting from Eq. (4.22) is shown in Figure 4.7.

In the case of a single biennial peak, $\nu_1 = \nu_2 = \frac{1}{2}$ and $a_1 = a_2$, Eq. (4.20) simplifies in a manner similar to Eq. (4.22). However, one must now keep terms up to $\mathcal{O}\left(\Delta^4\right)$, resulting in:

$$\begin{aligned} V(p,l) &\cong e^{-2\pi^2 l^2 \Delta^2} \left\{ 2 \sin\left(\pi[p - \beta]\right) \sin\left(\pi[p - \beta - l]\right) \right. \\ &\quad - \left[\epsilon_2 \left(1 - 2\pi^2 l^2 \Delta^2\right) + \epsilon_4\right] \sin\left(2\pi \left[p - \beta - l/2\right]\right) \\ &\quad \left. + \frac{3}{2}\, \epsilon_2^2 \cos\left(2\pi \left[p - \beta - l/2\right]\right) \right\}, \end{aligned} \tag{4.23}$$

where $\epsilon_4 = \left(\phi''''(\nu_1) + \phi''''(\nu_2)\right) \Delta^4/32$. As in the unit-complementary case, the persistence of a single narrow biennial peak, calculated using Eq. (4.23), bears a strong resemblance to the persistence from a single biennial mode, Eq. (4.7), and reduces to (4.7) when $\Delta = 0$. Unlike the single mode case, however, the persistence from a narrow peak is not a discontinuous step but varies smoothly, as shown in Figure 4.7. Finite width in a biennial spectral peak with slowly varying phase therefore acts to broaden the transition region of the persistence barrier.

Note that the assumptions used to obtain Eqs. (4.22) and (4.23) do not strictly require slowly varying phase. The sole requirement is that $\phi'(\nu_1) - \phi'(\nu_2) \ll 1$. While this is certainly satisfied by slow variations, $\phi'(\nu_i) \ll 1$, it is also satisfied by $\mathcal{O}(1)$ slopes in the phase, provided the slopes are similar at the two peaks. Thus, Eqs. (4.22) and (4.23) are valid when the phases of the two spectral peaks vary coherently.

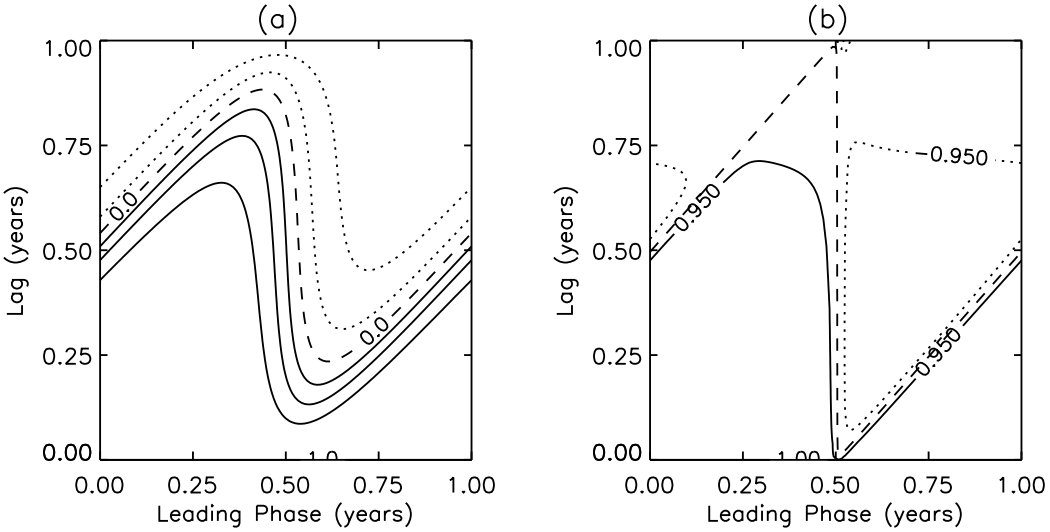

Figure 4.7: Persistence map calculated from Eqs. (4.22) and (4.23), time series whose spectra contain one or two narrow peaks with coherently varying phase angles. In both plots, $\beta = 0.5$, $\Delta = 0.1$, and $\epsilon_2 = \epsilon_4 = 0.05$. (a) Equation (4.22) with $\alpha = 2$, and $\hat{\nu} = -0.2$. (b) Equation (4.23), where the contour intervals are ± 1.0, ± 0.95, and 0.0.

4.3 A Measure for Persistence Barriers

4.3.1 Definition

In this section, I develop a means to quantitatively measure the persistence barrier in a time series. Since a persistence map is a two-dimensional field, such a measure necessarily involves some form of pattern recognition which reduces a two-dimensional field down to one, or a few, numbers. Like most pattern recognition algorithms, this measure compares the field of interest with some preconceived model. Here we shall compare the persistence map $C(p,l)$ of a time series $x(t)$ with a model persistence barrier, $T(p,l)$, that depends on several parameters. A least-squares fit of T to C provides an estimate of the parameters in T, and thus finds those parameters which best describe the barrier in C.

For the model barrier, let us use:

$$T(p,l) = A\, B(p,l) + (1-A)\, e^{-\frac{l}{\tau}}. \tag{4.24}$$

The first term in the model defines the deterministic persistence barrier, while the second term is the persistence map of the Ornstein-Uhlenbeck stochastic process [Eq. (4.17)]. The parameter A, the barrier amplitude, measures the degree to which T is best fit by a deterministic barrier.

To complete the model one must define B. One could use the persistence of the two-mode case with $\mathrm{m}^{\pm}$-complementary frequencies for $B(p,l)$ [Eq. (4.15)]. However, the parameters in Eq. (4.15) are difficult to interpret, and fitting to the equation is computationally difficult. Instead, let us use a model barrier which has a simpler form. The archetype for a barrier is a smooth transition in correlation from -1 to $+1$ at a particular location. A simple function that achieves this is $\tanh(x/\Delta p)$, where Δp measures the width of the transition region. Since persistence is periodic in phase, I construct the function

$$h(x,\, \Delta p) = \sum_{n=-\infty}^{\infty} \tanh\left(\frac{x+n}{\Delta p}\right). \tag{4.25}$$

Equation 4.25 is not itself periodic but contains continuous, periodic jumps. The difference between $h(x, \Delta p)$ at two different phases with constant Δp is, however, periodic. One

therefore defines $B(p, l)$ as:

$$B(p, l) \equiv h\left(p - p_b - l,\, \Delta p\right) - h\left(p - p_b,\, \Delta p\right) + 1. \tag{4.26}$$

The parameters p_b and Δp control the barrier position and width, respectively. We now have a theoretical barrier, T, that is completely described by a vector of four parameters, $\vec{a} = (A, \tau, \Delta p, p_b)$, where A is the amplitude, p_b is the barrier position, Δp is the barrier width, and τ is the stochastic decorrelation time.

Because T models a correlation function, it must satisfy the conditions $-1 \leq T(p, l) \leq 1$ and $T(p, l) = 1$ at $l = 0$. This leads to the following requirements on the parameters: $0 \leq A \leq 1$, $\Delta p \geq 0$, and $\tau \geq 0$. By construction, T is periodic in p_b with unit period. The parameters p_b, Δp, and τ all have units of time, while A is unitless. The mean of B is

$$\bar{B} = \int_0^1 dl \int_0^1 dp\, B(p, l) = 0, \tag{4.27}$$

and thus the mean of T, $\bar{T}$, is given by the mean of the stochastic part, $\bar{T} = (1 - A)\tau(1 - e^{-1/\tau})$.

4.3.2 Properties

To interpret the parameters, it is useful to first investigate the asymptotic behavior of T for large τ and Δp. Again, let us only consider lags restricted to $0 \leq l \leq 1$. For large τ, $T \sim AB + (1 - A) + \mathcal{O}(\tau^{-1})$; large decorrelation times thus result in the stochastic portion providing a constant background to the persistence map. For large Δp, $B \sim (1 - 2l) + \mathcal{O}(\Delta p^{-3})$, and the barrier reduces to a phase-independent, linear decay in lag. Numerically, these asymptotic results prove to be good approximations for $\tau > 10$ years or $\Delta p > 1$ year.

The model barrier, T, contains a rich variety of behaviors, reflected in the physical interpretation of its parameters. In particular, the three parameters A, τ, and Δp occasionally have different implications depending on their joint values. Let us now discuss some of the different behavior regimes of the model and the values of A, τ, and Δp associated with them.

When A is nearly one, i. e. $1 - A \ll 1$, the persistence describes a deterministic

process. The model contains regions of high positive and negative correlation, separated by a transition region whose width is determined by Δp. Thus, any loss of correlation is temporary and the barrier represents only a transition from correlation to anticorrelation instead of a loss of predictability. An example is shown in Figure 4.9(a). In the limit $A \to 1$, the decorrelation time, τ, has no meaning.

If the barrier amplitude is small, $A \ll 1$, the persistence map is that of a random process modified by a small phase-dependence [Figure 4.9(b)]. In this regime, the decorrelation time, τ, determines how rapidly correlation is lost. In the limit $A \to 0$, the barrier position, p_b, and barrier width, Δp, have no meaning.

For large barrier width, $\Delta p \geq 1$, the persistence becomes independent of phase [Figure 4.9(c)]. The values of A and τ together determine the decorrelation rate. In this regime, the deterministic part of the model has a linear decorrelation, which combines with the exponential decay from the stochastic portion.

When the decorrelation time becomes large, $\tau \gg 1$, the stochastic component of the model is approximately constant. The full model thus has regions containing high positive correlation and regions containing correlation with a value of $1 - 2A$. Thus, for appropriate values of A, the model varies from regions of high correlation to regions of correlation near zero. The transition region separating the two has a width determined by Δp. One can also think of Δp as the rate of decorrelation with respect to phase. Thus larger values of Δp indicate weaker barriers. An example of this case is shown in Figure 4.9(d) for $A = 0.6$.

4.3.3 Computation

To measure the persistence barrier of a given timeseries, one must first calculate the persistence map, $C(p, l)$, of the timeseries. By fitting $T(p, l)$ of Equation (4.24) to $C(p, l)$, one can find the set of parameters, $\vec{a} = (A, \tau, \Delta p, p_b)$, which best describes $C(p, l)$. I shall use the techniques introduced in chapter 3 to perform a nonlinear least-squares fit of $T(p, l)$ to $C(p, l)$. I will also make use of the correlation-like measure of goodness of fit in Equation (3.11) and the estimate of confidence intervals defined in Equation (3.16). The actual computations of these quantities, as well as numerical implementation of the model, involve some special issues not discussed in chapter 3. Appendix A contains a description of all of these special numerical issues and techniques used to compute barrier measures

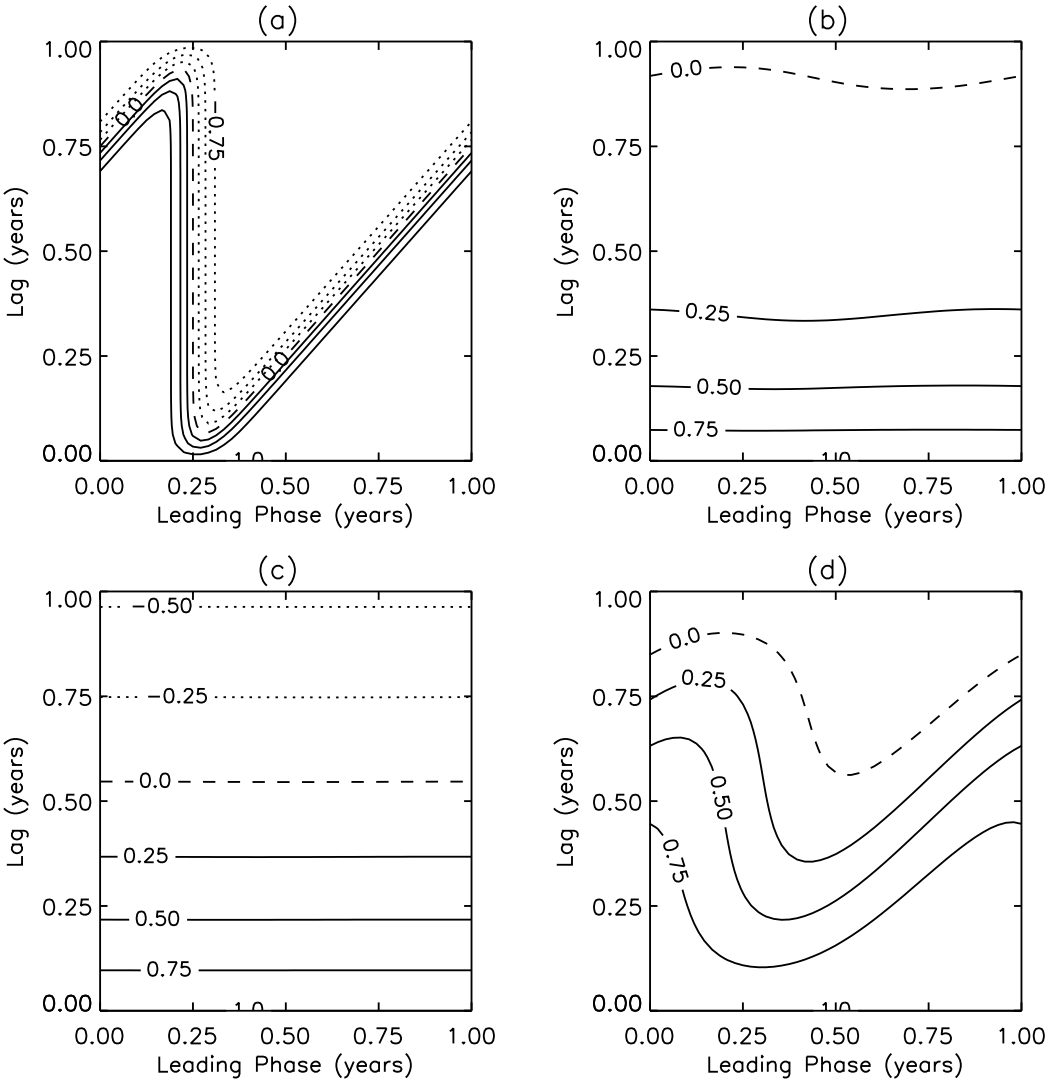

Figure 4.9: $T(p, l)$ at various parameter values. For all plots, $p_b = \frac{1}{4}$ yr. (a) $A = 1$, $\Delta p = 0.06$ yr, $\tau = 0.25$ yr. (b) $A = 0.03$, $\Delta p = \frac{1}{4}$ yr, $\tau = 0.25$ yr. (c) $A = 0.55$, $\Delta p = 1$ yr, $\tau = 0.25$ yr. (d) $A = 0.6$, $\Delta p = \frac{1}{4}$ yr, $\tau = 10$ yr.

throughout the remainder of this dissertation.

4.4 Summary

This chapter has analyzed the cyclostationary lag autocorrelation, or persistence, of a range of time series with the goal of gaining insight into the causes and meaning of the persistence barrier seen in ENSO data. In some time series, the persistence is found to be a function of the phase of the year. This work focuses attention on one particular pattern of phase-dependence, the persistence barrier, the pattern found in ENSO data.

A time series composed of a single sine wave only produces a phase-dependent persistence map if the frequency of the wave is a biennial cycle or one of its harmonics, $\nu = (2n+1)/2$. The phase-dependence takes the form of a single discontinuous barrier if the wave is a biennial cycle, i. e. $n = 0$. Otherwise, there are multiple discontinuous barriers in a single year.

A time series composed of two sine waves produces phase-dependent persistence maps under several circumstances. Considering the single sine wave result, it is not surprising that the persistence map is phase-dependent when one or both frequencies are harmonics of the biennial cycle. Surprisingly, phase-dependence also results when neither frequency is a biennial harmonic, provided the frequencies are $\mathrm{m}^{\pm}$-complementary, i. e. their sum or difference is an integer, m. In particular, unit-complementary frequencies, frequencies whose sum is 1, produce an ENSO-like barrier. This suggests that the barrier-producing harmonics of a time series with a broadband spectrum are either biennial cycles or unit-complementary pairs of harmonics.

Time series with narrow but continuous peaks can produce barriers when the peaks are centered on barrier-producing harmonics. If the phase angles of the harmonics are rapidly varying, however, the contributions of the continuum of modes around the peaks cancel, and the persistence depends only on lag. For either slowly varying or coherent phase angles, a barrier does result and is similar to that produced by discrete modes.

Thus, in the time series considered here, the existence of a barrier depends on barrier-producing harmonics having coherent phase angles. The position of that barrier is determined by those same coherent phase angles. This suggests that an arbitrary time series

will only have a barrier if its dominant harmonics are barrier producing and have coherent phases. Further, the phase angles also determine when the time series nears or crosses zero. The connection between phase angles and "zero-crossings" corroborates the work of Xue et. al. (XCZB94), who find that persistence barriers in ENSO indices occur during times when the index has variance near zero.

Following the analysis of persistence in section 4.2, section 4.3 proposed a quantitative measure of the properties of a persistence barrier. The measure is based on a model persistence barrier combining both deterministic and stochastic features, a model containing four parameters: a barrier amplitude, barrier position, barrier width, and stochastic decorrelation time. I shall make extensive use of this measure in chapters 5 and 6.

Chapter 5

Persistence Barriers in ENSO

5.1 Introduction

One of the key results of the previous chapter was a connection between persistence barriers and specific harmonic behavior. A time series composed of two types of harmonics, either a single biennial mode or two modes whose frequencies sum to unity, has a persistence map with a barrier. If two harmonics have frequencies adding to another integer or differing by any integer, then the resulting persistence map is phase-dependent, but does not look like the persistence barrier. One can generalize this result for discrete harmonics to time series containing one or more continuous narrow peaks. A single narrow spectral peak centered on a biennial mode or two narrow peaks centered on unit-complementary frequencies can generate a barrier. Furthermore, a barrier appears only if the phase angles of the modes vary either slowly or coherently over the narrow peak[s]. On the other hand, if the phase angles vary rapidly or incoherently, the persistence depends only on lag.

While one cannot analytically calculate the persistence of an arbitrary time series, one can identify two types of Fourier modes as barrier-producing: 1) biennial modes with frequency $\nu = \frac{1}{2}$ cycles/year; or 2) pairs of unit-complementary modes with frequencies $\nu_1 + \nu_2 = 1$ cycles/year. One sees that any barrier-producing harmonics must have frequencies between zero and 1.0 cycles/year, since $\nu < 0$ is equivalent to a phase-shift of π. Thus, only interannual frequencies give rise to barriers. I hypothesize that the persistence of a times series will only contain a barrier if two conditions are satisfied: 1) the dominant Fourier modes must be barrier-producing harmonics; and 2) the phase angles of all of the barrier-producing harmonics must be slowly varying or mutually-coherent. The first criteria implies dominant

interannual variability. One can think of the second criteria as follows: the phase angle of a Fourier mode specifies when that mode changes sign. If all of the barrier-producing harmonics have different phase angles, they will cancel each other, while coherent phase angles enhance the barrier-producing harmonics.

The purpose of this chapter is, in part, to numerically examine the aforementioned hypothesis.[1] Section 5.2.1 applies the measure defined in the last chapter to the Southern Oscillation Index [SOI] and the sea surface temperature in the NINO3 region of the central equatorial Pacific [NINO3 SST]. Section 5.2.2 numerically examines the hypothesis about conditions for producing a barrier using the SOI and NINO3 SST. Lastly, in section 5.3 I use the measure of barrier properties to study the interdecadal variability of ENSO.

5.2 ENSO Barriers

5.2.1 The Data

Let us now use the techniques of section 4.3 to measure the properties of the persistence barrier in ENSO. The two sets of data used are the Southern Oscillation Index [SOI], and the sea surface temperature from the NINO3 region [NINO3 SST] of the equatorial Pacific, 5°N-5°S by 150°W-90°W. Both datasets contain monthly mean values. The NINO3 SST data from 1950 to the present is available from the NOAA Climate Prediction Center. The historical NINO3 SST data, from 1896-1949, is based on a spatial average of data from the COADS SST analysis (OPRR87). The SOI is a standardized difference in the sea level pressure between Tahiti, at 17.6°S, 149.6°W, and Darwin, Australia, at 12.4°S, 130.9°E. The pressure at each location has been standardized by removing its cyclostationary mean and normalizing by its overall standard deviation. The SOI is then computed by taking the difference in the two standardized pressures and normalizing that difference by its overall standard deviation. To fill the gaps which exist in the Tahiti sea level pressure record before 1950, Torrence and Webster (TW98) perform a least squares regression using sea level pressure and sea surface temperature from other locations in the equatorial Pacific. The contiguous SOI data for 1896-1950 produced by this analysis is used in this work.

The results of fitting the model persistence barrier to the NINO3 SST and the

[1]The material in this chapter has been submitted to the Journal of the Atmospheric Sciences and was originally co-authored with Dr. Jeffrey B. Weiss (WW97b).

SOI are shown in Tables 5.1 and 5.2, respectively, in the row labeled "data". The reported uncertainty is the 68.3% confidence interval estimated using Equation (3.16) from chapter 3.[2] The NINO3 SST barrier has a strength, A, roughly three times larger than that of the SOI barrier, is about four times wider and has a much longer stochastic decorrelation time. The barriers in both datasets have positions at approximately the same time of year, mid to late March. The 68.3% confidence intervals are at most about 1/10 of the parameter values, indicating that the parameters are relatively well defined. The sole exception is the SOI barrier width whose uncertainty is large because the barrier is so narrow. Thus, both the NINO3 SST and the SOI have statistically significant barriers, since A is bounded away from zero. The shorter decorrelation time in the SOI is understood in terms of the faster typical time scale for atmospheric fluctuations, and the fact that the atmosphere is often thought of as the more random component of the coupled system.

One quantity not reported in Tables 5.1 and 5.2 are R^2, the descriptor of goodness of fit defined in section 3.3. For the NINO3 SST and SOI barriers, $R^2 = 0.966$ and 0.841, respectively. Evidently, it is not unusual for R^2 to have values greater than 0.9 (Kvå85). Thus, the persistence barrier of the SOI does not match the model too well, indicating the presence of other forms of phase-dependence in its persistence map.

Let us now investigate whether one can explain the measured persistence barriers by random fluctuations due to the finite size of the data sampling. This is done by integrating the Ornstein-Uhlenbeck [OU] process of Equation (4.17),

$$\dot{x} = -\frac{1}{\tau}x + \xi(t),$$

where ξ is Gaussian white noise. An infinitely long OU time series produces a persistence map with pure exponential decay; i. e. $A = 0$, Δp and p_b are undefined, and τ identical to that in Equation (4.17). The ENSO data, however, is not infinite, but is approximately 100 years of monthly mean data. To investigate the uncertainty associated with finite-sized time series, I have integrated 10,000 realizations of the OU process for 100 years, using appropriate techniques for the numerical integration of stochastic differential equations (Gar90). For each realization, I then calculated the best fit parameters for the persistence map obtained from

[2] See also appendix A.

Table 5.1: Results of the various Monte Carlos performed on the NINO3 SST. Entries are either the median values ± the 68% confidence interval of the distribution for the Monte Carlo, or the actual measure ± the estimated 68% confidence interval. The labels for each type of Monte-Carlo are explained in the text.

Type	A	p_b	Δp	τ
data	0.312 ± 0.018	0.258 ± 0.005	0.137 ± 0.011	1.072 ± 0.081
OU	0.048 ± 0.037	0.501 ± 0.337	0.136 ± 0.131	1.113 ± 0.219
FPS	0.263 ± 0.012	0.509 ± 0.337	0.345 ± 0.086	0.793 ± 0.028
UCPS	0.268 ± 0.013	0.492 ± 0.354	0.349 ± 0.087	0.803 ± 0.029
NCPS	0.312 ± 0.021	0.279 ± 0.011	0.148 ± 0.017	1.029 ± 0.078

Table 5.2: Results of the various Monte Carlos performed on the SOI. The notation is identical to that used in Table 5.1.

Type	A	p_b	Δp	τ
data	0.095 ± 0.012	0.275 ± 0.012	0.031 ± 0.029	0.398 ± 0.016
OU	0.049 ± 0.036	0.500 ± 0.336	0.072 ± 0.091	0.418 ± 0.060
FPS	0.037 ± 0.018	0.503 ± 0.338	0.007 ± 0.042	0.357 ± 0.010
UCPS	0.038 ± 0.019	0.460 ± 0.396	0.007 ± 0.042	0.361 ± 0.010
NCPS	0.084 ± 0.013	0.313 ± 0.032	0.068 ± 0.046	0.381 ± 0.007

monthly mean data. Two such experiments were performed: one with $\tau = 1$ year to compare with the NINO3 SST, and one with $\tau = 0.4$ years to compare with the SOI. The row labeled "OU" in Tables 5.1 and 5.2 show the median best fit parameters and the confidence intervals that account for 68.3% of the data from the 10,000 realizations.

This experiment shows that a red-noise time series with length similar to the ENSO datasets has a typical barrier strength of 0.05, independent of the other parameters. In both runs, $p_b \cong \frac{1}{2}$ yr. with a confidence interval of about $\frac{1}{3}$ yr., similar to the statistics of a flat, uniform distribution. The width, Δp, has a confidence interval which encloses zero. Regardless, neither p_b nor Δp are well-defined at the small value of A found in these experiments. Lastly, to 68% confidence, τ is at the value used in the given Monte Carlo experiment.

Let us now compare the Monte Carlo experiments performed on the OU process with the ENSO data. For the NINO3 SST barrier, such a comparison shows that the barrier strength is statistically different from that of the OU experiment. The NINO3 SST barrier thus cannot be explained by the finite length of the data record. In contrast, the confidence intervals of the OU experiment and the SOI barrier strength do overlap, indicating that there is some possibility that the SOI barrier is explained by a short red-noise time series.

The goodness of fit for both Monte Carlo experiments is $R^2 \sim 0.99$ with a 68% confidence on R^2 of $\sim \mathcal{O}(10^{-3})$. Therefore, the results of the OU experiment match the stochastic portion of the barrier model very well. Recall, however that $R^2 = 0.841$ for the SOI. Thus, the overlap between the measures of the SOI and the OU experiment is most likely due to the poorer fit between the SOI persistence and the barrier model.

5.2.2 Barrier-Generating Harmonics

In the introduction, I hypothesized that a timeseries generates a persistence barrier under certain circumstances. First, the timeseries must contain dominant interannual variability. Additionally, these dominant interannual modes must form unit-complementary pairs, i. e. $\nu_1 + \nu_2 = 1$ yr^{-1}, or one of the dominant interannual modes must be biennial, i. e. $\nu = \frac{1}{2}$ yr^{-1}. Finally, the phase-angles of the unit-complementary and biennial modes must be slowly varying or mutually-coherent. In this section, we will explore this hypothesis numerically.

Figures 5.1 and 5.2 contain the Fourier transform of the NINO3 SST and the SOI, respectively. The dotted line marks the biennial mode, $\nu = \frac{1}{2}$ yr^{-1}. Both spectra display dominant interannual harmonics, while modes with frequencies greater than 1 yr^{-1} are, on average, an order of magnitude weaker in power. Because of the noisiness of the spectrum, the power in the biennial mode of either dataset is not readily apparent. The power in the NINO3 SST and SOI at $\nu = \frac{1}{2}$ yr^{-1} is 0.01 and 0.001, respectively [using the same units as in Figures 5.1 and 5.2]. It is therefore unclear if the biennial mode has a greater effect on the persistence barrier than the unit-complementary modes.

To examine the effect of unit-complementary harmonics on persistence maps, let us filter them from the ENSO data in the following fashion. After calculating the Fourier transform of the data, one sets the amplitudes of certain harmonics to zero. One then inverts the modified transform, calculates the persistence map of the resulting data, and measures the map. In the first experiment, I filter out unit-complementary pairs by first selecting the mode with the largest power, then locating its complement. Both modes are filtered out, and the process repeated. The second experiment filters out all harmonics between the frequencies $\frac{1}{2} - \mathrm{d}\nu$ yr^{-1} and $\frac{1}{2} + \mathrm{d}\nu$ yr^{-1}. A third experiment filters all harmonics with $\nu \geq 1$ yr^{-1}, leaving only the interdecadal harmonics, all of which form unit-complementary pairs.

The five panels of Figure 5.3 summarize the results of the three filtering experiments on the NINO3 SST. The solid lines indicate the experiment where unit-complementary pairs are removed in order of decreasing power. The dashed lines display the result of filtering out modes as a function of distance from $\nu = \frac{1}{2}$ yr^{-1}. The dotted line in Figure 5.3 (a) denotes the percentage of the total spectral power contained in the harmonics with $\nu \geq 1$ yr^{-1}, i. e. the non-unit-complementary harmonics. Thus, the unit-complementary modes of the NINO3 SST contain 85% of the power.

Figures 5.3 (b)-(e) show the change in the barrier measures of the NINO3 SST as a function of power lost via the two filtering experiments. The bars labeled "error" indicate the 68% confidence interval averaged over all plotted data points. Values of A, p_b and Δp with confidence intervals below 0 and larger than 1 are not plotted. The barrier strength, A, decreases as the power in the unit-complementary modes decreases. The other measures also change as power is lost, with the decorrelation time, τ, decreasing, and the barrier width and

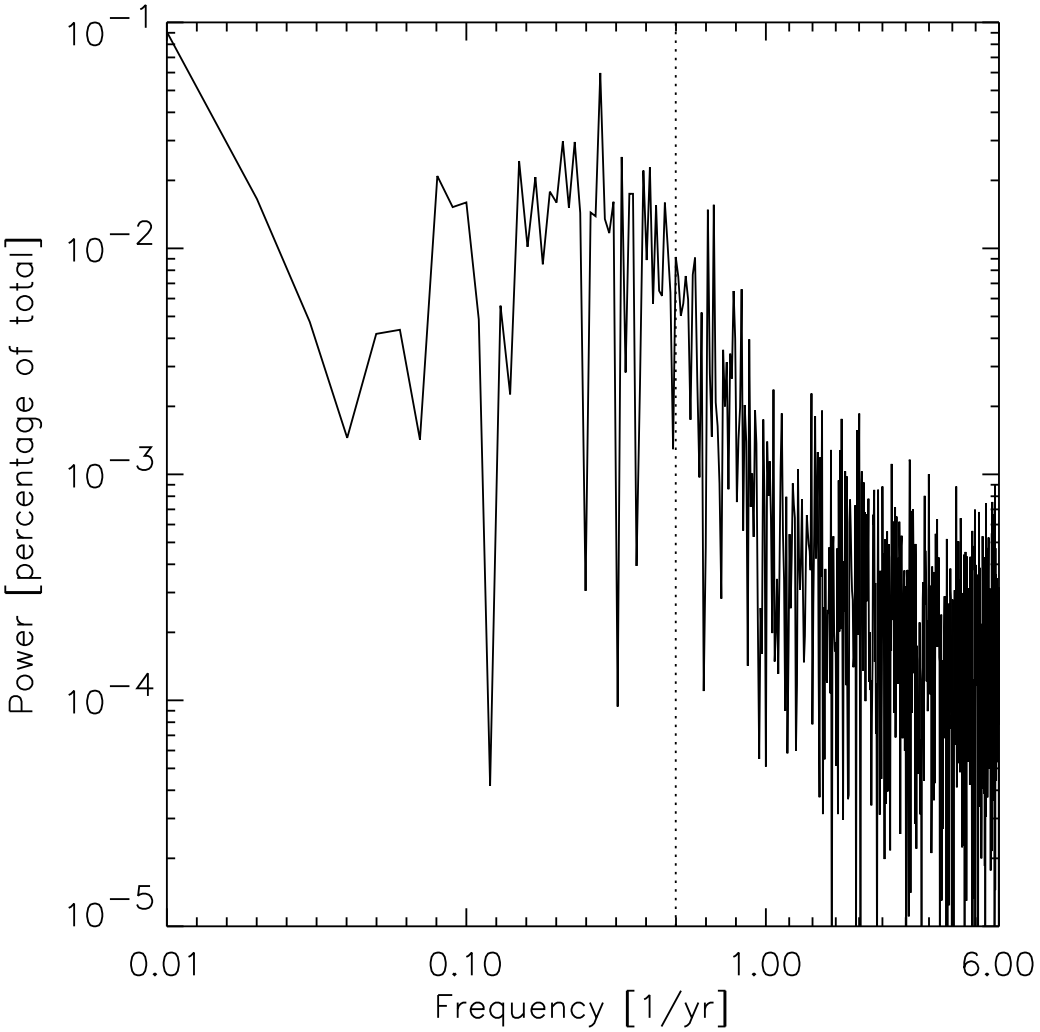

Figure 5.1: Fourier transform of the NINO3 SST. The dotted line indicates the biennial harmonic.

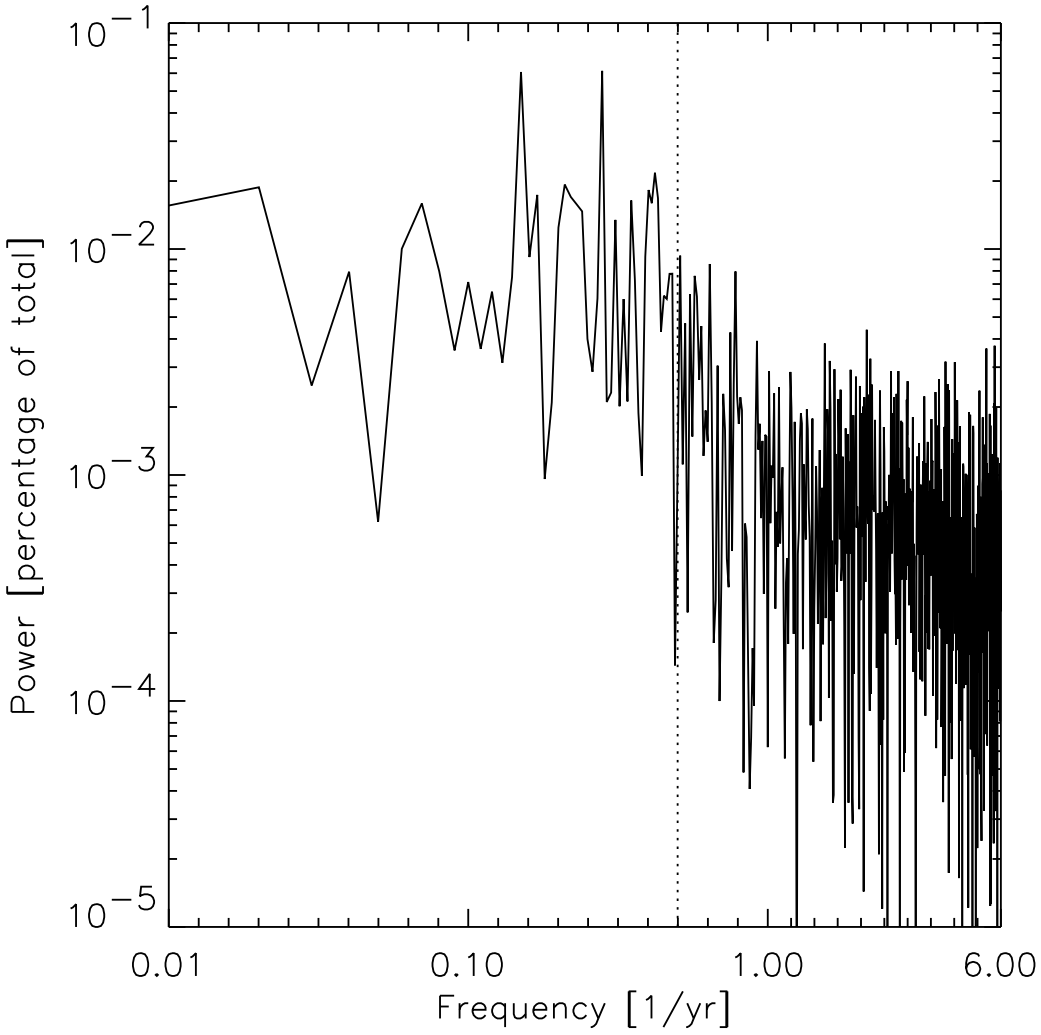

Figure 5.2: Fourier transform of the SOI, plotted as in Figure 5.1.

position, Δp and p_b, changing more and more drastically as A nears zero [I have truncated the plots of Δp and p_b at $A = 0$, where the latter two measures are ill-defined].

Most striking is the difference in behavior of the two different methods of filtering out harmonics, especially on A and τ. The barrier strength decreases much more rapidly when the harmonics are removed as a function of distance from $\nu = \frac{1}{2}$ yr^{-1}, while the converse is true for τ. Compare these results to those for the case in which all harmonics with $\nu \geq 1$ yr^{-1} are filtered out, indicated in the plots by the diamond. The barrier strength increases to $\frac{1}{2}$, accompanied by an increase to large τ, a slight increase in Δp and very little change in the barrier position. Because $\tau \gg 1$ year, the diamond is absent from Figure 5.3 (c).

The drastic increase in τ, seen when the intra-annual modes are removed from the NINO3 SST, reveals a quirk of the fitting method described in chapter 3 and appendix A. As noted in chapter 4, when the decorrelation time is large, i. e. $\tau \gg 1$ year, $T \sim AB + (1 - A) + \mathcal{O}(\tau^{-1})$. Large decorrelation times turn the stochastic portion of the model into a constant background, while A now describes the minimum value the persistence obtains. When the strength becomes large, fitting to a constant background best minimizes the error between the model persistence map [Equation (4.24)] and the data. In practice, we have noticed that the fit algorithm usually seeks large decorrelation times, with τ often greater than 10^6 years, when $A > 0.5$.

The selective filtering experiments performed on the NINO3 SST indicate two results. First, the intra-annual modes act to weaken the persistence barrier. This should not be surprising, since the harmonics with $\nu \geq 1$ yr^{-1} cannot form unit-complementary pairs. Second, the unit-complementary harmonics with frequencies near the biennial mode have the dominant effect on the barrier strength. They are not, however, solely responsible for generating the barrier.

Figure 5.4 contains the results of the three selective filtering experiments on the SOI. The graphs are plotted in the same fashion as Figure 5.3. The unit-complementary modes contain only 60% of the SOI's spectra power, as seen in Figure 5.4 (a). The behavior of the measures of the SOI barrier under the filtering experiments is similar to that of the NINO3 SST [Figure 5.4 (b)-(e)]. Removing modes via distance from $\nu = \frac{1}{2}$ yr^{-1} has a remarkably strong effect on A, and therefore on Δp, and p_b, as well. The decorrelation time

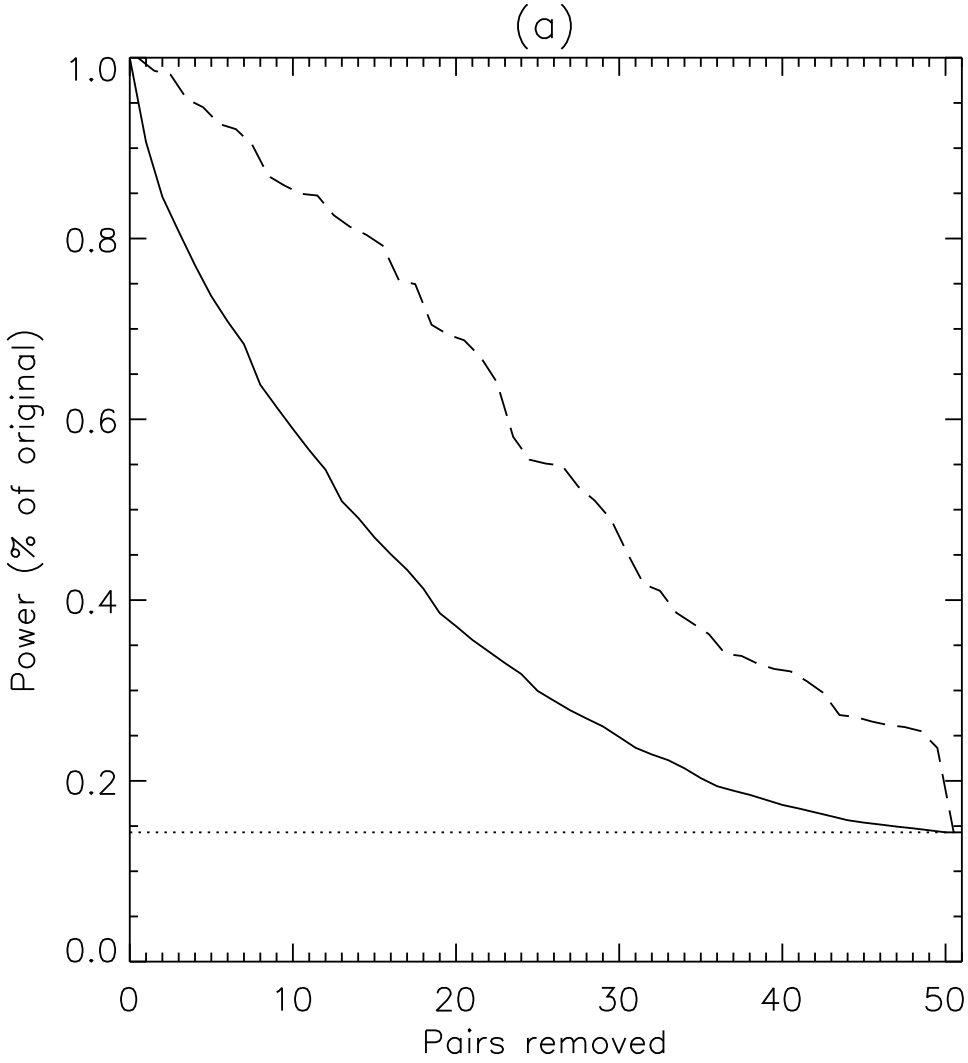

Figure 5.3: Plot of barrier measures for the NINO3 SST vs. power remaining after filtering particular unit-complementary modes. The solid lines describe filtering in order of decreasing power. The dashed line indicates filtering via distance from $\nu = \frac{1}{2}$ yr^{-1}. The dotted line and diamonds display filtering out all non-unit-complementary harmonics. The bars labeled "error" indicate a typical value of the 68% confidence interval. (a) The remaining power vs. the filtered harmonics.

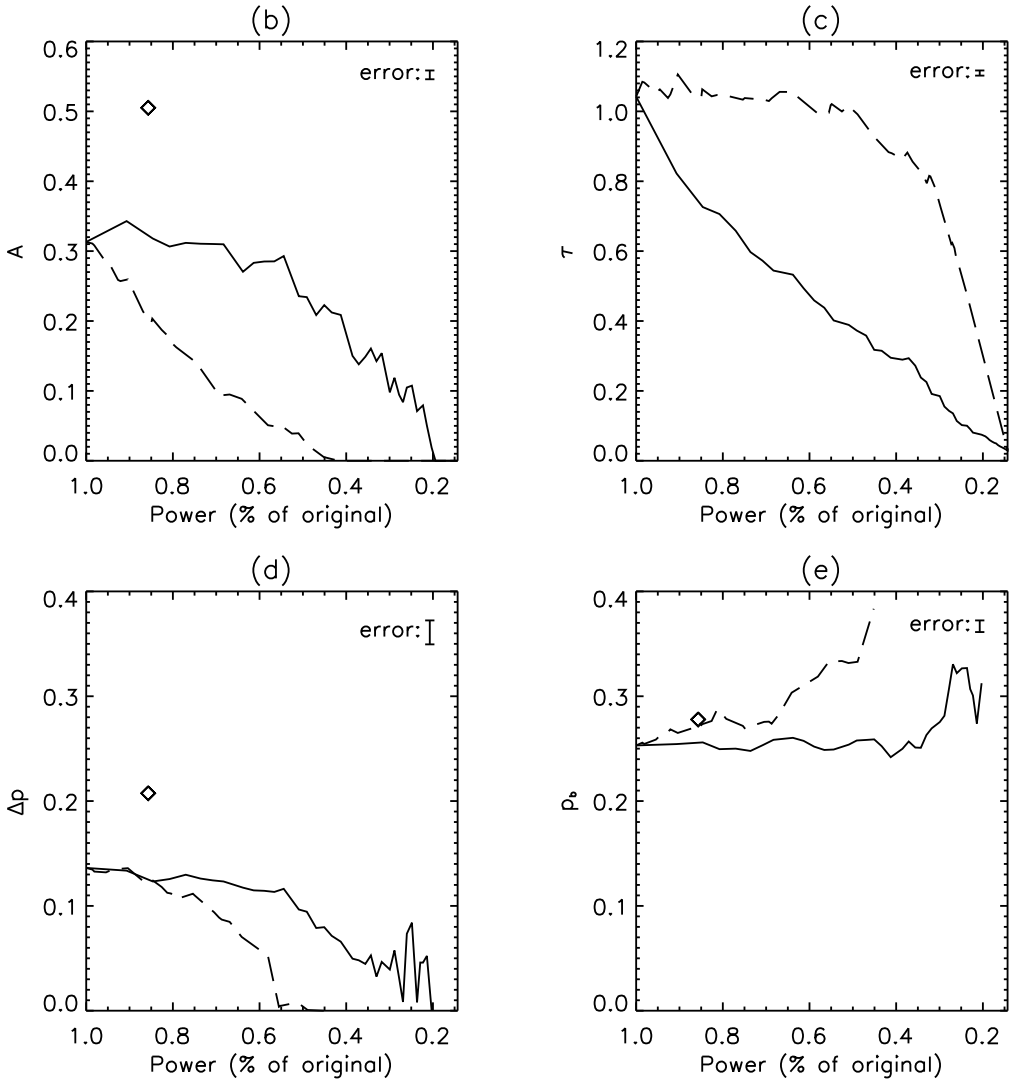

Figure 5.3: (b) The strength, A. (c) The decorrelation time, τ. (d) The barrier width, Δp. (e) The barrier position, p_b.

remains fairly constant until most of the power in the unit-complementary modes is gone. Filtering in order of decreasing power has a less pronounced effect; A initially increase, then decreases to a roughly constant value. It then increases again before finally dropping to zero. τ decreases monotonically, while Δp, and p_b, as always, follow the behavior of the barrier strength. Filtering the harmonics with $\nu > 1$ yr^{-1} strengthens the barrier to $A = \frac{1}{2}$. It also increases τ to large values, widens the barrier greatly, with $\Delta p = 0.3$ years, and has little effect on the barrier position.

There are certain similarities between the results of the selective filtering experiments for the SOI and those for the NINO3 SST. Removing the non-unit-complementary harmonics acts to strengthen and widen the barrier. It also increases the decorrelation time, τ, to a value larger than 10 years. As noted in chapter 4, the stochastic portion of the model [Equation (4.24)] is approximately constant at such large values of τ. Thus, A and Δp determine barrier strength when τ is large. A second similarity is the more rapid decrease in barrier strength when the near-biennial modes are removed. Unlike the NINO3 SST, filtering the near-biennial modes of the SOI rapidly reduces A to zero. Thus, the unit-complementary harmonics of the SOI which are near $\nu = \frac{1}{2}$ yr^{-1} generate the barrier.

In this section, we have numerically examined the connection between persistence barriers and barrier-generating modes. Filtering out all unit-complementary modes eliminates the persistence barrier, while filtering all non-unit-complementary modes strengthens the barrier. In both timeseries, power decreases as a function of frequency. Thus, unit-complementary pairs near the biennial mode contain power of similar magnitude. For those further away from the biennial mode, one of the harmonics in the pair has greater power. Chapter 4 revealed a connection between the amplitudes of each mode in a unit-complementary pair and the strength of the barrier. Similar amplitudes correspond to stronger barriers. Thus, the unit-complementary pairs in the ENSO data near the biennial mode play a greater role in producing the barrier.

5.2.3 Dependence on Fourier Phase Angles

In this section, we will examine the effect of the Fourier phase angles of the unit-complementary modes on the barrier. While the amplitude of a Fourier mode describes the power in a give mode, the corresponding phase angle describes when a mode changes sign.

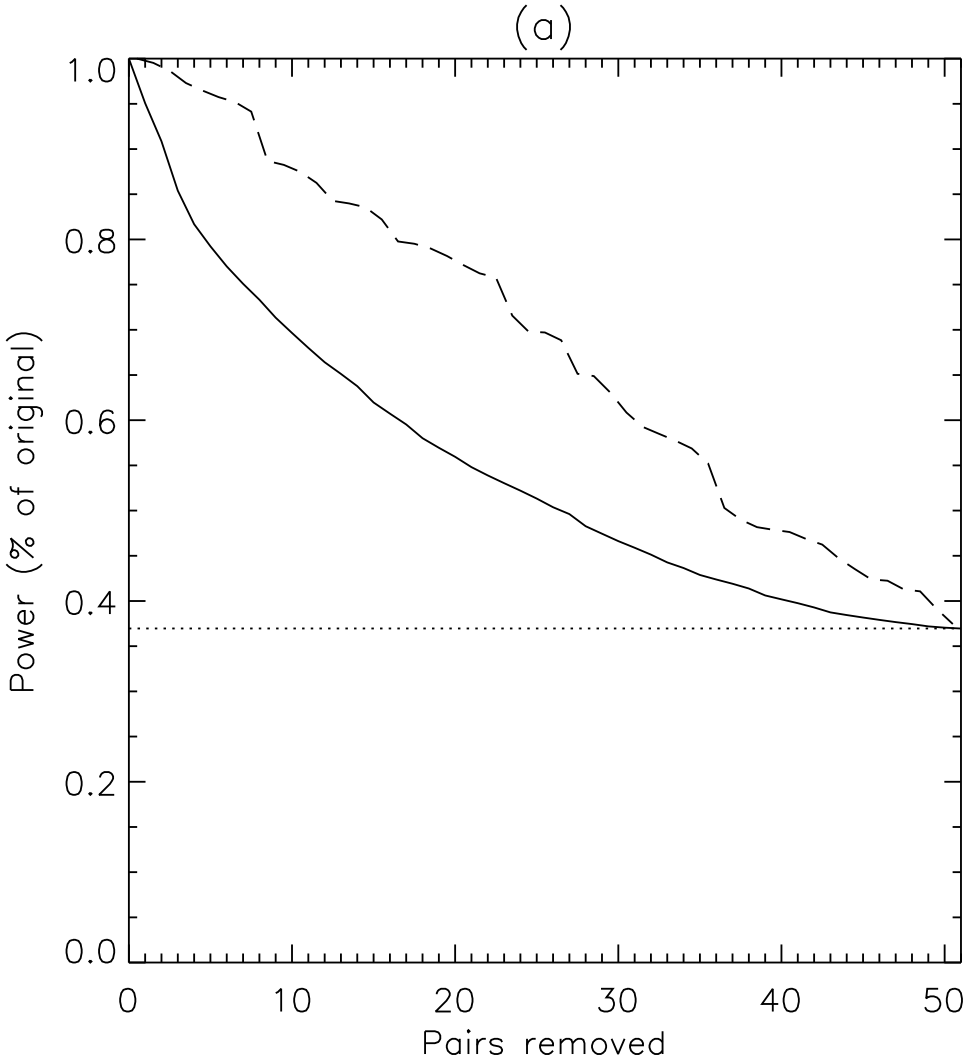

Figure 5.4: Plot of barrier measures for the SOI vs. power remaining after filtering certain unit-complementary harmonics. The data is graphed in the same manner as in Figure 5.3. (a) The remaining power vs. the filtered harmonics.

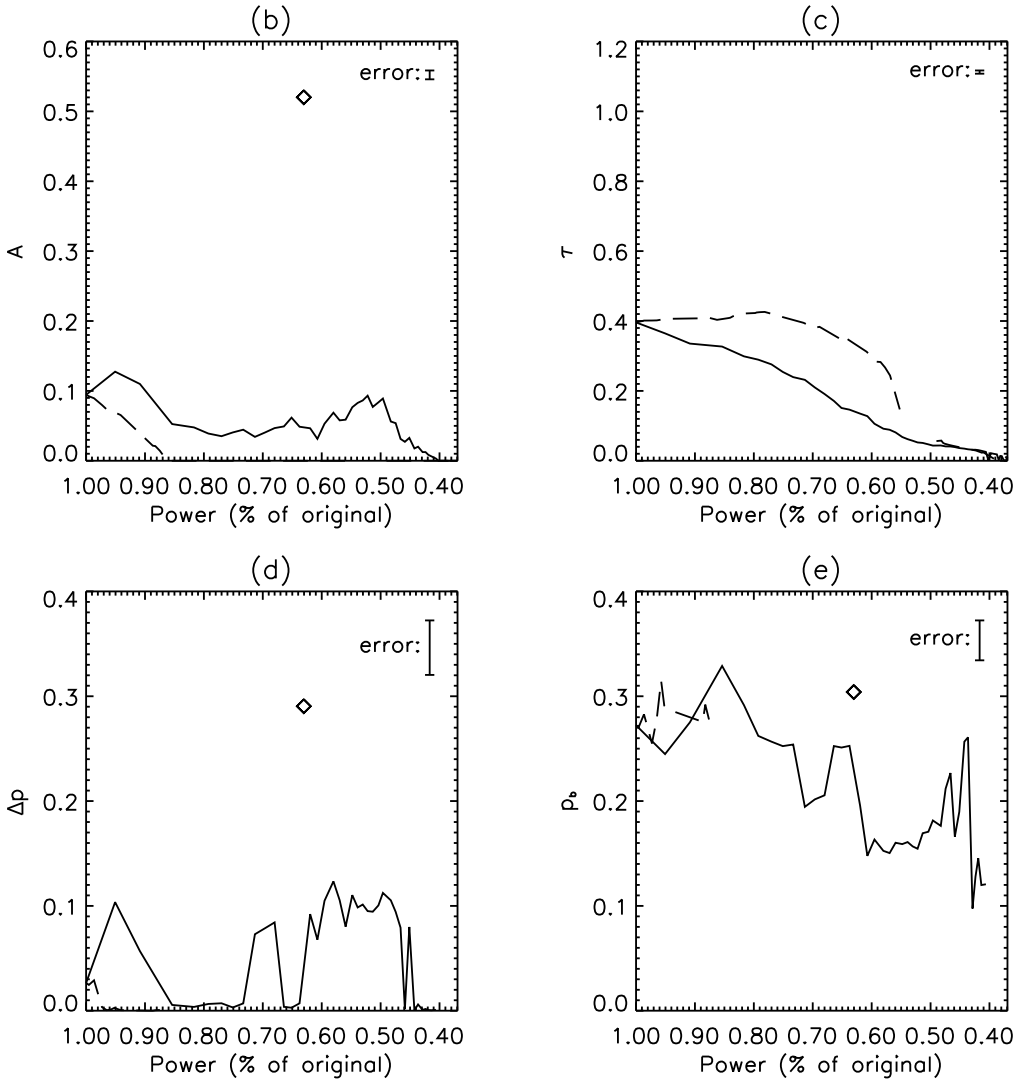

Figure 5.4: (b) The strength, A. (c) The decorrelation time, τ. (d) The barrier width, Δp. (e) The barrier position, p_b.

