

Stochastic Parameterization of Subgrid-Scale Processes: A Review of Recent Physically Based Approaches

Jonathan Demaeyer and Stéphane Vannitsem

Abstract We review some recent methods of subgrid-scale parameterization used in the context of climate modeling. These methods are developed to take into account (subgrid) processes playing an important role in the correct representation of the atmospheric and climate variability. We illustrate these methods on a simple stochastic triad system relevant for the atmospheric and climate dynamics, and we show in particular that the stability properties of the underlying dynamics of the subgrid processes have a considerable impact on their performances.

Keywords Stochastic parameterization • Response theory • Multiscale system • Homogenization • Subgrid processes

1 Introduction

From a global point of view, the Earth system is composed of a myriad of different interacting components. These components can be regrouped in compartments like the atmosphere, the hydrosphere, the lithosphere, the biosphere, and the cryosphere (Olbers, 2001).¹ Those compartments play a role on different timescales from seconds to ice ages. In this perspective, the resulting Earth's climate is a “concert” at which each compartment seems to play its own partition with its own tempo. Their respective contribution to the total variability of an observable, say, e.g., the global temperature, is, however, the outcome of complex interactions between the different components, leading to an emergent dynamics far from the one that could be generated by a linear additive superposition principle (Nicolis and Nicolis, 1981, 2012).

¹More recently, a new compartment has appeared, whose effect is not negligible at all and which is not predictable nor descriptive by evolution equations, namely the impact of the human activities.

J. Demaeyer (✉) • S. Vannitsem
Institut Royal Météorologique de Belgique, Avenue Circulaire, 3, 1180 Brussels, Belgium
e-mail: Jonathan.Demaeyer@meteo.be; Stephane.Vannitsem@meteo.be

A paradigmatic example is provided in the work of Hasselmann, detailed in his seminal paper of 1976, which states precisely that the slowly evolving components of the climate system, besides their own dynamics due their own physical processes, also integrate the impact of the faster components (Hasselmann, 1976). Hasselmann describes this process using the analogy of the Brownian motion where a macro-particle in a liquid integrates the effect of the collisions with the fluid's micro-particles, leading to the erratic trajectory of the former. The interest of this framework is that it provides a natural description of the “red noise” spectral density observed in most climatic records and observations (Ghil et al., 2002; Lovejoy and Schertzer, 2013). Subsequently, during the following decade, stochastic modeling for meteorology and climatology became an important research topic (Frankignoul, 1979; Frankignoul and Hasselmann, 1977; Frankignoul and Müller, 1979; Lemke, 1977; Lemke et al., 1980; Nicolis, 1981, 1982; Nicolis and Nicolis, 1981; Penland, 1989) before falling into disuse in what has been described as a “lull” of work in this field (Arnold et al., 2003). However, during that period, the ideas that correct parameterizations of subgrid processes are important to improve climate and weather models gained popularity (Newman et al., 1997; Penland, 1996; Penland and Matrosova, 1994). Stochastic parameterizations for the “turbulent” closure in 2-D large-eddy simulations on the sphere have also been considered (Frederiksen, 1999; Frederiksen and Davies, 1997). It led recently to the implementation of stochastic schemes to correct the model errors (Nicolis, 2003, 2004) made in large numerical weather prediction (NWP) models (Buizza et al., 1999; Shutts, 2005), improving the reliability of probabilistic forecasts and correcting partially their variability (Doblas-Reyes et al., 2009; Nicolis, 2005). The relation between multiplicative noise and the non-Gaussian character observed in some geophysical variables has also been considered (Sardeshmukh and Penland, 2015; Sura et al., 2005), as well as stochastic models for the climate extremes (Sura, 2013).

Since the beginning of the twenty-first century, a revival of the interest in stochastic parameterization methods have occurred, due to the availability of new mathematical methods to perform the stochastic reduction of ODEs systems. Almost simultaneously, a rigorous mathematical framework for the Hasselmann “program” was devised (Arnold, 2001; Kifer, 2001, 2003) and a new method based on the singular perturbation theory of Markov processes (Majda et al., 2001) was proposed. The latter approach is currently known as the Majda–Timofeyev–Vanden-Eijnden (MTV) method. Both methods have been tested and implemented successfully in geophysical models (Arnold et al., 2003; Culina et al., 2011; Franzke et al., 2005; Vannitsem, 2014). The revival of the Hasselmann program has also stressed the need to consider the occurrence of very rare events triggered by the noise that allow for the solutions of the system to jump from one local attractor to another one (Arnold, 2001). Such events display recurrence timescales that are few orders greater than the timescale of the climate variables considered, and thus induce transitions between different climatic states. The statistics of these transitions is then given by the so-called *Large Deviations* theory (Freidlin and Wentzell, 1984) [for recent developments on this matter, see Bouchet et al. (2016)]. In addition to these two methods, the modeling of the effects of subgrid scale through conditional

Markov chain has been considered (Crommelin and Vanden-Eijnden, 2008) and recently, new stochastic parameterization techniques have been proposed, based on an expansion of the backward Kolmogorov equation (Abramov, 2015) and on the Ruelle response theory (Wouters and Lucarini, 2012). The latter has been tested on a simple coupled ocean-atmosphere model (Demaeayer and Vannitsem, 2016), on stochastic triads (Wouters et al., 2016), and on an adapted version of the Lorenz'96 model (Vissio and Lucarini, 2016).

This renewal of interest for stochastic modeling and reduction methods illustrates how fruitful was the original idea of Hasselmann. However, in view of the availability of several possible approaches, one might wonder about their efficiency in different situations. Indeed, depending on the specific purpose that it needs to fulfill, some parameterizations might perform better than others. The present review aims to shed some light on these questions and to illustrate some of the aforementioned parameterization methods on a simple model for which most of the calculations can be made analytically.

In Sect. 2, we will present the general framework in which the problem of stochastic parameterizations is posed. In Sect. 3, we present the different parameterizations that we shall consider for the analysis model. The stochastic triad model used here and the comparison are presented in Sect. 4. Finally, the conclusions are given in Sect. 5.

2 The Parameterization Problem

Consider the following system of ordinary differential equations (ODEs):

$$\dot{Z} = T(Z) \tag{1}$$

where $Z \in \mathcal{R}^d$ is a set of variables relevant for the problem under interest for which the tendencies $T(Z)$ are known. And suppose that one wants to separate this set of variables into two different subset $Z = (X, Y)$, with $X \in \mathcal{R}^m$ and $Y \in \mathcal{R}^n$. In general, such a decomposition is made such that the subset X and Y have strongly differing *response times* $\tau_Y \ll \tau_X$ (Arnold et al., 2003), but we will assume here that this constraint is not necessarily met. System (1) can then be expressed as:

$$\begin{cases} \dot{X} = F(X, Y) \\ \dot{Y} = H(X, Y) \end{cases} \tag{2}$$

The timescale of the X sub-system is typically (but not always) longer than the one of the Y sub-system, and it is often materialized by a parameter $\delta = \tau_Y/\tau_X \ll 1$ in front of the time derivative \dot{Y} . The X and the Y variables represent, respectively, the resolved and the unresolved sub-systems. The resolved sub-system is the part of the full system that we would like to simulate, i.e., generate explicitly and numerically

its time-evolution. The general problem of model reduction consists thus to approximate the resolved component X as accurately as possible by obtaining a closed equation for the system X alone (Arnold et al., 2003). The term “accurately” here can have several meanings, depending on the kind of problem to solve. For instance, we can ask that the closed system for X has statistics that are very close to the ones of the X component of system (2). We can also ask that the closed system trajectories remain as close as possible to the trajectories of the full system for long times.

In general a parameterization of the sub-system Y is a relation \mathcal{E} between the two sub-systems:

$$Y = \mathcal{E}(X, t) \quad (3)$$

which allows to effectively close the equations for the sub-system X while retaining the effect of the coupling to the Y sub-system.

The problem of the model reduction is not new, and was considered first in celestial mechanics. Famous mathematicians have considered it and contributed to what is known nowadays as the theory of averaging (Sanders and Verhulst, 1985) and which forms the first step of the Hasselmann program (Arnold, 2001). The mathematical framework was set in the 1960s by the influential contribution of Bogoliubov and Mitropolski (1961). However, this averaging technique is a deterministic method which does not take into account the deviations from the average. The proposition of Hasselmann was thus to take into account these deviations by considering stochastic parameterization where the relation (3) can be considered in a statistical sense. In that framework, the Y sub-system and its effect on the sub-system X can be considered as a stochastic process, which possibly depends upon the state of the X sub-system. Different methods to achieve this program are now discussed in Sect. 3.

3 The Parameterization Methods

Let us now write system (2) as:

$$\begin{cases} \dot{X} = F_X(X) + \Psi_X(X, Y) \\ \dot{Y} = F_Y(Y) + \Psi_Y(X, Y) \end{cases} \quad (4)$$

where the coupling and the intrinsic dynamics are explicitly specified. In the present work, we shall focus on parameterizations that are defined in terms of stochastic processes. We will consider methods based

- on the Ruelle response theory (Demaeyer and Vannitsem, 2016; Wouters and Lucarini, 2012; Wouters et al., 2016).
- on the singular perturbation theory of Markov processes (Franzke et al., 2005; Majda et al., 2001).

- on the Hasselmann averaging methods (Arnold et al., 2003; Culina et al., 2011; Kifer, 2003; Vannitsem, 2014).
- on empirical methods (Arnold et al., 2013).

All these parameterizations can be written in the following form:

$$\dot{X} = F_X(X) + G(X, t) + \sigma(X) \cdot \tilde{\xi}(t) \quad (5)$$

where the matrix σ , the deterministic function G , and the random processes $\tilde{\xi}(t)$ have to be determined. The mathematical definition of these quantities obtained through averaging procedure and the measure being used to perform the averaging are usually both differing between the methods. These different choices are rooted in their different underlying hypothesis, as it will be discussed below. Specifically, the response theory method uses the measure of the uncoupled unresolved sub-system $\dot{Y} = F_Y(Y)$, the singular perturbation method uses the measure of the perturbation, and the averaging methods use the measure of the full unresolved sub-system $\dot{Y} = H(X, Y)$ with X considered as “frozen” (constant). Finally, the empirical methods use in general the output of the full unresolved Y sub-system, conditional or unconditional on the state X .

