Comments on "Noise-induced transitions in a simplified model of the thermohaline circulation"

Adam Hugh Monahan
Institut für Mathematik, Humboldt-Universität zu Berlin
Unter den Linden 6, 10099 Berlin, Germany
(email: monahan@mathematik.hu-berlin.de)

Axel Timmermann
KNMI, Postbus 201, 3730 AE De Bilt
The Netherlands
(email: timmera@knmi.nl)

Gerrit Lohmann
Geoscience Department, Bremen University
P.O 330 440, 28334 Bremen, Germany
(email: gerrit.lohmann@dkrz.de)

April 2, 2001

Starting from the classical Stommel (1961) two-box model for the North Atlantic meridional overturning circulation, Timmermann and Lohmann (2000, hereafter TL) consider the dynamics of the following system of equations:

$$\frac{d}{dt}y = -|1-y|y + \mu_0 + y\epsilon \tag{1}$$

$$\frac{d}{dt}\epsilon = -\frac{\epsilon}{\tau} + \frac{\sigma}{\tau}\xi, \tag{2}$$

where y represents the scaled salinity difference between the two boxes, μ_0 is the scaled salinity forcing, ξ is a white noise process, and ϵ is an Ornstein-Uhlenbeck (red noise) process with variance $\sigma^2/2\tau$ and autocorrelation e-folding time τ. TL associate ϵ with fluctuations of the temperature difference between the two boxes around the mean value (which is 1 in the scaled variables). The representation of variability in the temperature gradient by a red noise process is motivated by the findings of Lohmann and Schneider (1998) that in the Stommel model temperature differences vary on a much shorter timescale than salinity differences. As $\tau \to 0$, ϵ becomes Gaussian white noise, and the probability density function (PDF) of y satisfies a one-dimensional Fokker-Planck equation (FPE) from which the stationary PDF (that is, the PDF describing the system after the initial transients have died off) may be evaluated analytically. Because ϵ is multiplied by y in equation (1), the $\tau \to 0$ limit must be taken carefully. In technical jargon, equation (1) converges to a Stratonovich stochastic differential equation as $\tau \to 0$; a brief review of stochastic calculus and the Fokker-Planck equation is given in Penland (1996), and a more comprehensive discussion appears in Gardiner (1997). For $\tau \neq 0$, the PDF of y alone is no longer described by an FPE. One can, however, write down a FPE for the joint PDF of y and ϵ. Unfortunately, the stationary version of this FPE is a partial differential equation in two variables that cannot be solved analytically. In this note, we comment on two aspects of TL: first, the derivation and interpretation of equations (1) and (2), and second, the validity of the approximation used in TL to obtain analytic forms for the stationary PDF of y for $\tau \neq 0$.

We first comment on the derivation and interpretation of equations (1) and (2). The original system of equations in TL describes the dynamics of the salinity gradient, ΔS, when the temperature gradient ΔT is described as red noise fluctuations $\Delta T'$ around a mean value ΔT_0:

$$\frac{d}{dt^*}\Delta S = -\frac{c}{V}|\alpha \Delta T_0 + \alpha \Delta T' - \beta \Delta S|\Delta S + \frac{S_0}{h}(P-E) \quad (3)$$

$$\frac{d}{dt^*}\Delta T' = -\frac{1}{\tau^*}\Delta T' + \frac{\Sigma}{\tau^*}\xi, \quad (4)$$

where t^* is the dimensional time, V is the box volume, c is the proportionality constant between the density gradient and the meridional flux, α and β are respectively the thermal and haline expansion coefficients, S_0 is a reference salinity, h is the box depth, $P-E$ is the salinity forcing, τ^* is the temperature fluctuation relaxation timescale, Σ is a noise strength, and ξ is Gaussian white noise. TL argue that equations (1) and (2) follow from (3) and (4) under a suitable rescaling of variables. However, in equation (3), the process $\Delta T'$ appears only inside the absolute value sign, while in equation (1), the process appears *outside* the absolute value sign. In fact, equation (1) does not follow from (3). Furthermore, it can be shown that for (3), the stationary PDF of ΔS in the limit that $\Delta T'$ becomes a white noise process is a delta function at $\Delta S = 0$ (Peter Imkeller, personal communication), which is not physically reasonable.

