Correlation Effects in a Simple Stochastic Model of the Thermohaline Circulation

Adam Hugh Monahan*

Institut für Mathematik

Humboldt-Universität zu Berlin

Unter den Linden 6, 10099

Berlin, Germany.

monahana@uvic.ca

August 8, 2002

*Present Address: School of Earth and Ocean Sciences, University of Victoria, P.O. Box 3055 STN CSC, Victoria BC, Canada, V8P 5C2

Abstract

A simple model of the thermohaline circulation of the World Ocean is considered, in which fluctuations in internal oceanic mixing and in freshwater forcing are represented by stochastic processes. The effects on the stationary probability density function of correlations between fluctuations in mixing and freshwater forcing, and of finite autocorrelation time in oceanic mixing, are determined using a mixture of analytical and numerical techniques. The quantitative behaviour of the system is found to depend on the strength and correlation character of the noise processes, quite sensitively so in some regions of parameter space. The results of this analysis suggest the importance of accurately modelling high-frequency variability in nonlinear models of the climate system.

1 Introduction

To a first approximation, the circulation of the World Ocean can be divided into two components. The first, the wind-driven circulation, is the more familiar; it is associated with the oceanic gyre circulations dominating the surface flow of the World Ocean. As is suggested by its name, this circulation arises primarily through mechanical interaction with the overlying atmospheric circulation. In contrast, the thermohaline circulation (THC), characterising the circulation of the deep ocean, is predominantly buoyancy-driven. Throughout the bulk of the World Ocean, the vertical density gradient (stratification) is sufficiently strong that vertical motion is strongly suppressed. However, in a few regions, such as the Greenland, Iceland, and Norwegian seas in the Northern Hemisphere and the Ross and Weddell seas in the Southern Hemisphere, the intense cooling of the surface waters can erode the stratification to the point that large volumes of water can sink to the deep ocean. These convection events feed a system of slow, deep currents that constitute the circulation of the bulk of the World Ocean. Weak upwards flow distributed throughout the rest of the ocean returns this deep water to the surface, where currents in the thermocline layer flowing toward the convection sites complete the circuit. The present picture of

the THC holds that convection in the North Atlantic is the primary engine driving this circulation. Overviews of the global thermohaline circulation are provided in Weaver and Hughes [40] and in Schmitz [31]; Broecker [7] presents an introduction for non-specialists to the THC and its variability.

The thermohaline circulation is believed to play a central role in climate variability on timescales from decades to millennia, through both its internal dynamics and its response to external forcing (see e.g. [28] and [39]). In the Atlantic Ocean, as it operates today, the THC has the net effect of transporting a tremendous quantity of heat northward, with what are believed to be significant consequences for the climate of Northwest Europe [29]. However, the present configuration of the THC may not be the only one it can take. Investigations involving simple heuristic models (e.g. [27], [32], [35]), models of intermediate complexity (e.g. [10], [30]), and complex general circulation models (GCM; e.g. [14], [20], [26], [37]) indicate that the THC may display multiple, and very different, regimes of circulation, transitions between which are very rapid relative to the length of time spent within a regime. Furthermore, evidence of rapid shifts in the climate state abounds in the geological record, with timescales from those of glaciation cycles to higher frequency, millennial scale fluctuations (e.g. [2], [6], [9], [33]); it is widely believed that rearrangements of the THC play a central role in this variability.

Simple box models have demonstrated a remarkably good ability to reproduce the multiple regimes of the thermohaline circulation found in full GCMs. They are an attractive conceptual tool because of both their low computational cost and the small number of parameters that govern their dynamics. Considerable attention has been paid to both purely deterministic models (e.g. [19], [27], [32], [35]) and to models with a stochastic component (e.g. [8], [12], [18], [34], [36]). With deterministic models, the natural framework of analysis is dynamical systems theory, and discussion has centred on the nature and stability of attractors admitted by the model, and on the nature and location of bifurcations. Analysis of the stochastic class of models involves introducing the theory of stochastic differential equations (e.g. [11]). Attention in these models has primarily been focused on the nature of the spectrum and on gross

features of the stationary distribution.

A study combining perspectives from both stochastic analysis and dynamical systems theory was that of Timmermann and Lohmann [36], in which changes in the qualitative behaviour of the stationary distribution with changes in bifurcation parameters were mapped out. Monahan [21] (hereafter M02) extended the analysis of Timmermann and Lohmann using a somewhat more general model, and noted that for physically relevant noise levels, the stationary distribution was concentrated around one of the two regimes for most of the parameter range in which the distribution was technically bimodal. That is, although the model admits two THC regimes, the presence of noise renders one regime much more frequently occupied than the other. This phenomenon was denoted stabilisation by noise. It was shown in M02 that the partitioning of probability mass between the regimes is a complicated and potentially sensitive function of the noise levels.

In M02, fluctuations in internal oceanic mixing and in freshwater flux were parameterised as independent white-noise processes. In the present study, we extend the results of the earlier analysis by considering the effects of correlations between freshwater flux and internal mixing, and of finite autocorrelation time in mixing fluctuations. Section 2 describes the simple model of the THC considered in this study and recapitulates the basic results of M02. In Section 3, the effects of correlations between the fluctuating parameters are considered. Section 4 addresses the effects of nonzero autocorrelation times in the internal mixing process. A summary and conclusions are presented in Section 5.

2 The Stochastic Stommel Model

The model considered in M02 is a generalisation of the classical Stommel [35] two-box model of the THC, illustrated schematically in Figure 1. Two ocean boxes are considered, representing respectively mid- and high-latitude oceans. The boxes are characterised by average temperatures T_1 and T_2, and salinities S_1 and S_2. Heat exchanges with the atmosphere relax the box-averaged temperatures to the climato-

logical values T_{a1} and T_{a2}, while the salinities are driven by freshwater fluxes F_1^{oa} and F_2^{oa} . Density gradients between the boxes are assumed to establish large-scale currents which lead to an interbox exchange of salinity and temperature ; the Stommel ansatz is that the exchange is proportional to the absolute density gradient

$$q \propto |\alpha(T_1 - T_2) - \beta(S_1 - S_2)|, \tag{1}$$

where α and β are respectively the thermal and haline expansivity coefficients. The model studied in M02 further assumes the existence of a fluctuating interbox eddy exchange, denoted η. While this mixing process appears ad hoc, such a fluctuating eddy exchange term arises naturally from fluctuations in mechanical forcing in the more sophisticated model of the ocean circulation introduced by Maas [19].

Because the bulk exchange between the boxes depends only on the density gradient, the dynamics of the Stommel model depend only on the meridional (i.e. north-south) gradients of temperature and salinity. In nondimensionalised variables (defined in M02), the stochastic Stommel model is expressed:

$$\dot{x} = -|x - y|x - \eta x + \lambda(1 - x) \tag{2}$$

$$\dot{y} = -|x - y|y - \eta y + \mu + \sigma_2 \dot{W}_2 \tag{3}$$

$$\dot{\eta} = -\frac{1}{\tau}\eta + \frac{\sigma_1}{\tau}\dot{W}_1. \tag{4}$$

Here, x and y denote the nondimensionalised meridional temperature and salinity gradients, respectively; λ^{-1} is the nondimensional timescale on which the temperature gradient relaxes to the climatological value (which is 1 in these units); and μ is the meridional gradient of surface freshwater flux, which in principle can depend on the temperature gradient x. Fluctuations in the freshwater flux μ are parameterised as a white noise process, $\dot{W}_2$, scaled by the noise strength σ_2. The fluctuating diffusivity η is modelled as an Ornstein-Uhlenbeck (red noise) process with autocorrelation e-folding time τ and variance $\sigma_1^2/(2\tau)$. In the limit that $\tau \to 0$, η becomes white noise with strength σ_1.

The deterministic version of equations (2)-(4) was introduced by Stommel [35], and inspired considerable interest because it admits multiple stable fixed points for

a range of values of λ and μ. One of these resembles the present-day THC, in which water sinks at the poles and rises at lower latitudes, while in the other state the circulation is weaker and reversed. The existence of multiple THC regimes has since been demonstrated in a hierarchy of climate models, including full ocean/atmosphere/ice GCMs (e.g. [14], [20], [26], [27], [30], [37]). While the Stommel model is far too simple to provide a quantitatively accurate description of THC dynamics, it is sufficiently simple to admit analytic solutions while capturing what paleoclimate data and more sophisticated models suggest is an essential nonlinear feature of the THC (e.g. [9], [33], [40]).

Equations (2)-(4) are a stochastic differential equation (SDE) for the three-dimensional Markov process (x, y, η); the evolution of the joint probability density function (PDF) of (x, y, η) is given by a Fokker-Planck equation (FPE). A review of the theory of SDEs is given in Penland [25]; a more comprehensive discussion appears in Gardiner [11].

M02 considered the system (2)-(4) in the limits that $\tau \to 0$ and $\lambda \to \infty$. The second of these limits implies that $x = 1$; physically, this corresponds to a situation in which the timescale of the temperature dynamics is substantially shorter than that of the salinity dynamics, so it can be assumed that the temperature gradient x adjusts instantaneously to the climatological value. This idealisation, also considered in Cessi [8], renders the stationary Fokker-Planck equation of the resulting system analytically tractable.

We review briefly the main results of M02. The SDE that arises in the limits that $\tau \to 0$ and $\lambda \to \infty$ is

$$\dot{y} = -|1 - y|y - \sigma_1 y \circ \dot{W}_1 + \sigma_2 \dot{W}_2, \tag{5}$$

where the open circle denotes that the white noise $\dot{W}_1$ is to be interpreted in the Stratonovich sense. This is the appropriate interpretation for the white-noise limit of the autocorrelated process η [11]. The associated stationary Fokker-Planck equation for the stationary PDF, p^s, is

$$0 = -\frac{d}{dy}\left(\left[-|1 - y|y + \mu + \frac{\sigma_1^2}{2}y\right]p^s\right) + \frac{1}{2}\frac{d^2}{dy^2}([\sigma_1^2 y^2 + \sigma_2^2]p^s). \tag{6}$$

This FPE can be integrated to obtain a closed-form expression for p^s: for $y < 1$,

$$p^s(y) = N \exp\left[\frac{-2}{\sigma_1^2}\left\{-y + 1 + \left(\frac{\sigma_2}{\sigma_1} - \mu\frac{\sigma_1}{\sigma_2}\right)\left(\tan^{-1}\left(\frac{\sigma_1}{\sigma_2}y\right) - \tan^{-1}\left(\frac{\sigma_1}{\sigma_2}\right)\right) + \frac{\sigma_1^2 + 2}{4}\ln\left(\frac{\sigma_1^2 y^2 + \sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)\right\}\right], \quad (7)$$

and for $y > 1$,

$$p^s(y) = N \exp\left[\frac{-2}{\sigma_1^2}\left\{y - 1 - \left(\frac{\sigma_2}{\sigma_1} + \mu\frac{\sigma_1}{\sigma_2}\right)\left(\tan^{-1}\left(\frac{\sigma_1}{\sigma_2}y\right) - \tan^{-1}\left(\frac{\sigma_1}{\sigma_2}\right)\right) + \frac{\sigma_1^2 - 2}{4}\ln\left(\frac{\sigma_1^2 y^2 + \sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)\right\}\right], \quad (8)$$

where N is a normalisation constant. The stationary PDF has three extrema for

$$-2 + 4\sqrt{\mu} < \sigma_1^2 < 2\mu < 2, \quad (9)$$

two maxima at

$$y_- = \frac{1}{2} + \frac{\sigma_1^2}{4} - \sqrt{\left(\frac{1}{2} + \frac{\sigma_1^2}{4}\right)^2 - \mu} \quad (10)$$

$$y_+ = \frac{1}{2} - \frac{\sigma_1^2}{4} + \sqrt{\left(\frac{1}{2} - \frac{\sigma_1^2}{4}\right)^2 + \mu}, \quad (11)$$

$$(12)$$

and a local minimum at

$$y_o = \frac{1}{2} + \frac{\sigma_1^2}{4} + \sqrt{\left(\frac{1}{2} + \frac{\sigma_1^2}{4}\right)^2 - \mu}. \quad (13)$$

The peak around y_-, corresponding to a relatively weak meridional salinity gradient and strong overturning circulation, most closely corresponds to the present state of the THC. Note that the strength of the fluctuations in freshwater flux, σ_2, plays no role in determining the number or location of peaks in the system; this is because the white noise process $\dot{W}_2$ enters (5) additively. We will be primarily concerned with the range of parameter values within which p^s is bimodal; the populations corresponding to each of the peaks of the distribution will be referred to as *regimes*.

Based on an analysis of the average rate of transitions between regimes, it was argued in M02 that the physically-relevant range of noise strengths is $\sigma_1 < 0.2$,

$\sigma_2 < 0.2$. It was further noted that for these noise levels, for most of the range of μ in which p^s is bimodal, most of the probability mass of p^s is concentrated in one of the two regimes. This is illustrated in Figure 2. Contoured in Figure 2(a) is p^s as a function of μ for $\sigma_1 = \sigma_2 = 0.1$. The thick black lines denote the range of μ within which p^s is bimodal. For lower values of μ, the density is concentrated in the regime around y_-; there is a small window of μ within which the two regimes have comparable populations; and for higher values of μ, p^s is concentrated around y_+. Thus, for for a broad range of values of μ, while p^s is *technically* bimodal, it is *effectively* unimodal. In M02, this effect was denoted *stabilisation* by noise. Regime stabilisation is also apparent in Figure 2(b), which contours p^s as a function of σ_2 for $\mu = 0.205, \sigma_1 = 0.1$. For low values of σ_1, the regime around y_- dominates. The regime around y_+ dominates as σ_2 is increased, although the regime around y_- begins to be repopulated for larger σ_2.