If the dominant modes have slowly-varying or coherent phases, their sum will approach zero in a periodic fashion at a point described by the phase information. The work of Xue et. al. (XCZB94) showed that the persistence barrier in ENSO data is related to low cyclostationary variance. Their work thus supports a connection between coherent or slowly-varying phase and the persistence barrier.

Let us begin our exploration of the relationship between phase angles and persistence barriers with a phase-scrambling Monte Carlo experiment. After performing a Fourier transform on the data, one replaces the phase angles with a realization of uniformly distributed independent random numbers. One must take care to maintain certain symmetries in the phase or the resulting inverse Fourier transform will not produce a timeseries of real numbers. One then calculates the persistence map of the modified timeseries and the measures for this map. One performs this operation for 1000 realization and calculates distributions of the measures. The results of this Fourier phase-scrambling experiment are in the row labelled "FPS" in Tables 5.1 and 5.2. Because the distributions of the measures are not symmetric, I report the median of the distribution and the 68% confidence interval about the median.

Phase-scrambling the NINO3 SST alters all four of the measures [Table 5.1]. The position, p_b, is comparable to that found in the Ornstein-Uhlenbeck Monte-Carlo experiment, indicating a uniform distribution of p_b. I will therefore not discuss it either here or for the phase-scrambling experiment performed on the SOI. Similarly, the goodness of fit, R^2 [not reported in Tables 5.1 and 5.2], differs little from the values reported in section 5.2.1 for the unaltered data sets. I shall therefore ignore it as well.

For the other measures, the confidence intervals overlap neither those of the OU experiment nor the original data. There is a decrease in the barrier strength, A, and a doubling of the width, Δp, compared to the unaltered data. Additionally, the decorrelation time, τ decreases. Both the increase in width and drop in strength suggest a weakening of the barrier.

Phase-scrambling the SOI renders the barrier statistically indistinguishable from a stochastic process. As seen in Table 5.2, the median values of A and τ are smaller than those of the OU experiments, while Δp is statistically close to zero for both experiments. The confidence intervals of A and τ time are narrow for FPS, but do overlap those of the

OU experiment. Additionally, the confidence intervals of the strength and decorrelation time do not overlap those of the unaltered data. The measures of the original data are therefore statistically different from the results of phase scrambling.

The phase-scrambling Monte-Carlo experiments on the NINO3 SST and SOI reveal the importance of phase-coherency to the generation of persistence barriers. In a timeseries with a moderately strong barrier, e. g. the NINO3 SST, a lack of phase-coherency weakens and broadens the barrier. When a timeseries with a weak barrier lacks phase-coherency, the persistence is statistically identical to that of red-noise, as phase-scrambling the SOI demonstrates.

As an additional test, let us perform two modified phase-scrambling Monte-Carlo experiments. Instead of replacing all phase angles with random numbers, one replaces only a specific set of phase angles. The row labeled "UCPS" in Tables 5.1 and 5.2 are the results of randomizing the phases of only the unit-complementary modes. The complementary experiment, where one phase-scrambles all modes which are not unit-complementary, is labeled "NCPS" in the tables.

The results of these two Monte-Carlo experiments confirm the role of unit-complementary harmonics in generating a persistence barrier. The measures found in the UCPS experiment for both datasets are nearly identical to those of the FPS. Similarly, the measures of the NCPS experiment and those of the original data also overlap. The sole exception is the barrier position. In both datasets, p_b for the NCPS is later in the year than for the original barrier position. Additionally, the confidence intervals do not overlap. Chapter 4 examined the persistence of the sum of two sine waves whose frequencies were unit-complementary. The Fourier phases of the two waves determined the barrier position. Thus, the shift in position caused by the NCPS experiment merely indicates that the phases of the non-unit-complementary harmonics in the ENSO data play some role in determining the barrier position.

5.3 Persistence in Different Decades

An ideal use for the four measures of persistence barriers is investigating inter-decadal variability in the NINO3 SST and SOI. Using a window of some fixed length, one

calculates the persistence map for a windowed segment of the dataset. One then changes the starting point of the window in yearly increments. The four measures, A, τ, Δp, and p_b, then become a function of time and describe the interdecadal variability of the persistence barrier. The window should be large enough to provide sufficient data for calculating the persistence. Too small of a window increases errors in the persistence maps due to averaging a small data sample. Because of these errors, the model barrier [Equation (4.24)] will not fit well to the data, and moving the window can change the best-fit measures drastically. On the other hand, one cannot describe changes on timescales smaller than the window size. I have found that window of 20 years works best for the ENSO data. Anything smaller introduces the aforementioned "small-sample noise" into the measures.

Figure 5.6 contains plots of each measure of the NINO3 SST applied to the persistence of a running 20-year window of the data. The bars labeled "error" indicate the 68% confidence interval averaged over all plotted data points. The time axis of each plot, varying from 1906 to 1987, describes the center of the window. The dashed line indicates the value of each measure for the full timeseries. Due to the nonlinearity of the persistence itself, the measures from the full timeseries are not the same as the time average of the windowed measures.

The interdecadal variation in the NINO3 SST measures contains four regimes of behavior [Figure 5.6]. Dotted lines mark the edges of the four intervals. Because the measures are calculated from a running window, I have fixed these boundaries at convenient dates. They can readily be moved a few years in either direction. The conclusions will remain the same.

The first interval of interest in the NINO3 SST, from 1906-1915, contains a transition period. The barrier strength, A, decreases from 0.4 to 0.2 during this time period. The stochastic decorrelation time, τ, and width, Δp, experience similar decreases. τ drops from near 2 years to 0.4 years, while Δp ranges from 0.12 years to near zero. In contrast, the barrier position, p_b, increases from 0.3 to 0.38 years, or from mid-April to late-May.

The next regime of interdecadal behavior is from 1916-1945. The strength and decorrelation time both remain low, with typical values of $A \cong 0.25$ and $\tau \cong 0.4$ years. For most of the interval the width is around 0.1 years, though it does later drop to near zero. The position remains in the late boreal spring, with p_b gradually decreasing from 0.38 to 0.3

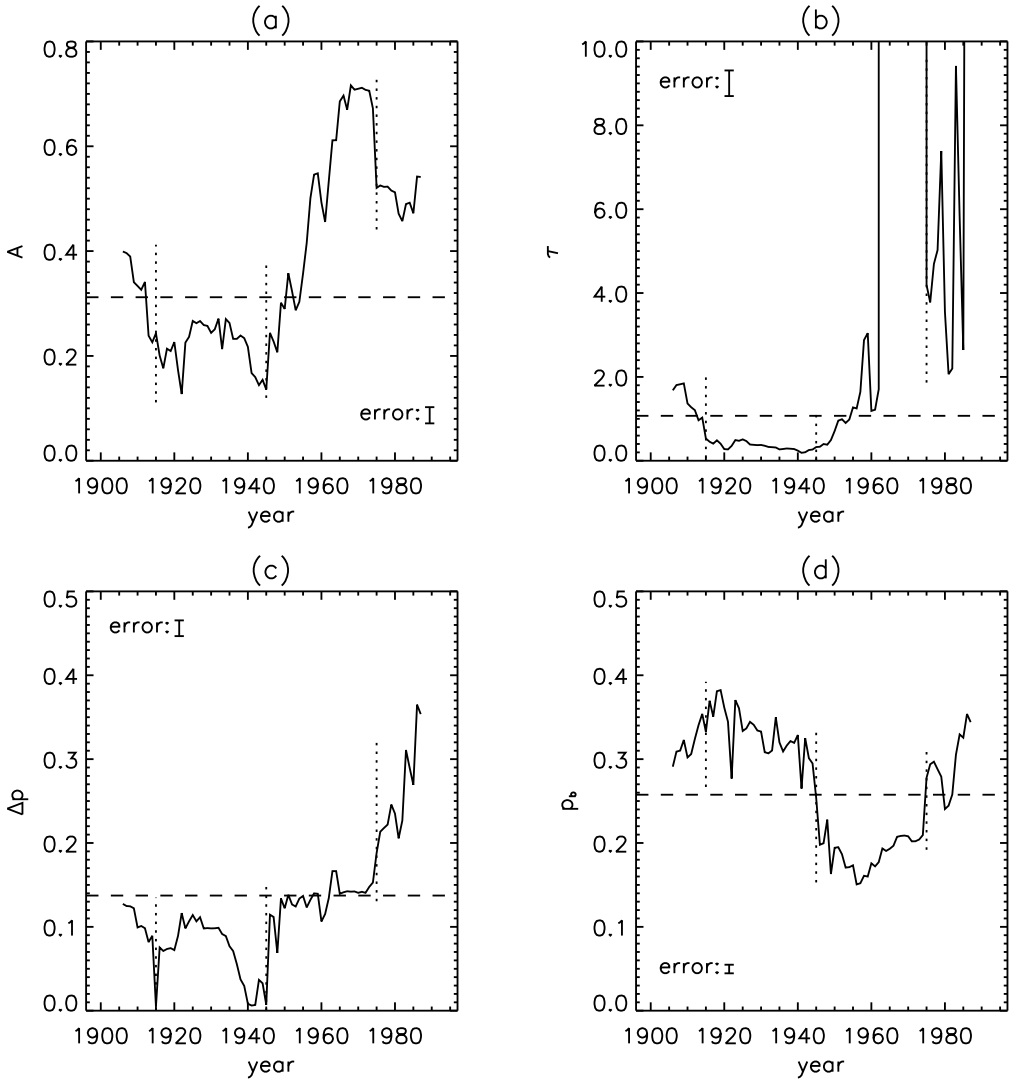

Figure 5.6: Interdecadal variation of NINO3 SST as exhibited by A, τ, Δp, and p_b, using a 20-year running window. The bars labeled "error" indicate a typical value of the 68% confidence interval. The dashed line is the value of the measure for the full timeseries. The dotted lines mark the boundaries between regions of qualitatively different behavior. (a) The strength, A. (b) The decorrelation time, τ. (c) The barrier width, Δp. (d) The barrier position, p_b.

years.

One can see the difference between the 1906-1915 and 1916-1945 intervals in Figures 5.8 and 5.10. Figures 5.8 (a) and 5.10 (a) contain the persistence maps of the windowed data centered at 1906 and 1920, respectively. Figures 5.8 (b) and 5.10 (b) show the model [Equation (4.24)] at the best-fit measures for the 1906 and 1920 maps, respectively. Also listed is the goodness of fit, R^2, for each best-fit model. In addition to having a poorer goodness of fit, the 1920 map looks noisier than the 1916 map. The weakening of barrier strength thus appears connected to the drop in stochastic decorrelation time as the noise in the barrier increases.

Between 1946 and 1975, there is a transition to a strong barrier, as exhibited by the large increase in A [Figure 5.6]. τ also increases, becoming large after 1960, when A reaches its largest value. Recall from section 5.2.2 that the stochastic term in the model barrier becomes constant at large τ, which is, in turn, connected to values of $A \geq 0.5$. The width remains near $\Delta p \cong 0.13$ years, the value for the full NINO3 SST dataset. p_b, however, drops below 0.2 years, i. e. the barrier is now located in March.

Figure 5.12 contains an example of the 1946-1975 period, displayed in the same style as Figure 5.8. The data used for these plots is the NINO3 SST window centered at 1970. The persistence map for this data [Figure 5.12 (a)] looks very much like the model barrier [Figure 5.12 (b)] and has a large R^2 to match. This supports the interpretation of a strong persistence barrier during the 1946-1975 interval.

During the final regime of interdecadal variation in the NINO3 SST measures, the barrier weakens. There is a drop in strength during 1976-1987 to $A \cong 0.5$. The decorrelation time varies widely, but is always larger than $\tau = 2$ years. Although the large A and τ indicate a strong barrier, the width increases throughout the interval, from $\Delta p \cong 0.13$ years to nearly triple that value by 1987. Lastly, the barrier position shifts back to May, with typically $p_b > 0.3$ years.

Figures 5.14 and 5.16 exemplify the weakening persistence barrier in the NINO3 SST between 1976 and 1987. The persistence map from the window centered at 1980 [Figure 5.14] shows a decrease in goodness of fit compared to the 1970 window [Figure 5.12]. The barrier for the 1987 window is wider still [Figure 5.16], and also exhibits a slight decrease in R^2 relative to the 1980 window. The barrier is therefore currently very weak.

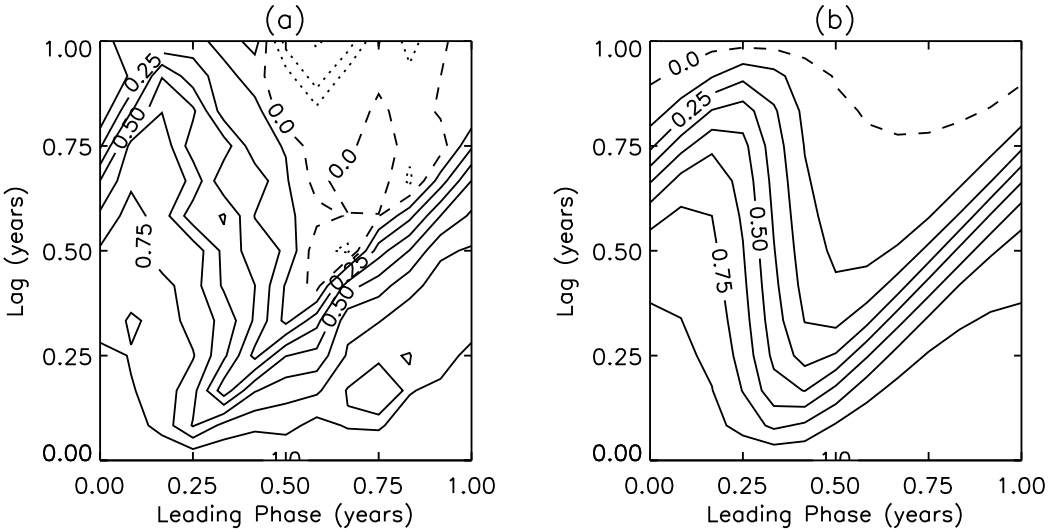

Figure 5.8: Persistence map of the NINO3 SST for a 20-year window centered at 1906. (a) Persistence map of the data. (b) The best-fit model barrier. The goodness of fit is $R^2 = 0.927$, compared to $R^2 = 0.966$ for the full timeseries.

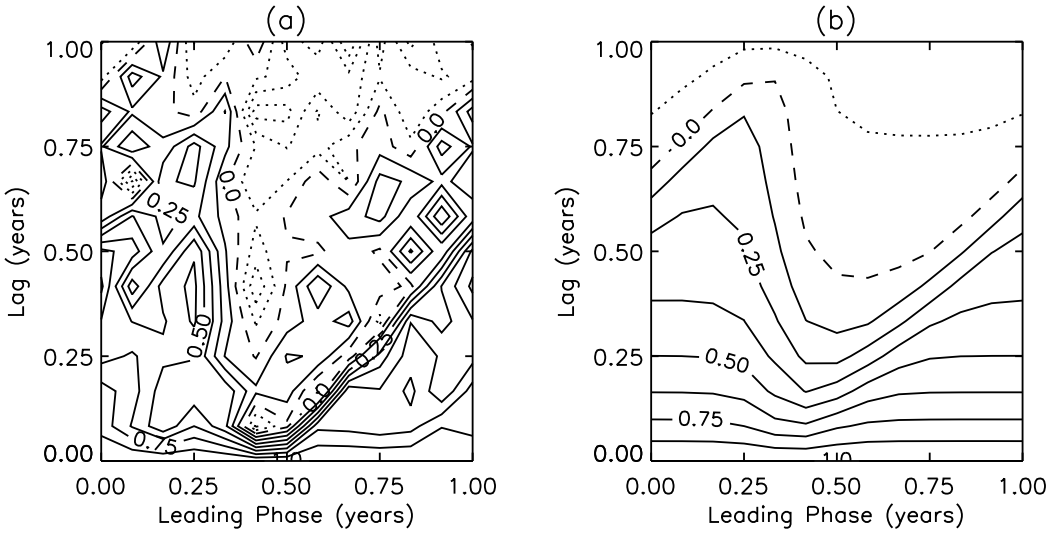

Figure 5.10: Persistence map of the NINO3 SST for a 20-year window centered at 1920. (a) Persistence map of the data. (b) The best-fit model barrier. The goodness of fit is $R^2 = 0.638$.

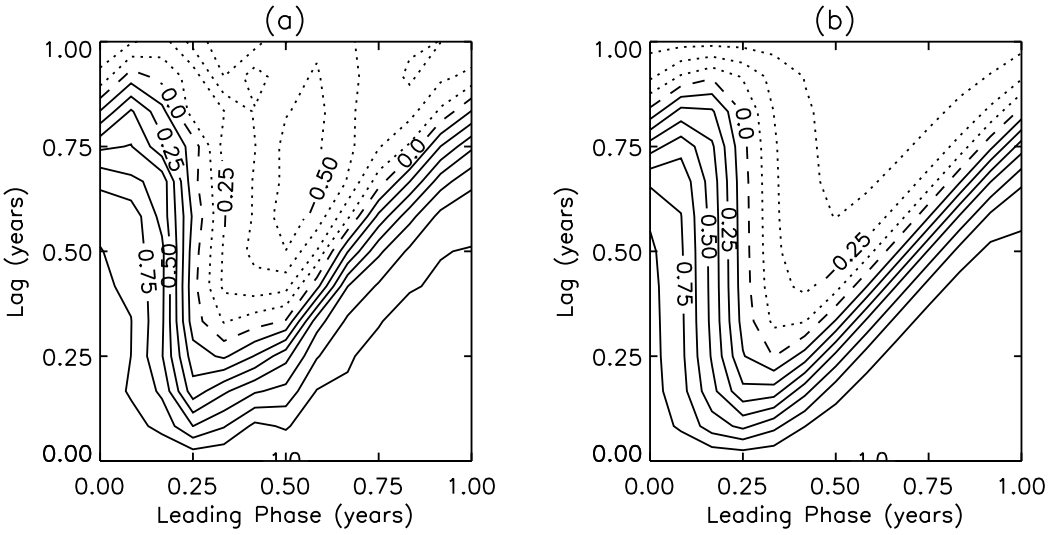

Figure 5.12: Persistence map of the NINO3 SST for a 20-year window centered at 1970. (a) Persistence map of the data. (b) The best-fit model barrier. The goodness of fit is $R^2 = 0.975$.

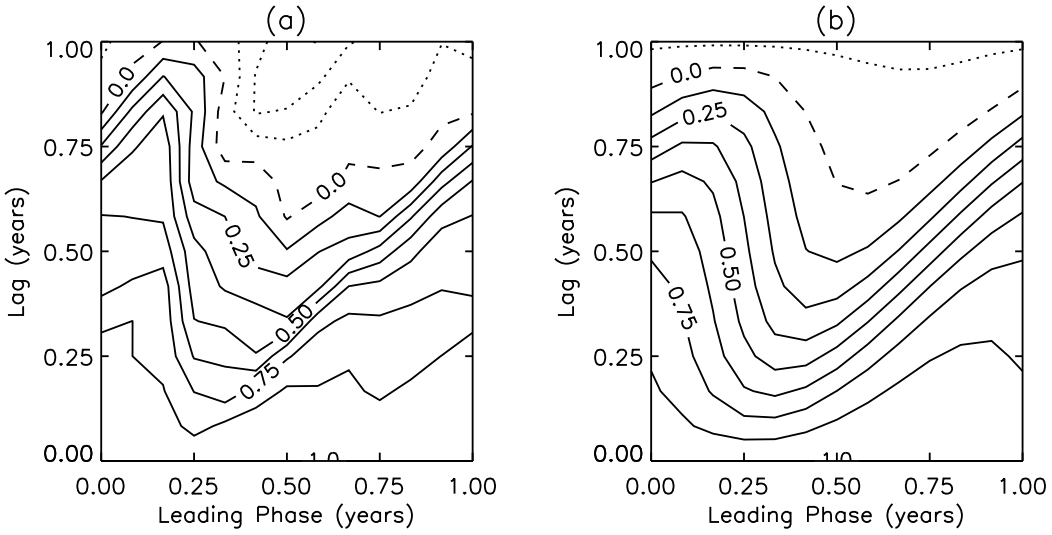

Figure 5.14: Persistence map of the NINO3 SST for a 20-year window centered at 1980. (a) Persistence map of the data. (b) The best-fit model barrier. The goodness of fit is $R^2 = 0.944$.

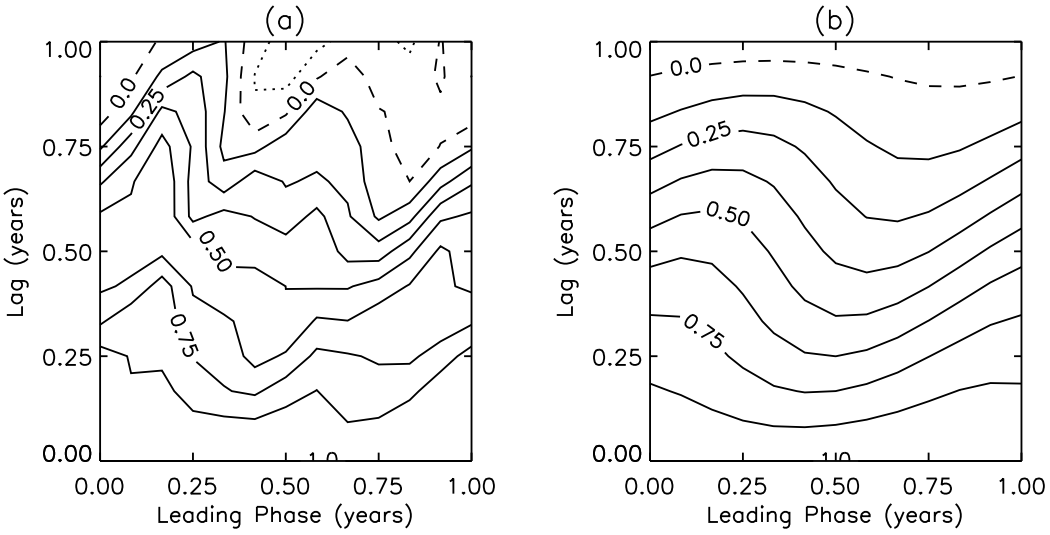

Figure 5.16: Persistence map of the NINO3 SST for a 20-year window centered at 1987. (a) Persistence map of the data. (b) The best-fit model barrier. The goodness of fit is $R^2 = 0.930$.

In summary, the interdecadal variations in the measures of the NINO3 SST persistence indicate four intervals of behavior in the barrier. From 1906 to 1915, the barrier is in decline, entering a period of weakness during 1916-1945. In the 1946-1975, the persistence barrier grows strong. During the present interval, 1976-1987, the barrier grows wide and weak.

Unlike the NINO3 SST, the windowed measures of the SOI contain no consistent intervals of behavior. Figure 5.18 shows the measures of the SOI for a 20-year window, plotted in the same fashion as Figure 5.6. The strength does increase after 1965, peaking at $A \sim 0.3$ near 1975, then decreasing down to 0.2 by 1987. Prior to 1965, however, A is simply noisy. The decorrelation time does have some consistent regimes of behavior, as seen in Figure 5.18 (b). However, there is no overlap of these intervals with intervals of behavior in the other measures. For example, τ is at its lowest during 1925-1945. However, the width, Δp, is near zero except between 1920 and 1935 and from 1955 to 1987.

Though we cannot draw many conclusions about barrier strength from Figure 5.18, we do see interesting behavior in the barrier position. Prior to 1935, the SOI barrier has a position sometime in May ($p_b \cong 0.35$ years). It then shifts to $p_b \sim 0.1$ years, or mid-February. After 1960, the barrier has a position in April, with $p_b \cong 0.25$ years.

Part of the difficulty of interpreting the interdecadal variability of the SOI measures comes from the strong presence of noise in the persistence maps. One sees this in the persistence maps for the SOI windows centered at 1931 and 1933, Figure 5.20 (a) and 5.22 (a). Both of these maps are similar to each other, as should be the case. The two windows differ only by four years of data. The best-fit models, however, differ drastically [Figure 5.20 (b) and 5.22 (b)]. The barrier strength and width change most, with $A = 0.27$ and $\Delta p = 0.23$ years for the 1931 window and $A = 0.12$ and $\Delta p = 0$ years for the 1933 data. For both data windows, goodness of fit is poor, with R^2 well below its value for the full SOI. Thus, moving the data window in the SOI can cause drastic changes in the best-fit measures, making interpretation of the measures difficult.

Not all of the SOI data windows produce noisy barriers that fit poorly to the model. The window centered at 1975, shown in Figure 5.24, contains a strong barrier. Though still somewhat noisy, the persistence map visually resembles the best-fit barrier and the goodness of fit is near its value for the full tiemseries. Note, too, that A reaches a maximum at 1975,

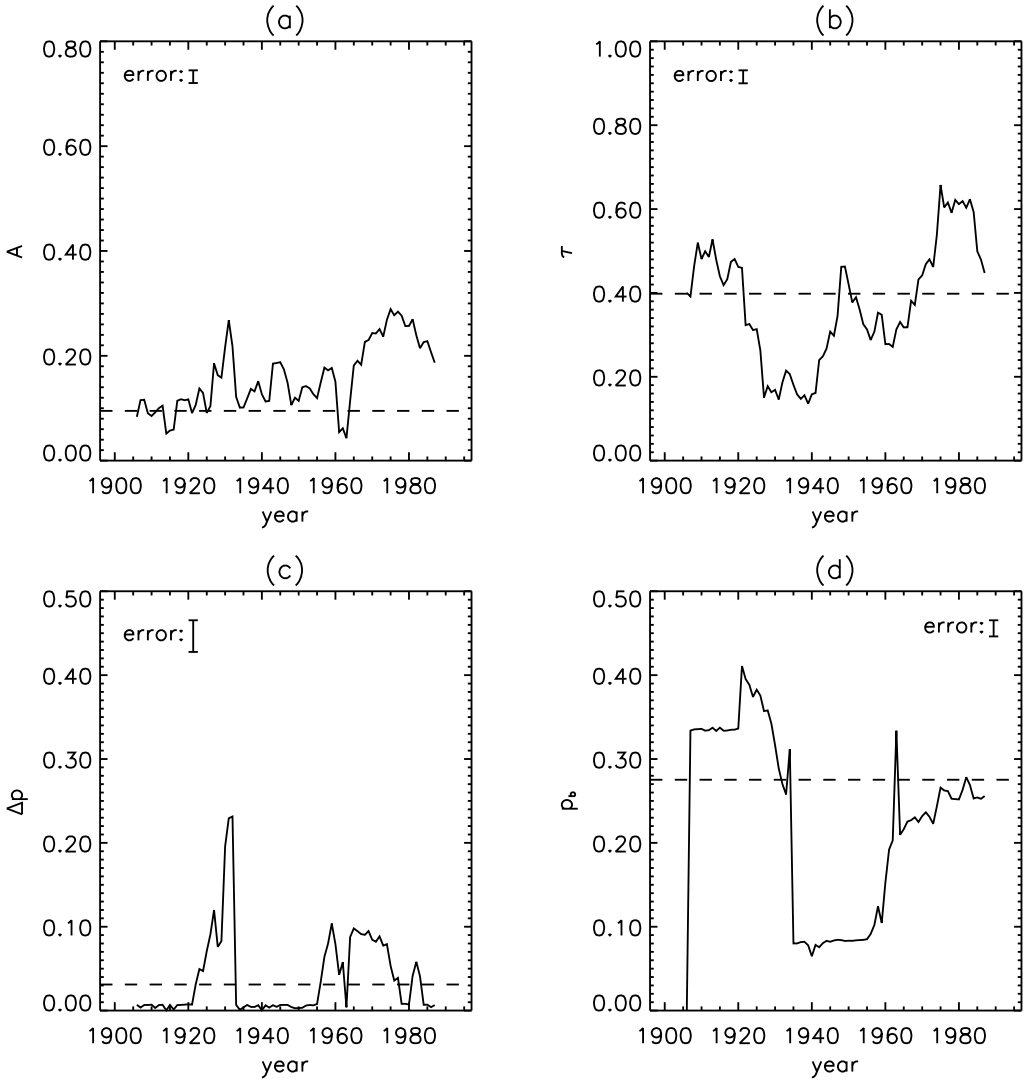

Figure 5.18: Interdecadal variation of the SOI as exhibited by A, τ, Δp, and p_b, using a 20-year running window. The data is graphed in the same manner as in Figure 5.6. (a) The strength, A. (b) The decorrelation time, τ. (c) The barrier width, Δp. (d) The barrier position, p_b.

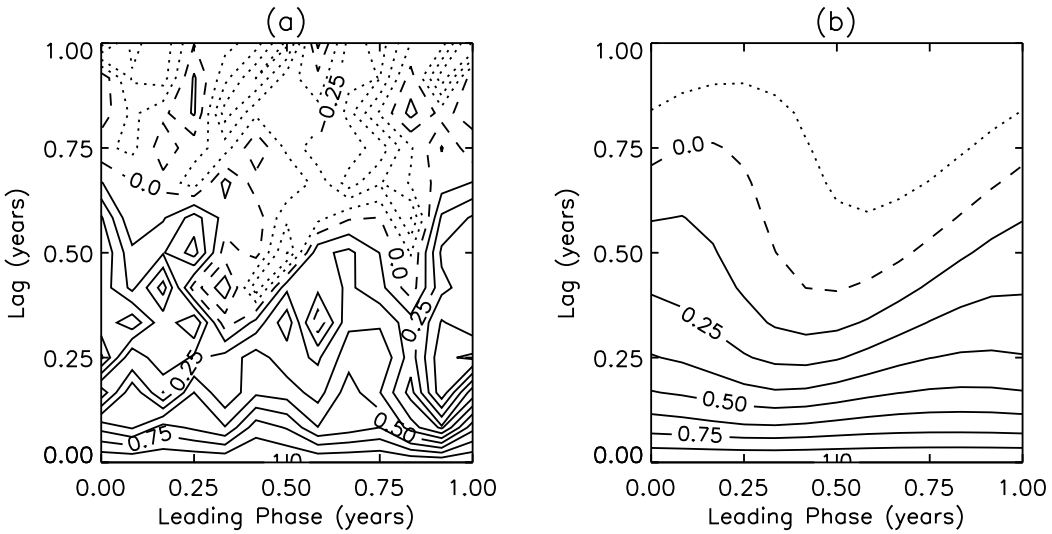

Figure 5.20: Persistence map of the SOI for a 20-year window centered at 1931. (a) Persistence map of the data. (b) The best-fit model barrier. The goodness of fit is $R^2 = 0.748$, compared to $R^2 = 0.841$ for the full timeseries.

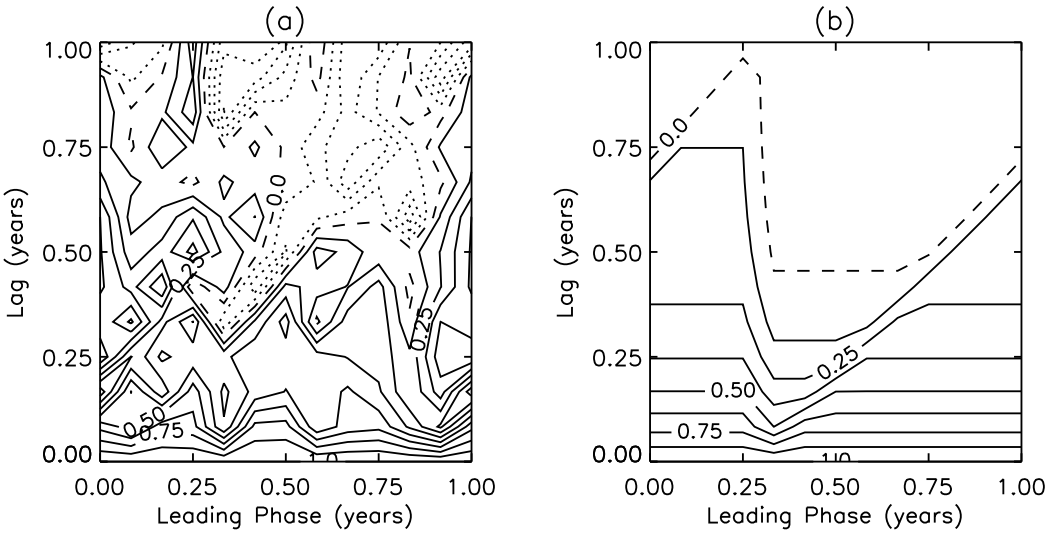

Figure 5.22: Persistence map of the SOI for a 20-year window centered at 1933. (a) Persistence map of the data. (b) The best-fit model barrier. The goodness of fit is $R^2 = 0.684$.

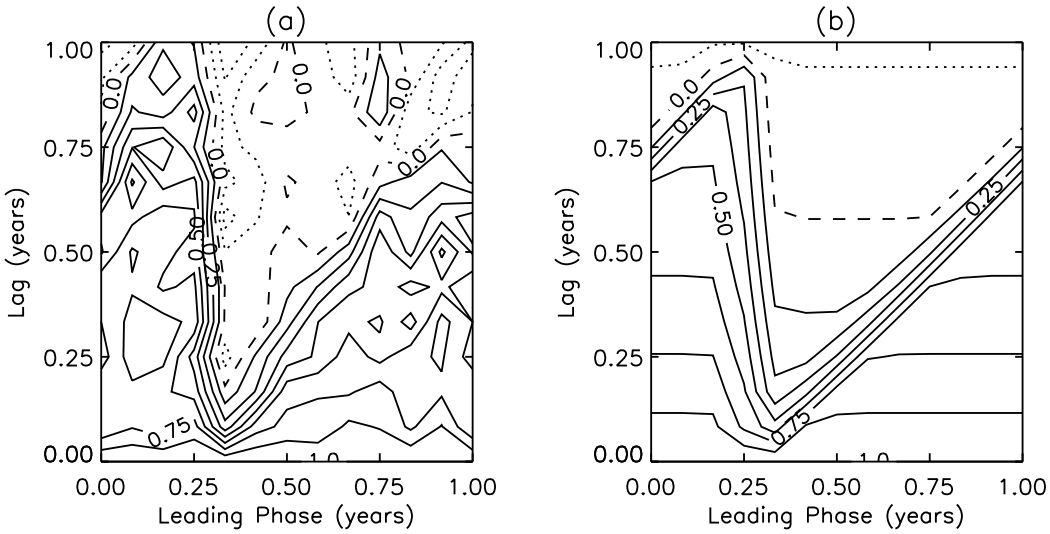

Figure 5.24: Persistence map of the SOI for a 20-year window centered at 1975. (a) Persistence map of the data. (b) The best-fit model barrier. The goodness of fit is $R^2 = 0.846$.

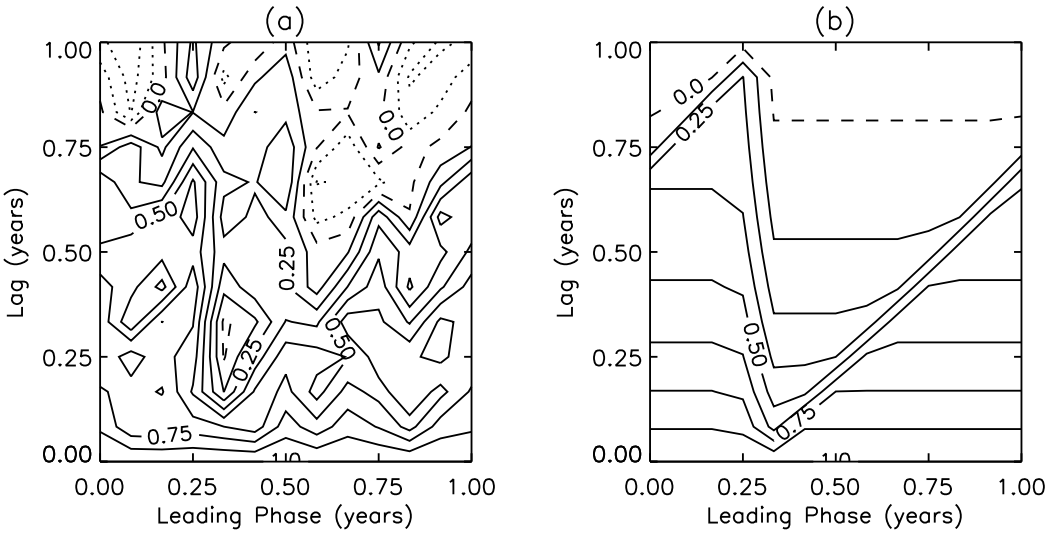

Figure 5.26: Persistence map of the SOI for a 20-year window centered at 1987. (a) Persistence map of the data. (b) The best-fit model barrier. The goodness of fit is $R^2 = 0.796$.

with a decorrelation time of 0.6 years. We therefore will interpret the increase in A from 1965 to 1975 as a strengthening of the SOI persistence barrier.

Contrasting Figure 5.24, Figure 5.26 shows the persistence map for the SOI window centered at 1987. The goodness of fit, R^2, exhibits a decrease relative to the 1975 window, though it is not as low as R^2 for the 1931 and 1933 windows. The decrease in A between 1976 and 1987 thus heralds a weakening of the SOI persistence barrier. This coincides with the present weakening of the NINO3 SST barrier.

Several other works have examined the interdecadal variability of ENSO. Wang and Wang (WW96) describe changes in the period of ENSO over three time periods. They find ENSO to have a dominant period of 3-4 years prior to 1910, 5-7 years during 1911-1960, and 5 years after 1961. The change in period at 1910 coincides with the beginning of the 1916-1945 interval we found for the NINO3 SST. This is the interval of low barrier strength and short decorrelation time. The second change in ENSO period, at 1960, falls in the middle of the transition to a strong NINO3 SST persistence barrier, 1946-1975. Note, however, that the NINO3 barrier reaches its largest strength after 1960. There is thus some correspondence with the observations of Wang and Wang (WW96).

Two other works examine the interdecadal variability of how ENSO phase-locks to the annual cycle. Torrence and Webster (TW98) find strong phase-locking between 1960 and 1977, with weak phase-locking after 1978. Balmaseda et. al. (BDA95) also find strong phase-locking to the annual cycle during the 1970's and weak locking during the 1980's. They also see weak phase-locking during the late 1910's and early 1920's. They additionally note weaker persistence during the 1920's and 1980's and stronger persistence in the 1970's. This is consistent with the behavior of the NINO3 SST measures, which show a strong barrier between 1960 and 1980 and a weakening barrier after 1980. The connection of barrier strength to how strongly ENSO phase-locks to the annual cycle additionally corroborates the work of section 5.2.2.

5.4 Summary

This chapter has made use of the results of chapter 4, especially the barrier measure. When applied to the SOI and NINO3 SST, the measures show that the persistence of both

data sets contain barriers which are statistically significant from the theoretical red-noise barrier to within 68% confidence. In section 5.2.1, comparisons of both datasets with a Monte-Carlo simulation of an Ornstein-Uhlenbeck process reveal a similar difference between the persistence of the ENSO data and that of a red-noise process. The "predictability barrier" is therefore an intrinsic statistical property of the ENSO data.

The selective filtering experiments performed on the NINO3 SST and the SOI in section 5.2.2 demonstrate the relationship of unit-complementary harmonics to persistence barriers. For the SOI, the near-biennial modes are entirely responsible for generating the barrier, whereas all unit-complementary harmonics of the NINO3 SST act to generate a persistence barrier. Following these experiments, the phase-scrambling Monte-Carlo experiments performed in section 5.2.3 demonstrate the necessity of phase-coherency between the barrier-generating modes. Both sections support the hypothesis that barrier-generating modes must have slowly-varying or mutually-coherent phase angles.

The importance of phase-coherency in the formation of a persistence barrier is also demonstrated in other work (XCZB94, TW98). Xue et. al. (XCZB94) show that the barrier is related to low cyclostationary variance. One can only have low cyclostationary variance, however, if the dominant harmonics of a timeseries have coherent or slowly-varying phases. Only then will their superposition approach zero in a periodic fashion. Additionally, Torrence and Webster (TW98) demonstrate a connection between changes in the ENSO persistence barrier and changes in whether ENSO phase-locks to the annual cycle. It is likely that this phase-locking to the annual cycle generates the necessary coherency between the unit-complementary harmonics in the SOI and NINO3 SST.

The chapter concludes with an application of the set of measures to the interdecadal variability of ENSO. The interdecadal variability in the measures shows a weak persistence barrier from 1915 to 1945 and another decline in the barrier after 1975. In contrast, the 1960's and 1970's show a strong barrier. This is consistent with other investigations of ENSO interdecadal variability, (BDA95,WW96,TW98) which find changes in ENSO behavior during the 1920's, 1960's and 1980's. Additionally, the barrier appears to have shifted location since the beginning of the century. At present, the persistence barrier in both ENSO datasets is weak.

The connection between the phase-coherency of the barrier-generating modes, the

ENSO persistence barrier, and phase-locking of ENSO to the annual cycle hints tantalizingly at some interesting underlying dynamics. Perhaps examining the barrier measures produced by a dynamical system following the quasiperiodic route to chaos will reveal these underlying dynamics. Such an exploration is precisely the work of the next chapter.

Chapter 6

Persistence and Dynamical Systems

6.1 On Dynamical Systems

6.1.1 Introduction

In the last two chapters, we examined the mathematical properties of persistence barriers and learned of a connection between phase-locking and persistence barriers. One question remains: what relationship, if any, is there between persistence barriers and actual dynamical systems behavior, such as Lyapunov exponents? Furthermore, the introduction [chapter 1] speculated a possible connection between ENSO models and the so-called "quasiperiodic route to chaos." We shall examine these issues in this chapter.

Rather than reiterate the contents of other chapters, I refer the reader back to section 1.2.1, which defined the concepts of quasiperiodicity, phase space, parameter space, and strange attractors. Section 2.1.2 described the standard notation used in dynamical systems analysis. Recall that one writes the equations of motion for a system according to Equation (2.1),

$$\dot{\vec{x}}\left(t\right) = \vec{F}\left(\vec{x}\left(t\right)\right),$$

where the dot denotes a derivative with respect to time. Another useful quantity is the Jacobian matrix defined in Equation (2.3),

$$J_{ij}\left(\vec{x}\left(t\right)\right) = \frac{\partial F_i}{\partial x_j}\left(\vec{x}\left(t\right)\right).$$

The vector $\vec{x}$ is a vector of the dependent variables of the system. When the dependent variables of the system are functions of not only time but spatial location, $\vec{x}$ is an infinite-

length vector containing the dependent variables evaluated at every spatial location. Such a set of coupled ordinary differential equations is known as a **flow** (Tab89).

There is another way to describe a dynamical system. If one conceptualizes time as a discrete variable instead of a continuous one, one can write the equations of motion using the notation of Equation (2.7):

$$\vec{x}_{n+1} = \vec{F}\left(\vec{x}_n\right),$$

where the subscript indicates a discrete timestep, t_n. The function $\vec{F}$ now maps the state of the system at one timestep to the state at the next timestep. Naturally, such a set of equations is known as a **mapping** (Tab89). This is another common way of describing a dynamical system. The Jacobian matrix of the system is, as for a flow, the derivative of $\vec{F}$ with respect to the components of $\vec{x}$.

For both flows and mappings, there exists a technique known as stability analysis (Tab89, NL92, BG90). There exist certain solutions of the equations of motion, solutions known as fixed points, defined by $\vec{F}\left(\vec{x}^{\star}\right) = 0$. One then evaluates the Jacobian, $\mathbf{J}\left(\vec{x}\right)$, at $\vec{x}^{\star}$ and computes it eigenvalues, which can be complex numbers. The signs of the real and imaginary parts of each eigenvalue then describe a particular type of motion near $\vec{x}^{\star}$. Appendix B performs a stability analysis for the damped forced pendulum and discusses it in greater detail.

The damped forced pendulum raises the issue of dynamical systems whose equations of motion are themselves dependent on time. Such systems are called non-autonomous. It is standard practice to reparameterize the time-dependent portions of the equations of motion as new dependent variables (Tab89). One can often simplify examinations of the phase space of such systems using Poincaré Sections [see section 1.2.1], especially when the time-dependence is periodic. Through the Poincaré Section, one converts an N-dimensional flow to an $(N-1)$-dimensional mapping with the same dynamics as the original flow.

6.1.2 The Quasiperiodic Route to Chaos

One of the fundamental works examining the quasiperiodic route to chaos [hereafter called "QP-chaos"] are the papers by Bak, Bohr, and Jensen (JBB84, BBJ84, hereafter called BBJ). A later work by Bak (Bak86) describes some of the same results in a more accessible

form. The introduction contained a rather vague description of QP-chaos [section 1.2.1], most notably of the so-called "Devil's Staircase" and "critical curves." Let us now examine these concepts in a more mathematical form.

BBJ describe QP-chaos in terms of a quantity know as "winding number." The concept of winding number is appropriate to systems whose attractors form a 2-torus in phase space (Tab89). Let us call the axis of the torus that forms a circle through the center of its volume the "minor" axis and denote the axis through the hole in the torus as the "major" axis [Figure 6.1]. The winding number describes the number of revolutions a trajectory makes about the minor axis before completing a single revolution about the major axis.

Let us denote the angle of the trajectory about the minor axis as $\vartheta(t)$ and refer to this angle as the **winding angle**. The mathematical definition of the winding number is then (JBB84, Bak86, Tab89, TCZ95):

$$W \equiv \lim_{t\to\infty} \frac{\vartheta(t) - \vartheta(t_0)}{t - t_0}. \tag{6.1}$$

The winding angle may be one of the directions of phase space, i. e. one of the dependent variables of the system. Then again, it may not. In general one must first discern what coordinates describe the 2-torus for the system. Only then does the definition of the winding angle in terms of the system's dependent variables become clear.

One can now describe QP-chaos in terms of the value of the winding number. When the winding number is rational — say, $W = \frac{P}{Q}$, where P and Q are integers — the system takes Q revolutions about the major axis to make P revolutions about the minor axis (JBB84, Tab89). If motion about the major axis represents the period of a driving force or some other fundamental physical timescale of the system, the system is said to be "mode-locked." Mode-locking is, of course, periodic motion; trajectories stay in a specific region on the 2-torus. When W is an irrational number, the motion of the system is still regular, but never repeats. Trajectories eventually cover the surface of the 2-torus. This is quasiperiodic motion.

A very simple system studied by BBJ is the so-called "sine circle map," defined as:

$$\vartheta_{n+1} = f(\vartheta_n) \bmod 2\pi, \tag{6.2}$$

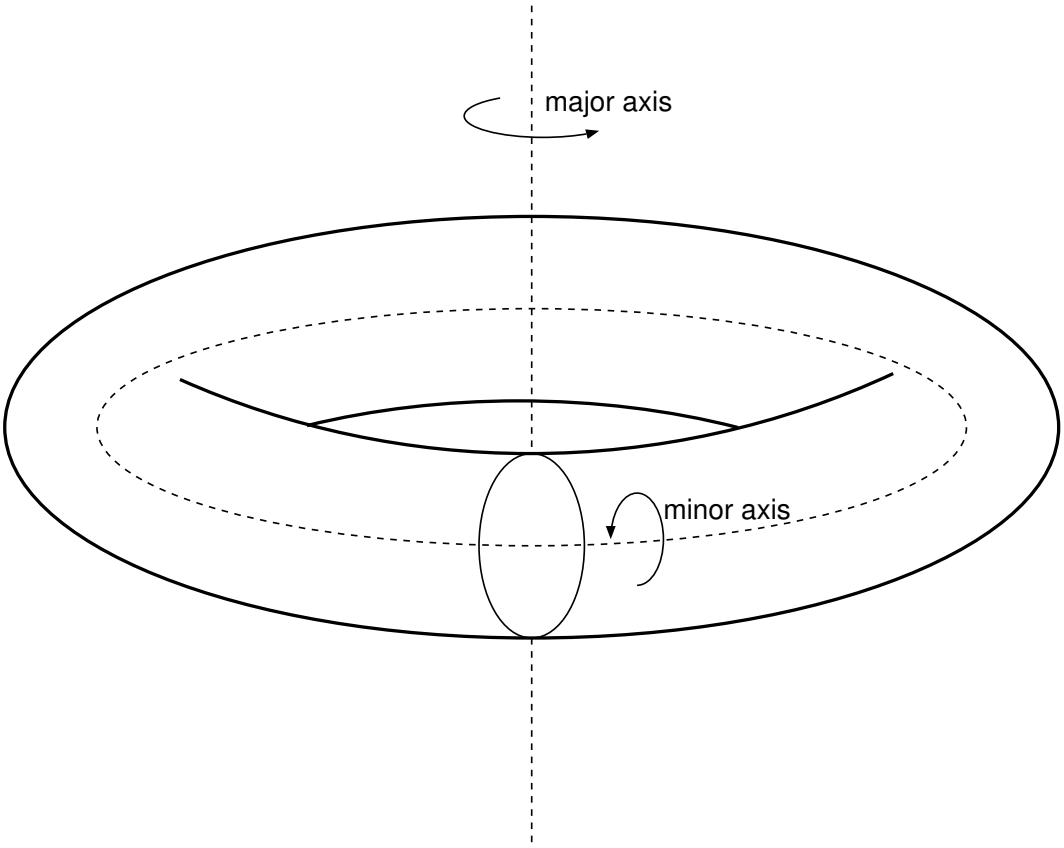

Figure 6.1: A 2-torus and its axes.

where

$$\begin{aligned} f(\vartheta_n) &\equiv \vartheta_n + \Omega + \frac{K}{2\pi}\sin(\vartheta_n), \\ &\equiv \vartheta_n + \Omega + Kg(\vartheta_n). \end{aligned} \tag{6.3}$$

One can cast many dynamical systems into the above form by specifying different $g(\vartheta)$, K, and Ω (AB84, BBJC84, Tab89). For the case of $g(\vartheta) = \frac{1}{2\pi}\sin(\vartheta)$, the sine-circle map, Ω is the "natural" frequency of the system and K is equivalent to the amplitude of a periodic driving force (JBB84, BBJ84). BBJ demonstrate the Devil's Staircase with a plot of the winding number, W, as a function of Ω. At particular values of Ω, W remains constant at a rational number. Each of these constant regions forms a "step" on the Devil's Staircase. "Gaps" between the steps, where W becomes irrational, indicate quasiperiodic states. At $K = 1$, the Devil's Staircase becomes "complete." That is, all steps meet, without overlap or gaps, and the system only exhibits mode-locked behavior.

A map of rational W as a function of both K and Ω is called an "Arnol'd map" by BBJ (JBB84); I shall use that term in this chapter. BBJ refer to the regions of rational W as "Arnol'd Tongues." Between the Arnol'd Tongues are areas of quasiperiodic behavior. At $K = 1$, all of the tongues touch. The function $K(\Omega) = 1$ is also termed the "critical curve" and is a more complex function in other systems (JBB84, AB84). Beyond the critical curve — in this case, for $K > 1$ — the mode-locked regions overlap and the system becomes chaotic.

In general, the presence of mode-locking and a critical curve along which exists a complete Devil's Staircase is a sign of the quasiperiodic route to chaos. Section 6.2 examines QP-chaos in the damped forced pendulum and its relationship to persistence. After a description of the equations of motion for the damped forced pendulum in section 6.2.1, section 6.2.2 examines existing work on the pendulum, adding a few analytical results from a stability analysis in appendix B. Section 6.2.3 compares Arnol'd maps of the pendulum with corresponding maps of the persistence barrier measures from chapter 4. Section 6.2.4 attempts a connection between the instantaneous Lyapunov exponents and persistence barriers. The chapter concludes with a brief examination of the Vallis model (Val88) in section 6.3.