In the rest of the section, we shall describe more precisely each of the above methods.

3.1 The Method Based on Response Theory

This method is based on the Ruelle response theory (Ruelle, 1997, 2009) and was proposed by Wouters and Lucarini (2012, 2013). In this context, system (4) must be considered as two intrinsic sub-dynamics for X and Y that are weakly coupled. The response theory quantifies the contribution of the “perturbation” Ψ_X , Ψ_Y to the invariant measure² $\tilde{\rho}$ of the fully coupled system (4) as:

$$\tilde{\rho} = \rho_0 + \delta_{\psi} \rho^{(1)} + \delta_{\psi, \psi} \rho^{(2)} + O(\Psi^3) \quad (6)$$

where ρ_0 is the invariant measure of the uncoupled system which is also supposed to be an existing, well-defined SRB measure. As shown in Wouters and Lucarini (2012), this theory gives the framework to parameterize the effect of the coupling on the component X . The parameterization is based on three different terms having a response similar, up to order two, to the couplings Ψ_X and Ψ_Y :

$$\dot{X} = F_X(X) + M_1(X) + M_2(X, t) + M_3(X, t) \quad (7)$$

²The theory assumes that for the system under consideration, a SRB measure (Young, 2002) exists (e.g., an Axiom-A system).

where

$$M_1(X) = \left\langle \Psi_X(X, Y) \right\rangle_{\rho_{0,Y}} \quad (8)$$

is an averaging term. $\rho_{0,Y}$ is the measure of the uncoupled system $\dot{Y} = F_Y(Y)$. The term $M_2(X, t) = \sigma_R(X, t)$ is a correlation term:

$$\left\langle \sigma_R(X, t) \otimes \sigma_R(X, t + s) \right\rangle = \mathbf{g}(X, s) = \left\langle \Psi'_X(X, Y) \otimes \Psi'_X(\phi_X^s(X), \phi_Y^s(Y)) \right\rangle_{\rho_{0,Y}} \quad (9)$$

where \otimes is the outer product, $\Psi'_X(X, Y) = \Psi_X(X, Y) - M_1(X)$ is the centered perturbation, and ϕ_X^s, ϕ_Y^s . There are two flows and two systems $\dot{X} = F_X(X)$ and $\dot{Y} = F_Y(Y)$. The process σ_R is thus obtained by taking the square root of the matrix \mathbf{g} , which is here accomplished by a decomposition of Ψ'_X on a proper basis (Wouters and Lucarini, 2012). The M_3 term is a memory term:

$$M_3(X, t) = \int_0^\infty ds h(X(t-s), s). \quad (10)$$

involving the memory kernel

$$h(X, s) = \left\langle \Psi_Y(X, Y) \cdot \nabla_Y \Psi_X(\phi_X^s(X), \phi_Y^s(Y)) \right\rangle_{\rho_{0,Y}} \quad (11)$$

All the averages are thus taken with $\rho_{0,Y}$, the invariant measure of the unperturbed system $\dot{Y} = F_Y(Y)$. This particular choice of the measure is due to the perturbative nature of the method and simplifies the averaging procedure in many cases. The terms M_1 , M_2 and M_3 , are derived (Wouters and Lucarini, 2012) such that their responses up to order two match the response of the perturbation Ψ_X and Ψ_Y . Consequently, this ensures that for a *weak coupling*, the response of the parameterization (7) on the observables will be approximately the same as the coupling.

The advantages of this simplified averaging procedure (by using $\rho_{0,Y}$) should be tempered by the additional cost induced by the computation of the memory term, the latter implying that this parameterization is a non-Markovian one (Chekroun et al., 2015). However, the integral (10) in this memory term must only be evaluated from 0 up to the timescale τ_Y of the fast variable, due to the exponential decrease of the integrand. Moreover, in some cases, this non-Markovian parameterization can be effectively replaced by a Markovian one (Wouters et al., 2016).

3.2 Singular Perturbation Theory Method

Singular perturbation methods were developed in the 1970s for the analysis of the linear Boltzmann equation in some asymptotic limit (Ellis et al., 1975; Grad, 1969; Majda et al., 2001; Papanicolaou, 1976). Here, these methods are applicable if the problem can be cast into a Fokker–Planck equation. The procedure described in Majda et al. (2001) requires assumptions on the timescales of the different terms of system (4). In terms of the small parameter $\delta = \tau_Y/\tau_X$ defined in Sect. 2, the fast variability of the unresolved component Y is considered of order $O(\delta^{-2})$ and modeled as an Ornstein–Uhlenbeck process. The Markovian nature of the process defined by Eq. (4) and its singular behavior in the limit of an infinite timescale separation ($\delta \rightarrow 0$) allow then to apply the method.

More specifically, the parameter δ serves to distinguish terms with different timescales and is then used as a small perturbation parameter. In this setting, the backward Fokker–Planck equation reads (Majda et al., 2001):

$$-\frac{\partial \rho^\delta}{\partial s} = \left[\frac{1}{\delta^2} \mathcal{L}_1 + \frac{1}{\delta} \mathcal{L}_2 + \mathcal{L}_3 \right] \rho^\delta \quad (12)$$

where the function $\rho^\delta(s, X, Y|t)$ is defined with the final value problem $f(X)$: $\rho^\delta(t, X, Y|t) = f(X)$. The function ρ^δ can be expanded in terms of δ and inserted in Eq. (12). The zeroth order of this equation ρ^0 can be shown to be independent of Y and its evolution given by a closed, averaged backward Fokker–Planck equation (Kurtz, 1973):

$$-\frac{\partial \rho^0}{\partial s} = \bar{\mathcal{L}} \rho^0 \quad (13)$$

This equation is obtained in the limit $\delta \rightarrow 0$ and gives the sought limiting, averaged process $X(t)$. Note that this procedure does not necessarily require the presence of the explicit small parameter δ in the original Eq. (4). Since δ disappears from Eq. (13), one can simply use the parameter to identify the fast terms to be considered, and eventually consider $\delta = 1$ (Franzke et al., 2005).

The parameterization obtained by this procedure is given by Franzke et al. (2005):

$$\dot{X} = F_X(X) + G(X) + \sqrt{2} \sigma_{\text{MTV}}(X) \cdot \tilde{\xi}(t) \quad (14)$$

with

$$G(X) = \int_0^\infty ds \left[\langle \psi_Y(X, Y) \cdot \nabla_Y \psi_X(X, \phi_Y^s(Y)) \rangle_{\bar{\rho}} + \langle \psi_X(X, Y) \cdot \nabla_X \psi_X(X, \phi_Y^s(Y)) \rangle_{\bar{\rho}} \right] \quad (15)$$

$$\sigma_{\text{MTV}}(X) = \left(\int_0^\infty ds \langle \Psi'_X(X, Y) \Psi'_X(X, \phi_Y^s(Y)) \rangle_{\tilde{\rho}} \right)^{1/2} \quad (16)$$

with the same notation as in the previous subsection. The measure $\tilde{\rho}$ is the measure of the $O(\delta^{-2})$ perturbation, i.e., the source of the fast variability of the unresolved Y component. This measure thus depends on which terms of the unresolved component are considered as “fast,” and some assumptions should here be made. For instance, it is customary to consider as the fast terms the quadratic terms in Y and to replace them by Ornstein–Uhlenbeck processes whose measures are used to compute the averages (Franzke et al., 2005; Majda et al., 2001).

Finally, if one assumes that the source of the fast variability in the sub-system is given by the “intrinsic” term $F_Y(Y)$ (such that $\tilde{\rho} = \rho_{0,Y}$) and if the perturbation Ψ_X only depends on Y , this parameterization is simply given by the integration of the function $g(s)$ and $h(X, s)$ of the response theory parameterization given by Eqs. (9) and (11). This can be interpreted as an averaging of the latter parameterization when the timescale separation is infinite and X can thus be considered as constant over the timescale of the integrand. Therefore, M_2 can be modeled as a white noise and the memory term is Markovian.

3.3 Hasselmann Averaging Method

Since the initial work of Hasselmann in the 1970s (Hasselmann, 1976), various approaches have been considered to average directly the effects of the “fast” evolving variables on the “slow” ones. These methods assume in general a sufficient timescale separation between the resolved and unresolved components of the systems, and a direct average can be performed as,

$$\dot{X} = \bar{F}(X) = \langle F(X, Y) \rangle_{\rho_{Y|X}} \quad (17)$$

where $\rho_{Y|X}$ is the measure of the system

$$\dot{Y} = H(X, Y) \quad (18)$$

conditional on the value of X . In this approach, X is thus viewed as a constant parameter for the unresolved dynamics. In other words, this particular framework assumes that since X is slowly evolving with respect to the typical timescale of Y , it can be considered as “frozen” while Y evolves. With some rigorous assumptions, this approach has been mathematically justified (Kifer, 2003) and applied successfully to idealized geophysical models (Arnold et al., 2003) with non-trivial invariant measures. In the same vein, an approximation has been proposed in Abramov (2013) for the average (17), assuming that F is at most quadratic,

$$\langle F(X, Y) \rangle_{\rho_{Y|X}} = F(X, \bar{Y}(X)) + \frac{1}{2} \frac{\partial^2 F}{\partial Y^2}(X, \bar{Y}(X)) : \Sigma(X) \quad (19)$$

where “:” means the element-wise matrix product with summation and where

$$\bar{Y}(X) = \langle Y \rangle_{\rho_{Y|X}} \quad (20)$$

$$\Sigma(X) = \langle (Y - \bar{Y}(X) \otimes (Y - \bar{Y}(X))) \rangle_{\rho_{Y|X}}. \quad (21)$$

The approximation to the second order is particularly well suited for the application to atmospheric and climate flows for which the quadratic terms are usually the main non-linearities associated with the advection in the system.

