

# Stochastic parameterization of subgrid-scale processes: A review of recent physically-based approaches

Jonathan Demaeyer\* and Stéphane Vannitsem

Royal Meteorological Institute of Belgium, Avenue Circulaire, 3, 1180 Brussels, Belgium

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 has a considerable impact on their performances.

## I. 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 [1]<sup>1</sup>. 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 [2, 3].

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 [4]. 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 [5, 6]. Subsequently, during the following decade, stochastic modeling for meteorology and climatology became an important research topic [2, 7–14] before falling into disuse in what has been described as a “lull” of work in this field [15]. However, during that period, the idea that correct parameterizations of sub-grid processes are important to improve climate and weather models gained popularity [16–18]. Stochastic parameterizations for the “turbulent” closure in 2-D large-eddy simulations on the sphere have also been considered [19, 20]. It led recently to the implementation of stochastic schemes to correct the model errors [21, 22] made in large numerical weather prediction (NWP) models [23, 24], improving the reliability of probabilistic forecasts and correcting partially their variability [25, 26]. The relation between multiplicative noise and the non-Gaussian character observed in some geophysical variables has also been considered [27, 28], as well as stochastic models for the climate extremes [29].

Since the beginning of the 21st 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 [30–32] and a new method based on the singular perturbation theory of Markov processes [33] 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 [15, 34–36]. 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 [31]. Such events display recurrence timescales that are few orders greater than the timescale of the climate variables considered, and thus induce transitions between different climates. The statistics of these transitions is then given by the so-called *Large Deviations* theory [37] (for recent developments on this matter, see Ref. [38]). In addition to these two methods, the modeling of the effects of subgrid scale through conditional Markov chain has been considered [39] and recently, new stochastic parameterization techniques have been proposed, based on an expansion of the backward Kolmogorov equation [40] and on the Ruelle response theory [41]. The latter has been tested on a simple coupled ocean-atmosphere model [42], on stochastic triads [43] and on an adapted version of the Lorenz’96 model [44].

---

\* E-mail: Jonathan.Demaeyer@meteo.be

<sup>1</sup> More recently, a new compartment have 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.

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 Section II, we will present the general framework in which the problem of stochastic parameterizations is posed. In Section III, 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 Section IV. Finally, the conclusions are given in Section V.

## II. THE PARAMETERIZATION PROBLEM

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

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

where  $Z \in \mathbb{R}^d$  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  $X \in \mathbb{R}^m$  and  $Y \in \mathbb{R}^n$ . In general, such a decomposition is made such that the subset  $X$  and  $Y$  have strongly differing *response times*  $\tau_Y \ll \tau_X$  [15], but we will assume here that this constraint is not necessarily met. System (1) can then be expressed as:

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

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 [15]. 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 remains 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  $\Xi$  between the two sub-systems:

$$Y = \Xi(X, t) \quad (3)$$

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 mathematician have considered it and contributed to what is known nowadays as the theory of averaging [45] and which forms the first step of the Hasselmann program [31]. The mathematical framework was set in the 1960s by the influential contribution of Bogoliubov and Mitropolski [46]. 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 Section III.

## III. 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, following the Hasselmann program, we shall focus on parameterizations that are defined in terms of stochastic processes. We will consider methods based

- on the Ruelle response theory [41–43].
- on the singular perturbation theory of Markov processes [33, 34].
- on the Hasselmann averaging methods [15, 32, 35, 36].
- on empirical methods [47].

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  $\sigma$ , 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.

### A. The method based on response theory

This method is based on the Ruelle response theory [48, 49] and was proposed by Wouters and Lucarini [41, 50]. 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”  $\Psi_X$ ,  $\Psi_Y$  to the invariant measure<sup>2</sup>  $\tilde{\rho}$  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  $\rho_0$  is the invariant measure of the uncoupled system which is also supposed to be an existing, well defined SRB measure. As shown in [41], 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  $\Psi_X$  and  $\Psi_Y$ :

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

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:

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

where  $\otimes$  is the outer product,  $\Psi'_X(X, Y) = \Psi_X(X, Y) - M_1(X)$  is the centered perturbation and  $\phi_X^s$ ,  $\phi_Y^s$  are the flow of the uncoupled system  $\dot{X} = F_X(X)$  and  $\dot{Y} = F_Y(Y)$ . The process  $\sigma_R$  is thus obtained by taking the square root of

---

<sup>2</sup> The theory assumes that for the system under consideration, a SRB measure [51] exists (e.g. an Axiom-A system).

the matrix  $\mathbf{g}$ , which is here accomplished by a decomposition of  $\Psi'_X$  on a proper basis [41]. 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 simplify the averaging procedure in many case. The terms  $M_1$ ,  $M_2$  and  $M_3$  are derived [41] such that their responses up to order two match the response of the perturbation  $\Psi_X$  and  $\Psi_Y$ . 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 [52]. However, the integral (10) in this memory term must only be evaluated from 0 up to the timescale  $\tau_Y$  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 [43].

## B. Singular perturbation theory method

Singular perturbation methods were developed in the 1970s for the analysis of the linear Boltzmann equation in some asymptotic limit [33, 53–55]. Here, these methods are applicable if the problem can be cast into a Fokker-Planck equation. The procedure described in Ref. [33] requires assumptions on the timescales of the different terms of system (4). In term of the small parameter  $\delta = \tau_Y/\tau_X$  defined in Section II, the fast variability of the unresolved component  $Y$  are 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  $\delta$  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 [33]:

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

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  $\rho^\delta$  can be expanded in term of  $\delta$  and inserted in Eq. (12). The zeroth order of this equation  $\rho^0$  can be shown to be independent of  $Y$  and its evolution given by a closed, averaged backward Fokker-Planck equation [56]:

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

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  $\delta$  in the original equation (4). Since  $\delta$  disappears from Eq. (13), one can simply use the parameter to identify the fast terms to be considered, and eventually consider  $\delta = 1$  [34].

The parameterization obtained by this procedure is given by [34]:

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

with

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

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

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 replaced them by Ornstein-Uhlenbeck processes whose measures is used to compute the averages [33, 34].

Finally, if one assume 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  $\Psi_X$  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.

### C. Hasselmann averaging method

Since the initial work of Hasselmann in the 1970s [4], 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 [32] and applied successfully to idealized geophysical models [15] with non-trivial invariant measures. In the same vein, an approximation has been proposed in Ref. [57] 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 Ref. [57], an approach based on the fluctuation-dissipation theorem was 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 has been derived in Refs. [15, 40] 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 [15]). Such a requirement may not be always fulfilled, like for instance if the fast system conditional on the state  $X$  is unstable and does not possess any attractor (see Section IV for an example).

