Disentangling multi-level systems: averaging, correlations and memory

Jeroen Wouters and Valerio Lucarini
Meteorologisches Institut, University of Hamburg, KlimaCampus, Grindelberg 7, 20144 Hamburg, Germany
E-mail: jeroen.wouters@zmaw.de and valerio.lucarini@zmaw.de

Received 28 October 2011
Accepted 7 February 2012
Published 6 March 2012

Online at stacks.iop.org/JSTAT/2012/P03003
doi:10.1088/1742-5468/2012/03/P03003

Abstract. We consider two weakly coupled systems and adopt a perturbative approach based on the Ruelle response theory to study their interaction. We propose a systematic way of parameterizing the effect of the coupling as a function of only the variables of a system of interest. Our focus is on describing the impacts of the coupling on the long term statistics rather than on the finite-time behavior. By direct calculation, we find that, at first order, the coupling can be surrogated by adding a deterministic perturbation to the autonomous dynamics of the system of interest. At second order, there are additionally two separate and very different contributions. One is a term taking into account the second-order contributions of the fluctuations in the coupling, which can be parameterized as a stochastic forcing with given spectral properties. The other one is a memory term, coupling the system of interest to its previous history, through the correlations of the second system. If these correlations are known, this effect can be implemented as a perturbation with memory on the single system. In order to treat this case, we present an extension to Ruelle’s response theory able to deal with integral operators. We discuss our results in the context of other methods previously proposed for disentangling the dynamics of two coupled systems. We emphasize that our results do not rely on assuming a time scale separation, and, if such a separation exists, can be used equally well to study the statistics of the slow variables and that of the fast variables. By recursively applying the technique proposed here, we can treat the general case of multi-level systems.

Keywords: driven diffusive systems (theory), coarse-graining (theory), stochastic processes (theory), stationary states

ArXiv ePrint: 1110.6113
1. Introduction

The dynamics of complex systems can often be partitioned in such a way as to highlight a multi-level structure, where each level consists of a subset of variables that interact strongly with each other and are weakly coupled with variables belonging to other levels of the system. Often, the dynamics of each level takes place on distinct spatial and temporal scales, so it is possible to hierarchically arrange the variables from the most to the least relevant, on the basis of the properties of interest. In practical applications, the observation or the simulation at high resolution of processes occurring on small temporal and spatial scales is very challenging. It is crucial to devise strategies for accounting, at least in an approximate way, for the impact of the fast, almost inaccessible variables on the slow variables, which are more usually the main object of investigation. In the language of dynamical systems theory, the overarching goal is to define rigorously or empirically an effective dynamics for the slow manifold of the system able to take into account the influence of the, enslaved, fast manifold. Otherwise, the presence of such very different scales make these systems hard to simulate, so they are often referred to as stiff problems [9], and hard to observe in practical terms, because no overall optimal spatial and temporal resolution for the collection of data can be defined, with the result that some variables are poorly described, whereas other variables have redundant observations.

In these terms, the climate system provides an especially interesting case study. In the study of geophysical fluid dynamics, one often encounters a wide range of scales. The spatial scales involved in atmospheric dynamics can range from $10^7$ m down to $10^{-6}$ m. Temporal scales can vary from $10^{16}$ to $10^{-6}$ s [19, 33]. Here, one has to take into account coupling between different subsystems, which feature dynamics taking place in vastly different ranges, as clarified when considering separately the atmosphere, the oceans, the biosphere, and the cryosphere [24, 31]. We still do not have a standard formulation of a climate model, because, depending on the class of problems that we want to tackle,
we need to resort to different approximations and even different formulations of the relevant dynamics. The state-of-the-art climate models used for the IPCC projections [32] cannot be used to study the past million years of Earth’s history, not only because they are too computationally expensive, but also because they do not take into appropriate consideration the processes which dominate the Earth’s dynamics on such long time scales. On the other hand, Earth models of intermediate complexity [10,26], which are tailored to such problems, are definitely not competitive if our aim is to describe the variability of the present climate and the short term climate change. The problem of providing a satisfactory description of the dynamical processes that cannot be explicitly be represented has long been addressed at various levels of rigor by the climate community [23,34,13,12,3,31,25]. The solutions that are being used in practical applications today go by the name of parameterizations. Processes that are not explicitly described are approximated via functions (deterministic or, more recently, stochastic) of the variables of the model, basically along the lines of classical closure theories in fluid dynamics for the Reynolds stress tensor. Given the approximate and empirical nature of the parameterizations, a problem that often arises is as regards their robustness with respect to changes in the modeling setup considered, such as in the case where additional variables are included in the model or different large scale conditions are considered. Retuning the parameterization is definitely a delicate aspect of climate modeling and a major hurdle to be overcome to achieve the goal of obtaining flexible climate models able to obtain so-called seamless simulation of the statistical properties and seamless (probabilistic) prediction of the state.

In this paper, we tackle this class of problems from a rather general point of view. We will consider a two-level system, consisting of two weakly coupled systems $X$ and $Y$. We focus our attention on the $X$ system. We want to derive ab initio a suitable parameterization of its coupling with the $Y$ system.

For this purpose, various reduction techniques have been proposed in the literature. A classic approach is the averaging method [3,14], which assumes that there is a vast time scale separation between the two systems $X$ and $Y$. Such a method allows one to derive a renormalized dynamics for the $X$ system whose trajectories converge on finite time scales to those of the original dynamics. An interesting use of the averaging method for deriving a simplified dynamics has been proposed in [1], where a sort of mean field ansatz is adopted. In statistical mechanics, projector operator techniques which allow us to formally eliminate certain variables have found a great deal of success, as in the case of the Mori–Zwanzig equation [36].

In the present paper, we take a different route, based on the response theory developed by Ruelle [28,30,17]. Such a theory provides explicit formulas for computing how the long-time averages of axiom A dynamical systems (described by the so-called SRB measure [35]) change as a result of small perturbations to the flow. This approach is especially useful for studying the impact of changes in the internal parameters of a system or of small modulations to the external forcing [27,6,2,18,22]. According to the theory, the changes in the observables can be expressed basically as a power series in orders of the strength of the forcing, so one should assume, in our case, that the coupling between the $X$ and $Y$ systems—which acts as a perturbation to the dynamics of $X$—should be small. We derive explicit expressions for the perturbations in the statistics of the $X$ system due to the coupling at the first and second order. We explain how these can be parameterized in terms of a surrogate deterministic forcing, a multiplicative noise, and a memory term, so
that we construct an approximated projected dynamics on the $X$ variables. Our approach does not make use of any large time scale separation between the $X$ and the $Y$ variables, and does not even assume that the $X$ variables are slower than the $Y$ variables.

This paper is structured as follows. In section 2 we give a short review of Ruelle’s response theory at higher orders. In section 3 we then use this theory to derive the response of the system $X$ to the coupling with the system $Y$. We derive the first-order and second-order terms and present an expression for a surrogate dynamics of $X$ alone. The correction to the unperturbed dynamics will include a deterministic forcing, a stochastic forcing, and a forcing given by a term which takes into account the history of the system. We present also a diagrammatic representation of the most relevant contributions. In section 4 we briefly discuss our results in comparison to classic methods of reduction of variables such as the averaging method [3,14] and the projection operator method [36]. In section 5 we present our conclusions and perspectives for future work. In an appendix, we describe how to extend the Ruelle response formula to forcings describing memory effects, which we use to treat the memory terms resulting at the second order of perturbation.