6.2 The Damped Forced Pendulum

6.2.1 Equations of Motion

The standard equation of motion for an idealized damped forced pendulum is:

$$m\ddot{\theta} + \eta\dot{\theta} + m\omega^2 \sin\theta = a' + b'\cos(2\pi\nu t'), \tag{6.4}$$

where m is the mass of the pendulum bob, η is a damping coefficient, ω is the natural frequency of the pendulum for small amplitude motion,[1] a' is a constant driving torque, b' is the amplitude of a periodic driving torque with frequency ν, and a dot represents a derivative with respect to time, t'. Given units of kg for m and seconds for t', the units of ν and η are $\sec^{-1}$ and kg/sec, respectively, ω^2 has units of rad/$\sec^2$, and a' and b' have units of (kg rad)/$\sec^2$. Because this work draws analogies between the pendulum and climatological phenomena driven by an annual cycle, it is convenient to rescale (6.4) using $\phi(t) = \theta(\nu t')$. This results in the following unitless form of the pendulum equation:

$$\ddot{\phi} + \beta\dot{\phi} + \gamma\sin\phi = D_0 + D_1\cos(2\pi t). \tag{6.5}$$

Rewriting (6.5) in standard form:

$$\dot{\phi} = p, \tag{6.6}$$

$$\dot{p} = -\beta p - \gamma\sin\phi + D_0 + D_1\cos(\tilde{\tau}), \tag{6.7}$$

$$\dot{\tilde{\tau}} = 2\pi, \tag{6.8}$$

where p is the unitless angular momentum and $\tilde{\tau}$ is the forcing timescale. The pendulum thus has a 3-dimensional phase space which is periodic in the $\tilde{\tau}$-direction.

One may wonder what range the parameters of (6.6-6.8) have. It is straightforward to show that a change in sign in any one of the parameters is equivalent to a coordinate

[1] $\omega^2 = \frac{mg}{L}$, where L is the distance from the pivot to the pendulum bob and g is the acceleration due to gravity.

transform. The symmetries are as follows:

$$\beta \to -\beta \quad \Leftrightarrow \quad \tilde{\tau} \to -\tilde{\tau}, \tag{6.9}$$

$$\gamma \to -\gamma \quad \Leftrightarrow \quad \phi \to \phi + \pi, \tag{6.10}$$

$$D_0 \to -D_0 \quad \Leftrightarrow \quad \phi \to -\phi \quad \text{and} \quad D_1 \to -D_1, \tag{6.11}$$

$$D_1 \to -D_1 \quad \Leftrightarrow \quad \tilde{\tau} \to \tilde{\tau} + \pi. \tag{6.12}$$

Note that these symmetries are related to two other symmetries: $\phi \to -\phi \Leftrightarrow D_0 \to -D_0$ and $D_1 \to -D_1$; and $\tilde{\tau} \to \tilde{\tau} + \pi \Leftrightarrow \phi \to -\phi$ and $p \to -p$. The latter symmetry has interesting consequences, as we shall see. Because the changes in the sign of the parameters simply induces a coordinate transform, let us subsequently assume that the parameters β, γ, D_0, and D_0 have positive values.

There is another common rescaling of Eq. (6.4), used in works describing its relationship to Josephson junctions (DBHL82). It is of the form:

$$\dot{\theta}' = p', \tag{6.13}$$

$$\dot{p}' = -\hat{\beta} p' - \sin\theta' + \hat{D}_0 + \hat{D}_1 \cos(\tau'), \tag{6.14}$$

$$\dot{\tau}' = 2\pi\hat{\nu}. \tag{6.15}$$

The parameters, $\hat{\nu}$, $\hat{\beta}$, $\hat{D}_0$, and $\hat{D}_1$, are related to the parameters of Eqs. (6.6-6.8) by the following equations:

$$\beta = \frac{\hat{\beta}}{\hat{\nu}}, \tag{6.16}$$

$$\gamma = \frac{1}{\hat{\nu}^2}, \tag{6.17}$$

$$D_0 = \frac{\hat{D}_0}{\hat{\nu}^2}, \tag{6.18}$$

$$D_1 = \frac{\hat{D}_1}{\hat{\nu}^2}. \tag{6.19}$$

One may wonder why this work does not use the rescaling of Eqs. (6.13-6.15). For Josephson junctions, the experimenter applies the driving force and therefore can easily tune it. The natural frequency of the Josephson junction, however, is a fixed property of that junction.

In the atmospheric sciences, the "natural frequency" of ENSO is a subject of debate. The annual cycle is not, nor is it tunable. Since it is the annual cycle that drives the dynamics of the oceans and atmosphere, it is only natural to parameterize the pendulum equations with a fixed driving force.

6.2.2 Examination of Dynamical Behavior

6.2.2.1 2-Tori, Winding Number

The damped forced pendulum can undergo both rotational and librational motion. In the case of the former, the angular direction, ϕ, either increases or decreases continuously in an oscillatory fashion. However, this coordinate has the property $\phi = \phi \bmod 2\pi$, similar to the circle map defined in Eq. (6.2). Thus, motion of a rotationally-oscillating pendulum through its phase space forms a 2-torus [Figure 6.2]. Let us associate the τ-direction with motion along the major axis of the torus, letting ϕ describe the winding angle of Eq. (6.1). The winding number for the pendulum, then, is

$$W_r = \lim_{t\to\infty} \frac{\phi(t) - \phi_0}{t}, \tag{6.20}$$

where ϕ_0 is the value of ϕ at time $t = 0$. The subscript "r" indicates that this winding number applies to rotational states of the pendulum. Note that $W_r \simeq \bar{p}$, where the bar indicates an average over time.

For the case of librational oscillations, where motion is bound to $-\pi < \phi < \pi$, Eq. (6.20) will always return a winding number of zero. Nevertheless, the motion of the pendulum still forms a 2-torus with motion in the τ-direction along the major axis [Figure 6.3]. The other axis is associated with the winding angle $\psi(t)$, defined as:

$$\psi(t + dt) \equiv \psi(t) + d\psi, \tag{6.21}$$

where

$$d\psi \equiv \tan^{-1}\left(\frac{p(t+dt) - \bar{p}}{\phi(t+dt) - \bar{\phi}}\right) - \tan^{-1}\left(\frac{p(t) - \bar{p}}{\phi(t) - \bar{\phi}}\right), \tag{6.22}$$

$\bar{\phi}$ and $\bar{p}$ are average values of ϕ and p, respectively, and the function $\tan^{-1}\left(\frac{y}{x}\right)$ ranges from

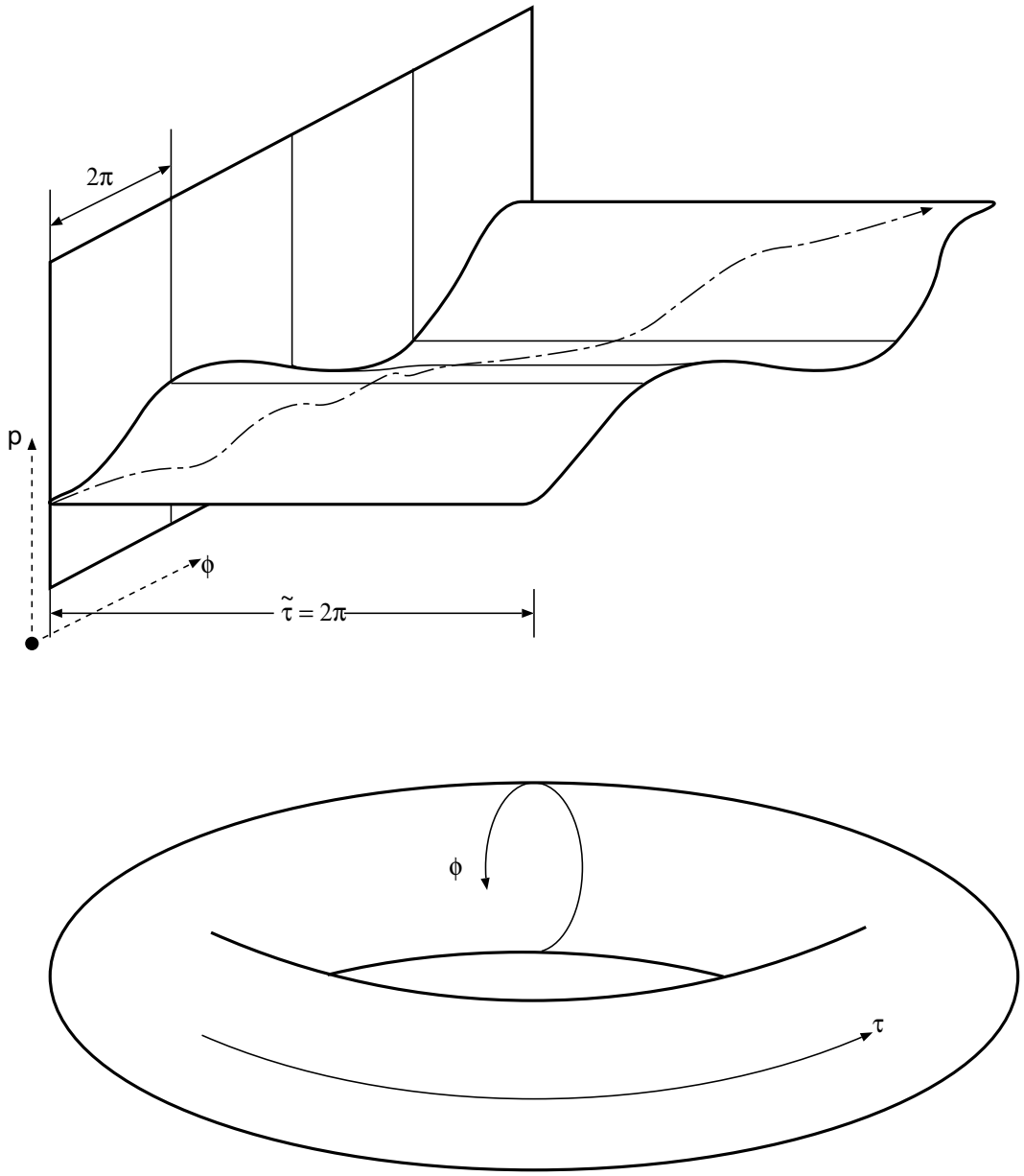

Figure 6.2: Image of the 2-torus for rotational motion. Shown both as a 2-torus [lower graph] and "cut open" [upper graph].

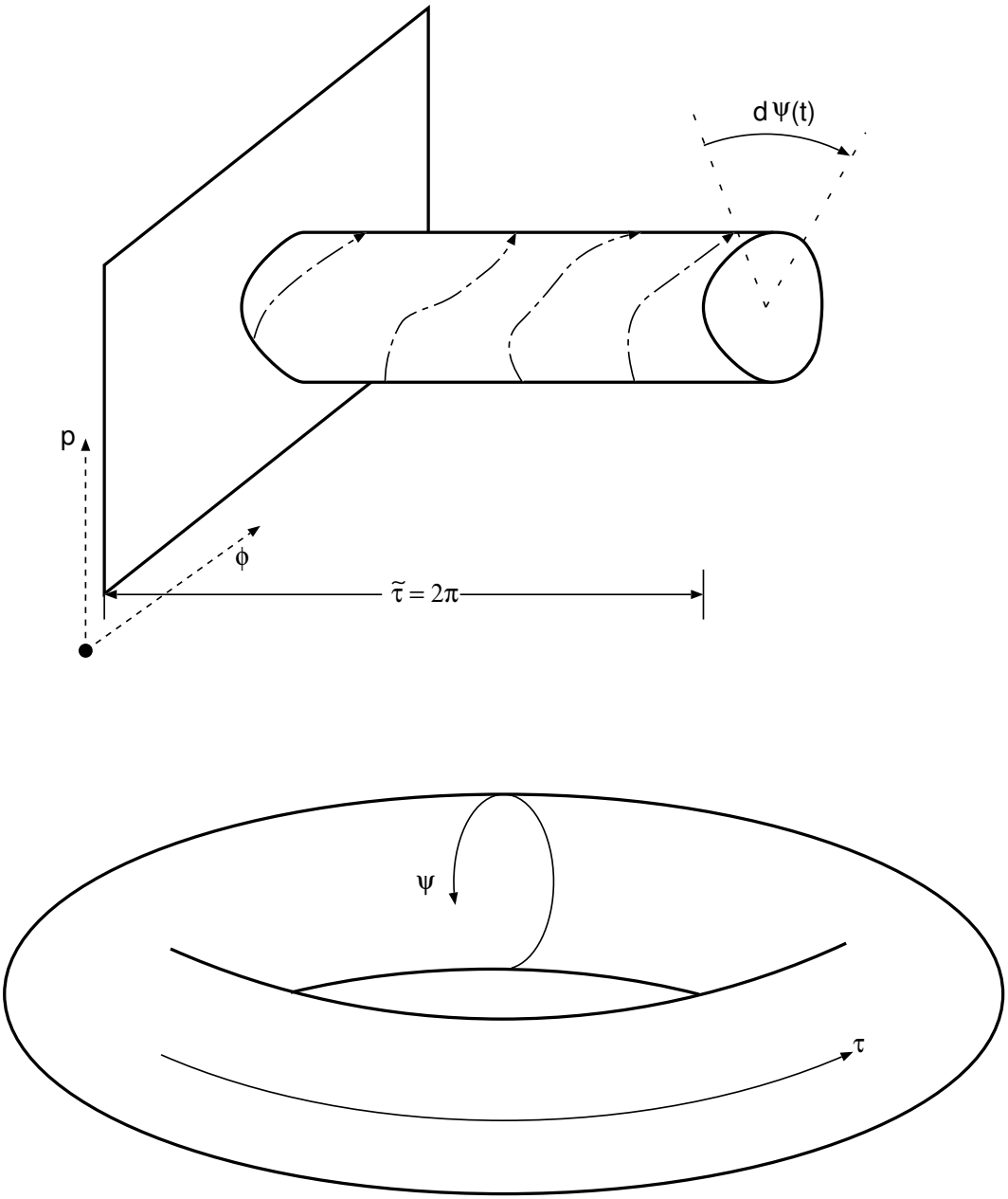

Figure 6.3: Image of the 2-torus for librational motion. Shown both as a 2-torus [lower graph] and "cut open" [upper graph]. Also shown is the winding angle, $\psi(t)$, for this 2-torus.

0 to 2π, using the signs of x and y to determine the quadrant of the winding angle.

Because of the occurrence of a second 2-torus, one must define another winding number for librational states,

$$W_l \equiv \lim_{t\to\infty} \frac{\psi(t) - \psi_0}{t}. \tag{6.23}$$

As before, ψ_0 is the value of ψ at $t = 0$ and the subscript "l" indicates that this winding number applies to librational states of the pendulum. Note that it is not equivalent to Eq. (6.20), even in the case of rotational motion, nor does $\overline{p}$ approximate it.

6.2.2.2 Some Hints: Stability Analysis

In general, Equations (6.6-6.8) do not have a closed analytical solution and must be integrated numerically. There are, however, certain special cases where (6.6-6.8) can be solved analytically. These are presented in appendix B. The appendix also describes the fixed points of (6.6-6.8) and performs a stability analysis for these special cases. These cases are rather simple; however, the goal is to gain insight into the general behavior of the damped forced pendulum using simplifying limits.

The two "gravityless" cases considered in section B.1 are quite trivial. Nevertheless, these are the only two cases for which one can analytically find the winding number [section B.3.1]. Most interesting is the case of a gravityless, damped pendulum with constant forcing. The pendulum undergoes rotation and has a winding number, $W_r = \frac{D_0}{\beta}$. Note, however, that this relation appears to hold in numerical results for $\gamma \neq 0$. Figure 6.4 contains the Arnol'd map of the pendulum for $0.0950712 \leq \beta \leq 4.75356$, $0.2 \leq D_0 \leq 20$, $\gamma = 4.398150$ and $D_1 = 45$. The map is plotted in the same manner as those of section C.2. The algorithm described in section C.1.3 chooses the appropriate winding coordinate. There are bands of period-2, period-3, period-4, period-$\frac{3}{2}$, period-$\frac{8}{5}$, and period-$\frac{4}{3}$ behavior plainly visible. All of the bands fall along lines of constant $\frac{D_0}{\beta}$.

The next most complicated special case is that of $D_1 = 0$, no periodic forcing. An integral of the motion [Equation (B.3)] exists when, additionally, there is no damping. Using this integral of the motion to solve for $\phi(t)$, however, is difficult and possibly intracable [section B.1]. A stability analysis is possible for the full $D_1 = 0$, with fixed points at $\phi^\star = \pi - \arcsin\left(\frac{D_0}{\gamma}\right)$ and $\phi^\star = \arcsin\left(\frac{D_0}{\gamma}\right)$ for all $D_0 < \gamma$ [section B.2]. The former

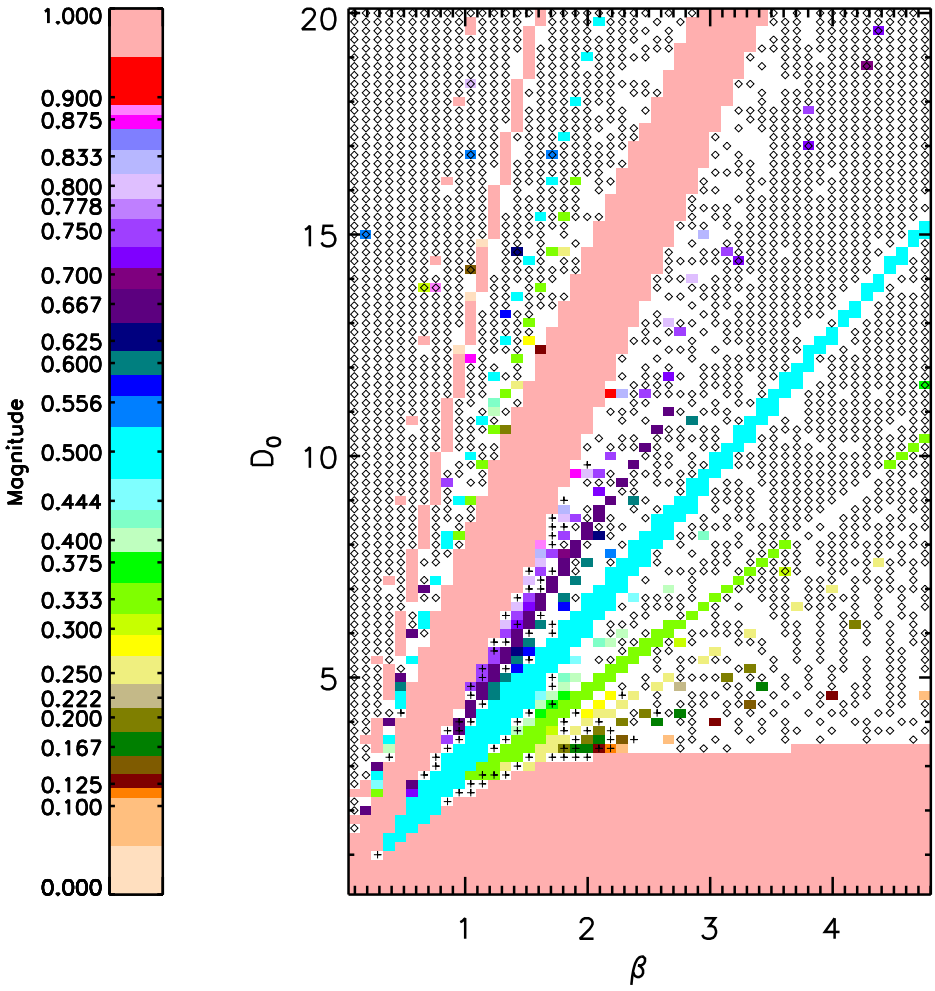

Figure 6.4: Arnol'd map for the damped forced pendulum, with $\gamma \approx 4.4$ and $D_1 = 45$ [Tableau #10], plotted in the same fashion as Figure C.1. "+" indicates positive GLE; "◇" indicates zero GLE. Gridpoints set to the background color indicate zero winding number. A detailed description of the plotting method is in section C.2.

fixed point is always a saddle-node point, while the latter is some form of stable or elliptic point, depending on the strength of the damping. Notice that the presence of a constant driving force shifts both fixed points towards $\phi = \frac{\pi}{2}$ as D_0 increases, rendering the separatrix asymmetric. At $D_0 = \gamma$, there is a bifurcation. This may or may not indicate a transition to chaos.

For $D_1 \neq 0$, the pendulum has no true fixed points, due to Eq. (6.15). However, section B.3.2 of appendix B used the transform $\tilde{\tau} = 2\pi\,(n + \delta)$, where $0 < \delta < 1$ and n is an integer, to convert the pendulum equations to a mapping, one strikingly similar to the standard map (Tab89). This new mapping has a new parameter, δ, and fixed points at $\phi^\star = \pi - \arcsin\left(\frac{\alpha}{\gamma}\right)$ and $\phi^\star = \arcsin\left(\frac{\alpha}{\gamma}\right)$, where $\alpha = D_0 + D_1 \cos(2\pi\delta)$. Although this does not represent the true dynamics of the damped forced pendulum, it is suggestive. The periodic driving eliminates the fixed points described in the last paragraph. Perhaps, for sufficiently small D_1, those erstwhile fixed points become basins of attraction, one a stable basin, the other a "saddle-node" basin. As D_1 increases, those basins of attraction would eventually begin to overlap. Indeed, the aforementioned mapping hints at a relationship between the parameters γ, D_0, and D_1 and a set of bifurcations in the true pendulum.

6.2.2.3 The Pendulum in Crisis

The idea of two basins of attraction "colliding" as one tunes the parameters of the system is actually a route to chaos known as "crisis" (GW85). Crisis-induced chaos is a form of intermittency, a behavior in which a system erratically alternates between two stable states. Gwinn and Westervelt examined the pendulum for crisis-induced intermittency (GW86). They use the parameterization of Equations (6.13-6.19) with the constant driving, $\hat{D}_0$, set to zero. They find that the crisis is due to the collision of an attractor and a saddle-node point, a situation speculated in the previous section. Gwinn and Westervelt also state that the crisis is controlled by the value of the parameter $\hat{D}_1$ [Eqs. (6.13-6.15)].

6.2.2.4 Bifurcations: Period Doubling and Saddle-Node

As noted earlier, works using the damped forced pendulum to examine Josephson junctions parameterize the equations in the form of Equations (6.13-6.19). The fundamental work in this area is that of D'Humieres et. al. (DBHL82, hereafter DBHL), who examine

the pendulum with $\hat{D}_0 = 0$, using a voltage controlled oscillator to "integrate" the system. DBHL also find intermittency in the pendulum due to two unstable phase-locked states. They do not, however, discuss the type of intermittency.

Gwinn and Westervelt (GW85) do examine types of intermittency. In addition to crisis-induced intermittency, there is also "Pommeau-Manneville Type-1" intermittency, generated by a saddle-node bifurcation (GW86). The bifurcation generates a pair of attractors from one. As in crisis-induced intermittency, the periodic driving controls the Pommeau-Manneville intermittency (GW86). Tominaga et. al. (THM90) examine this dependence in greater detail. They find that, prior to the saddle-node bifurcation, the pendulum contains a limit cycle.

Appendix B speculates a tie between a bifurcation and $\hat{D}_1$, as well as noting the change in symmetry induced by the periodic driving. Kerr et. al. (KWB$^+$85) examine the symmetry breaking in detail. They note that the symmetry[2] $\tilde{\tau} \rightarrow \tilde{\tau} + \pi \Leftrightarrow \phi \rightarrow -\phi$ and $p \rightarrow -p$does not always hold, producing a second solution. Gwinn and Westervelt (GW86) note that the symmetry breaking generates the saddle-node bifurcation leading to Pommeau-Manneville intermittency. Kerr et. al. (KWB$^+$85), however, see an effect on period-doubling bifurcations in the pendulum.

DBHL also noticed the tie between symmetry breaking and the period-doubling route to chaos in the pendulum. They see symmetry-breaking in the form of $\phi(t)$ gaining a larger amplitude in one direction than the other. This, DBHL, note, occurs just before the pendulum goes chaotic. They also find several other interesting properties of the period-doubling cascade. It occurs for both librational and rotational states. They see a period-3 state that undergoes a period-doubling cascade. They also see "incomplete" period-doubling, where the doubling reverses itself before the system goes chaotic. [Figure C.3 contains an example of such an "incomplete" period-doubling.] They hypothesize that the relationship between the pendulum's parameters and the "true" period-doubling routes to chaos is a non-monotonic function. Both DBHL and Gwinn and Westervelt (GW86) examine the bifurcations as a function of $\hat{D}_1$.

One thing missing from all of these descriptions of intermittency and period-doubling chaos in the pendulum is constant forcing. All examine changes in behavior as a function of

[2]or, under the parameterization they use, $\tau' \rightarrow \tau' + \pi \Leftrightarrow \theta \rightarrow -\theta$ and $p' \rightarrow -p'$,

mainly $\hat{D}_1$ and $\hat{\nu}$. DBHL did note changes in the symmetry breaking, a shift in the locations of attractors, and a change in the onset of chaos when they added a small $\hat{D}_0$. They also saw a dependence of the winding number on the ratio $\frac{\hat{D}_0}{\hat{\beta}}$, confirming the work here. They specifically state, however, that they do not find any examples of the quasiperiodic route to chaos. As we shall see, QP-chaos has a strong connection to the constant driving, D_0.

6.2.2.5 Quasiperiodicity and Critical Curves

There are several works examining a connection between the damped forced pendulum and the quasiperiodic route to chaos [QP-chaos] (AB84, JBB84, BBJ84, BBJC84, AL85b, AL85a, AL86). These works, too, use the parameterization of Equations (6.13-6.19), but with $\hat{D}_0 \neq 0$. In addition to attempting to cast the pendulum equations into the form of a circle map for the case of strong damping, Azbel and Bak (AB84) find a transition from libration to rotation when $\hat{D}_0$ is larger than a critical value. Bak et. al. (BBJC84) reveal that tuning $\hat{\beta}$ and $\hat{D}_0$ changes the winding number. They further find that the Devil's Staircase and the critical curve exist in the $\hat{\beta}$-$\hat{D}_0$ plane of parameter space. Chaos occurs at values smaller than the critical $\hat{\beta}$ and $\hat{D}_0$. Continuing this work, Alstrøm and Levinsen (AL85b) compute several critical curves.

6.2.3 Persistence Barriers: Parameter Space Explorations

The discussion in the previous section suggests that intermittency and period-doubling is restricted to the $D_0 = 0$ subspace of the pendulum parameter space, while QP-chaos occurs only when $D_0 \neq 0$. It was also noted that the QP-chaos critical curve exists in the $\hat{\beta}$-$\hat{D}_0$ subspace of parameter space. Note that this corresponds to variations in β, γ, and D_0 for the parameterization used by this work.

Let us make use of the existing body of work to guide an exploration of parameter space for the damped forced pendulum. Appendix C holds the results of this exploration, as well as all details behind the numerical integration of Eqs. (6.6-6.8) and the computation of the Arnol'd maps. In this exploration, QP-chaos does indeed occur only when $D_0 \neq 0$. Because section C.2 details the pendulum's behavior in each examined region of parameter space, I shall instead provide a summary of the results.

One thing not fully discussed in section C.2 is the choice of a cutoff value of R^2, the

nonlinear correlation coefficient [Equation (3.11)]. The cutoff determines parameter points at which the persistence does not fit the barrier model [Equation (4.24)] well. To determine a cutoff value, let us assemble all values of R^2 from Figures C.1-C.24 in section C.2. Using this set of values, let us now examine the distribution function for R^2. Figure 6.5 contains such a histogram, using 50 bins over the range of valid R^2 values. The median value of this probability distribution function [PDF], is almost exactly 1.0, with a confidence interval of ± 0.057 at the 95% level. However, this value describes the bounds of 95% of the area centered on the median. Let us instead use the percentile value at 95%, the point in the PDF bounding 95% of the area of the PDF. This value is $R^2 \cong 0.98$, i. e. there is a 95% chance that a point in Figures C.1-C.24 has $R^2 \geq 0.98$. Therefore, let $R^2_{\text{cutoff}} = 0.98$.

In the QP-region of parameter space, large areas of period-1 behavior separate some QP areas from the mode-locked and chaotic regions. Recall that, in the circle map, QP oscillations occur between the mode-locked Arnol'd Tongues. The tongues widen, and where they overlap, chaos ensues. The behavior shown in section C.2 implies a QP-route to chaos which is converse that of the circle map. The mode-locked Arnol'd Tongues and the period-1 region all converge toward a common point in parameter space. Chaotic behavior appears mostly at this convergence point, as well as in between the Arnol'd Tongues. Chaos may also occur along the boundary between period-1 and mode-locked areas.

In the mode-locked islands, the barrier strength, A, is large, while barrier width, Δp, is typically no larger than 0.3 yr. These values of A and Δp hold even for the chaotic points. The mode-locked, period-2 islands always have a strong, narrow barrier, as expected. In the areas with only QP behavior, no chaos, no mode locking, a curious relationship appears. The barrier measures A and Δp show a dependence on the ratios $\frac{D_0}{\beta}$ and $\frac{D_0}{\gamma}$ in these islands of purely QP behavior. For $\frac{D_0}{\beta}$, the two measures decrease as the ratio decreases, while they increase with decreasing $\frac{D_0}{\gamma}$. Furthermore, for all QP areas of all the examined cross-sections through parameter space, the stochastic decay time, τ, decreases as some function of β.

The situation for the $D_0 = 0$ regions of parameter space is not so obvious. There are areas where the barrier strength, A, is large yet the other measures, Δp, p_b, and τ, are undefined. At yet other points with large A, Δp is also large. Though large barrier width often indicates a "flat barrier" — i. e. a purely linear decay in place of the barrier-like feature — it may also signify a persistence map that does not depend solely on lag yet

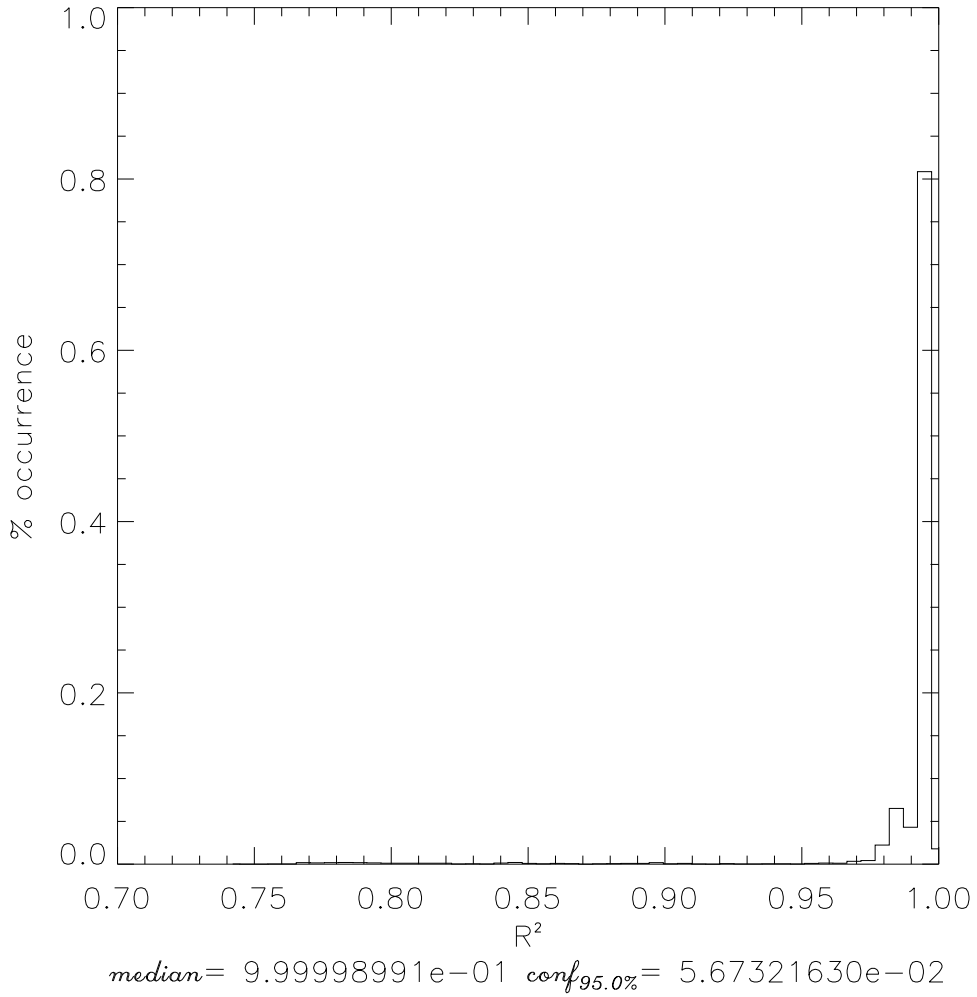

Figure 6.5: Histogram of R^2, the nonlinear correlation coefficient, from Figures C.1-C.24. [The text describes the production of this histogram in more detail.] The distribution is nonzero over the entire displayed range of the graph.

contains some feature that does not resemble the barrier.[3] There are also some points in this region of parameter space where $A > 0.5$ but Δp is small, indicating a strong persistence barrier. In general, however, the only consistent pattern in the $D_0 = 0$ regions of parameter space is the division into period-1 areas and chaotic areas. There is no clear, consistent pattern in A, τ, Δp, and p_b as there is in the QP-region of parameter space, where $D_0 \neq 0$.

The individual cross-sections of parameter space hint at a connection between QP-chaos and the persistence barrier. To better examine this connection, let us construct a two-dimensional probability distribution function [PDF] that is a function of the largest global Lyapunov exponent [hereafter GLE-1] and each of the barrier measures. Let us exclude from the PDF those points that do not satisfy two conditions. The first is the criteria $R^2 \geq 0.98$. For the second, for each measure, exclude those points for which that measure is ill-defined. Specifically, the measures are ill-defined if the upper and lower bounds of the confidence intervals for A, Δp, and p_b are, respectively, larger than one and less than zero. In the case of τ, which has no upper bound, let us consider τ ill-defined when the lower bound of the confidence interval is less than zero. Finally, let us also exclude period-1 states. Because there are so many of them, the period-1 points will dominate the PDF's. This not only masks any non-period-1 behavior, it is completely uninteresting, since the measures for period-1 states are fixed at predefined values. Figures 6.6 and 6.7 contain PDF's of barrier measure vs. GLE-1 [i. e. largest GLE] plotted in the fashion just described, using 25 bins in each direction.

Figure 6.6 contains the PDF for the $D_0 = 0$ area of parameter space. The barrier width occurs almost entirely at a value of $\Delta p = 0$ yr., independent of GLE-1. The bulk of the barrier amplitude occurs at values of $A < 0.1$ and GLE-1 above 2 bits/year. The barrier position, p_b, occurs most frequently in two peculiar bands, also at GLE-1 above 2 bits/year. A similar band occupies $0.1 \leq \tau \leq 1.0$ years. Notice that the PDF is a bit more uniform in A for GLE-1 between 0 and 2 bits/year, though most of the occurrence of A is still below 0.15 and never exceeds 0.7. The other features of the four PDF's, though interesting in their own right, do not hold nearly as much importance as the behavior of A for in the low-GLE-1 interval.

The cause for so much scrutiny of one region of the PDF for A becomes clear

[3] Chapter 4 discusses this possibility.

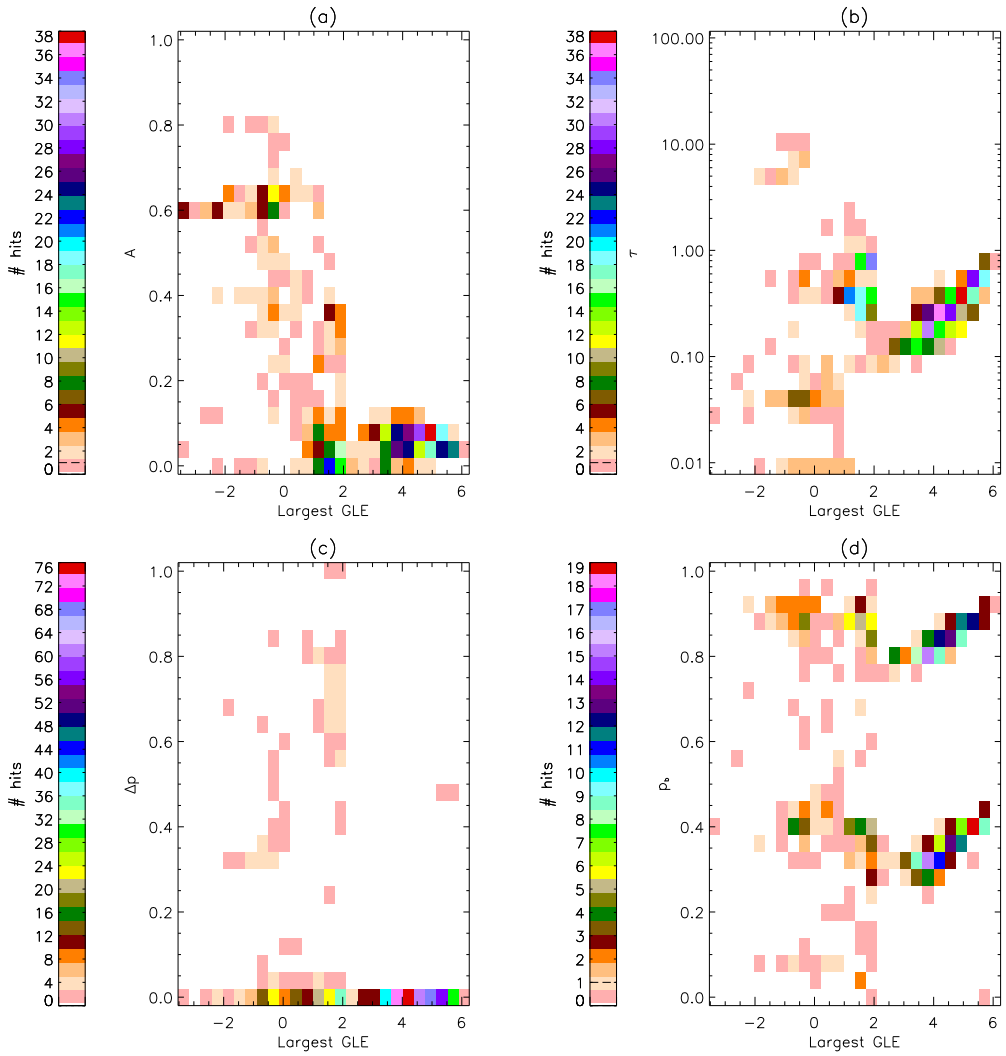

Figure 6.6: Joint histogram of the largest global Lyapunov exponent [GLE] and each of the barrier measures, A, τ, Δp, and p_b, evaluated in the $D_0 = 0$ region of the pendulum's parameter space. The color bar of each graph indicates the color coding for the distribution displayed by that graph only. In all graphs, the background color indicates where the distribution is zero for a given bin. (a) GLE vs. the barrier amplitude, A; contains 857 points. (b) GLE vs. the stochastic decay time, τ; contains 780 points. (c) GLE vs. the barrier width Δp; contains 732 points. (d) GLE vs. the barrier position, p_b; contains 721 points.

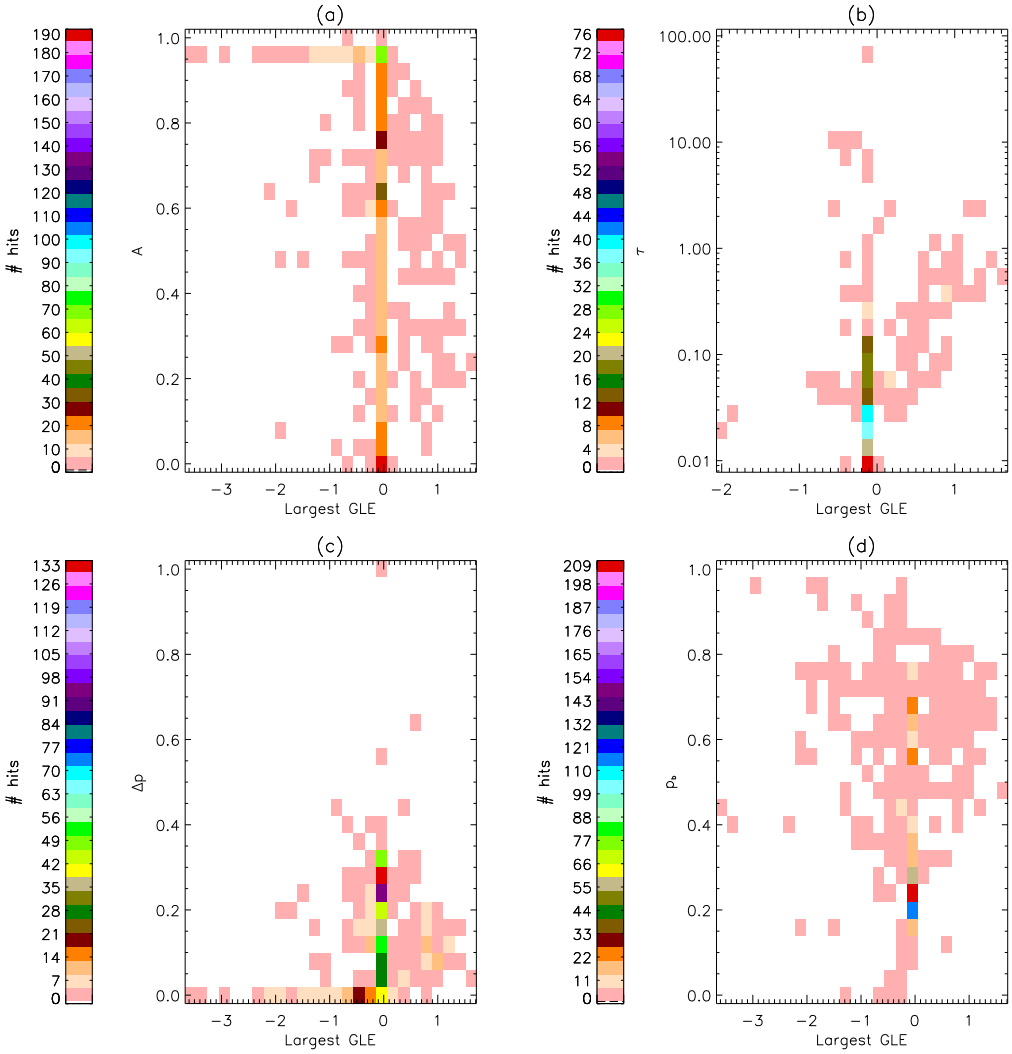

Figure 6.7: Joint histogram of the largest global Lyapunov exponent [GLE] and each of the barrier measures, A, τ, Δp, and p_b, evaluated in the quasiperiodic region of the pendulum's parameter space. The histograms are drawn in the same fashion as Figure 6.6. (a) GLE vs. the barrier amplitude, A; contains 1074 points. (b) GLE vs. the stochastic decay time, τ; contains 428 points. (c) GLE vs. the barrier width Δp; contains 869 points. (d) GLE vs. the barrier position, p_b; contains 865 points.

when one examines Figure 6.7, the PDF for the $D_0 \neq 0$ region of parameter space. GLE-1 never exceeds 1.5 bits/year, while the stochastic decay time, τ, has only half as many points as the other measures. In all four PDF's, the peak occurrence is at zero GLE-1. Obviously, this is due to the presence of numerous QP states for $D_0 \neq 0$. For negative GLE-1, $A \approx 1$ and $\Delta p \approx 0$ almost exclusively, indicating strong persistence barriers whenever there is periodic behavior.[4] Notice the chaotic points, especially the barrier width and amplitude. $\Delta p < 0.2$ yr. for most positive GLE-1. At the same time, A is fairly uniform, though $0.2 \leq A \leq 0.8$ mostly. The number of $A > 0.5$ also decreases as GLE-1 increases. Nevertheless, this was certainly not the case for the $D_0 = 0$ section of parameter space, where $A < 0.1$ for the bulk of chaotic states. The uniformity of the barrier amplitude and low values of barrier width for chaotic states is rather suggestive. It implies a higher occurrence of strong persistence barriers for QP-chaos than for other forms of chaos. It does not, alas, confirm such a connection. The PDF's simply do not cover enough of pendulum's parameter space.

6.2.4 Routes to Chaos and ILE Distributions

Let us now turn our attention to the use of the instantaneous Lyapunov exponents [ILE] defined in chapter 2. Recall from that chapter the utility of probability distribution functions of the ILE. In this section, I shall examine individual points in parameter space, including a QP-route to chaos, using both persistence maps and PDF's of the ILE spectrum. A two-dimensional, joint PDF between the phase of the year and the ILE may reveal a connection to persistence maps.

For each point in parameter space, let us integrate the pendulum equations and compute the ILE spectrum using the specifications of section C.1 in appendix C. Let us then display the following data: the persistence map, a segment from and the power spectrum of the pendulum momentum minus its cyclostationary mean,[5] the Poincaré Section, and the ILE in a special form. First, one sorts the ILE at each point in phase space [which corresponds directly to a point in time] into two sets, one containing the maximum ILE at that point, the other containing the minimum. Next, one sorts these two sets of maximum

[4] This is mode-locked, non-period-1 behavior.

[5] Recall from section C.1 that we must use the pendulum momentum, p, to compute the persistence map. The displayed timeseries and power spectrum are therefore of p.

ILE and minimum ILE according to the phase of the year. Since both the phase space and the ILE spectra are computed using time in units of years, with a fixed resolution of 120 points/year, this step is fairly easy. Finally, one assembles the set of ILE at each phase into a histogram. The resulting two-dimensional PDF displays the [time] phase-dependence of the maximally and minimally growing directions in [dynamical] phase-space, independent of location in phase-space.

On a final note, in this work, I shall construct the histograms using 75 bins divided over the full range of each set of ILE. I will also discuss only the maximum ILE. There are only three exponents in the spectra, either global or instantaneous, one of which is associated with the $\tilde{\tau}$-direction. This exponent is always zero. The sum of the remaining two exponents, in the case of both local and global varieties, must equal the dissipation of the system, a constant. Hence describing both exponents is redundant.

We shall examine six different points in parameter space. Three are a subset of the bifurcation-map in Figures C.17-C.18. The other three are a QP-route to chaos from Figures C.5-C.6. The parameter values for all six points, their maximum global Lyapunov exponents, their winding numbers, and the values of R^2 for their persistence maps, all labeled by figure number, are in Table 6.1. I have stated the parameters out to their full precision, since truncating the parameters will change the behavior of the system.

Figures 6.8-6.10 display the pendulum at a point in parameter space taken from Figures C.17-C.18. Note that these figures are a version of the bifurcation diagram from Gwinn and Westervelt (GW86). The system has a winding number of $\frac{1}{2}$, yet the Poincaré Section seems to indicate a period-4 oscillation. The power spectrum, however, indicates a more complicated periodicity, with numerous harmonics of both a period-2 and period-4 oscillation. Many of these harmonics are intra-annual in nature, a fact reflected in the persistence map. Note that the persistence contains a [time] phase-dependence that does not match the barrier model. However, the regions of high negative correlation are small enough that the barrier model still fits well, as indicated by the nonlinear correlation coefficient, $R^2 \approx 0.99$. The phase-dependent probability distribution functions of the maximum and minimum ILE are not terribly revealing. The peak of the PDF for the maximum ILE [henceforth denoted $\mathrm{ILE}_{\max}$] falls below 0 yr^{-1} and away from the persistence barrier position. Note that the lines indicating the barrier position $\pm$ the barrier width fall atop each other.

Table 6.1: Table of parameter values for six points in the pendulum's parameter space, accompanied by their maximum global Lyapunov exponents [GLE], the values of R^2 for their persistence maps, and their winding numbers [either W_l or W_r, whichever is appropriate]. Each of the points is identified by the figure numbers displaying its phase space.

	β	γ	D_0	D_1
Fig. 6.8-6.10	4.712388980385	88.82643960980	0	96.06117224770226
Fig. 6.11-6.13	4.712388980385	88.82643960980	0	96.463101838696829
Fig. 6.14-6.16	4.712388980385	88.82643960980	0	96.664066634194114
Fig. 6.17-6.19	3.737924059107871	12.74484039397	18.38315778425752	12.74484039397
Fig. 6.20-6.22	3.737924059107871	12.74484039397	16.91495217087256	12.74484039397
Fig. 6.23-6.25	3.737924059107871	12.74484039397	12.51033533071768	12.74484039397

	max. GLE	W	R^2
Fig. 6.8-6.10	-0.241	0.5	0.9991
Fig. 6.11-6.13	0.497	0.302	0.9991
Fig. 6.14-6.16	-3.399	0.333	0.9996
Fig. 6.17-6.19	0.0	0.713	0.9636
Fig. 6.20-6.22	-0.076	0.625	0.9838
Fig. 6.23-6.25	0.601	0.207	0.9989

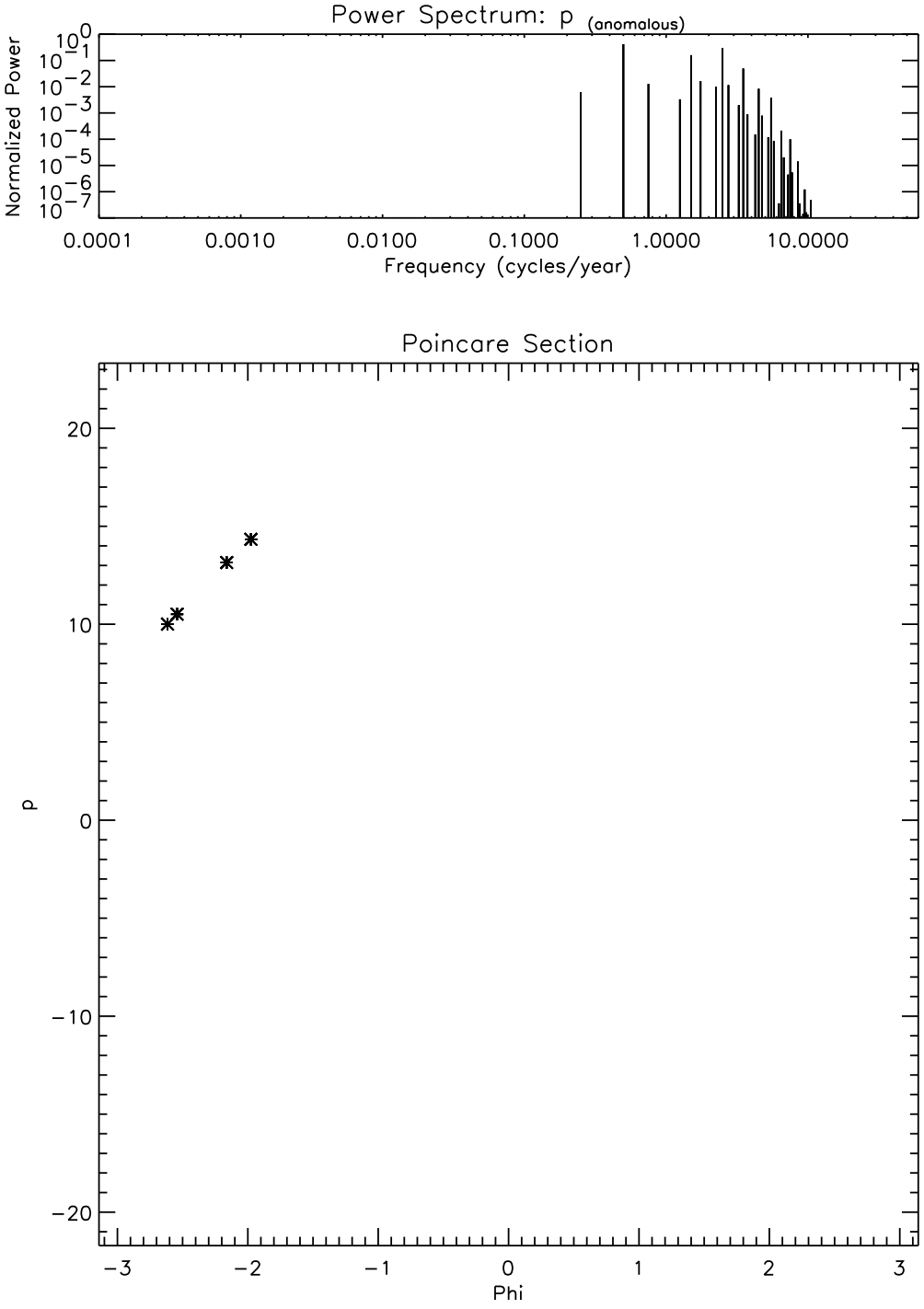

Figure 6.8: Power spectrum of the pendulum momentum, minus its cyclostationary mean, and the pendulum's Poincaré Section, using the parameters $\beta \approx 4.7$, $\gamma \approx 88.8$, $D_0 = 0$, and $D_1 \approx 96.06$.