## D. 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 Section I but we will consider here a method based on state dependent AR(1) processes proposed in Ref. [47]. In this case, a timeserie  $r(t)$  of the coupling part  $\Psi_X$  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  $\phi$  and the state dependent standard deviation  $\sigma_e(X)$  which is modeled as  $\sigma_e(X) = \sigma_0 + \sigma_1 |X|$  with the parameters  $\sigma_0$  and  $\sigma_1$  given by a binning procedure. The parameter  $\Delta t$  is the time step of integration of Eq. (24). Other empirical parameterizations have been proposed by the authors of Ref. [47], notably one with the function  $\mathcal{U}(X) = (1 + e(t))\mathcal{U}_{\text{det}}(X)$  which resembles the SPPT<sup>3</sup> parameterization used in the ECMWF<sup>4</sup> Numerical Weather Prediction model [23]. However, the study shows no substantial differences with the parameterization (25).

## IV. APPLICATIONS AND RESULTS

In this section, we will illustrate the various parameterizations described in Section III to the following example

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

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

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

with  $X, y_1, y_2 \in \mathbb{R}$ . The matrices involved are defined as

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

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  $\delta > 0$  quantify the time-scale 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 Ref. [33] (see also Subsection III B). Additionally, the parameter  $\varepsilon > 0$  controls the coupling strength between the two sub-systems. In this setup the coupling are thus proportional to the ratio  $\varepsilon/\delta$ , and therefore the characterization of the coupling as “weak” depends directly on the time-scale separation.

The deterministic part of Eq. (27) is a well known model called a triad and encountered in fluid dynamics [58–60], and in simplified geophysical flows e.g. [33, 43]. Due to the presence of invariant manifolds, its mathematical structure can be found in higher-order model. See Ref. [42] for an example of such structure in the framework of a coupled

<sup>3</sup> Acronym for *Stochastically Perturbed Parameterization Tendencies Scheme*.

<sup>4</sup> Acronym for *European Center for Medium-Range Weather Forecasts*.

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<sup>5</sup>.

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. With the dissipation and stochastic terms set to zero, the 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 [33, 60]

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

It allows for the following configurations of their signs :  $(+, -, -)$ ,  $(+, +, -)$ ,  $(+, -, +)$ ,  $(-, +, +)$ ,  $(-, +, -)$ ,  $(-, -, +)$ . These different configurations are associated with different kinds of energy exchange scenario and different stability properties [58].

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

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

and consider various values of the parameters  $\delta$  and  $\varepsilon$ . 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 version have been integrated over  $4.5 \times 10^5$  timeunits with a timestep  $\Delta t = 0.01$  after a transient period of  $5.0 \times 10^4$  timeunits to let the system relax to its stationary state. The state  $X$  has been recorded every 0.1 timeunit, giving a dataset of  $4.5 \times 10^6$  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 [61, 62]. The 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 [43] 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 [25]. 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 [43, 47]. 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 [34, 39, 40, 57, 63]. 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 the proposed methods and consider first the different measures used for averaging in system (27).

### A. Stability and measures

All the ingredients needed to compute the parameterizations presented in Section III 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

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

---

<sup>5</sup> Except for the empirical methods which by definition use this kind of procedures.

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 [43]

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

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

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

In order for these processes to be stable, the real part of the eigenvalues of the matrix  $\mathbf{T}$  must be negative [64] 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  $\beta \in \mathbb{R}$ ). On the other hand, the system  $\dot{Y} = H(X, Y)$  has the eigenvalues  $\lambda_\pm = (-a \pm \sqrt{\Delta(X)})/\delta^2$  with

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

Therefore, if

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

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 null one. Consequently, equation (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  $\varepsilon$  and  $\delta$  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  $\varepsilon$  and  $\delta$  parameters considered (see Subsection IV C 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 [58].

For the interested reader, the exact calculation of the parameterization of Section III using the covariance and correlation matrices are detailed in the appendix (see Section A).

## B. 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  $\varepsilon$ :  $0.05$ ,  $0.125$  and  $0.4$ . The probability densities associated with these different systems are represented on Figs. 1 and 3. 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 [47]

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

between the distribution  $P$  of the full coupled system and the distribution  $Q$  of the parameterizations is worth computing to quantify the differences. This distance quantifies the similarity between the densities. It is depicted for  $\delta = 0.1$  on Fig. (2), 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  $\varepsilon$ , it is the Hasselmann averaging method which performs best. The empiric 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  $\mathcal{U}_{\text{det}}(X)$  or to the inadequacy of the fitting function  $\sigma_0 + \sigma_1 |X|$  for the standard deviation  $\sigma_\varepsilon(X)$  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 an decreased timescale separation  $\delta = 0.4$ , the same conclusions are reached as for  $\delta = 0.1$ , 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  $\delta \rightarrow 0$ .

### C. 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 on Figs. 5 and 7, and the Hellinger distances are shown on Figs. 6 and 8. First, we must remark that the parameterization based on the Hasselmann's averaging method is not defined for most of the  $\delta$  and  $\varepsilon$  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 Subsection IV A. 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  $\delta$  and  $\varepsilon$ , 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.

### D. Discussion

The results obtained so far with these two types of triads highlight the utility of the parameterization schemes discussed here. First, the empiric 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 gives very good results. This difference is quite intriguing and interesting. It indicates that different physically based parameterization should be considered depending on the kind of problems encountered. In particular, the stability properties of the system considered seems 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. In extenso, 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.

## V. CONCLUSIONS

The parameterization of subgrid-scale processes is an important tool available for meteorologist and climatologist, for stochastic modeling, model reduction, or to improve the statistical properties of the forecasting systems. The variety of different approaches available bear witness of the richness of the field but at the same time can also lead to