An alternate interpretation of (1) is as follows. We consider the Stommel (1961) model:

$$\frac{d}{dt^*}\Delta S = \left(-\frac{c}{V}|\alpha \Delta T - \beta \Delta S| + \eta\right)\Delta S + \frac{S_0}{h}(P-E) \quad (5)$$

$$\frac{d}{dt^*}\Delta T = \left(-\frac{c}{V}|\alpha\Delta T - \beta\Delta S| + \eta\right)\Delta T + \frac{F_{oa}}{C_p}, \tag{6}$$

where C_p is the oceanic heat capacity. The freshwater forcing $P - E$ is an unspecified function of ΔT, and the atmosphere-ocean heat flux F_{oa} has the net effect of relaxing ΔT to some value ΔT_0. The quantity η in equations (5) and (6) is a parameterisation of the eddy transport of temperature and salinity between the boxes; this eddy mixing may be associated with transport due to, for example, the wind-driven gyres or quasigeostrophic eddies. The process η is *not* constrained to be positive, so eddy transport between the boxes may be upgradient (see e.g. Nakamura and Chao, 2000). A similar term appears in response to fluctuations in mechanical forcing in the model of the thermally and wind-driven ocean circulation introduced by Maas (1994). In general, η should have a nonzero mean value, so that the eddy transport is on average downgradient. The goal here, however, is to describe a meaningful interpretation of the model analysed in TL, in which η is of mean zero.

As in Cessi (1994), we assume that the timescale on which ΔT is relaxed to ΔT_0 by the thermal forcing F_{oa} is sufficiently small, relative to the timescales of salinity dynamics, that $\Delta T \simeq \Delta T_0$. Further, we model η as a red noise (Ornstein-Uhlenbeck) process with autocorrelation e-folding timescale τ^* and variance $\Sigma^2/2\tau^*$. Then we obtain the system

$$\frac{d}{dt^*}\Delta S = -\frac{c}{V}|\alpha\Delta T_0 - \beta\Delta S|\Delta S + \eta\Delta S + \frac{S_0}{h}(P - E) \tag{7}$$

$$\frac{d}{dt^*}\eta = -\frac{\eta}{\tau^*} + \frac{\Sigma}{\tau^*}\xi(t^*), \tag{8}$$

where $\xi(t)$ is a white-noise process. Defining the nondimensional quantities:

$$t = \frac{c\alpha\Delta T_0}{V}t^* \tag{9}$$

$$y = \frac{\beta}{\alpha\Delta T_0}\Delta S \tag{10}$$

$$\epsilon = \frac{V}{c\alpha\Delta T_0}\eta, \tag{11}$$

equations (7) and (8) reduce to (1) and (2) where

$$\mu = \frac{\beta V S_0}{ch(\alpha\Delta T_0)^2}(P-E) \tag{12}$$

$$\tau = \frac{c\alpha\Delta T_0}{V}\tau^* \tag{13}$$

$$\sigma = \left(\frac{V}{c\alpha\Delta T_0}\right)^{1/2}\Sigma. \tag{14}$$

In the above, we have used the fact that for a white noise process,

$$\xi(t^*) = \left(\frac{t}{t^*}\right)^{1/2}\xi(t) \tag{15}$$

under a rescaling of time, which follows from the fact that the white noise must be delta-correlated in both the dimensional and nondimensional variables.

Thus, a meaningful physical interpretation can be made of equations (1) and (2), but this interpretation differs from that of TL.