In Abramov (2013), an approach based on the fluctuation–dissipation theorem is proposed to estimate the mean state $\bar{Y}(X)$ and the covariance matrix $\Sigma(X)$.

The deterministic parameterization (17) can be recast in a stochastic parameterization following the same principle. Such a parameterization is derived in Arnold et al. (2003), Abramov (2015) and reads

$$\dot{X} = \bar{F}(X) + \sigma_A(X) \cdot \xi(t) \quad (22)$$

with

$$\sigma_A(X) = \left(2 \int_0^\infty ds \left\langle (F(X, \phi_{Y|X}^s(Y)) - \bar{F}(X)) (F(X, Y) - \bar{F}(X)) \right\rangle_{\rho_{Y|X}} \right)^{1/2} \quad (23)$$

where $\phi_{Y|X}^s$ is the flow of the system (18) for X constant (“frozen”). A drawback of such an approach is that it requires that the measure $\rho_{Y|X}$ exists and is well-defined [ideally a SRB measure (Arnold et al., 2003)]. Such a requirement may not be always fulfilled, for instance, if the fast system conditional on the state X is unstable and does not possess any attractor (see Sect. 4 for an example).

3.4 Empirical Methods

The empirical methods are generally based on the statistical analysis of the timeseries Y of the full system (4). Many procedures exist as discussed in Sect. 1 but we will consider here a method based on state-dependent AR(1) processes proposed in Arnold et al. (2013). In this case, a timeserie $r(t)$ of the coupling part Ψ_X of the X tendency must first be computed with (4). The parameterization is then given by

$$\dot{X} = F_X(X) + \mathcal{U}(X) \quad (24)$$

with

$$\mathcal{U}(X) = \mathcal{U}_{\text{det}}(X) + e(X(t), t). \quad (25)$$

The function $\mathcal{U}_{\text{det}}(X)$ represents the deterministic part of the parameterization and is obtained by a least-squares fit of the timeserie $r(t)$ versus the timeserie $X(t)$ with the cubic function $\mathcal{U}_{\text{det}}(X) = p_0 + p_1 X + p_2 X^2 + p_3 X^3$. The ‘‘stochastic’’ part $e(X(t), t)$ is then given by the following state-dependent AR(1) process:

$$e(X(t), t) = \phi \frac{\sigma_e(X(t))}{\sigma_e(X(t - \Delta t))} e(X(t - \Delta t), t - \Delta t) + \sigma_e(X(t)) (1 - \phi^2)^{1/2} z(t) \quad (26)$$

where $z(t)$ is a standard Gaussian white noise process. The parameters of the process e are determined by considering the residual timeserie $r(t) - \mathcal{U}_{\text{det}}(X(t))$ to compute the lag-1 autocorrelation ϕ and the state-dependent standard deviation $\sigma_e(X)$ which is modeled as $\sigma_e(X) = \sigma_0 + \sigma_1 |X|$ with the parameters σ_0 and σ_1 given by a binning procedure. The parameter Δt is the time step of integration of Eq. (24). Other empirical parameterizations have been proposed by Arnold et al. (2013), notably one with the function $\mathcal{U}(X) = (1 + e(t)) \mathcal{U}_{\text{det}}(X)$ which resembles the SPPT³ parameterization used in the ECMWF⁴ Numerical Weather Prediction model (Buizza et al., 1999). However, the study shows no substantial differences with the parameterization (25).

4 Applications and Results

In this section, we will illustrate the various parameterizations described in Sect. 3 to the following example:

$$\begin{cases} \dot{X} = -DX + q\xi(t) + \frac{\varepsilon}{\delta} Y^T \cdot \mathbf{C} \cdot Y \\ \dot{Y} = \frac{1}{\delta^2} (\mathbf{A} \cdot Y + \delta \mathbf{B}_Y \cdot \xi_Y(t)) + \frac{\varepsilon}{\delta} X \mathbf{V} \cdot Y \end{cases} \quad (27)$$

where $D > 0$, $q > 0$ and

$$Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}. \quad (28)$$

The matrices involved are defined as

³Acronym for *Stochastically Perturbed Parameterization Tendencies Scheme*.

⁴Acronym for *European Center for Medium-Range Weather Forecasts*.

$$\mathbf{C} = \begin{bmatrix} 0 & B \\ B & 0 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} -a & \beta \\ -\beta & -a \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} 0 & B_1 \\ B_2 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{B}_Y = \begin{bmatrix} q_Y & 0 \\ 0 & q_Y \end{bmatrix} \quad (29)$$

with $a, \beta, q_Y > 0$. The process $\xi(t)$ and $\xi_Y(t)$ are uncorrelated standard Gaussian white noise processes.

The X and Y variables represent, respectively, the resolved and the unresolved sub-systems. The parameter $\delta > 0$ quantifies the timescale separation of the terms of the tendencies of the two components, with the three timescales $O(1), O(\delta^{-1})$, and $O(\delta^{-2})$ as in Majda et al. (2001) (see also Sect. 3.2). Additionally, the parameter $\varepsilon > 0$ controls the coupling strength between the two sub-systems. In this setup the coupling is thus proportional to the ratio ε/δ , and therefore the characterization of the coupling as “weak” depends directly on the timescale separation.

The deterministic part of Eq. (27) is a well-known model called a triad encountered in fluid dynamics (Ohkitani and Kida, 1992; Smith and Waleffe, 1999; Waleffe, 1992), and in simplified geophysical flows, e.g. Majda et al. (2001), Wouters et al. (2016). Due to the presence of invariant manifolds, its mathematical structure can be found in higher-order model. See Demaeyer and Vannitsem (2016) for an example of such structure in the framework of a coupled ocean-atmosphere model. In the present study, the interest of the stochastic triad model (27) is that, $H(X, Y)$ being linear in Y , the measure $\rho_{0,Y}$ and $\rho_{Y|X}$ can be analytically computed since both $\dot{Y} = F_Y(Y)$ and $\dot{Y} = H(z, Y)|_{z=X}$ are two-dimensional Ornstein–Uhlenbeck processes. Therefore, for this simple case, the set of methods proposed in the previous section can be applied exactly without resorting to a binning procedure of the output of the Y sub-system.⁵

As energy conservation is a rule in physical systems in the absence of dissipation and fluctuations, we will adopt this rule for the current system. System (27) conserves the “energy” $(X^2 + y_1^2 + y_2^2)/2$ if the coefficient B, B_1 , and B_2 are chosen such that (Majda et al., 2001; Smith and Waleffe, 1999)

$$2B + B_1 + B_2 = 0. \quad (30)$$

It allows for the following configurations of their signs: $(+, -, -)$, $(+, +, -)$, $(+, -, +)$, $(-, +, +)$, $(-, +, -)$, $(-, -, +)$. These different configurations are associated with different kinds of energy exchange scenarios and different stability properties (Waleffe, 1992).

We will focus on the two configurations $(-, -, +)$ and $(-, +, +)$, with parameters

1. $B = -0.0375, B_1 = -0.025, B_2 = 0.1$
2. $B = -0.0375, B_1 = 0.025, B_2 = 0.05$

⁵Except for the empirical methods which by definition use this kind of procedures.

and consider various values of the parameters δ and ε . The other parameters are fixed to $a = 0.01$, $D = 0.01$, and $\beta = 0.01/12$. Once the parameterizations have been developed, the different model versions have integrated over 4.5×10^5 timeunits with a timestep $\Delta t = 0.01$ after a transient period of 5.0×10^4 timeunits to let the system relax to its stationary state. The state X has been recorded every 0.1 timeunit, giving a dataset of 4.5×10^6 points for the analysis. The parameterizations given by Eqs. (7), (14), and (22) have been integrated with a second order Runge–Kutta (RK2) stochastic scheme which converges to the Stratonovich calculus (Hansen and Penland, 2006; Rümelin, 1982). Equation (24) has been integrated with a deterministic RK2 scheme where the stochastic forcing $e(X, t)$ is considered constant during the timestep. The memory term M_3 appearing in the parameterization (7) and given by the integral (10) over the past of X has been computed numerically at each timestep. Although it increases considerably the integration time, this method is adopted in order to clarify the memory effect in Eq. (7). A Markovianization of this parameterization is possible (Wouters et al., 2016) but in the present case it would have required some assumptions that would blur the comparison of the methods.

The relative performances of the parameterizations can be tested in multiple ways, by comparing the climatology (the average state) or the variability (variance) of the systems (Nicolis, 2005). Another method is to look at the predictive skill score of the models, that is the ability of the parameterizations to provide skillful forecast compared to original system, as in Arnold et al. (2013), Wouters et al. (2016). On longer term, the good representation of the “climate” of a model by the parameterizations can be assessed by looking at the stationary probability densities and comparing them using some score (Abramov, 2012, 2013, 2015; Croomelin and Vanden-Eijnden, 2008; Franzke et al., 2005). The decorrelation properties of the models and the parameterizations can also be tested, to provide information about the correct representation of the timescales of the models. All those different aspects can be significant, depending on the purpose of the parameterization scheme. However, for the brevity of the present work, we shall focus on the probability densities and whether or not they are correctly reproduced by the parameterizations.

We present now the results obtained by with the proposed methods and consider first the different measures used for averaging in system (27).