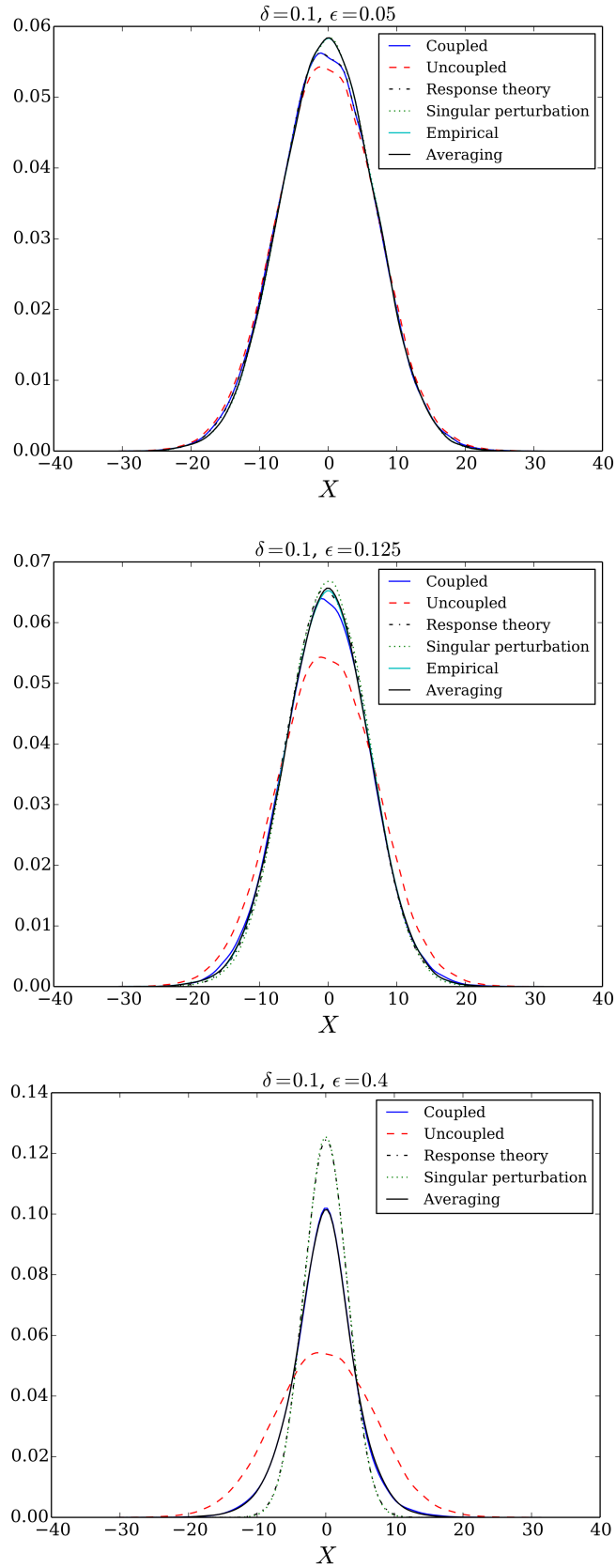


Figure 1: Probability densities of the full coupled dynamics (27), the uncoupled dynamics  $\dot{X} = -DX + q\xi(t)$  and the parameterized model versions for the timescale separation  $\delta = 0.1$  and for the triad parameters of case 1. The empirical parameterization density is not represented for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

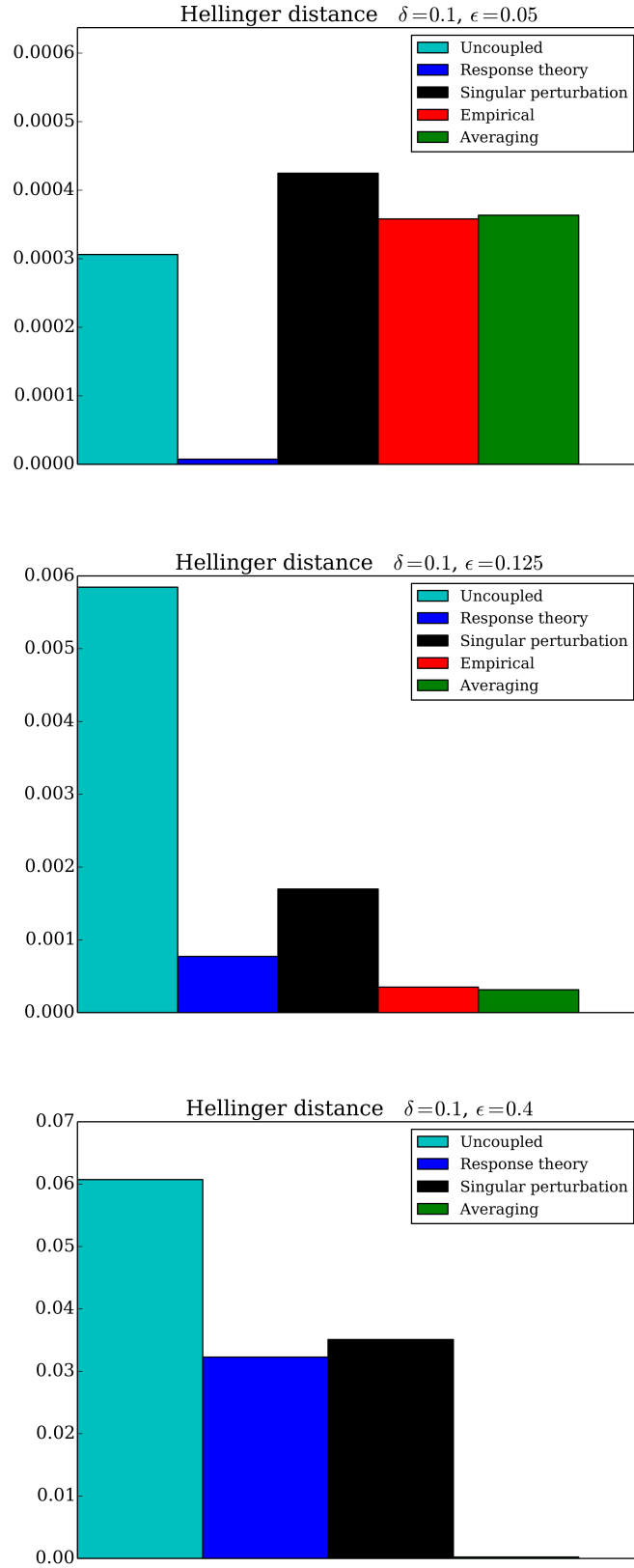


Figure 2: Hellinger distance (37) between the densities of the different parameterized models and the full coupled system density for case 1. A small distance indicates that the two densities concerned are very similar. The Hellinger distance between the full coupled system and the uncoupled system distribution is depicted as reference. In case  $\epsilon = 0.4$ , the empirical parameterization is not represented due to its divergence.