2. Response theory

Ruelle [29] recently derived explicit formulas for describing the smooth dependence of the SRB measure of axiom A dynamical systems on small perturbations of the flow. Such response theory boils down to a Kubo-like perturbative expression connecting the terms describing the linear and nonlinear responses of the system as expectation values of observables for the unperturbed measure. At order $n$ of nonlinearity, such expectation values can be written as $n$-tuple convolutions of a causal Green function with the time-delayed perturbative fields, so at every order, Kramers-Kronig relations can be written for the Fourier transform of the Green function, the so-called susceptibility [17]. In order to clarify the goal of response theory and set our notation, we present a simple derivation of the linear and higher order nonlinear response formulas obtained by introducing the so-called Dyson interaction picture for the operators. Given a continuous time dynamical system on a phase space $M$ with dynamics

$$\dot{x}_t = F(x_t) \quad \text{for } x \in M$$

and a corresponding flow function $f^t$ relating $x_t$ to the initial conditions $x_0$, $x_t = f^t(x_0)$, we assume that this dynamical system possesses a so-called Sinai–Ruelle–Bowen (SRB) measure. This implies that there exists a set $K \subset M$ of full Lebesgue measure on which long term averages coincide for every continuous observable $A : M \to R$:

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T \dd{\tau} A(f^\tau(x_0)) =: \rho_0(A) \quad \forall x_0 \in K.$$  

(2)

The existence of such an invariant measure has been proven for axiom A dynamical systems. For a discussion of dynamical systems that have an SRB measure, see [35]. Taking instead of an initial value $x_0$ an absolutely continuous initial measure on $K$, we can also write $\rho_0$ as

$$\rho_0(A) = \lim_{T \to \infty} \frac{1}{T} \int_0^T \dd{\tau} \int l(dx)A(f^\tau(x)).$$

(3)
We want to know how such averages, given by the invariant measure \( \rho_0 \), are affected by small changes in the dynamics of the system. Let us assume that the dynamics of the system is altered by adding a time-independent perturbation \( \Psi(x) \):

\[
\dot{x}_t = F(x_t) + \Psi(x_t),
\]

with \( \tilde{f}^t \) the perturbed flow function, so that we end up with a modified SRB measure \( \tilde{\rho} \):

\[
\tilde{\rho}(A) = \lim_{T \to \infty} \frac{1}{T} \int_0^T \mathrm{d}\tau \int l(dx)A(\tilde{f}^\tau(x)).
\]

If the strength of the perturbation is small, the corresponding change in the SRB measure of the system will also in general be small. The occurrence of resonances has been studied in [30, 5, 7], but this is not expected to occur in systems of physical interest (see [8] and references therein). Hence, we can expand the perturbed SRB measure in powers of the perturbation \( \Psi \). In order to do so, we first expand the evolved observable \( A(\tilde{f}^t(x)) \). This amounts to constructing a Dyson series, where the perturbation to the dynamical system plays the role of the interaction Hamiltonian. Using the chain rule, the time dependence of an observable \( A \) at time \( t \) is described by

\[
\frac{dA(\tilde{f}^t(x_0))}{dt} = \mathcal{L}A(\tilde{f}^t(x_0)) = (\mathcal{L}_0 + \mathcal{L}_1)A(\tilde{f}^t(x_0)),
\]

where

\[
\mathcal{L}_0 A(x) = F(x) \cdot \nabla_x A(x), \quad \mathcal{L}_1 = \Psi(x) \cdot \nabla_x A(x), \quad \mathcal{L} = \mathcal{L}_0 + \mathcal{L}_1.
\]

By defining an interaction observable as \( A_I(\tau, x_0) = \Pi_0(-\tau)A(\tilde{f}^\tau(x_0)) = \exp(-\mathcal{L}_0 \tau)\exp(\mathcal{L} \tau)A(x_0) \) with \( \Pi_0(\tau) = \exp(\mathcal{L}_0 \tau) \), we can make an expansion of \( A \) in orders of the perturbation \( \Psi \). This observable \( A_I \) obeys the differential equation

\[
\frac{dA_I(\tau, x_0)}{d\tau} = \mathcal{L}_I(\tau)A_I(\tau, x_0),
\]

where

\[
\mathcal{L}_I(\tau) = \Pi_0(-\tau)\mathcal{L}_1 \Pi_0(\tau).
\]

Note that the operator \( \mathcal{L}_I \) is of order \( \Psi \). A formal solution to equation (8) is given by Duhamel’s formula:

\[
A_I(\tau, x_0) = A_I(0, x_0) + \int_0^\tau \mathrm{d}s_1 \mathcal{L}_I(s_1)A_I(s_1, x_0).
\]

By iterating this solution and making use of the fact that \( A_I(\tau, x_0) = \exp(-\mathcal{L}_0 \tau)A(\tilde{f}^\tau(x_0)) \), we get an expression for the evolved observable \( A(\tilde{f}^t(x_0)) \) in terms of \( A(0, x_0) \):

\[
A(\tilde{f}^t(x_0)) = \Pi_0(\tau)A(x_0) + \int_0^\tau \mathrm{d}s_1 \Pi_0(\tau - s_1)\mathcal{L}_1 \Pi_0(s_1)A(x_0)
\]

\[
+ \int_0^\tau \mathrm{d}s_1 \int_0^{s_1} \mathrm{d}s_2 \Pi_0(\tau - s_1)\mathcal{L}_1 \Pi_0(s_1 - s_2)\mathcal{L}_1 \Pi_0(s_2)A(x_0) + O(\Psi^3).
\]

doi:10.1088/1742-5468/2012/03/P03003
Such an expression for the expectation values under perturbed dynamics $A(\tilde{f}^T(x_0))$ in terms of the unperturbed dynamics $\Pi_0(\tau)A(x_0) = A(f^\tau(x_0))$ allows us to expand the SRB measure $\tilde{\rho}$ around $\rho_0$. The perturbed SRB measure $\tilde{\rho}$ can be written as

$$\tilde{\rho}(A) = \lim_{T \to \infty} \frac{1}{T} \int_0^T d\tau \int l(dx) \left( \Pi_0(\tau)A(x) + \int_0^\tau ds_1 \Pi_0(\tau - s_1)\mathcal{L}_1\Pi_0(s_1)A(x) + O(\Psi^2) \right)$$

$$= \rho_0(A) + \delta \rho^{(1)}(A) + O(\Psi^2), \quad (11)$$

where the first term in this sum gives the unperturbed measure $\rho_0$, while the second term is the linear response to $\Psi$. We can rewrite it as

$$\delta \rho^{(1)}(A) = \lim_{T \to \infty} \frac{1}{T} \int_0^T ds_1 \int_0^T d\tau \int l(dx)\Pi_0(\tau - s_1)\mathcal{L}_1\Pi_0(s_1)A(x)$$

$$= \lim_{T \to \infty} \frac{1}{T} \int_0^T ds_1 \int_0^T s_1 \int_0^u l(dx)\Pi_0(u)\mathcal{L}_1\Pi_0(s_1)A(x)$$

$$= \int_0^\infty ds_1 \rho_0(\mathcal{L}_1\Pi_0(s_1)A) = \int_0^\infty ds_1 \int \rho_0(dx)\Psi(x) \cdot \nabla_x A(f^{s_1}(x))$$

$$= \int_0^\infty ds_1 G^{(1)}_{A,\Psi}(s_1). \quad (12)$$

This shows that the linear response can be written in terms of averages taken in the unperturbed system. Continuing the calculation, we obtain the following expression for the higher order terms:

$$\tilde{\rho}(A) = \rho_0(A) + \int_0^\infty ds_1 \rho_0(\mathcal{L}_1\Pi_0(s_1)A)$$

$$+ \int_0^\infty ds_1 \int_0^{s_1} ds_2 \rho_0(\mathcal{L}_1\Pi_0(s_1)\mathcal{L}_1\Pi_0(s_2)A) + O(\Psi^3)$$

$$= \rho_0(A) + \delta \rho^{(1)}(A) + \delta \rho^{(2)}(A) + O(\Psi^3), \quad (13)$$

where

$$\delta \rho^{(2)}(A) = \int_0^\infty ds_1 \int_0^{s_1} ds_2 \rho_0(\mathcal{L}_1\Pi_0(s_1)\mathcal{L}_1\Pi_0(s_2)A)$$

$$= \int_0^\infty ds_1 \int_0^\infty ds_2 \int \rho_0(dx)\Psi(x) \cdot \nabla_x \Psi(f^{s_1}(x)) \cdot \nabla f^{s_2}(x) A(f^{s_1+s_2}(x))$$

$$= \int_0^\infty ds_1 \int_0^\infty ds_2 G^{(2)}_{A,\Psi}(s_1, s_2). \quad (14)$$

The question of whether the response terms remain finite is in general difficult to answer. To have a finite response, the functions $G^{(1)}_{A,\Psi}$ and $G^{(2)}_{A,\Psi}$ should rapidly drop off to zero for large times. For axiom A dynamical systems this has been proven to be the case by splitting the response into contributions in unstable and stable directions and making use of mixing properties in the unstable directions and dampening of perturbations in the stable directions. See [30] for a discussion of the first-order term and [29] for the second-order term.
3. Two-level systems

We will now make use of the response formulas derived in section 2 to study how the statistical properties of a dynamical system which is composed, in the unperturbed case, of two independent subsystems change when a weak coupling is introduced. Let us assume that the variables $x$ of the dynamical system can be separated into two groups of variables $X$ and $Y$:

$$ x = \begin{pmatrix} X \\ Y \end{pmatrix}, $$

and so the unperturbed dynamical system can be considered as given by the uncoupled dynamics

$$ \dot{X} = F_X(X), \quad \dot{Y} = F_Y(Y). \quad \text{(15)} $$

A weak coupling between the two subsystems is then introduced as follows:

$$ \begin{aligned} 
\dot{X} &= F_X(X) + \Psi_X(X, Y), \\
\dot{Y} &= F_Y(Y) + \Psi_Y(X, Y). 
\end{aligned} \quad \text{(16)} $$

The unperturbed and perturbed evolution operators $L_0$ and $L_1$ are now given by

$$ L_0 = F_X(X) \cdot \nabla_X + F_Y(Y) \cdot \nabla_Y \quad \text{(17)} $$

and

$$ L_1 = \Psi_X(X, Y) \cdot \nabla_X + \Psi_Y(X, Y) \cdot \nabla_Y = \Psi^T \cdot \nabla, \quad \text{(18)} $$

where $\Psi^T = (\Psi_X \quad \Psi_Y)$ and $\nabla^T = (\nabla_X \quad \nabla_Y)$.

3.1. First order

We derive from equations (7) and (13) that the first-order response to $L_1$ is given by

$$ \delta^{(1)}_\Psi \rho(A) = \int_0^\infty ds_1 \rho_0(A \Pi_0(s_1)) ds_1 \rho_0(dx) \Psi^T(x) \nabla(A \circ f^{s_1})(x), \quad \text{(19)} $$

where $\rho_0$ is the SRB measure corresponding to the uncoupled dynamics of equation (15). The response is an integral over time of the response function $G_{A,\Psi}^{(1)}(\tau)$ describing the delayed impact of a forcing in the past over a time span of $\tau$:

$$ \delta^{(1)}_\Psi \rho(A) = \int_0^\infty ds_1 G_{A,\Psi}^{(1)}(s_1) \quad G_{A,\Psi}^{(1)}(s_1) = \int \rho_0(dx) \Psi^T(x) \nabla(A \circ f^{s_1})(x). \quad \text{(21)} $$

On exploiting the fact that the unperturbed dynamics of the $X$ and $Y$ subsystems are uncoupled, this becomes

$$ G_{A,\Psi}^{(1)}(\tau) = \int \rho_0(dx) \Psi^T(x) \cdot \nabla A \left( \left[ \begin{array}{c} f_X^\tau(X) \\ f_Y^\tau(Y) \end{array} \right] \right), \quad \text{(22)} $$

doi:10.1088/1742-5468/2012/03/P03003
where \( f^X_\tau \) and \( f^Y_\tau \) represent the unperturbed flows of the \( X \) and \( Y \) subsystems, respectively. We are only interested in the evolution of observables on the \( X \) system so we consider observables \( A \) that depend on \( X \) variables only:

\[
A \left( \left( \begin{array}{c} X \\ Y \end{array} \right) \right) = A(X).
\] (23)

Therefore, the response function \( G^{(1)}_{A,\Psi} \) can be written as

\[
G^{(1)}_{A,\Psi}(\tau) = \int \rho_0 \left( \frac{dX}{dY} \right) \Psi_X(X,Y) \cdot \nabla_X A(f^X_\tau(X)).
\] (24)

Since the dynamics of the unperturbed system is decoupled, its SRB measure \( \rho_0 \) can be written as a product of the SRB measures \( \rho_{0,X}(dX) \) and \( \rho_{0,Y}(dY) \). We have

\[
G^{(1)}_{A,\Psi}(\tau) = \int \rho_{0,X}(dX) \langle \Psi_X(X,Y) \rangle_{\rho_{0,Y}} \cdot \nabla_X A(f^X_\tau(X)).
\] (25)

Such a first-order perturbation can be represented in a diagrammatic way, as depicted in figure 1, where the arrow indicates the direction of the interaction and the time variable refers to the presence of a time-delayed, causal effect on the system of interest of the external perturbation.

Equation (25) can be used to derive a parameterization of the first-order effect of coupling the \( X \) and \( Y \) subsystems on the \( A(X) \) observables, expressed as a function of the \( X \) variables only. In other words, we want to derive an expression for a perturbation vector field \( M_1 \) such that if we consider the dynamical system

\[
\frac{dX}{dt} = F_X(X) + M_1(X),
\] (26)

where \( F_X(X) \) is the same as in equation (15), up to the first order, the expectation values of the \( A \) observables over the SRB invariant measure \( \tilde{\rho}_1 \) of the dynamical system given in equation (26) and over the invariant measure \( \rho_{X,Y} \) are identical. The invariant measure \( \rho_0(dX) \) for the unperturbed \( M_1 = 0 \) system is, obviously, identical to the partial measure \( \rho_{0,X}(dX) \) introduced in equation (24). Therefore, we must impose that the first-order effect in equation (26) due to \( M_1 \) is the same as the first-order correction due to the coupling between the \( X \) and the \( Y \) variables introduced in equation (16). We discover
that on choosing
\[ M_1(X) := \langle \Psi_X(X, Y) \rangle_{\rho_0, Y} \tag{27} \]
and plugging this expression into equation (12), we obtain that up to the first order in \( \Psi \) and \( \forall A \),
\[ \delta \rho^{(1)}_{M_1}(A) = \delta^{(1)}_{\Psi} \rho(A), \tag{28} \]
and so
\[ \tilde{\rho}_{X,Y}(A) = \rho_{0, X}(A) + \delta^{(1)}_{\Psi} \rho(A) + O(\Psi^2) = \tilde{\rho}_1(A) + O(\Psi^2) = \tilde{\rho}_0(A) + \delta \rho^{(1)}_{M_1}(A) + O(\Psi^2). \tag{29} \]

Concluding, the effect on \( X \) of its coupling with \( Y \) can be described at first order with a mean field theory constructed by taking the \( Y \)-ensemble average of the coupling, where the averaging is taken on the uncoupled \( Y \) dynamics.