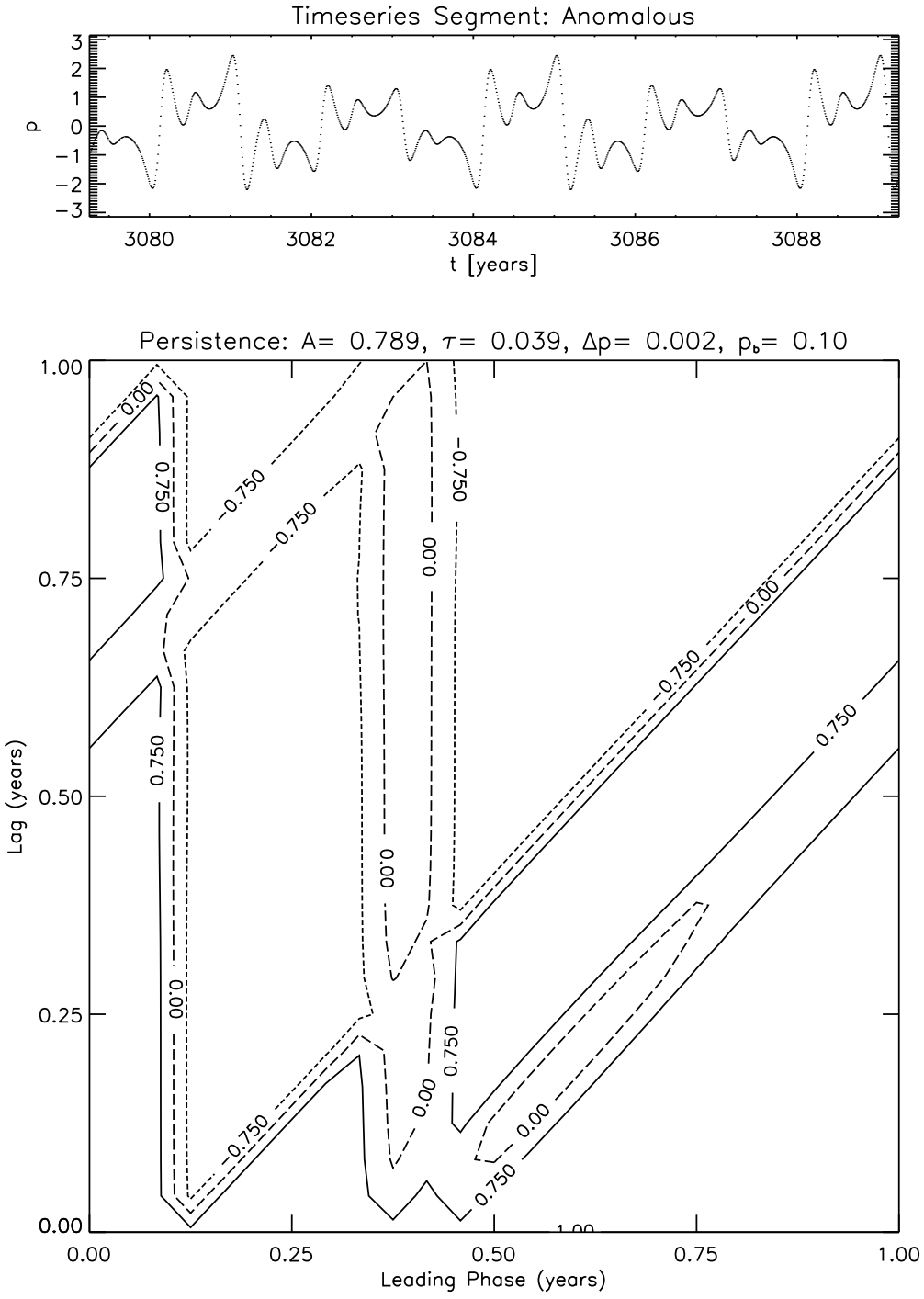

Figure 6.9: Persistence map and timeseries segment for the damped forced pendulum with parameters $\beta \approx 4.7$, $\gamma \approx 88.8$, $D_0 = 0$, and $D_1 \approx 96.06$. The timeseries segment is from the pendulum momentum minus its cyclostationary mean. The persistence map is computed from the pendulum momentum. Contour levels are 1.0, ± 0.75, and 0.0.

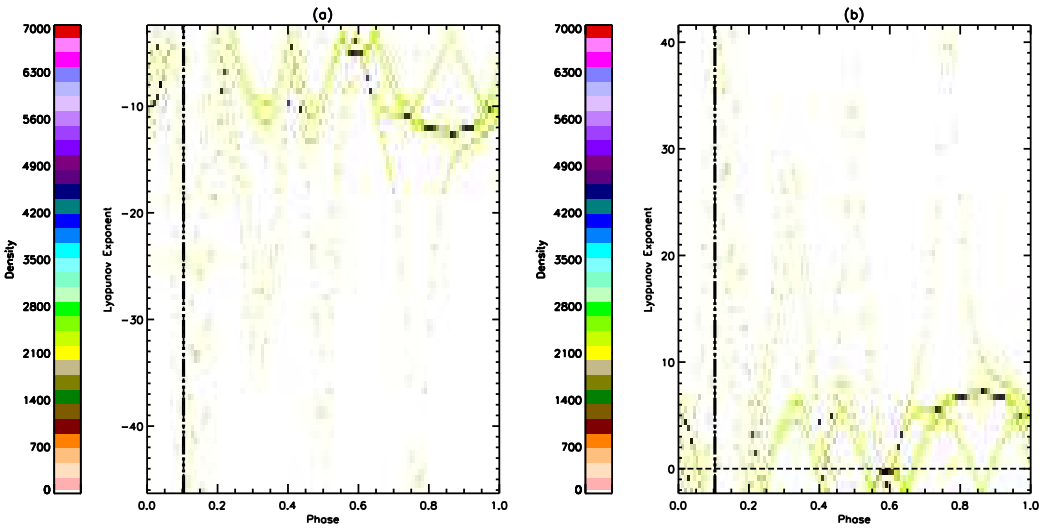

Figure 6.10: Phase-dependent probability distribution functions of the maximum and minimum instantaneous Lyapunov exponent [ILE] for the damped forced pendulum with parameters $\beta \approx 4.7$, $\gamma \approx 88.8$, $D_0 = 0$, and $D_1 \approx 96.06$. [See the text for a description of how these PDF's are generated.] The dashed line indicates a value of 0 bits/year for the ILE. The dot-dashed line indicates the barrier position, p_b, while the two "triple-dot-dashed" lines indicate $p_b \pm \Delta p$. [All three lines coincide for this plot.]

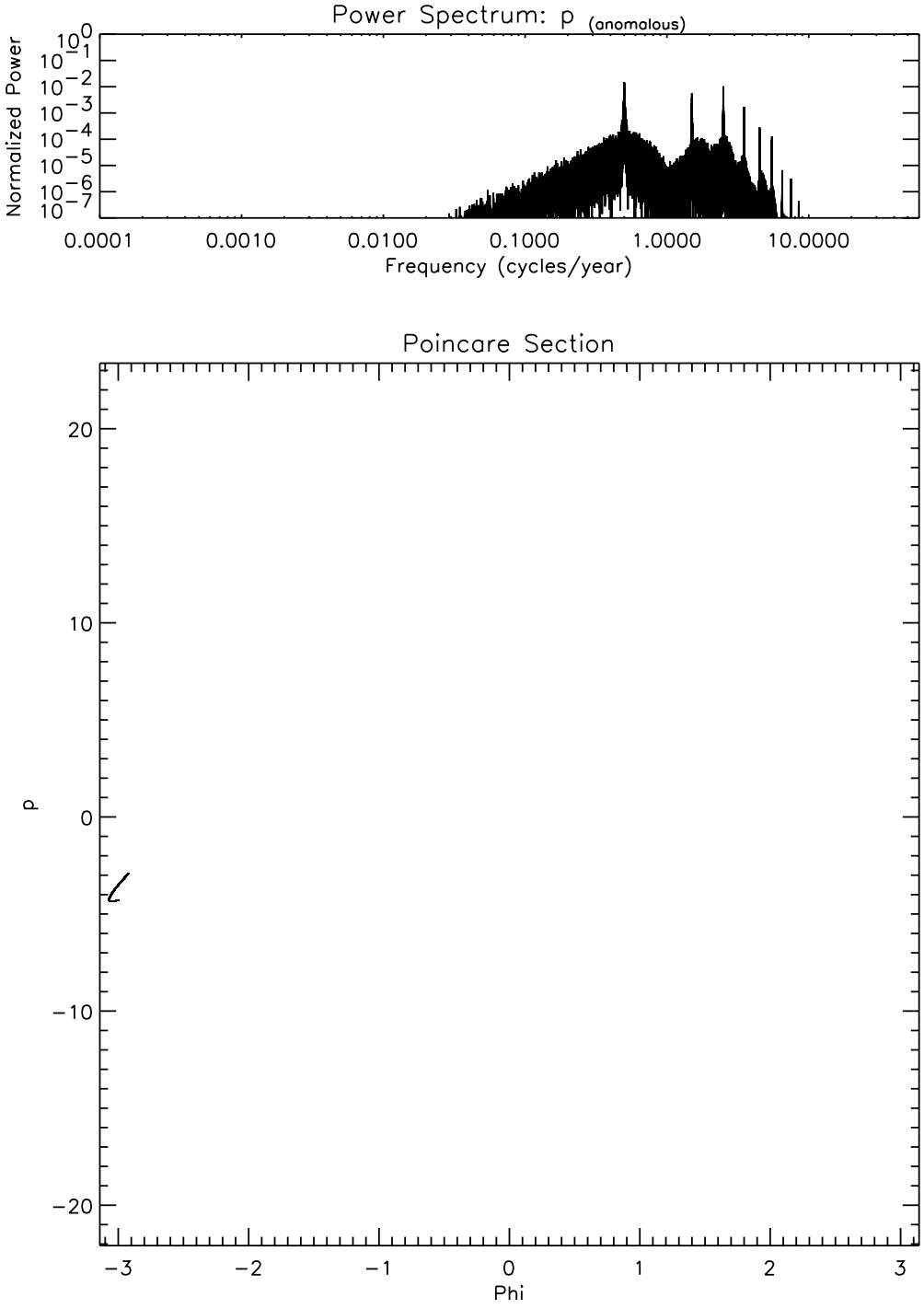

Figure 6.11: Power spectrum of the pendulum momentum, minus its cyclostationary mean, and the pendulum's Poincaré Section, using the parameters $\beta \approx 4.7$, $\gamma \approx 88.8$, $D_0 = 0$, and $D_1 \approx 96.46$.

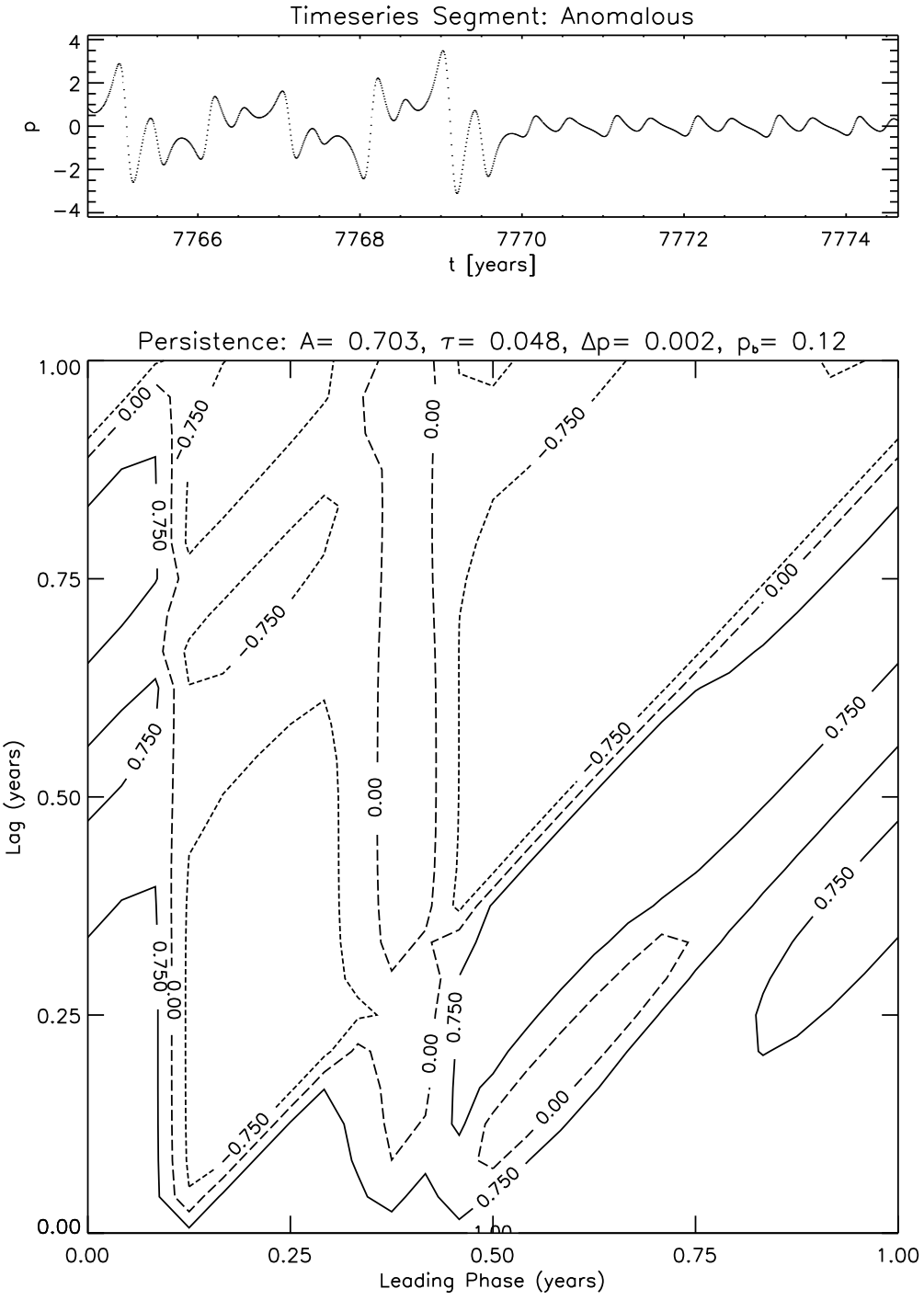

Figure 6.12: Persistence map and timeseries segment for the damped forced pendulum with parameters $\beta \approx 4.7$, $\gamma \approx 88.8$, $D_0 = 0$, and $D_1 \approx 96.46$. The timeseries segment is of the pendulum momentum minus its cyclostationary mean. The persistence map is computed from the pendulum momentum. Contour levels are as in Figure 6.9.

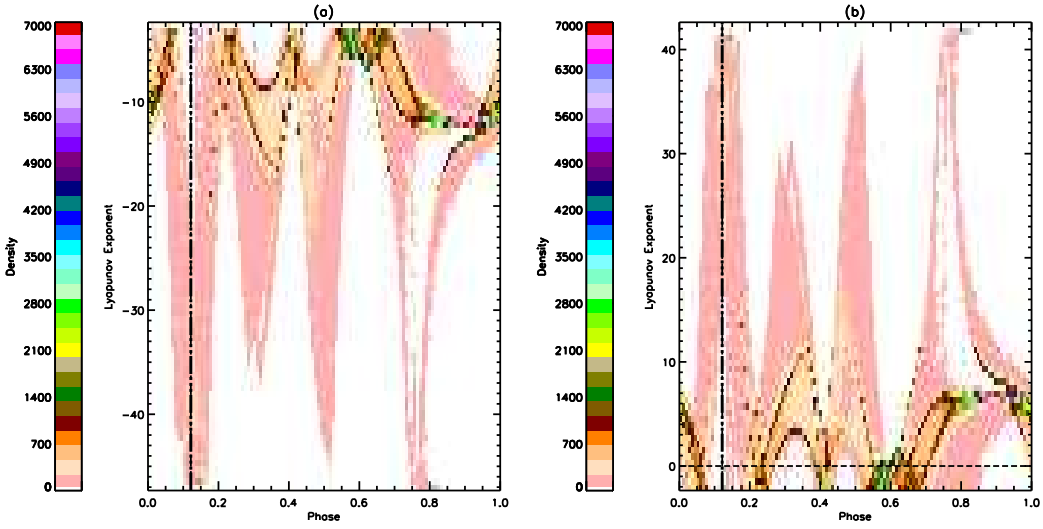

Figure 6.13: Phase-dependent probability distribution functions of the maximum and minimum ILE for the damped forced pendulum with parameters $\beta \approx 4.7$, $\gamma \approx 88.8$, $D_0 = 0$, and $D_1 \approx 96.46$, plotted in the same manner as Figure 6.10. (a) PDF of the minimum ILE. (b) PDF of the maximum ILE.

Figures 6.11-6.13 continue the examination of this region of parameter space with a chaotic point, whose GLE-1 [the maximum global Lyapunov exponent] equals 0.497 bits/yr. Note the folded shape of the Poincaré Section, indicating the onset of chaos. The power spectrum displays a strong period-2 harmonic, with numerous intra-annual harmonics. As for the previous parameter set, the persistence map contains a "non-barrier" feature. Once again, only a small area of this feature deviates from the model barrier, producing a nonlinear correlation coefficient of $R^2 \approx 0.99$. The PDF of the maximum ILE has grown a bit more interesting, though the lines indicating the p_b and $p_b \pm \Delta p$ still fall atop each other. The barrier position does coincide with an overall increase in $\mathrm{ILE}_{\max}$ to values greater than 10 yr^{-1}. The peak of the PDF, however, still falls at a different phase [of the year] and near $\mathrm{ILE}_{\max} = 0$.

The last parameter space point from this series is shown in Figures 6.14-6.16. This is a period-3 oscillation, as indicated by the winding number [Table 6.1], the Poincaré Section, and the power spectrum. However, there are again strong intra-annual harmonics, reflected as well in the persistence map. Note that $\Delta p > 1$ year for this persistence map, indicating a fit to a linearly decaying model barrier. The phase-dependent feature actually resembles a "double barrier." Notice how the PDF of $\mathrm{ILE}_{\max}$ reflects this behavior, with one peak in $\mathrm{ILE}_{\max}$ near the position of the "first" barrier and another, equally strong peak near the position of the "second" barrier. A high-density ridge, moving in the direction of increasing $\mathrm{ILE}_{\max}$, precedes both of these.

Figures 6.8-6.16 examined three neighboring points in parameter space, each differing slightly in D_1 [Table 6.1]. The chosen parameters are points along a bifurcation route to chaos described by Gwinn and Westervelt (GW86). All of the chosen parameters produce a pendulum with strong intra-annual behavior. Hence, the persistence maps contained [time] phase-dependent behavior in a form other than a persistence barrier. Most interesting, though, was the PDF of $\mathrm{ILE}_{\max}$ for one of these cases. We found a relationship between increasing $\mathrm{ILE}_{\max}$ and position of barrier-like behavior. It would indeed be exciting if this connection bore out for a state of the pendulum with a true persistence barrier.

The next set of parameters, shown in Figures 6.17-6.25, examine a quasiperiodic [QP] route to chaos, using points from Figures C.5-C.6 of appendix C. The first parameter set exhibits QP-behavior, as seen in Figures 6.17-6.19. The Poincaré Section forms a closed loop

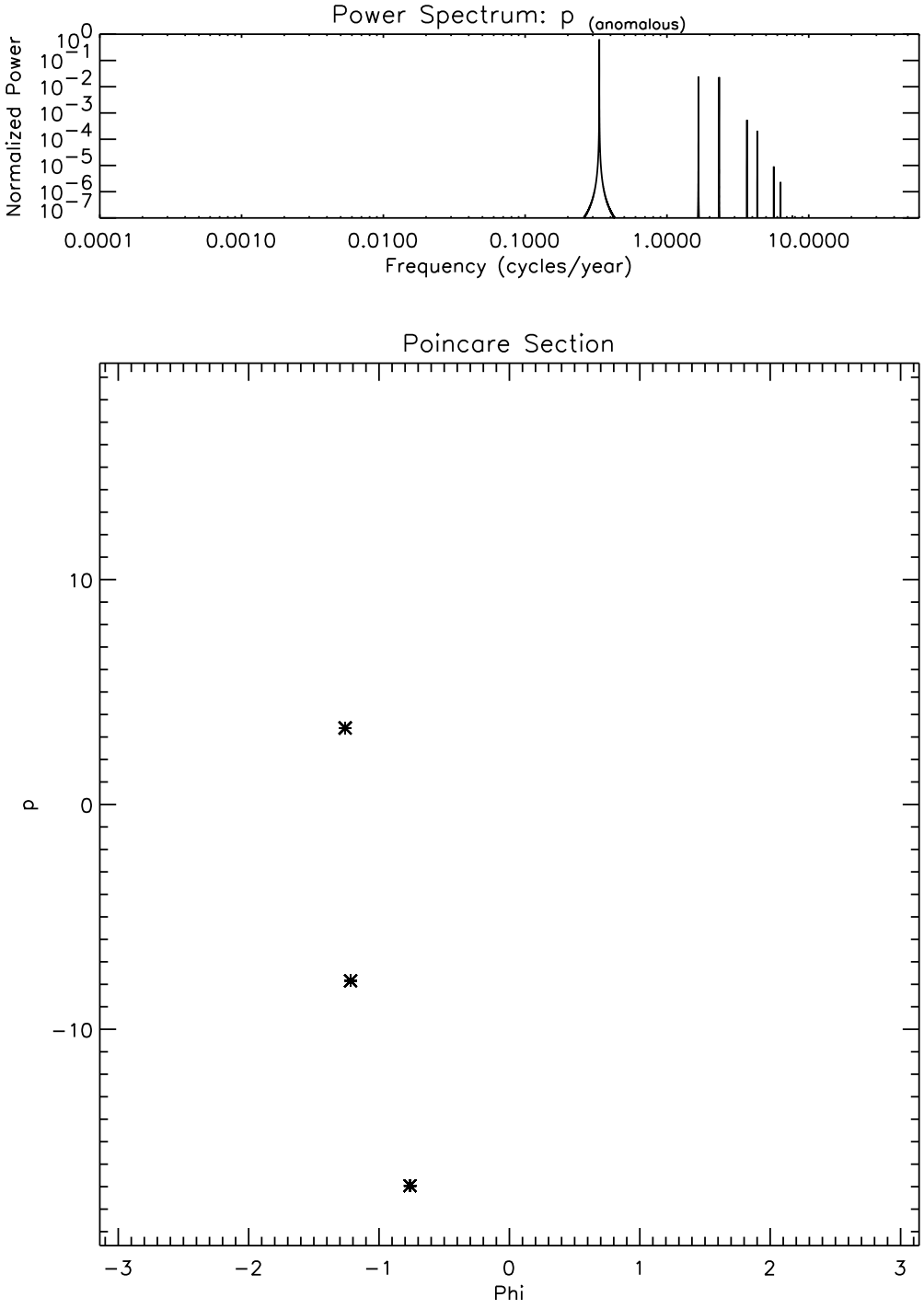

Figure 6.14: Power spectrum of the pendulum momentum, minus its cyclostationary mean, and the pendulum's Poincaré Section, using the parameters $\beta \approx 4.7$, $\gamma \approx 88.8$, $D_0 = 0$, and $D_1 \approx 96.66$.

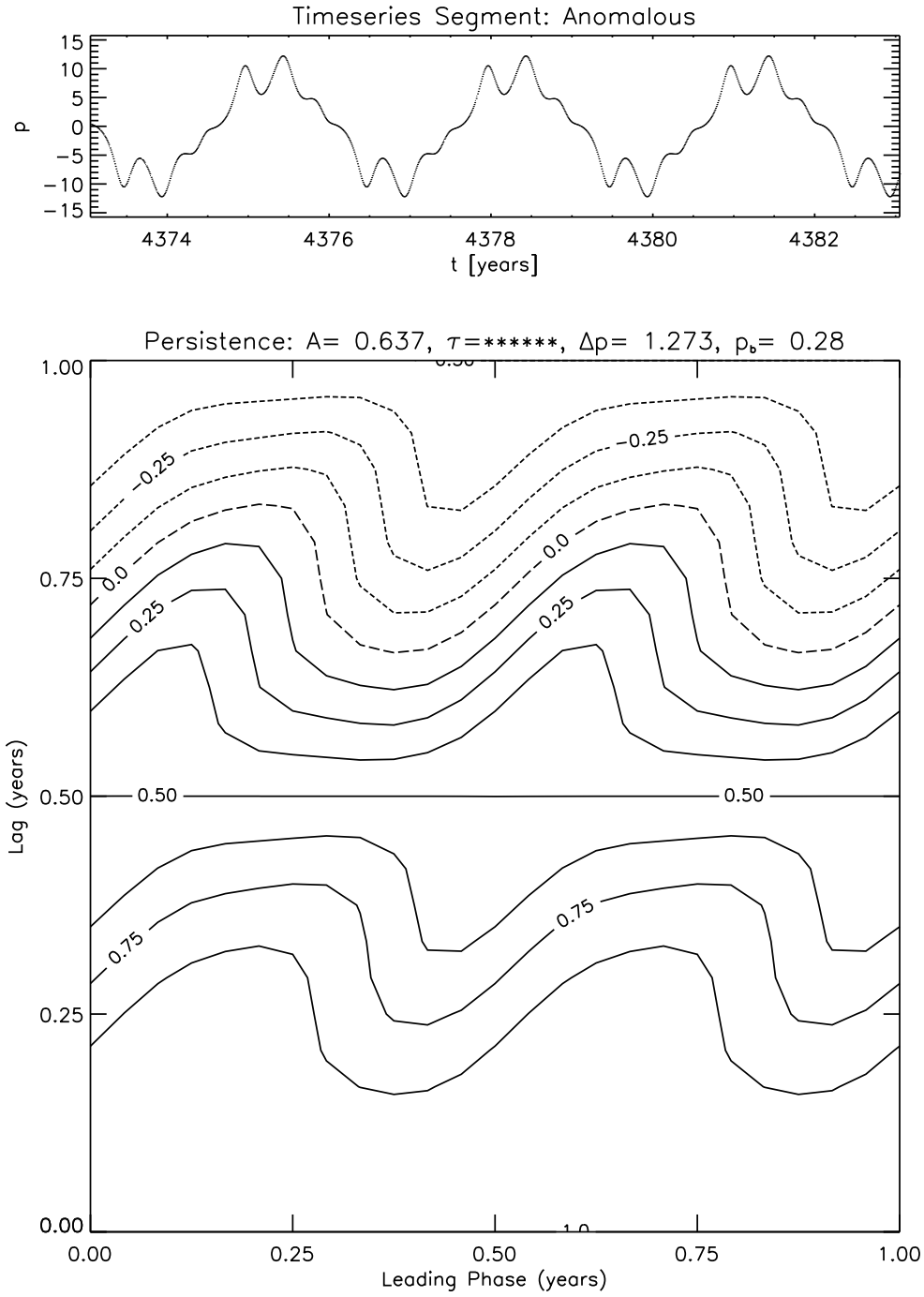

Figure 6.15: Persistence map and timeseries segment for the damped forced pendulum with parameters $\beta \approx 4.7$, $\gamma \approx 88.8$, $D_0 = 0$, and $D_1 \approx 96.66$. The timeseries segment is of the pendulum momentum minus its cyclostationary mean. The persistence map is computed from the pendulum momentum. Note that the stochastic decay time, τ, for this map is $\mathcal{O}\left(10^6\right)$ yrs. Contours in this and all subsequent persistence maps are in steps of $\frac{1}{8}$.

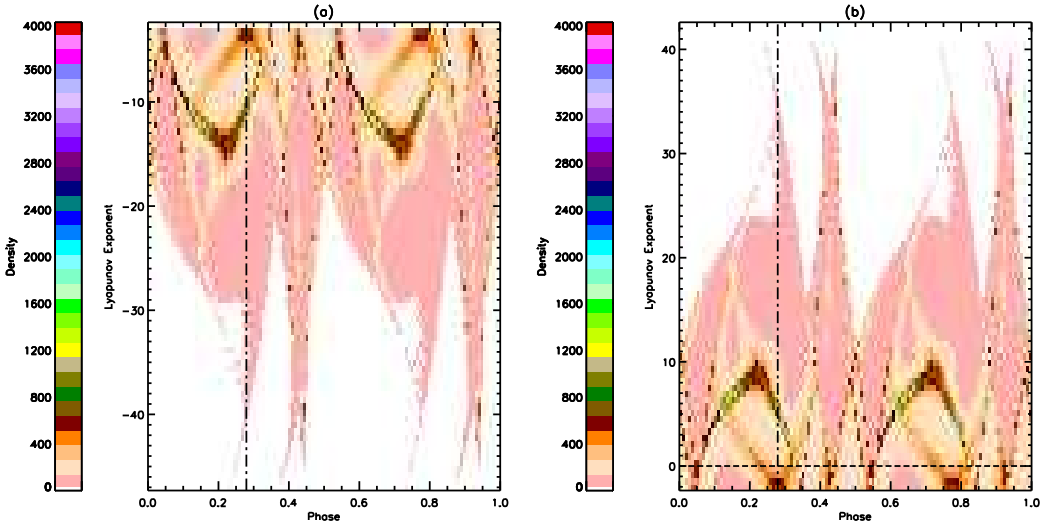

Figure 6.16: Phase-dependent probability distribution functions of the maximum and minimum ILE for the damped forced pendulum with parameters $\beta \approx 4.7$, $\gamma \approx 88.8$, $D_0 = 0$, and $D_1 \approx 96.66$, plotted in the same manner as Figure 6.10. (a) PDF of the minimum ILE. (b) PDF of the maximum ILE.

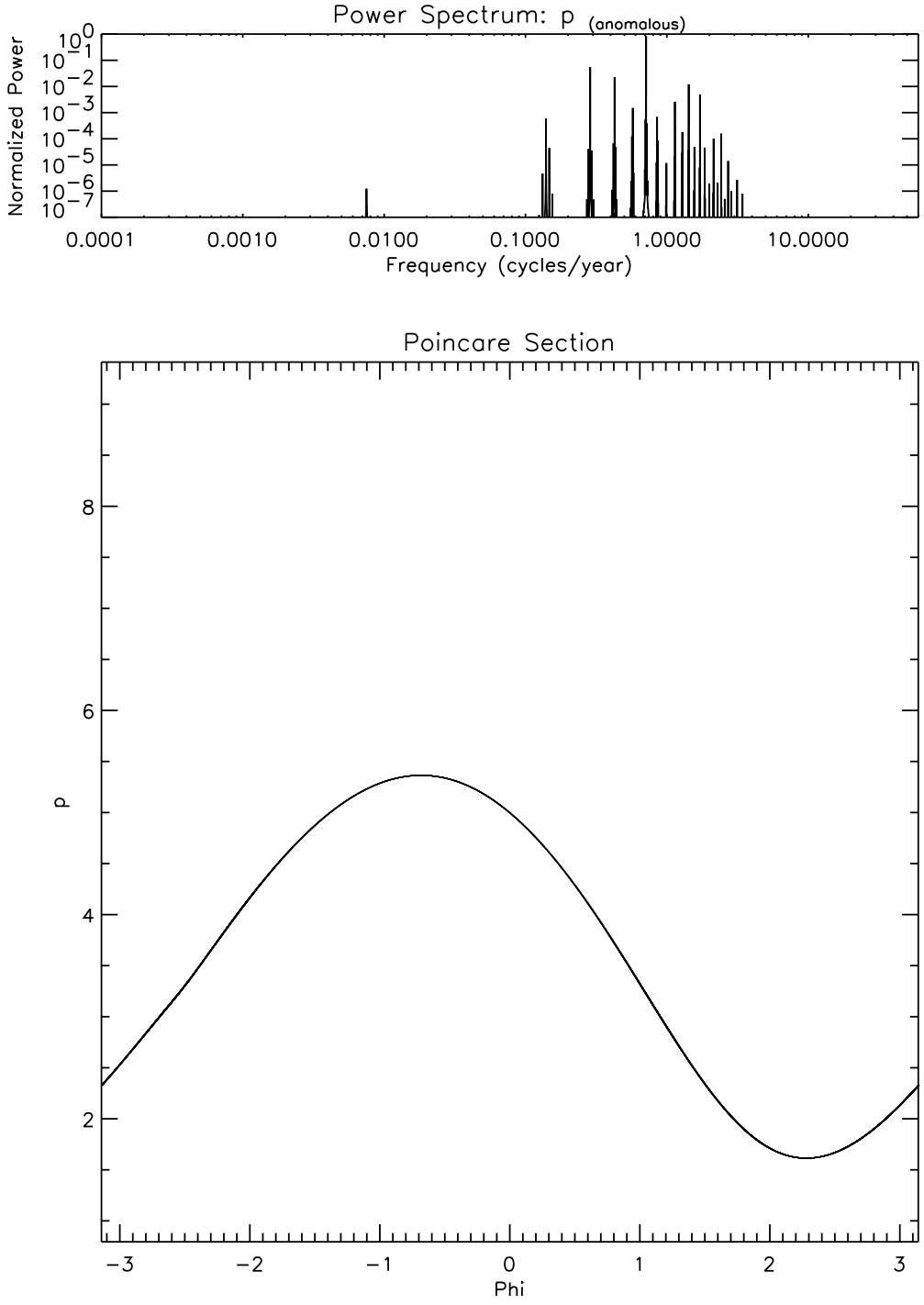

Figure 6.17: Power spectrum of the pendulum momentum, minus its cyclostationary mean, and the pendulum's Poincaré Section, using the parameters $\beta \approx 3.7$, $\gamma = D_1 \approx 12.7$, and $D_0 \approx 18.3$.

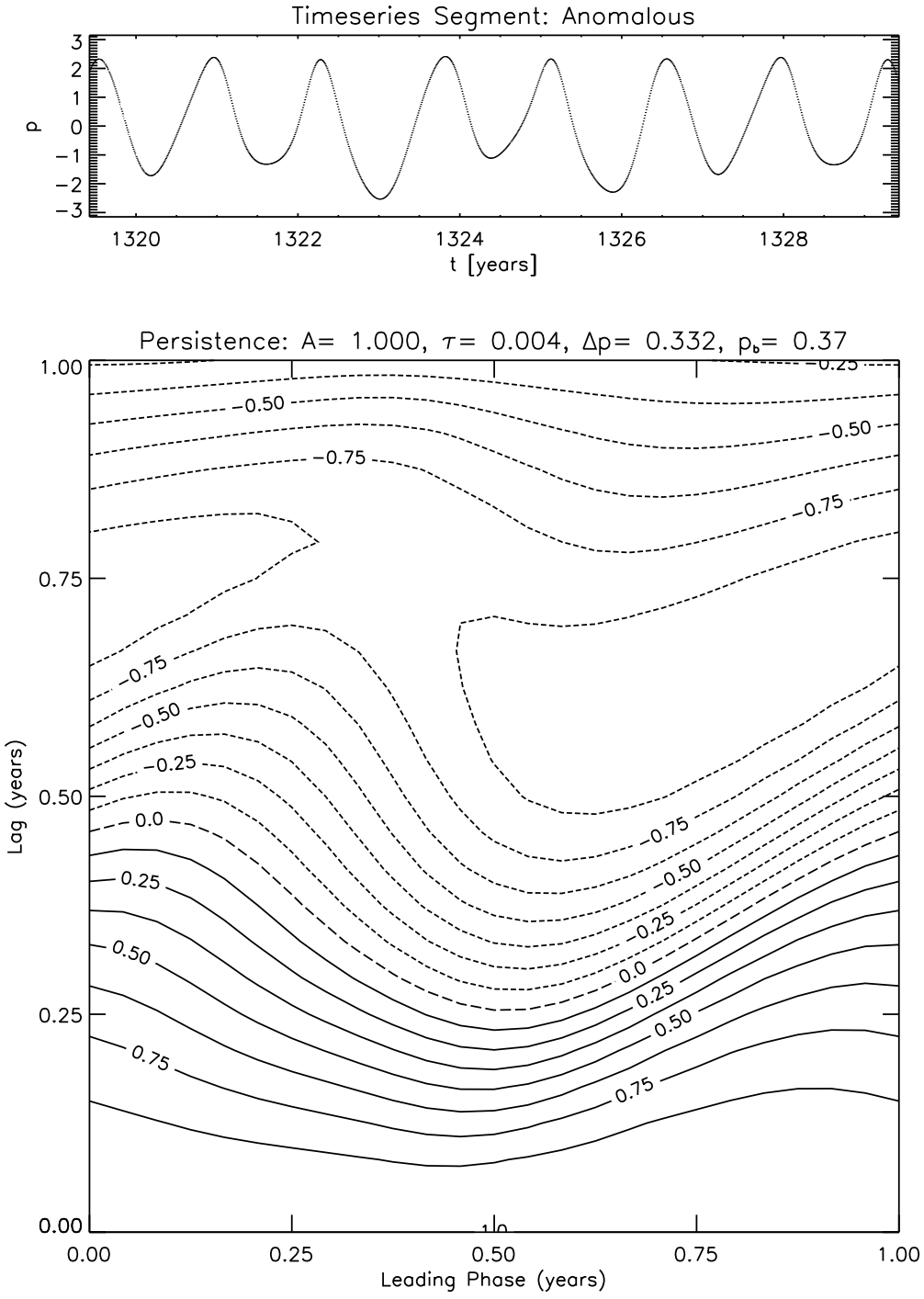

Figure 6.18: Persistence map and timeseries segment for the damped forced pendulum with parameters $\beta \approx 3.7$, $\gamma = D_1 \approx 12.7$, and $D_0 \approx 18.3$. The timeseries segment is of the pendulum momentum minus its cyclostationary mean. The persistence map is computed from the pendulum momentum.

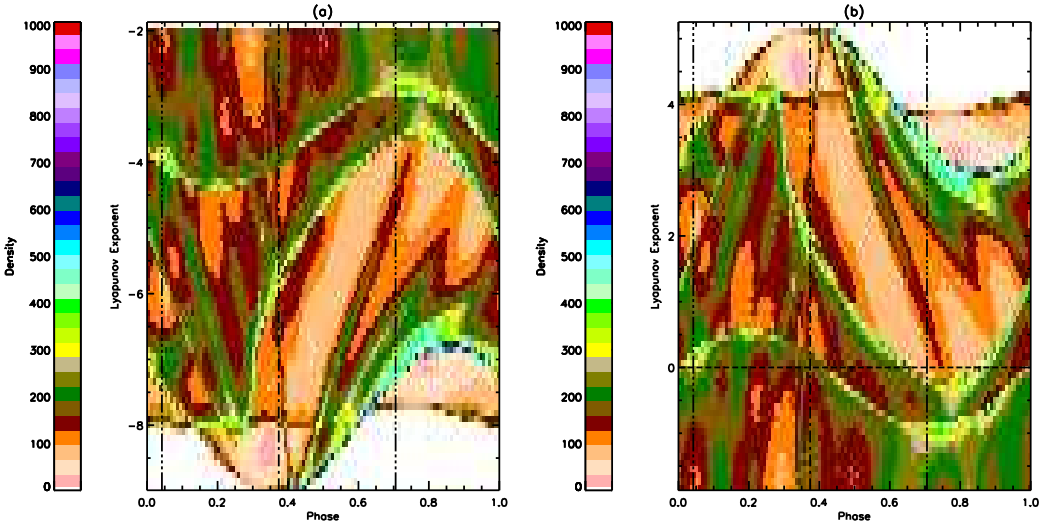

Figure 6.19: Phase-dependent probability distribution functions of the maximum and minimum ILE for the damped forced pendulum with parameters $\beta \approx 3.7$, $\gamma = D_1 \approx 12.7$, and $D_0 \approx 18.3$, plotted in the same manner as Figure 6.10. As in Figure 6.10, the dot-dashed line indicates the barrier position, p_b, while the two "triple-dot-dashed" lines indicate $p_b \pm \Delta p$. (a) PDF of the minimum ILE. (b) PDF of the maximum ILE.

with no folds, while the winding number is zero [Table 6.1]. The persistence map contains a strong but broad barrier, with a width of roughly $\frac{1}{3}$ yr. Note that, while this feature visually resembles a barrier, it contains strong negative correlation at high lag, causing a lowered value of $R^2 = 0.96$. The PDF's of the ILE are most intriguing, however. The value of $\mathrm{ILE}_{\max}$ peaks right at the barrier position, p_b. For phases between $p_b - \Delta p$ and p_b, $\mathrm{ILE}_{\max}$ has a fairly uniform density distribution. Between p_b and $p_b + \Delta p$ there is a high-density ridge at the upper bound of the value of $\mathrm{ILE}_{\max}$. This ridge tends towards decreasing $\mathrm{ILE}_{\max}$ as a function of phase. Described more generally, the direction of maximal error growth takes on a wide range of ever larger growth rates just before the barrier position. After the barrier position, direction of maximal error growth takes on a definite, decreasing rate of error growth.

The next step along this QP-route to chaos is a mode-locked period-8 oscillation [Figures 6.20-6.22]. The winding number is actually $\frac{5}{8}$, but the Poincaré Section contains only 8 points. Notice how those point fall on the same curve formed by the QP-state! The persistence map also resembles that of the QP-state, with $A \sim 1$ and $\Delta p \sim 0.3$. The barrier in this map, however, more closely resembles the barrier model. Unfortunately, the PDF of $\mathrm{ILE}_{\max}$ contains a dizzying array of criss-crossing ridges, rendering interpretation such as that of the previous paragraph difficult. Such an interpretation is not really necessary; one should expect most periodic timeseries with a period-8 oscillation to exhibit some form of persistence barrier.[6]

The final stop on this QP-route to chaos is, of course, a chaotic pendulum. Figures 6.23-6.25 display the results. Not only is the largest GLE 0.6 bits/year, the Poincaré Section contains a fold, a telltale indicator of chaotic behavior. Note, too, the noisiness and flatness of the power spectrum. It is therefore surprising to find a moderately strong, broad barrier in the persistence. Here, as in the other points of this QP-sequence, $\Delta p \sim 0.3$. As for the QP-state, the PDF of $\mathrm{ILE}_{\max}$ displays a connection to persistence barriers. This time, the ridges of high density form an arc, peaked at p_b. Prior to the barrier position, the ridge is broader, narrowing and increasing in $\mathrm{ILE}_{\max}$ value towards p_b. On the other side, the narrower ridge shows decreasing $\mathrm{ILE}_{\max}$ value. The ridge begins to spread out after

[6]The reasons for this come easily from chapter 4. The lone exception to this statement is a timeseries of a single cosine or sine with a period of 8.

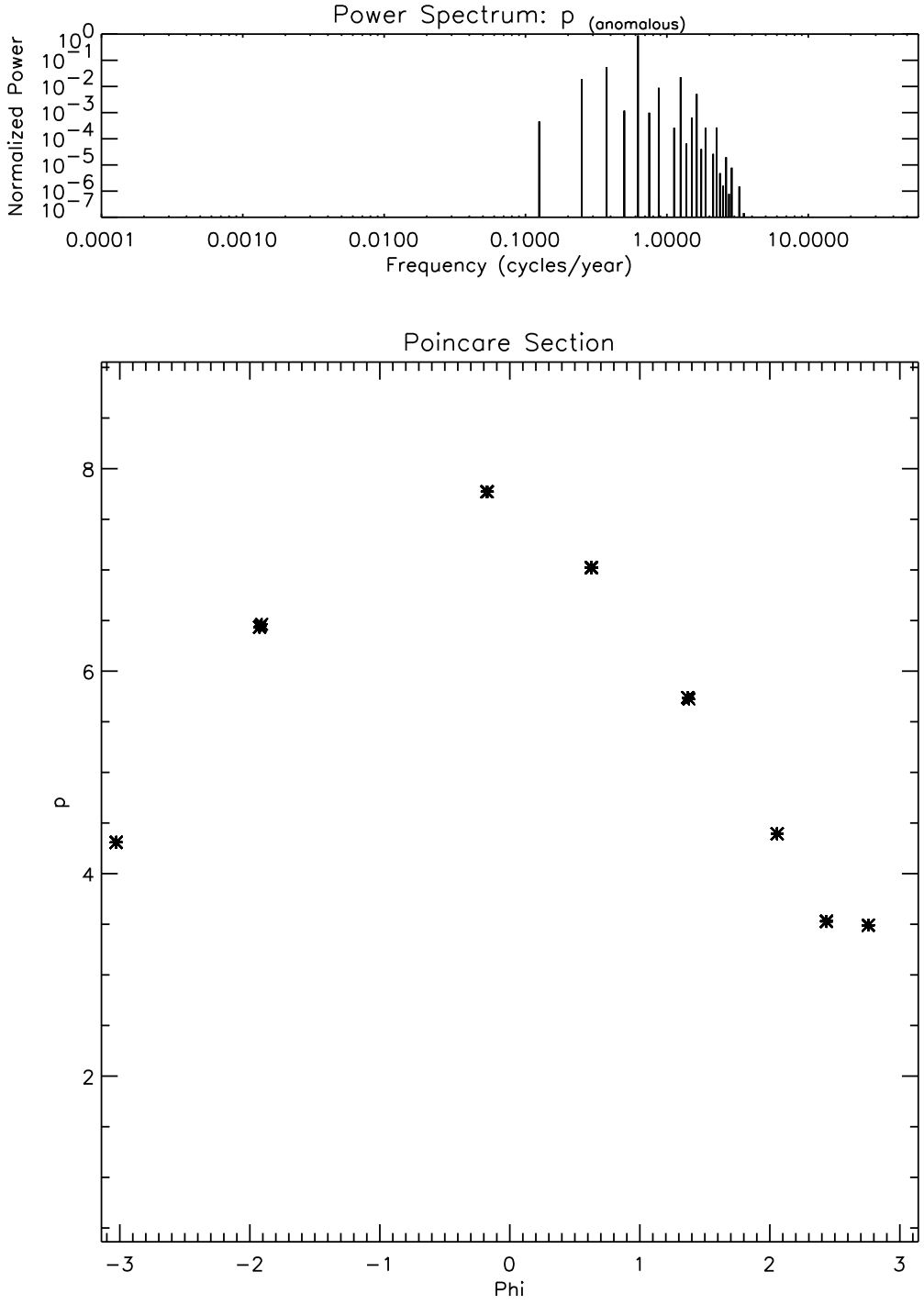

Figure 6.20: Power spectrum of the pendulum momentum, minus its cyclostationary mean, and the pendulum's Poincaré Section, using the parameters $\beta \approx 3.7$, $\gamma = D_1 \approx 12.7$, and $D_0 \approx 16.9$.

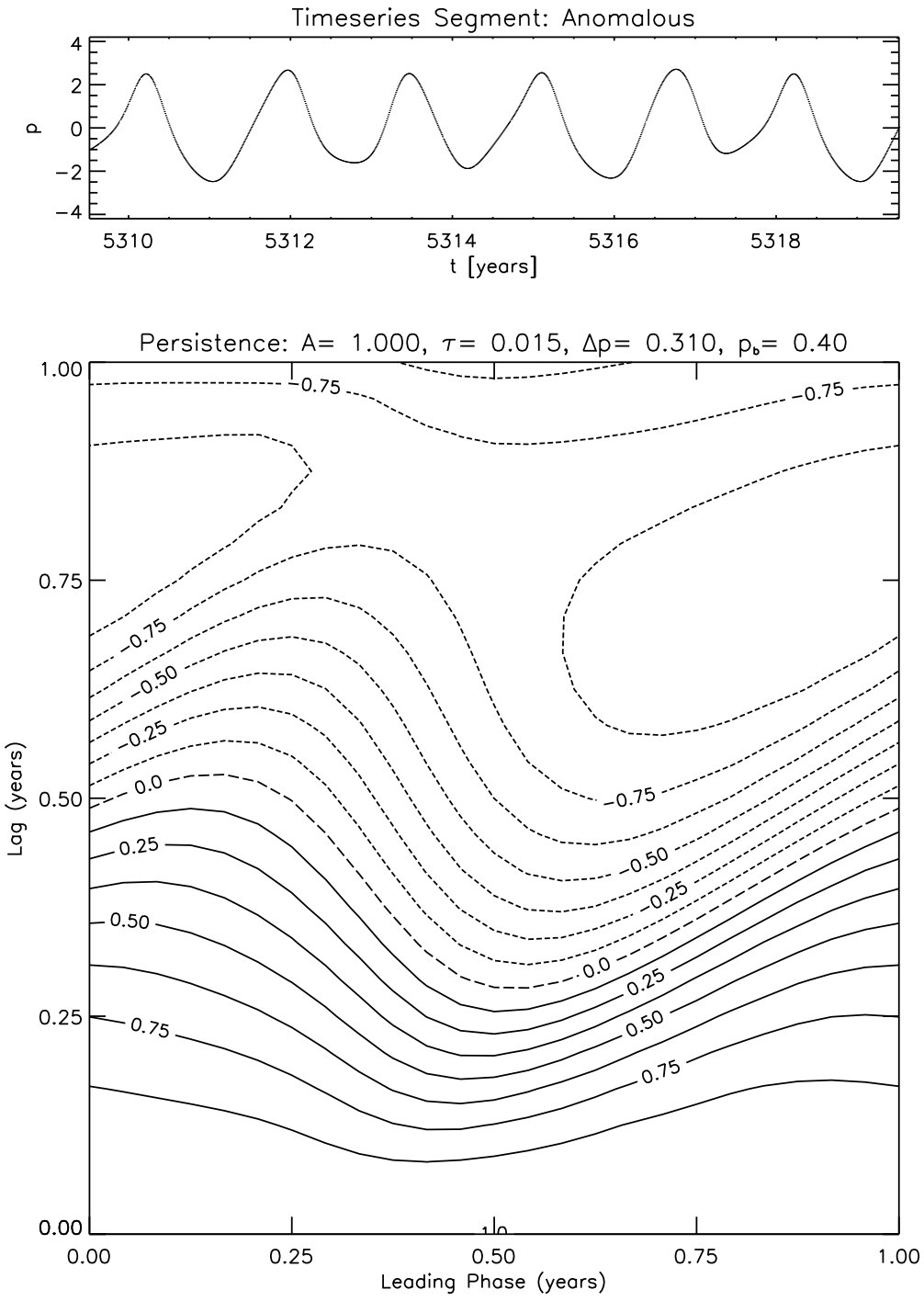

Figure 6.21: Persistence map and timeseries segment for the damped forced pendulum with parameters $\beta \approx 3.7$, $\gamma = D_1 \approx 12.7$, and $D_0 \approx 16.9$. The timeseries segment is of the pendulum momentum minus its cyclostationary mean. The persistence map is computed from the pendulum momentum.

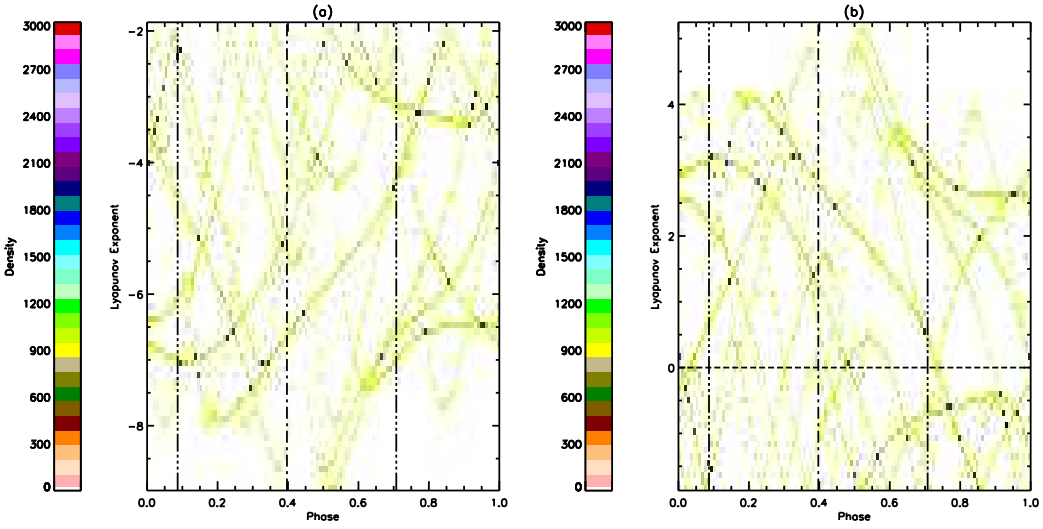

Figure 6.22: Phase-dependent probability distribution functions of the maximum and minimum ILE for the damped forced pendulum with parameters $\beta \approx 3.7$, $\gamma = D_1 \approx 12.7$, and $D_0 \approx 16.9$, plotted in the same manner as Figure 6.19. (a) PDF of the minimum ILE. (b) PDF of the maximum ILE.