The partitioning of probability mass between the two regimes is a complicated function of the parameters $(\mu, \sigma_1, \sigma_2)$. It can be characterised by the surface $\mu_{0.5}(\sigma_1, \sigma_2)$, defined as the value of μ for given (σ_1, σ_2) at which the probability mass is equal on either side of y_o:

$$\int_{-\infty}^{y_o} dy\, p^s(y; \sigma_1, \sigma_2, \mu_{0.5}) = \int_{y_o}^{\infty} dy\, p^s(y; \sigma_1, \sigma_2, \mu_{0.5}). \tag{14}$$

This surface is contoured in Figure 3. For $\mu < \mu_{0.5}$, the regime around y_- is more populated than that around y_+; the converse is true for $\mu > \mu_{0.5}$.

The asymptotic behaviour of a deterministic system with multiple attractors is determined by the basin of attraction in which the initial conditions lie. With the addition of stochastic fluctuations, the long-time behaviour is characterised by the stationary distribution, irrespective of the initial conditions. In the present example, stabilisation by noise preferentially populates one regime in p^s, leading to very different long-time behaviour than that characterising the deterministic system. However, this is only physically relevant if the characteristic escape time of the regime within which the system starts is substantially smaller than the relevant timescales of the physical phenomenon under consideration. If the escape time from the regime in

which the system started is longer than, for example, the lifetime of the Earth, then for all intents and purposes the system will remain in this regime, and the stationary PDF is not relevant for the discussion of the system. The regime escape times for the system (5) can be calculated; it was shown in M02 that for a broad range of noise parameters the escape times are smaller than relevant timescales of THC dynamics, so stabilisation of regimes is relevant.

Rahmstorf [27] demonstrated the existence of multiple THC regimes in a full oceanic GCM, and showed furthermore that the hysteresis loop obtained by varying the freshwater forcing parameter (equivalent to μ in the present study) above and below the region in which there exist multiple regimes could be fit to the hysteresis curve obtained from a simple Stommel-like model. By construction, noise was suppressed in this OGCM experiment; transitions between regimes were interpreted as occurring at deterministic bifurcation points. However, the ocean-atmosphere system contains substantial high-frequency variability, and we have seen that the presence of noise in a simple Stommel model can stabilise one regime relative to the other. Figures 2 and 3 illustrate that the stabilisation of a regime generally occurs some distance from the deterministic bifurcation points. Consequently, we would expect that fluctuations can induce transitions between regimes before the deterministic bifurcation points are reached. This is demonstrated in Figure 4, which compares the deterministic and the stochastic hysteresis curves obtained at two different sets of noise levels (see Figure caption). As anticipated, the hysteresis loops in the presence of fluctuations are substantially narrowed relative to the deterministic loop. Transitions between regimes occur not at the deterministic bifurcation points, but at random points within the region in which the regime to which the transition occurs has been stabilised. A rigorous analysis of the effects of noise on the hysteresis structure of nonlinear systems is given by Berglund and Gentz ([4], [5])

In the following two sections, we will consider the effects on the stationary distribution p^s of relaxing some of the assumptions used to obtain equation (5). First, we will investigate the effects of allowing the fluctuations in internal mixing and in freshwater forcing to be correlated. Secondly, the effects of a nonzero autocorrelation

time τ in the mixing process η will be addressed.

3 Correlated Mixing and Freshwater Flux Fluctuations

In the previous section, the fluctuations in internal mixing and freshwater forcing were represented as independent white noise processes. To a certain extent, however, the high-frequency variability parameterised by these processes shares a common source: internal atmospheric dynamics. The processes parameterised by the fluctuating mixing $\sigma_1 \dot{W}_1$ include those associated with the wind-driven circulation, and the fluctuations in freshwater forcing $\sigma_2 \dot{W}_2$ represent, among other processes, variations in atmospheric moisture transport associated with mid-latitude storminess. Other processes drive fluctuations in internal eddy activity (e.g. instability of oceanic jets) and freshwater fluxes (e.g. ice melting/freezing), so these fluctuations are not entirely mutually dependent. We proceed to investigate the extent to which mutual dependence between the mixing and freshwater forcing fluctuations changes the results described in the previous section.

The simplest model of dependence between $\dot{W}_1$ and $\dot{W}_2$ is linear dependence, i.e.,

$$W_2 = \gamma W_1 + (1-\gamma^2)^{1/2} W_3, \tag{15}$$

where W_3 is a Wiener process independent of W_1 and $\gamma \in [-1, 1]$. The SDE for y can then be written:

$$\dot{y} = -|1-y|y + \mu + (-\sigma_1 y + \gamma\sigma_2) \circ \dot{W}_1 + (1-\gamma^2)^{1/2}\sigma_2 \dot{W}_3. \tag{16}$$

The associated Fokker-Planck equation for the stationary density p^s is:

$$0 = -\left(-|1-y|y + \mu + \frac{1}{2}\sigma_1^2 y - \frac{1}{2}\gamma\sigma_1\sigma_2\right) p^s + \frac{1}{2}\frac{d}{dy}\left([\sigma_1^2 y^2 - 2\gamma\sigma_1\sigma_2 y + \sigma_2^2]p^s\right), \tag{17}$$

which can be expressed

$$\frac{dp^s}{p^s} = 2\left(\frac{-|1-y|y + \mu - \frac{1}{2}\sigma_1^2 y + \frac{1}{2}\gamma\sigma_1\sigma_2}{\sigma_1^2 y^2 - 2\gamma\sigma_1\sigma_2 y + \sigma_2^2}\right) dy. \tag{18}$$

The right-hand side of (18) can be integrated to obtain a closed-form expression for the stationary distribution p^s: for $y < 1$:

$$\begin{aligned} p^s(y) &= N \exp\left[-\frac{2}{\sigma_1^2} \left\{ 1 - y + \left(-\gamma \frac{\sigma_2}{\sigma_1} + \frac{2+\sigma_1^2}{4} \right) \ln \left(\frac{\sigma_1^2 y^2 - 2\gamma\sigma_1\sigma_2 y + \sigma_2^2}{\sigma_1^2 - 2\gamma\sigma_1\sigma_2 + \sigma_2^2} \right) \right.\right. \\ &\quad + \frac{1}{\sqrt{1-\gamma^2}} \left(\frac{\sigma_2}{\sigma_1}(1-2\gamma^2) + \gamma - \frac{\sigma_1}{\sigma_2}\mu \right) \\ &\quad \left.\left. \times \left(\tan^{-1}\left(\frac{\sigma_1 y - \sigma_2\gamma}{\sigma_2\sqrt{1-\gamma^2}} \right) - \tan^{-1}\left(\frac{\sigma_1 - \sigma_2\gamma}{\sigma_2\sqrt{1-\gamma^2}} \right) \right) \right\} \right], \end{aligned} \tag{19}$$

and for $y > 1$:

$$\begin{aligned} p^s(y) &= N \exp\left[-\frac{2}{\sigma_1^2} \left\{ y - 1 + \left(\gamma \frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2-2}{4} \right) \ln \left(\frac{\sigma_1^2 y^2 - 2\gamma\sigma_1\sigma_2 y + \sigma_2^2}{\sigma_1^2 - 2\gamma\sigma_1\sigma_2 + \sigma_2^2} \right) \right.\right. \\ &\quad - \frac{1}{\sqrt{1-\gamma^2}} \left(\frac{\sigma_2}{\sigma_1}(1-2\gamma^2) + \gamma + \frac{\sigma_1}{\sigma_2}\mu \right) \\ &\quad \left.\left. \times \left(\tan^{-1}\left(\frac{\sigma_1 y - \sigma_2\gamma}{\sigma_2\sqrt{1-\gamma^2}} \right) - \tan^{-1}\left(\frac{\sigma_1 - \sigma_2\gamma}{\sigma_2\sqrt{1-\gamma^2}} \right) \right) \right\} \right], \end{aligned} \tag{20}$$

where N is a normalisation constant. Clearly, in the limit that $\gamma \to 0$, (19) and (20) reduce to (7) and (8). It is also clear that the correlation between fluctuations in freshwater forcing and mixing has a non-trivial effect on the the structure of the stationary PDF.

Extrema of p^s occur when the numerator of the RHS of equation (18) vanishes. The PDF is bimodal when the following inequalities obtain:

$$\sigma_1^2 - \gamma\sigma_1\sigma_2 < 2\mu < 2\left(\frac{1}{2} + \frac{\sigma_1^2}{4} \right)^2 - \gamma\sigma_1\sigma_2 \tag{21}$$

$$\sigma_1^2 < 2 \tag{22}$$

$$-\mu < \left(\frac{1}{2} - \frac{\sigma_1^2}{4} \right)^2 + \frac{1}{2}\gamma\sigma_1\sigma_2. \tag{23}$$

In this range of parameter values, p^s has maxima at :

$$y_- = \frac{1}{2} + \frac{\sigma_1^2}{4} - \sqrt{\left(\frac{1}{2} + \frac{\sigma_1^2}{4} \right) - \mu - \frac{1}{2}\gamma\sigma_1\sigma_2} \tag{24}$$

$$y_+ = \frac{1}{2} - \frac{\sigma_1^2}{4} + \sqrt{\left(\frac{1}{2} - \frac{\sigma_1^2}{4} \right) + \mu + \frac{1}{2}\gamma\sigma_1\sigma_2}, \tag{25}$$

and a local minimum at

$$y_o = \frac{1}{2} + \frac{\sigma_1^2}{4} + \sqrt{\left(\frac{1}{2} + \frac{\sigma_1^2}{4}\right) - \mu - \frac{1}{2}\gamma\sigma_1\sigma_2}. \tag{26}$$

When $\gamma = 0$, the strength of fluctuations in freshwater forcing, σ_2, has no effect on the number or location of extrema in p^s. The coupling of mixing and freshwater forcing fluctuations, however, induces a dependence on both γ and σ_2 of the parameter range in which there is bimodality, through the combination $\gamma\sigma_2$. Figure 5 displays the phase diagram of the system as a function of μ, σ_1, and $\gamma\sigma_2$. Within the shaded area, p^s is bimodal, while outside it is unimodal. The solid black line outlines the region of multiple equilibria in the limit of uncorrelated fluctuations, $\gamma = 0$. The coupling of mixing and freshwater forcing fluctuations greatly increases the range of values of μ for which there exists some set of parameter values $(\gamma, \sigma_1, \sigma_2)$ such that p^s is bimodal.

Regime stabilisation still occurs for $\gamma \neq 0$; indeed, allowing γ to be nonzero introduces another parameter which affects the partition of the probability mass of p^s between the regimes around y_- and y_+. Figure 6(a) contours p^s as a function of γ for $(\mu, \sigma_1, \sigma_2) = (0.17, 0.05, 0.1)$. As γ increases, the probability mass shifts from the regime around y_- to that around y_+. However, increasing γ can also shift probability mass from the y_+ regime to that around y_-, as is illustrated in Figure 6(b) for $(\mu, \sigma_1, \sigma_2) = (0.2, 0.1, 0.05)$.

A comprehensive picture of the partitioning of probability mass between the two regimes can be obtained by considering the surface $\mu_{0.5}(\sigma_1, \sigma_2, \gamma)$, defined as in equation (14). Figure 7 contours $\mu_{0.5}$ as a function of σ_1 and σ_2 for different values of γ; clearly, γ has a nontrivial quantitative effect on $\mu_{0.5}$. Generally speaking, as γ is decreased (increased) from zero, $\mu_{0.5}$ becomes a less (more) sensitive function of σ_1 and σ_2, and the range of values taken by $\mu_{0.5}$ decreases (increases). Indeed, increasing γ above zero introduces ranges of σ_1, σ_2 in which $\mu_{0.5}$ is lower than any value obtained for $\gamma = 0$.

It was demonstrated in M02 that the partitioning of probability mass between THC regimes is a complicated function of the noise strengths σ_1, σ_2. The preceding analysis demonstrates that regime populations can also depend sensitively on the

strength of the correlation γ between the fluctuations in mixing and freshwater forcing. The next section considers the effects on p^s of a finite autocorrelation timescale in the mixing fluctuations.

4 Red Noise Mixing

Generally speaking, the timescales of internal ocean eddy variability are greater than those of atmospheric eddy variability. In the analyses presented above, the noise processes associated with internal oceanic eddy mixing and freshwater flux have both been represented as white-noise processes. As was anticipated in equations (2)-(4), eddy mixing may be more appropriately represented as a process with a nonzero autocorrelation time. In this section, we will consider the effects of red-noise internal mixing on the structure of the stationary PDF of y.