3.2. Second order

From equation (13) we can derive that at second order the response is a double integral of a response function \( G^{(2)}_{A, \Psi} \) by plugging the expressions given in equations (17) and (18) into equation (14):
\[ \delta^{(2)}_{\Psi, \rho}(A) = \int_0^\infty \int_0^\infty ds_1 \int_0^\infty ds_2 \rho(L_1 \Pi_0(s_1) L_1 \Pi_0(s_2) A) = \int_0^\infty \int_0^\infty ds_1 \int_0^\infty ds_2 G^{(2)}_{A, \Psi, \Psi}(s_1, s_2) 
\]
\[ = \delta^{(2)}_{\Psi, \Psi, a} \rho(A) + \delta^{(2)}_{\Psi, \Psi, b} \rho(A) = \int_0^\infty \int_0^\infty ds_1 \int_0^\infty ds_2 G^{(2)}_{A, \Psi, \Psi, a}(s_1, s_2) + G^{(2)}_{A, \Psi, \Psi, b}(s_1, s_2), \tag{30} \]
where we separate such a second-order contribution into the sum of two terms, \( G^{(2)}_{A, \Psi, \Psi, a} \) and \( G^{(2)}_{A, \Psi, \Psi, b} \), defined as
\[ G^{(2)}_{A, \Psi, \Psi, a}(s_1, s_2) = \int \rho_0(dx) \Psi_X(X, Y) \cdot \nabla_X \Pi_0(s_1) (\Psi_X(X, Y) \cdot \nabla_X) (A(f^s(X))), \]
\[ G^{(2)}_{A, \Psi, \Psi, b}(s_1, s_2) = \int \rho_0(dx) \Psi_X(Y, Y) \cdot \nabla_Y \Pi_0(s_1) (\Psi_X(X, Y) \cdot \nabla_X) (A(f^s(X))). \tag{31} \]

The Green function \( G^{(2)}_{A, \Psi, \Psi, a}(s_1, s_2) \) responsible for second contribution \( \delta^{(2)}_{\Psi, \Psi, a} \) can be depicted in a diagrammatic mode as in figure 2. This term describes the second-order effect of the perturbation of the \( Y \) dynamics on the statistical properties of the \( X \) system, and does not take into account the modifications on the statistical properties of \( Y \) due to its coiling with \( X \). Instead, figure 3 provides a graphical representation of the \( G^{(2)}_{A, \Psi, \Psi, b}(s_1, s_2) \), responsible for the second-order contribution \( \delta^{(2)}_{\Psi, \Psi, b} \). In this case, the second-order effect comes as a result of the modification to the first-order effect shown in figure 1 due to the impact of the \( X \) dynamics on the \( Y \) systems: note that in this case, as opposed to the case in figures 1–2, one arrow points from the \( X \) to the \( Y \) system.

Our goal is now to provide a second-order parameterization of the coupling in terms of the \( X \) variables alone, such that the invariant measure \( \tilde{\rho}_2 \) of the resulting system is

doi:10.1088/1742-5468/2012/03/P03003

9
such that $\forall A$,

$$\rho_{X,Y}(A) = \hat{\rho}_2(A) + O(\Psi^3).$$

(32)

In the following, we show that if we decompose the coupling term $\Psi_X(X,Y)$ into its mean w.r.t. $\rho_{0,Y}$ plus fluctuations in the form of a sum of separable contributions:

$$\Psi_X(X,Y) = M(X) + \sum_{i,j} a_{ij} \Psi'_{X,1,i}(X) \Psi'_{X,2,j}(Y)$$

(33)

this can be accomplished by the following reduced system:

$$\dot{X}(t) = F_X(X(t)) + M_1(X(t)) + \sum_{i,j} a_{ij} \Psi'_{X,1,i}(X(t)) \sigma_j(t) + \int_0^\infty d\tau h(\tau, X(t-\tau))$$

(34)

where $\sigma$ mimics the correlations present in the fluctuations:

$$\langle \sigma_j(t) \sigma_l(t+\tau) \rangle = \langle \Psi_{X,2,j}(Y) \Psi'_{X,2,l}(f^{\tau}(Y)) \rangle,$$

(35)

$$\langle \sigma_j(t) \rangle = 0.$$  

(36)

and $h$ is a kernel representing the memory effected by the presence of unresolved variables

$$h_{ij}(\tau, X) = \langle \Psi_{Y,i}(X,Y) \partial_{Y,i} \Psi_{X,j}(f^{\tau}(X), f^{\tau}(Y)) \rangle_{\rho_{0,Y}}.$$  

(37)

doi:10.1088/1742-5468/2012/03/P03003
For simplicity, we indicate all of the perturbative terms in equation (34) as $M_j(X(t))$, $M_2$ being the stochastic term and $M_3$ being the memory term. We will now derive the expression for $M_2$ and $M_3$ by matching the corrections up to $O(\Psi^2)$ to the unperturbed dynamics of the previous part with what is given in equations (21) and (30). Obviously, since equation (34) has a stochastic component, its invariant measure $\hat{\rho}_2$ will be accordingly defined. Please note that in the following, when we refer to first-order or second-order perturbations to the measure $\rho_0$ due to the $M_j$ terms in equation (34), the expectation values over the realizations of the stochastic processes are implicitly considered. Note that in [20] we have explained how to treat the impact of stochastic perturbations on deterministic dynamical systems along the lines of the Ruelle response theory.

The first-order response in equation (34) can be written as $\sum_k \delta_{M_k}^{(1)} \rho(A)$, where

$$
\delta_{M_k}^{(1)} \rho(A) = \int_0^\infty ds_1 \int_0^\infty \rho_0(X) (dX) \int_0^\infty ds_1 G_{A,M_k}^{(1)}(s_1)
= \int_0^\infty ds_1 \int_0^\infty \rho_0(X) (dX) M_k(X) \cdot \nabla_X A(f_X^1(X)).
$$

The full second-order response can be written as $\sum_{i,j} \delta_{M_i,M_j}^{(2)} \rho(A)$, where

$$
\delta_{M_i,M_j}^{(2)} \rho(A) = \int_0^\infty ds_1 \int_0^\infty ds_2 \int_0^\infty \rho_0(X) (dX) G_{A,M_i,M_j}^{(2)}(s_1, s_2)
= \int_0^\infty ds_1 \int_0^\infty ds_2 \int_0^\infty \rho_0(X) (dX) M_i(X) \cdot \nabla_X M_j(f_{X_2}^2(X)) \cdot \nabla_{f_{X_2}^2(X)} A(f_{X_1}^1(X))
$$

Combining equations (29) and (32), we impose that

$$
\delta_{\Psi}^{(1)} \rho(A) + \delta_{\Psi,\Psi,a}^{(2)} \rho(A) + \delta_{\Psi,\Psi,b}^{(2)} \rho(A) = \sum_k \delta_{M_k}^{(1)} \rho(A) + \sum_{i,j} \delta_{M_i,M_j}^{(2)} \rho(A) + O(\Psi^3).
$$

Our procedure will lead to the simplifying of equation (40) into the following:

$$
\delta_{\Psi}^{(1)} \rho(A) + \delta_{\Psi,\Psi,a}^{(2)} \rho(A) + \delta_{\Psi,\Psi,b}^{(2)} \rho(A) = \delta_{M_1}^{(1)} \rho(A) + \delta_{M_2}^{(1)} \rho(A) + \delta_{M_1,M_2}^{(2)} \rho(A) + \delta_{M_1,M_2}^{(2)} \rho(A),
$$

where the terms missing on the right-hand side in equation (41) with respect to equation (40) are of order $O(\Psi^3)$ or higher, or equal zero under averaging over noise realizations.