We now comment on the validity of the approximation used by TL to obtain an analytic solution of the PDF of y for $\tau \neq 0$. TL employ an approximation due to Jung and Hänggi (1987), known as the Unified Coloured Noise Approximation (UCNA), to reduce the system (1)-(2) to an approximate 1-dimensional system, whose associated stationary Fokker-Planck equation admits and analytic solution. TL calculate the following UCNA expression for the stationary PDF of y:

$$\begin{aligned} p_s(y,\tau) &= N_1 y^{-1}\left|1+\tau\left(\frac{\mu_0}{y}-y\right)\right| e^{\frac{2}{\sigma^2}\left[-\frac{1}{2}\tau\left(-1+y+\frac{\mu_0}{y}\right)^2-\ln y+y-\frac{\mu_0}{y}\right]} \quad ,0<y\leq 1 \\ p_s(y,\tau) &= N_2 y^{-1}\left|1+\tau\left(\frac{\mu_0}{y}+y\right)\right| e^{\frac{2}{\sigma^2}\left[-\frac{1}{2}\tau\left(1-y+\frac{\mu_0}{y}\right)^2+\ln y-y-\frac{\mu_0}{y}\right]} \quad ,y>1 \end{aligned} \tag{16}$$

where N_1, N_2 are appropriate normalisations. Figures 6 and 7 of TL plot p_s as a function of y for different values of τ and σ. For small values of τ, the results resemble those of the white noise limit. For increasing τ, p_s given by (16) displays qualitatively different behaviour. In particular, nodes and new extrema appear in the PDF. These results are interpreted as "noise-induced transitions", and the analogy of quantum-mechanical tunneling is used to describe the passage of the system across the node of the PDF.

Another approach to determine an approximation of the PDF associated with (1)-(2) is to integrate the equations numerically; the simplest algorithm is a forward Euler discretisation (Kloeden and Platen, 1992). Denoting the discrete time step by δ so that

$$t_k = k\delta \tag{17}$$

the forward Euler discretisation of (1)-(2) is

$$y_{t_k} = y_{t_{k-1}} + \delta(-|1 - y_{t_{k-1}}|y_{t_{k-1}} + \mu_0 + y_{t_{k-1}}\epsilon_{t_{k-1}}) \tag{18}$$

$$\epsilon_{t_k} = \epsilon_{t_{k-1}} - \delta\frac{\epsilon_{t_{k-1}}}{\tau} + \sqrt{\delta}\frac{\sigma}{\tau}W_{t_{k-1}} \tag{19}$$

where W_{t_k} is a sequence of zero-mean, unit variance Gaussian random variables. Equations (18) and (19) were integrated for 150000 time units with a timestep $\delta = 0.05$ for the parameter values τ, σ used in Figures 6 and 7 of TL. Figures 1 and 2 display Gaussian kernel density estimates of the PDFs obtained from the simulation, along with the UCNA approximations (16). The numerical results are robust to reduction of the stepsize δ, and inspection of the time series indicates that the record is long enough for estimation of the PDF from the time series to be appropriate.

Comparing the numerical and UCNA results in Figures 1 and 2, it is clear that for small values of τ, the stationary PDFs produced by numerical integration and by the UCNA are in close agreement. However, for $\tau \sim O(1)$, there are marked differences. In particular, the

PDFs produced by numerical integration do not display any nodes or new extrema. Instead, the result of raising τ for fixed σ is seen to be a shift of the PDF toward the right-hand peak. The differences between the numerical and UCNA approximations occur because of a breakdown of the validity of the UCNA for τ of $O(1)$.