We consider the system (2)-(3) in the limits that $\lambda \to \infty$ and that W_1 and W_2 are independent. We then obtain the SDE in the variables (y, η):

$$\dot{y} = -|1-y|y - \eta y + \mu + \sigma_2 \dot{W}_2 \tag{27}$$

$$\dot{\eta} = -\frac{1}{\tau}\eta + \frac{\sigma_1}{\tau}\dot{W}_1. \tag{28}$$

Denoting by $q^s(y, \eta)$ the stationary joint PDF of y and η, the associated stationary FPE is

$$0 = \partial_y[(|1-y|y + \eta y - \mu)q^s] + \partial_\eta \left[\frac{1}{\tau}\eta q^s\right] + \frac{1}{2}\sigma_2^2 \partial_{yy} q^s + \frac{1}{2}\frac{\sigma_1^2}{\tau^2}\partial_{\eta\eta} q^s. \tag{29}$$

The stationary distribution of y, $p^s(y)$, is the marginal distribution:

$$p^s(y) = \int_{-\infty}^{\infty} d\eta \; q^s(y, \eta) \tag{30}$$

For $\tau \neq 0$, the stationary FPE (29) is a partial differential equation in two variables which does not admit an analytic solution. This system, with $\sigma_2 = 0$, was considered in Timmermann and Lohmann [36] and Monahan et al. [22]. The first of these studies employed an approximation scheme known as the Unified Colored Noise Approximation (UCNA, [17]) to obtain the stationary distribution of y. However,

as was noted in Monahan at al. [22], the UCNA is only valid for τ much less than the timescale of the deterministic dynamics of y; in particular, it breaks down for $\tau \simeq O(1)$ and greater. These are values of τ which are in principle of interest, and thus other approaches must be found to estimate $p^s(y)$. Unfortunately, as is discussed in Hänggi and Jung [13], analytic methods for the study of the stationary distribution of a system subject to red-noise fluctuations do not exist for general τ, and so we must take recourse to numerical schemes. Before presenting the results of the numerical analysis, we will consider the dynamics of the deterministic component of the system (27)-(28), and investigate the structure of the PDFs obtained by linearising these equations around the deterministic fixed points.

4.1 Red Noise Heuristics

As is described in Hänngi and Jung [13], some intuition into the system (27)-(28) can be obtained by considering the vector field associated with the deterministic drift:

$$\begin{aligned} \mathbf{f} &= (f_y, f_\eta) \\ &= \left(-|1-y|y - \eta y + \mu, -\frac{1}{\tau}\eta \right). \end{aligned} \tag{31}$$

For $\mu \in (0, .25)$, the vector field has two stable fixed points: $(y, \eta) = (y_-, 0)$ and $(y_+, 0)$, and one unstable fixed point: $(y, \eta) = (y_o, 0)$, where

$$y_- = \frac{1}{2} - \sqrt{\frac{1}{4} - \mu} \tag{32}$$

$$y_+ = \frac{1}{2} + \sqrt{\frac{1}{4} + \mu} \tag{33}$$

$$y_o = \frac{1}{2} + \sqrt{\frac{1}{4} - \mu}. \tag{34}$$

$$\tag{35}$$

The fixed points y_o and y_+ meet at a bifurcation point at $\mu = 0$; below this value of μ, only the stable fixed point y_- survives. Similarly, y_o and y_- meet in a saddle-node bifurcation at $\mu = 0.25$; above this value, only the stable fixed point y_+ survives. The bifurcation diagram is illustrated in Figure 8. Figure 9 displays a plot of $\mathbf{f}$ for

$\mu = 0.19, \tau = 1$. The direction of $\mathbf{f}$ is indicated by arrows, and its magnitude is contoured. The solid dots indicate the fixed points. The thick line is the stable manifold of the unstable fixed point $(y_o, 0)$, which is also the separatrix of the basins of attraction of the stable fixed points; this curve satisfies the differential equation:

$$\frac{d\eta}{dy} = \frac{1}{\tau} \frac{\eta}{|1-y|y + \eta y - \mu} \quad , \quad \eta(y_o) = 0 \tag{36}$$

Figure 10 displays the vector field $\mathbf{f}$ for $\mu = 0.19$ and $\tau = 0.05, 0.5, 5, 50$. For $\tau << 1$, $|f_\eta| >> |f_y|$ except in a boundary layer around the line $\eta = 0$. For $\tau \sim O(1)$, the components f_η and f_y are of comparable magnitude over large regions of the state space. For $\tau >> 1$, $|f_y| >> |f_\eta|$ except in a boundary layer around the line $f_y = 0$, that is, near $\eta = \mu/y - |1-y|$.

The full SDE (27)-(28) can be interpreted as a diffusion in this vector field. The vector field $\mathbf{f}$ drives the system toward the fixed points (via boundary layers for $\tau << 1$ and $\tau >> 1$), while the stochastic fluctuations drive it away; the stationary distribution arises as the equilibrium between these tendencies. Transitions between regimes occur when a trajectory crosses the separatrix. Note that while τ has no effect on the number or location of the fixed points, it has a substantial effect on the structure of $\mathbf{f}$.

4.2 Small Noise Approximation

For small values of the noise strengths σ_1 and σ_2, most of the mass of the stationary joint PDF $q^s(y, \eta)$ will be concentrated around the deterministic fixed points. In this limit, a local description of the dynamics of the nonlinear SDE (28)-(27) around each of these fixed points is provided by the corresponding linearised system.

We consider first the linearised dynamics around the fixed point $(y_+, 0)$. Defining $\hat{y} = y - y_+$, $\hat{\eta} = \eta - 0$, the linearised system is:

$$\frac{d}{dt} \begin{pmatrix} \hat{y} \\ \hat{\eta} \end{pmatrix} = \begin{pmatrix} 1 - 2y_+ & -y_+ \\ 0 & -\frac{1}{\tau} \end{pmatrix} \begin{pmatrix} \hat{y} \\ \hat{\eta} \end{pmatrix} + \begin{pmatrix} \sigma_2 & 0 \\ 0 & \frac{\sigma_1}{\tau} \end{pmatrix} \begin{pmatrix} \dot{W}_1 \\ \dot{W}_2 \end{pmatrix} \tag{37}$$

We denote the matrices associated with the linearised drift and diffusion by A and B respectively. The stationary covariance of the linearised SDE, C_o, satisfies the

Lyapunov equation [11]:

$$AC_o + C_o A^T + BB^T = 0. \tag{38}$$

This equation can be solved using the eigenvalue decomposition of A:

$$A = u\Lambda v^T \tag{39}$$

where

$$u = \frac{1}{1-2y_+ + \frac{1}{\tau}} \begin{pmatrix} 1 & y_+ \\ 0 & 1-2y_+ + \frac{1}{\tau} \end{pmatrix} \tag{40}$$

$$\Lambda = \begin{pmatrix} 1-2y_+ & 0 \\ 0 & -\frac{1}{\tau} \end{pmatrix} \tag{41}$$

$$v = \begin{pmatrix} 1-2y_+ + \frac{1}{\tau} & 0 \\ -y_+ & 1 \end{pmatrix}, \tag{42}$$

and

$$uv^T = u^T v = I. \tag{43}$$

The Lyapunov equation (38) can be expressed:

$$D_{ij} = -\frac{(v^T BB^T v)_{ij}}{\lambda_i + \lambda_j} \tag{44}$$

where

$$D = v^T C_o v, \tag{45}$$

and $\lambda_1 = 1 - 2y_+$, $\lambda_2 = -1/\tau$ are the eigenvalues of A. The stationary covariance matrix C_o follows from the relation

$$C_o = uDu^T. \tag{46}$$

After some algebra, we obtain

$$C_o = \begin{pmatrix} \frac{1}{2(2y_+-1)}\left(\sigma_2^2 + \frac{y_+^2\sigma_1^2}{1+\tau(2y_+-1)}\right) & -\frac{1}{2}\frac{\sigma_1^2 y_+}{1+\tau(2y_+-1)} \\ -\frac{1}{2}\frac{\sigma_1^2 y_+}{1+\tau(2y_+-1)} & \frac{\sigma_2^2}{2\tau} \end{pmatrix} \tag{47}$$

For perturbations sufficiently small that the linearised system is a good approximation to the full system, the stationary PDF of y around y_+ can be approximated as a

Gaussian with mean y_+ and variance

$$\mathrm{var}(y) = \frac{1}{2(2y_+ - 1)} \left(\sigma_2^2 + \frac{y_+^2 \sigma_1^2}{1 + \tau(2y_+ - 1)} \right) \tag{48}$$

A similar calculation for the system linearised around the fixed point $(y_-, 0)$ yields the covariance matrix:

$$C_o = \begin{pmatrix} \frac{1}{2(1-2y_-)} \left(\sigma_2^2 + \frac{\sigma_1^2 y_-^2}{1+\tau(1-2y_-)} \right) & -\frac{1}{2} \frac{\sigma_1^2 y_-}{1+\tau(1-2y_-)} \\ -\frac{1}{2} \frac{\sigma_1^2 y_-}{1+\tau(1-2y_-)} & \frac{\sigma_2^2}{2\tau} \end{pmatrix} \tag{49}$$

Thus, for noise levels sufficiently small that the mass of the stationary PDF is concentrated around the deterministic fixed points, the width of the PDFs around the fixed points decreases monotonically as τ is increased.

While the linearisations can approximate the shapes of the peaks of the stationary PDF around the fixed points, they cannot estimate the relative amplitudes of these peaks. The evolution of the fraction of the total probability mass around y_-, denoted N_-, can be obtained from the FPE (29) by integrating over the region $y < y_o$:

$$\frac{d}{dt} N_- = \int_{-\infty}^{y_o} dy \int_{-\infty}^{\infty} d\eta \; \nabla \cdot \mathbf{J}, \tag{50}$$

where the FPE has been expressed in terms of the divergence of the probability density current $\mathbf{J}$ (see Appendix). Using the divergence theorem, equation (50) can be expressed as a surface integral over the line $y = y_o$:

$$\frac{d}{dt} N_- = \int_{-\infty}^{\infty} d\eta \; J_y(y_o, \eta) \tag{51}$$

The stationary distribution is such that the total probability flux across this line vanishes. Thus, the partitioning of probability mass between the two regimes depends on the structure of q^s in the neighbourhood of the line $y = y_o$; in this region, the PDFs obtained from the linearised systems cannot be expected to be accurate. To estimate the global structure of the stationary distribution, we turn in the next subsection to numerical techniques. Of course, if the escape time of the system from the regime in which it started is substantially greater than any relevant THC timescale, then for all intents and purposes the system will remain in this regime and the PDF of the linearised system will be a good approximation.

4.3 Numerical Estimates of $p^s(y)$

Approximate solutions q^s to the Fokker-Planck equation (29) are obtained using a finite-difference scheme described in the Appendix. Estimates of the stationary PDF can also be obtained through long numerical integrations of the original SDE (27)-(28), as in Monahan et al. [22]. This approach is not practical, however, for lower noise levels as the characteristic escape times of the regimes become very large, and very long integrations need to be carried out to obtain good estimates of the stationary distribution. For the noise levels of interest in the present study, it is much more efficient to estimate q^s directly from the Fokker-Planck equation.

Figure 11(a) contours p^s as a function of μ for $\sigma_1 = \sigma_2 = 0.1, \tau = 1$. Comparing this Figure with Figure 2, it is seen that there is a slight quantitative change, but no qualitative change, in regime stabilisation. The value of τ can have an effect on the stabilisation of regimes, as is illustrated in Figure 11(b), which contours p^s as a function of τ for $\mu = 0.175$ and $\sigma_1 = \sigma_2 = 0.075$. A more comprehensive picture of the effect of τ on the partitioning of the probability mass between THC regimes is presented in Figure 12, which contours the surface $\mu_{0.5}$ as a function of σ_1 and σ_2 for $\tau = 0.1, 0.5, 1, 5$. Inspection of this Figure indicates that, in general, as τ increases, $\mu_{0.5}$ decreases and becomes a less sensitive function of σ_1.

Figure 13(a) plots the numerical estimate of p^s for $\mu = 0.235$ and $\sigma_1 = \sigma_2 = 0.1$, along with the approximation from the linearised system calculated in the previous subsection. The numerical estimate and the linearised approximation are essentially indistinguishable. Plotted in Figure 13(b) are the numerical estimate and linearised approximation of p^s for $\mu = 0.19$ and $\sigma_1 = \sigma_2 = 0.1$. For the linearised approximation, the relative amplitudes of the two peaks was estimated from the numerical estimate of p^s. While the linearised approximations cannot estimate the partitioning of probability mass between the two regimes, they provide a reasonable approximation of the shape of the distribution within each regime.

Figure 5 demonstrates that the presence of multiplicative white noise can produce peaks of the stationary PDF that do not correspond to stable fixed points of the

deterministic system. The same is true of red noise, as is illustrated in Figure 14. This Figure plots p^s for $\tau = 1$, $\mu = 0.255$, $\sigma_2 = 0.15$, and $\sigma_1 = 0.1, 0.2$, and 0.3. At this value of μ, the deterministic vector field has only the single fixed point $(y_+, 0)$. For $\sigma_1 = 0.1$, p^s is unimodal, with probability mass concentrated around y_+. As σ_1 increases, a distinct shoulder develops in p^s at low values of y. For $\sigma_1 = 0.3$, a second peak has appeared which does not correspond to a fixed point of the deterministic system.

A heuristic explanation of the origin of this second peak in p^s follows from inspection of the vector field $\mathbf{f}$; this is plotted in Figure 15. The magnitude of $\mathbf{f}$ has two minima: one at the fixed point $(y_+, 0)$, and a second near $(y, \eta) = (0.5, 0)$. Near the second minimum, the deterministic trajectories slow down without stopping. While trajectories starting from all initial conditions in the (y, η) state space are ultimately driven toward the fixed point $(y_+, 0)$, there are regions of the state space in which the paths to the fixed point proceed via the region around $(0.5, 0)$ in which the speed has a local minimum. In the presence of fluctuations, the system will occasionally be driven into the parts of state space that are attracted to the fixed point via this region of reduced speed. Passing through this region, probability mass is accumulated as the trajectories slow down. For low noise levels, trajectories rarely escape from the region around the fixed point, and the probability mass around the region of reduced speed is small. As the noise level is increased, trajectories pass through the low speed region more frequently and the probability mass in this area increases. Figure 16 displays a sample trajectory of 50 time units duration in the (y, η) state space. The transition from the regime around y_+ occurs when the trajectory moves into a region of the state space in which $\mathbf{f}$ directs it toward the reduced-speed region. The trajectory moves around in this region for some time before gradually moving back toward the fixed point.