From equation (29), we have that $\delta_{\Psi}^{(1)} \rho(A) = \delta_{M_1}^{(1)} \rho(A)$. We now consider the next order of perturbation. We first observe that $\delta_{\Psi,\Psi,a}^{(2)} \rho(A)$ can be written as

$$
\delta_{\Psi,\Psi,a}^{(2)} \rho(A) = \int_0^\infty ds_1 \int_0^\infty ds_2 \int_0^\infty \rho_0(dx)((\Psi_X(X,Y) - M(X)) \cdot \nabla_X \times \Pi_0(s_1)((\Psi_X(X,Y) - M(X)) \cdot \nabla_X) A(f_{X_2}^2))
+ \int_0^\infty ds_1 \int_0^\infty ds_2 \int_0^\infty \rho_0(dx)(M(X) \cdot \nabla_X) \Pi_0(s_1)((M(X) \cdot \nabla_X) A(f_{X_2}^2))
= \int_0^\infty \rho_0(dx)(\Psi_X(X,Y) \cdot \nabla_X) \Pi_0(s_1)(\Psi_X(X,Y) \cdot \nabla_X) A(f_{X_2}^2) + \delta_{M_1,M_2}^{(2)} \rho(A)
= \delta_{\Psi,\Psi,f}^{(2)} \rho(A) + \delta_{M_1,M_2}^{(2)} \rho(A),
$$

where $\delta_{\Psi,\Psi,f}^{(2)} \rho(A)$ is given by

$$
\delta_{\Psi,\Psi,f}^{(2)} \rho(A) = \int_0^\infty ds_1 \int_0^\infty \rho_0(dx)((\Psi_X(X,Y) - M(X)) \cdot \nabla_X)
\times \Pi_0(s_1)((\Psi_X(X,Y) - M(X)) \cdot \nabla_X) A(f_{X_2}^2) + \delta_{M_1,M_2}^{(2)} \rho(A).
$$
where \( \Psi'_X(X, Y) = \Psi_X(X, Y) - M(X) \) gives the fluctuations of \( \Psi \) around its \( Y \) averaged value \( M \). We now construct the term \( M_2(X(t)) \) such that its first-order impact \( \delta^{(1)}_{M_2} \rho(\mathcal{A}) \) vanishes, whereas its second contribution \( \delta^{(2)}_{M_2} \rho(\mathcal{A}) \) corresponds exactly to \( \delta^{(2)}_{\Psi, \Psi, f} \rho(\mathcal{A}) \) introduced in equation (42), which we refer to as the fluctuation term.

**The fluctuation term.** To analyze the \( \delta^{(2)}_{\Psi, \Psi, f} \rho(\mathcal{A}) = \iint ds_1 ds_2 G^{(2)}_{A, \Psi, \Psi, f}(s_1, s_2) \) term in equation (42), we first consider the case where the coupling fluctuations \( \Psi_X \) are separable in \( X \) and \( Y \) variables:

\[
\Psi'_X(X, Y) = \Psi'_{X,1}(X) \Psi'_{X,2}(Y).
\]

In this case the response function \( G^{(2)}_{A, \Psi, \Psi, f} \) can now be written as

\[
G^{(2)}_{A, \Psi, \Psi, f}(s_1, s_2) = \int \rho_{0,X}(dX)g(s_1)\Psi'_{X,1}(X)\nabla_X \Psi'_{X,1}(f^s_{X}(X))\nabla f^s_{X}(X)(A \circ f^s_{X})(f^s_{X}(X)),
\]

where

\[
g(s_1) = \langle \Psi'_{X,2}(Y)\Psi'_{X,2}(f^s_{Y}(Y)) \rangle_{\rho_{0,Y}}.
\]

We now impose that the second-order contributions \( \delta^{(2)}_{\Psi, \Psi, f} \rho(\mathcal{A}) \) and \( \delta^{(2)}_{M_2} \rho(\mathcal{A}) \) are identical, so that

\[
G^{(2)}_{A, \Psi, \Psi, f}(s_1, s_2) = G^{(2)}_{A, M_2}(s_1, s_2) = \int \rho_{0,X}(dX)M_2(T-s_1-s_2, X)\nabla_X M_2(T-s_1, f^s_{X}(X))\nabla f^s_{X}(X)(A(f^s_{X}+s_2(X)))
\]

for some time-dependent perturbation \( M_2 \). This can be achieved by choosing the perturbation \( M_2 \) to be a multiplicative noise term described as follows:

\[
M_2(t, X) = \Psi'_{X,1}(X)\sigma(t), \quad \langle \sigma(t_1)\sigma(t_2) \rangle = g(t_2 - t_1), \quad \langle \sigma(t) \rangle = 0,
\]

where the average in the second equality is an average over realizations of the noise. Taking the average of the model response over the realizations of the noise gives

\[
\langle G^{(2)}_{A, M_2}(s_1, s_2) \rangle = G^{(2)}_{A, \Psi, \Psi, f}(s_1, s_2) \Rightarrow \delta^{(2)}_{M_2, M_2} \rho(\mathcal{A}) = \delta^{(2)}_{\Psi, \Psi, f} \rho(\mathcal{A}).
\]

These results can be easily generalized also to the case where the coupling term \( \Psi'_X \) is not separable. In this case, \( \Psi'_X \) can be decomposed as a product of Schauder bases of functions in \( X \) and \( Y \):

\[
\Psi'_X(X, Y) = \sum_{i,j} a_{ij} \Psi'_{X,1,i}(X) \Psi'_{X,2,j}(Y).
\]

The response function is now given by

\[
G^{(2)}_{A, \Psi, \Psi, f}(s_1, s_2) = \int \rho_{0,X}(dX) \times \sum_{i,j,k,l} a_{ij}a_{kl}g_{j,l}(s_1)\Psi'_{X,1,i}(X)\nabla_X \Psi'_{X,1,k}(f^s_{X}(X))\nabla f^s_{X}(A \circ f^s_{X})(f^s_{X}(X))
\]

\[
g_{j,l}(s_1) = \langle \Psi'_{X,2,j}(Y)\Psi'_{X,2,l}(f^s_{Y}(Y)) \rangle.
\]
The response can in this case be modeled with a linear combination of multiplicative noise

$$M_2(X) = \sum_{i,j} a_{ij} \Psi'_{i,i}(X) \sigma_j(t),$$

(49)

$$\langle \sigma_j(t_1) \sigma_i(t_2) \rangle = g_{j,i}(t_2 - t_1),$$

(50)

$$\langle \sigma_j(t) \rangle = 0.$$  (51)

Note also that, since $\langle \sigma_j(t) \rangle = 0 \ \forall j$, we have that $\delta^{(1)}_{\delta_2} \rho(A) = \delta^{(2)}_{\delta_1} \rho_2(A) = \delta^{(2)}_{\delta_2} \rho(A) = 0$, where we have already anticipated that $M_3$ is not written as a stochastic term. At this stage, we have shown that $\delta^{(1)}_{\delta_1} \rho(A) + \delta^{(2)}_{\delta_1,a} \rho(A) = \delta^{(1)}_{\delta_1} \rho(A) + \delta^{(2)}_{\delta_1} \rho(A) + \delta^{(2)}_{\delta_2} \rho(A)$.