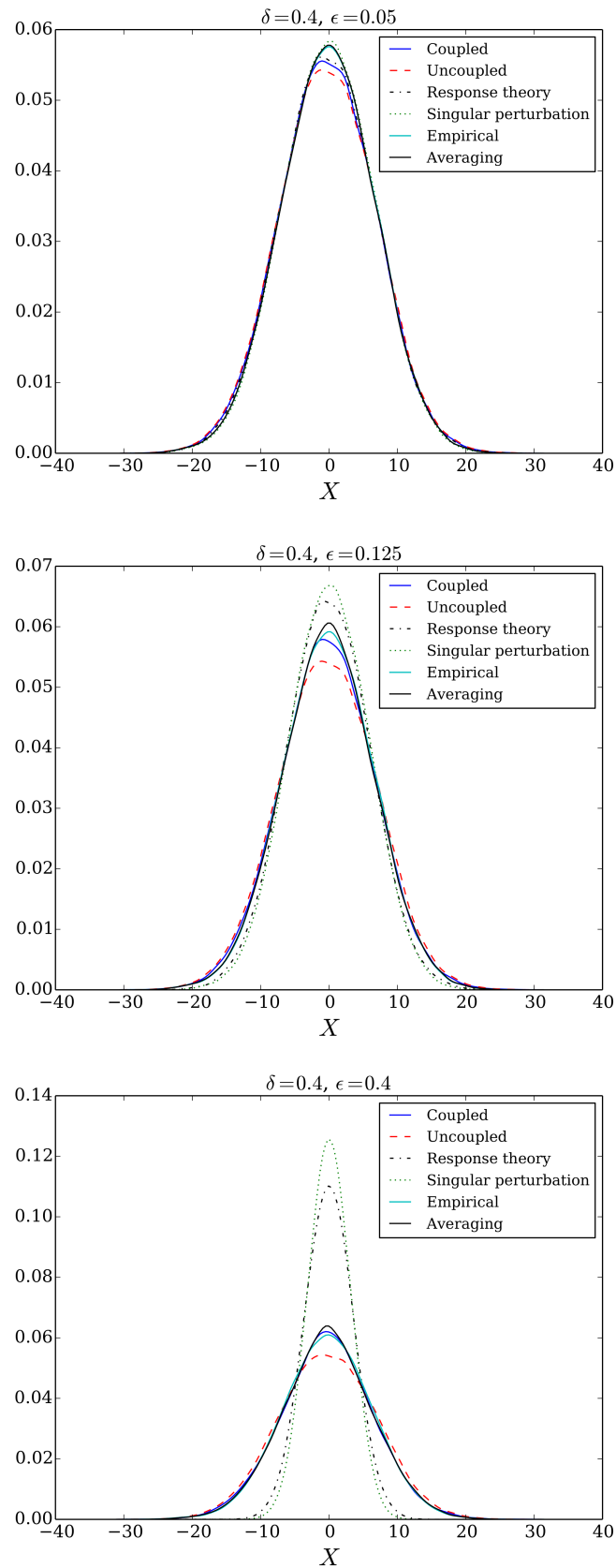


Figure 3: Same as Fig. 1 but for the timescale separation  $\delta = 0.4$ .

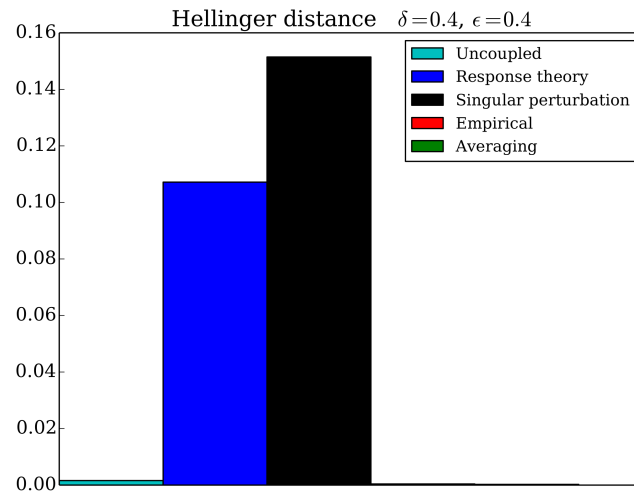
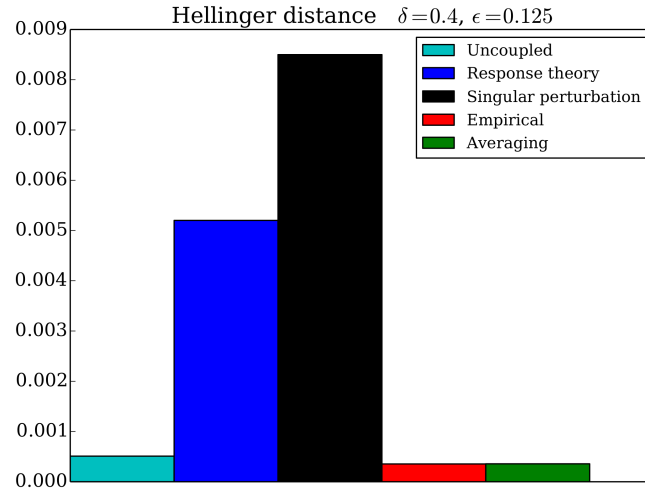
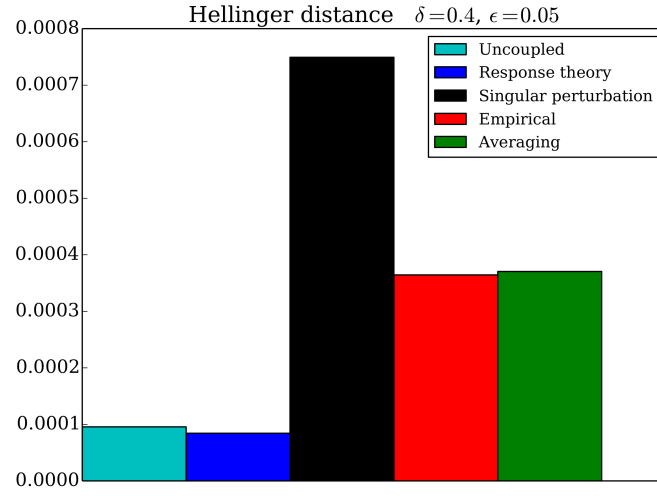


Figure 4: Same as Fig. 2 but for the timescale separation  $\delta = 0.4$ .

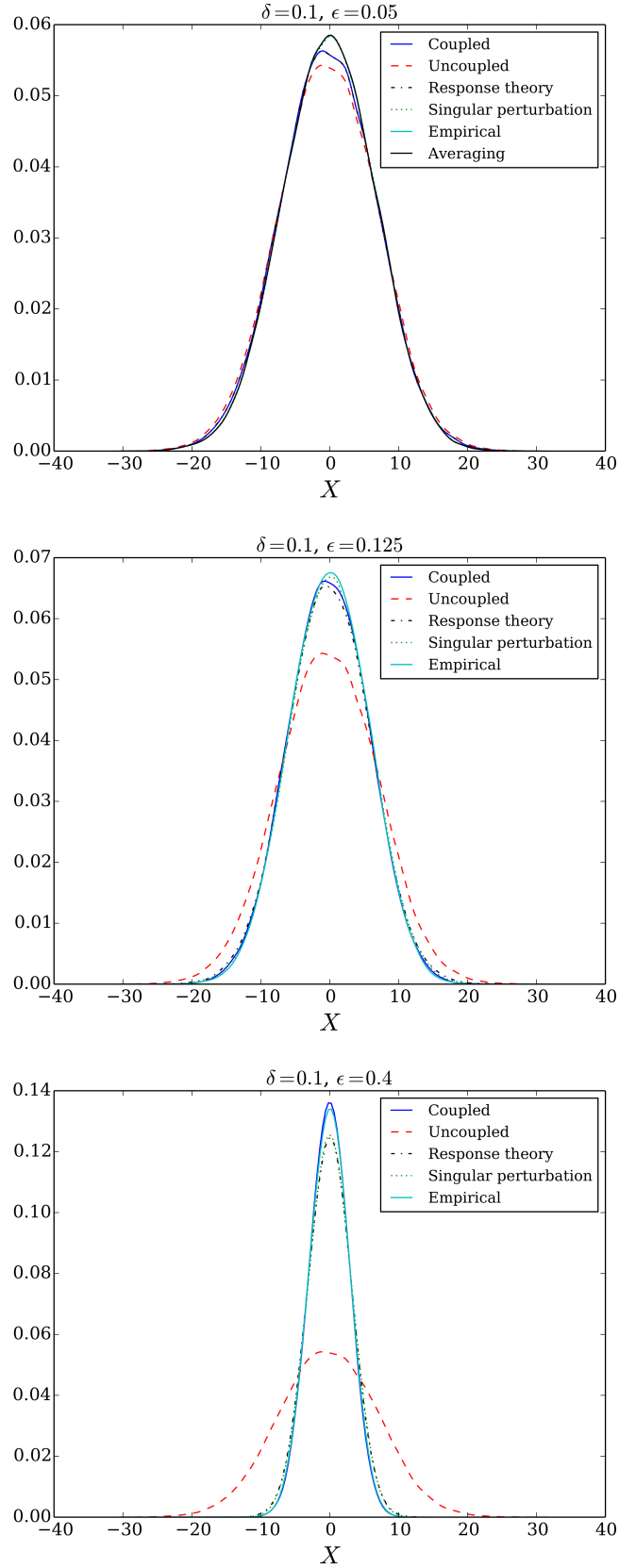


Figure 5: Probability densities of the coupled full dynamics (27), the uncoupled dynamics  $\dot{X} = -DX + q\xi(t)$  and the parameterized models for the timescale separation  $\delta = 0.1$  and for the triad parameters of case 2. The direct averaging parameterization density is only represented for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