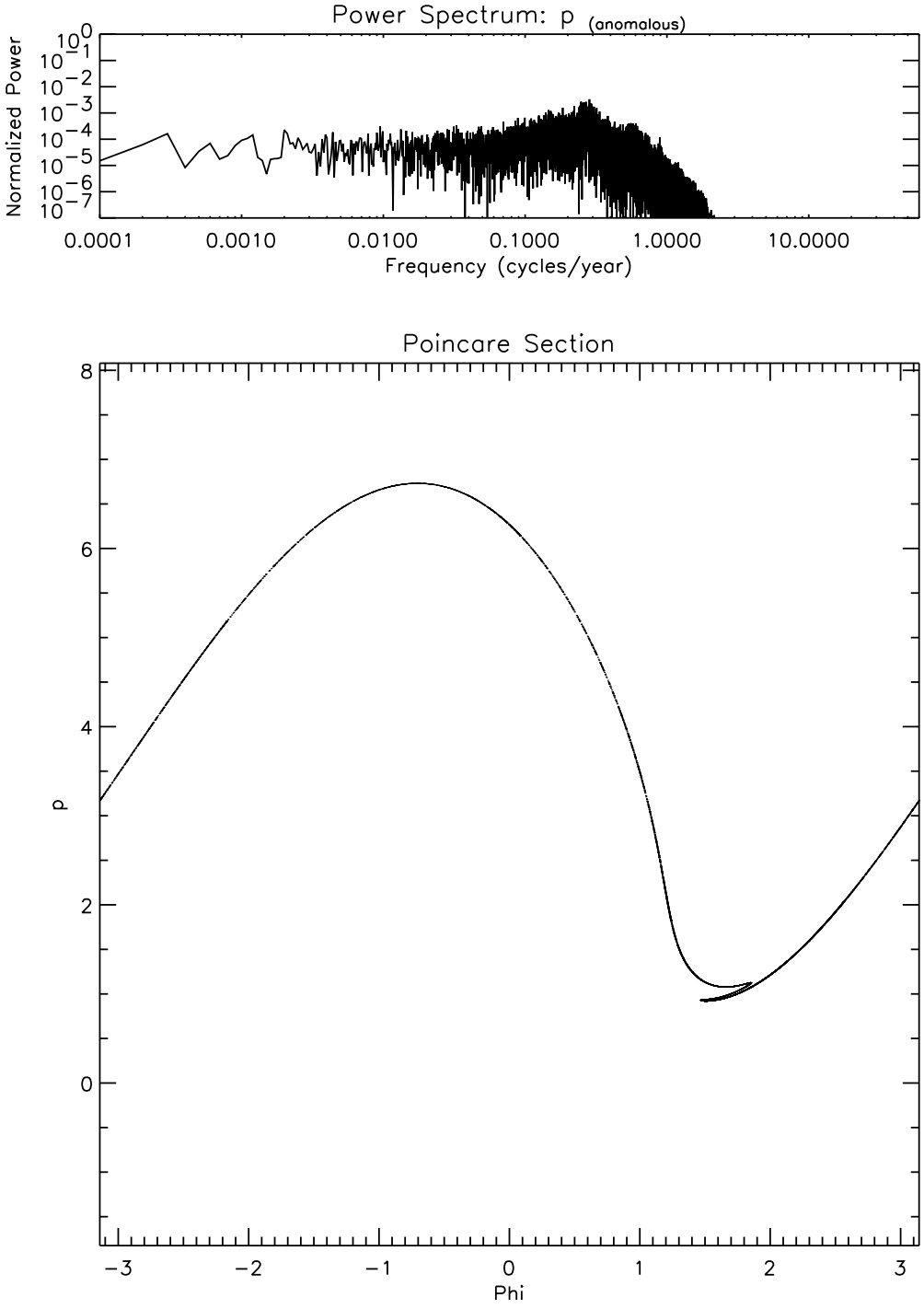

Figure 6.23: Power spectrum of the pendulum momentum, minus its cyclostationary mean, and the pendulum's Poincaré Section, using the parameters $\beta \approx 3.7$, $\gamma = D_1 \approx 12.7$, and $D_0 \approx 12.5$.

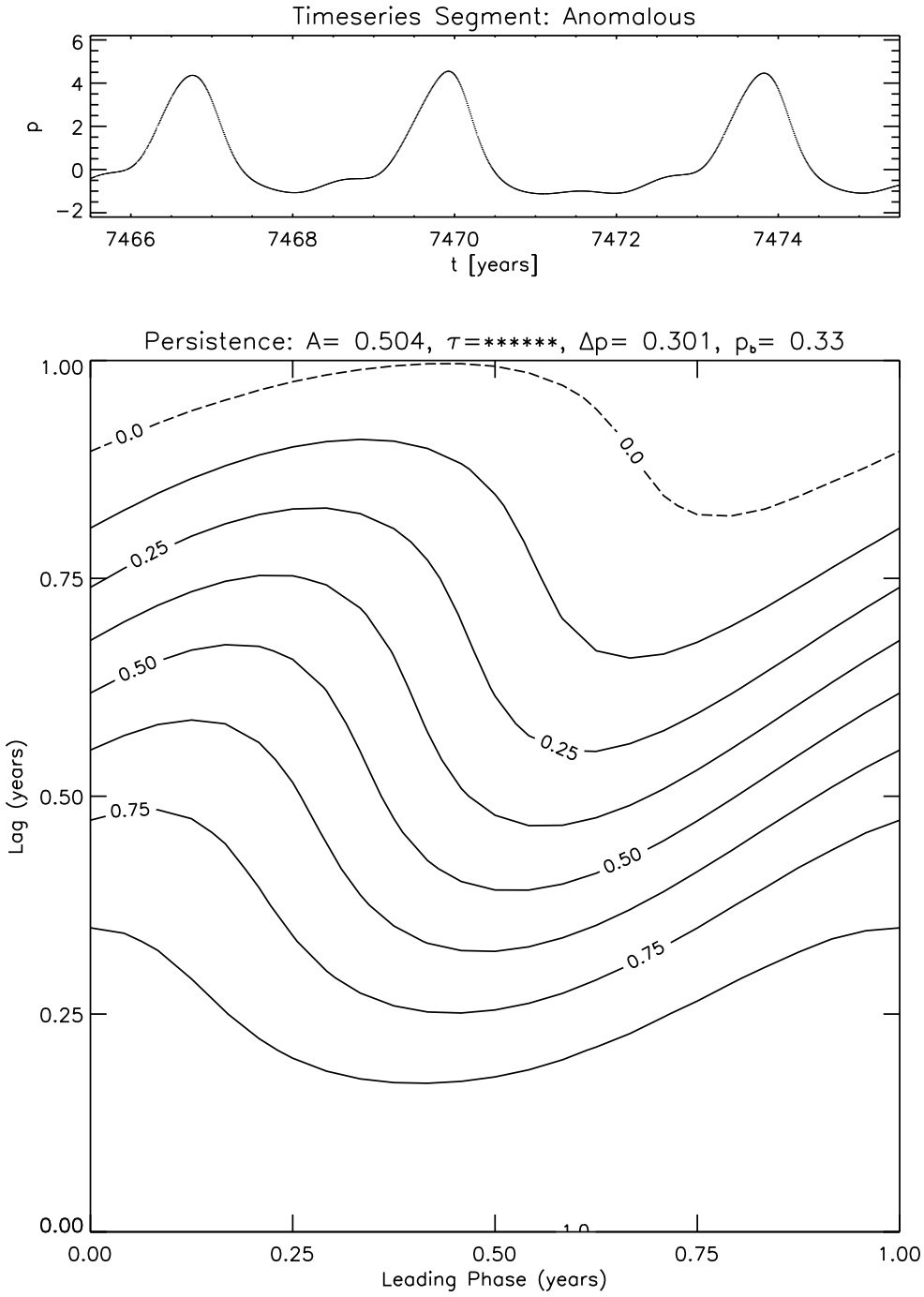

Figure 6.24: Persistence map and timeseries segment for the damped forced pendulum with parameters $\beta \approx 3.7$, $\gamma = D_1 \approx 12.7$, and $D_0 \approx 12.5$. The timeseries segment is of the pendulum momentum minus its cyclostationary mean. The persistence map is computed from the pendulum momentum. Note that the stochastic decay time, τ, for this map is $\mathcal{O}\left(10^{10}\right)$ yrs.

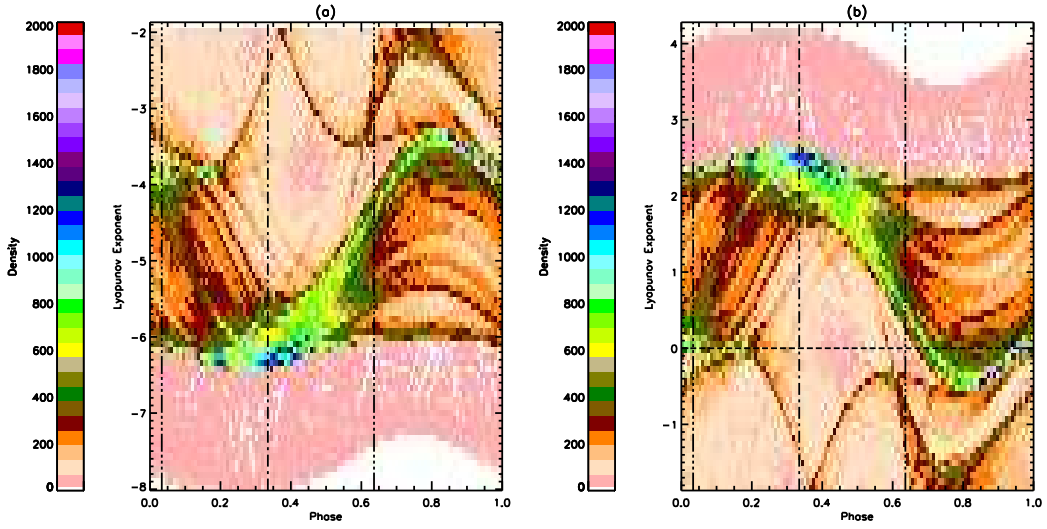

Figure 6.25: Phase-dependent probability distribution functions of the maximum and minimum ILE for the damped forced pendulum with parameters $\beta \approx 3.7$, $\gamma = D_1 \approx 12.7$, and $D_0 \approx 12.5$, plotted in the same manner as Figure 6.19. (a) PDF of the minimum ILE. (b) PDF of the maximum ILE.

$p_b + \Delta p$. Here, in a chaotic mode, there is again this change in the error growth behavior of the system, a change centered on the barrier position.

Figures 6.17-6.25, examined a quasiperiodic [QP] route to chaos in the damped forced pendulum by changing D_0. All points along this route exhibit a wide but strong persistence barrier. The most important result, however, is a connection between the persistence barrier and "high-density ridges" in the phase-dependent probability distributions of the maximal ILE. The barrier position, p_b, indicates the phase at which the timeseries loses correlation the most rapidly. Prior to p_b, the rate of maximal error growth, $\mathrm{ILE}_{\max}$, has a broad range of values. That range both narrows and increases, with the strongest error growth occurring at p_b. Past p_b, the maximal error growth rate maintains its range but decreases. In a sense, the system cyclically "loses predictability" up to the phase of the year indicated by p_b, "regaining predictability" thereafter.

6.3 The Vallis Model

Let us now turn our attention to a simple ENSO model, that of Vallis (Val88). It is a simple two-point, two layer model, one which Vallis rescales to obtain:

$$\dot{u} = \hat{B} T_{\mathrm{n3}} - \hat{C}\,(u - u^{\star})\,, \tag{6.24}$$

$$\dot{T}_m = -u T_{\mathrm{n3}} - (T_m - 1)\,, \tag{6.25}$$

$$\dot{T}_{\mathrm{n3}} = u T_m - T_{\mathrm{n3}}, \tag{6.26}$$

where $\hat{B}$ is a thermal damping timescale, and $\hat{C}$ is a frictional damping timescale. The system's variables, u, T_m, and T_{n3}, are the nondimensionalized zonal wind speed, the mean upper ocean temperature, and the upper ocean temperature at the center of the basin, respectively. [Negative u represents easterlies.] The quantity $u^{\star}$ describes the effect of the trade winds along the Equator. In a sense, $u^{\star}$ is the wind forcing of the basin. Note that I have chosen the notation "T_{n3}" to serve as a mnemonic, for the NINO3 region is roughly in the center of the Pacific Ocean basin.

Vallis examines Equations (6.24-6.26) using parameter values of $\hat{B} = 102$ and $\hat{C} = 3$ and setting $u^{\star}$ to a variety of things (Val88). For the special case $u^{\star} = 0$, he performs a

stability analysis, noting that one can reduce Eqs. (6.24-6.26) to the Lorenz model (Lor63) with a substitution of variables and a parameter rescaling. He also examines constant forcing with[7] $u^\star = -0.47335104$, as well as stochastic forcing. The results of his analysis are present in his paper and are therefore not duplicated here.

For the purpose of this work, let us consider a periodic forcing, $u^\star = u_0^\star + u_1^\star \cos(2\pi t)$, where t is in units of model years. Equations (6.24-6.26) then become:

$$\dot{u} = \hat{B}T_{\mathrm{n3}} - \hat{C}u + \hat{u}_{\mathrm{bar}} + \hat{u}_{\mathrm{pd}} \cos(\tilde{\tau}), \tag{6.27}$$

$$\dot{T}_m = -uT_{\mathrm{n3}} - (T_m - 1), \tag{6.28}$$

$$\dot{T}_{\mathrm{n3}} = uT_m - T_{\mathrm{n3}}, \tag{6.29}$$

$$\dot{\tilde{\tau}} = 2\pi, \tag{6.30}$$

where the parameters $\hat{u}_{\mathrm{bar}} = \hat{C}u_0^\star$ and $\hat{u}_{\mathrm{pd}} = \hat{C}u_1^\star$. $\hat{u}_{\mathrm{bar}}$ represents the constant forcing of an average zonal wind, hence the notation. The Jacobian of this system is:

$$\mathbf{J}(\vec{x}) = \begin{pmatrix} -\hat{C} & 0 & \hat{B} & \hat{u}_{\mathrm{pd}} \sin(\tilde{\tau}) \\ -T_{\mathrm{n3}} & -1 & -u & 0 \\ T_m & u & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{6.31}$$

Notice that the dispersion for this system, $-\left(\hat{C} + 2\right)$, is a constant with respect to time. I shall integrate this system in the same manner used for the damped forced pendulum. I shall also compute the Lyapunov exponents and persistence map in the same fashion used for the pendulum.

Table 6.2 describes the examined regions of parameter space for Equations (6.27-6.30). The listed values of $\hat{u}_{\mathrm{bar}}$ and $\hat{u}_{\mathrm{pd}}$, the amplitudes of the constant and periodic forcing, are based on $u^\star = -0.47335104$ and $\hat{C} = 3$. The range of values for $\hat{B}$ are based on temperatures of 10° C and 15° C, rescaled as described in Vallis (Val88). This temperature range represents possible mean temperatures along the thermocline.

Figures 6.26 and 6.27 contain the barrier measures as a function of parameter

[7] I have rescaled the value Vallis uses to reflect time in units of model years.

Table 6.2: Table of parameter values examined for Equations (6.27-6.30). Each set of parameters is identified by the number of the figure displaying it. The third row is a single point in parameter space. The first two rows are two-dimensional cross-sections of parameter space. For these two rows, the specified values of the maximum global Lyapunov exponent [GLE] and the values of R^2 for their persistence maps are the maximum and minimum seen over the cross-section.

	$\hat{B}$	$\hat{C}$	$\hat{u}_{\text{bar}}$	$\hat{u}_{\text{pd}}$	max. GLE	R^2
Fig. 6.26	88.51800773 to 132.7770116 in 11 steps	3	-1.42005312	-0.284010624 to -2.84010624 in 10 steps	0.63–0.8 $\left[\frac{\text{bits}}{\text{year}}\right]$	0.64–0.84
Fig. 6.27	102	3	-0.284010624 to -2.84010624 in 10 steps	-0.284010624 to -2.84010624 in 10 steps	0.45–0.8 $\left[\frac{\text{bits}}{\text{year}}\right]$	0.3–0.87
Figs. 6.28 –6.30	102	3	-1.42005312	-1.42005312	0.705	0.798

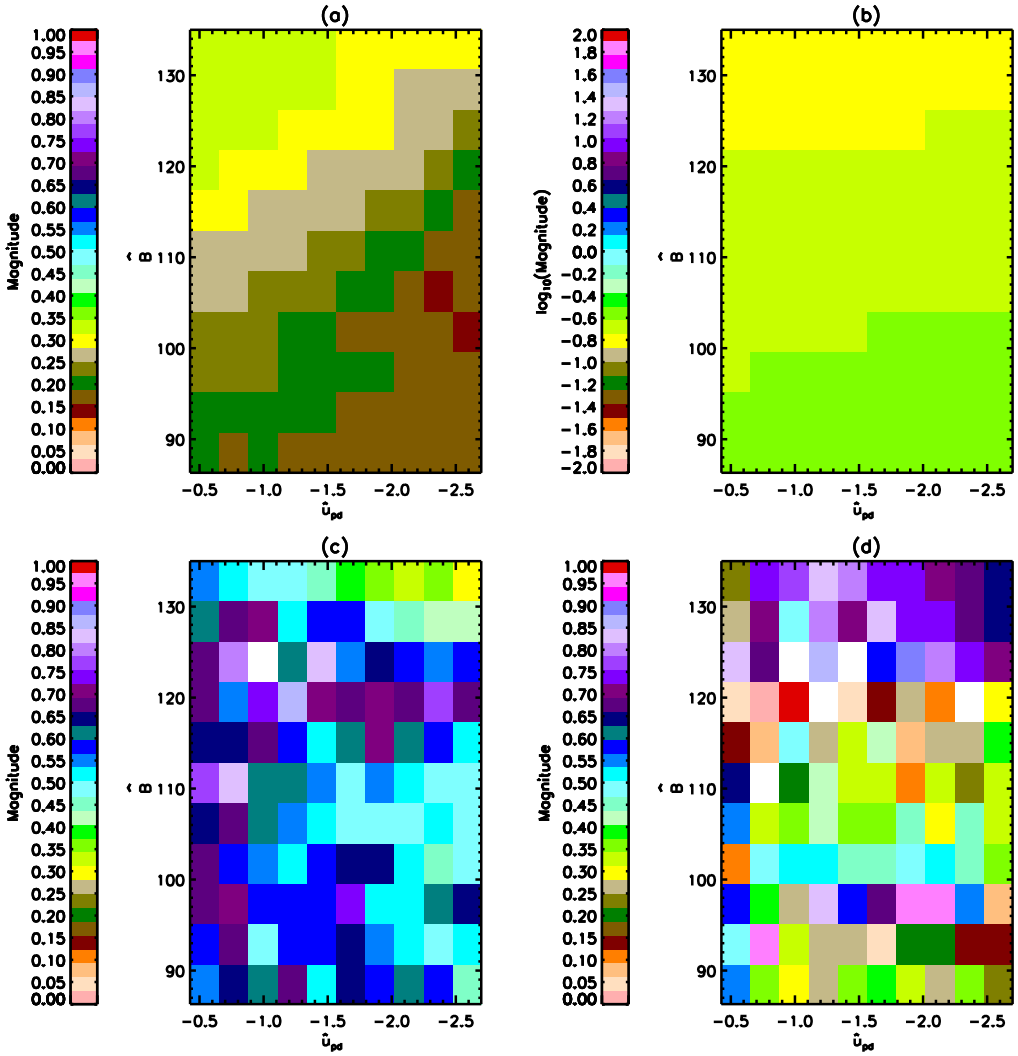

Figure 6.26: The barrier measures for Eqs. (6.27-6.30), with $\hat{C} = 3$ and $\hat{u}_{\rm bar} \approx 1.42$. A detailed description of the plotting method is in section C.2. (a) The barrier amplitude, A. (b) $\log_{10}(\tau)$, the stochastic decay time displayed on a logarithmic scale. (c) The barrier width, Δp. (d) The barrier position, p_b.

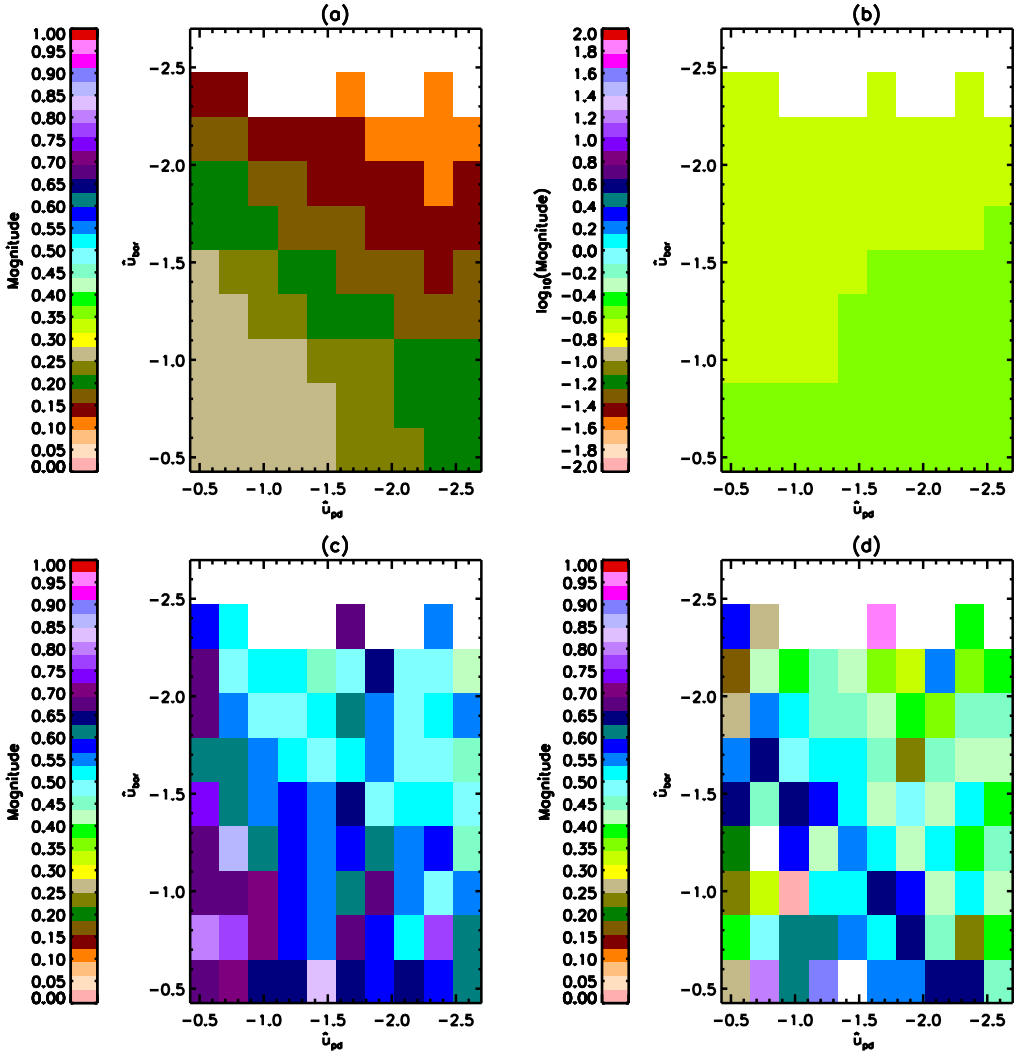

Figure 6.27: The barrier measures for Eqs. (6.27-6.30), with $\hat{B} = 102$ and $\hat{C} = 3$, plotted as in Figure 6.26. (a) The barrier amplitude, A. (b) $\log_{10}(\tau)$, the stochastic decay time displayed on a logarithmic scale. (c) The barrier width, Δp. (d) The barrier position, p_b.

points on a cross-section of parameter space. The first figure varies the thermal damping coefficient, $\hat{B}$, over a range corresponding to the temperatures 10° C and 15° C. The second varies the magnitude of the constant forcing, $\hat{u}_{\rm bar}$. Both figures vary the amplitude of the periodic driving. These figures differ from the tableaus of appendix C in one respect: there is no corresponding map of winding number. The concept of winding number for a three-dimensional system is not straightforward. I have found no satisfactory definition for such a winding number, and so shall ignore it. Furthermore, the largest global Lyapunov exponent is positive over all gridpoints in both figures. Therefore, I have simply listed in Table 6.2 the minimum and maximum values of the largest GLE for these parameter grids.

The actual values of each measure are rather uninteresting. In both figures, the stochastic decay time, τ, is roughly 0.25 yr., while the barrier width, Δp, is almost always larger than 0.5 yr. The barrier position, p_b, shows no clear pattern. The barrier amplitude in Figures 6.26 and 6.27 is in the range $0.15 < A < 0.35$ and $0.10 < A < 0.25$, respectively. Thus, the persistence contains a combination of a linear decay and an exponential decay in this region of parameter space. Or, it contains some other form of phase-dependence, as indicated by the broad range of nonlinear correlation coefficient [Table 6.2].

Figures 6.28-6.30 display an example of a single point in parameter space, with $\hat{B} = 102$, $\hat{C} = 3$, and $\hat{u}_{\rm bar} = \hat{u}_{\rm pd} \approx -1.42$. The three Poincaré Sections through parameter space reveal a strange attractor that strongly resembles that of the Lorenz system (Lor63). The persistence map, computed using $T_{\rm n3}$ [in analogy to the NINO3 SST], shows an almost exclusive dependence on lag. The distribution function of the maximal and minimal ILE's, computed according to the method of the previous section, also do not depend on the phase of the year.

This section attempted an examination of the simple Vallis model of ENSO (Val88), using a periodic driving force. The purpose was to see if this system manifested any of the behavior seen for the damped forced pendulum. Unfortunately, for the parameters used, the system contained no persistence barriers, nor any quasiperiodic or mode-locked behavior. Perhaps this system does contain quasiperiodic states in some other region of its parameter space; however, a search for them must be left to other work.

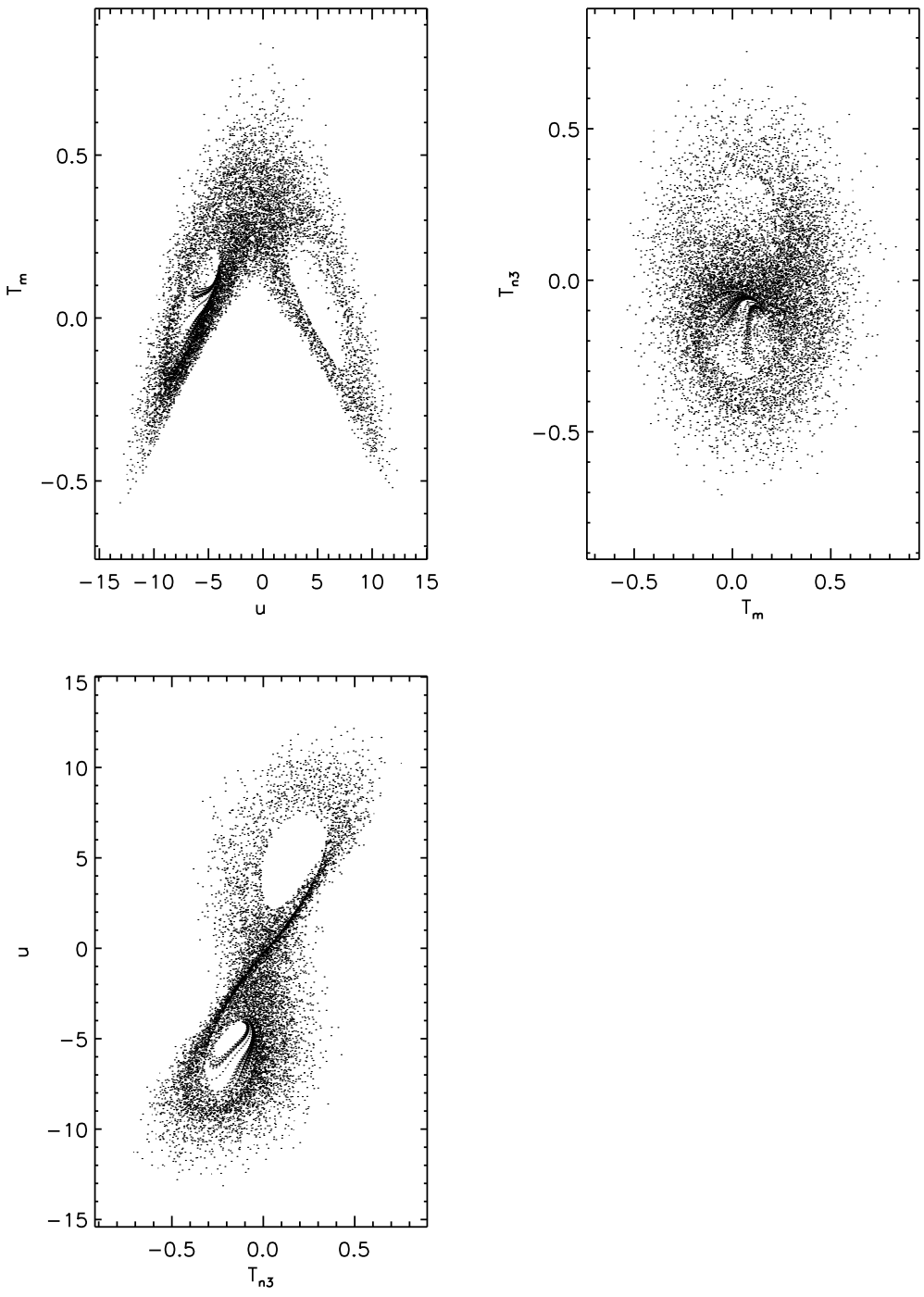

Figure 6.28: The three possible Poincaré Sections of phase space for Equations (6.27-6.30) with $\hat{B} = 102$, $\hat{C} = 3$, and $\hat{u}_{\rm bar} = \hat{u}_{\rm pd} \approx -1.42$.

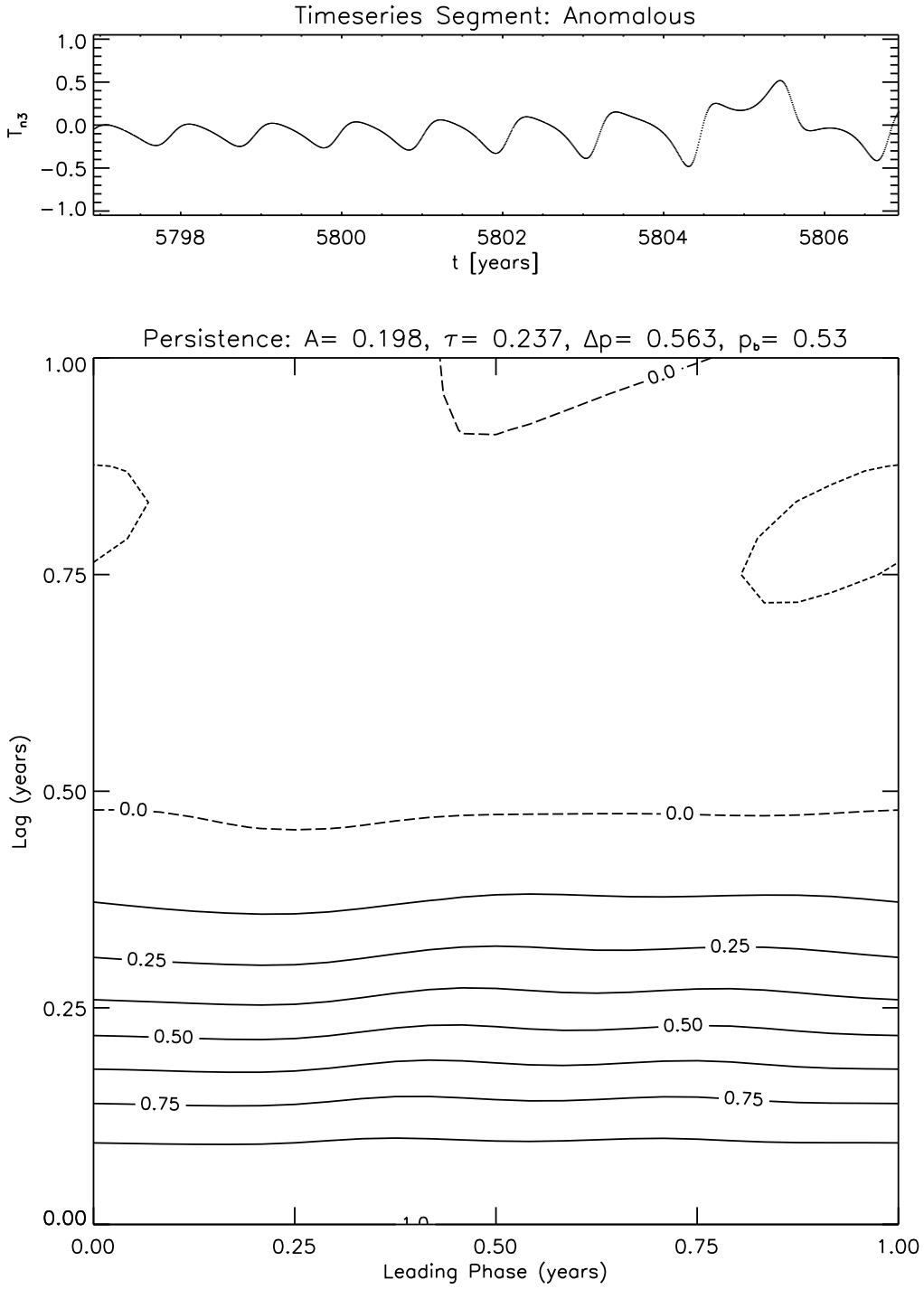

Figure 6.29: Persistence map and timeseries segment for Equations (6.27-6.30) with $\hat{B} = 102$, $\hat{C} = 3$, and $\hat{u}_{\rm bar} = \hat{u}_{\rm pd} \approx -1.42$. The timeseries segment is of $T_{\rm n3}$ minus its cyclostationary mean. The persistence map is computed from $T_{\rm n3}$.

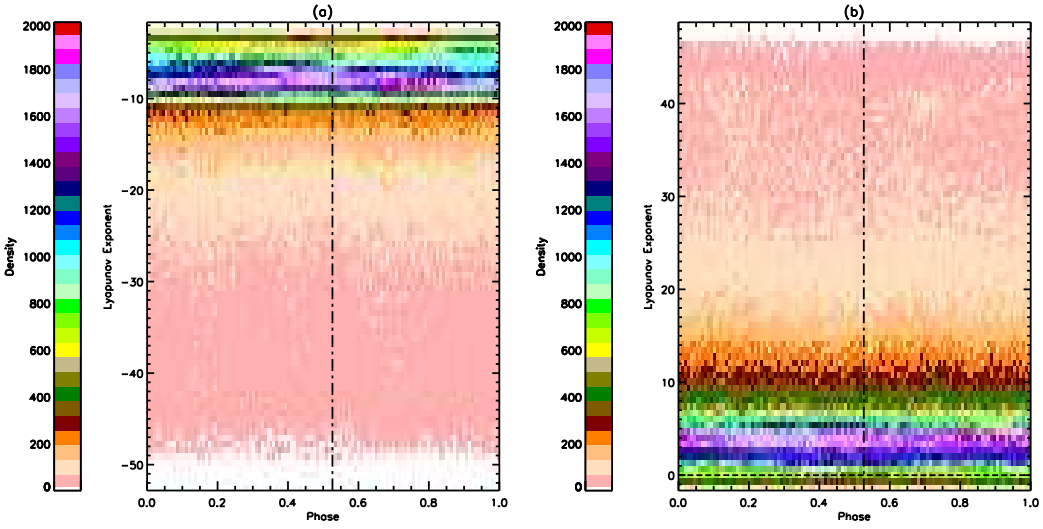

Figure 6.30: Phase-dependent probability distribution functions of the maximum and minimum ILE for Equations (6.27-6.30) with $\hat{B} = 102$, $\hat{C} = 3$, and $\hat{u}_{\rm bar} = \hat{u}_{\rm pd} \approx -1.42$, plotted in the same manner as Figure 6.19. (a) PDF of the minimum ILE. (b) PDF of the maximum ILE.

6.4 Summary

In this chapter, we have examined the damped forced pendulum and a version of the Vallis ENSO model (Val88) containing periodic forcing. The chapter began with an overview of the quasiperiodic route to chaos. Section 6.2.1 described the equations of motion for the damped forced pendulum. A discussion of behavior seen in the pendulum by other works, with a connection to the analysis of the pendulum in appendix B, was presented in section 6.2.2.

In section 6.2.3, we saw the first sign of a connection between the quasiperiodic route to chaos and persistence barriers. The quasiperiodic route to chaos occurs in the region of the pendulum's parameter space where $D_0 \neq 0$. For quasiperiodic states, the barrier measures A and Δp occasionally showed a dependence on the ratios $\frac{D_0}{\beta}$ and $\frac{D_0}{\gamma}$. The barrier amplitude, A was typically higher and Δp, the barrier width, was typically smaller in the quasiperiodic area of parameter space than in the $D_0 = 0$ region. A joint distribution function of the largest global Lyapunov exponent vs. each barrier measure corroborates this. Specifically, for chaotic states, the barrier amplitude has the greatest density at values below 0.1 when $D_0 = 0$, while it is more uniformly distributed between 0 and 1 for quasiperiodic chaos.

Section 6.2.4 revealed a further connection between persistence barriers and Lyapunov exponents. At any given point in phase-space, there is a maximal instantaneous Lyapunov exponent, or $\mathrm{ILE}_{\max}$, indicating the rate of most rapid error growth. A special distribution function of $\mathrm{ILE}_{\max}$, dependent on both the magnitude of the exponent and the phase of the year in which that value occurs, shows a striking connection to the barrier position, p_b, and the barrier width, Δp. At times prior to p_b, the distribution of $\mathrm{ILE}_{\max}$ at a given phase narrows and shifts to more rapid error growth. This shift peaks at p_b. At times after p_b, the distribution of $\mathrm{ILE}_{\max}$ at a given phase changes its shape little but shifts downward, indicating weaker error growth. For some states of the pendulum, $\mathrm{ILE}_{\max}$ begins to disperse [i. e. the distribution at a given phase broadens] after $p_b + \Delta p$.

The connection between $\mathrm{ILE}_{\max}$ and p_b is most clear in the three examples of quasiperiodic chaos. For some of the $D_0 = 0$ examples, $\mathrm{ILE}_{\max}$ also seemed to mimic the phase-dependence in the persistence map, even though this dependence was not in the form

of a barrier. Generally speaking, at a particular phase of the year, there appears to be an increase in the typical rate of error growth prior to a more rapid loss of correlation in the persistence map. Following the aforementioned point of most rapid drop in persistence, the typical error growth rate decreases. This suggests that, for chaotic or seemingly chaotic data, a barrier in the data's persistence map indicates a cyclical loss of predictability in the form of more rapid error growth.

For the parameter values examined, the periodically driven Vallis model of section 6.3 was always chaotic and contained no persistence barriers. Interestingly, though, the distribution of $\mathrm{ILE}_{\mathrm{max}}$ was, like the persistence map, "flat." It contained no variation as a function of phase of year. I therefore posit the following: suppose there is a set of parameters that produces a persistence barrier in this system. I speculate that the distributions of $\mathrm{ILE}_{\mathrm{max}}$ will contain a phase-dependence like that seen in the pendulum. That is, there will be an increase in the maximal error grows of the system corresponding to the persistence barrier position.

Chapter 7

Conclusions

The introductory chapter raised a question: Could the "barriers" in persistence exhibited by the El Niño-Southern Oscillation [ENSO] somehow be connected to low-order chaotic behavior? The main results of this dissertation have made definite progress towards an answer to this question. A mathematical understanding led to a set of measures for barrier-like features in persistence. These measures provided the necessary link to local rates of error growth, especially for systems exhibiting the quasiperiodic route to chaos.

The work began with an overview of the dynamics of error growth, the Lyapunov exponents. Chapter 2 summarized the work of several fundamental papers that define the Lyapunov exponents. Lyapunov exponents describe the exponential growth rate of an initial volume of perturbations evolving under linearized perturbation dynamics. The exponent depends not on the size of the initial volume but the subspace it occupies in phase space. The global Lyapunov exponent spectrum, described in section 2.2.3, are a set of numbers that uniquely describe the growth rates of any volume in any subspace of phase space. Sections 2.2.4 and 2.2.5 laid the groundwork for the definition of local Lyapunov exponents in section 2.3. These local exponents describe error growth in a specific region of the attractor of the system. One version of local exponents, the "instantaneous" Lyapunov exponents, describes error growth "at a point" in phase space. From the instantaneous Lyapunov exponent [ILE] spectrum, one can compute both the global exponent spectrum as well as any local exponent over any portion of the attractor. Distribution functions of the ILE become quite useful in chapter 6.

Chapter 3 focused on the task of fitting a model to data. It described some of the basic principles of regression using models which are nonlinear functions of their parameters.

Section 3.2 then discussed the method of genetic algorithms, including how one could use a genetic algorithm to compute an initial parameter value for a "traditional" nonlinear regression method. The remainder of the chapter defined an estimate for confidence intervals on the parameters, along with a generalized "correlation coefficient" usable with nonlinear models. By using nonlinear regression as a form of pattern recognition, one could determine numerically which persistence maps contain a barrier-like feature.

Chapter 4 undertook a detailed mathematical analysis of persistence. Persistence belongs to a class of statistics known as cyclostationary statistics, statistics which account for any natural periodicity in a data set. Cyclostationary statistics are always periodic functions of the phase of the year [or whatever other periodic cycle the data contains]. Persistence is, technically, the cyclostationary lag autocorrelation.

The mathematical analysis of persistence involved computing the persistence of various timeseries and examining the results for insight into the meaning of the "persistence barrier," a particular form of phase-dependence seen in the persistence of ENSO data. A time series composed of a single sine wave only produces a phase-dependent persistence map if the frequency of the wave is a biennial cycle or a harmonic thereof, i. e. $\nu = (2n+1)/2$. The phase-dependence takes the form of a single discontinuous "barrier" if the wave is a biennial cycle, i. e. $n = 0$. Otherwise, there are multiple discontinuous barriers in a single year.

A time series composed of two sine waves produces phase-dependent persistence maps under two circumstances. Obviously if one or both frequencies are harmonics of the biennial cycle, the resulting persistence map depends on the phase of the year. Surprisingly, phase-dependence also results when neither frequency is a biennial harmonic, provided the frequencies are $m^{\pm}$-complementary; that is, their sum or difference is an integer, m. In particular, unit-complementary frequencies, frequencies whose sum is 1, produce a persistence barrier. This suggests that the "barrier-producing harmonics" of a time series with a broadband spectrum are either biennial cycles or unit-complementary pairs of harmonics.

Time series with narrow but continuous peaks can produce barriers when the peaks are centered on the barrier-producing harmonics. If the phase angles of the harmonics are rapidly varying, however, the contributions of the continuum of modes around the peaks cancel, and the persistence depends only on lag. For either slowly varying or coherent phase

angles, a barrier does result and is similar to that produced by discrete modes. Thus, the existence of a barrier depends on barrier-producing harmonics having coherent phase angles. The position of that barrier is determined by those same coherent phase angles. This suggests that an arbitrary time series will only have a barrier if its dominant harmonics are barrier producing and have coherent phases.

Using this analysis of persistence, I then proposed a quantitative measure of the properties of a persistence barrier. The measure is based on a model persistence barrier combining both deterministic and stochastic features, a model containing four parameters: a barrier amplitude, A, a stochastic decorrelation time, τ, a barrier width, Δp, and a barrier position, p_b. Computation of the measures makes extensive use of the techniques described in chapter 3.

Building on the results of chapter 4, chapter 5 applied the barrier measures to the SOI and NINO3 SST. The measures show that the persistence of both data sets contain barriers which are statistically significant from the theoretical red-noise barrier to within 68% confidence. Comparisons of both datasets to a Monte-Carlo simulation of a red-noise process reveal a similar difference. The "predictability barrier" is therefore an intrinsic statistical property of the ENSO data.

Using the SOI and NINO3 SST, I then examined numerically the relationship of the barrier-generating harmonics and their phase angles to persistence barriers. By selectively filtering certain harmonics from the data, one can use the barrier measures to determine which harmonics are responsible for generating the barrier. In the SOI, the near-biennial modes generate the barrier, whereas all unit-complementary harmonics of the NINO3 SST act in unison to produce a persistence barrier. Phase-scrambling Monte-Carlo experiments randomized the phase-angles associated with the barrier-generating harmonics, testing the phase-coherency criteria. For both the NINO3 SST and the SOI, phase-scrambling weakens or eliminates the persistence barrier, confirming the necessity for coherent Fourier phases amongst the barrier-generating harmonics.

The chapter concluded with an application of the measures to the interdecadal variability of ENSO. The persistence of a running 20-year window of the ENSO data was computed and measured. The resulting timeseries of measures described the interdecadal variability of the persistence barrier. The measures revealed a weak persistence barrier from

1915 to 1945 and another decline in the barrier after 1975. In contrast, the 1960's and 1970's showed a strong barrier. At present, the persistence barrier in both ENSO datasets is weak. Additionally, the barrier position, p_b, changed location over the course of this century. During the middle of the century, p_b was around March, while the barrier was positioned around May during other times.

Other works (XCZB94, TW98) have demonstrated a connection between the persistence barrier and phase-locking of ENSO to the annual cycle. It is likely that such phase-locking generates the necessary coherency in the unit-complementary harmonics of the ENSO data. Note that phase-locking is also a prominent feature of the quasiperiodic route to chaos. Under this route to chaos, a system locks to a harmonic of an external forcing. The chaotic behavior ensues when the system attempts to simultaneously lock to two different harmonics of the forcing. The system instead randomly alternates between the two phase-locked behaviors. Nevertheless, some effect of the two original phase-locked states should remain. A connection between persistence barriers and the quasiperiodic route to chaos therefore seems likely.

In chapter 6, I examined the damped forced pendulum and a version of the Vallis ENSO model (Val88) containing periodic forcing. The driven Vallis model showed no examples of quasiperiodic behavior for the examined parameters. Additionally, the barrier measures indicated either a poor fit to the barrier model or a persistence map dependent only on lag. Let us therefore focus on the results from the damped forced pendulum.

The damped forced pendulum contains a rich variety of behavior, including period-doubling bifurcations, crisis-induced intermittency, intermittency induced by saddle-node bifurcations, and quasiperiodic routes to chaos. A discussion of behavior seen in the pendulum by other works was presented in section 6.2.2. Using these works as a starting point, I explored several cross-sections of the pendulum's parameter space, the results of which are assembled in appendix C. The quasiperiodic route to chaos occurs in the region of the pendulum's parameter space where $D_0 \neq 0$.

The first task in the analysis of the pendulum involved R^2, the nonlinear correlation coefficient defined in chapter 3. R^2 describes the goodness of fit between a persistence map and the barrier model used to compute the measures. By assembling the values of R^2 computed at all points in the pendulum parameter space into a distribution, we saw that

$R^2 \geq 0.98$ for 95% of the parameter points. Thus, any points in parameter space with $R^2 < 0.98$ were interpreted as fitting poorly to the barrier model.

In the quasiperiodic region of parameter space, quasiperiodic states consistently fit poorly to the barrier model. In contrast, for both mode-locked and chaotic states, the barrier amplitude, A was typically larger than 0.5 while Δp, the barrier width, was typically smaller than 0.4 yr. In some cases, these two measures indicated an even stronger barrier than do the aforementioned values. Again, strong, narrow persistence barriers appeared in mode-locked and chaotic states alike. In the $D_0 = 0$ region, however, there was not a clear pattern to the barrier amplitude and width.

To examine this behavior more closely, I assembled joint distribution functions of the largest global Lyapunov exponent vs. each barrier measure. Two such distribution functions were examined, one for the $D_0 = 0$ region, the other for the quasiperiodic area of parameter space. For chaotic states, the barrier amplitude has the greatest density at values below 0.1 when $D_0 = 0$, while it is more uniformly distributed between 0 and 1 for quasiperiodic chaos [Figures 6.6-6.7]. In both joint distributions, the barrier width, Δp, remained near zero.

The more uniform distribution of A for chaotic states in the $D_0 \neq 0$ region of parameter space is suggestive. It hints at a connection between quasiperiodic chaos and strong persistence barriers. It is not a definitive connection, however. The histograms upon which these results are based examine only a smattering of the possible points in the pendulum's parameter space. There could easily be a similar connection, say, to period-doubling bifurcations, a connection not revealed by this limited exploration of parmeter space.

Section 6.2.4 revealed a further connection between persistence barriers and Lyapunov exponents. At any given point in phase-space, there is a maximal instantaneous Lyapunov exponent, or $\mathrm{ILE}_{\max}$, indicating the rate of most rapid error growth. A special distribution function of $\mathrm{ILE}_{\max}$, dependent on both the magnitude of the exponent and the phase of the year in which that value occurs, showed a strong connection to the barrier position, p_b, and the barrier width, Δp. These distributions often contained "ridges," bands of high occurrence. At times prior to p_b, such ridges in the distribution of $\mathrm{ILE}_{\max}$ converged and shifted to more rapid error growth as a function of phase of the year. This shift peaked

at p_b. At times after p_b, the ridges changed their shape and distribution little but shifted downward as a function of phase, indicating weaker error growth. For some states of the pendulum, $\mathrm{ILE}_{\max}$ began to disperse [i. e. the distribution at a given phase broadened] after $p_b + \Delta p$.

The reader should note that I have painted the description of the behavior of $\mathrm{ILE}_{\max}$ in very broad strokes indeed. The use of the term "ridges" is not accurate for all states of the pendulum. There are some points in parameter space at which the error growth behavior manifests "smears" in the $\mathrm{ILE}_{\max}$ distribution instead of ridges. Regardless, there is nevertheless an increase in the values of $\mathrm{ILE}_{\max}$ that peaks at p_b. Only when $\Delta p > 0.5$ yr., when the persistence contains only a dependence on lag, does this not hold.

The connection between $\mathrm{ILE}_{\max}$ and p_b was most clear in the three examples of quasiperiodic chaos. For some of the $D_0 = 0$ examples, $\mathrm{ILE}_{\max}$ also seemed to mimic the phase-dependence in the persistence map, even though this dependence was not in the form of a barrier. Generally speaking, at a particular phase of the year, there appears to be an increase in the typical rate of error growth prior to a more rapid loss of correlation in the persistence map. Following the aforementioned point of most rapid drop in persistence, the typical error growth rate decreases. This suggests that, for chaotic or seemingly chaotic data, a barrier in the data's persistence map indicates a cyclical loss of predictability in the form of more rapid error growth.

Once again, it is highly tempting to posit a strong connection between persistence barriers and Lyapunov exponents. The results from phase-dependent distributions of $\mathrm{ILE}_{\max}$, the maximal instantaneous Lyapunov exponent, do seem to indicate this. Said results also point to a connection between persistence barriers and an increase in local error growth rates. At the very least, for the systems displayed in Figures 6.17-6.19 and 6.23-6.25, the persistence barrier is also a "predictability barrier." That is, the barrier coincides with an increase in the maximal error growth rate in phase space.

In summary, joint distribution functions of the largest global Lyapunov exponent and the barrier measures hint at a connection between the quasiperiodic route to chaos and strong persistence barriers. Furthermore, phase-dependent distributions of the maximal instantaneous Lyapunov exponent show a connection between persistence barriers and predictability. The latter connection is seen in a specific set of parameters on the quasiperiodic

route to chaos.

Because the aforementioned connections occur only in certain parts of parameter space, one direction for future work should involve a more detailed study of parameter space. This work examined the largest global Lyapunov exponent [GLE], the winding number, and the persistence barrier measures for pendulum parameters varying over a two-dimensional grid. Imagine now replacing the grid of fixed values with a Monte Carlo experiment that chooses points in parameter space at random. One could do two varieties of this Monte Carlo experiment, one with the restriction $D_0 \neq 0$ and a second for which D_0 is fixed at 0. One could then use the distribution of winding number and of the largest GLE to determine if the $D_0 = 0$ subspace truly contains no quasiperiodic routes to chaos. This, however, is a secondary experiment. The primary goal is to construct the joint distributions of GLE and barrier measure, comparing the distributions from the $D_0 = 0$ subspace to those of the $D_0 \neq 0$ parameter points in the fashion done here.

There are two difficulties with the experiment described above. First, none of the pendulum parameters have any upper bound. The experimenter, however, will need to impose such a bound to make the Monte Carlo experiment practical. Furthermore, the number of runs needed to uniformly cover even the three-dimensional $D_0 = 0$ subspace of parameter space will be quite large, perhaps even prohibitively so.

Far trickier is any attempt to prove a connection between quasiperiodic chaos and persistence barriers using distributions of $\mathrm{ILE}_{\max}$. Though containing some common traits, said distributions will differ in appearance and behavior from one point in parameter space to another. A visual examination of distributions of $\mathrm{ILE}_{\max}$ using several points in parameter space, both on the quasiperiodic route to chaos and on other routes, is the next logical step.

The barrier measure used in this work depends on a somewhat limited model. The barrier model of chapter 4 only described phase-dependence of a particular type. There are others. Recall from chapter 4 that unit-complementary harmonics, harmonics whose sum equals 1, generate persistence barriers. There are, however, m-complementary harmonics which produce other phase-dependence patterns. For the purpose of examining the pendulum's parameter space, it may be of use to have a model that incorporates other "m-complementary patterns" besides the persistence barrier.

Finally, I return again to the introduction. Recall that most estimates of low-

dimensional chaos in physical data often underestimate the dimensionality of the attractor (Lor91). Abarbanel and Lall (AL96) discovered that, as long as a data set sufficiently samples the attractor of the system, estimates of dimension and Lyapunov exponent, along with phase space reconstruction, need not require a huge dataset. Lorenz (Lor91) noted that an appropriate choice of observable may capture the essential dynamics of the full system. Consider, then, a multivariate phase space reconstruction using SST, sea level pressure, and rainfall timeseries from numerous locations across the Pacific. Let us also add a few timeseries from the Indian monsoon and the Atlantic Ocean, simply to ensure a broad sampling of the phase space of the Earth-system. It would indeed be interesting to see what this multivariate phase space reconstruction would produce for an estimate of the largest GLE.