By construction, the UCNA assumes a timescale separation between the processes y and ϵ; it is only valid when ϵ varies much more rapidly than y. This implies that the UCNA is a small τ approximation, but how small is "small"? TL note that the domain of validity of the UCNA is given by the following pair of inequalities:

$$\kappa(y,\tau) = \frac{1}{\sqrt{\tau}} - \sqrt{\tau}\left(y - \frac{\mu_0}{y}\right) >> \bar{t}^{-1} \quad ,0 < y \leq 1 \tag{20}$$

$$\kappa(y,\tau) = \frac{1}{\sqrt{\tau}} - \sqrt{\tau}\left(-y - \frac{\mu_0}{y}\right) >> \bar{t}^{-1} \quad ,y > 1, \tag{21}$$

where $\bar{t}$ is a "characteristic timescale" of variability of y. Defining a "typical" value of y, TL argue for the global (in y) satisfaction of these conditions. In fact, their validity must be considered locally in y. In particular, the UCNA certainly fails wherever κ vanishes. For $0 < y \leq 1$, this occurs for

$$y = \frac{1}{2\tau} + \sqrt{\left(\frac{1}{2\tau}\right)^2 + \mu_0} \tag{22}$$

(a second root is discarded because it occurs for $y < 0$, outside the domain of consideration). For $\tau < \frac{1}{1-\mu_0}$, this root also falls outside of $[0,1]$. However, for $\tau \geq \frac{1}{1-\mu_0}$, κ vanishes for $y \in [0,1]$, and the UCNA fails within the domain of interest. In particular, it is clear from (16) that κ vanishes at precisely the values y where p_s has a node. Thus, the emergence of zeros in the PDF of y is an artifact of the breakdown of the UCNA. In fact, the UCNA fails not just at the points at the zeros of κ, but in a surrounding neighbourhood, as is demonstrated by the differences between the numerical and UCNA PDFs for $\tau = 0.8 < \frac{1}{1-\mu_0}$. For $y > 1$,

κ never vanishes, but because the overall amplitude of the PDF is a function of its global structure, at best only the shape of the PDF for $y > 1$ will agree with that produced by the UCNA. Thus, for τ of $O(1)$, the UCNA breaks down locally in y, with global consequences for the structure of the PDF.

Calculating the stationary PDF of a system in a one-dimensional potential subject to coloured noise remains an unsolved problem in physics. A number of different approximations have been proposed, but they are valid only in the limit of small or of very large τ (Horsthemke and Lefever, 1984; Hänggi and Jung, 1995). To obtain the stationary PDFs of y in the case where its timescale is of the same order of magnitude as ϵ, at present we must take recourse to numerical methods. We note that an essential conclusion of TL is unchanged, namely that increasing σ populates the left-hand peak of the stationary PDF of y at the expense of the right-hand peak, while increasing τ has the opposite effect.

References

Cessi, P., 1994: A simple box model of stochastically forced thermohaline flow. *J. Phys. Oceanogr*, **24**, 1911-1920.

Gardiner, C.W., 1997: *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. Springer, 442 pp.

Hänggi, P. and P. Jung, 1995: Colored noise in dynamical systems. *Adv. Chem. Phys.*, **89**, 239-326.

Horsthemke, W., and R. Lefever, 1984: *Noise-Induced Transitions: Theory and Applications in Physics, Chemistry, and Biology*. Springer, 318 pp.

Jung, P. and P. Hänggi, 1987: Dynamical systems: A unified coloured-noise approximation. *Phys. Rev.*, **A35**, 4464-4466.

Kloeden, P.E., and E. Platen, 1992: *Numerical Solution of Stochastic Differential Equations.* Springer, 632 pp.

Lohmann , G., and J.Schneider, 1998: Dynamics and predictability of Stommel's box model: A phase space perspective with implications for decadal climate variability. *Tellus*, **51**A, 326-336.

Maas, L.R.M., 1994: A simple model for the three-dimensional, thermally and wind-driven ocean circulation. *Tellus*, **46A**, 671-680.

Nakamura, M., and Y. Chao, 2000: On the eddy isopycnal thickness diffusivity of the Gent-McWilliams subgrid mixing parameterisation. *J. Climate*, **13**, 502-510.