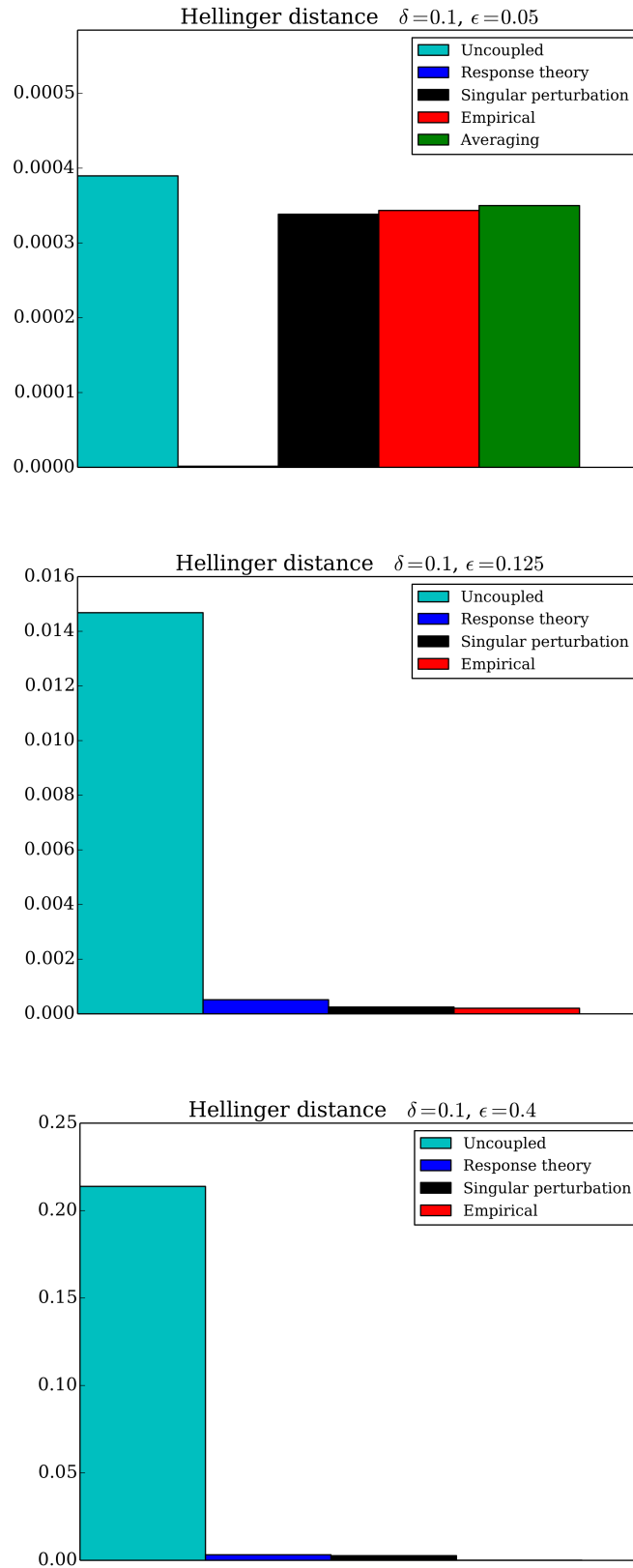


Figure 6: Hellinger distance (37) between the densities of the different parameterized models and the full coupled system density for case 2. A small distance indicates that the two densities concerned are very similar. The Hellinger distance between the full coupled system and the uncoupled system distribution is depicted as reference. In case  $\epsilon = 0.4$ , the empirical parameterization is not represented due to its divergence.

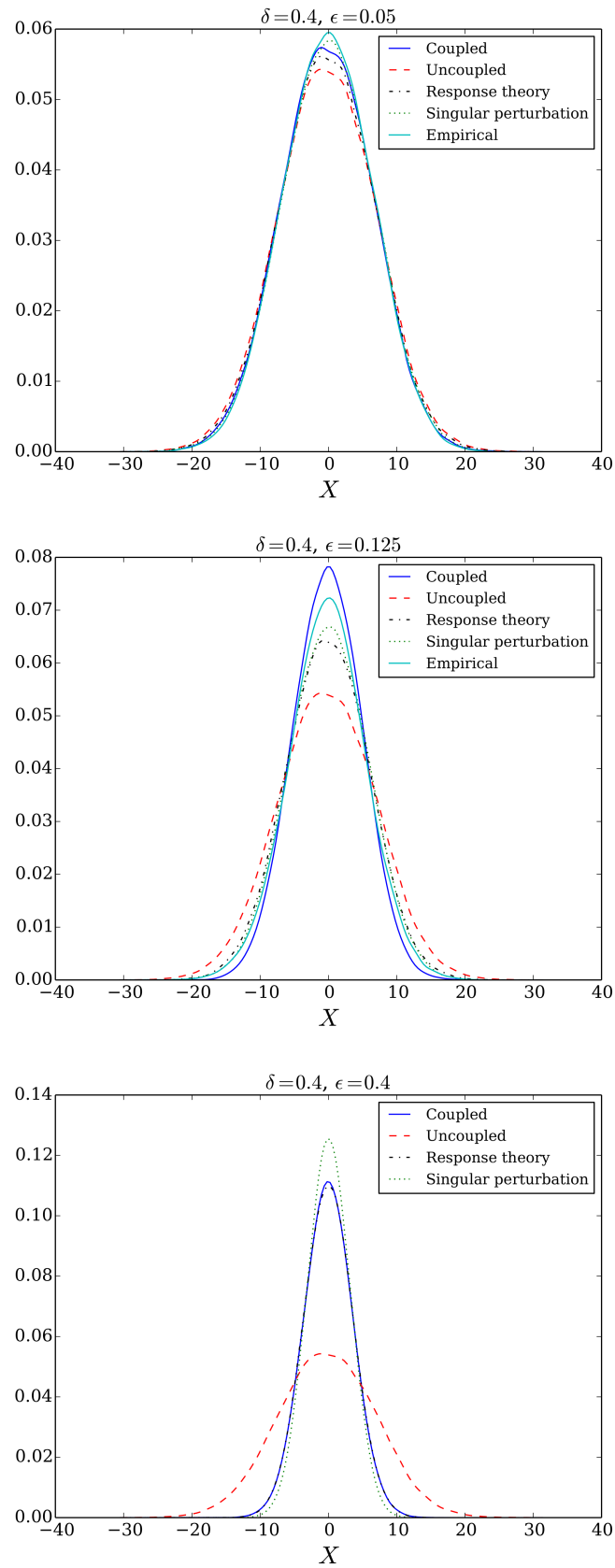


Figure 7: Same as Fig. 5 but for the timescale separation  $\delta = 0.4$ .