4.1 Stability and Measures

All the ingredients needed to compute the parameterizations presented in Sect. 3 can be derived with the help of the covariance and the correlation of the Y variables in the framework of two different systems related to the unresolved dynamics, namely the unperturbed dynamics $\dot{Y} = F_Y(Y)$ and the unresolved dynamics $\dot{Y} = H(X, Y)$ with X frozen. The measure of the former is necessary to derive the response theory and the singular perturbation based parameterizations, while the latter is needed for the Hasselmann averaging method. These two systems are both two-dimensional

Ornstein–Uhlenbeck processes of the form

$$\dot{Y} = \mathbf{T} \cdot Y + \mathbf{B} \cdot \xi_Y(t) \quad (31)$$

for which, respectively, $\mathbf{T} = \mathbf{A}/\delta^2$ and $\mathbf{T} = \mathbf{A}/\delta^2 + (\varepsilon X/\delta) \mathbf{V} \cdot Y$. In both cases, we have $\mathbf{B} = \mathbf{B}_Y/\delta$. Their measure is then given by Wouters et al. (2016)

$$\rho(Y) = \frac{1}{\mathcal{Z}} \exp\left(-\frac{1}{2} Y^\top \cdot \Sigma^{-1} \cdot Y\right) \quad (32)$$

where \mathcal{Z} is a normalization factor and where Σ is the covariance matrix solution of

$$\mathbf{T} \cdot \Sigma + \Sigma \cdot \mathbf{T}^\top = -\mathbf{B} \cdot \mathbf{B}^\top. \quad (33)$$

In order for these processes to be stable, the real part of the eigenvalues of the matrix \mathbf{T} must be negative (Gardiner, 2009) for every state X that the full coupled system (27) can possibly achieve. The eigenvalues of the system $\dot{Y} = F_Y(Y)$ are $\lambda_\pm = (-a \pm i\beta)/\delta^2$ and it is thus always stable (since $a > 0$ and $\beta \in \mathcal{R}$). On the other hand, the system $\dot{Y} = H(X, Y)$ has the eigenvalues $\lambda_\pm = (-a \pm \sqrt{\Delta(X)})/\delta^2$ with

$$\Delta(X) = -(B_1 X \delta \epsilon + \beta) (\beta - B_2 X \delta \epsilon) \quad (34)$$

Therefore, if

$$\text{Re}\left(\sqrt{\Delta(X)}\right) > a \quad (35)$$

for some X , the Ornstein–Uhlenbeck process is unstable, and it is then called an *explosive* process. For any initial condition, the process diverges, and thus the only possible stationary measure is the trivial one. Consequently, Eq. (33) gives nonphysical solutions, the stationary covariance matrix does not exist, and the parameterizations depending upon cannot be derived.

For the system (27), if $\text{sgn}(B_1 B_2) = -1$, as in case 1, then the process is stable for every X if $a^2 > -\frac{(B_1+B_2)^2 \beta^2}{4B_1 B_2}$. For case 1, this inequality is satisfied, and thus the process (31) is stable for every X . Moreover, depending on the sign of $\Delta(X)$, the process for X fixed is a stochastic focus (if $\Delta(X) > 0$) or a stochastic damped oscillator (if $\Delta(X) < 0$). Here, it is a focus if

$$X \in [\min(-\beta/\delta \varepsilon B_1, \beta/\delta \varepsilon B_2), \max(-\beta/\delta \varepsilon B_1, \beta/\delta \varepsilon B_2)] \quad (36)$$

and an oscillator otherwise. That is, for the considered ε and δ parameters value, the system (27) is an oscillator for most of the X values.

If $\text{sgn}(B_1 B_2) = 1$, as in case 2, then the condition (35) must be satisfied for every state X . For case 2, this inequality was not satisfied for every state X for most of the

values of the ε and δ parameters considered (see Sect. 4.3 below). The stability is therefore reversed as the system is non-oscillating for most of the X values.

To summarize, if B_1 and B_2 are of opposite sign, the dynamics of $\dot{Y} = H(X, Y)$ is stable and generally oscillatory. If B_1 and B_2 have the same sign, then the dynamics is unstable in most cases and generally hyperbolic. This is a consequence of the well-known difference of stability of the triads depending on their energy exchange properties (Waleffe, 1992).

For the interested reader, the exact calculation of the parameterization of Sect. 3 using the covariance and correlation matrices is detailed in the Appendix (see section “[Appendix: Practical Computation of the Parameterizations](#)”).

4.2 The $(-, -, +)$ Stochastic Triad (Case 1)

Let us now consider case 1 corresponding to the $(-, -, +)$ stochastic triad for two different values of the timescale separation $\delta = 0.1$ and 0.4 . For each of these timescale separation, we considered three values of the coupling strength ε : 0.05, 0.125, and 0.4. The probability densities associated with these different systems are represented in Figs. 1 and 2. For a timescale separation $\delta = 0.1$, the fully coupled dynamics given by Eq. (27) is quite well represented by all the proposed parameterizations. Since it is hard to distinguish the different density curves, a score such as the Hellinger distance (Arnold et al., 2013)

$$H(P, Q) = \frac{1}{2} \int (\sqrt{dP} - \sqrt{dQ})^2 \quad (37)$$

between the distribution P of the full coupled system and the distribution Q of the parameterizations is worth computing to quantify the differences (the smaller the better). It is depicted for $\delta = 0.1$ on Fig. 3, and it shows that for a very small coupling parameter $\varepsilon = 0.05$, the best parameterization is the response theory given by Eq. (7). For larger values of ε , it is the Hasselmann averaging method which performs best. The empirical method gives a good correction of the uncoupled dynamics for $\varepsilon = 0.125$ but diverges for $\varepsilon = 0.4$. This may be due to instabilities introduced by the cubic deterministic parameterization $\mathcal{Z}_{\text{det}}(X)$ or to the inadequacy of the fitting function $\sigma_0 + \sigma_1 |X|$ for the standard deviation $\sigma_e(X)$ in the AR(1) process (26). Indeed, in general, this model fits quite well the statistics in the neighborhood of $X = 0$, but the standard deviation reaches a plateau for higher values of X . A more complicated fitting function would thus be necessary to get a stable dynamics. For a timescale separation $\delta = 0.4$, the same conclusions are reached, but the singular perturbation method performs not very well in all cases, as illustrated in Fig. 4 that for $\varepsilon = 0.125$ and 0.4 . The response based and singular perturbation methods are even less effective than the uncoupled dynamics. It is not surprising for the latter since it is supposed to be valid in the limit $\delta \rightarrow 0$.

Fig. 1 Probability densities of the full coupled dynamics (27), the uncoupled dynamics $\dot{X} = -DX + q\xi(t)$, and the parameterized model versions for the timescale separation $\delta = 0.1$ and for the triad parameters of case 1. The empirical parameterization density is not represented for $\epsilon = 0.4$ due to its divergence