In the presence of red noise, then, the phenomenon of stabilisation is qualitatively unchanged. Quantitatively, increasing τ tends to shift the mass of the stationary distribution p^s to higher values of y, or, equivalently, to reduce $\mu_{0.5}$. Furthermore, as is the case for multiplicative white noise, the red noise fluctuations can induce

peaks in the stationary distribution that do not correspond to fixed points of the associated deterministic system. This can also occur in a deterministic system with many degrees of freedom (e.g. [23]).

5 Summary and Conclusions

In M02, the phenomenon of regime stabilisation by noise in a simple model of the THC was noted, and it was demonstrated that stabilisation can have a marked effect on diagnoses of the stability of the present climate. By relaxing some of the assumptions used to obtain the model considered in M02, the present study extends this earlier work. The effects on the partition of probability mass between THC regimes of nonzero correlations between fluctuations in internal oceanic eddy mixing and freshwater fluxes from the atmosphere, and of eddy mixing with a nonzero autocorrelation time, were considered using analytical and numerical approaches. It was demonstrated that while regime stabilisation was qualitatively unchanged, there were potentially substantial quantitative changes in the partitioning of probability mass between THC regimes.

The possibility that the THC can display rapid shifts between very different regimes of circulation is of interest in the study of both past and future climates, and in particular response of the climate system to anthropogenic forcing (e.g. [15], [16]). The structure of multiple regimes of thermohaline circulation has been investigated using models throughout the hierarchy of climate models, from simple conceptual models such as the one in this study (e.g., [8], [27], [32], [35]), through models of intermediate complexity (e.g., [10], [30]), to full GCMs (e.g., [14], [20], [26], [27], [37]). In general, the models of intermediate complexity and GCMs used in the study of THC regimes are purely deterministic; the effects of variability not explicitly resolved by the model are ignored or represented by deterministic functions of the resolved variables. A generic feature of such models is the systematic underestimation of internal variability in the system [3]. Some studies have considered the effects on THC variability of high-frequency fluctuations parameterised as stochastic processes

(e.g. [1], [24], [38], [41]), but did not undertake extensive studies of the dependence of this variability on noise parameters. The study of M02 demonstrated that the regime dynamics of the system can depend sensitively on the strength of fluctuations in the system; the present study demonstrates a further sensitivity to correlations within and between fluctuations. These results point to the potential importance of the accurate representation of the details of internal variability in sophisticated models of the climate system.

Acknowledgements

The author would like to thank Peter Imkeller for his suggestions and helpful comments. This work was supported by DFG Schwerpunktprogramm "Interagierende stochastische Systeme von hoher Komplexität".

Appendix: Numerical Scheme

This appendix describes the numerical scheme used to solve the stationary Fokker-Planck Equation (29). The y and η coordinates were discretised on uniform meshes:

$$\eta_i = \eta_{min} + (i-1)\delta_\eta \qquad i = 1, ..., M \tag{52}$$

$$y_j = y_{min} + (j-1)\delta_y \qquad j = 1, ..., N. \tag{53}$$

The FPE for the joint PDF q^s can be written as the divergence in (η, y) space of a probability current $\mathbf{J} = (F, G)$:

$$0 = \partial_\eta F + \partial_y G, \tag{54}$$

where

$$F = \frac{1}{\tau}\eta q^s + \frac{1}{2}\frac{\sigma_1^2}{\tau^2}\partial_\eta q^s \tag{55}$$

$$G = (|1-y|y + \eta y - \mu)q^s + \frac{1}{2}\sigma_2^2 \partial_y q^s. \tag{56}$$

The discrete FPE is

$$0 = \frac{F_{i+\frac{1}{2},j} - F_{i-\frac{1}{2},j}}{\delta_\eta} + \frac{G_{i,j+\frac{1}{2}} - G_{i,j-\frac{1}{2}}}{\delta_y}, \tag{57}$$

where

$$F_{i,j} = \frac{1}{\tau}\frac{\eta_{i+\frac{1}{2}}q^s_{i+\frac{1}{2},j} + \eta_{i-\frac{1}{2}}q^s_{i-\frac{1}{2},j}}{2} + \frac{1}{2}\frac{\sigma_1^2}{\tau^2}\frac{q^s_{i+\frac{1}{2},j} - q^s_{i-\frac{1}{2},j}}{\delta_\eta} \quad (58)$$

$$G_{i,j} = \frac{([|1-y_{j+\frac{1}{2}}| + \eta_i]y_{j+\frac{1}{2}} - \mu)q^s_{i,j+\frac{1}{2}} - ([|1-y_{j-\frac{1}{2}}| + \eta_i]y_{j-\frac{1}{2}} - \mu)q^s_{i,j-\frac{1}{2}}}{2} + \frac{1}{2}\sigma_2^2\frac{q^s_{i,j+\frac{1}{2}} - q^s_{i,j-\frac{1}{2}}}{\delta_y}. \quad (59)$$

The discretised FPE can then be written:

$$0 = a_{i,j}q^s_{i+1,j} + b_{i,j}q^s_{i,j+1} + c_{i,j}q^s_{i,j} + d_{i,j}q^s_{i-1,j} + e_{i,j}q^s_{i,j-1}, \quad (60)$$

where

$$a_{i,j} = \frac{1}{2\delta_\eta^2}\left(\frac{1}{\tau}\eta_{i+1}\delta_\eta + \frac{\sigma_1^2}{\tau^2}\right) \quad (61)$$

$$b_{i,j} = \frac{1}{2\delta_y^2}\left((|1-y_{j+1}|y_{j+1} + \eta_i y_{j+1} - \mu)\delta_y + \sigma_2^2\right) \quad (62)$$

$$c_{i,j} = -\frac{1}{\delta_\eta^2}\frac{\sigma_1^2}{\tau^2} - \frac{1}{\delta_y^2}\sigma_2^2 \quad (63)$$

$$d_{i,j} = \frac{1}{2\delta_\eta^2}\left(-\frac{1}{\tau}\eta_{i-1}\delta_\eta + \frac{\sigma_1^2}{\tau^2}\right) \quad (64)$$

$$e_{i,j} = \frac{1}{2\delta_y^2}\left(-(|1-y_{j-1}|y_{j-1} + \eta_i y_{j-1} - \mu)\delta_y + \sigma_2^2\right). \quad (65)$$

We impose the boundary conditions that the normal component of the probability flux vanishes at $\eta = \eta_{min}, \eta_{min} + M\delta_\eta; y = y_{min}, y_{min} + N\delta_y$. That is,

$$F_{\frac{3}{2},j} = F_{M-\frac{1}{2},j} = 0 \quad (66)$$

$$G_{i,\frac{3}{2}} = G_{i,N-\frac{1}{2}} = 0. \quad (67)$$

We must also apply the normalisation constraint

$$\sum_{i=1}^{M}\sum_{j=1}^{N} q^s_{i,j}\delta_\eta\delta_y = 1. \quad (68)$$

Together, equations (60),(66)-(67), and (68) define a linear system that can be inverted to solve for $q^s_{i,j}$.

For a given discretisation, this numerical approach can be expected to be most accurate for $\tau \sim O(1)$. As is illustrated in Figure 10, for both $\tau << 1$ and $\tau >> 1$, an important part of the deterministic component of the dynamics is concentrated in narrow boundary layers, the resolution of which requires a fine local mesh.

References

[1] M. Aeberhardt, M. Blatter, and T.F. Stocker, *Variability on the century time scale and regime changes in a stochastically forced zonally averaged ocean-atmosphere model, Geophys. Res. Lett.* **27** (2000) 1303.

[2] R. Alley *Ice-core evidence of abrupt climate changes, Proc. Nat. Acad. Science* **97** (2000) 1331.

[3] J.J. Barsugli and D.S. Battisti, *The basic effects of atmosphere-ocean thermal coupling on midlatitude variability, J. Atmos. Sci.* **55** (1998) 477.

[4] N. Berglund and B. Gentz, *Beyond the Fokker-Planck equation: Pathwise control of noisy bistable systems, J. Phys.* **A35** (2002) 2057.

[5] N. Berglund and B. Gentz, *The effect of additive noise on dynamical hysteresis, Nonlinearity* **15** (2002) 605.

[6] G. Bond and co-authors, *A pervasive millennial-scale cycle in North Atlantic Holocene and glacial climates, Science* **278** (1997) 1257.

[7] W. Broecker, *Chaotic climate, Sci. Amer.* **273** (1995) 44.

[8] P. Cessi, *A simple box model of stochastically-forced thermohaline flow, J. Phys. Oceanogr.* **24** (1994) 1911.

[9] P.U. Clark, N.G. Pisias, T.F. Stocker, and A.J. Weaver, *The role of the thermohaline circulation in abrupt climate change, Nature* **415** (2002) 863.

[10] A. Ganopolski and S. Rahmstorf, *Simulation of rapid glacial climate changes in a coupled climate model, Nature* **409** (2001) 153.

[11] C.W. Gardiner, **Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences** (Springer, 1997).

[12] S.M. Griffies and E. Tziperman, *A linear thermohaline oscillator driven by stochastic atmospheric forcing, J. Climate* **8** (1995) 2440.

[13] P. Hänggi and P. Jung, *Colored noise in dynamical systems, Adv. Chem. Phys.* **89** (1995) 239.

[14] T.M. Hughes and A.J. Weaver, *Multiple equilibria of an asymmetric two-basin ocean model, J. Phys. Oceanogr.* **24** (1994) 619.

[15] Intergovernmental Panel on Climate Change, *Workshop on rapid non-linear climate change, Tech. Report* (1998).

[16] Intergovernmental Panel on Climate Change, **Climate Change 2001: The Scientific Basis** (Cambridge University Press, 2001).

[17] P. Jung and P. Hänggi, *Dynamical systems: a unified coloured-noise approximation, Phys. Rev.* **A35** (1987) 4464.

[18] G. Lohmann and J. Schneider, *Dynamics and predictability of Stommel's box model: a phase space perspective with implications for decadal climate variability, Tellus* **48A** (1999) 326.

[19] L. Maas *A simple model for the three-dimensional, thermally and wind-driven ocean circulation, Tellus* **46A** (1994) 671.

[20] S. Manabe and R. Stouffer, *Two stable equilibria of a coupled ocean-atmosphere model, J. Climate* **1** (1988) 841.

[21] A.H. Monahan, *Stabilisation of climate regimes by noise in a simple model of the thermohaline circulation, J. Phys. Oceanogr.* **32** (2002) 2072.

[22] A.H. Monahan, A. Timmermann, and G. Lohmann, *Comments on "Noise-induced transitions in a simplified model of the thermohaline circulation", J. Phys. Oceanogr.* **32** (2002) 1112.

[23] H. Mukougawa, *A dynamical model of "quasi-stationary" states in large-scale atmospheric motions, J. Atmos. Sci.* **45** (1988) 2868.

[24] L. Mysak, T. Stocker, and F. Huang, *Century-scale variability in a randomly-forced, two-dimensional thermohaline ocean circulation model, Clim. Dyn.* **8** (1993) 103.

[25] C. Penland, *A stochastic approach to nonlinear dynamics: a review, Bull. Am. Met. Soc.*, submitted.

[26] S. Rahmstorf, *Bifurcations of the Atlantic thermohaline circulation in response to changes in the hydrological cycle, Nature* **378** (1995) 145.

[27] S. Rahmstorf, *On the freshwater forcing and transport of the Atlantic thermohaline circulation, Clim. Dyn.* **12** (1996) 799.

[28] S. Rahmstorf, *Decadal variability of the thermohaline ocean circulation,* in **Beyond El Niño: Decadal and Interdecadal Climate Variability**, ed. A. Navarra (Springer, 1999) pp. 309-332.

[29] S. Rahmstorf, *Rapid transitions of the thermohaline ocean circulation: a modelling perspective,* in **Reconstructing Ocean History: A Window Into the Future**, eds. F. Abrantes and A. Mix (Kluwer, 1999) pp. 139-149.

[30] A. Schmittner and A.J. Weaver, *Dependence of multiple climate states on ocean mixing parameters, Geophys. Res. Lett.* **28** (2001) 1207.

[31] W.J. Schmitz, *On the interbasin-scale thermohaline circulation, Rev. Geophys.* **33** (1995) 151.

[32] J. Scott, J. Marotzke, and P. Stone, *Interhemispheric thermohaline circulation in a coupled box model, J. Phys. Oceanogr.* **29** (1999) 351.

[33] T. Stocker and O. Marchal, *Abrupt climate change in the computer: Is it real?, Proc. Nat. Acad. Science* **97** (2000) 1362.

[34] H. Stommel and W. Young, *The average T-S relation of a stochastically-forced box model, J. Phys. Oceanogr.* **23** (1993) 151.

[35] H. Stommel, *Thermohaline convection with two stable regimes of flow, Tellus* **13** (1961) 224.

[36] A. Timmermann and G. Lohmann, *Noise-induced transitions in a simplified model of the thermohaline circulation, J. Phys. Oceanogr.* **30** (2000) 1891.

[37] E. Tziperman, *Proximity of the present-day thermohaline circulation to an instability threshold, J. Phys. Oceanogr.* **30** (2000) 90.

[38] X. Wang, P. Stone, and J. Marotzke, *Global thermohaline circulation. Part I: Sensitivity to atmospheric moisture transport, J. Climate* (1999) **12** 71.