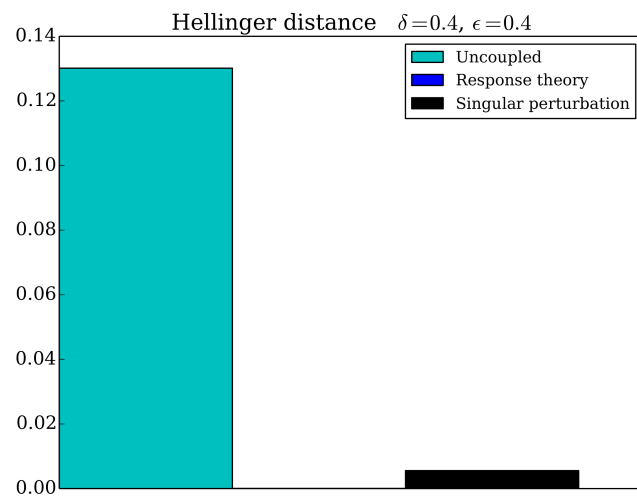
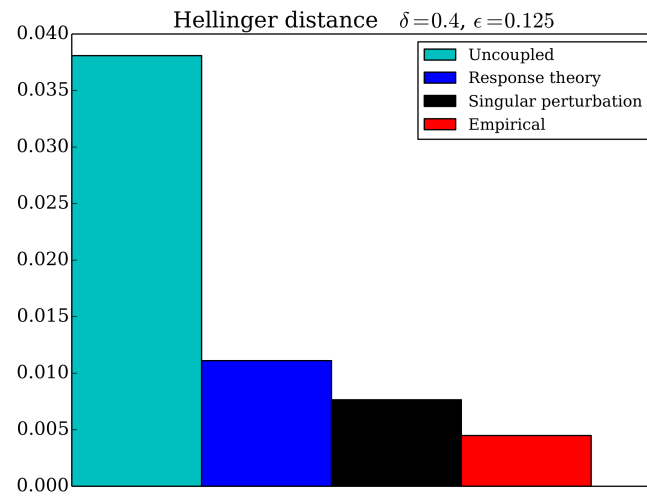
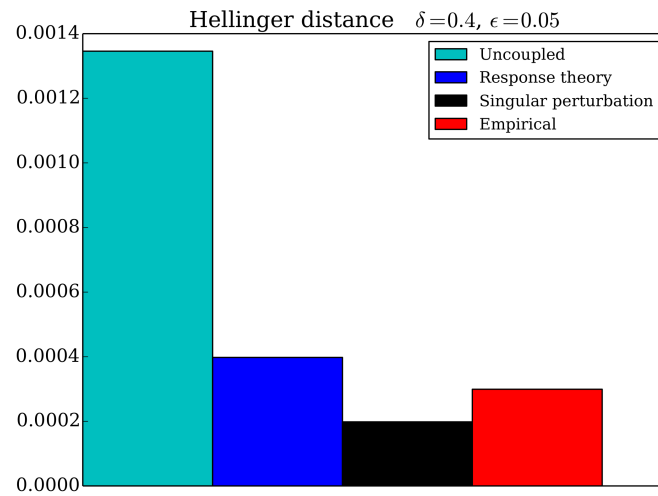


Figure 8: Same as Fig. 6 but for the timescale separation  $\delta = 0.4$ .

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 included perturbative methods like the Ruelle response theory [41] or the singular perturbation theory [33], averaging methods like the Hasselmann method [4, 15] or empirical methods [47]. As expected, these parameterizations provided good 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 control the behavior and performance of model error parameterizations [25].

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 sub-grid scale processes on their performances. This type of analysis is currently under way in the context of a coupled ocean-atmosphere system [65].

## Appendix A: 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.

### 1. 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:

$$\dot{X} = -D X + q \xi(t) + M_1(X) + M_2(X, t) + M_3(X, t) \quad (\text{A1})$$

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

$$\rho_{0,Y}(Y) = \frac{1}{\mathcal{Z}} \exp\left(-\frac{1}{2} Y^\top \cdot \Sigma^{-1} \cdot Y\right) \quad (\text{A2})$$

where  $\Sigma$  is the covariance matrix solution of

$$\mathbf{A} \cdot \Sigma + \Sigma \cdot \mathbf{A}^\top = -\mathbf{B}_Y \cdot \mathbf{B}_Y^\top \quad (\text{A3})$$

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 [64]:

$$\Sigma = \langle Y \otimes Y \rangle = \frac{q_Y^2}{2a} \mathbf{I} \quad (\text{A4})$$

$$\langle \phi_Y^t(Y) \otimes \phi_Y^s(Y) \rangle = \mathbf{E}(t-s) \cdot \Sigma \quad , \quad t > s \quad (\text{A5})$$

$$\langle \phi_Y^t(Y) \otimes \phi_Y^s(Y) \rangle = \Sigma \cdot \mathbf{E}(s-t)^\top \quad , \quad t < s \quad (\text{A6})$$

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

$$\mathbf{E}(t) = \exp(\mathbf{A}t/\delta^2) = e^{-at/\delta^2} \begin{bmatrix} \cos(\beta t/\delta^2) & \sin(\beta t/\delta^2) \\ -\sin(\beta t/\delta^2) & \cos(\beta t/\delta^2) \end{bmatrix} \quad (\text{A7})$$

The various terms  $M_i$  are then computed as follow.

#### a. The term $M_1$

It is the average term:

$$M_1(X) = \langle \Psi_X(X, Y) \rangle \quad (\text{A8})$$

We have thus

$$M_1(X) = 2B \frac{\varepsilon}{\delta} \langle y_1 y_2 \rangle = 2B \frac{\varepsilon}{\delta} \Sigma_{12} = 0 \quad (\text{A9})$$

by using Eq. (A4).

b. The term  $M_2$

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

$$M_2(t) = \sigma_R(t) \quad (\text{A10})$$

with

$$\langle \sigma_R(t) \sigma_R(t') \rangle = g(t - t') \quad (\text{A11})$$

and the correlation function

$$g(s) = \langle \Psi'_X(Y) \Psi'_X(\phi_Y^s(Y)) \rangle \quad (\text{A12})$$

where  $\Psi'_X(Y) = \Psi_X(Y) - M_1$ . The result in the present case is given by the formula (see Ref. [42]) :

$$\mathbf{g}(s) = \frac{\varepsilon^2}{\delta^2} \text{Tr} \left( (\mathbf{C} + \mathbf{C}^\top) \cdot \boldsymbol{\Sigma} \cdot \mathbf{E}(s)^\top \cdot \mathbf{C}^\top \cdot \mathbf{E}(s) \cdot \boldsymbol{\Sigma} \right) = \frac{\varepsilon^2}{\delta^2} \frac{q_Y^4}{a^2} B^2 e^{-2as/\delta^2} \cos(2\beta s/\delta^2) \quad (\text{A13})$$