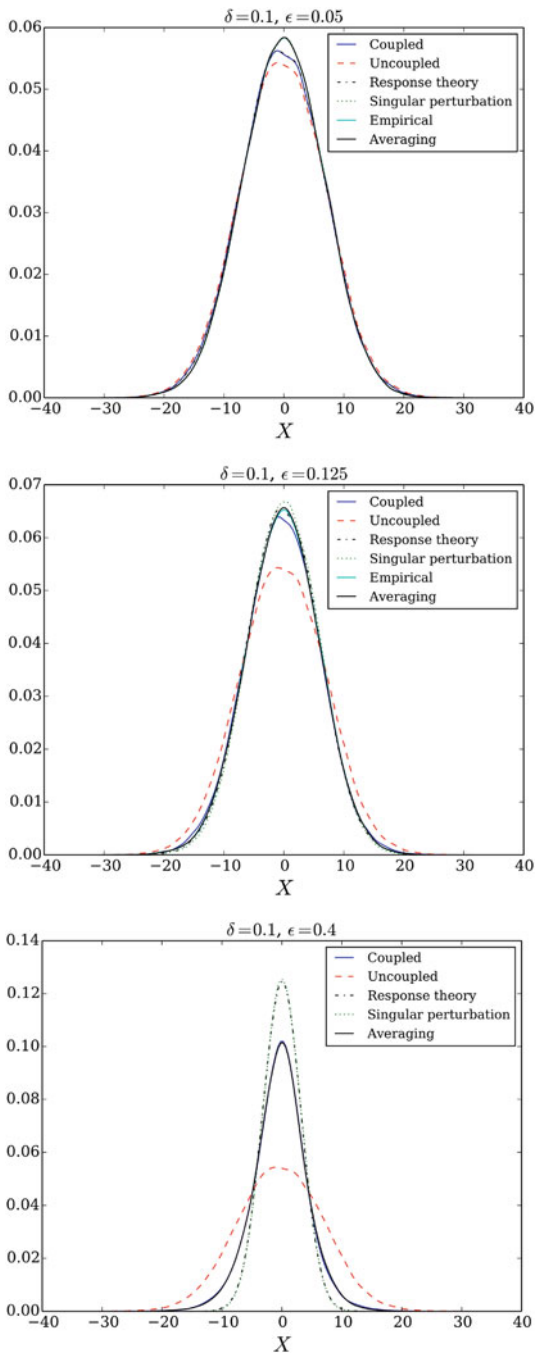


Fig. 2 Same as Fig. 1 but for the timescale separation $\delta = 0.4$

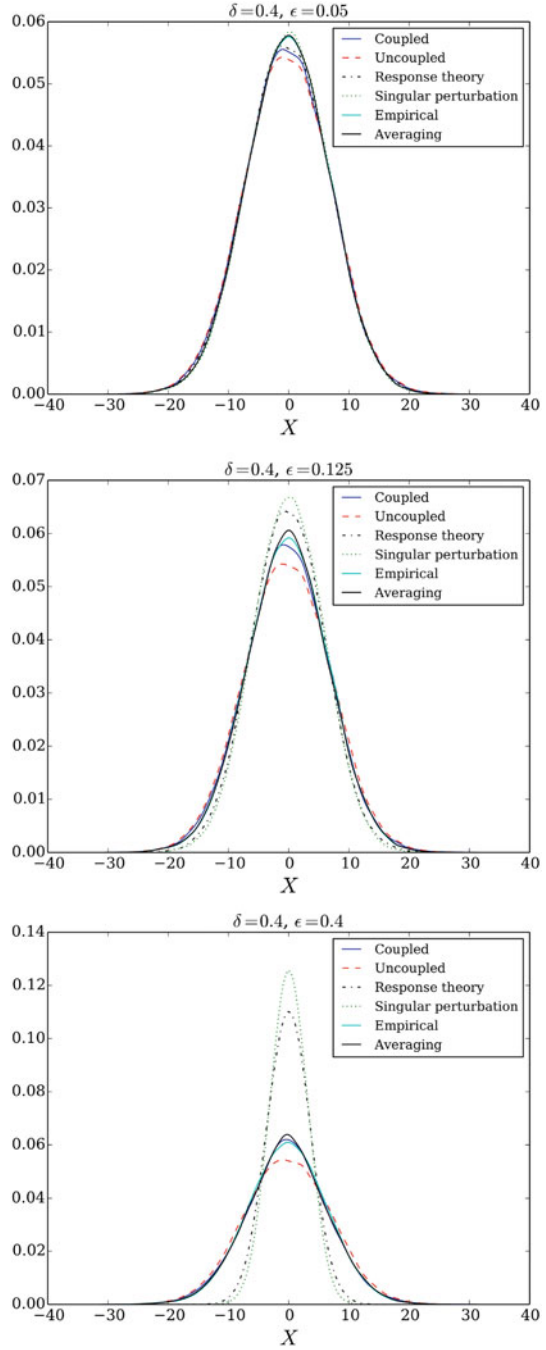


Fig. 3 Hellinger distance (37) between the densities of the different parameterized models and the full coupled system density for case 1. A small distance indicates that the two densities concerned are very similar. The Hellinger distance between the full coupled system and the uncoupled system distribution is depicted as reference. In case $\epsilon = 0.4$, the empirical parameterization is not represented due to its divergence

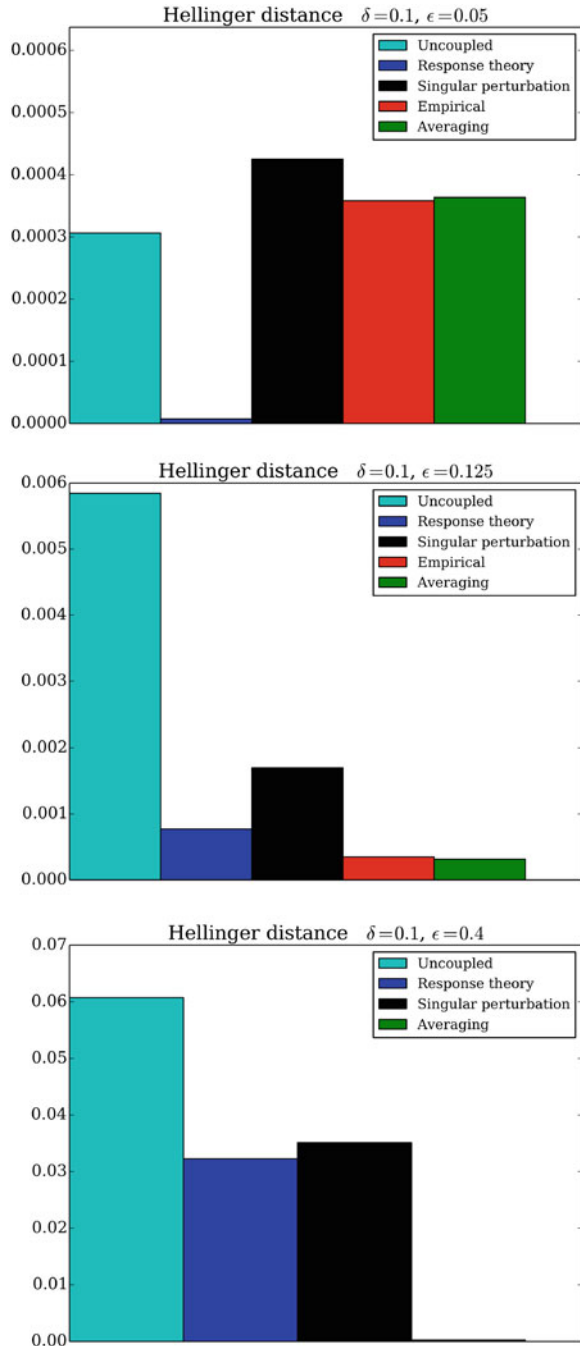


Fig. 4 Same as Fig. 3 but for the timescale separation $\delta = 0.4$

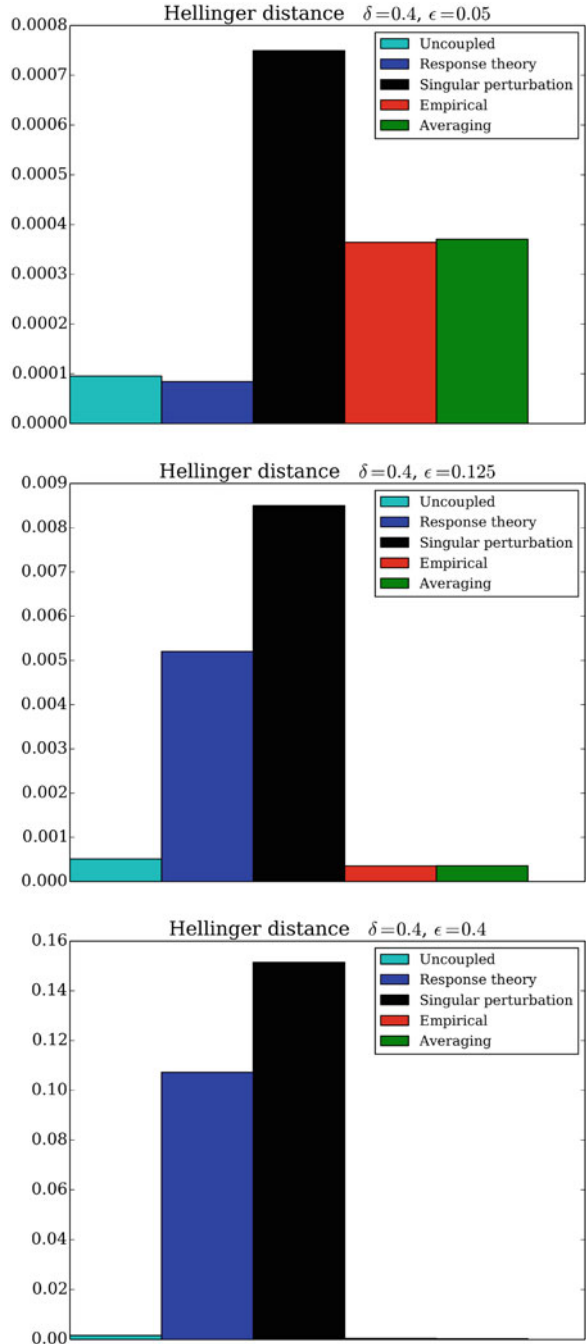


Fig. 5 Probability densities of the coupled full dynamics (27), the uncoupled dynamics $\dot{X} = -DX + q\xi(t)$, and the parameterized models for the timescale separation $\delta = 0.1$ and for the triad parameters of case 2. The direct averaging parameterization density is only represented for $\epsilon = 0.05$ because the system diverges for the other values