[39] A.J. Weaver, C. Bitz, A. Fanning, and M. Holland, *Thermohaline circulation: High-latitude phenomena and the difference between the Pacific and Atlantic, Annu. Rev. Earth Planet. Sci.* **27** (1999) 231.

[40] A.J. Weaver and T. Hughes, *Stability and variability of the thermohaline circulation and its link to climate, Trends in Phys. Oceanogr.* **1** (1992) 15.

[41] A.J. Weaver, J. Marotzke, P.F. Cummins, and E. Sarachik, *Stability and variability of the thermohaline circulation, J. Phys. Oceanogr.* **23** (1993) 41.

Figure Captions

Figure 1: Schematic diagram of the Stommel two-box model.

Figure 2: Contour plot of p^s as a function of (a) μ for $\sigma_1 = \sigma_2 = 0.1$ (contours: 1,2,...,6) and (b) σ_2 for $\mu = 0.205$ and $\sigma_1 = 0.1$ (contours: 0.5,1,1.5,...,10.5)

Figure 3: Contour plot of $\mu_{0.5}$ as a function of σ_1 and σ_2.

Figure 4: Deterministic (thick lines) and stochastic (thin lines) hysteresis loops for (a) $\sigma_1 = \sigma_2 = 0.05$. (b) $\sigma_1 = \sigma_2 = 0.1$

Figure 5: Plot of the region of parameter space in which p^s is bimodal (shaded), as a function of μ, σ_1, and $\gamma\sigma_2$. The solid lines denote the boundary of the region of bimodality for $\gamma = 0$.

Figure 6: Contour plots of p^s as a function of γ for (a) $(\mu, \sigma_1, \sigma_2) = (0.17, 0.05, 0.1)$ and (b) $(\mu, \sigma_1, \sigma_2) = (0.2, 0.1, 0.05)$.

Figure 7: Contour plots of $\mu_{0.5}$ as a function of σ_1, σ_2, and γ.

Figure 8: Bifurcation diagram of the y component of the fixed points of the deterministic vector field (31).

Figure 9: Plot of the vector field $\mathbf{f}$ for $\mu = 0.19$, $\tau = 1$. The direction of $\mathbf{f}$ is indicated by the arrows; the magnitude is contoured. The solid dots indicate the fixed points, and the thick line is the stable manifold of the unstable fixed point. Contours: 0.1,0.3,0.5, ... The lowest contour surrounds the fixed points.

Figure 10: As in Figure 9, for $\tau = 0.05, 0.5, 5, 50$. For $\tau = 0.05$, contours: 1,2,3, ...

Figure 11: Contour plots of $p^s(y)$ as (a) a function of μ for $\tau = 1$, $\sigma_1 = \sigma_2 = 0.1$. Contours: 1,2,...,6 (b) a function of τ for $\mu = 0.175$, $\sigma_1 = \sigma_2 = 0.075$. Contours: 1,2, ... 8.

Figure 12: Contour plots of $\mu_{0.5}$ as a function of σ_1, σ_2 for $\tau = 0.1, 0.5, 1, 5$.

Figure 13: Numerical estimate of $p^s(y)$ (thick line) and linearised approximation (thin line) for $\sigma_1 = \sigma_2 = 0.1$ and (a) $\mu = 0.235$, (b) $\mu = 0.19$.

Figure 14: Plots of p^s for $\tau = 1$, $\mu = 0.255$, $\sigma_2 = 0.15$, and $\sigma_1 = 0.1$ (thin line), $\sigma_1 = 0.2$ (dashed line), and $\sigma_3 = 0.3$ (thick line).

Figure 15: As in Figure 9, for $\tau = 1$ and $\mu = 0.255$.

Figure 16: Sample trajectory in the (y, η) space for $\mu = 0.255$, $\tau = 1$,$\sigma_1 = 0.3$, and $\sigma_2 = 0.15$. The duration of the trajectory is 50 time units.

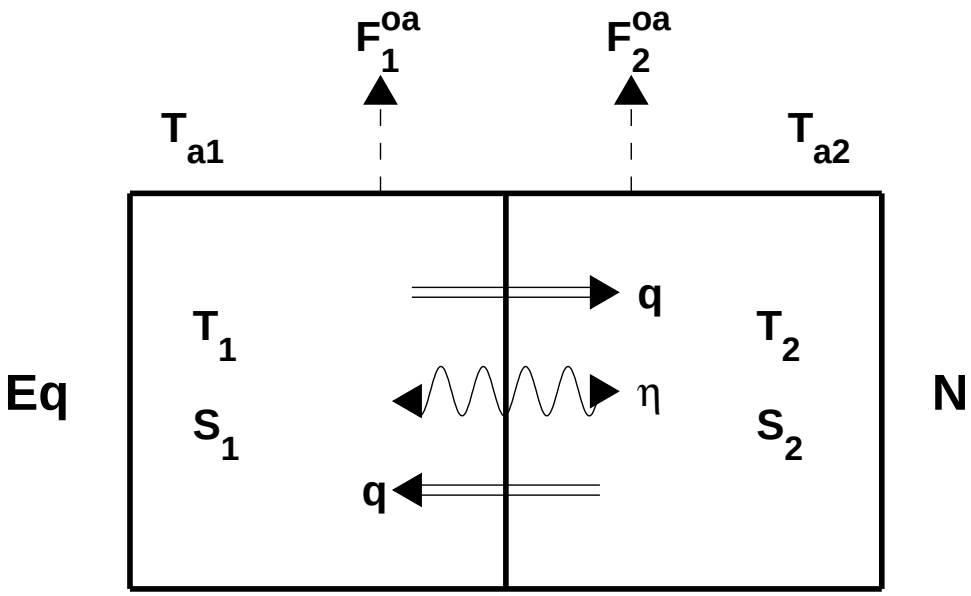

Figure 1: Schematic diagram of the Stommel two-box model.

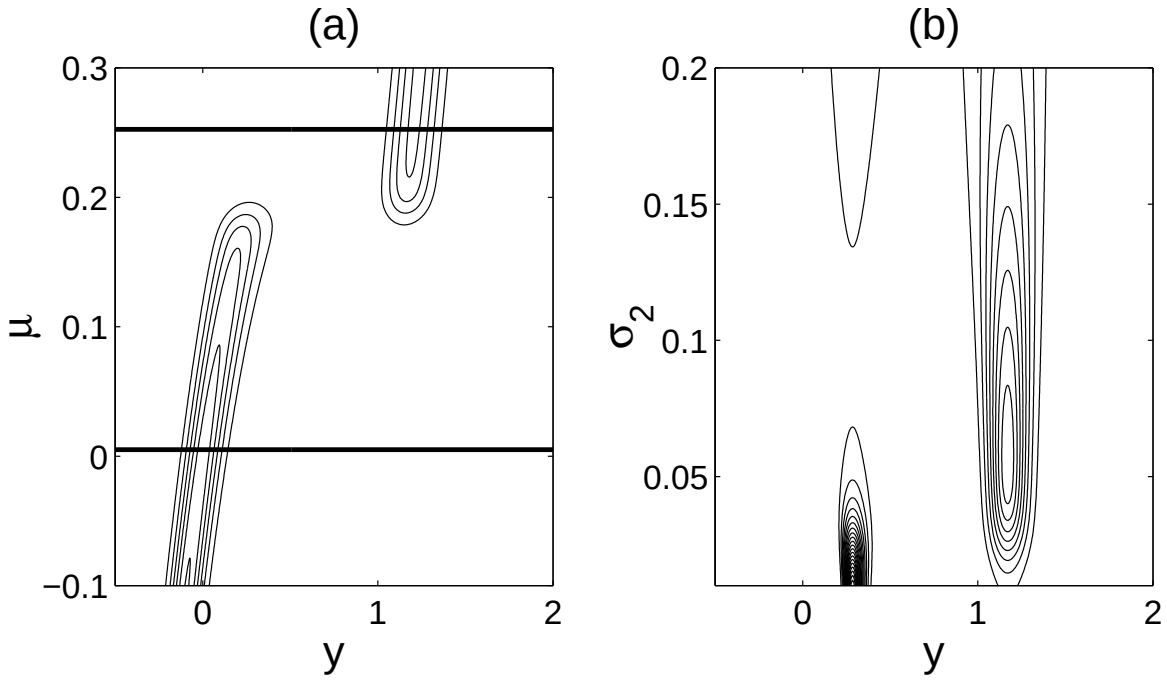

Figure 2: Contour plot of p^s as a function of (a) μ for $\sigma_1 = \sigma_2 = 0.1$ (contours: 1,2,...,6) and (b) σ_2 for $\mu = 0.205$ and $\sigma_1 = 0.1$ (contours: 0.5,1,1.5,...,10.5)

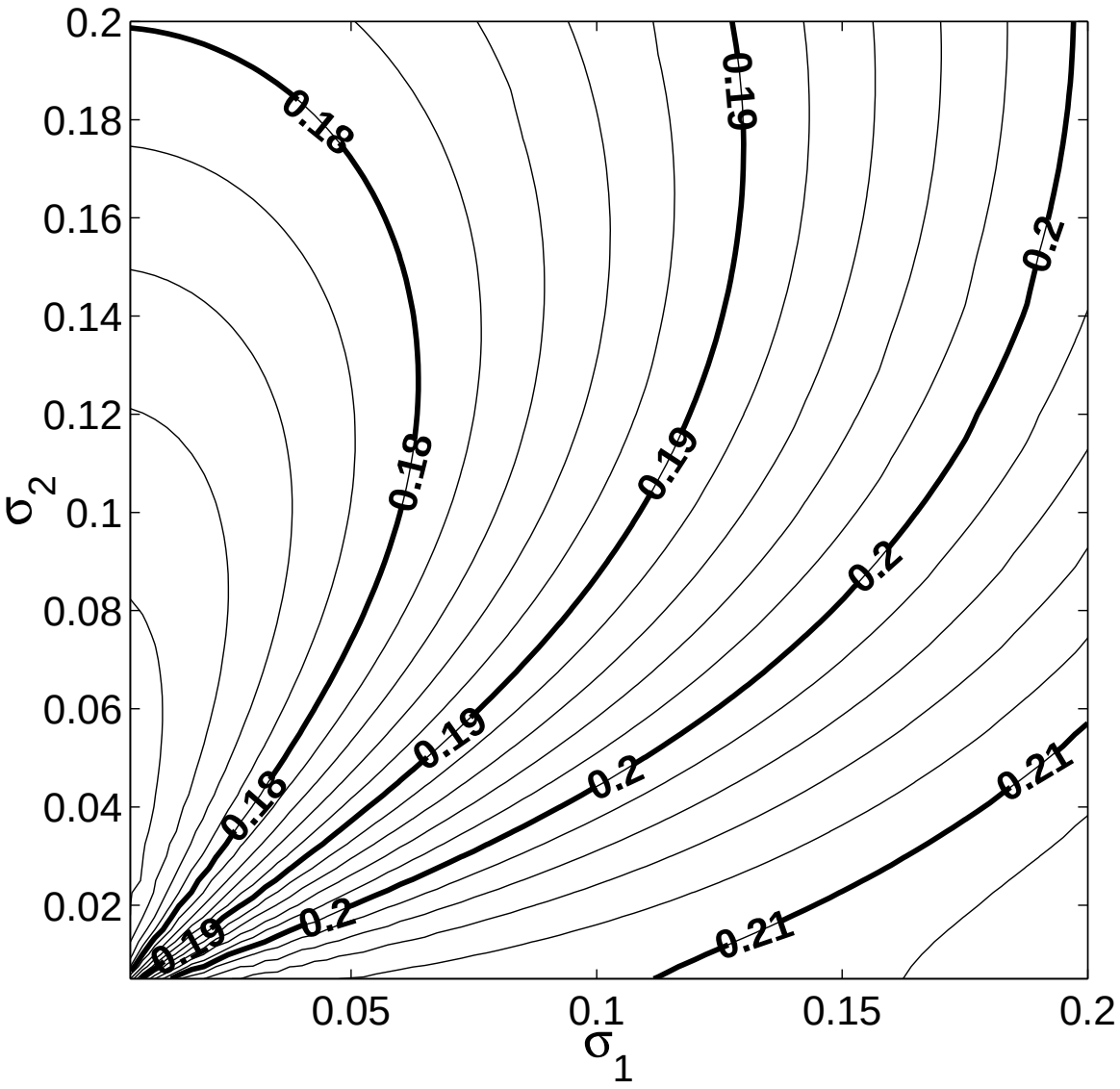

Figure 3: Contour plot of $\mu_{0.5}$ as a function of σ_1 and σ_2.