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

c. The term  $M_3$

This is the memory term, defined by

$$M_3(X, t) = \int_0^\infty ds h(X(t-s), s) \quad (\text{A14})$$

with the memory kernel

$$h(X, s) = \langle \Psi_Y(X, Y) \cdot \nabla_Y \Psi_X(Y^s) \rangle \quad (\text{A15})$$

which in the present case is given by the formula (see Ref. [42])

$$h(X, s) = \frac{\varepsilon^2}{\delta^2} \text{Tr} \left( (X \mathbf{V} \cdot \boldsymbol{\Sigma}) \cdot (\mathbf{E}(s)^\top \cdot (\mathbf{C} + \mathbf{C}^\top) \cdot \mathbf{E}(s)) \right) \quad (\text{A16})$$

$$= \frac{\varepsilon^2}{\delta^2} \frac{q_Y^2}{a} X B (B_1 + B_2) e^{-2as/\delta^2} \cos(2\beta s/\delta^2) \quad (\text{A17})$$

The fact that the memory kernel (A17) and the correlation function (A13) present the same form imply that a Markovian parameterization is available [43] even if by definition, Eq. (A1) is a non-Markovian parameterization.

## 2. The singular perturbation method

With this parameterization, the parameter  $\delta$  serves to distinguish terms with different timescale and is then used as a small perturbation parameter [33, 34]. The parameterization is given by:

$$\dot{X} = -D X + q \xi(t) + G(X) + \sqrt{2} \sigma_{\text{MTV}}(X) \cdot \tilde{\xi}(t) \quad (\text{A18})$$

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

$$G(X) = \int_0^\infty ds \langle \Psi_Y(X, Y) \cdot \nabla_Y \Psi_X(X, \phi_Y^s(Y)) \rangle_{\bar{\rho}} \quad (\text{A19})$$

$$\sigma_{\text{MTV}}(X) = \left( \int_0^\infty ds \langle \Psi'_X(X, Y) \Psi'_X(X, \phi_Y^s(Y)) \rangle_{\bar{\rho}} \right)^{1/2} \quad (\text{A20})$$

We see that the quantities appearing in this parameterization can easily be obtained from the functions  $h$  and  $g$  of the section A 1. Indeed we have,

$$G(X) = \int_0^\infty ds h(X, s) = \varepsilon^2 X q_Y^2 \frac{B(B_1 + B_2)}{2(a^2 + \beta^2)} \quad (\text{A21})$$

$$\mathbf{S}_{\text{MTV}}(X) = \int_0^\infty ds g(s) = \varepsilon^2 \frac{q_Y^4 B^2}{2a(a^2 + \beta^2)} \quad (\text{A22})$$

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

### 3. Averaging method

In this approach, we consider the system (2) and the parameterization from Ref. [57]:

$$\dot{X} = \bar{F}(X) \quad (\text{A23})$$

with

$$\bar{F}(X) = \langle F(X, Y) \rangle_{\rho_{Y|X}} = F(X, \bar{Y}(X)) + \frac{1}{2} \frac{\partial^2 F}{\partial Y^2}(X, \bar{Y}(X)) : \Sigma(X) \quad (\text{A24})$$

and

$$\bar{Y}(X) = \langle Y \rangle_{\rho_{Y|X}} \quad (\text{A25})$$

$$\Sigma(X) = \langle (Y - \bar{Y}(X)) \otimes (Y - \bar{Y}(X)) \rangle_{\rho_{Y|X}} \quad (\text{A26})$$

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

$$\rho_{0,Y}(Y) = \frac{1}{\mathcal{Z}} \exp\left(-\frac{1}{2} Y^\top \cdot \Sigma^{-1}(X) \cdot Y\right) \quad (\text{A27})$$

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

$$\mathbf{T}(X) \cdot \Sigma + \Sigma \cdot \mathbf{T}(X)^\top = -\frac{1}{\delta^2} \mathbf{B}_Y \cdot \mathbf{B}_Y^\top \quad (\text{A28})$$

with

$$\mathbf{T}(X) = \mathbf{A}/\delta^2 + \varepsilon X \mathbf{V}/\delta \quad (\text{A29})$$

With the help of  $\bar{Y}(X) = 0$  and  $\Sigma(X)$ , we can now rewrite Eq. (A24) as

$$\begin{aligned} \bar{F}(X) &= F(X, 0) + \frac{\varepsilon}{\delta} \mathbf{C} : \Sigma(X) \\ &= -D X + q\xi(t) + \frac{B(B_1 + B_2) q_Y^2 X \varepsilon^2}{2(a^2 + \beta^2 - X\beta\delta\varepsilon B_2 + X\delta\varepsilon B_1(\beta - X\delta\varepsilon B_2))} \end{aligned} \quad (\text{A30})$$

This forms a deterministic averaging parameterization. It can be extended into a stochastic parameterization [40] as follow:

$$\dot{X} = \bar{F}(X) + \sigma_A(X) \cdot \xi(t) \quad (\text{A31})$$

with

$$\sigma_A(X) = \sqrt{S(X)} \quad (\text{A32})$$

and

$$S(X) = 2 \int_0^\infty ds \langle (F(X, Y^s(X)) - \bar{F}(X)) \otimes (F(X, Y) - \bar{F}(X)) \rangle_{\rho_{Y|X}} \quad (\text{A33})$$

We thus have

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

where we have extended the result of Eq. (A13) to the stationary Ornstein-Uhlenbeck process  $\dot{Y} = H(X, Y)$  for  $X$  “frozen”. The function  $S(X)$  can be computed analytically using mathematical softwares but is a very complicate function that is not worth displaying in this review. This can however be provided upon query to Dr. J. Demaeyer.

---

[1] D. Olbers, in *Stochastic Climate Models* (Springer, 2001) pp. 3–63.