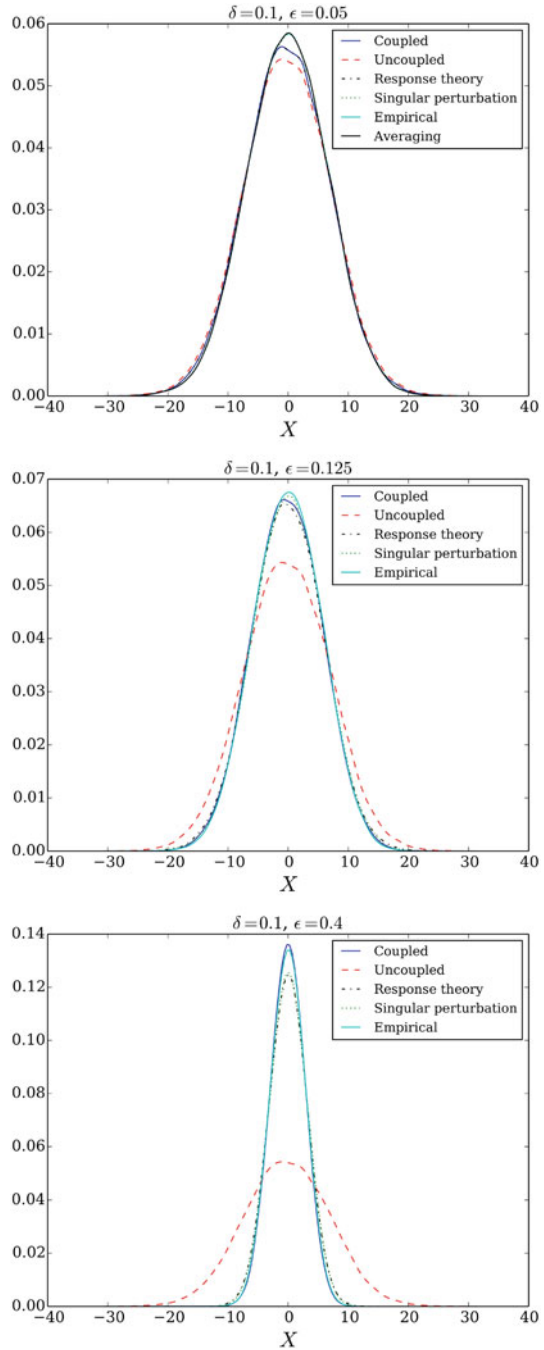


Fig. 6 Same as Fig. 5 but for the timescale separation $\delta = 0.4$

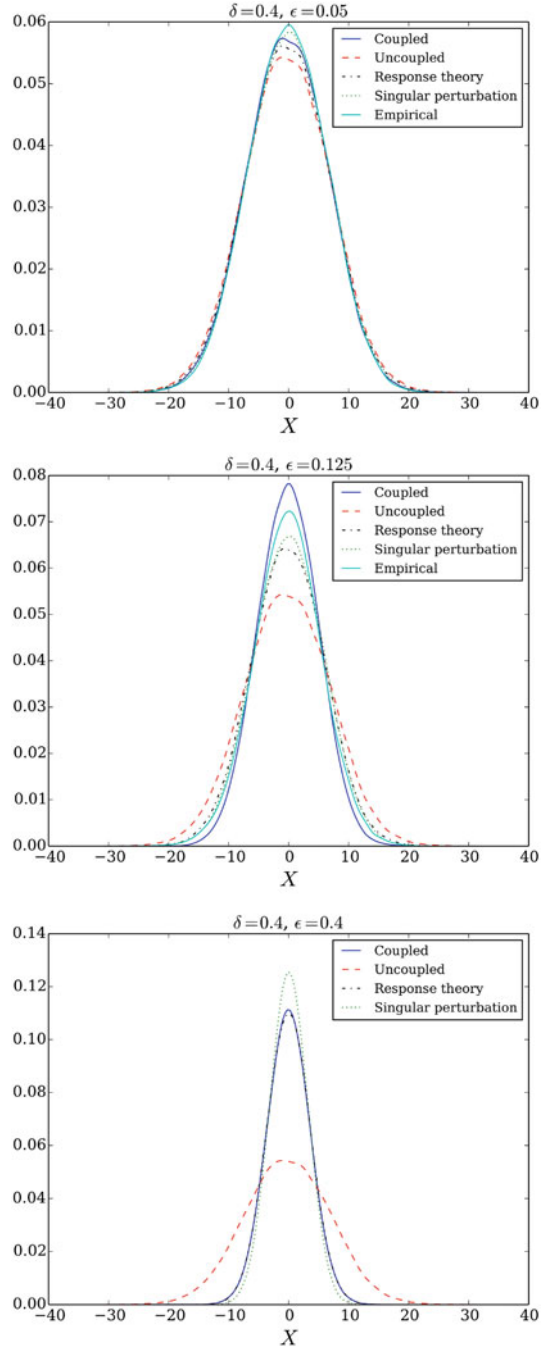


Fig. 7 Hellinger distance (37) between the densities of the different parameterized models and the full coupled system density for case 2. A small distance indicates that the two densities concerned are very similar. The Hellinger distance between the full coupled system and the uncoupled system distribution is depicted as reference. In case $\epsilon = 0.4$, the empirical parameterization is not represented due to its divergence

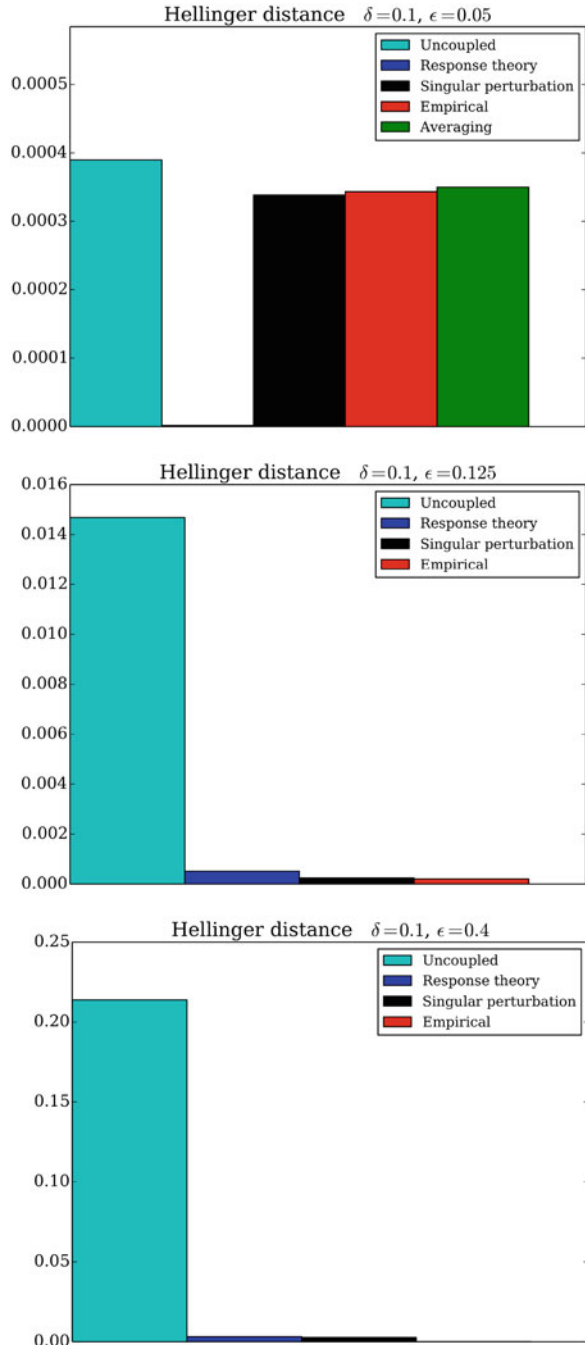
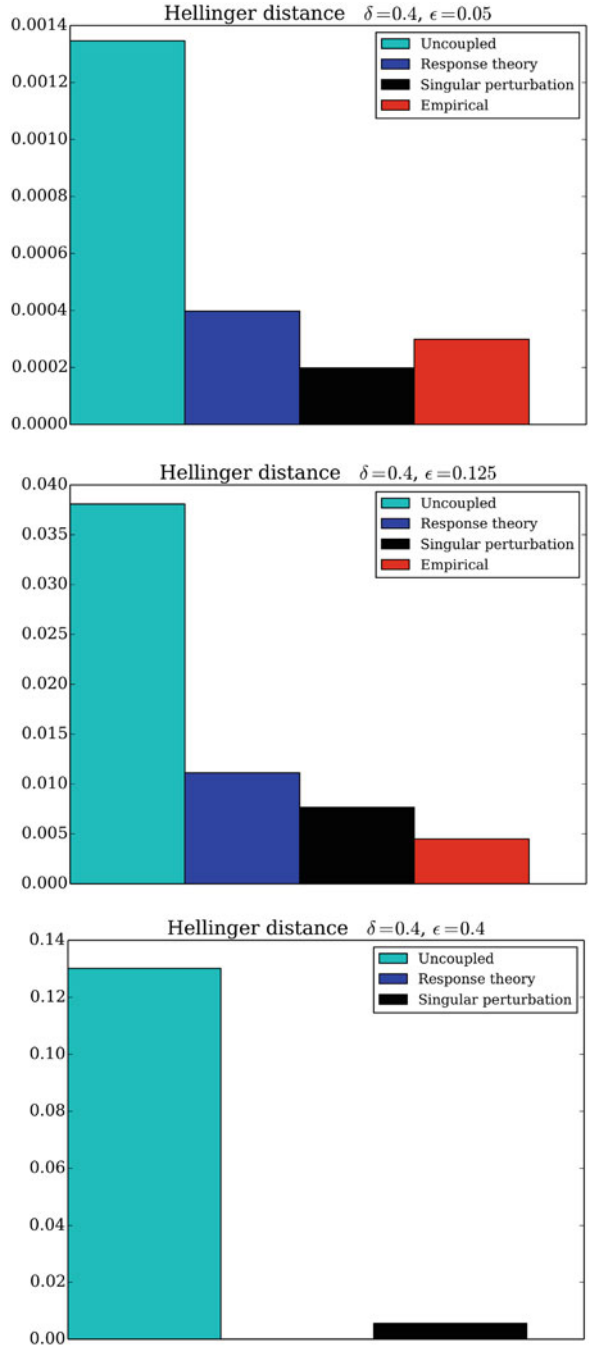


Fig. 8 Same as Fig. 7 but for the timescale separation $\delta = 0.4$



4.3 The $(-, +, +)$ Stochastic Triad (Case 2)

We now consider the parameters of case 2, for which the system (27) is a $(-, +, +)$ stochastic triad. The probability densities are depicted in Figs. 5 and 6, and the Hellinger distances are shown in Figs. 7 and 8. First, we must remark that the parameterization based on the Hasselmann's averaging method is not defined for most of the δ and ε parameters values. It is due to the fact that the dynamics of the unresolved component Y with X considered as a parameter is unstable, as shown in Sect. 4.1. Indeed, this linear system undergoes a bifurcation at some value X^* which destabilizes the dynamics $\dot{Y} = H(X, Y)$ with X frozen. Therefore, the measure $\rho_{Y|X}$ is not defined for some ranges of the full X dynamics and the method fails. The only case where this destabilization does not occur is for $\delta = 0.1$ and $\varepsilon = 0.05$, but the parameterization does not perform well. For these parameter values, the only parameterization that performs very well is the one based on response theory. For the other values of the parameters δ and ε , all the parameterizations have good performances. A particularly unexpected result is the very good correction provided by the response theory and singular perturbation based methods for the extreme case $\delta = 0.4$ and $\varepsilon = 0.4$ (see the bottom panel of Fig. 8). This have to be contrasted with their bad performances in the case of the other triad (see the bottom panel of Fig. 4). Note that for this extreme case, the direct averaging method fails and the empirical method is unstable and diverges.

4.4 Discussion

The results obtained so far with these two types of triads highlight the utility of the parameterization schemes discussed here. First, the empirical parameterization gives usually good results when it does not destabilize the dynamics. However, this method requires a case by case time-consuming statistical analysis whose complexity increases with the dimensionality of the problem considered. Physically based parameterizations do not require such an analysis, and the best approach in the present system is the Hasselmann averaging one, but it requires that the dynamics of the unresolved system be stable. It was thus very effective to correct the dynamics of the $(-, -, +)$ triad, but not the other triad $(-, +, +)$. In this latter case, the perturbative methods like the singular perturbation method or the response theory method give very good results. This difference is quite intriguing and interesting. It indicates that different physically-based parameterizations should be considered depending on the kind of problems encountered. In particular, the stability properties of the system considered seem to play an important role. This conclusion holds whatever the timescale separation and for the most realistic values of the coupling strength between the components ($\varepsilon = 0.125$ and 0.4). However, for very small values of the coupling strength, the response based method seems to be the best approach in all cases.