The memory term. In order to complete the construction of the parameterization leading to the surrogate dynamics given in equation (41), we need to show that $\delta^{(2)}_{\rho \rho,b} \rho(A) = \delta^{(1)}_{\rho \rho,b} \rho(A)$, and that the two terms $\delta^{(2)}_{\rho \rho,b} \rho(A)$ and $\delta^{(2)}_{\rho \rho,b} \rho(A)$ are $O(\Psi^3)$ or higher. The $G^{(2)}_{A,\Psi,\Psi,\rho,b}$ term in equations (31) can be seen as a modification to the first-order response on the $X$ system related to the coupling on $Y$ due to the fact that the $Y$ system is modified by the interaction with the $X$ system. After integrating over the $Y$ variables, we can express $\delta^{(2)}_{\rho \rho,b} \rho(A)$ as

$$\delta^{(2)}_{\rho \rho,b} \rho(A) = \int_0^\infty ds_1 \int_0^\infty ds_2 \int \rho_{0,X}(dX) \langle \Psi_{Y,i}(X,Y) \partial_{Y,i} \Psi_{X,j}(f^{s_2}(X),f^{s_2}(Y)) \rangle_{\rho_{0,Y}} \times \partial_{f^{s_2}(X),j}(A(f^{s_1+s_2}(X))).$$

(52)

This can be expressed as

$$\delta^{(2)}_{\rho \rho,b} \rho(A) = \int_0^\infty ds_1 \int_0^\infty ds_2 \int \rho_{0,X}(dX) h_j(s_2, X) \partial_{f^{s_2}(X),j}(A(f^{s_1+s_2}(X))),$$

(53)

where we introduce the following integral kernel:

$$h_j(s, X) = \langle \Psi_{Y,i}(X,Y) \partial_{Y,i} \Psi_{X,j}(f^s(X), f^s(Y)) \rangle_{\rho_{0,Y}}.$$  (54)

This term describes a memory effect: the evolution of the $Y$ variables between times $T - s_1 - s_2$ and $T - s_1$ keeps track (at first order) of the previous state of the $X$ system. Therefore, when we consider the impact of the $Y$ variables on the $X$ system, at second order we get a contribution coming from the history of $X$ itself. Obviously, the shorter the decorrelation time of the $Y$ system, the weaker the influence of this term.

A last step needed at this stage is to provide a response formula for integral forcings, since the formulas derived in section 2 are only valid for perturbations without memory. In the appendix we derive such a generalized expression, which includes as a special case that which has been discussed in section 2. The linear response to a perturbation which can be written as follows:

$$\dot{X}(t) = F_X(X(t)) \rightarrow \dot{X}(t) = F_X(X(t)) + \int_0^\infty d\tau h(t - \tau, X(t - \tau))$$

(55)

results as, $\forall$ observables $A$,

$$\delta^{(1)} \rho(A) = \int_0^\infty ds_1 \int_0^\infty ds_2 \int \rho_{0,X}(dX) h_j(s_2, X) \partial_{f^{s_2}(X),j}(A(f^{s_1+s_2}(X))).$$

(56)

doi:10.1088/1742-5468/2012/03/P03003
This indeed implies that on choosing $M_3 = \int_0^\infty d\tau h(\tau, X(t-\tau))$ in equation (34), we obtain $\delta^{(2)}_{\Psi,\Psi,b}\rho(A) = \delta^{(1)}_{M_3}\rho(A)$. It is important to note that, by construction, even if $\delta^{(1)}_{M_3}\rho(A)$ is a first-order perturbation of the expectation value of $A$, it is $O(\Psi^2)$. Therefore, all terms such as $\delta^{(2)}_{M_3, M_3}\rho(A)$, $\delta^{(1)}_{M_1, M_3}\rho(A)$ presented in equation (40) are $O(\Psi^3)$ or higher, so equation (41) is indeed correct. Analogously to the case for the noise discussed in the previous section, the term $h_j(s, X)$ can be estimated from a trajectory of the system and then used to provide the suitable parameterization. This concludes our derivation and leads to the result anticipated in equation (34).

3.3. The independent coupling case

An especially interesting result can be obtained when considering the not uncommon situation where the coupling is independent of the variable that it is affecting, i.e. $\Psi_X(Y)$ depends only on $Y$ and $\Psi_Y(X)$ depends only on $X$. In this case, the reduced dynamical system able to mimic the statistics of the full system for the observables depending on $X$ variables only up to the second order on the coupling strength can be written as

$$\dot{X}(s) = F_X(X(s)) + M_1 + \sigma(s) + \int_0^\infty d\tau h(s - \tau, X'(t - \tau)).$$  \hspace{1cm} (57)

The first-order term $M_1$ is now a constant drift term and the fluctuation term is an additive noise term with correlations

$$\langle \sigma(t_1)\sigma(t_2) \rangle = \langle (\Psi_X(Y) - M_1)(\Psi_X(f^{t_2-t_1}(Y)) - M_1) \rangle_{\rho_0,Y}. \hspace{1cm} (58)$$

The integral kernel of the memory term simplifies to

$$h(s, X) = \Psi_Y(X)\langle \partial_Y \Psi_X(f^s(Y)) \rangle_{\rho_0,Y}. \hspace{1cm} (59)$$

where the role of the impact of the memory effect of the $Y$ variables on the memory of the $X$ system is especially clear.

4. Comparison with other methods of reduction of variables

The problem of reduction of variables is a classic topic in dynamical systems theory and statistical mechanics. We will briefly discuss the relation between these methods and what is presented here. A well-developed approach is the so-called averaging method [3,14]. This method deals with coupled systems of the form

$$\dot{X} = f(X, Y),$$  \hspace{1cm} (60)

$$\dot{Y} = \frac{1}{\epsilon}g(X, Y).$$  \hspace{1cm} (61)

In the limit where the time scale difference between the two goes to infinity, on taking $\epsilon \to \infty$, the trajectories of the original system converge on finite time scales to the trajectories of an averaged version of the $X$ system, namely

$$\dot{X}' = \rho_{Y\mid X}(f(X', Y)). \hspace{1cm} (62)$$

doi:10.1088/1742-5468/2012/03/P03003
where \( Y \) is average over the measure \( \rho_{Y|X} \), which is the SRB measure corresponding to the dynamics of \( Y \) for fixed \( X \):

\[
\dot{Y} = g_X(Y), \\
g_X(Y) = g(X,Y).
\]

The fluctuations around this average trajectory can be approximated by an Ornstein–Uhlenbeck process. One should note that while we have discussed here long term averages in the form of SRB measures, the averaging method describes convergence on finite time scales. Another point of difference is the weak coupling assumption that we have made here. This assumption is not made in the averaging method, where the limit of an infinitely fast \( Y \) process is taken instead. Such a limit basically takes care of getting rid of the higher order effects discussed in this paper.