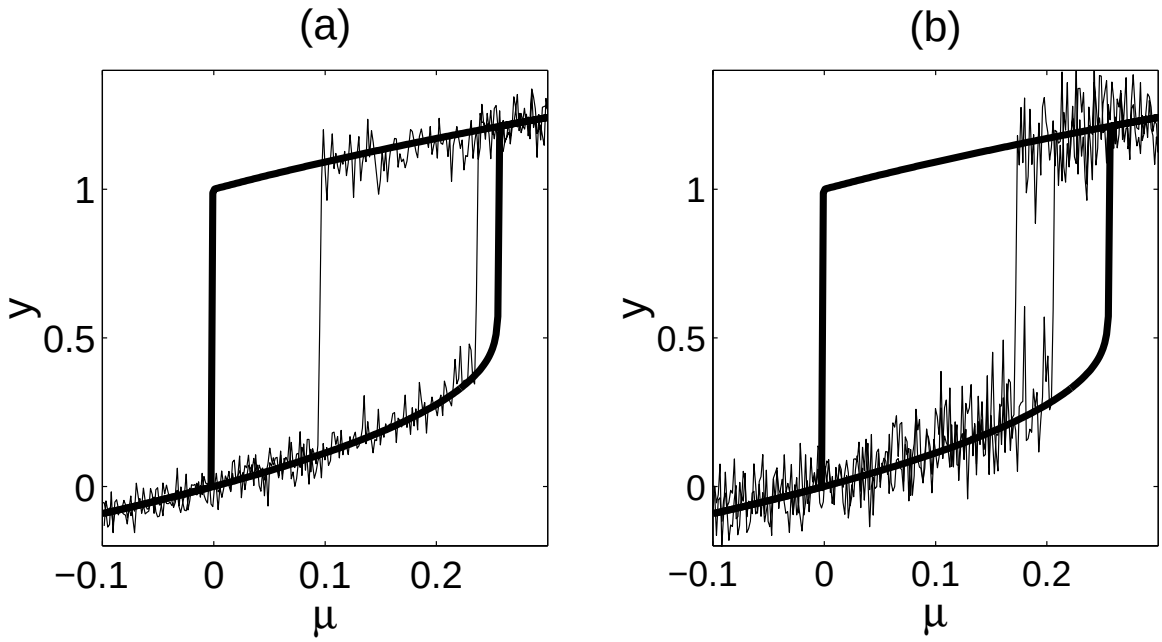

Figure 4: Deterministic (thick lines) and stochastic (thin lines) hysteresis loops for (a) $\sigma_1 = \sigma_2 = 0.05$. (b) $\sigma_1 = \sigma_2 = 0.1$

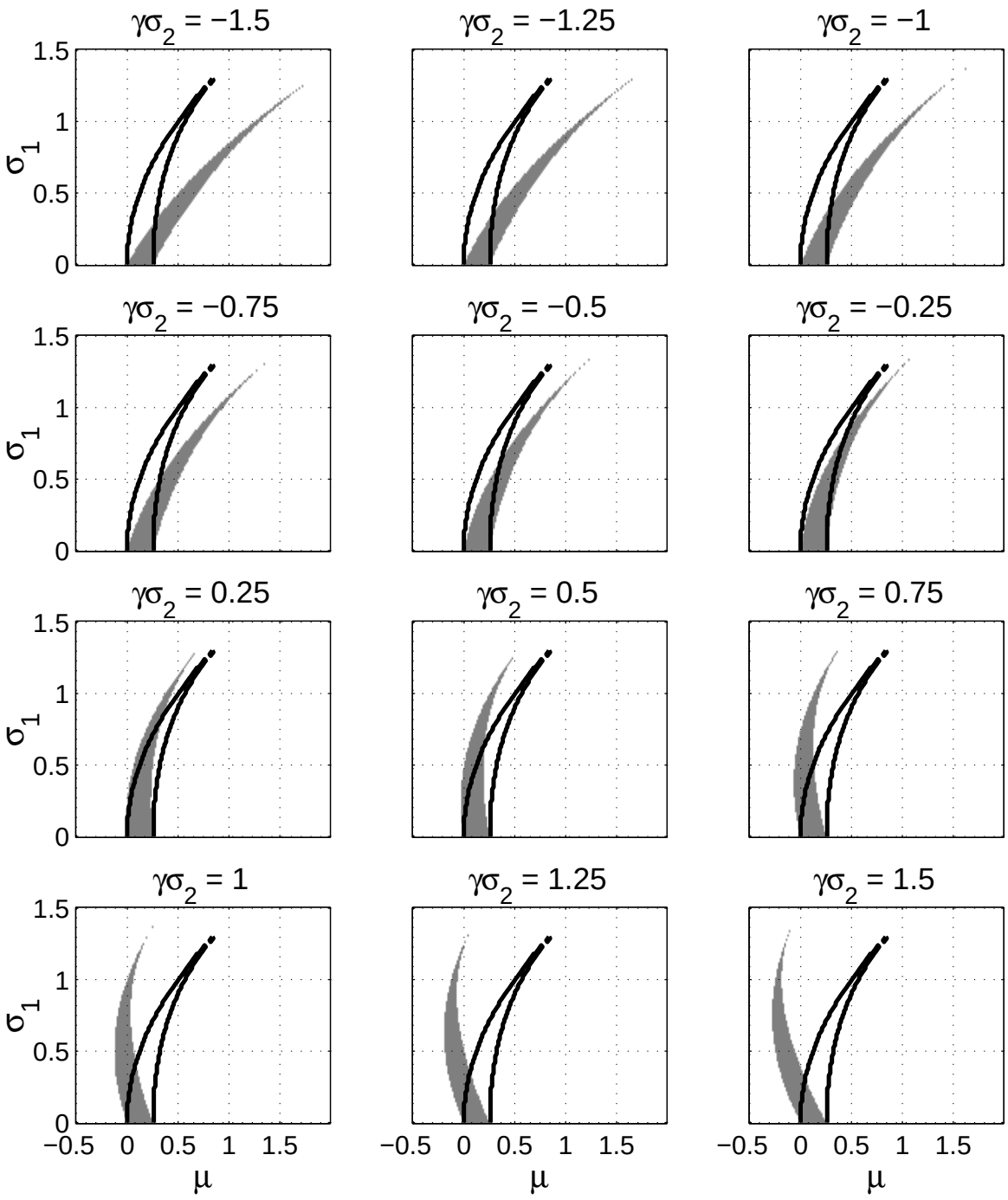

Figure 5: Plot of the region of parameter space in which p^s is bimodal (shaded), as a function of μ, σ_1, and $\gamma\sigma_2$. The solid lines denote the boundary of the region of bimodality for $\gamma = 0$.

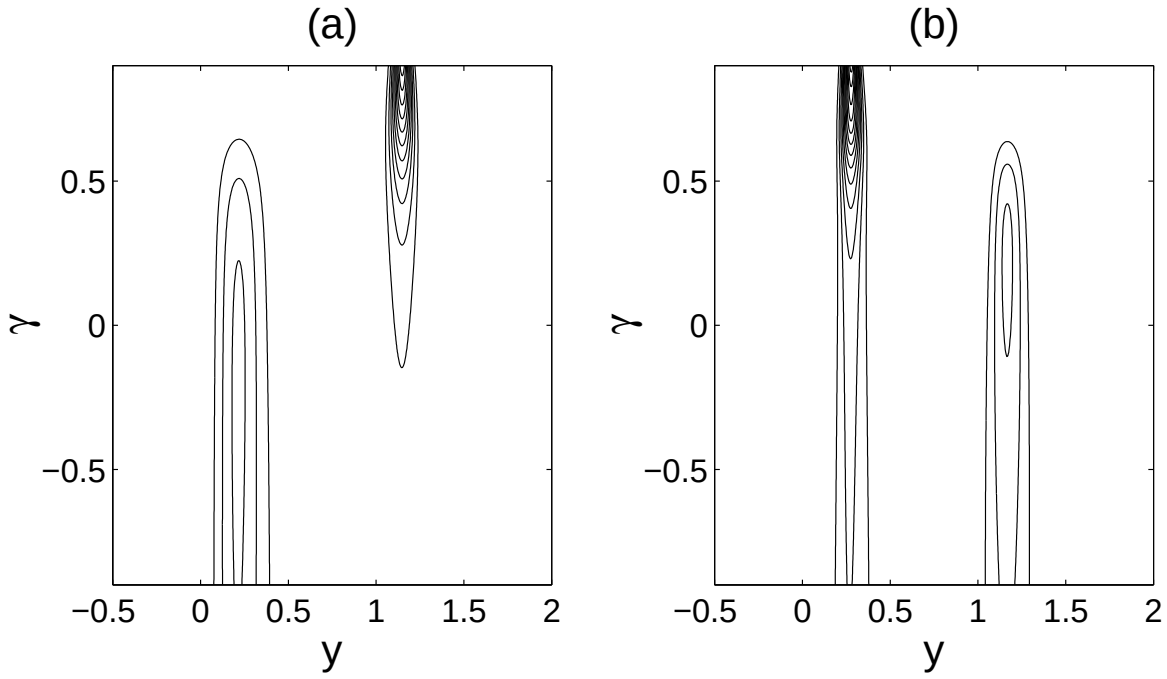

Figure 6: Contour plots of p^s as a function of γ for (a) $(\mu, \sigma_1, \sigma_2) = (0.17, 0.05, 0.1)$ and (b) $(\mu, \sigma_1, \sigma_2) = (0.2, 0.1, 0.05)$.

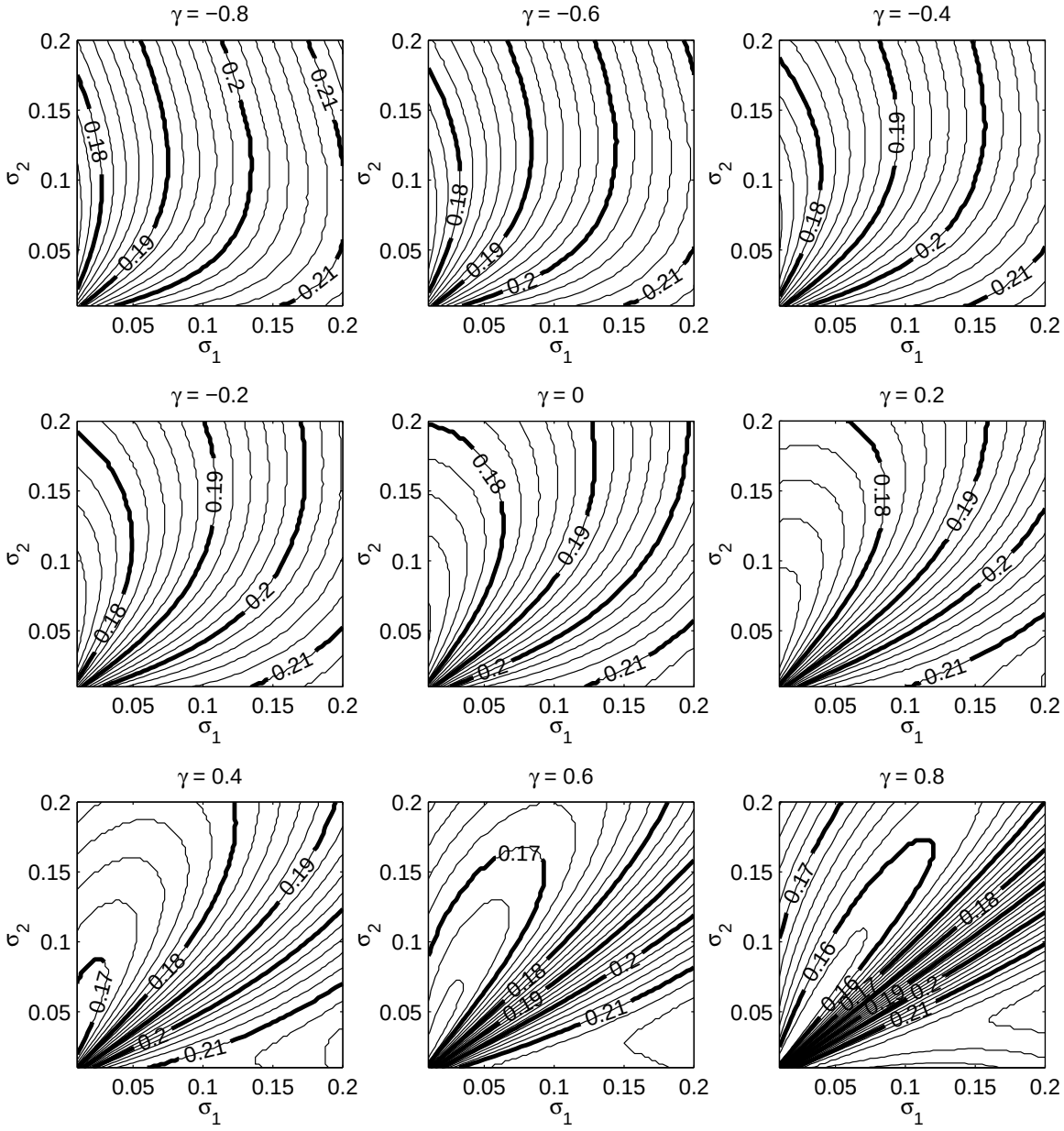

Figure 7: Contour plots of $\mu_{0.5}$ as a function of σ_1, σ_2, and γ.