Bibliography

[AB84] M. Y. Azbel and P. Bak. Analytical results on the periodically driven damped pendulum. Application to sliding charge-density waves and Josephson junctions. Phys. Rev. B, 30(7):3722–3727, Oct. 1984.

[AB93] C. Amitrano and R. S. Berry. Probability distributions of local Lyapunov exponents for Hamiltonian systems. Phys. Rev. E, 47(5):3158–3173, May 1993.

[ABK91a] H. D. I. Abarbanel, R. Brown, and M. B. Kennel. Lyapunov exponents in chaotic systems: Their importance and their evaluation using observed data. Int. J. Mod. Phys. B, 5(9):1347–1375, 1991.

[ABK91b] H. D. I. Abarbanel, R. Brown, and M. B. Kennel. Variation of Lyapunov exponents on a strange attractor. J. Nonlinear Sci., 1:175–199, 1991.

[ABK92] H. D. I. Abarbanel, R. Brown, and M. B. Kennel. Local Lyapunov exponents computed from observed data. J. Nonlinear Sci., 2:343–365, 1992.

[AL85a] P. Alstrøm and M. T. Levinsen. Fractal structure of the complete devil's staircase in dissipative systems described by a driven damped-pendulum equation with a distorted potential. Phys. Rev. B, 32(3):1503–1511, Aug. 1985.

[AL85b] P. Alstrøm and M. T. Levinsen. Josephson junction at the onset of chaos: A complete devil's staircase. Phys. Rev. B, 31(5):2753–2758, Mar. 1985.

[AL86] P. Alstrøm and M. T. Levinsen. Cross-over for the sine map and the driven damped pendulum. Phys. Scr., 34:373–377, 1986.

[AL96] H. D. I. Abarbanel and U. Lall. Nonlinear dynamics of the Great Salt Lake: System identification and prediction. Clim. Dyn., 12:287–297, 1996.

[BAD94] M. A. Balmaseda, D. L. T. Anderson, and M. K. Davey. ENSO prediction using a dynamical ocean model coupled to statistical atmospheres. Tellus, 46A(4):497–511, 1994.

[Bak86] P. Bak. The devil's staircase. Physics Today, 39(12):39–45, Dec. 1986.

[Bat88] D. S. Battisti. Dynamics and thermodynamics of a warming event in a coupled tropical atmosphere-ocean model. J. Atmos. Sci., 45(20):2889–2919, Oct. 1988.

[BBJ84] T. Bohr, P. Bak, and M. H. Jensen. Transition to chaos by interaction of resonances in dissipative systems. II. Josephson junctions, charge-density waves, and standard maps. Phys. Rev. A, 30(4):1970–1981, Oct. 1984.

[BBJC84] P. Bak, T. Bohr, M. H. Jensen, and P. V. Christiansen. Josephson junctions and circle maps. Solid State Commun., 51(4):231–234, 1984.

[BDA95] M. A. Balmaseda, M. K. Davey, and D. L. T. Anderson. Decadal and seasonal dependence of ENSO prediction skill. J. Climate, 8:2705–2715, Nov. 1995.

[BG90] G. L. Baker and J. P. Gollub. Chaotic Dynamics: An Introduction. Cambridge University Press, 1990.

[BGGS80a] G. Benettin, L. Galgani, A. Giorgilli, and J.-M. Strelcyn. Lyapunov characteristic exponents for smooth dynamical systems and for Hamiltonian systems; A method for computing all of them. Part 1: Theory. Meccanica, 15:9–20, March 1980.

[BGGS80b] G. Benettin, L. Galgani, A. Giorgilli, and J.-M. Strelcyn. Lyapunov characteristic exponents for smooth dynamical systems and for Hamiltonian systems; A method for computing all of them. Part 2: Numerical application. Meccanica, 15:21–30, March 1980.

[BGS76] G. Benettin, L. Galgani, and J.-M. Strelcyn. Kolmogorov entropy and numerical experiments. Phys. Rev. A, 14(6):2338–2345, Dec. 1976.

[BH89] D. S. Battisti and A. C. Hirst. Interannual variability in a tropical atmosphere-ocean model: Influence of the basic state, ocean geometry and nonlinearity. J. Atmos. Sci., 46(12):1687–1712, Jun. 1989.

[Bje66] J. Bjerknes. A possible response of the atmospheric Hadley circulation to equatorial anomalies of ocean temperature. Tellus, 18:820–829, 1966.

[Bje69] J. Bjerknes. Atmospheric teleconnections from the equatorial Pacific. Mon. Wea. Rev., 97(3):163–172, Mar. 1969.

[BLKR91] T. P. Barnett, M. Latif, E. Kirk, and E. Roeckner. On ENSO physics. J. Climate, 4:487–515, May 1991.

[Blu91] M. B. Blumenthal. Predictability of a coupled ocean-atmosphere model. J. Climate, 4:766–784, Aug. 1991.

[BVWS89] J. A. Blackburn, S. Vik, B. Wu, and H. J. T. Smith. Driven pendulum for studying chaos. Rev. Sci. Instrum., 60(3):422–426, Mar. 1989.

[BZC95a] G. Bürger, S. E. Zebiak, and M. A. Cane. Quasi-fixed points and periodic orbits in the Zebiak-Cane ENSO model with applications in Kalman filtering. Part I: Monthly quasi-fixed points. Mon. Wea. Rev., 123:2802–2813, Sept. 1995.

[BZC95b] G. Bürger, S. E. Zebiak, and M. A. Cane. Quasi-fixed points and periodic orbits in the Zebiak-Cane ENSO model with applications in Kalman filtering. Part II: Periodic orbits. Mon. Wea. Rev., 123:2814–2824, Sept. 1995.

[CMZ90] M. A. Cane, M. Münnich, and S. E. Zebiak. A study of self-excited oscillations of the tropical ocean-atmosphere system. Part I: Linear analysis. J. Atmos. Sci., 47(13):1562–1577, Jul. 1990.

[CMZ91] M. A. Cane, M. Münnich, and S. E. Zebiak. A study of self-excited oscillations of the tropical ocean-atmosphere system. Part II: Nonlinear cases. J. Atmos. Sci., 48(10):1238–1248, May. 1991.

[CWLJ94] P. Chang, B. Wang, T. Li, and L. Ji. Interactions between the seasonal cycle and the Southern Oscillation — frequency entrainment and chaos in a coupled ocean-atmosphere model. Geophys. Res. Lett., 21(25):2817–2820, Dec. 1994.

[CZ85] M. Cane and S. E. Zebiak. A theory for El Niño and the Southern Oscillation. Science, 228:1085–1086, May 1985.

[CZBC95] D. Chen, S. E. Zebiak, A. J. Busalacchi, and M. A. Cane. An improved procedure for El Niño forecasting: Implications for predictability. Science, 269:1699–1702, Sept. 1995.

[DBHL82] D. D'Humieres, M. R. Beasley, B. A. Huberman, and A. Libacher. Chaotic states and routes to chaos in the forced pendulum. Phys. Rev. A, 26(6):3483–3496, Dec. 1982.

[EKRC86] J. P. Eckman, S. O. Kamphosrt, D. Ruelle, and S. Ciliberto. Liapunov exponents from time series. Phys. Rev. A, 34(6):4971–4979, Dec. 1986.

[EY93] B. Eckhardt and D. Yao. Local Lyapunov exponents in chaotic systems. Physica D, 65:100–108, 1993.

[GA96] N. Gupte and R. E. Amritkar. Enhancing chaos in chaotic maps and flows. Phys. Rev. E, 54(5):4580–4585, Nov. 1996.

[Gar90] C. W. Gardiner. Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences. Springer Verlag, 2 edition, 1990.

[Gla61] E. G. Gladyshev. Periodically correlated random sequences. Soviet Math., 2:385–388, 1961.

[Gol89] D. E. Goldberg. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, 1989.

[GSO87] I. Goldhirsch, P. L. Sulem, and S. Orsag. Stability and Lyapunov stability of dynamical systems: A differential approach and a numerical method. Physica D, 27:311–337, 1987.

[GW85] E. G. Gwinn and R. M. Westervelt. Intermittent chaos and low-frequency noise in the driven damped pendulum. Phys. Rev. Lett., 54(15):1613–1616, Apr. 1985.

[GW86] E. G. Gwinn and R. M. Westervelt. Fractal basin boundaries and intermittency in the driven damped pendulum. Phys. Rev. A, 33(6):4143–4155, Jun. 1986.

[GW88] N. E. Graham and W. B. White. The El Niño cycle: A natural oscillator of the Pacific ocean-atmosphere system. Science, 240:1293–1301, Jun. 1988.

[HDHM94a] D. Heng, R. Doerner, B. Hübinger, and W. Martienssen. Approaching nonlinear dynamics by studying the motion of a pendulum. I. Observing trajectories in state space. Internat. J. Bifur. Chaos Appl. Sci. Engrg., 4(4):751–760, 1994.

[HDHM94b] D. Heng, R. Doerner, B. Hübinger, and W. Martienssen. Approaching nonlinear dynamics by studying the motion of a pendulum. II. Analyzing chaotic motion. Internat. J. Bifur. Chaos Appl. Sci. Engrg., 4(4):761–771, 1994.

[HDHM94c] D. Heng, R. Doerner, B. Hübinger, and W. Martienssen. Approaching nonlinear dynamics by studying the motion of a pendulum. III. Predictability and control of chaotic motion. Internat. J. Bifur. Chaos Appl. Sci. Engrg., 4(4):773–784, 1994.

[HH98] R. L. Haupt and S. E. Haupt. Practical Genetic Algorithms. Wiley, 1998.

[HN96] J.-P. Huang and G. R. North. Cyclic spectral analysis of fluctuations in a GCM simulation. J. Atmos. Sci., 53(3):370–379, Feb. 1996.

[JBB84] M. H. Jensen, P. Bak, and T. Bohr. Transition to chaos by interaction of resonances in dissipative systems. I. Circle maps. Phys. Rev. A, 30(4):1960–1969, Oct. 1984.

[JN93a] F.-F. Jin and J. D. Neelin. Modes of interannual tropical ocean-atmosphere interaction—a unified view. Part I: Numerical results. J. Atmos. Sci., 50(21):3477–3503, Nov. 1993.

[JN93b] F.-F. Jin and J. D. Neelin. Modes of interannual tropical ocean-atmosphere interaction—a unified view. Part II: Analytical results in the weak-coupling limit. J. Atmos. Sci., 50(21):3504–3522, Nov. 1993.

[JN93c] F.-F. Jin and J. D. Neelin. Modes of interannual tropical ocean-atmosphere interaction—a unified view. Part III: Analytical results in fully coupled cases. J. Atmos. Sci., 50(21):3523–3540, Nov. 1993.

[JNG94] F.-F. Jin, D. Neelin, and M. Ghil. El Niño on the devil's staircase: Annual subharmonic steps to chaos. Science, 264:70–72, Apr. 1994.

[JTX95] M. S. Jolly, R. Téman, and C. Xiong. Convergence of a chaotic attraction with increased spatial resolution of the Ginzgurg-Landau equation. Chaos Solitions Fractals, 5(10):1833–1845, 1995.

[KGV83] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecci. Optimization by simulated annealing. Science, 220(4598):671–680, May 1983.

[Kir84] S. Kirkpatrick. Optimization by simulated annealing: Quantitative studies. J. Stat. Phys., 34(5/6):975–986, 1984.

[Kle91] R. Kleeman. A simple model of the atmosphere response to ENSO sea surface temperature anomalies. J. Atmnos. Sci., 48(1):3–18, Jan. 1991.

[Kvå85] T. O. Kvålseth. Cautionary note about R^2. Am. Stat., 39(4):279–285, Nov. 1985.

[KWB+85] W. C. Kerr, M. B. Williams, A. R. Bishop, K. Fesser, P. S. Lomdahl, and S. E. Trullinger. Symmetry and chaos in the motion of the damped driven pendulum. Z. Phys. B – Condensed Matter, 59:103–110, 1985.

[LGPA93] A. Lepri, G. Giacomelli, A. Politi, and F T. Arecchi. High-dimensional chaos in delayed dynamical systems. Physica D, 70:235–249, 1993.

[Lor63] E. N. Lorenz. Deterministic nonperiodic flow. J. Atmos. Sci., 20:130–141, Mar. 1963.

[Lor65] E. N. Lorenz. A study of the predictability of a 28-variable atmospheric model. Tellus, 17(3):321–333, 1965.

[Lor69] E. N. Lorenz. The predictability of a flow which posesses many scales of motion. Tellus, 21(3):289–307, 1969.

[Lor84] E. N. Lorenz. The local structure of a chaotic attractor in four dimensions. Physica D, 13D:90–104, 1984.

[Lor91] E. N. Lorenz. Dimension of weather and climate attractors. Nature, 353:241–244, Sep. 1991.

[Mee90] G. A. Meehl. Seasonal cycle forcing of El Niño-Southern Oscillation in a global coupled ocean-atmosphere GCM. J. Climate, 3:72–98, Jan. 1990.

[MGH80] J. J. Moré, B. S. Garbow, and K. E. Hillstrom. User guide for MINPACK-1. Report ANL-80-74, Argonne National Laboratory, Argonne, Illinois, 1980.

[MK96] A. M. Moore and R. Kleeman. The dynamics of error growth and predictability in a coupled model of ENSO. Q. J. R. Meteorol. Soc., 122:1405–1446, 1996.

[Nee90] J. D. Neelin. A hybrid coupled general circulation model for El Niño studies. J. Atmos. Sci, 47(5):674–693, Mar. 1990.

[Nes89] J. Nese. Quantifying local predictability in phase space. Physica D, 35:237–250, 1989.

[NL92] Y. I. Neimark and P. S. Landa. Stochastic and Chaotic Oscillations. Kluwer Academic Publishers, 1992.

[OPRR87] A. H. Oort, Y. H. Pan, R. W. Reynolds, and C. F. Ropelewski. Historical trend in the surface temperature over the oceans based on the COADS. Climate Dynamics, 2:29–38, 1987.

[Ose68] V. I. Oseledec. A multiplicative ergodic theorem. Ljapunov characteristic numbers for dynamical systems. Trans. Moscow Math. Soc., 19:197–231, 1968.

[Phi90] S. G. Philander. El Niño, La Niña, and the Southern Oscillation. Academic Press, Inc., 1990.

[PM93] C. Penland and T. Magorian. Prediction of Niño 3 sea surface temperatures using linear inverse modeling. J. Climate, 6(6):1067–1076, Jun. 1993.

[PM94] C. Penland and L. Matrosova. A balance condition for stochastic numerical models with application to the El Niño-Southern Oscillation. J. Climate, 7(9):1352–1372, Sept. 1994.

[PS95] C. Penland and P. D. Sardeshmukh. Error and sensitivity analysis of geophysical eigensystems. J. Climate, 8(8):1988–1998, Aug. 1995.

[PTVF92] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. Numerical Recipies in C: The Art of Scientific Computing. Cambridge University Press, 2 edition, 1992.

[RC82] E. M. Rasmusson and T. H. Carpenter. Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Niño. Mon. Wea. Rev., 110:354–384, May 1982.

[Rya97] T. P. Ryan. Modern Regression Methods. Wiley, 1997.

[SP83] J. Shukla and D. A. Paolino. The Southern Oscillation and long-range forecasting of the summer monsoon rainfall over India. Mon. Wea. Rev., 111(9):1830–1837, Sep. 1983.

[Spi95] M. R. Spiegel. Schaum's Outline of Theory and Problems of Probability and Statistics. McGraw-Hill, 1995.

[SS88a] P. S. Schopf and M. J. Suarez. Vacillations in a coupled ocean-atmosphere model. J. Atmos Sci., 45(3):549–566, Feb. 1988.

[SS88b] M. J. Suarez and P. S. Schopf. A delayed action oscillator for ENSO. J. Atmos. Sci., 45(21):3283–3287, Nov. 1988.

[Tab89] M. Tabor. Chaos and Integrability in Nonlinear Dynamics: An Introduction. John Wiley & Sons, 1989.

[TCZ95] E. Tziperman, M. A. Cane, and S. E. Zebiak. Irregularity and locking to the seasonal cycle in an ENSO prediction model as explained by the quasi-periodicity route to chaos. J. Atmos. Sci., 52(3):293–306, Feb. 1995.

[THM90] H. Tominaga, T. Horita, and H. Mori. Dynamic scaling laws for bifurcations of chaos in dissipative differential systems. Prog. Theor. Phys., 84(4):553–557, Oct. 1990.

[TL95] A. Trevisan and R. Legnani. Transient error growth and local predictability: A study in the Lorenz system. Tellus, 47A(1):103–117, 1995.

[Tro65] A. J. Troup. The "southern oscillation". Quart. J. Roy. Meteor. Soc., 91:490–506, 1965.

[TSCJ94] E. Tziperman, L. Stone, M. Cane, and H. Jarosh. El Niño chaos: Overlapping of resonances between the seasonal cycle and the Pacific ocean-atmosphere oscillator. Science, 264:72–74, Apr. 1994.

[TW98] C.. Torrence and P. Webster. The annual cycle of predictability in ENSO indices and coupled models. Submitted to Q. J. R. Meteorol. Soc., 1998.

[Val88] G. K. Vallis. Conceptual models of El Niño and the Southern Oscillation. J. Geophys. Res., 93(C11):13,979–13,991, Nov. 1988.

[VGR90] D. M. Vavriv, G. A. Gromov, and V. B. Ryabov. Interaction of chaotic and noisy oscillations in quasilinear systems. Sov. Phys. Tech. Phys., 35(11):1231–1236, Nov. 1990.

[VN94] S. Vannitsem and C. Nicolis. Predictability experiments on a simplified thermal convection model: The role of spatial scales. J. Geophys. Res., 99(D5):10,377–10,385, May 1994.

[Wei] J. B. Weiss. Private communication.

[Wri79] P. B. Wright. Persistence of rainfall anomalies in the central Pacific. Nature, 277:371–374, Feb. 1979.

[Wri85] P. B. Wright. The Southern Oscillation: An ocean-atmosphere feedback system? Bull. Amer. Meteor. Soc., 66(4):398–412, Apr. 1985.

[WSSV85] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano. Determining Lyapunov exponents from a time series. Physica D, 16:285–317, 1985.

[WW96] B. Wang and Y. Wang. Temporal structure of the Southern Oscillation as revealed by waveform and wavelet analysis. J. Climate, 9:1586–1598, Jul. 1996.

[WW97a] J. P. Weiss and J. B. Weiss. Quantifying the persistence barrier in ENSO, Part I: Theory. Submitted to J. Atmos. Sci., Dec. 1997.

[WW97b] J. P. Weiss and J. B. Weiss. Quantifying the persistence barrier in ENSO, Part II: Barriers in ENSO. Submitted to J. Atmos. Sci., Dec 1997.

[WY92] P. J. Webster and S. Yang. Monsoon and ENSO: Selectively interactive systems. Q. J. R. Meterol. Soc., 118:877–926, 1992.

[Wyr75] K. Wyrtki. El Niño — the dynamic response of the equatorial Pacific ocean to atmospheric forcing. J. Phys. Oceanogr., 5:572–584, Oct. 1975.

[XCZB94] Y. Xue, M. A. Cane, S. E. Zebiak, and M. B. Blumenthal. On the prediction of ENSO: a study with a low-order Markov model. Tellus, 46A:512–528, 1994.

[Yam96] Y. Y. Yamaguchi. Slow relaxation at critical point of second order phase transition in a highly chaotic Hamiltonian system. Prog. Theor. Phys., 95(4):717–731, Apr. 1996.

[YN93] S. Yoden and M. Nomura. Finite-time Lyapunov stability and its application to atmospheric predictability. J. Atmos. Sci, 50(11):1531–1543, Jun. 1993.

[ZC87] S. Zebiak and M. Cane. A model El Niño-Southern Oscillation. Mon. Wea. Rev., 115:2262–2278, Oct. 1987.

[ZPE93] X. Zeng, R. A. Pielke, and R. Eykholt. Chaos theory and its applications to the atmosphere. Bull. Amer. Meteor. Soc., 74(4):631–644, Apr. 1993.

Appendix A

The Barrier Measure: Issues and Implementation

A.1 Optimization

A.1.1 Implementation of the Model

Chapter 4 defined a method for measuring the persistence barrier of a timeseries. The method involved computing the persistence map, $C\,(p,l)$, of the timeseries, then performing a nonlinear least squares fit to a "model barrier," $T\,(p,l)$. Recall the definition of the barrier model from Equations (4.24-4.26):

$$T(p,l) = A\;B(p,l) + (1-A)\;e^{-\frac{l}{\tau}},$$

where

$$B(p,l) \equiv h\,(p - p_b - l,\;\Delta p) - h\,(p - p_b,\;\Delta p) + 1,$$

and

$$h(x,\;\Delta p) = \sum_{n=-\infty}^{\infty} \tanh\left(\frac{x+n}{\Delta p}\right).$$

The parameters, A, τ, Δp, and p_b, for which the model best fits $C\,(p,l)$ define the quantitative measures of a persistence barrier.

There are several issues surrounding the computation of $T\,(p,l)$. First, one must clearly truncate the infinite sum in $h\,(x,\;\Delta p)$. Limiting the sum to the bounds $\pm\mathrm{Int}(\frac{20}{\Delta p})$ produces "errors" of $\mathcal{O}(10^{-16})$, where Int (x) is the function returning the integer part of x. That is, using additional terms from the infinite sum changes h by less than $\mathcal{O}(10^{-16})$, the precision of floating point numbers on most modern computers. All computations of the

barrier measures reported in this work make use of the $\pm\text{Int}(\frac{20}{\Delta p})$ truncation.

Section 3.3.4 discussed the usefulness of rescaling the parameters of a nonlinear model to ease computation. This can be especially important when one uses a gradient method, such as the Levenberg-Marquardt algorithm, to perform the minimization. Recall from section 3.1 that gradient methods require the derivatives of the model with respect to its parameters. Such derivatives of $T(p,l)$ are nontrivial expressions. Additionally, the model is only valid for $A \in [0,1]$ and τ, $\Delta p \geq 0$. To implement these limits in a natural way while simplifying the derivatives of $T(p,l)$ with respect to the parameters, let us make the following parameter rescalings:

$$\begin{aligned} \tilde{\alpha} &= \arccos(2A-1), \\ \tilde{\rho} &= \frac{1}{\sqrt{\tau}}, \\ \tilde{v} &= \frac{1}{\sqrt{\Delta p}}. \end{aligned}$$

All computations of the barrier measures use these rescalings during the Levenberg-Marquardt phase of the regression. They are **not** used, however, during the genetic-algorithm phase or during confidence interval estimation.

A.1.2 The Genetic Algorithm

To perform the nonlinear least-squares fit of the model barrier, $T(p,l)$, to the data map, $C(p,l)$, this research uses the two-stage regression described in section 3.2.3. The second stage is the MINPACK library routine `lmder.f`, an implementation of the Levenberg-Marquardt gradient method (MGH80).[1] The first stage is the genetic algorithm [GA] described in sections 3.2.2 and 3.2.3. As noted in section 3.2.3, the GA used in this research implements crossover as defined in Equations (3.2.2) and (3.2.3), with $0 \leq w \leq 1.49$. Its other properties include:

- $N_{\text{iters}} = 100$,
- $N_{\text{pop}} = 40$,
- $N_{\text{elite}} = 3$,

[1] The MINPACK library is available from NETLIB: `http://www.netlib.org/`.

- $N_{\text{pop}_0} = 5N_{\text{pop}}$,
- $N_{\text{breed}} = \frac{1}{2}N_{\text{pop}}$,
- $\mu = 0.30$. [This is **not** the mutation rate; see below.]

Refer to sections 3.2.2 and 3.2.3 for the meaning of these numbers. All of these values are empirical; they permit the GA to locate the neighborhood of the global minimum well while computing the model as few times as possible.

The GA implements mutation in a rather peculiar and nonstandard way that nevertheless works well for the problem at hand. The quantity called "μ" is not a mutation rate, as mentioned above, but a control parameter thereof. To perform mutation, the GA randomly chooses a "non-elite gene" $\mu \cdot N_{\text{pop}}$ times for each gene. The actual algorithm is as follows:

(1) Loop through all genes [i. e. parameters], beginning with the first. At the i^{th} gene:

  (a) Loop $\mu \cdot N_{\text{pop}}$ times. Each time through the loop:

    (i) Randomly choose an individual from the non-elite members of the population.

    (ii) Replace the i^{th} gene of that member with a new, random value.

  (b) Repeat.

(2) Move to the $(i+1)^{\text{th}}$ gene. Repeat.

Although this process of altering genes happens $\mu \cdot N_{\text{pop}} \cdot N_{\text{parameters}}$ times, it does not describe the total number of mutated individuals in the population. After all, the same individual may be randomly chosen multiple times. Even the same gene in the same individual may change multiple times. So, as the GA iterates, the mutation rate changes randomly from generation to generation. This may not be the best way to implement a random mutation rate; however, it is the simplest to code. It furthermore works well for fitting the barrier model to persistence maps from data.

A.1.3 Special Handling

There are some additional computation issues related to neither the implementation of the model nor the workings of the genetic algorithm. The first issue is related to an empirically seen behavior of the `lmder.f` routine. Recall from section 4.3.2 that the model barrier, $T(p, l)$ changes very little for $\Delta p > 1$ year. Consequently, the least-squares error changes little when Δp is in this range. Once $\Delta p > 1$ year, the Levenberg-Marquardt algorithm simply keeps increasing it, leading to values of Δp often greater than $\mathcal{O}(10^3)$. This renders interpreting the results difficult at best. To remedy the situation, any values[2] of $\Delta p > 4$ years are reduced according to the following "schedule:"

$$\{1000,\ 100,\ 50,\ 25,\ 10,\ 9,\ \ldots\ ,\ 2,\ 1,\ 0.9,\ \ldots\ ,\ 0.2,\ 0.1,\ 0.01,\ 0.001\}\,.$$

As long as the least-squared error, χ^2, remains smaller than the error tolerance originally specified to the Levenberg-Marquardt algorithm, Δp is reduced further.[3] In this way, the value of Δp remains small without altering the degree of fit between the model and the data.

There are two additional issues surrounding the computation of confidence intervals, computations performed using Equation (3.16) from section 3.3.3. Section 3.3.4 mentioned these issues, both related to the curvature matrix. The curvature matrix, $\mathbf{D}$, contains the second partial derivatives of χ^2, the least-squared error, with respect to the parameters. Because we have an analytical form of the barrier model, $T\,(p, l)$, we could compute $\mathbf{D}$ in analytic closed form. This is not, however, the best way to compute $\mathbf{D}$.

Consider the function $f\,(x) = \tanh\left(\frac{x-p_b}{\Delta p}\right)$, a function that also describes a cross-section through $T\,(p, l)$. As seen in Figure A.1, changes in Δp and p_b smaller than the data resolution do not alter χ^2. Because of this "dependence" between the data resolution and Δp and p_b, the curvature matrix will be locally flat over intervals the size of the data resolution. The curvature matrix will therefore be singular due to zero diagonals.

The definition of confidence intervals, as seen in section 3.3.1, assumes non-flat curvature in the vicinity of the minimum. That vicinity needn't be infinitesimally local, however. In fact, Eq. (3.16) will certainly work using finite-difference estimates of the derivatives to

[2] This limit is partially empirical and partially related to machine precision.
[3] This error tolerance is 10^{-12}.

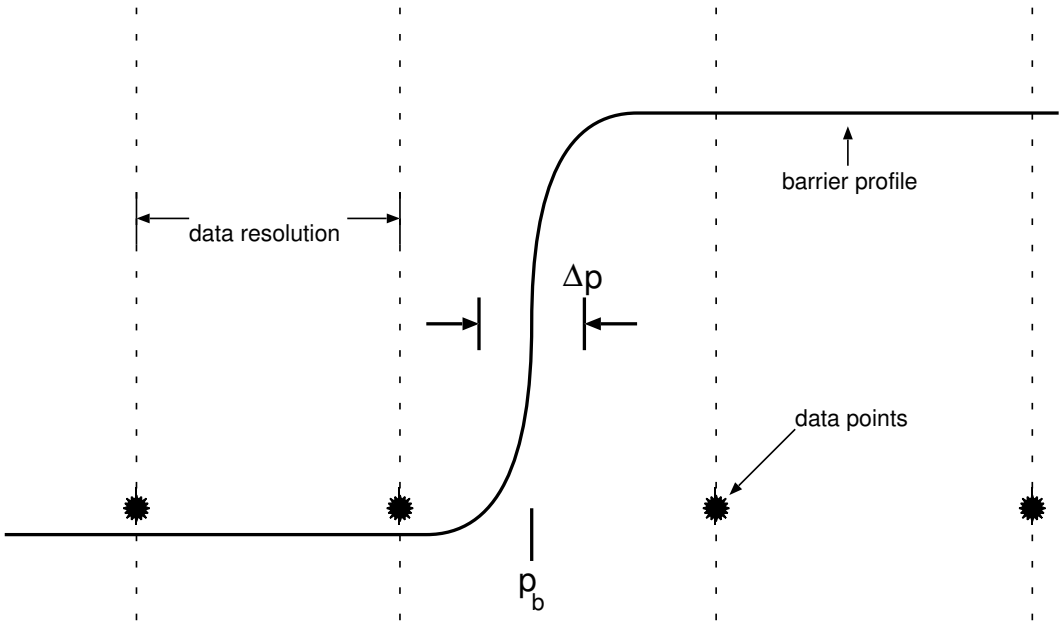

Figure A.1: Schematic of the effect of data resolution on the curvature matrix.

compute the curvature matrix. The curvature along Δp and p_b **is** nonzero over scales larger than the resolution of the data. Therefore, the algorithm for computing confidence intervals uses finite-difference derivatives for computing $\mathbf{D}$. The stepsize of the finite-difference derivative is initially half of the data resolution. The algorithm doubles this stepsize dynamically until the curvature matrix is nonsingular or, in the case of A, Δp and p_b, until the stepsize is larger than one. If the stepsize is smaller than one, but extends beyond one of the bounds, several things happen. For p_b, the algorithm wraps the lower end of the step. For A, τ, and Δp the algorithm performs an approximation, assuming the derivative is symmetric and computing appropriately.

When A is near 0 or 1, Δp and p_b or τ, respectively, are undefined. There is little sense in computing the full curvature matrix under either of these cases. Thus, the algorithm that computes confidence intervals only calculates a submatrix of $\mathbf{D}$ corresponding to the defined parameters when A is within $\pm 10^{-4}$ of 0 and 1. Only that submatrix is inverted and used in Eq. (3.16). This eliminates almost all of the cases of singular curvature matrices, permitting the entire "barrier measuring" algorithm to run unattended.

A.2 Persistence and the Pendulum

One of the key techniques of chapter 6 is the use of "parameter-space maps." One integrates the dynamical system and computes such quantities as winding number or global Lyapunov exponent, varying two of the system's parameters over a range of values. The resulting map describes the behavior of the system through a two-dimensional slice of its parameter space. Chapter 6 uses maps of winding number, the so-called Arnol'd maps, maps of the largest global Lyapunov exponent, and maps of the persistence barrier measures.

Unlike the persistence maps considered in chapter 4-5, persistence maps of dynamical systems may not have any connection to the barrier model. There may be phase-dependence in the persistence map, but a dependence that does not match the barrier model. Additionally, some timeseries from dynamical systems may have a unit period, in which case the persistence map is undefined. Finally, the system may change over timescales so large that the persistence is flat and near unity. The code therefore performs several checks before computing barrier measures.

The first of these checks is of the timeseries itself. First, the cyclostationary variance defined in Equation (4.2), $\sigma^2(p)$, is computed. If $\sigma^2(p)$ is "consistently small," i. e. near zero for most or all values of p, then the timeseries most likely has an annual period and the persistence will be undefined. Clearly, computing barrier measures is meaningless in such cases.

An average value, $\overline{\sigma^2(p)}^p$, of the cyclostationary variance should reasonably describe whether or not $\sigma^2(p)$ is "consistently small." If $\overline{\sigma^2(p)}^p$ is below some threshhold, the persistence is assumed to be undefined and the barrier measures are assigned a set of constant values. For the threshhold, I use the standard error on the **full** variance, σ^2. The statistical standard error on a variance is proportional to that variance, hence a standard error on $\sigma^2(p)$ is useless for comparison purposes.

One may ask why I don't use the standard error on the mean of $\sigma^2(p)$ over p. The standard error on a mean describes errors due to finite sample size, errors **in that mean**, not in the cyclostationary variance, $\sigma^2(p)$. Extend this concept further: the standard error on the full variance [not the cyclostationary one] describes errors in that variance due to finite sample size. Such errors will, naturally, be smaller than the equivalent errors in the cyclostationary variance, since the full variance uses a larger sample [i. e. the full data set]. Clearly, if $\sigma^2(p)$ is always smaller than the standard error on the full variance, it is indistinguishable from sample-size errors.

Although the reasoning behind the aforementioned test for annual timeseries is long, the numerical implementation is quite simple. One simply computes[4] and compares $\overline{\sigma^2(p)}^p$ to $\frac{\sigma^2}{\sqrt{2N_{\text{data}}}}$. This also executes quickly, hence the reason for using it.

If the persistence is defined, the code performs one additional check. First, the mean-squared error between the persistence map and a matrix with all elements equal to 1 is computed. If that mean-squared error is smaller than the error tolerance used by the Levenberg-Marquardt algorithm, a fixed set of barrier measures and confidence intervals is returned. The values returned are based on the results that the minimization algorithm would produce. This check for flatness is far faster than a full least-squares fit to the model. Both this and the aforementioned check noticeably shorten the run time of the "parameter-

[4]Note that $\overline{\sigma^2(p)}^p \neq \sigma^2$! The full variance is computed about the **full** mean, whereas the cyclostationary variance is computed about the **cyclostationary** mean.

space mapper" code.

Appendix B

The Pendulum: Special Cases

B.1 Integration of Special Cases

To gain some intuition into the behavior of the damped forced pendulum in general, one should examine limiting cases. The special cases of Equations (6.6-6.8) come in 2 categories: "frictionless" and "gravityless." These correspond to $\beta = 0$ and $\gamma = 0$, respectively, each of which must be integrated in a slightly different fashion. In the three cases presented below, ϕ_0 and p_0 denote the initial conditions of the system.

Case 1: $\gamma = 0$ and $\beta = 0$

This case is trivial to integrate, and in some sense really isn't a pendulum. It is presented for completeness:

$$\begin{aligned} \phi(t) &= \phi_0 + p_0 t + \frac{1}{2} D_0 t^2 + \frac{D_1}{4\pi^2} \left[1 - \cos(2\pi t)\right], \\ p(t) &= p_0 + D_0 t + \frac{D_1}{2\pi} \sin(2\pi t). \end{aligned} \tag{B.1}$$

Case 2: $\gamma = 0$ and $\beta \neq 0$

While this case also does not really describe a pendulum, the results provide interesting information about the relationship between the winding number of the pendulum and the parameters β and D_0. Using the transform $\hat{p} = p \cdot e^{\beta t}$ to integrate Eqs. (6.6-6.8),

$$\begin{aligned} \phi(t) &= \phi_0 + \frac{p_0}{\beta} - \frac{D_0}{\beta^2} + \frac{D_0}{\beta} t + \left[\zeta - \frac{p_0}{\beta} + \frac{D_0}{\beta^2}\right] \cdot e^{-\beta t} \\ &\quad + \zeta \left[\frac{\beta}{2\pi} \sin(2\pi t) - \cos(2\pi t)\right], \\ p(t) &= \frac{D_0}{\beta} + \left[p_0 - \frac{D_0}{\beta} - \beta\zeta\right] \cdot e^{-\beta t} + \zeta \left[\beta \cos(2\pi t) + 2\pi \sin(2\pi t)\right], \end{aligned} \tag{B.2}$$

where $\zeta = \frac{2\pi D_1}{4\pi^2+\beta^2}$.

Case 3: $D_1 = 0$ and $\beta = 0$

When $\gamma \neq 0$, Eqs. (6.6-6.8) becomes far more difficult to integrate, due to the highly nonlinear $\sin(\phi)$ term. One can find the complete integration of the standard pendulum, for which $\beta = D_0 = D_1 = 0$, in many texts (BG90, Tab89). In these derivations, one makes use of an integral of the motion, the total energy of the system, to compute $\phi(t)$ in terms of Jacobi elliptic integral functions. Interestingly enough, an integral of the motion also exists for $D_0 \neq 0$. By multiplying Eq. (6.7) by ϕ, one can show that

$$\frac{d}{dt}\left[\frac{1}{2}p^2 - D_0\phi - \gamma\cos(\phi)\right] = 0,$$

from which one defines the integral of the motion,

$$\begin{aligned} \hat{E} &\equiv \frac{1}{2}p^2 - D_0\phi - \gamma\cos(\phi) = \text{constant} \\ &= \frac{1}{2}\dot{\phi}^2 - D_0\phi - \gamma\cos(\phi). \end{aligned} \tag{B.3}$$

Note that (B.3) reduces to the total energy of the system for $D_0 = 0$.

One can use (B.3) to solve for $\phi(t)$. After some algebra, one finds:

$$t\sqrt{2} = \int_{\phi_0}^{\phi} \frac{d\phi'}{\sqrt{\hat{E} + D_0\phi' + \gamma\cos(\phi')}}. \tag{B.4}$$

The author has made several attempts to integrate this, including coordinate transforms of ϕ, without success. If one instead lets $\mathcal{I}(\phi)$ be the integral on the right-hand side of (B.4), evaluated from 0 to ϕ, one can write the solution as $\phi(t) = \mathcal{I}^{-1}\left(t\sqrt{2} + \mathcal{I}(\phi_0)\right)$. A stability analysis, presented in the next section, provides a clearer description of the motion hidden in (B.4).

B.2 Stability Analysis at $D_1 = 0$

When one cannot integrate the equations of motion, one can use a linear stability analysis of the system to learn something about the dynamics (Tab89, NL92, BG90). As noted in section 6.1.1, one computes the eigenvalues of the Jacobian matrix of the system,

$\mathbf{J}(\vec{x})$, evaluated at the fixed points, $\vec{x}^\star$. The signs of the real and imaginary parts of each eigenvalue then describe a particular type of motion near $\vec{x}^\star$, with a nonzero imaginary portion denoting some form of rotation around the fixed point, and a negative or positive real portion indicating a stable or unstable fixed point, respectively (Tab89, NL92, BG90). In combination, the eigenvalues may indicate more complicated behavior, such as the "hyperbolic" or "saddle" point, where the real part of one eigenvector is positive and that of another is negative. Some examples are sketched in Figure B.1. A complete list of eigenvalue types and their corresponding dynamics can be found in Tabor (Tab89).

It is possible to perform a stability analysis for the case of only $D_1 = 0$. The system is autonomous and has equations of motion described by:

$$\dot{\phi} = p, \tag{B.5}$$

$$\dot{p} = -\beta p - \gamma \sin(\phi) + D_0. \tag{B.6}$$

The Jacobian matrix is thus

$$\mathbf{J}(\vec{x}) = \begin{pmatrix} 0 & 1 \\ -\gamma\cos(\phi) & -\beta \end{pmatrix}, \tag{B.7}$$

and the fixed points, $\vec{x}^\star$, are

$$\phi^\star = \arcsin\left(\frac{D_0}{\gamma}\right) \pm 2\pi k, \quad \pm(2k+1)\pi - \arcsin\left(\frac{D_0}{\gamma}\right) \tag{B.8}$$

$$p^\star = 0, \tag{B.9}$$

where k is an integer.[1] From Eqs. (6.10) and (6.11), $D_0 > 0$ and $\gamma > 0$, since negative values are equivalent to a coordinate transformation.

As noted earlier, one uses the eigenvalues of $\mathbf{J}(\vec{x})$ to perform the linear stability analysis. These are easy to compute, and have the general form:

$$\lambda_1 = -\frac{1}{2}\left(\beta - \sqrt{\beta^2 - R}\right), \tag{B.10}$$

[1] Note that $|\arcsin(y)| \leq \pi/2$, which does not describe all possible values of $\sin(x) = y$. In addition to the well-known relation $\sin(x \pm 2\pi k) = \sin(x)$, there also exists the identity $\sin(\pm(2k+1)\pi - x) = \sin(x)$, hence the second set of fixed points in (B.8).

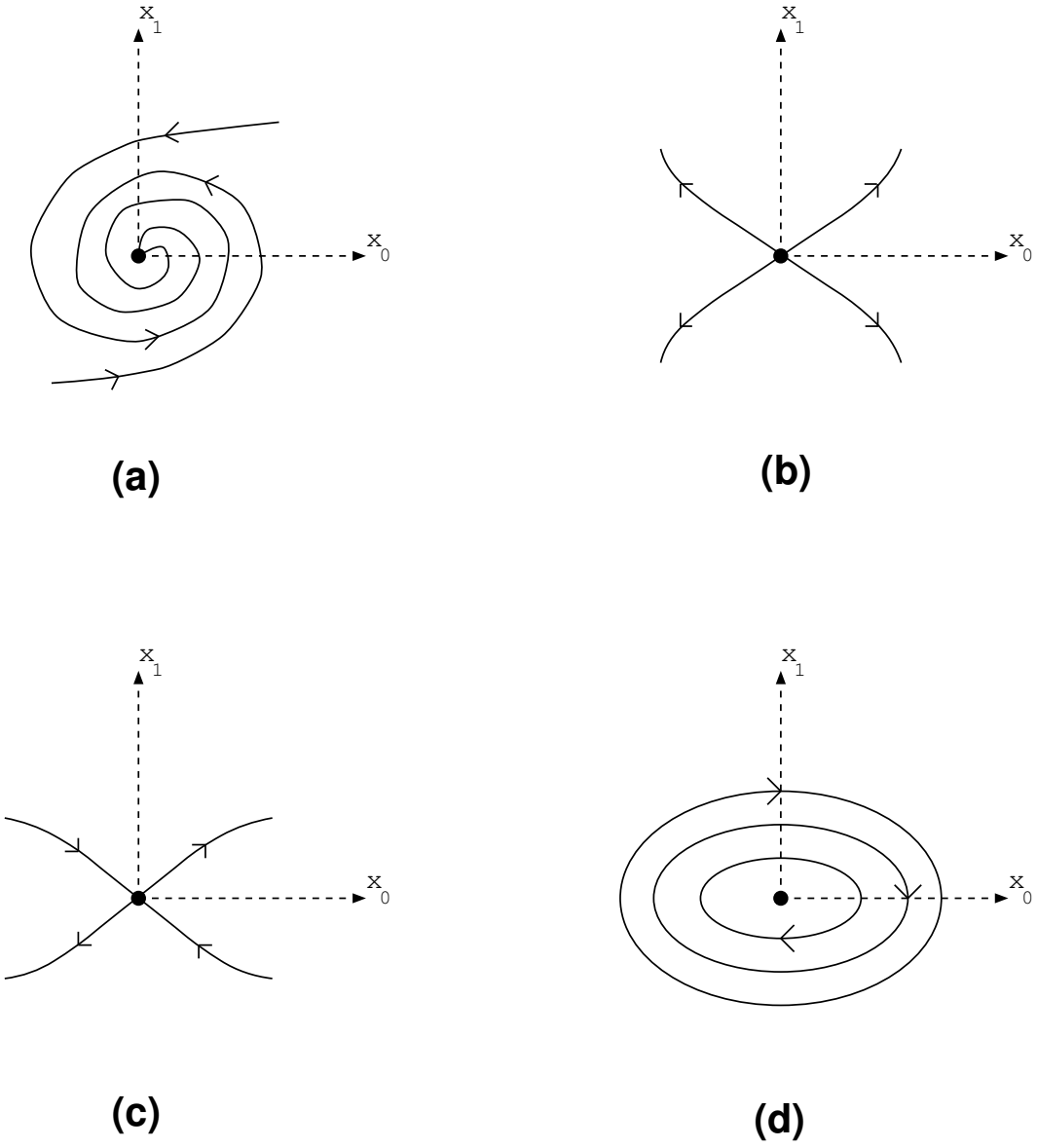

Figure B.1: Behaviors around a fixed point in a 2-D phase space, after Tabor's Figure 1.9-1.10 (Tab89). (a) A stable spiral point. Both eigenvalues are complex numbers. (b) An unstable node. (c) A saddle node. The eigenvalues are real for these two cases. (d) An elliptic point. Both eigenvalues are pure imaginary numbers.

$$\lambda_2 \quad = \quad -\frac{1}{2}\left(\beta + \sqrt{\beta^2 - R}\right), \tag{B.11}$$

where the fixed point at which one evaluates $\mathbf{J}(\vec{x})$ determines R. At $\phi^\star = \pm(2k+1)\pi - \arcsin\left(\frac{D_0}{\gamma}\right)$, $R = -4\sqrt{\gamma^2 - D_0^2}$. Thus both eigenvalues are always real numbers with $\lambda_1 > 0$ and $\lambda_2 < 0$. This describes a saddle-point. For $\phi^\star = \arcsin\left(\frac{D_0}{\gamma}\right) \pm 2\pi k$, one finds $R = +4\sqrt{\gamma^2 - D_0^2}$. The behavior around this fixed point depends on the values of the parameters.

Case 1: $D_0 = 0$

This particular case is performed by Tabor for both the damped and undamped pendulum (Tab89). Those results are summarized here. The strength of the damping creates three types of fixed point at $\phi^\star = \pm 2\pi k$. For $\beta = 0$, the eigenvalues λ_1 and λ_2 are pure imaginary numbers, indicating an elliptic point. The presence of damping turns this into a stable point. Specifically, $\beta^2 < R$ produces a stable spiral point, $\beta^2 > R$ describes a stable node, and $\beta^2 = R$ is a stable improper node.

Case 2: $D_0 \neq 0$ and $\gamma > D_0$

The presence of the torque represented by D_0 shifts both fixed points, $\phi^\star = \arcsin\left(\frac{D_0}{\gamma}\right)$ and $\phi^\star = \pi - \arcsin\left(\frac{D_0}{\gamma}\right)$, towards $\frac{\pi}{2}$. The behavior around both fixed points remains the same. The second fixed point is still a saddle point, as noted earlier, while the first again depends on the damping. Undamped motion is still around an elliptic point, and $\beta^2 < R$, $\beta^2 = R$, and $\beta^2 > R$ are still stable spiral, stable improper, and stable nodes, respectively.

Case 3: $D_0 \neq 0$ and $\gamma = D_0$

Both fixed points have now shifted to $\phi^\star = \frac{\pi}{2}$ and $R = 0$. Hence $\lambda_1 = 0$ and $\lambda_2 = -\frac{\beta}{2}$, which describes a neutrally stable point. The transition from two fixed points to a single fixed point indicates a saddle-node bifurcation (Tab89).

Case 4: $D_0 \neq 0$ and $\gamma < D_0$

The fixed points described by Eqs. (B.8) and (B.9) have now both ceased to exist. Whereas the pendulum could undergo libration or rotation in the previous two cases, it can now only undergo rotation.

The motion described by Equation (B.4) is now readily apparent. The integral in (B.4) is probably some cousin of the Jacobi elliptic integral functions, the inverse of which

is some periodic function. The phase space still contains a separatrix (Tab89), though it is now asymmetric. At the $\gamma = D_0$ bifurcation, the separatrix ceases to exist.

B.3 Mappings

B.3.1 Winding Numbers and Connection to Circle Map

Section B.1 examined two "gravityless" special cases of the damped forced pendulum. These two cases, specified in Equations (B.1-B.2), are more than mere mathematical exercises. By computing winding numbers for Equations (B.1-B.2), one may learn which parameters have a connection to the quasiperiodic route to chaos. One can also re-express the flows of Eqs. (B.1-B.2) as a one-dimensional mapping in the form of a circle map. Recall the definition of the circle map in Eqs. (6.2-6.3) :

$$\begin{aligned} \vartheta_{n+1} &= f(\vartheta_n) \bmod 2\pi, \\ f(\vartheta_n) &\equiv \vartheta_n + \Omega + Kg(\vartheta_n). \end{aligned}$$

The first step is to let $t = t_n = n + \delta$ where $0 < \delta < 1$. One can absorb the phase shift, δ, into the initial conditions, ϕ_0 and p_0, without loss of generality. The remainder of the derivation is quite lengthy. I report only the results here.

Equation (B.1) is the case where $\beta = \gamma = 0$. The winding number, computed using (B.1) in Eq. (6.20), comes in two cases. For $D_0 = 0$, the winding number is $W_r = p_0$, while for $D_0 \neq 0$, $W_r \to \infty$. The conversion to a one-dimensional circle map is also straightforward. One finds:

$$\begin{aligned} \Omega &= \frac{1}{2} D_0, \\ K &= p_0, \\ g(\phi_n) &= \sqrt{1 + \frac{2D_0}{p_0^2}(\phi_n - \phi_0)}, \end{aligned}$$

where $\phi_{n+1} = (\phi_n + \Omega + Kg(\phi_n)) \bmod 2\pi$ defines the mapping. The choice of K may seem odd at first. However, the square root in $g(\phi_n)$ originates in the solution to a quadratic, the sign of which must be the sign of p_0. The choice of K automatically selects the proper root.

The interesting result is the dependence of the "natural frequency," Ω, of the mapping on D_0, the constant torque.

The case of $\beta \neq 0$ and $\gamma = 0$, described by Equation (B.2), splits into two subcases, depending on the value of D_0. For $D_0 = 0$, the circle map, $\phi_{n+1} = (\phi_n + \Omega + Kg(\phi_n)) \bmod 2\pi$, has the form:

$$\begin{aligned} \Omega &= \left[\frac{p_0}{\beta} - \frac{2\pi D_1 \beta}{\beta^2 + 4\pi^2}\right] \cdot \left[1 - e^{-\beta}\right], \\ K &= \left[e^{-\beta} - 1\right], \\ g(\phi_n) &= \phi_n - \phi_0. \end{aligned}$$

The winding number for this case, a librational state, is zero when $D_1 = 0$. For $D_1 \neq 0$, computing the new winding coordinate analytically via Equations (6.21-6.22) is non trivial and the finding the analytic form of W_l [Eq. (6.23)] may be intractable.

For $D_0 \neq 0$, at any value of D_1, the pendulum is in a rotational state and the winding number is $W_r = \frac{D_0}{\beta}$. The circle map shows a similar dependence:

$$\begin{aligned} \Omega &= \frac{D_0}{\beta}, \\ K &= \frac{D_0}{\beta^2}\left[1 - e^{-\beta}\right], \\ g(\phi_n) &= h^{-1}\left(\Xi \cdot e^{\Xi - \beta[\phi_n - \phi_0]}\right), \end{aligned}$$

where

$$\Xi = \frac{\beta}{D_0}\left[p_0 - \frac{2\pi D_1 \beta}{\beta^2 + 4\pi^2}\right] - 1,$$

and

$$h(x) \equiv xe^x.$$

The winding number is actually a function of only the damping and the constant torque. Numerical experiments for the full pendulum also indicate a dependence of the winding number on the ratio $\frac{D_0}{\beta}$, as seen in section 6.2.2.

B.3.2 Stability Analysis — An Attempt

Although one cannot perform a stability analysis for the full equations of motion for the pendulum, Eqs. (6.6-6.8), there is an approximation one can perform. As in the previous section, let us make the transformation: $t = t_n = n + \delta$, where $0 < \delta < 1$, and $\dot{\phi} = \frac{d\phi}{dt} \sim \phi_{n+1} - \phi_n$. Applying similar approximations to the other two equations of motion, one converts the pendulum to the following mapping:

$$\phi_{n+1} = \phi_n + p_n, \tag{B.12}$$

$$p_{n+1} = (1-\beta)\, p_n - \gamma \sin(\phi_n) + D_0 + D_1 \cos(2\pi\delta), \tag{B.13}$$

$$\tilde{\tau}_{n+1} = \tilde{\tau}_n + 2\pi. \tag{B.14}$$

The third equation is trivial and irrelevant to the dynamics. Note the similarity between the first two equations and the standard map (Tab89).