In [1] the averaging method is used to derive a simplified reduced model. The \( X \)-dependent measure \( \rho_{Y|X} \) is first replaced by the measure \( \rho_{Y|X^*} \) for a fixed value \( X^* \) for \( X \). To improve the statistics of the reduced model a linear term in \( X \) is added:

\[
\dot{X}' = \rho_{Y|X} (f(X',Y)) + (X - X^*)\delta\rho_{Y|X^*} (f(X',Y)),
\]

where \( \delta\rho_{Y|X} \) is the response of the measure \( \rho_{Y|X^*} \) to a change in the value of the fixed \( X \). Such a feedback mechanism, where a change in \( X \) affects the \( X \) dynamics through the presence of \( Y \), is also present in the memory effect that we find here. However, as the limit of infinite time scale differences is taken in the averaging method, there the feedback happens instantly from the point of view of the slow process and no memory is present.

Projector operator techniques have found much use in the statistical mechanics for deriving Fokker–Planck equations. A celebrated result is the Mori–Zwanzig equation, which consists of an integro-differential equation, where the integral term describes the memory effect due to projected out variables [36]. Consider the setting of a dynamical system with Liouville operator \( L \), where we have a projector operator \( P \) projecting out irrelevant parameters. Every observable \( A \) can then be written as a 2-vector containing \( A_1 = PA \) and \( A_2 = QA = (1 - P)A \). The Liouville equation for this vector is given by

\[
\frac{d}{dt} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix},
\]

where \( L_{11} = PLP \), \( L_{12} = PLQ \), \( L_{21} = QLP \) and \( L_{22} = QLQ \). By using Duhamel’s formula for the differential equation for \( A_2 \), a solution for \( A_2 \) can be found where the term with \( L_{21}A_1 \) is seen as a perturbation to the free evolution of \( A_2 \) under \( L_{22} \). On plugging this solution back into the differential equation for \( A_1 \), the following equation of motion for \( A_1 \) is obtained:

\[
\frac{d}{dt} A_1(t) = L_{11}A_1(t) + L_{12} \int_0^t ds \exp(L_{22}(t-s))L_{21}A_1(s) + L_{12} \exp(L_{22}t)A_2(0).
\]
5. Conclusions

In this paper we have adopted the Ruelle response theory to study the effect of a weak coupling between two dynamical systems. We have derived explicit formulas for computing the changes in the expectation value of a general observable of one of the two systems (we call this system $X$, and refer to the other system as $Y$) up to the second order in the coupling strength. We have also presented a simple diagrammatic representation of such terms, which is suggestive of the possibility of reframing the Ruelle theory in graphical terms along the lines of the Feynman representation of quantum electrodynamics.

These results allow us to define a surrogate perturbed dynamics for the system $X$ such that up to second order its statistical properties are the same as those of the coupled system when a general observable of $X$ alone is considered. In other words, we have attempted to derive a parameterization of the coupling. We have found that the invariant measure for the coupled system restricted to the $X$ system can be approximated by the invariant measure of a reduced system with three forcing terms $M_i$ added to the unperturbed flow $F_X(X(t))$, $M_1$ being a deterministic correction which fully describes the change in the invariant measure up to first order, $M_2$ being a stochastic forcing expressed as a sum of multiplicative noise terms with non-trivial correlation properties, and $M_3$ being an integral expression which describes a memory term. The explicit expressions for the $M_j$ terms are given in equation (34). The $M_1$ term basically describes a mean field theory, where the effective $X$ dynamics is modified by adding the effect of the $Y$ averaged dynamics. The $M_2$ term, instead, takes care of representing the effect of the fluctuations of the $Y$ systems. Finally, the $M_3$ term describes the non-Markovian effect due to the $X$ system impacting the state of the $Y$ state at a given time in the past, which in turn perturbs the $X$ system at a later time. Note that, if the coupling is independent of the variable that it is affecting, and so the perturbation on $X(Y)$ depends only on $Y(X)$ variables, the parameterization becomes especially attractive, as $M_1$ reduces to a constant, $M_2$ reduces to an additive noise term and $M_3$ has a greatly simplified integral kernel.

The choice that we have made for the parameterization of the unresolved variables is not unique. Indeed many other parameterizations yielding the same expectation values up to the second order may exist. The parameterization chosen here has the advantage of being fairly straightforward to derive from the different response terms that we have calculated. It is interesting that the parameterization obtained in such a generic way has the form of a generalized Langevin equation, which has been used to describe the effects of re-collisions with fluid particles in a Brownian motion [15, 36]. Similarly, colored noise and delayed differential equations have also proven to be useful parameterizations in climate science. For the former, see [34] and references therein; for the latter, see for example [11, 4]. The discussion in this paper suggests that for a weakly coupled system another approach is possible: by performing directly a linear response calculation, foregoing the parameterization.

It is important to note that in our derivation no assumptions have been made on the dynamics of $X$ and $Y$ systems in terms of differing time scales and/or of decorrelation time, our only hypothesis being the assumption that the coupling is weak in some sense: there is—a priori—symmetry between the $X$ and $Y$ systems; we just select for which one we want to derive an effective dynamics for. Actually, the characteristics of the unperturbed dynamics of the two systems are expected to play an important role in the
relative importance of the fluctuation and memory terms. In a system that is highly mixing, correlations in time disappear quickly and the fluctuation term will act as white noise. The memory term, on the other hand, will quickly drop off to zero in this case. If the system decorrelates slowly, there will be a more pronounced memory effect.

We want to stress that our method does not aim at describing within a certain accuracy and for a certain time horizon the individual orbits of the system of interest, but rather focuses on providing an approximate method for reconstructing its long term averages. Therefore, the method discussed in this paper has points of contact with and differences from techniques for the reduction of variables previously proposed in the literature such as the averaging method, which assumes a very large time scale separation but no limit on the strength of the coupling, and the projection operator method which is more formal and has less straightforward practical applicability.

Obviously, if the system consists of more than two levels it can be reduced to only one by repeatedly applying the approach described in this paper, so our method has a good degree of generality when reduction of variables is sought.

These findings will be of particular interest in the scientific fields dealing with multi-level systems and, in particular, in the context of climate modeling. First of all we derived an explicit strategy for substituting unresolved processes that is robust to changes in parameters or scales. Furthermore the parameters (such as the mean coupling, the noise correlations and the memory kernel) are explicitly formulated in terms of statistical properties of just the unresolved processes. By comparing the fluctuation and memory terms we see that different strategies will be applicable whether the unresolved process is fast or slow. Hence, this is relevant when modeling, e.g., the impact of the atmosphere on the oceans versus the other way around.

As future research, we will be implementing the proposed modeling strategies in low-dimensional systems, such as the Lorenz ’96 system [16]. We hope that this will lead to applications on more realistic models. Of particular interest will be how well this derivation based on weak couplings will apply to couplings in the climate simulations and whether the interplay between the fluctuations and memory can be observed. On a theoretical note, the connection with the theory of generalized Langevin equations will be explored further.

Acknowledgments

The authors acknowledge Matteo Colangeli for fruitful discussions and an anonymous referee for insightful comments.