Penland, C., 1996: A stochastic model of IndoPacific sea surface temperature anomalies. *Physica D*, **98**, 534-558.

Stommel, H.M., 1961: Thermohaline convection with two stable regimes of flow. *Tellus*, **13**, 224-230.

Timmermann, A. and G. Lohmann, 2000: Noise-induced transitions in a simplified model of the thermohaline circulation. *J. Phys. Oceanogr.*, **30**, 1891-1900.

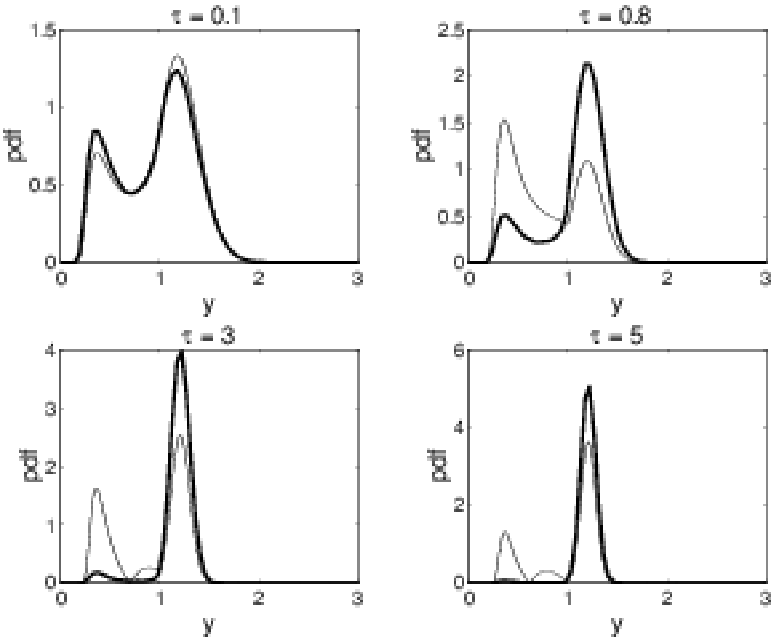

Figure 1: Stationary PDF of y from numerical integration (thick line) and from the UCNA approximation (thin line), for $\sigma = 0.3$.

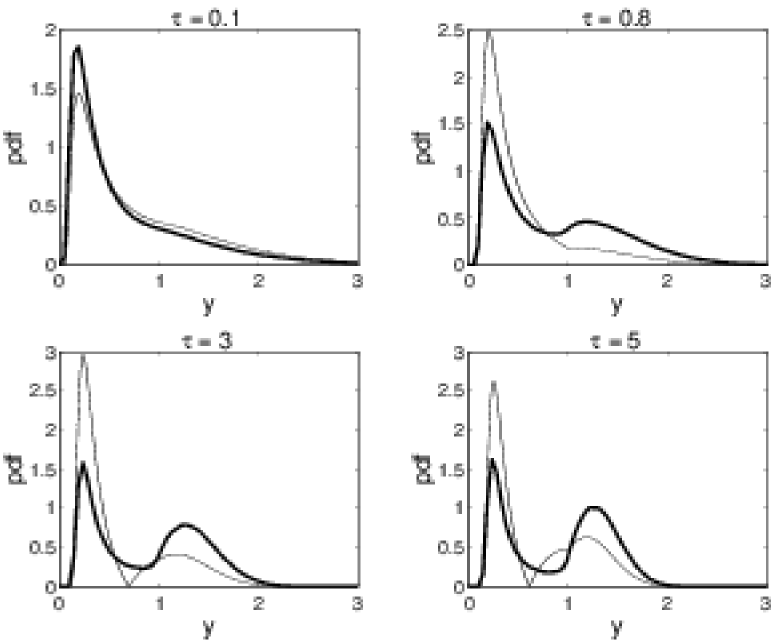

Figure 2: As in Figure 1, for $\sigma = 1$.