One can perform a stability analysis, similar to that of section B.2, on Equations (B.12-B.14). The Jacobian of the mapping is

$$\mathbf{J}(\vec{x}_n) = \begin{pmatrix} 1 & 1 \\ -\gamma\cos(\phi_n) & -\beta \end{pmatrix}, \tag{B.15}$$

and the fixed points are

$$\phi^\star = \arcsin\left(\frac{\alpha}{\gamma}\right) \pm 2\pi k, \quad \pm(2k+1)\,\pi - \arcsin\left(\frac{\alpha}{\gamma}\right), \tag{B.16}$$

$$p^\star = 0, \tag{B.17}$$

where $\alpha = D_0 + D_1\cos(2\pi\delta)$. These exist if $\alpha \leq \gamma$. The eigenvalues of the Jacobian, evaluated at these fixed points, are

$$\lambda_1 = \frac{1}{2}\left(2 - \beta - \sqrt{\beta^2 + R(\phi^\star)}\right), \tag{B.18}$$

$$\lambda_2 = \frac{1}{2}\left(2 - \beta + \sqrt{\beta^2 + R(\phi^\star)}\right), \tag{B.19}$$

where, once again,

$$R\left(\phi^{\star}\right) = \begin{cases} -4\sqrt{\gamma^2 - \alpha^2}, & \phi^{\star} = \arcsin\left(\frac{\alpha}{\gamma}\right), \\ +4\sqrt{\gamma^2 - \alpha^2}, & \phi^{\star} = \pi - \arcsin\left(\frac{\alpha}{\gamma}\right). \end{cases}$$

Both eigenvalues are real for the fixed point at $\phi^{\star} = \pi - \arcsin\left(\frac{\alpha}{\gamma}\right)$. λ_2 is always positive, while the sign of λ_1 depends on the size of the parameters. Hence this fixed point could now be a saddle-node or an unstable fixed point. At $\phi^{\star} = \arcsin\left(\frac{\alpha}{\gamma}\right)$, there are numerous possible behaviors, depending on the values of all four parameters. Note, too, that both D_0 and D_1 shift the fixed points towards $\frac{\pi}{2}$, with a bifurcation at $\gamma = \alpha = D_0 + D_1 \cos\left(2\pi\delta\right)$.

At $D_1 = 0$, Equations (B.15-B.19) **do not** reduce back down to the results of section B.2. Thus, this mapping doesn't reflect the true dynamics of the system. Nevertheless, it hints at some interesting results. First, there are possibly several saddle-node bifurcations, all functions of D_0, D_1, and γ. Second, the term $D_1 \cos\left(2\pi\delta\right)$ represents a time-dependent term in the true equations of motion. Thus, there aren't actually any fixed points when $D_1 \neq 0$. Instead, the fixed points that were once at $\phi^{\star} = \arcsin\left(\frac{D_0}{\gamma}\right)$ and $\phi^{\star} = \pi - \arcsin\left(\frac{D_0}{\gamma}\right)$ are now perturbed by the driving. Consequently, the attractors that were once located at these fixed points can overlap, generating chaos via crisis. The parameter D_1, the driving amplitude, possibly induces crisis-based intermittency (GW86).

Appendix C

Parameter Space Explorations

C.1 Method of Computation

C.1.1 Integrating the Equations of Motion

This appendix begins by describing the details behind numerically integrating the two dynamical systems of chapter 6, the damped forced pendulum and the Vallis model (Val88). The integration algorithm is a modified version of the integrator from Press et. al. (PTVF92), a "4/5" Runge-Kutta integrator with adaptive stepsizing. The "4/5" indicates a fourth-order integrator with fifth-order error correction. The modifications made to the algorithm permit it to integrate a dynamical system with a toroidal phase space.

Consider the pendulum. Although the pendulum angle has the property $\phi \pm 2\pi k = \phi$, where k is an integer, using an "unwrapped" ϕ generates numerical roundoff errors as ϕ grows large. Clearly, this is disastrous for study of a chaotic system. The roundoff errors would induce an unwanted stochastic forcing. Furthermore, once ϕ is sufficiently large, the integrator will cease functioning. Let $d\phi$ be the step taken by the integrator. If $\mathcal{O}(\phi) - \mathcal{O}(d\phi)$ is larger than machine precision, the calculation $\phi \pm d\phi$ will return ϕ. Thus, one must map any ϕ outside of the interval $[-\pi, \pi]$ back into this interval.

The custom portion of the Runge-Kutta integrator maps any periodic variables back into their range. It is usable on systems with multiple periodic variables, not just on the damped forced pendulum. Furthermore, this "wrapping" integrator keeps track of the numbers of times and the directions of the wrapping. One can therefore "unwrap" the phase space for later uses, such as the computation of the winding number.

Let us consider the units of time for each model to be years. All integrations in

this thesis use a resolution of 120 points per model year. Using the previous data point as its initial condition, the integrator evolves the equations of motion for $\frac{1}{120}$ years, thereby generating the next data point. The error tolerance for all integrations is 10^{-9}. For this error tolerance, a maximum number of adaptive steps set to 1000 per data point works well for resolutions as large as 12 points per year.[1]

C.1.2 Lyapunov Exponents

Chapter 2 covered most of the issues surrounding the calculation of both the global and instantaneous Lyapunov exponent spectra, or GLE and ILE, respectively. All calculations of the GLE use the algorithm described in sections 2.2.3-2.2.5. With a slight modification, one can use the Lyapunov eigenvectors from the GLE algorithm to compute the ILE via Equation (2.48).

For both the GLE and ILE, let us integrate an orthonormal basis using the linearized equations for perturbations [Eq. (2.2)], integrating them simultaneously with the equations of motion. After each integration from one data point to the next, using the method of the last section, one performs several computations. The evolved perturbations are orthonormalized via Gram-Schmidt orthonormalization. The ILE at that point in phase space are computed and stored. The current value of the GLE spectrum is also computed and stored. At the end of the run, one has a timeseries of values, one for the ILE, one for the GLE.[2]

For the GLE, the timeseries of values displays the convergence of the spectrum at time increases. The reported values of the GLE are actually the average of the "convergence series" over the final 10% of its length. Accompanying these averages are the corresponding standard deviations, a useful set of numbers for determining how well the GLE have indeed converged. Another useful check is a comparison of the sum of the GLE to the dissipation of the system. For both the pendulum and the Vallis model (Val88), the trace of the Jacobian is constant, and therefore describes the dissipation of the system [Equations (2.21) and (2.49)] at all times. Typically, I find that the sum of the GLE and the dissipation differ by about $\mathcal{O}\left(10^{-9}\right)$, the RK error tolerance. The standard deviations are usually on the order of $\frac{1}{\text{run length}}$.

[1] See Press et. al. (PTVF92) for further information about the purpose and function in the algorithm of the maximum number of adaptive steps.

[2] In practice, I use two separate programs for the ILE and GLE.

In the case of the ILE spectrum, one can also compare the sum of the spectrum at each point in phase space to the trace of the Jacobian. Empirical studies for many parameter values showed no deviation from the dissipation greater than $\mathcal{O}\left(10^{-9}\right)$. One can also integrate the ILE spectrum timeseries using standard techniques (PTVF92) and thereby estimate the GLE spectrum [see Eq. (2.40)]. Comparing the GLE computed in this way to the dissipation of the system show a typical "error" of $\mathcal{O}\left(10^{-7}\right)$. Note that, in this work, the GLE has units of bits/year, while the ILE uses units of 1/year.

For all runs, the computation begins with an alignment integration of 100 model years. This integrates away all transients, permits the system to seek its attractor, and allows the initial orthonormal basis to align with the Lyapunov eigenvectors. The "production run" is always 10,000 model years in length. Both this value and the choice of a 120 points/year resolution come from an empirical study of the errors in the Lyapunov exponents. The run length is sufficiently large enough to permit the GLE spectrum to converge. Also, these are the largest resolution and shortest run length for which estimations of the GLE from the ILE are accurate. That is, the integration of each ILE timeseries matches the corresponding GLE computed directly.

C.1.3 Winding Number

The definition of a winding number for the Vallis model (Val88) is not clear. In the case of the damped forced pendulum, there are two possible definitions, depending on the state of the system. To determine numerically if the system is on the librational or rotational 2-torus [section 6.2.2.1], consider Equation (6.20). If the system undergoes libration, W_r, the winding number for the rotational 2-torus, should be zero to within the standard error on the mean. Therefore, let us compute the standard error on the mean of ϕ, the winding angle for the rotational 2-torus. If W_r is smaller than the standard error on $\overline{\phi}$, assume that the system is on the librational 2-torus and use W_l [Eq. (6.23)]. If this is not the case, W_r is the proper winding number.

C.1.4 Persistence Maps

The persistence maps and measures are computed in the standard fashion. There is one additional detail to the computation. Calculating the persistence for 1,200,000 data

points, then fitting the model barrier to a 120 by 120 persistence map, is computationally prohibitive. To speed calculations, I subsample the timeseries every 5 points, beginning with point 6. This is analogous to collecting monthly data on the 15^{th} day of each month. This reduced-resolution timeseries now has only 24 points/year, greatly speeding calculations.

One issue remains: which variable, ϕ or p, should one use to compute the persistence? The pendulum angle, ϕ, is periodic. Thus, its transitions from $-\pi$ to π and vice versa will introduce an additional frequency component. The pendulum momentum, p, has no such difficulties, yet still describes the dynamics of the system. Therefore, let us compute the persistence map of p [the pendulum momentum].

C.2 Tableaus: Damped Forced Pendulum

This section contains an exploration of the damped forced pendulum's behavior, using Equations (6.6-6.8) from chapter 6. The results of this exploration are images of two-dimensional cross-sections through parameter space. Images of the winding number, the sign of the largest global Lyapunov exponent [GLE], and the barrier measures form a tableau for a given set of parameters. The tableaus are orthogonal to the axes of parameter space, i. e. only two of the parameters in Eqs. (6.6-6.8) vary, with the other two at fixed values.

The first image in each tableau examines phase-locking, using both winding number and the largest GLE. Both are computed using the methods described in the previous section. The color of each gridpoint represents winding numbers, W, with a particular rational value. W is considered rational if $\frac{P}{Q} - 10^{-4} \leq W \leq \frac{P}{Q} + 10^{-4}$, where P and Q are integers between 1 and 10. If the winding number of a gridpoint does not meet this criteria, it is set to the background color. The resulting image is a type of Arnol'd map. Because winding numbers greater than 1 are equivalent to those between 0 and 1, I first compute $W \bmod 1$ before plotting the Arnol'd map.

For the GLE, the associated standard deviation [described in section C.1.2] is subtracted from the largest GLE. Doing so provides a conservative estimate of when the largest GLE is positive and, hence, when the system is chaotic. If this "adjusted" largest GLE is between -10^{-4} and 10^{-4}, it is set to zero. If the largest GLE, processed as described, is positive at a particular gridpoint, a "+" is plotted there. If it is zero, a "$\diamond$" is plotted over

the gridpoint. For negative values, no symbol is plotted atop the gridpoint.

Occasionally, a gridpoint with irrational W may not contain a "$\diamond$". Also, a "+" will often be on a gridpoint with rational winding number. These are most likely due to the criteria used to deem W irrational and determine the sign of the largest GLE. The winding number most likely has an irrational value, but one within $\pm 10^{-4}$ of a rational number.

The next four images of the tableau contain the barrier measures, computed as described in section A.2. The color of a gridpoint represents the value of that measure. For the stochastic decay time, τ, values greater than 100 years are truncated. Similarly, values of the barrier width, Δp, are truncated at 1 year. Gridpoints set to the background color have a confidence interval larger than that measure's range of valid values.[3] They are also set to the background color if the nonlinear correlation coefficient, R^2, defined in chapter 3, is less[4] than 0.98.

There is one lone exception to this description: period-1 states. Recall from section A.2 that timeseries with an annual period are given a stock set of measures. For the barrier position and width, p_b and Δp, respectively, the confidence intervals are set to 10^{30} and therefore appear as the background color in their respective maps. The barrier amplitude, A, is set to a value of 10^{-8}, while the decay time, τ, is assigned 10^{30}. Thus, any gridpoint for which A is zero, τ is 100, and p_b and Δp, are the background color indicates a period-1 state.

The parameter space of the damped forced pendulum was examined in 15 regions. Three of them are uninteresting and therefore only described verbally. Accompanying the other 12 tableaus are listings of the parameter values for that set followed by a brief description of the images. The parameter values are **exact**, as this is a chaotic system and any rounded values may produce different results. The actual grid is computed using steps of $\frac{\mu_{\max}-\mu_{\min}}{N_{\text{steps}}-1}$, where $\mu_{\max}$ and $\mu_{\min}$ are the upper and lower bounds, respectively, of one of the pendulum parameters [represented here by a "μ"].

Period-1 Regions [not shown]

For the regions:

- γ varies from 39.47841760436 to 631.6546816697 for 21 steps,

[3] In the case of τ, any confidence interval larger than τ itself fulfills this criteria.
[4] This cutoff is not arbitrary. Section 6.2.3 discusses the reason for this choice.

- D_1 varies from 1 to 78.9568352087 for 21 steps,
- $D_0 = 0$ and $\beta = 1.570796326795$ or $\beta = 6.283185307180$,

and the region:

- β varies from 1.570796326795 to 6.283185307180 for 21 steps,
- D_1 varies from 1.5791367041742002 to 78.9568352087 for 51 steps,
- $D_0 = 0$ and $\gamma = 39.47841760436$,

the system is period-1 except for a small number of points. These regions were an attempt to reproduce the state-space diagram of D'Humieres et. al. (DBHL82) with the parameterization used in this work. Converting between the parameter sets using Equations (6.16-6.19) mapped the two varying parameters $\hat{\nu}$ and $\hat{D}_1$ onto β, γ, and D_1. Most likely, the two-parameter cross sections I chose did not capture any interesting dynamics.

Tableau #1: Figures C.1 and C.2

- β varies from 1.570796326795 to 6.283185307180 for 21 steps,
- D_1 varies from 25.266187266779998 to 1263.309363339 for 51 steps,
- $D_0 = 0$ and $\gamma = 631.6546816697$.

This set of parameters also came from an attempt to reproduce the state-space diagram of D'Humieres et. al. (DBHL82) with the parameterization in Equations (6.6-6.8). This particular choice does show something interesting, a period-1 region and a chaotic region. The barrier amplitude, A, is below 0.2 in the chaotic region, with the strongest A occurring in regions of large β and D_1. The width, Δp, is consistently near zero in the chaotic region, while the stochastic decay time and barrier position, p_b and τ, contain a mosaic of constant values. The p_b values are between either 0.3 and 0.5 yr. or 0.85 and 1.0 yr., while $0.1 < \tau < 1.0$ yr.

Tableau #2: Figures C.3 and C.4

- γ varies from 54.8311355616 to 164.4934066848 in 6 steps,
- D_1 varies from 67.19884649888 to 109.6622711232 in 85 steps,

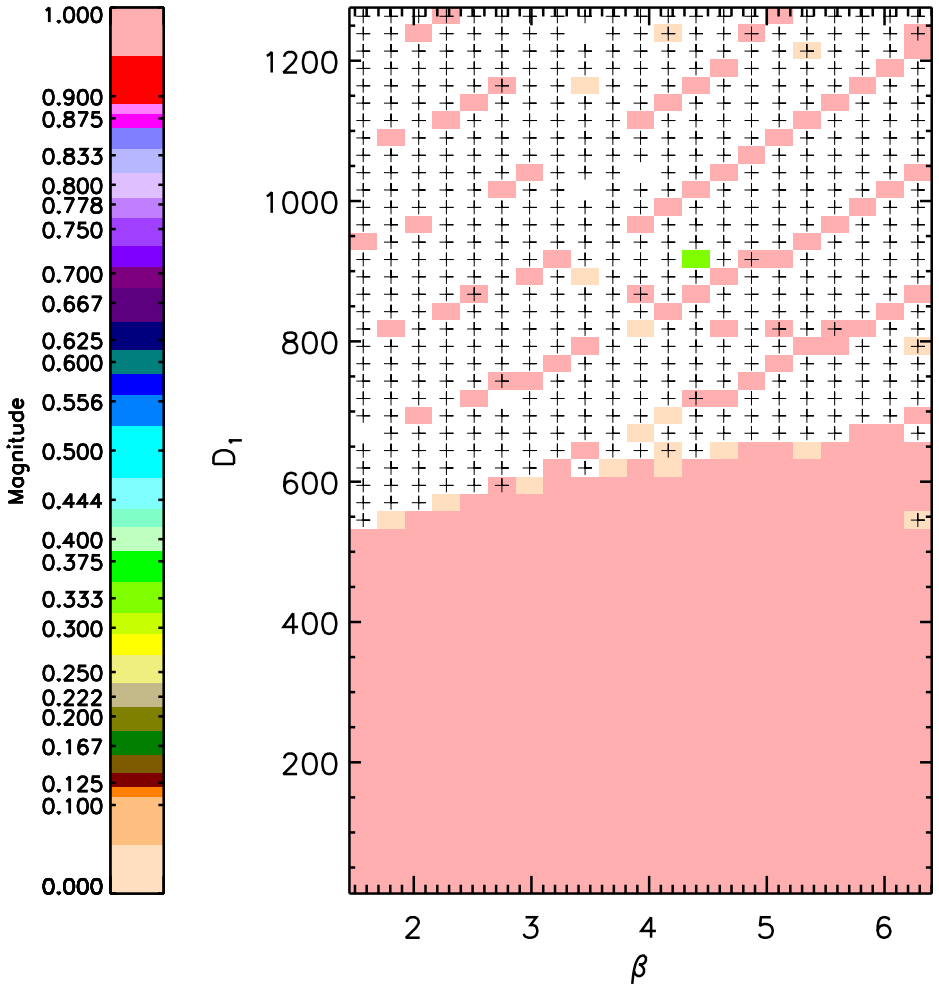

Figure C.1: Arnol'd map for the damped forced pendulum, with $D_0 = 0$ and $\gamma \approx 631.65$ [Tableau #1]. "+" indicates positive GLE; "◇" indicates zero GLE. Gridpoint set to the background color indicate zero winding number. A detailed description of the plotting method is in section C.2.

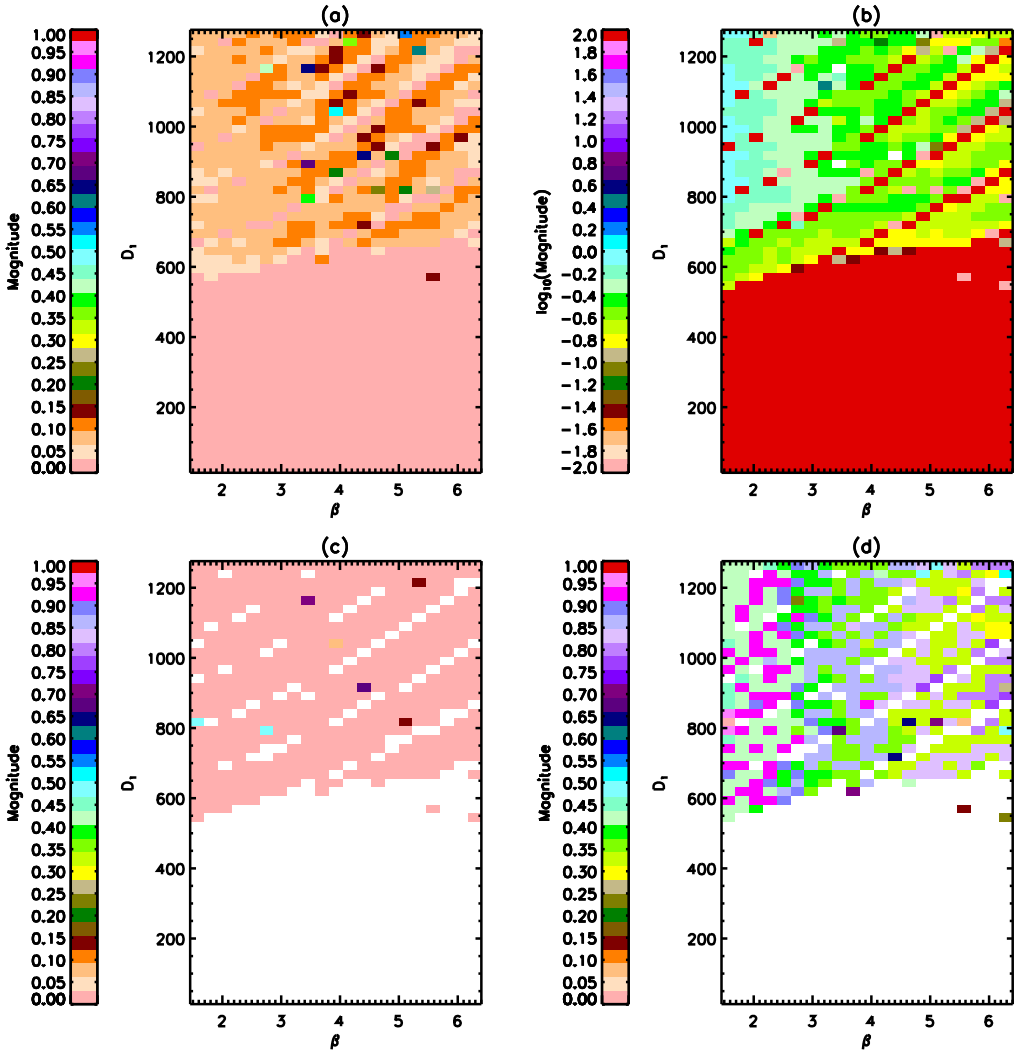

Figure C.2: The barrier measures for Tableau #1 [$D_0 = 0$ and $\gamma \approx 631.65$]. A detailed description of the plotting method is in section C.2. (a) The barrier amplitude, A. (b) $\log_{10}(\tau)$, the stochastic decay time displayed on a logarithmic scale. (c) The barrier width, Δp. (d) The barrier position, p_b.

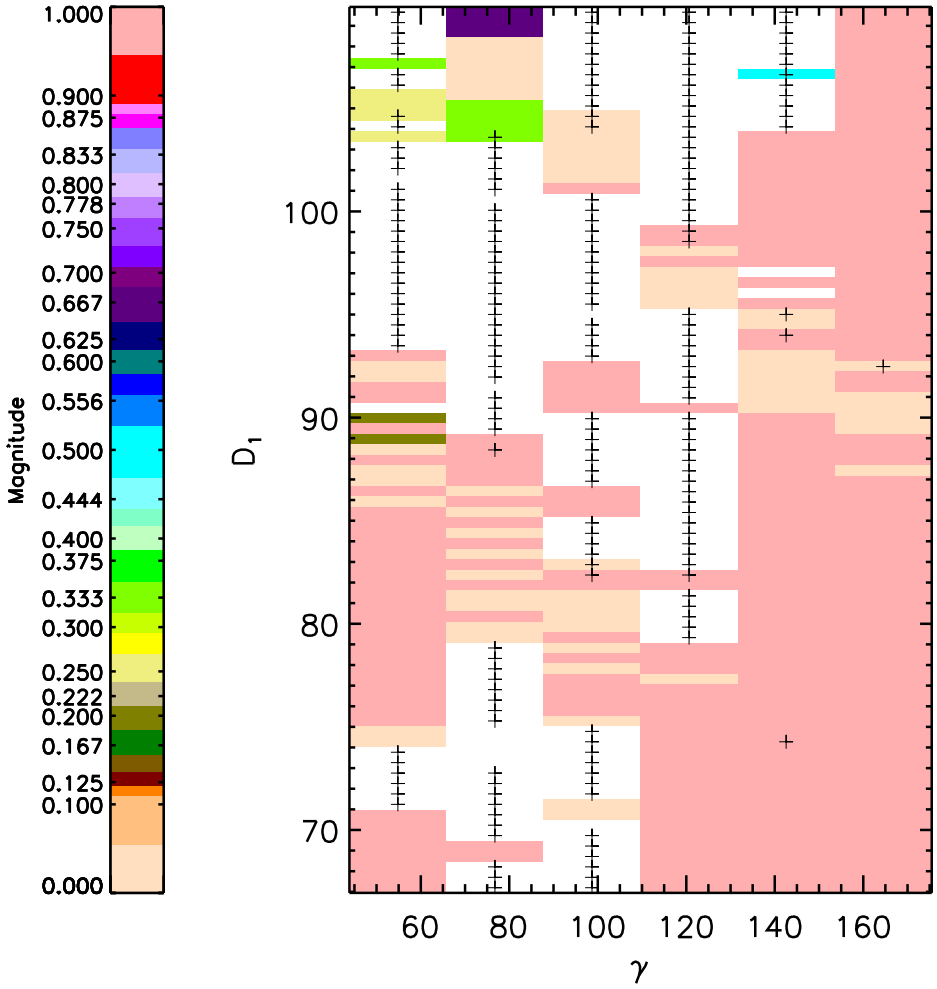

Figure C.3: Arnol'd map for the damped forced pendulum, with $\beta \approx 2.09$ and $D_0 = 0$ [Tableau #2], plotted in the same fashion as Figure C.1. A detailed description of the plotting method is in section C.2.

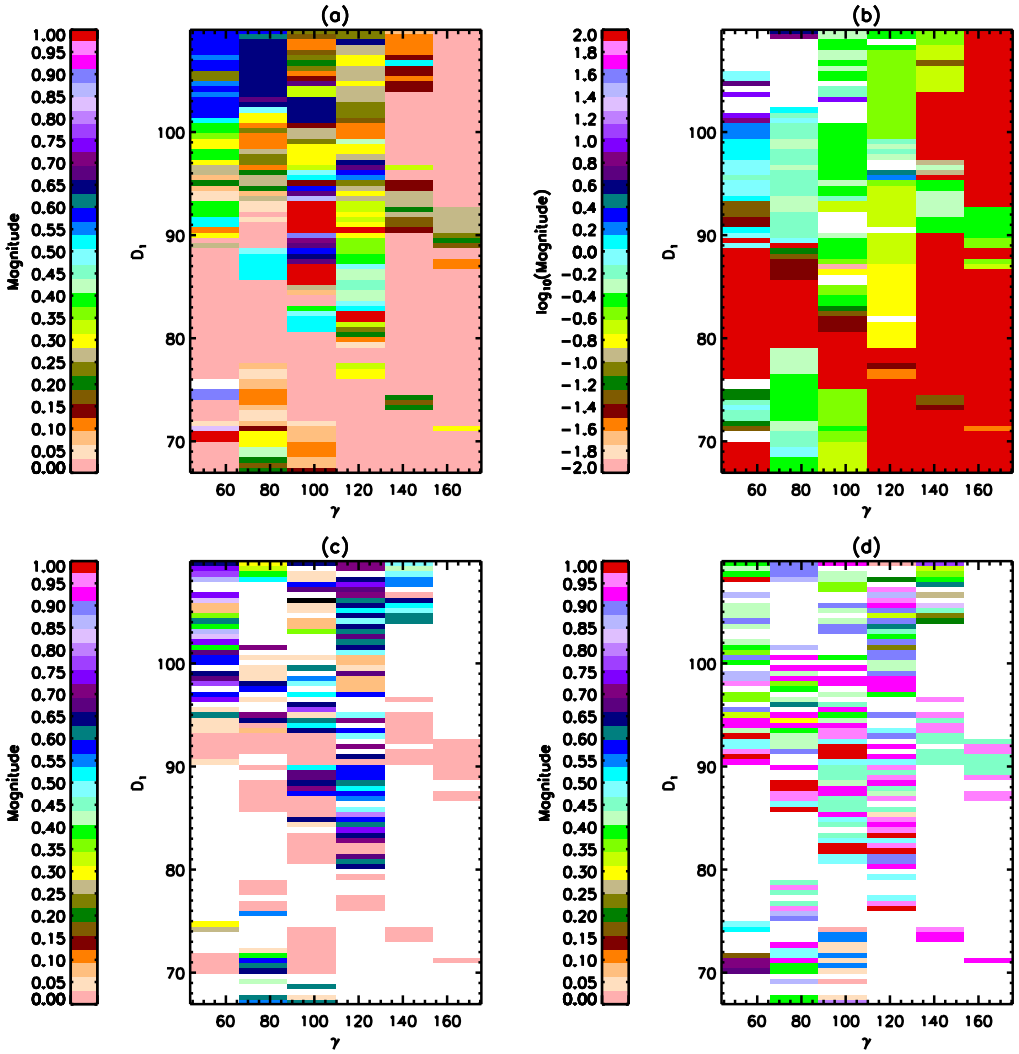

Figure C.4: The barrier measures for Tableau #2 [$\beta \approx 2.09$ and $D_0 = 0$]. A detailed description of the plotting method is in section C.2. (a) The barrier amplitude, A. (b) $\log_{10}(\tau)$, the stochastic decay time displayed on a logarithmic scale. (c) The barrier width, Δp. (d) The barrier position, p_b.

- $\beta = 2.094395102393$and $D_0 = 0$.

This set of parameters was inspired by a table of behaviors by Kerr et. al. (KWB$^+$85) examining some period-doublings. Near $\gamma \approx 80$ and $D_1 \approx 100$, there is an "incomplete" period-doubling of the type seen by D'Humieres et. al. (DBHL82), a transition from period-3 to period-1 to period-$\frac{3}{2}$. Oddly, there are period-1 regions with strong A and zero Δp. Most of the bands of strong barrier amplitude have $\Delta p > .5$, indicating a nearly flat barrier. It is also possible that the persistence contained some feature not represented by the barrier model yet $\Delta p > .5$ satisfied the least-squares fit to within $R^2 = 0.98$. Located near the incomplete period-doubling is a block of $A \sim 0.65$ with corresponding "undefined" Δp. That is, the confidence interval on the barrier width is larger than the range of valid values. p_b and τ are also undefined in this region, possibly indicating a poor fit. In general, however, there is no clear pattern to the behavior.

Tableau #3: Figures C.5 and C.6

- β varies from 0.22847946571564 to 5.711986642891 in 26 steps,
- D_0 varies from 0.764690423638 to 19.11726059095 in 26 steps,
- $\gamma = D_1 = 12.74484039397$.

This tableau, and all subsequent ones up to Tableau #8 [Figures C.5-C.16] are based on the QP-chaos critical curves of Alstrøm and Levinsen (AL85b). Whenever β and D_0 are the varying parameters, the Arnol'd maps contain diagonal lines of constant winding number. Some other common features to Tableaus #3-#8 are quasiperiodic regions where, additionally, $A \approx 0$. They also contain a large period-1 region, as well as a prominent period-2 band somewhere within the mode-locked area. For brevity, I shall henceforth use the abbreviation "QP" for "quasiperiodic."

In the present tableau, Figures C.5 and C.6, the QP behavior occupies a corner of large forcing, D_0, and weak damping. There is mode locked behavior in an area of large D_0 and large β, sandwiched between two period-1 regions. Note that there is no clear critical curve. Chaotic gridpoints occur between the Arnol'd Tongues.

As for the persistence, much of the QP area fits poorly to the barrier model, as indicated by the white area common to all four plots of the barrier measures. The stochastic

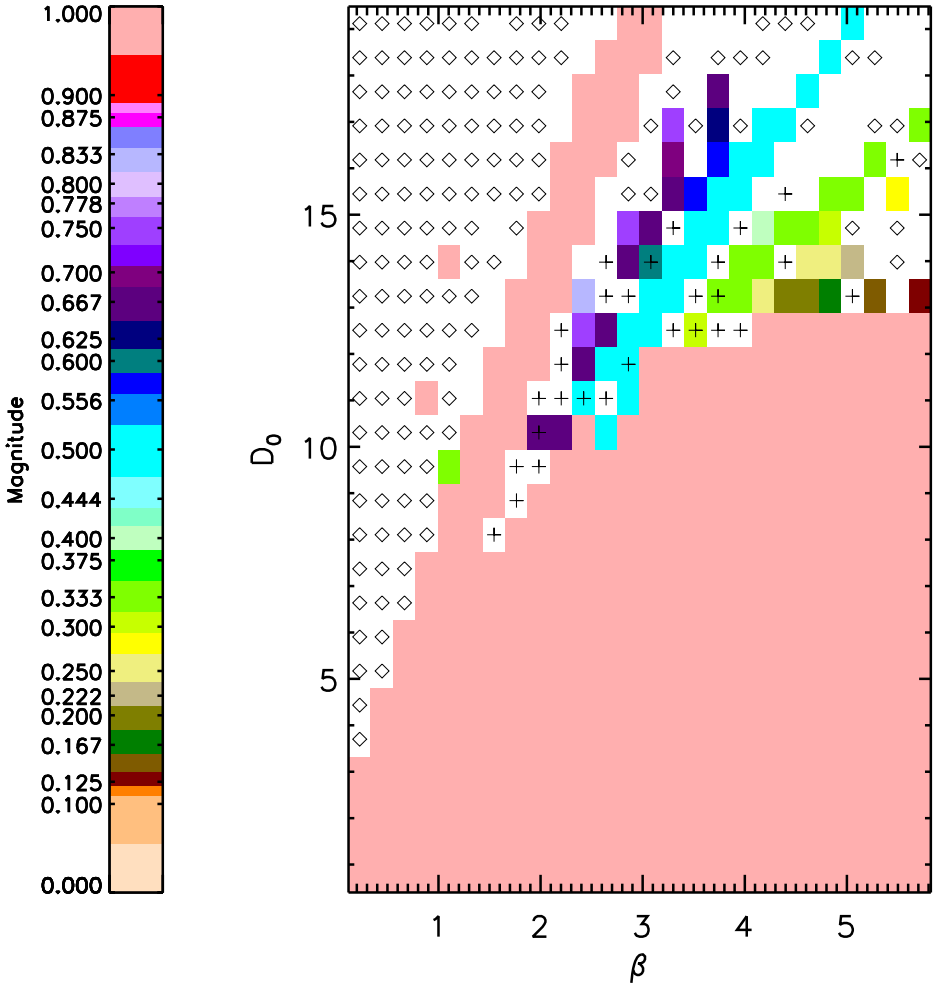

Figure C.5: Arnol'd map for the damped forced pendulum, with $\gamma = D_1 \approx 12.7$ [Tableau #3], plotted in the same fashion as Figure C.1. A detailed description of the plotting method is in section C.2.

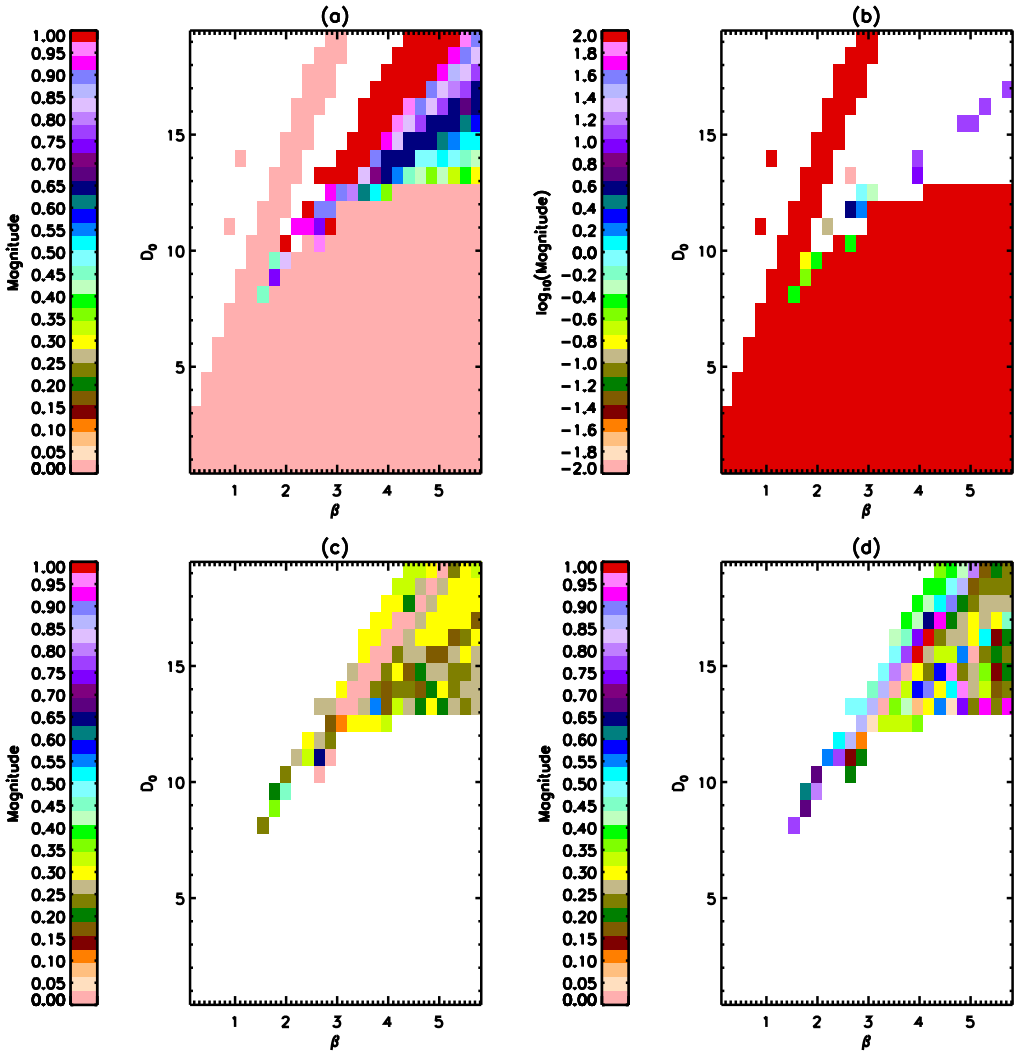

Figure C.6: The barrier measures for Tableau #3 [$\gamma = D_1 \approx 12.7$]. A detailed description of the plotting method is in section C.2. (a) The barrier amplitude, A. (b) $\log_{10}(\tau)$, the stochastic decay time displayed on a logarithmic scale. (c) The barrier width, Δp. (d) The barrier position, p_b.

decay time, τ, is ill-defined except, by design, in the period-1 regions. In the mode-locked area, where, mostly, it is not ill defined, the barrier width tends to have values of roughly $0.25 < \Delta p < 0.45$. The barrier amplitude is also typically larger than $\frac{1}{2}$ in the mode-locked regime, indicating strong but broad persistence barriers. The sole exception to this is, not surprisingly, the period-2 island, where $A \approx 1$ and $\Delta p \approx 0$.

Tableau #4: Figures C.7 and C.8

- β varies from 0.22847946571564 to 5.711986642891 in 26 steps,
- D_0 varies from 1.2744840393968 to 31.86210098492 in 26 steps,
- $\gamma = 12.74484039397$ and $D_1 = 50.97936157587$.

As in the previous tableau, there is again a QP area in the large-D_0 and weak-β corner, while mode-locked and chaotic behavior is sandwiched between period-1 regions. Additionally, there are again diagonal bands of constant winding number, with prominent period-2 mode-locking. This time, however, the fit to the model barrier does meet the criteria $R^2 \geq 0.98$. Notice that $\tau < 0.1$ yrs. in the QP area. Note, too, that the whole of both mode-locked islands have $A > 0.40$ and $\Delta p < 0.2$ at almost all gridpoints, even chaotic ones. This tableau certainly hints at a connection between strong persistence barriers and QP-chaos.

Tableau #5: Figures C.9 and C.10

- β varies from 0.28559933214452 to 7.139983303613 in 26 steps,
- D_0 varies from 3.262679140856 to 81.56697852140 in 26 steps,
- $\gamma = D_1 = 50.97936157587$.

This tableau exhibits common features found in Tableaus #3 and #4: $\tau < 0.1$ yrs. in the QP areas, period-2 islands, islands with mode-locking and chaos. Notice that the chaotic points occur along the border of the period-1 regions and in between Arnol'd Tongues. Also common with Tableau #4 are the barrier measures. The mode-locked area shows typical barrier measures of $A > 0.5$ and $\Delta p < 0.2$, even at chaotic points. Furthermore, p_b has a consistent value near 0.7-0.8 yr. over much of the mode-locked island.

Tableau #6: Figures C.11 and C.12

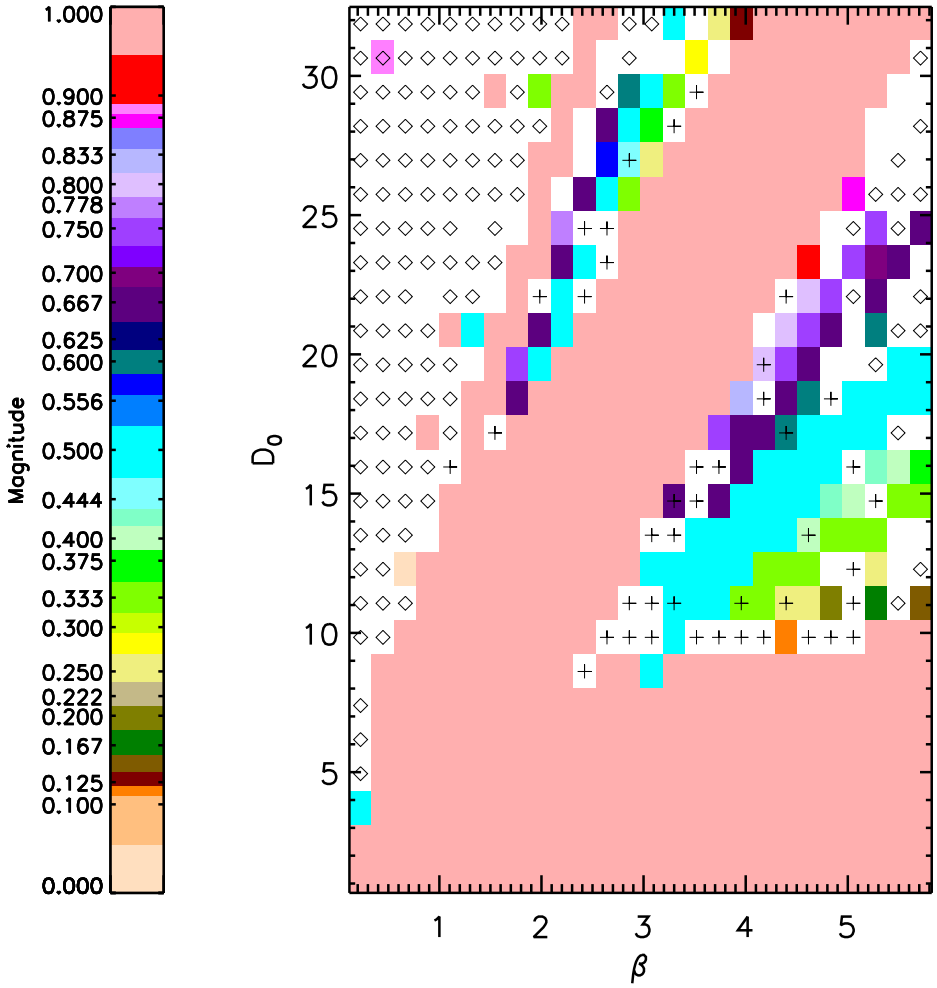

Figure C.7: Arnol'd map for the damped forced pendulum, with $\gamma \approx 12.7$ and $D_1 \approx 51.0$ [Tableau #4], plotted in the same fashion as Figure C.1. A detailed description of the plotting method is in section C.2.

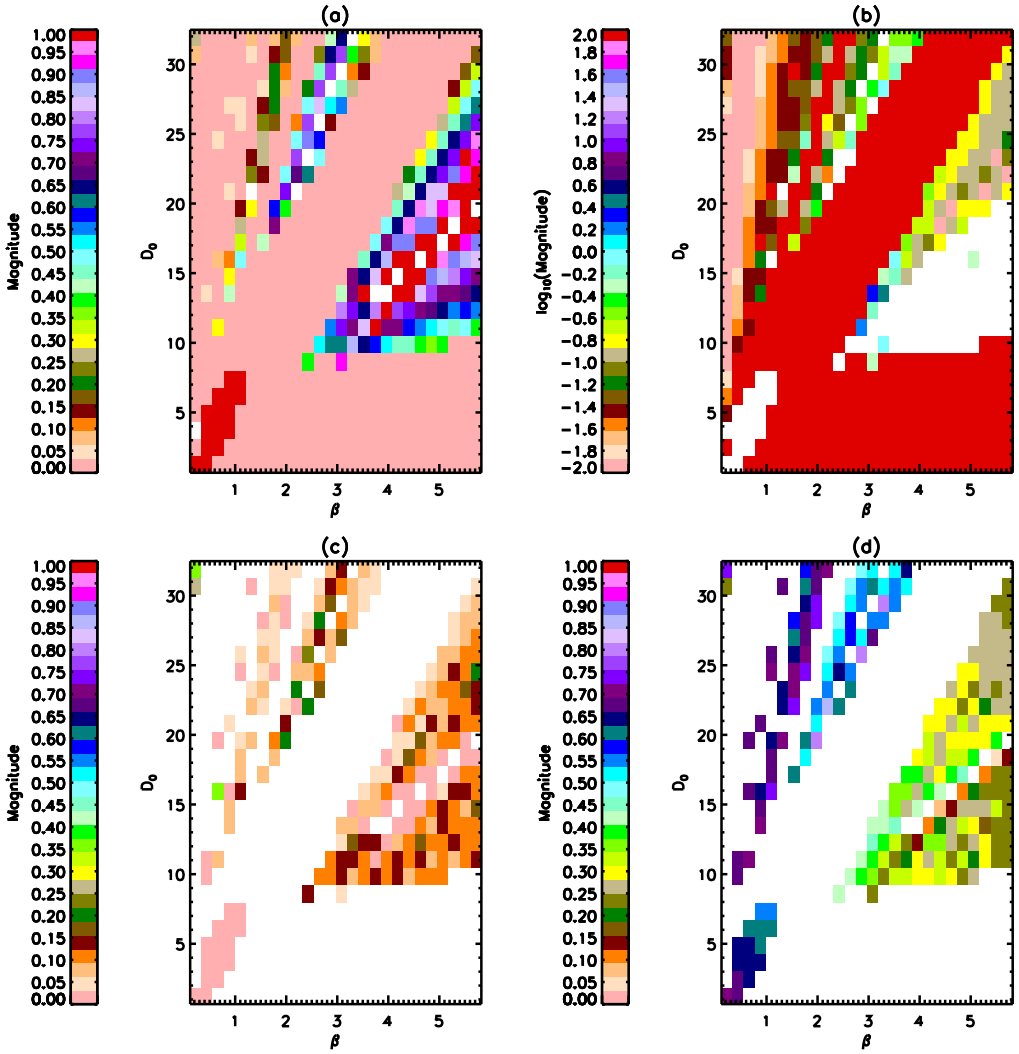

Figure C.8: The barrier measures for Tableau #4 [$\gamma \approx 12.7$ and $D_1 \approx 51.0$]. A detailed description of the plotting method is in section C.2. (a) The barrier amplitude, A. (b) $\log_{10}(\tau)$, the stochastic decay time displayed on a logarithmic scale. (c) The barrier width, Δp. (d) The barrier position, p_b.

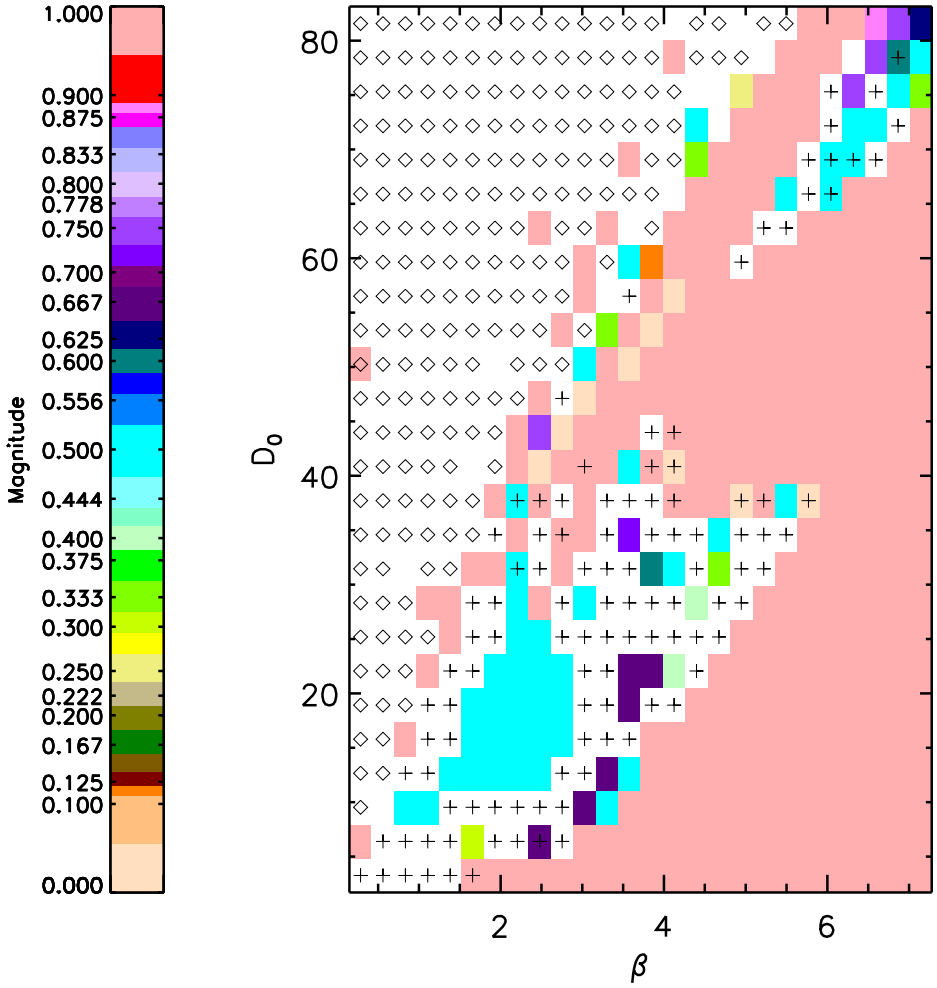

Figure C.9: Arnol'd map for the damped forced pendulum, with $\gamma = D_1 \approx 51.0$ [Tableau #5], plotted in the same fashion as Figure C.1. A detailed description of the plotting method is in section C.2.

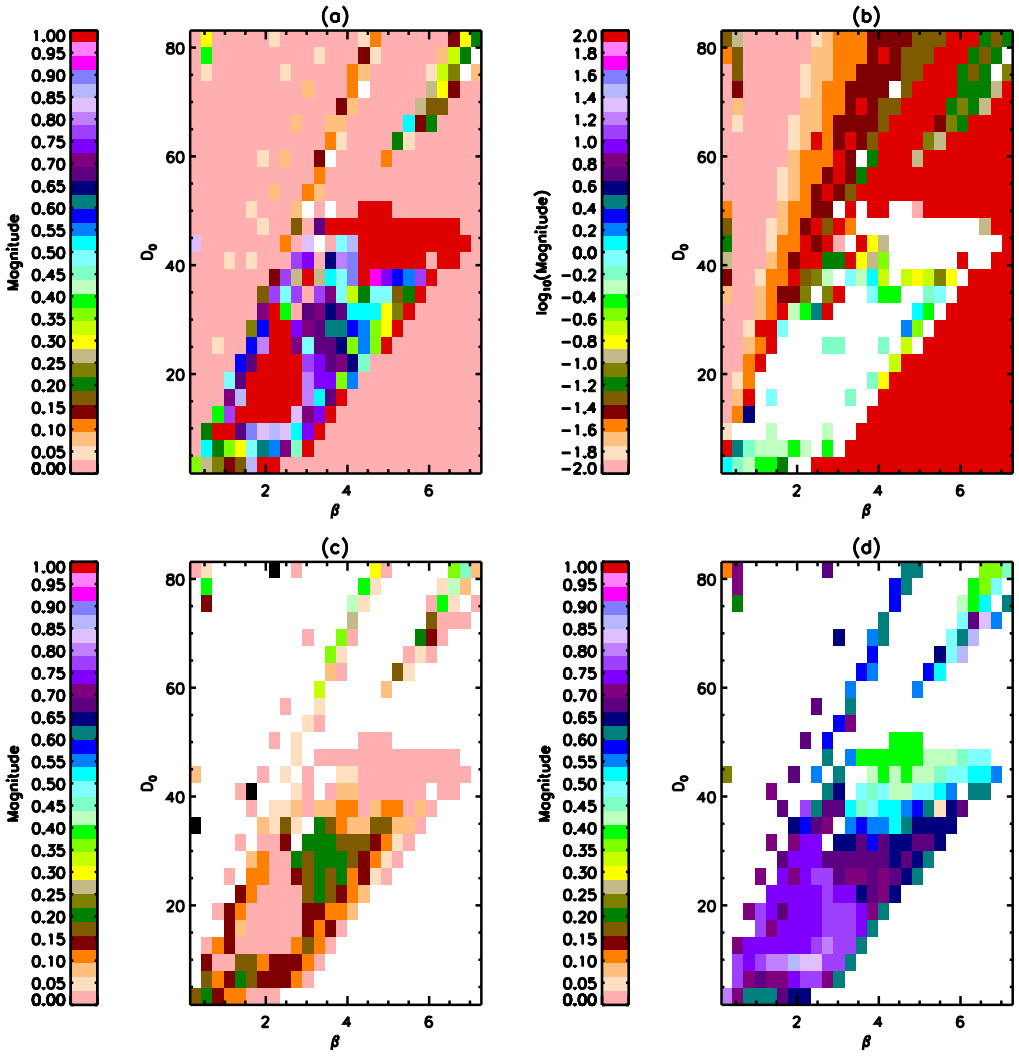

Figure C.10: The barrier measures for Tableau #5 [$\gamma = D_1 \approx 51.0$]. A detailed description of the plotting method is in section C.2. (a) The barrier amplitude, A. (b) $\log_{10}(\tau)$, the stochastic decay time displayed on a logarithmic scale. (c) The barrier width, Δp. (d) The barrier position, p_b.