A question that is left open in the present work is to determine precisely which stability property is giving the contrasting observed result. More specifically, is it the hyperbolic instability of the $(-, +, +)$ triad which makes the perturbative approach and the response based parameterization perform so well, or is it simply the fact that it is unstable? On the other hand, is it the damped oscillatory behavior of the $(-, -, +)$ triad which makes the Hasselmann's method works well, or is it simply the fact that it is stable? Such questions should be addressed in the case of a more complex, globally stable system, which allows to have locally stable and unstable fast dynamics.

5 Conclusions

The parameterization of subgrid-scale processes is an important tool in model reduction, in order to improve the statistical properties of the forecasting systems. The variety of approaches available bear witness of the richness of the field but at the same time can also lead to questions on the best choice for the problem at hand. The purpose of the present review was to describe briefly some of the most recent methods and to illustrate them on a simple stochastic triad example. The methods covered include perturbative methods like the Ruelle response theory (Wouters and Lucarini, 2012), the singular perturbation theory (Majda et al., 2001), averaging methods like the Hasselmann method (Arnold et al., 2003; Hasselmann, 1976) and an empirical method (Arnold et al., 2013). As expected, these parameterizations provided contrasting results depending on the timescale separation and on the coupling between the resolved variables and the subgrid one. But more importantly, our results in the context of this simple triad stress the importance of the underlying stability properties of the unresolved system. It thus confirms a known result that the structure of the Jacobian and of the Hessian of a given system controls the behavior and performance of model error parameterizations (Nicolis, 2005).

Further comparisons of the different methods are needed in the context of more sophisticated systems in order to analyze the role of the stability properties of the subgrid scale processes on their performances. This type of analysis is currently under way in the context of a coupled ocean-atmosphere system (De Cruz et al., 2016).

Appendix: Practical Computation of the Parameterizations

In the following section, for illustrative purposes, we detail the computation that we have made to obtain the result of the present review. We start with the method based on response theory.

Response Theory Method

We consider the system (27) with the form (4) in mind. In this case, the influence of the Y sub-system on the X sub-system is parameterized as:

$$\dot{X} = -DX + q\xi(t) + M_1(X) + M_2(X, t) + M_3(X, t) \quad (38)$$

where then terms M_1 , M_2 , and M_3 are, respectively, given by Eqs. (8), (9, and (10). The average in these formula are performed with the measure $\rho_{0,Y}$ of the unperturbed Y dynamics $\dot{Y} = F_Y(Y)$. Since this latter is an Ornstein–Uhlenbeck process, its measure is the Wiener measure

$$\rho_{0,Y}(Y) = \frac{1}{\mathcal{Z}} \exp\left(-\frac{1}{2} Y^\top \cdot \Sigma^{-1} \cdot Y\right) \quad (39)$$

where Σ is the covariance matrix solution of

$$\mathbf{A} \cdot \Sigma + \Sigma \cdot \mathbf{A}^\top = -\mathbf{B}_Y \cdot \mathbf{B}_Y^\top \quad (40)$$

and \mathcal{Z} is a normalization factor.

The covariance and correlation of the stationary process $\dot{Y} = F_Y(Y)$ are thus straightforward to compute (Gardiner, 2009):

$$\Sigma = \langle Y \otimes Y \rangle = \frac{q_Y^2}{2a} \mathbf{l} \quad (41)$$

$$\langle \phi_Y^t(Y) \otimes \phi_Y^s(Y) \rangle = \mathbf{E}(t-s) \cdot \Sigma, \quad t > s \quad (42)$$

$$\langle \phi_Y^t(Y) \otimes \phi_Y^s(Y) \rangle = \Sigma \cdot \mathbf{E}(s-t)^\top, \quad t < s \quad (43)$$

where \mathbf{l} is the identity matrix, ϕ_Y^t is the flow of $\dot{Y} = F_Y(Y)$, and the matrix $\mathbf{E}(t)$ is the exponential

$$\mathbf{E}(t) = \exp(\mathbf{A}t/\delta^2) = e^{-at/\delta^2} \begin{bmatrix} \cos(\beta t/\delta^2) & \sin(\beta t/\delta^2) \\ -\sin(\beta t/\delta^2) & \cos(\beta t/\delta^2) \end{bmatrix} \quad (44)$$

The various terms M_i are then computed as follows.

The Term M_1

It is the average term:

$$M_1(X) = \langle \psi_X(X, Y) \rangle \quad (45)$$

We have thus

$$M_1(X) = 2B\frac{\varepsilon}{\delta}\langle y_1, y_2 \rangle = 2B\frac{\varepsilon}{\delta}\Sigma_{12} = 0 \quad (46)$$

by using Eq. (41).

The Term M_2

It is the noise/correlation term which is defined here as:

$$M_2(t) = \sigma_R(t) \quad (47)$$

with

$$\langle \sigma_R(t)\sigma_R(t') \rangle = g(t - t') \quad (48)$$

and the correlation function

$$g(s) = \langle \Psi'_X(Y)\Psi'_X(\phi_Y^s(Y)) \rangle \quad (49)$$

where $\Psi'_X(Y) = \Psi_X(Y) - M_1$. The result in the present case is given by the formula [see Demaeyer and Vannitsem (2016)]:

$$g(s) = \frac{\varepsilon^2}{\delta^2} \text{Tr}((C + C^T) \cdot \Sigma \cdot E(s)^T \cdot C^T \cdot E(s) \cdot \Sigma) = \frac{\varepsilon^2}{\delta^2} \frac{q_Y^4}{a^2} B^2 e^{-2as/\delta^2} \cos(2\beta s/\delta^2) \quad (50)$$

The term M_2 must thus be devised as a process with the same correlation.

The Term M_3

This is the memory term, defined by

$$M_3(X, t) = \int_0^\infty ds h(X(t-s), s) \quad (51)$$

with the memory kernel

$$h(X, s) = \langle \Psi_Y(X, Y) \cdot \nabla_Y \Psi_X(\phi_Y^s(Y)) \rangle \quad (52)$$

which in the present case is given by the formula [see Demaeyer and Vannitsem (2016)]

$$h(X, s) = \frac{\varepsilon^2}{\delta^2} \text{Tr} \left((XV \cdot \Sigma) \cdot (\mathbf{E}(s))^T \cdot (\mathbf{C} + \mathbf{C}^T) \cdot \mathbf{E}(s) \right) \quad (53)$$

$$= \frac{\varepsilon^2}{\delta^2} \frac{q_Y^2}{a} XB(B_1 + B_2) e^{-2as/\delta^2} \cos(2\beta s/\delta^2) \quad (54)$$

The fact that the memory kernel (54) and the correlation function (50) present the same form implies that a Markovian parameterization is available (Wouters et al., 2016) even if by definition, Eq. (38) is a non-Markovian parameterization.

The Singular Perturbation Method

With this parameterization, the parameter δ serves to distinguish terms with different timescale and is then used as a small perturbation parameter (Franzke et al., 2005; Majda et al., 2001). The parameterization is given by:

$$\dot{X} = -DX + q\xi(t) + G(X) + \sqrt{2} \sigma_{\text{MTV}}(X) \cdot \tilde{\xi}(t) \quad (55)$$

with notably $\langle \xi(t)\tilde{\xi}(t') \rangle = 0$ and

$$G(X) = \int_0^\infty ds \langle \Psi_Y(X, Y) \cdot \nabla_Y \Psi_X(X, \phi_Y^s(Y)) \rangle_{\tilde{\rho}} \quad (56)$$

$$\sigma_{\text{MTV}}(X) = \left(\int_0^\infty ds \langle \Psi'_X(X, Y) \Psi'_X(X, \phi_Y^s(Y)) \rangle_{\tilde{\rho}} \right)^{1/2} \quad (57)$$

We see that the quantities appearing in this parameterization can easily be obtained from the functions h and g of section “[Response Theory Method](#)”. Indeed we have

$$G(X) = \int_0^\infty ds h(X, s) = \varepsilon^2 X q_Y^2 \frac{B(B_1 + B_2)}{2(a^2 + \beta^2)} \quad (58)$$

$$\sigma_{\text{MTV}}(X) = \int_0^\infty ds g(s) = \varepsilon^2 \frac{q_Y^4 B^2}{2a(a^2 + \beta^2)} \quad (59)$$

where we notice that the parameter δ has disappeared, since this parameterization is valid in the limit $\delta \rightarrow 0$.