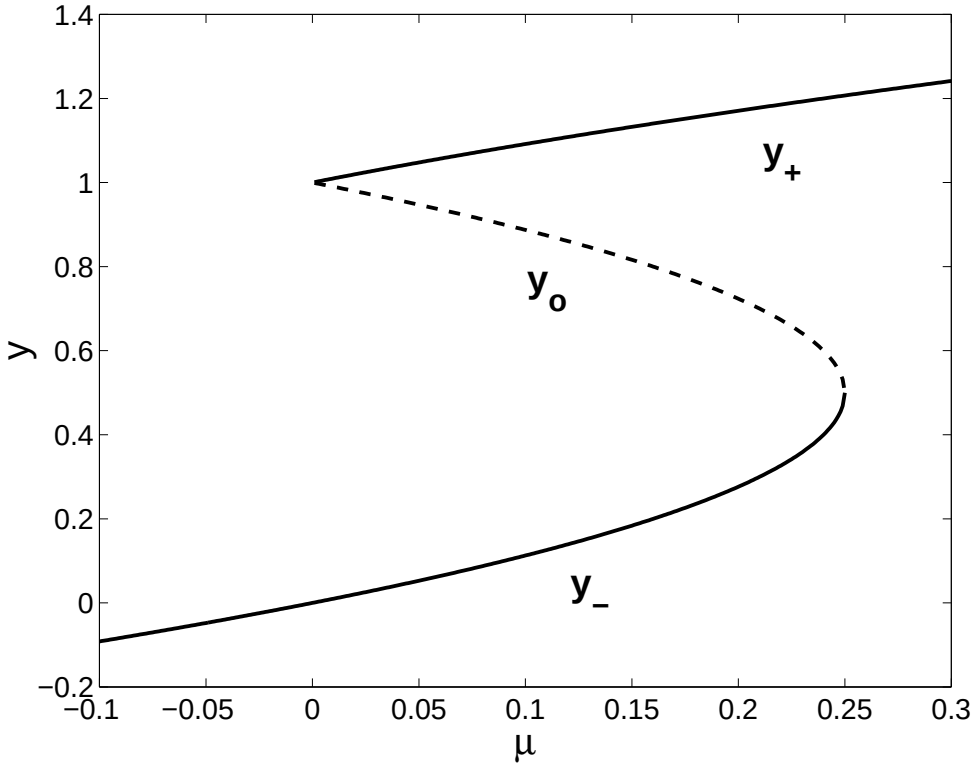

Figure 8: Bifurcation diagram of the y component of the fixed points of the deterministic vector field (31).

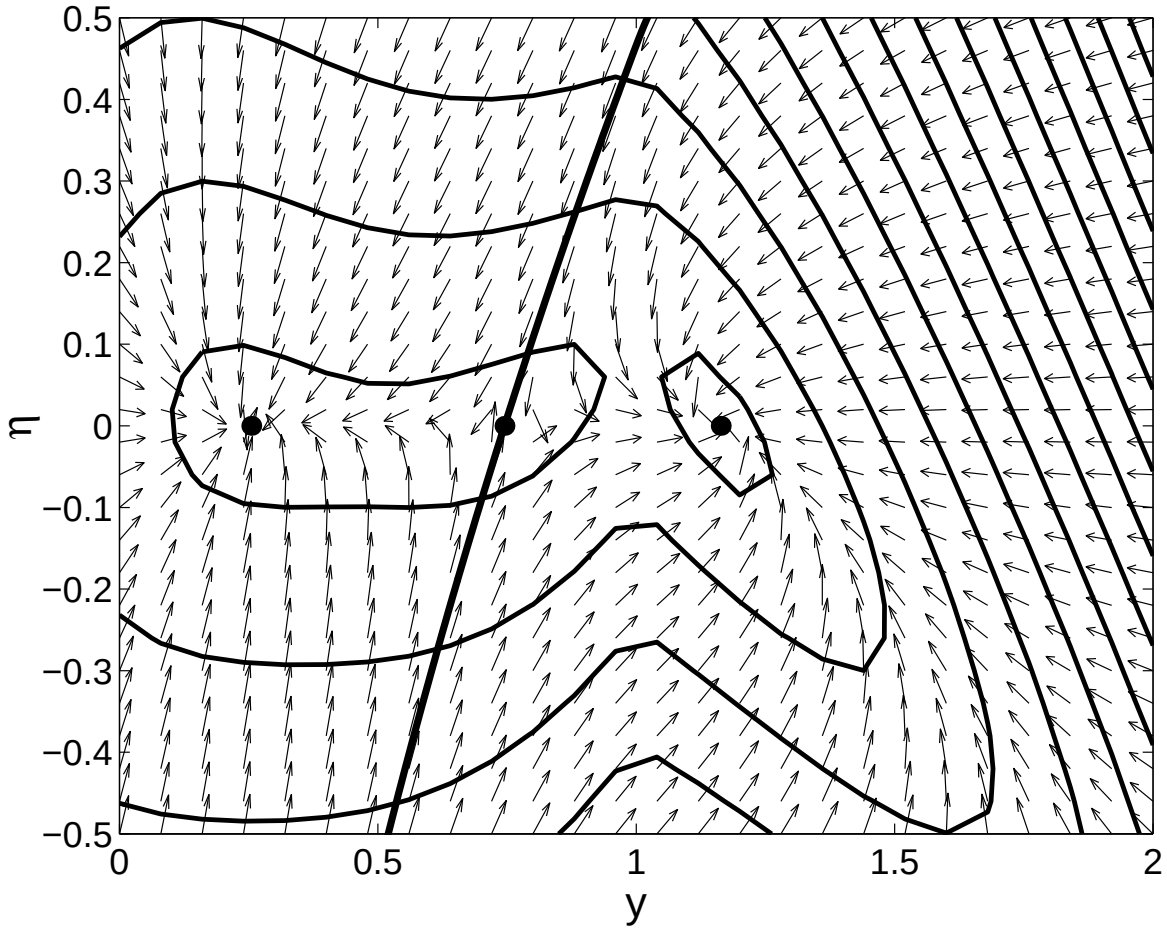

Figure 9: Plot of the vector field $\mathbf{f}$ for $\mu = 0.19$, $\tau = 1$. The direction of $\mathbf{f}$ is indicated by the arrows; the magnitude is contoured. The solid dots indicate the fixed points, and the thick line is the stable manifold of the unstable fixed point. Contours: 0.1,0.3,0.5, ... The lowest contour surrounds the fixed points.

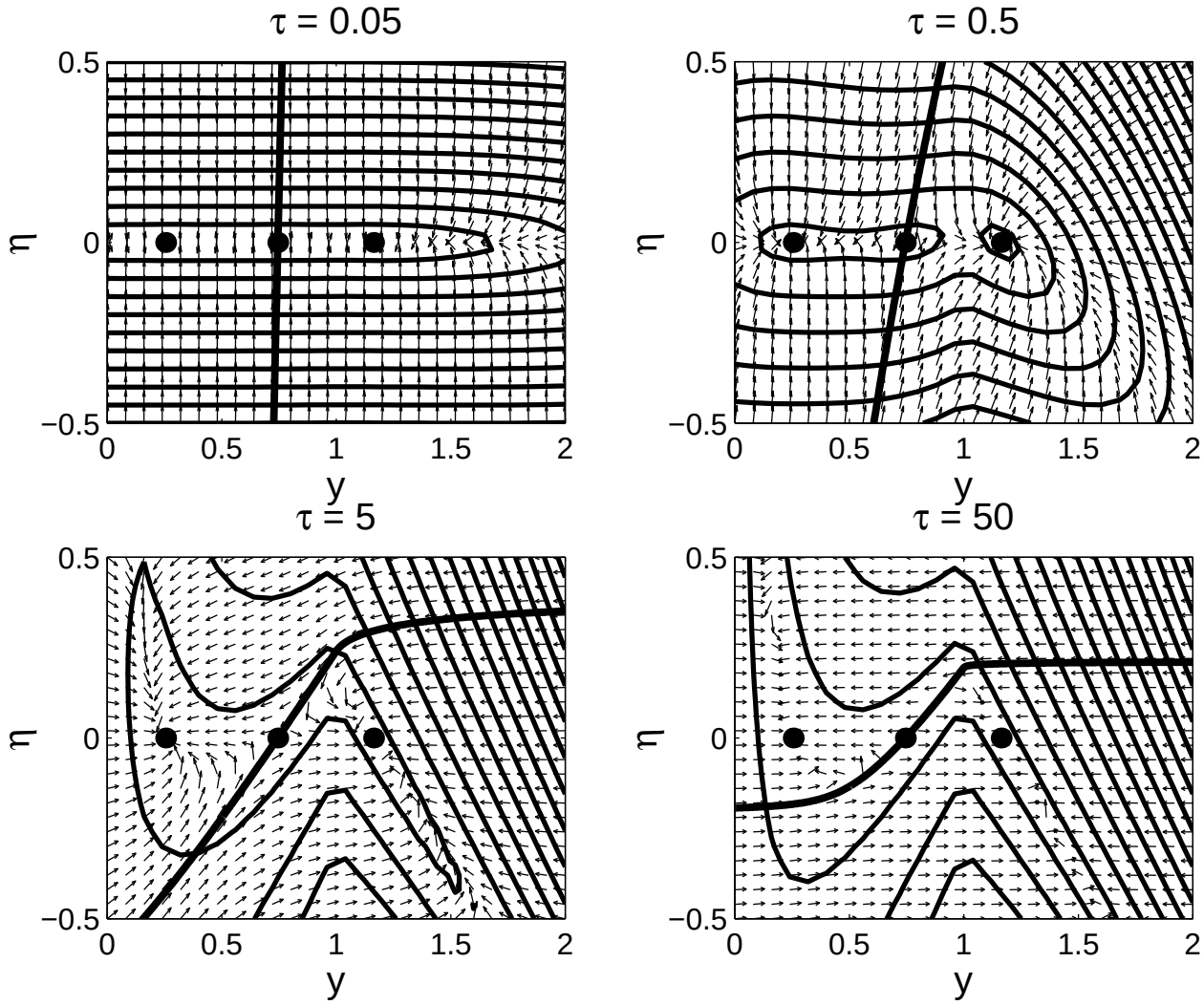

Figure 10: As in Figure 9, for $\tau = 0.05, 0.5, 5, 50$. For $\tau = 0.05$, contours: 1,2,3, ...

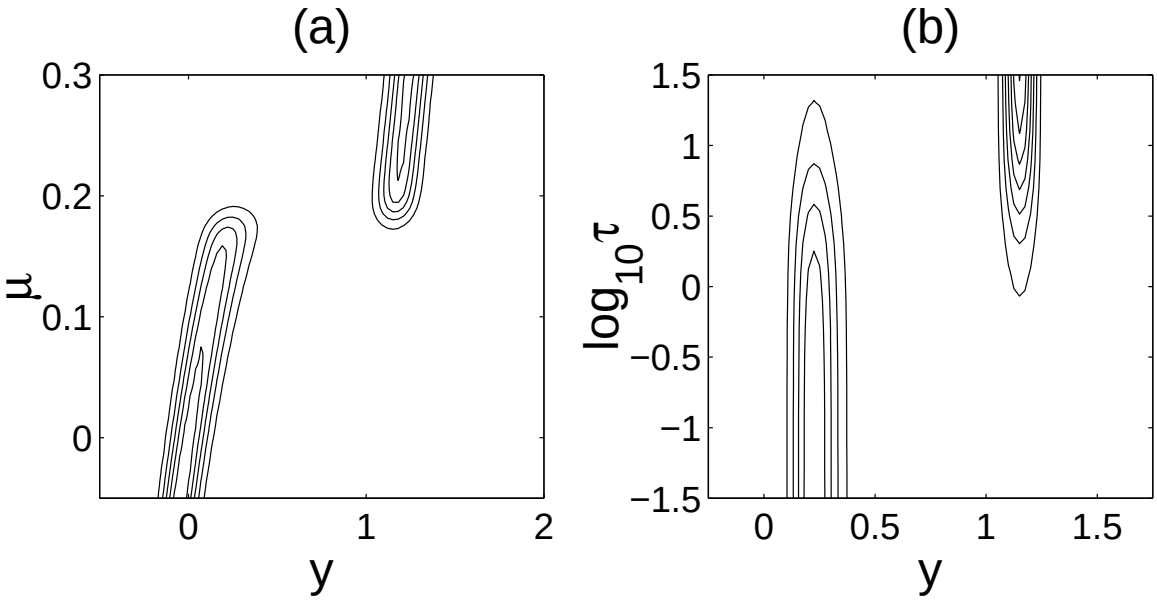

Figure 11: Contour plots of $p^s(y)$ as (a) a function of μ for $\tau = 1$, $\sigma_1 = \sigma_2 = 0.1$. Contours: 1,2,...,6 (b) a function of τ for $\mu = 0.175$, $\sigma_1 = \sigma_2 = 0.075$. Contours: 1,2, ... 8.