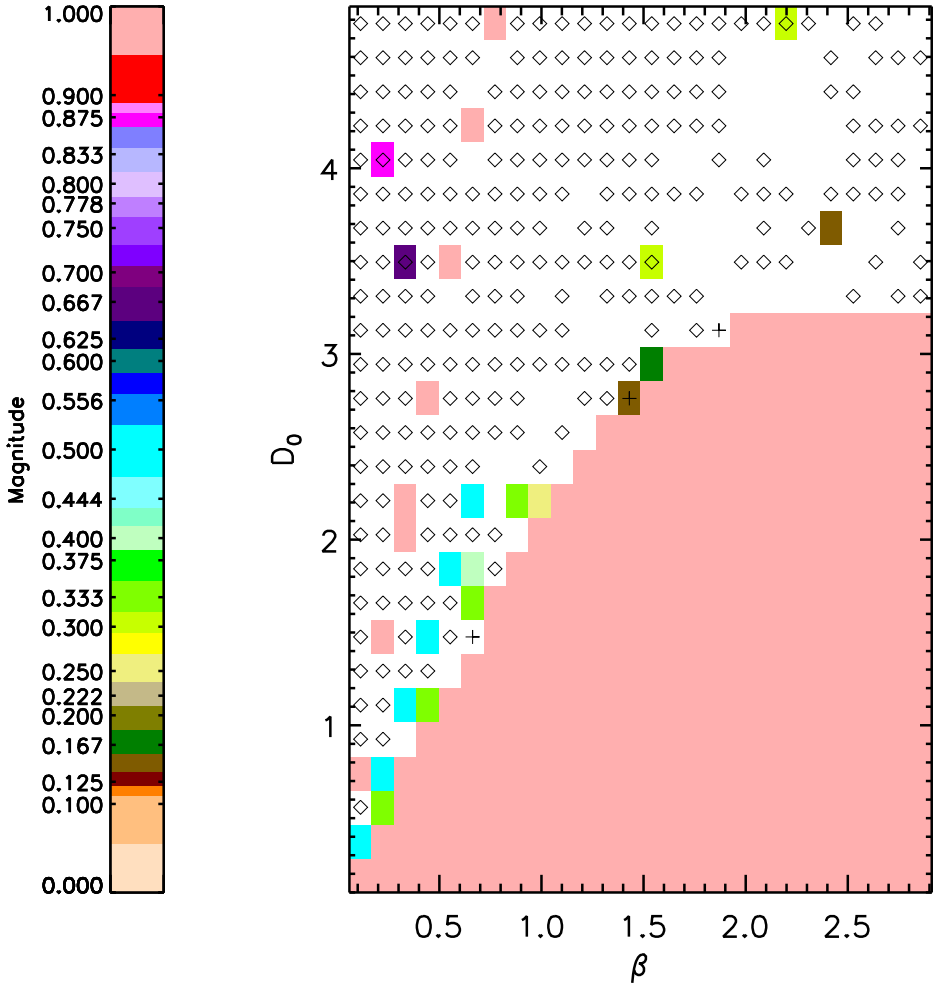

Figure C.11: Arnol'd map for the damped forced pendulum, with $\gamma = D_1 \approx 3.19$ [Tableau #6], plotted in the same fashion as Figure C.1. A detailed description of the plotting method is in section C.2.

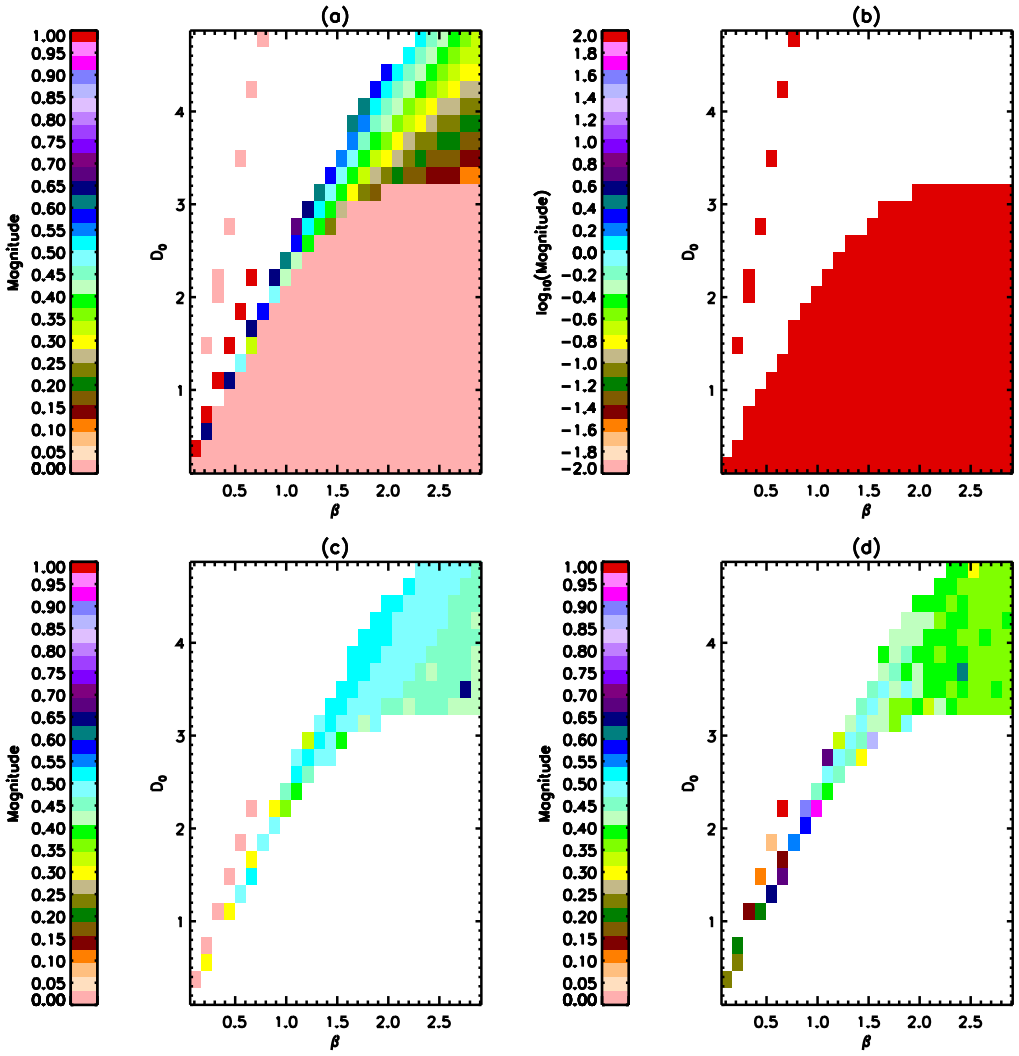

Figure C.12: The barrier measures for Tableau #6 [$\gamma = D_1 \approx 3.19$]. A detailed description of the plotting method is in section C.2. (a) The barrier amplitude, A. (b) $\log_{10}(\tau)$, the stochastic decay time displayed on a logarithmic scale. (c) The barrier width, Δp. (d) The barrier position, p_b.

- β varies from 0.1142397328578 to 2.855993321445 in 26 steps,
- D_0 varies from 0.1911726059092 to 4.77931514773 in 26 steps,
- $\gamma = D_1 = 3.186210098492$.

This tableau differs from the previous three. With only few exceptions, there are no chaotic points and almost no mode-locked points. Part, but not all, of the QP region is ill fitting [i. e. $R^2 < 0.98$]. However, there is an almost monotonic dependence of the barrier amplitude, A, on the ratio of $\frac{D_0}{\beta}$, decreasing as the ratio decreases. This occurs in the QP area of phase space. The width is almost constant in same region, with $\Delta p \sim 0.5$ yr., indicating two possibilities. First, the persistence may indeed have only a linear decay. Second, the persistence may have contained some feature not represented by the barrier model, as noted previously. p_b is also fairly constant in one region, with a value near 0.45 yr.

Tableau #7: Figures C.13 and C.14

- γ varies from 0.6372420196985 to 25.48968078794 in 40 steps,
- D_0 varies from 0.764690423638 to 19.11726059095 in 25 steps,
- $\beta = 5.711986642891$ and $D_1 = 12.74484039397$.

As with Tableau #6, this area of phase space is divided into period-1 and QP behavior with only a smattering of chaotic and mode-locked behavior. Here, there is a dependence of the barrier measures A and Δp on the ratio $\frac{D_0}{\gamma}$. In the QP region, both measures are initially large but decrease to 0 as the ratio $\frac{D_0}{\gamma}$ increases. Additionally, note that $p_b = 0.25$ yr. over almost all of the QP area.

Tableau #8: Figures C.15 and C.16

- β varies from 0.22847946571564 to 5.711986642891 in 25 steps,
- γ varies from 0.6372420196985 to 25.48968078794 in 40 steps,
- $D_0 = 19.11726059095$ and $D_1 = 12.74484039397$.

This tableau shows the features of Tableaus #3-#5, even though D_0 is constant. There is a large QP region, a period-1 region, and an island of mode-locked and chaotic behavior.

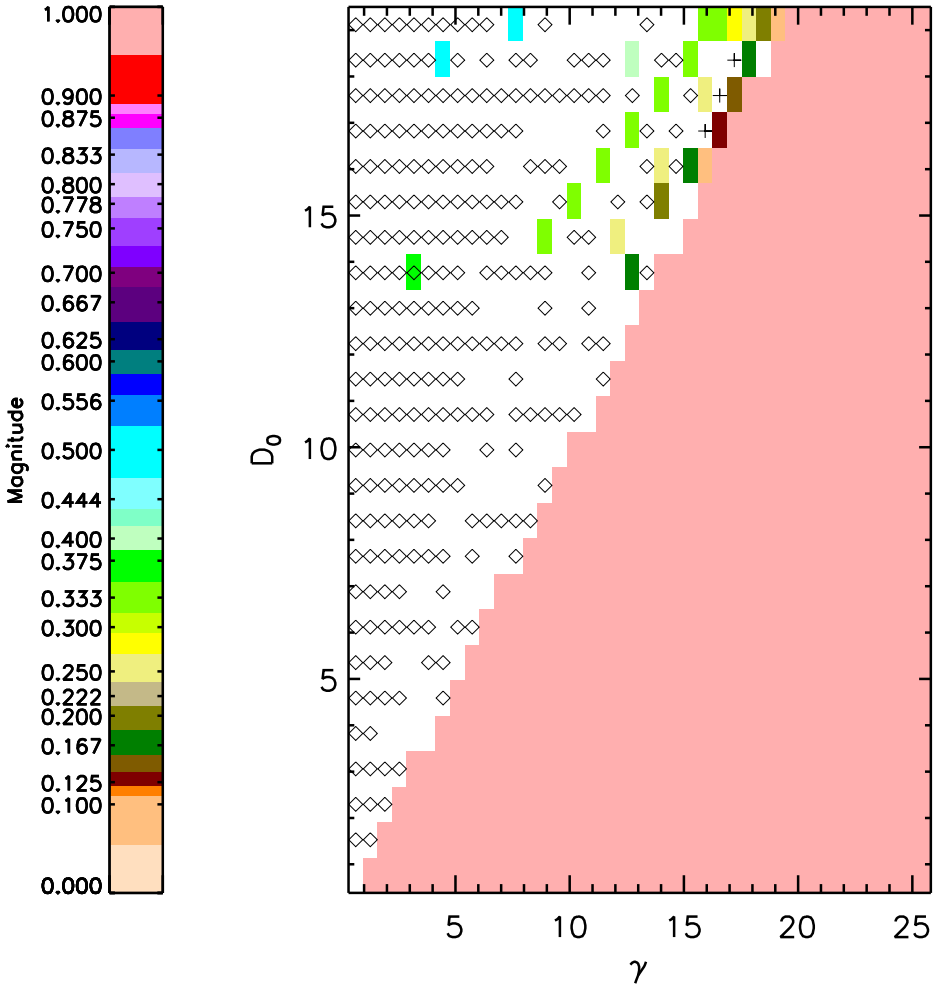

Figure C.13: Arnol'd map for the damped forced pendulum, with $\beta \approx 5.71$ and $D_1 \approx 12.7$ [Tableau #7], plotted in the same fashion as Figure C.1. A detailed description of the plotting method is in section C.2.

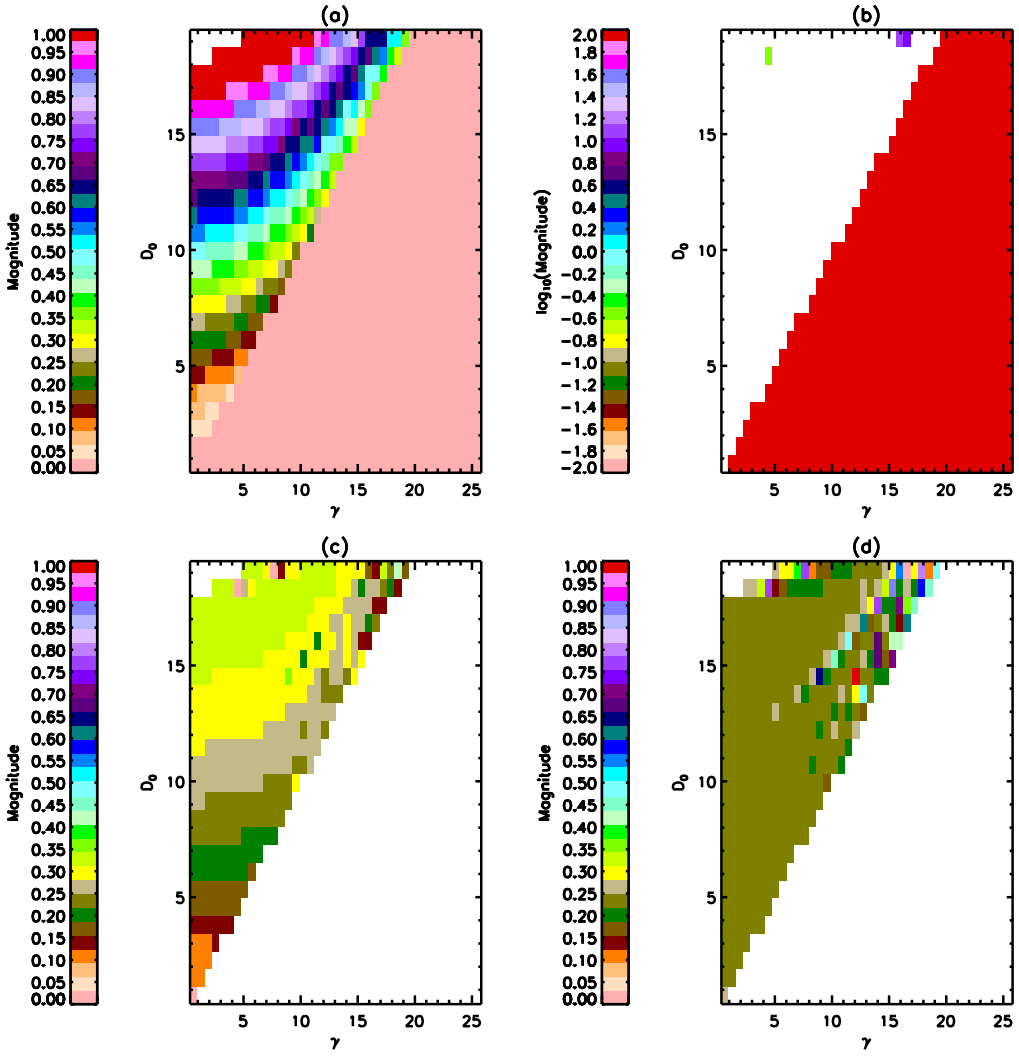

Figure C.14: The barrier measures for Tableau #7 [$\beta \approx 5.71$ and $D_1 \approx 12.7$]. A detailed description of the plotting method is in section C.2. (a) The barrier amplitude, A. (b) $\log_{10}(\tau)$, the stochastic decay time displayed on a logarithmic scale. (c) The barrier width, Δp. (d) The barrier position, p_b.

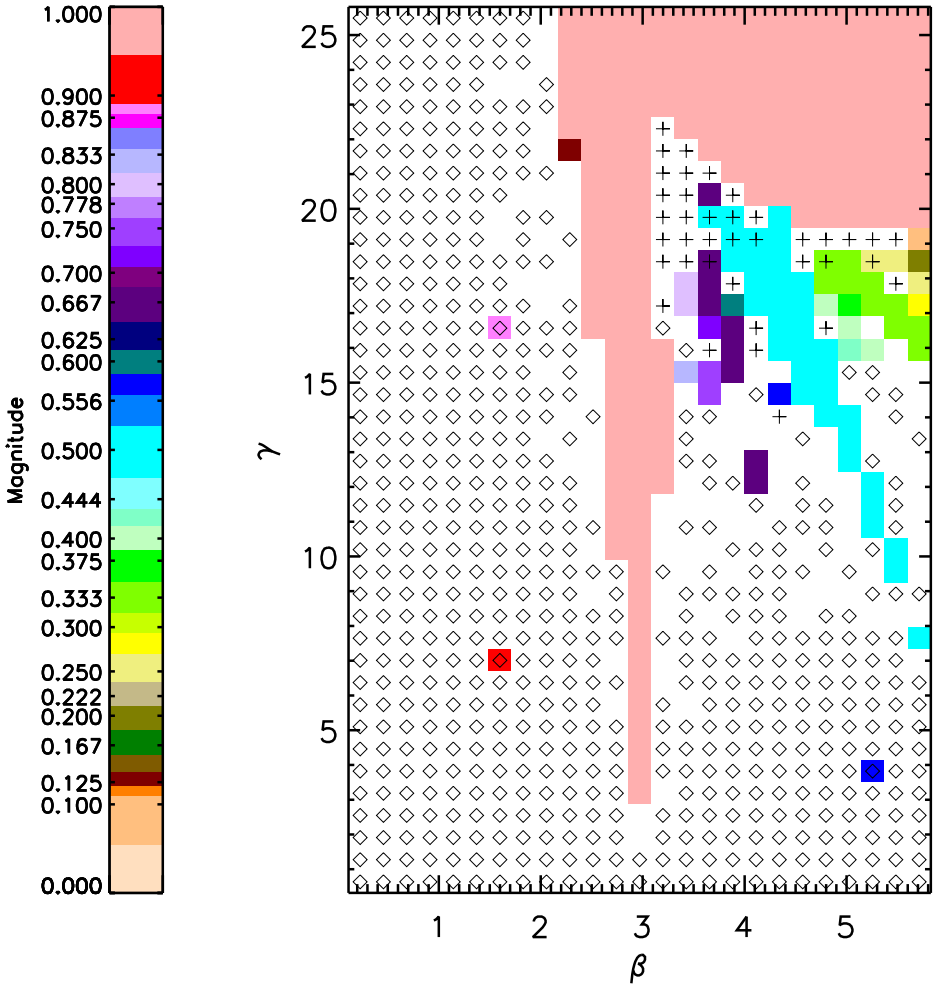

Figure C.15: Arnol'd map for the damped forced pendulum, with $D_0 \approx 19.1$ and $D_1 \approx 12.7$ [Tableau #8], plotted in the same fashion as Figure C.1. A detailed description of the plotting method is in section C.2.

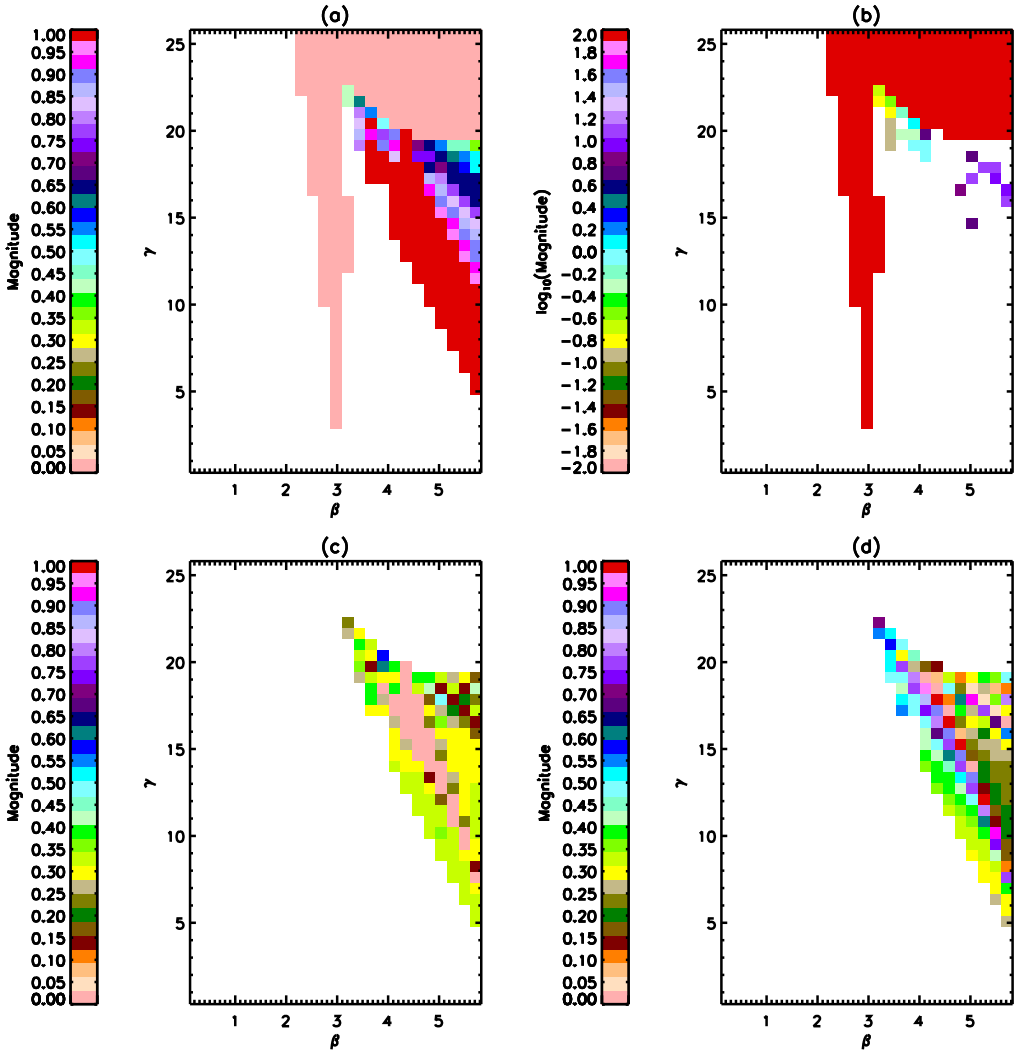

Figure C.16: The barrier measures for Tableau #8 [$D_0 \approx 19.1$ and $D_1 \approx 12.7$]. A detailed description of the plotting method is in section C.2. (a) The barrier amplitude, A. (b) $\log_{10}(\tau)$, the stochastic decay time displayed on a logarithmic scale. (c) The barrier width, Δp. (d) The barrier position, p_b.

The QP region is almost completely ill-fitting [$R^2 < 0.98$]. There is a large period-2 island as well as readily visible period-4 and period-$\frac{8}{5}$ areas. The chaotic points occur where these mode-locked areas and the period-1 area converge. The barrier measures indicate wide but strong persistence barriers in the mode-locked/chaotic area, with $A > 0.7$ and $\Delta p \approx 0.3$ except at the period-2 island. There are again large areas of constant p_b.

Tableau #9: Figures C.17 and C.18

- β varies from 2.356194490192 to 4.712388980385 in 2 steps,
- D_1 varies from 88.82643960980 to 133.2396594147 in 222 steps,
- $\gamma = 88.82643960980$ and $D_0 = 0$.

It may seem odd to vary β in only two steps. The larger value of β, the range of D_1, and the values of γ and D_0 are equivalent to the parameters used in the bifurcation diagram of Gwinn and Westervelt (GW86). The results look similar to those of Tableau #2. There is a period-2 area, with the expected barrier measures. There is also a set of period-3 points, with $A \approx 0.75$ but with undefined Δp, τ, and p_b, possibly indicating a poor fit to the barrier model and, therefore, the presence of some other phase-dependent feature in the persistence. There are also some areas of linearly decaying persistence described by large values of A and Δp. Or, these points where A and Δp are large may also indicate persistence with some other phase-dependent feature. In general, however, there is no strikingly clear pattern of behavior as there was for the $D_0 \neq 0$ tableaus.

Tableau #10: Figures C.19 and C.20

- β varies from 0.475356 to 4.75356 in 10 steps,
- D_0 varies from 2 to 20 in 10 steps,
- $\gamma = 4.398150$ and $D_1 = 45$.

This regions is presented for completeness. It covers the same area of parameter space as Figure 6.4, but in coarser steps. It shows the same behavior as Figures C.5-C.10, where β and D_0 were also the varying parameters.

Tableau #11: Figures C.21 and C.22

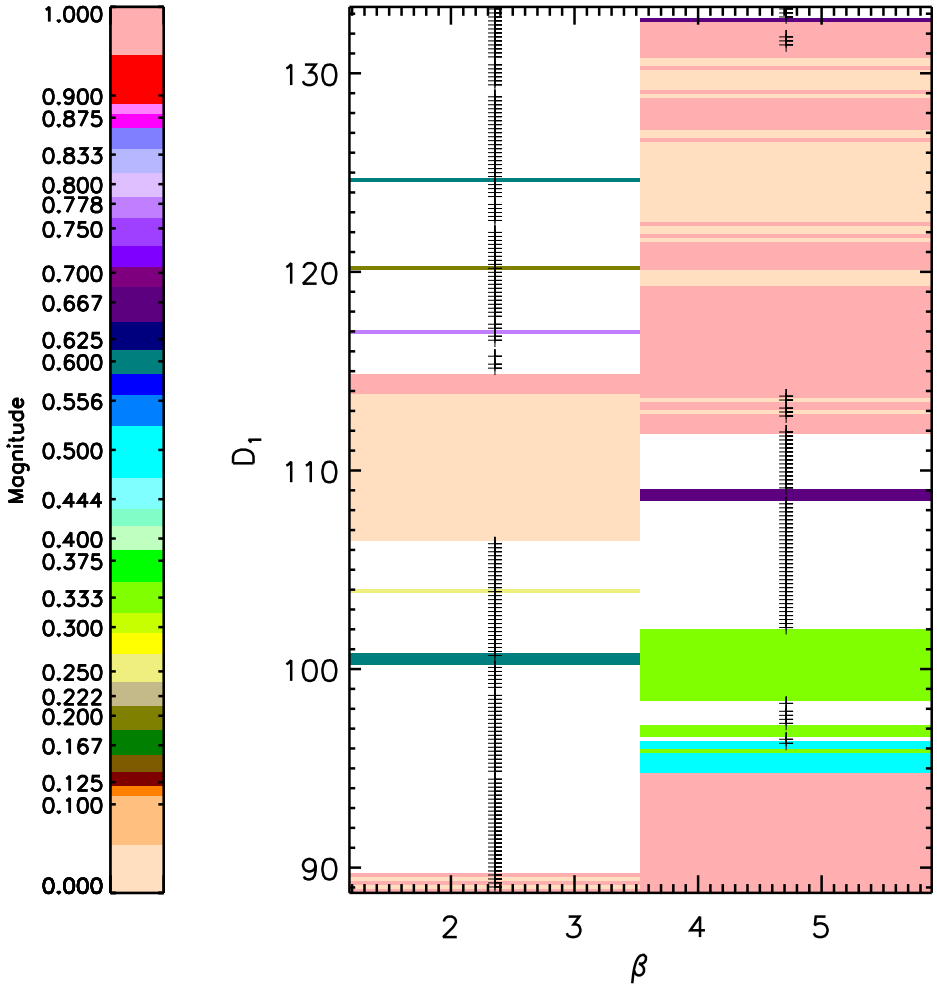

Figure C.17: Arnol'd map for the damped forced pendulum, with $\gamma \approx 88.8$ and $D_0 = 0$ [Tableau #9], plotted in the same fashion as Figure C.1. A detailed description of the plotting method is in section C.2.

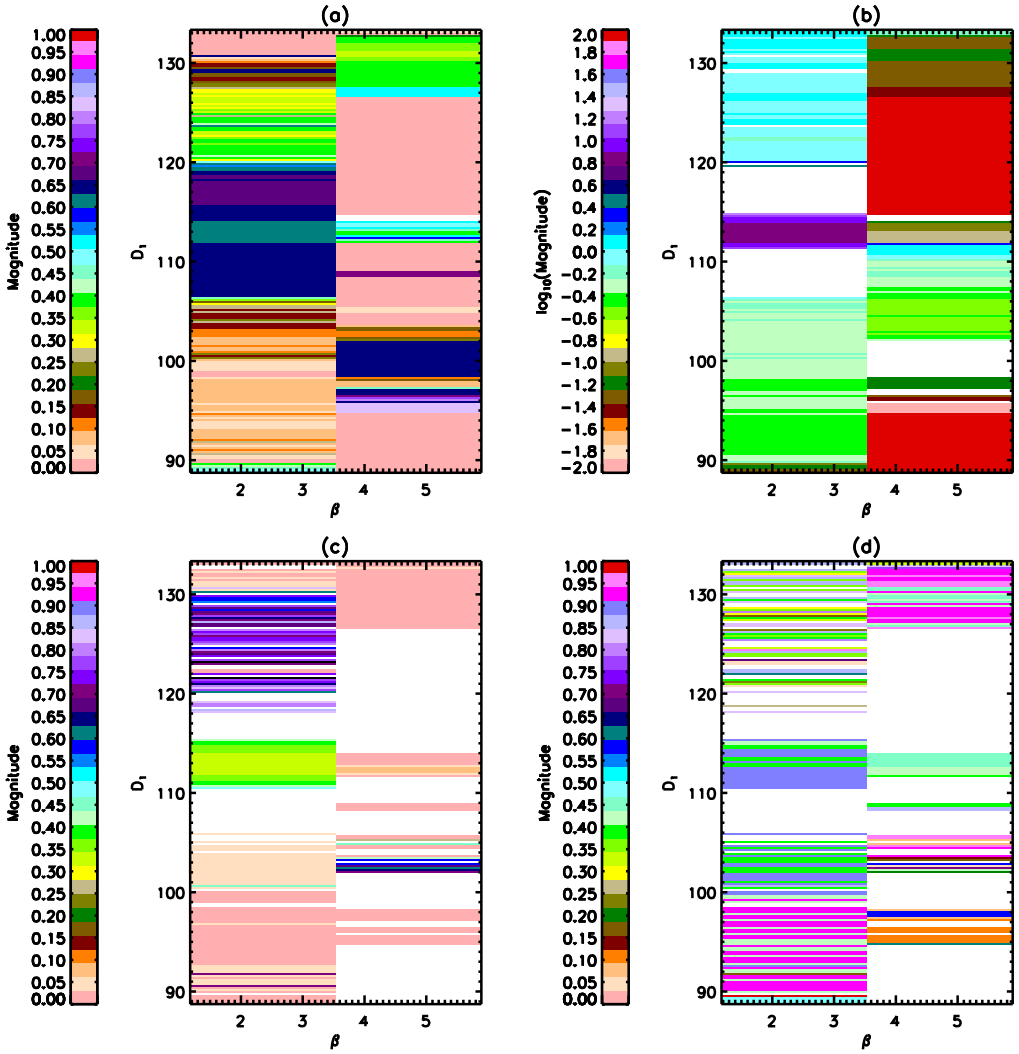

Figure C.18: The barrier measures for Tableau #9 [$\gamma \approx 88.8$ and $D_0 = 0$]. A detailed description of the plotting method is in section C.2. (a) The barrier amplitude, A. (b) $\log_{10}(\tau)$, the stochastic decay time displayed on a logarithmic scale. (c) The barrier width, Δp. (d) The barrier position, p_b.

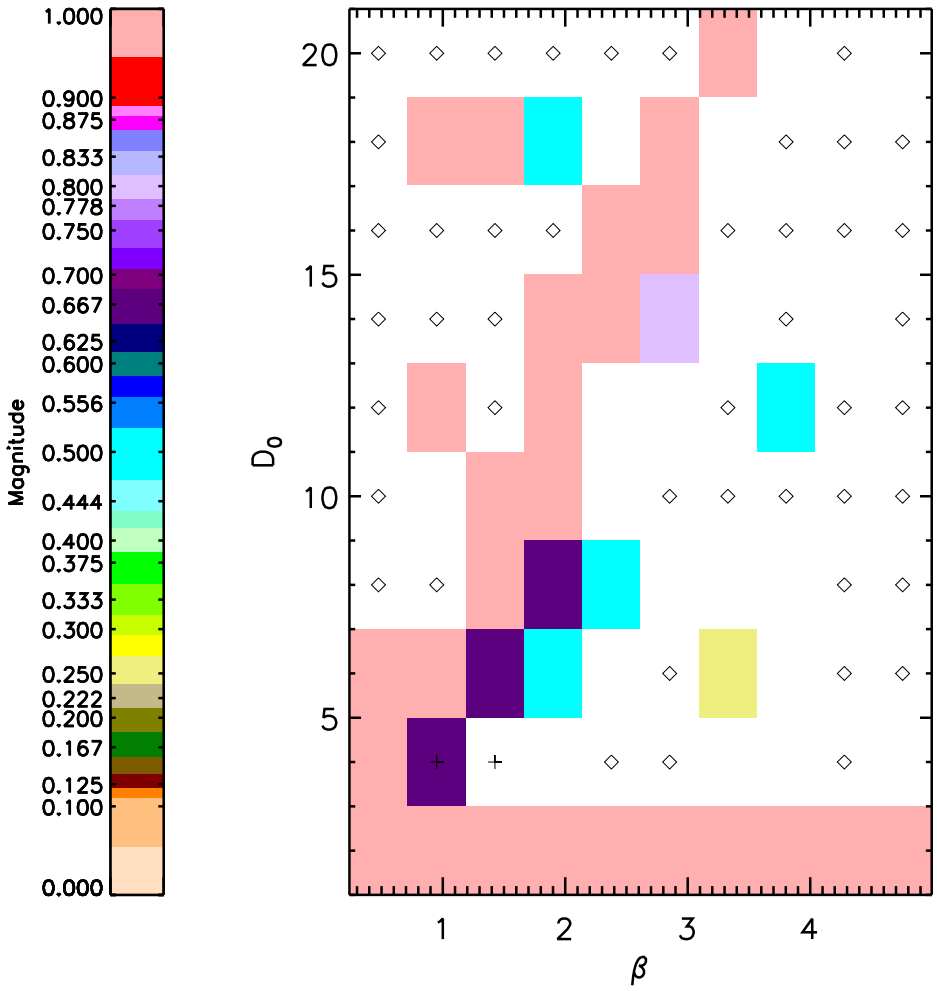

Figure C.19: Arnol'd map for the damped forced pendulum, with $\gamma \approx 4.4$ and $D_1 = 45$ [Tableau #10], plotted in the same fashion as Figure C.1. A detailed description of the plotting method is in section C.2.

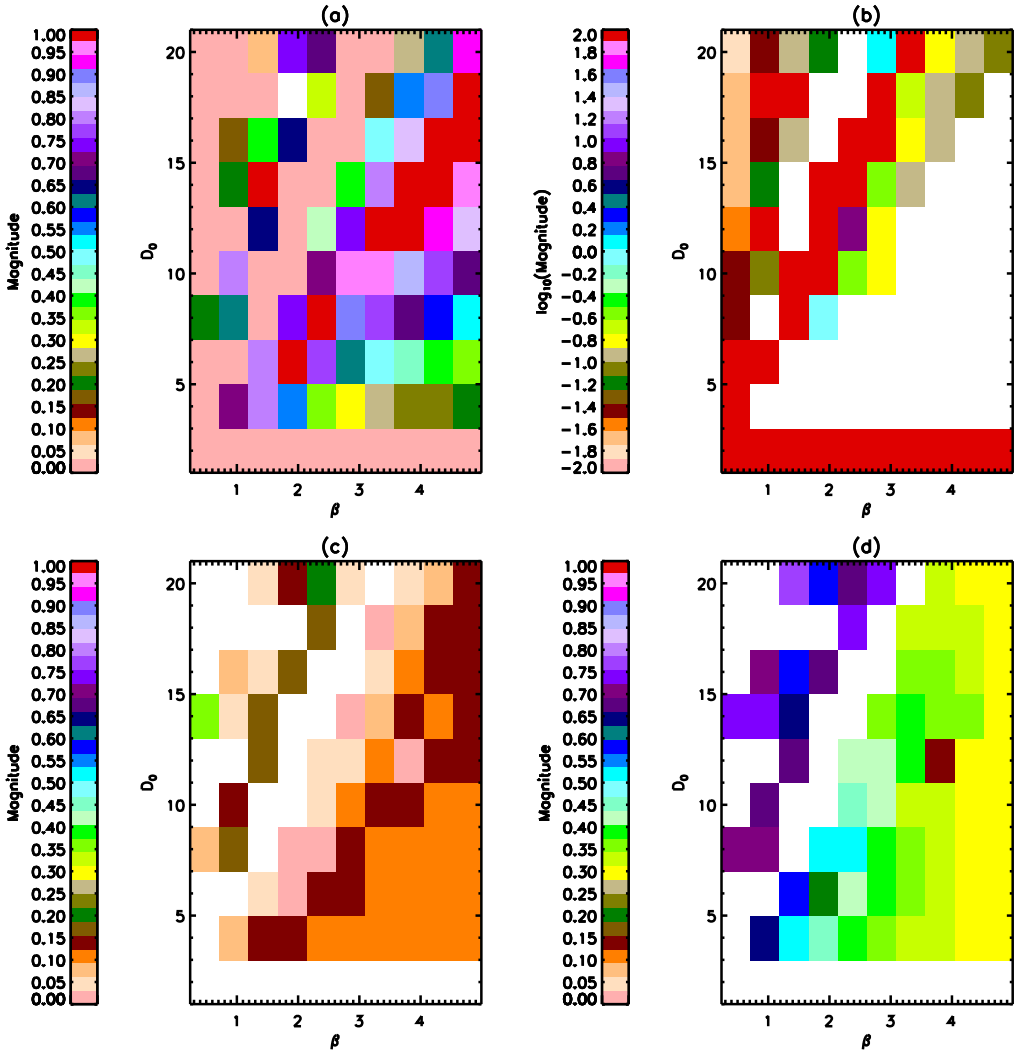

Figure C.20: The barrier measures for Tableau #10 [$\gamma \approx 4.4$ and $D_1 = 45$]. A detailed description of the plotting method is in section C.2. (a) The barrier amplitude, A. (b) $\log_{10}(\tau)$, the stochastic decay time displayed on a logarithmic scale. (c) The barrier width, Δp. (d) The barrier position, p_b.

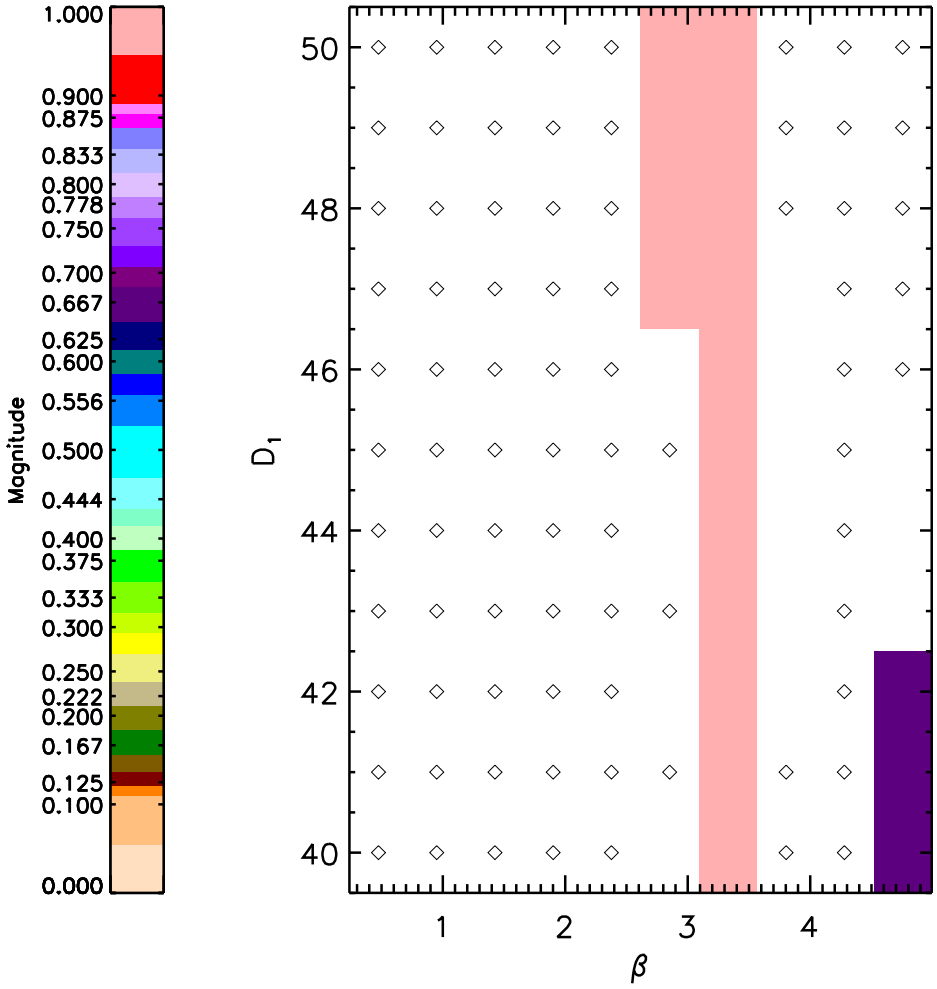

Figure C.21: Arnol'd map for the damped forced pendulum, with $\gamma \approx 4.4$ and $D_0 = 20$ [Tableau #11], plotted in the same fashion as Figure C.1. A detailed description of the plotting method is in section C.2.

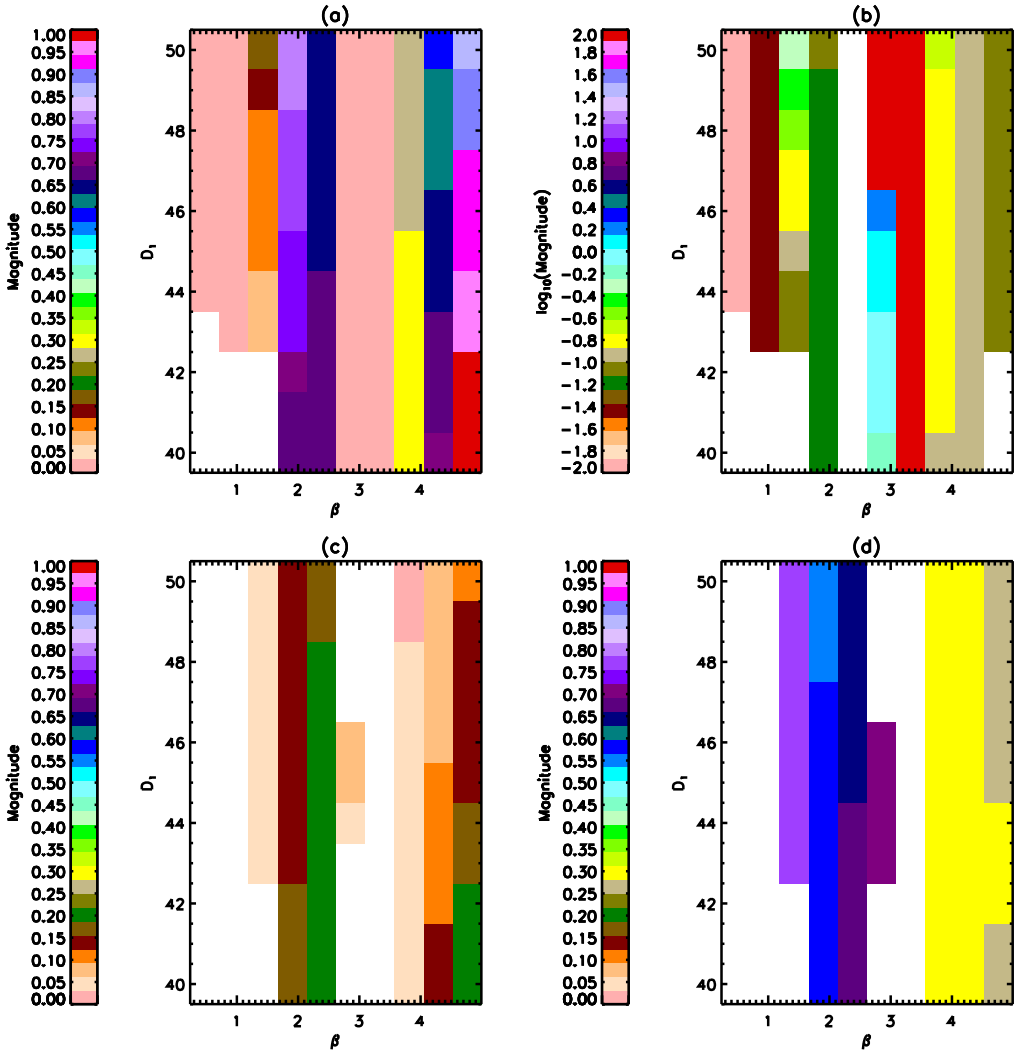

Figure C.22: The barrier measures for Tableau #11 [$\gamma \approx 4.4$ and $D_0 = 20$]. A detailed description of the plotting method is in section C.2. (a) The barrier amplitude, A. (b) $\log_{10}(\tau)$, the stochastic decay time displayed on a logarithmic scale. (c) The barrier width, Δp. (d) The barrier position, p_b.

- β varies from 0.475356 to 4.75356 in 10 steps,
- D_1 varies from 40 to 50 in 11 steps,
- $\gamma = 4.398150$ and $D_0 = 20$.

This exploration of parameter space is a variant of the previous one. There is clear QP behavior and a period-1 area, but no mode-locking or chaos. There is a dependence of all of the barrier measures except p_b on β, with also a mild dependence on D_1. The barrier amplitude is mainly larger than 0.7, with one band near 0.3. The barrier position, p_b, comes in two bands, one near 0.25 yr., the other near 0.75 yr. The barrier width, Δp, remains below 0.2 yr. The stochastic decay time has some rather interesting behavior, occurring in D_0-independent bands. The value of τ at these bands decreases from 0.1 yrs., in two separate bands, as a function of β.

Tableau #12: Figures C.23 and C.24

- γ varies from 10 to 100 in 10 steps,
- D_1 varies from 10 to 100 in 10 steps,
- $\beta = 1$ and $D_0 = 1$.

Although $D_0 \neq 0$, there is no QP behavior in this region of parameter space. There are large areas of period-1 and chaotic behavior, with some mode-locking. All of the barrier measures show wide variation in the chaotic region, with no clear pattern emerging.

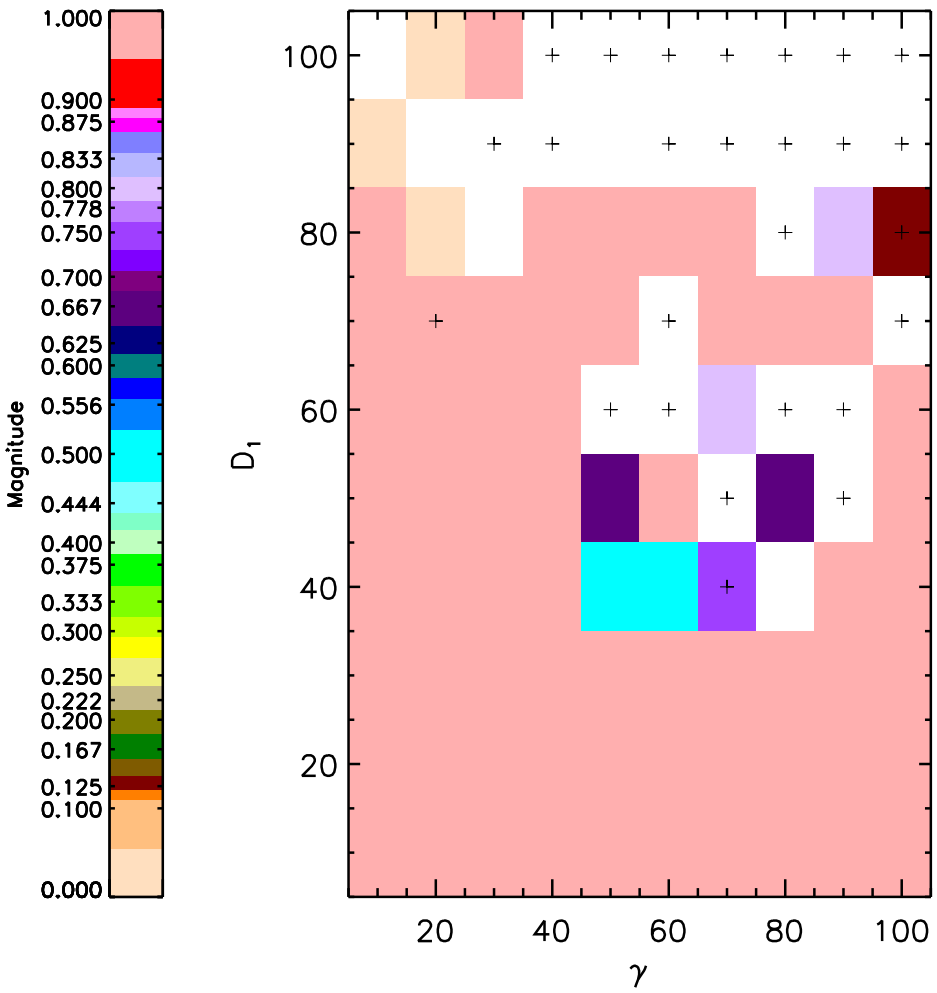

Figure C.23: Arnol'd map for the damped forced pendulum, with $\beta = 1$ and $D_0 = 1$ [Tableau #12], plotted in the same fashion as Figure C.1. A detailed description of the plotting method is in section C.2.

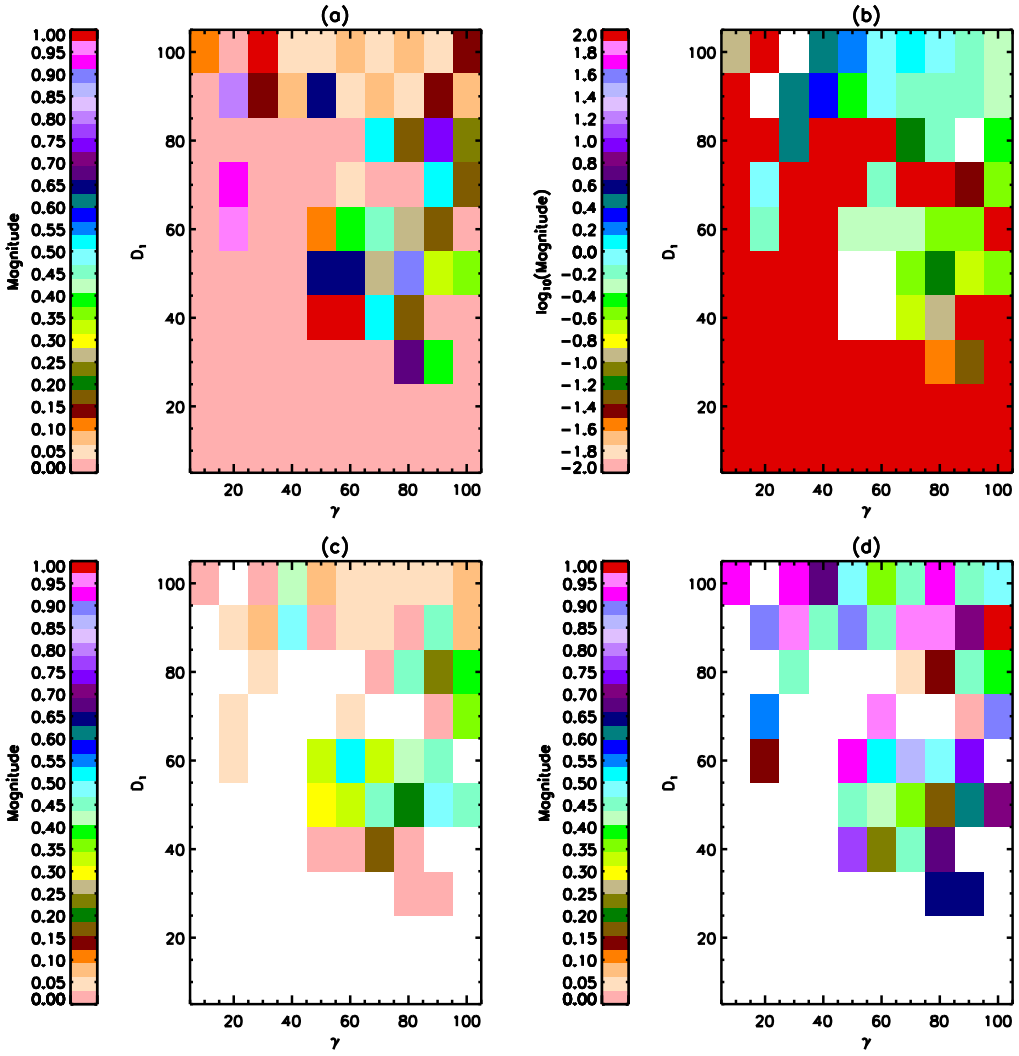

Figure C.24: The barrier measures for Tableau #12 [$\beta = 1$ and $D_0 = 1$]. A detailed description of the plotting method is in section C.2. (a) The barrier amplitude, A. (b) $\log_{10}(\tau)$, the stochastic decay time displayed on a logarithmic scale. (c) The barrier width, Δp. (d) The barrier position, p_b.