Averaging Method

In this approach, we consider the system (2) and the parameterization (Abramov, 2013):

$$\dot{X} = \bar{F}(X) \quad (60)$$

with

$$\bar{F}(X) = \langle F(X, Y) \rangle_{\rho_{Y|X}} = F(X, \bar{Y}(X)) + \frac{1}{2} \frac{\partial^2 F}{\partial Y^2}(X, \bar{Y}(X)) : \Sigma(X) \quad (61)$$

and

$$\bar{Y}(X) = \langle Y \rangle_{\rho_{Y|X}} \quad (62)$$

$$\Sigma(X) = \langle (Y - \bar{Y}(X)) \otimes (Y - \bar{Y}(X)) \rangle_{\rho_{Y|X}} \quad (63)$$

where $\rho_{Y|X}$ is the measure of the system $\dot{Y} = H(X, Y)$ with X “frozen.” It is the measure of an Ornstein–Uhlenbeck process

$$\rho_{0,Y}(Y) = \frac{1}{\mathcal{Z}} \exp\left(-\frac{1}{2} Y^\top \cdot \Sigma^{-1}(X) \cdot Y\right) \quad (64)$$

where \mathcal{Z} is a normalization factor and $\Sigma(X)$ is the stationary covariance matrix solution of

$$\mathsf{T}(X) \cdot \Sigma + \Sigma \cdot \mathsf{T}(X)^\top = -\frac{1}{\delta^2} \mathsf{B}_Y \cdot \mathsf{B}_Y^\top \quad (65)$$

with

$$\mathsf{T}(X) = \mathsf{A}/\delta^2 + \varepsilon X \mathsf{V}/\delta. \quad (66)$$

With the help of $\bar{Y}(X) = 0$ and $\Sigma(X)$, we can now rewrite Eq. (61) as

$$\begin{aligned} \bar{F}(X) &= F(X, 0) + \frac{\varepsilon}{\delta} \mathsf{C} : \Sigma(X) \\ &= -DX + q\xi(t) + \frac{B(B_1 + B_2) q_Y^2 X \varepsilon^2}{2(a^2 + \beta^2 - X\beta\delta\varepsilon B_2 + X\delta\varepsilon B_1(\beta - X\delta\varepsilon B_2))} \end{aligned} \quad (67)$$

This forms a deterministic averaging parameterization. It can be extended into a stochastic parameterization (Abramov, 2015) as follows:

$$\dot{X} = \bar{F}(X) + \sigma_A(X) \cdot \xi(t) \quad (68)$$

with

$$\sigma_A(X) = \sqrt{S(X)} \quad (69)$$

and

$$S(X) = 2 \int^{\infty} \left(2 \int_0^{\infty} ds \left\langle (F(X, \phi_{Y|X}^s(Y)) - \bar{F}(X)) (F(X, Y) - \bar{F}(X)) \right\rangle_{\rho_{Y|X}} \right)^{1/2} \quad (70)$$

We thus have

$$S(X) = 2 \frac{\varepsilon^2}{\delta^2} \int_0^{\infty} ds \operatorname{Tr} \left((C + C^T) \cdot \Sigma(X) \cdot \exp[\mathbf{T}(X)^T s] \cdot C^T \cdot \exp[\mathbf{T}(X)s] \cdot \Sigma(X) \right)$$

where we have extended the result of Eq. (50) to the stationary Ornstein–Uhlenbeck process $\dot{Y} = H(X, Y)$ for X “frozen”. The function $S(X)$ can be computed analytically using mathematical software but is a very complicated function that is not worth displaying in this review. This can, however, be provided upon query to the authors.

References

- Abramov, R.V. 2012. *Multiscale Modeling & Simulation* 10(1): 28.
- Abramov, R.V. 2013. *Multiscale Modeling & Simulation* 11(1): 134.
- Abramov, R. 2015. *Fluids* 1(1): 2.
- Arnold, L. 2001. *Stochastic climate models*, 141–157. New York: Springer.
- Arnold, L., P. Imkeller, and Y. Wu. 2003. *Dynamical Systems* 18(4): 295.
- Arnold, H., I. Moroz, and T. Palmer. 2013. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 371(1991): 20110479.
- Bogoliubov, N.N., and Y.A. Mitropolski. 1961. In *Asymptotic methods in the theory of non-linear oscillations*, ed. Bogoliubov, N.N., and Y.A. Mitropolski, vol. 1. New York: Gordon and Breach.
- Bouchet, F., T. Grafke, T. Tangarife, and E. Vanden-Eijnden. 2016. *Journal of Statistical Physics* 162(4): 793.
- Buizza, R., M. Miller, and T. Palmer. 1999. *Quarterly Journal of the Royal Meteorological Society* 125(560): 2887.
- Chekroun, M.D., H. Liu, and S. Wang. 2015. *Approximation of stochastic invariant manifolds: stochastic manifolds for nonlinear SPDEs I*. Cham: Springer.
- Crommelin, D., and E. Vanden-Eijnden. 2008. *Journal of the Atmospheric Sciences* 65(8): 2661.
- Culina, J., S. Kravtsov, A.H. Monahan. 2011. *Journal of the Atmospheric Sciences* 68(2): 284.
- De Cruz, L., J. Demaeyer, and S. Vannitsem. 2016. *Geoscientific Model Development* 9(8): 2793.
- Demaeyer, J., and S. Vannitsem. 2017. Stochastic parametrization of subgrid-scale processes in coupled ocean-atmosphere systems: benefits and limitations of response theory. *Quarterly Journal of the Royal Meteorological Society*, 143(703), 881–896.
- Doblas-Reyes, F., A. Weisheimer, M. Déqué, N. Keenlyside, M. McVean, J. Murphy, P. Rogel, D. Smith, and T. Palmer. 2009. *Quarterly Journal of the Royal Meteorological Society* 135(643): 1538.
- Ellis, R.S., and M.A. Pinsky. 1975. *Journal de Mathématiques Pures et Appliquées* 54(9): 125.
- Frankignoul, C. 1979. *Dynamics of Atmospheres and Oceans* 3(2–4): 465.
- Frankignoul, C., and K. Hasselmann. 1977. *Tellus* 29(4): 289.
- Frankignoul, C., and P. Müller. 1979. *Journal of Physical Oceanography* 9(1): 104.
- Franzke, C., A.J. Majda, and E. Vanden-Eijnden. 2005. *Journal of the Atmospheric Sciences* 62(6): 1722.
- Frederiksen, J.S. 1999. *Journal of the Atmospheric Sciences* 56(11): 1481.

- Frederiksen, J.S., and A.G. Davies. 1997. *Journal of the Atmospheric Sciences* 54(20): 2475.
- Freidlin, M.I., and A.D. Wentzell. 1984. *Random perturbations of dynamical systems*, 15–43. New York: Springer.
- Gardiner, C.W. 2009. *Handbook of stochastic methods*, 4th ed. Berlin: Springer.
- Ghil, M., M. Allen, M. Dettinger, K. Ide, D. Kondrashov, M. Mann, A.W. Robertson, A. Saunders, Y. Tian, F. Varadi, et al. 2002. *Reviews of Geophysics* 40(1): 3-1–3-41.
- Grad, H. 1969. *Transport Theory* 1: 269.
- Hansen, J.A., and C. Penland. 2006. *Monthly Weather Review* 134(10): 3006.
- Hasselmann, K. 1976. *Tellus* A28(6): 473–485.
- Kifer, Y. 2001. *Stochastic climate models*, 171–188. Berlin: Springer.
- Kifer, Y. 2003. *Stochastics and Dynamics* 3(02): 213.
- Kurtz, T.G. 1973. *Journal of Functional Analysis* 12(1): 55.
- Lemke, P. 1977. *Tellus* 29(5): 385.
- Lemke, P., E. Trinkl, and K. Hasselmann. 1980. *Journal of Physical Oceanography* 10(12): 2100.
- Lovejoy, S., and D. Schertzer. 2013. *The Weather and climate: emergent laws and multifractal cascades*. Cambridge: Cambridge University Press.
- Majda, A.J., I. Timofeyev, and E. Vanden Eijnden. 2001. *Communications on Pure and Applied Mathematics* 54(8): 891.
- Newman, M., P.D. Sardeshmukh, and C. Penland. 1997. *Journal of the Atmospheric Sciences* 54(3): 435.
- Nicolis, C. 1981. *Physics of solar variations*, 473–478. Berlin: Springer.
- Nicolis, C. 1982. *Tellus* 34(1): 1.
- Nicolis, C. 2003. *Journal of the Atmospheric Sciences* 60(17): 2208.
- Nicolis, C. 2004. *Journal of the Atmospheric Sciences* 61(14): 1740.
- Nicolis, C. 2005. *Quarterly Journal of the Royal Meteorological Society* 131(609): 2151.
- Nicolis, C., and G. Nicolis. 1981. *Tellus* 33(3): 225.
- Nicolis, G., and C. Nicolis. 2012. *Foundations of complex systems: emergence, information and prediction*. Singapore: World Scientific.
- Ohkitani, K., and S. Kida. 1992. *Physics of Fluids A: Fluid Dynamics (1989–1993)* 4(4): 794.
- Olbers, D. 2001. *Stochastic climate models*, 3–63. New York: Springer.
- Papanicolau, G.C. 1976. *Journal of Mathematics* 6(4): 653–674.
- Penland, C. 1989. *Monthly Weather Review* 117(10): 2165.
- Penland, C. 1996. *Physica D: Nonlinear Phenomena* 98(2): 534.
- Penland, C., and L. Matrosova. 1994. *Journal of Climate* 7(9): 1352.
- Ruelle, D. 1997. *Communications in Mathematical Physics* 187(1): 227.
- Ruelle, D. 2009. *Nonlinearity* 22(4): 855.
- Rüemelin, W. 1982. *SIAM Journal on Numerical Analysis* 19(3): 604.
- Sanders, J., F. Verhulst, and J. Murdock (eds.). 2007. *Averaging Methods in Nonlinear Dynamical Systems. Collection: Applied Mathematical Sciences*, vol. 59, 2nd ed. New York: Springer.
- Sardeshmukh, P.D., and C. Penland. 2015. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 25(3): 036410.
- Shutts, G. 2005. *Quarterly Journal of the Royal Meteorological Society* 131(612): 3079.
- Smith, L.M., and F. Waleffe. 1999. *Physics of Fluids* 11: 1608.
- Sura, P. 2013. *Extremes in a changing climate*, 181–222. Dordrecht: Springer.
- Sura, P., M. Newman, C. Penland, and P. Sardeshmukh. 2005. *Journal of the Atmospheric Sciences* 62(5): 1391.
- Vannitsem, S. 2014. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 372: 20130282.
- Vissio, G., and V. Lucarini. 2016. arXiv preprint arXiv:1612.07223.
- Waleffe, F. 1992. *Physics of Fluids A: Fluid Dynamics (1989–1993)* 4(2): 350.

- Wouters, J., and V. Lucarini. 2012. *Journal of Statistical Mechanics: Theory and Experiment* 2012(03): P03003.
- Wouters, J., and V. Lucarini. 2013. *Journal of Statistical Physics* 151(5): 850.
- Wouters, J., S.I. Dolaptchiev, V. Lucarini, and U. Achatz. 2016. *Nonlinear Processes in Geophysics* 23(6): 435.
- Young, L.S. 2002. *Journal of Statistical Physics* 108(5): 733.