Appendix. The response to perturbations with memory

As in [21], we derive the response of a discrete dynamical system

\[ x_{t+1} = f(x_t). \]

Consider a perturbation \( \chi \) that has a memory of the variable at \( l \) time steps in the past:

\[ \tilde{x}_{t+1} = f(\tilde{x}_t) + \chi_l(\tilde{x}_{t-l}). \]
We can expand the evolution of $\tilde{x}$ around the evolution of $x$. By induction, one can show that

$$\tilde{x}_{t+k} = f^k(\tilde{x}_t) + \sum_{j=1}^{k} D f^{k-j}(f^j(\tilde{x}_t)) \chi_l(\tilde{x}_{t-l+j-1}) + O(\chi^2), \quad (A.1)$$

or for an observable $A$ that

$$A(\tilde{x}_{t+k}) = A(f^k(\tilde{x}_t)) + \sum_{j=0}^{k} D (A \circ f^{k-j})(f^j(\tilde{x}_t)) \chi_l(\tilde{x}_{t-l+j-1}) + O(\chi^2). \quad (A.2)$$

From this we derive the following linear response:

$$\tilde{\rho}(A) = \lim_{T \to \infty} \frac{1}{T} \sum_{k=1}^{T} A(\tilde{x}_{t+k})$$

$$= \rho(A) + \frac{1}{T} \sum_{k=1}^{T} \sum_{j=1}^{k} \chi_l(f^{j-1}(\tilde{x}_{t-l})) D (A \circ f^{k-j})(f^{j+l}(\tilde{x}_{t-l})) + O(\chi^2) \quad (A.3)$$

$$= \rho(A) + \int \rho(dx) \sum_{i \geq 0} \chi_l(x) D (A \circ f^i)(f^{l+1}(x)) + O(\chi^2). \quad (A.4)$$

Let us consider the case where the memory dependence is given by a sum of memory terms at different times in the past:

$$\tilde{x}_{t+1} = f(\tilde{x}_t) + \sum_{l=0}^{\infty} \chi_l(\tilde{x}_{t-l}). \quad (A.5)$$

The response can be derived from equation (A.4) by linearity:

$$\delta \rho(A) = \int \rho(dx) \sum_{l=0}^{\infty} \sum_{i \geq 0} \chi_l(x) D (A \circ f^i)(f^{l+1}(x)) + O(\chi^2). \quad (A.6)$$

In the continuous case, for an equation of the form

$$\frac{d}{dt} \tilde{x}_t = F(\tilde{x}_t) + \int_0^\infty \chi_\tau(\tilde{x}_{t-\tau}) \quad (A.7)$$

this response becomes

$$\delta \rho_T = \int \rho(dx) \int_0^\infty ds \int_0^\infty d\tau \chi_\tau(x) \partial_{f^s(x)} (A \circ f^s)(f^\tau(x)) + O(\chi^2). \quad (A.8)$$
Disentangling multi-level systems: averaging, correlations and memory

References

[1] Abramov R V, *Suppression of chaos at slow variables by rapidly mixing fast dynamics through linear energy-preserving coupling*, 2012 Commun. Math. Sci. 10 595–624

[2] Abramov R V and Majda A, *New approximations and tests of linear fluctuation-response for chaotic nonlinear forced-dissipative dynamical systems*, 2008 J. Nonlinear Sci. 18 303

[3] Arnold L, *Hasselmann’s program revisited: the analysis of stochasticity in deterministic climate models*, 2001 Stoch. Clim. Models 49 141

[4] Bhattacharya K, Ghil M and Vulis I L, *Internal variability of an energy-balance model with delayed albedo effects*, 1982 J. Atmos. Sci. 39 1747

[5] Cessac B, *Does the complex susceptibility of the Hénon map have a pole in the upper-half plane? a numerical investigation*, 2007 Nonlinearity 20 2883

[6] Cessac B and Sepulchre J-A, *Linear response, susceptibility and resonances in chaotic toy models*, 2007 Physica D 225 13

[7] Cessac B and Sepulchre J-A, *Linear response, susceptibility and resonances in chaotic toy models*, 2007 Physica D 225 13

[8] Colangeli M, *Nonequilibrium response from the dissipative Liouville equation*, 2010 J. Stat. Mech. P12019

[9] Fatkullin I and Vanden-Eijnden E, *A computational strategy for multiscale systems with applications to Lorenz ’96 model*, 2004 J. Comput. Phys. 200 605

[10] Fraedrich K, Jansen H, Kirk E, Luksch U and Lunkeit F, *The Planet Simulator: towards a user friendly model*, 2005 Meteorol. Z. 14 299

[11] Ghil M and Childress S, 1987 Topics in Geophysical Fluid Dynamics: Atmospheric Dynamics, Dynamo Theory, and Climate Dynamics (Berlin: Springer)

[12] Hasselmann K, *Stochastic climate models part i. theory*, 1976 Tellus 28 473

[13] Imkeller P and von Storch J-S, 2001 Stochastic Climate Models (Progress in Probability) 1st edn (Basel: Birkhäuser)

[14] Kifer Y, *Some recent advances in averaging*, 2004 Modern Dynamical Systems and Applications (Cambridge: Cambridge University Press) p 385

[15] Kubo R, Toda M and Hashitsume N, 1988 Statistical Physics II: Nonequilibrium Statistical Mechanics (Berlin: Springer)

[16] Lorenz E N, 1996 *Predictability: A Problem Partly Solved* (GARP Publication Series vol 16) (Geneva: WMO) pp 132–6

[17] Lucarini V, *Response theory for equilibrium and non-equilibrium statistical mechanics: causality and generalized Kramers–Kronig relations*, 2008 J. Stat. Phys. 131 543

[18] Lucarini V, *Evidence of dispersion relations for the nonlinear response of the Lorenz ’63 system*, 2009 J. Stat. Phys. 134 381

[19] Lucarini V, *Modelling complexity: the case of climate science*, 2011 arXiv:1106.1265

[20] Lucarini V, *Stochastic perturbations to dynamical systems: a response theory approach*, 2012 J. Stat. Phys. online, doi:10.1007/s10955-012-0422-0

[21] Lucarini V, Kuna T, Wouters J and Faranda D, *Relevance of sampling schemes in light of Ruelle’s linear response theory*, 2012 Nonlinearity at press

[22] Lucarini V and Sarno S, *A statistical mechanical approach for the computation of the climatic response to general forcings*, 2011 Nonlinear Process Geophys. 18 7

[23] Orrell D, *Model error and predictability over different timescales in the Lorenz ’96 systems*, 2003 J. Atmos. Sci. 60 2219

[24] Paixoto J and Oort A, 1992 Physics of Climate (New York: AIP)

[25] Palmer T N and Williams P, 2009 Stochastic Physics and Climate Modelling (Cambridge: Cambridge University Press)

[26] Rahmstorf S et al, *Thermohaline circulation hysteresis: a model intercomparision*, 2005 Geophys. Res. Lett. 32 12

[27] Reich C H, *Linear response of the Lorenz system*, 2002 Phys. Rev. E 66 036103

[28] Ruelle D, *Differentiation of SRB states*, 1997 Commun. Math. Phys. 187 227

[29] Ruelle D, *Nonequilibrium statistical mechanics near equilibrium: computing higher-order terms*, 1998 Nonlinearity 11 5

[30] Ruelle D, *A review of linear response theory for general differentiable dynamical systems*, 2009 Nonlinearity 22 855

[31] Saltzman B, 2001 Dynamical Paleoclimatology, Generalized Theory of Global Climate Change 1st edn (New York: Academic)

doi:10.1088/1742-5468/2012/03/P03003
Disentangling multi-level systems: averaging, correlations and memory

[32] Solomon S (ed), 2007 Climate Change 2007: The Physical Science Basis (Cambridge: Cambridge University Press)

[33] Vallis G K et al, 2006 Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-scale Circulation (Cambridge: Cambridge University Press)

[34] Wilks D S, Effects of stochastic parametrizations in the Lorenz ’96 system, 2005 Q. J. R. Meteorol. Soc. 131 389

[35] Young L-S, What are SRB measures, and which dynamical systems have them?, 2002 J. Stat. Phys. 108 733

[36] Zwanzig R, 2001 Nonequilibrium Statistical Mechanics (Oxford: Oxford University Press)