- [2] C. Nicolis and G. Nicolis, *Tellus* **33**, 225 (1981).
- [3] G. Nicolis and C. Nicolis, *Foundations of complex systems: emergence, information and prediction* (World Scientific, 2012).
- [4] K. Hasselmann, *Tellus A* **28** (1976).
- [5] M. Ghil, M. Allen, M. Dettinger, K. Ide, D. Kondrashov, M. Mann, A. W. Robertson, A. Saunders, Y. Tian, F. Varadi, *et al.*, *Reviews of geophysics* **40** (2002).
- [6] S. Lovejoy and D. Schertzer, *The weather and climate: emergent laws and multifractal cascades* (Cambridge University Press, 2013).
- [7] P. Lemke, *Tellus* **29**, 385 (1977).
- [8] C. Frankignoul and K. Hasselmann, *Tellus* **29**, 289 (1977).
- [9] C. Frankignoul, *Dynamics of Atmospheres and Oceans* **3**, 465 (1979).
- [10] C. Frankignoul and P. Müller, *Journal of Physical Oceanography* **9**, 104 (1979).
- [11] P. Lemke, E. Trinkl, and K. Hasselmann, *Journal of Physical Oceanography* **10**, 2100 (1980).
- [12] C. Nicolis, in *Physics of Solar Variations* (Springer, 1981) pp. 473–478.
- [13] C. Nicolis, *Tellus* **34**, 1 (1982).
- [14] C. Penland, *Monthly Weather Review* **117**, 2165 (1989).
- [15] L. Arnold, P. Inkeller, and Y. Wu, *Dynamical Systems* **18**, 295 (2003).
- [16] C. Penland and L. Matrosova, *Journal of climate* **7**, 1352 (1994).
- [17] C. Penland, *Physica D: Nonlinear Phenomena* **98**, 534 (1996).
- [18] M. Newman, P. D. Sardeshmukh, and C. Penland, *Journal of the atmospheric sciences* **54**, 435 (1997).
- [19] J. S. Frederiksen and A. G. Davies, *Journal of the atmospheric sciences* **54**, 2475 (1997).
- [20] J. S. Frederiksen, *Journal of the atmospheric sciences* **56**, 1481 (1999).
- [21] C. Nicolis, *Journal of the Atmospheric Sciences* **60**, 2208 (2003).
- [22] C. Nicolis, *Journal of the Atmospheric Sciences* **61**, 1740 (2004).
- [23] R. Buizza, M. Milleer, and T. Palmer, *Quarterly Journal of the Royal Meteorological Society* **125**, 2887 (1999).
- [24] G. Shutts, *Quarterly Journal of the Royal Meteorological Society* **131**, 3079 (2005).
- [25] C. Nicolis, *Quarterly Journal of the Royal Meteorological Society* **131**, 2151 (2005).
- [26] F. Doblas-Reyes, A. Weisheimer, M. Déqué, N. Keenlyside, M. McVean, J. Murphy, P. Rogel, D. Smith, and T. Palmer, *Quarterly Journal of the Royal Meteorological Society* **135**, 1538 (2009).
- [27] P. Sura, M. Newman, C. Penland, and P. Sardeshmukh, *Journal of the atmospheric sciences* **62**, 1391 (2005).
- [28] P. D. Sardeshmukh and C. Penland, *Chaos: An Interdisciplinary Journal of Nonlinear Science* **25**, 036410 (2015).
- [29] P. Sura, in *Extremes in a Changing Climate* (Springer, 2013) pp. 181–222.
- [30] Y. Kifer, in *Stochastic Climate Models* (Springer, 2001) pp. 171–188.
- [31] L. Arnold, in *Stochastic climate models* (Springer, 2001) pp. 141–157.
- [32] Y. Kifer, *Stochastics and Dynamics* **3**, 213 (2003).
- [33] A. J. Majda, I. Timofeyev, and E. Vanden Eijnden, *Communications on Pure and Applied Mathematics* **54**, 891 (2001).
- [34] C. Franzke, A. J. Majda, and E. Vanden-Eijnden, *Journal of the atmospheric sciences* **62**, 1722 (2005).
- [35] J. Culina, S. Kravtsov, and A. H. Monahan, *Journal of the Atmospheric Sciences* **68**, 284 (2011).
- [36] S. Vannitsem, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **372**, 20130282 (2014).
- [37] M. I. Freidlin and A. D. Wentzell, in *Random Perturbations of Dynamical Systems* (Springer, 1984) pp. 15–43.
- [38] F. Bouchet, T. Grafke, T. Tangarife, and E. Vanden-Eijnden, *Journal of Statistical Physics* **162**, 793 (2016).
- [39] D. Crommelin and E. Vanden-Eijnden, *Journal of the Atmospheric Sciences* **65**, 2661 (2008).
- [40] R. Abramov, *Fluids* **1**, 2 (2015).
- [41] J. Wouters and V. Lucarini, *Journal of Statistical Mechanics: Theory and Experiment* **2012**, P03003 (2012).
- [42] J. Demaeyer and S. Vannitsem, *Quarterly Journal of the Royal Meteorological Society*, in press (2016), [10.1002/qj.2973](https://doi.org/10.1002/qj.2973).
- [43] J. Wouters, S. I. Dolaptchiev, V. Lucarini, and U. Achatz, *Nonlinear Processes in Geophysics* **23**, 435 (2016).
- [44] G. Vissio and V. Lucarini, arXiv preprint arXiv:1612.07223 (2016).
- [45] J. Sanders and F. Verhulst, *Applied Mathematical Sciences* **59** (1985).
- [46] N. N. Bogoliubov and Y. A. Mitropolski, *Asymptotic Methods in the Theory of Non-Linear Oscillations*, by NN Bogoliubov and YA Mitropolski. New York: Gordon and Breach, 1961. **1** (1961).
- [47] H. Arnold, I. Moroz, and T. Palmer, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **371**, 20110479 (2013).
- [48] D. Ruelle, *Communications in Mathematical Physics* **187**, 227 (1997).
- [49] D. Ruelle, *Nonlinearity* **22**, 855 (2009).
- [50] J. Wouters and V. Lucarini, *Journal of Statistical Physics* **151**, 850 (2013).
- [51] L.-S. Young, *Journal of Statistical Physics* **108**, 733 (2002).
- [52] M. D. Chekroun, H. Liu, and S. Wang, *Approximation of Stochastic Invariant Manifolds: Stochastic Manifolds for Nonlinear SPDEs I* (Springer, 2015).
- [53] H. Grad, *Transport theory* **1**, 269 (1969).
- [54] R. S. Ellis and M. A. Pinsky, *J. Math. Pures Appl* **54**, 125 (1975).
- [55] G. C. Papanicolaou, *Journal of Mathematics* **6** (1976).
- [56] T. G. Kurtz, *Journal of Functional Analysis* **12**, 55 (1973).
- [57] R. V. Abramov, *Multiscale Modeling & Simulation* **11**, 134 (2013).

- [58] F. Waleffe, *Physics of Fluids A: Fluid Dynamics* (1989-1993) **4**, 350 (1992).
- [59] K. Ohkitani and S. Kida, *Physics of Fluids A: Fluid Dynamics* (1989-1993) **4**, 794 (1992).
- [60] L. M. Smith and F. Waleffe, *Physics of fluids* **11**, 1608 (1999).
- [61] J. A. Hansen and C. Penland, *Monthly weather review* **134**, 3006 (2006).
- [62] W. Rümelin, *SIAM Journal on Numerical Analysis* **19**, 604 (1982).
- [63] R. V. Abramov, *Multiscale Modeling & Simulation* **10**, 28 (2012).
- [64] C. W. Gardiner, *Handbook of stochastic methods*, 4th ed. (Springer Berlin, 2009).
- [65] L. De Cruz, J. Demaeyer, and S. Vannitsem, *Geoscientific Model Development* **9**, 2793 (2016).