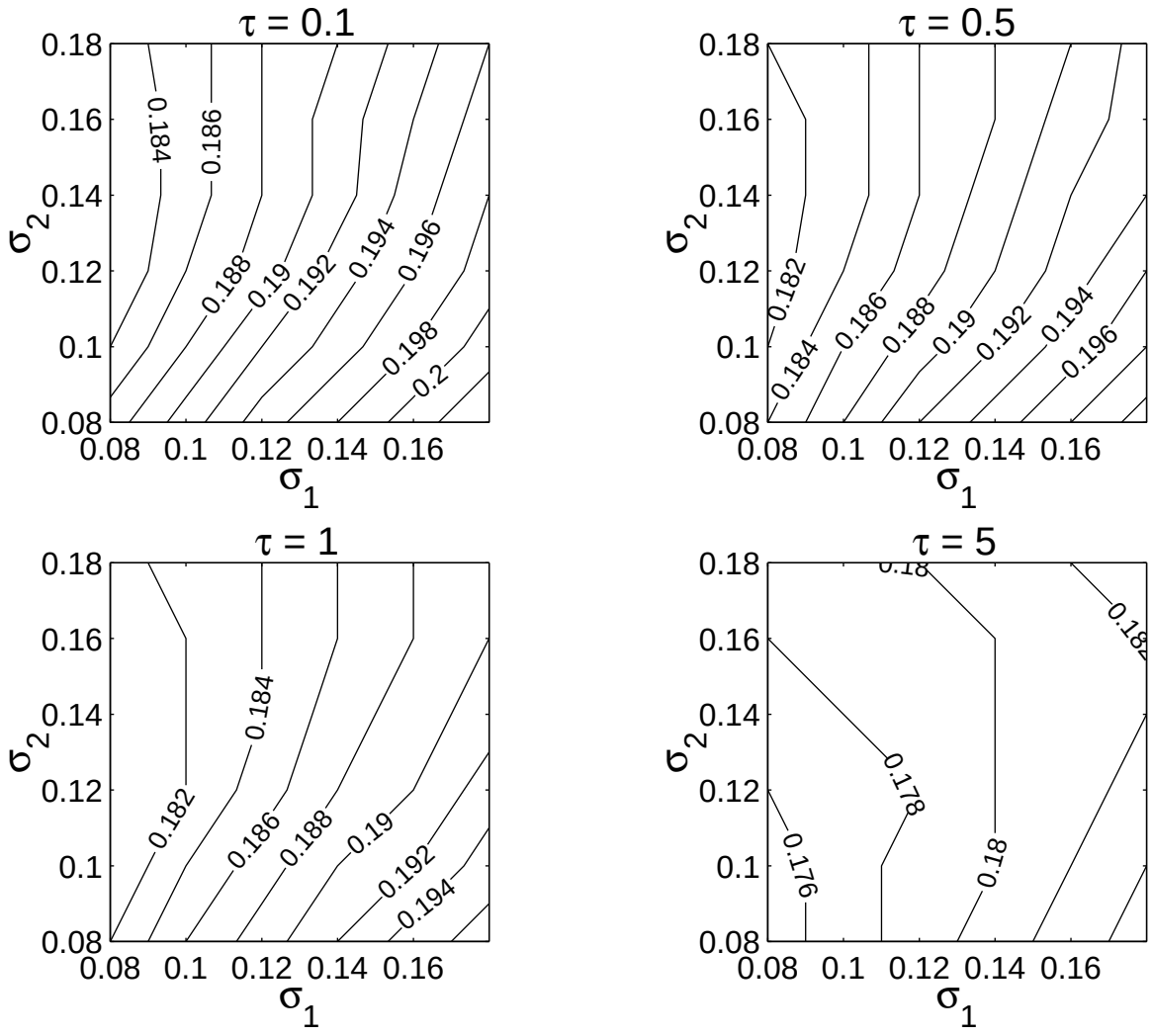

Figure 12: Contour plots of $\mu_{0.5}$ as a function of σ_1, σ_2 for $\tau = 0.1, 0.5, 1, 5$.

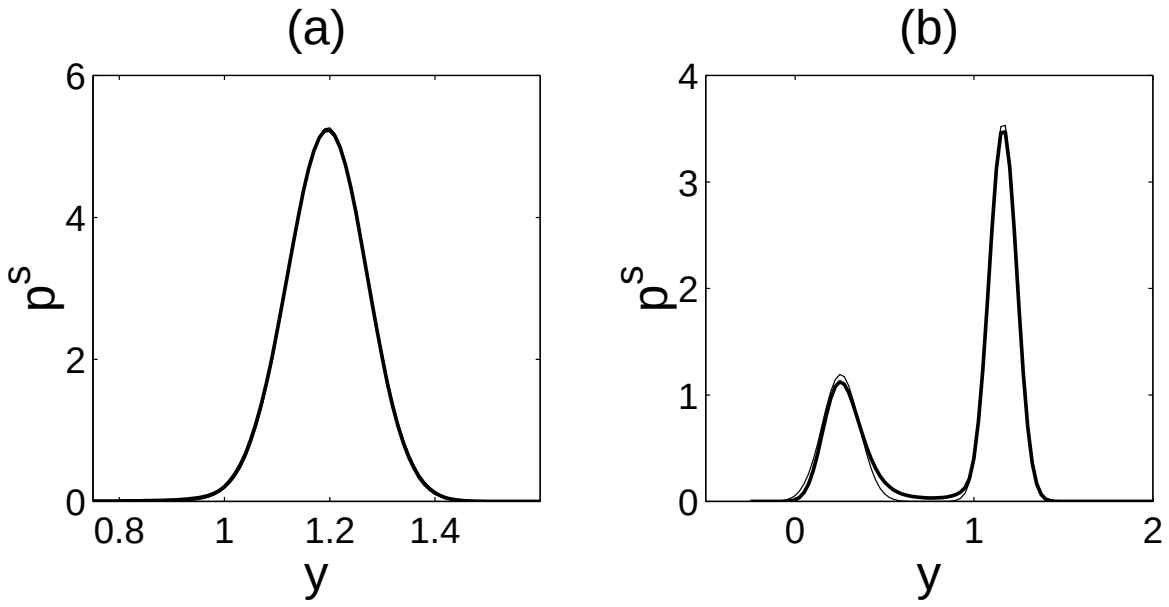

Figure 13: Numerical estimate of $p^s(y)$ (thick line) and linearised approximation (thin line) for $\sigma_1 = \sigma_2 = 0.1$ and (a) $\mu = 0.235$, (b) $\mu = 0.19$.

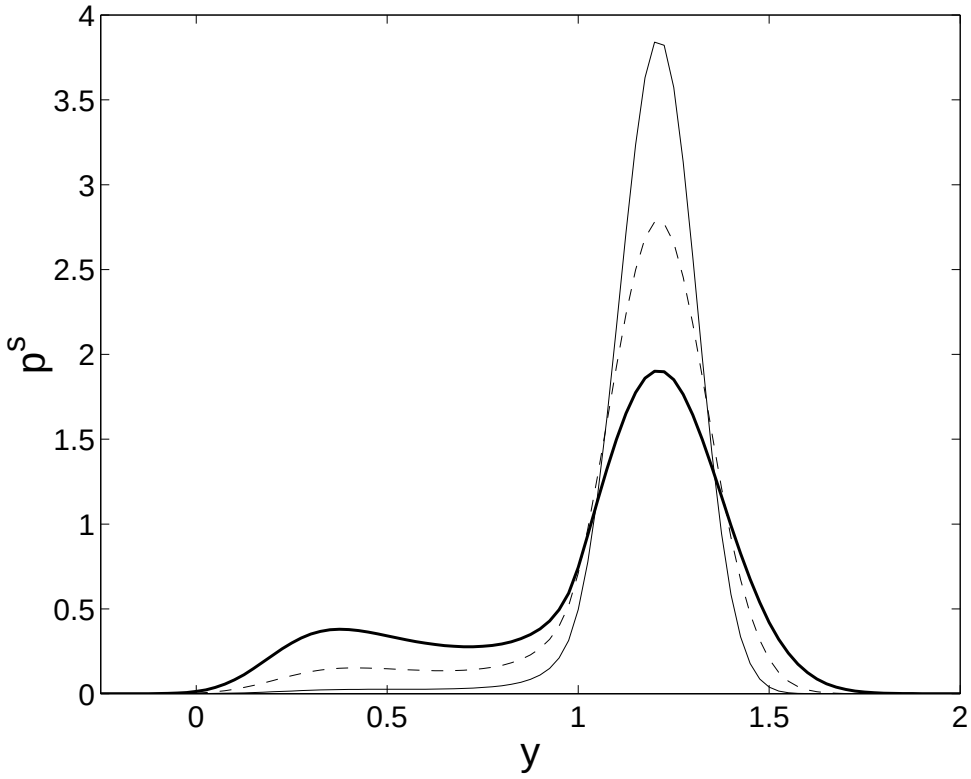

Figure 14: Plots of p^s for $\tau = 1$, $\mu = 0.255$, $\sigma_2 = 0.15$, and $\sigma_1 = 0.1$ (thin line), $\sigma_1 = 0.2$ (dashed line), and $\sigma_3 = 0.3$ (thick line).

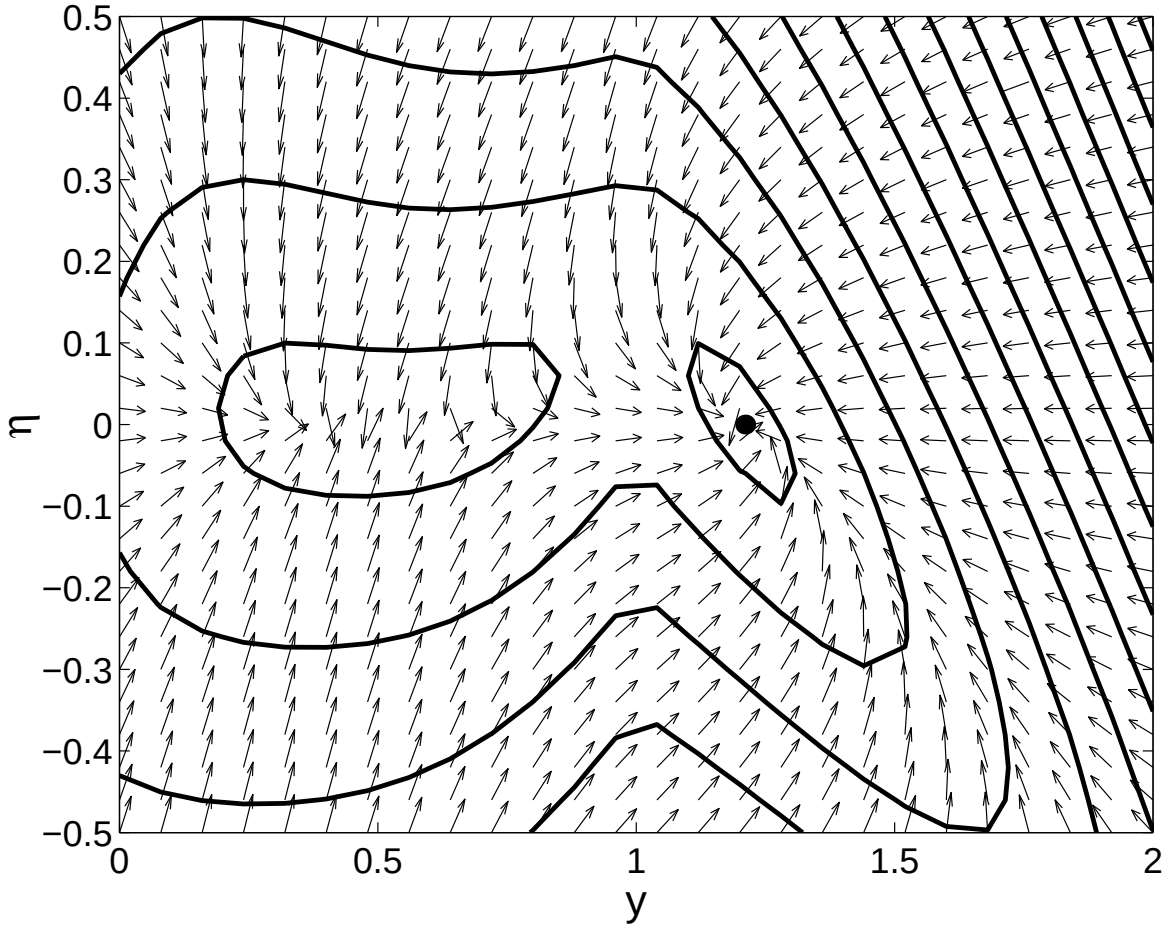

Figure 15: As in Figure 9, for $\tau = 1$ and $\mu = 0.255$.

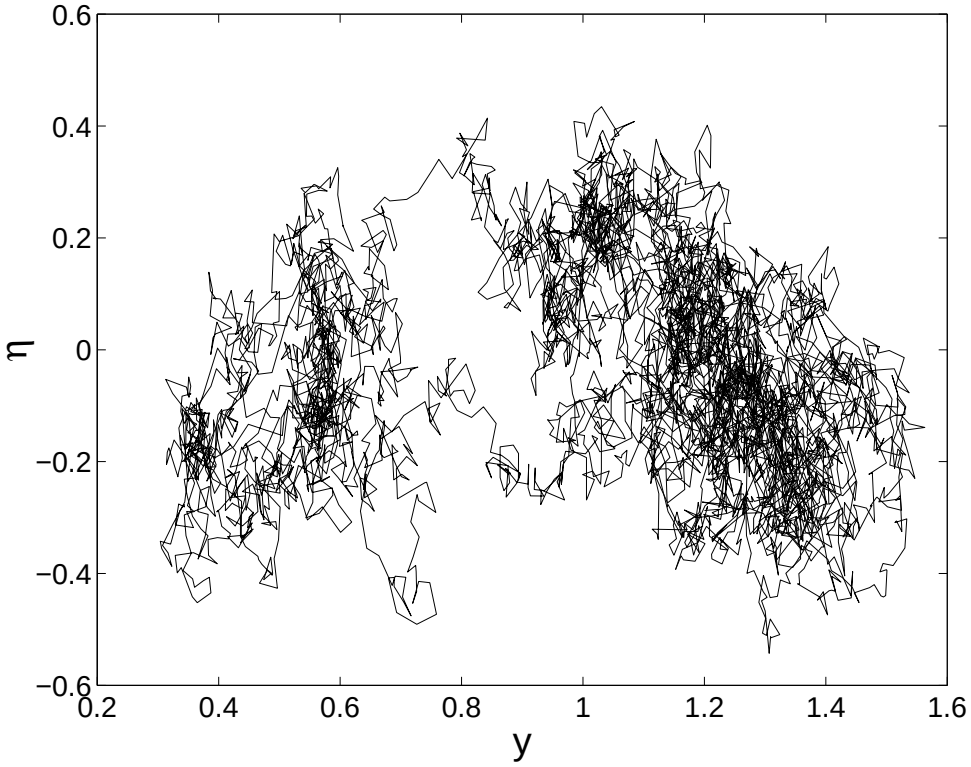

Figure 16: Sample trajectory in the (y, η) space for $\mu = 0.255$, $\tau = 1$,$\sigma_1 = 0.3$, and $\sigma_2 = 0.15$. The duration of the trajectory is 50 time units.