Response formulae for n-point correlations in statistical mechanical systems and application to a problem of coarse graining

Response formulae for *n*-point correlations in statistical mechanical systems and application to a problem of coarse graining

2017 J. Phys. A: Math. Theor. 50 355003

(http://iopscience.iop.org/1751-8121/50/35/355003)

OPEN ACCESS
IOP Publishing
Journal of Physics A: Mathematical and Theoretical
J. Phys. A: Math. Theor. **50** (2017) 355003 (21pp)
https://doi.org/10.1088/1751-8121/aa812c

Response formulae for *n*-point correlations in statistical mechanical systems and application to a problem of coarse graining

Valerio Lucarini[1,2,3] **and Jeroen Wouters**[1,2,4]

[1] Department of Mathematics and Statistics, University of Reading, Reading, RG66AX, United Kingdom
[2] Centre for the Mathematics of Planet Earth, University of Reading, Reading, RG66AX, United Kingdom
[3] CEN, University of Hamburg, Hamburg, 20144, Germany
[4] School of Mathematics and Statistics, The University of Sydney, Sydney, Australia

E-mail: v.lucarini@reading.ac.uk

Received 22 February 2017, revised 14 July 2017
Accepted for publication 20 July 2017
Published 7 August 2017

Abstract

Predicting the response of a system to perturbations is a key challenge in mathematical and natural sciences. Under suitable conditions on the nature of the system, of the perturbation, and of the observables of interest, response theories allow to construct operators describing the smooth change of the invariant measure of the system of interest as a function of the small parameter controlling the intensity of the perturbation. In particular, response theories can be developed both for stochastic and chaotic deterministic dynamical systems, where in the latter case stricter conditions imposing some degree of structural stability are required. In this paper we extend previous findings and derive general response formulae describing how n–point correlations are affected by perturbations to the vector flow. We also show how to compute the response of the spectral properties of the system to perturbations. We then apply our results to the seemingly unrelated problem of coarse graining in multiscale systems: we find explicit formulae describing the change in the terms describing the parameterisation of the neglected degrees of freedom resulting from applying perturbations to the full system. All the terms envisioned by the Mori–Zwanzig theory—the deterministic, stochastic, and non-Markovian terms—are affected at first order in the perturbation. The obtained results provide a more comprehensive understanding of the response of statistical mechanical systems

to perturbations. They also contribute to the goal of constructing accurate and robust parameterisations and are of potential relevance for fields like molecular dynamics, condensed matter, and geophysical fluid dynamics. We envision possible applications of our general results to the study of the response of climate variability to anthropogenic and natural forcing and to the study of the equivalence of thermostatted statistical mechanical systems.

Keywords: response theory, correlations, coarse graining, invariant measure, parameterisation, reduced order model, thermostat

1. Introduction

1.1. Response theories

Understanding how a system responds to perturbations is a key challenge in mathematical and natural sciences and has long been the subject of extensive analysis through formal, experimental, and numerical investigations. A fundamental step in the direction of developing a comprehensive response theory can be found in the early work of Kubo (1957) (see also Kubo *et al* (1988)), who studied the impact of imposing weak perturbations to a statistical mechanical system originally at the thermodynamic equilibrium described by the canonical ensemble. While the proposed theory had been criticised from an early stage—see the famous argument by van Kampen (1971) as discussed in Marconi *et al* (2008)—it has been extremely successful in describing many physical phenomena (Lucarini *et al* 2005, Marconi *et al* 2008). The Kubo response theory leads to response formulae that express the change in the expectation value of a given observable Ψ of the system as a perturbative series. The zeroth order term is the expectation value of the observable Ψ in the unperturbed system, while the first order term, corresponding to the linear response, is expressed in terms of an explicitly determined causal Green's function, which contains comprehensive information on the interplay between the background dynamics of the system and the applied perturbation. It is important to note that the Green's function itself is constructed as an expectation value of an observable on the unperturbed measure, with the ensuing effect that the unperturbed system contains all the information needed for estimating its response to general forcings. This provides the basis for the cornerstone of Kubo's response theory, the so-called fluctuation-dissipation theorem (FDT), which links forced and free fluctuations in the linear perturbative regime. This structure extends to higher order terms with a simple generalisation, see e.g. Lucarini and Colangeli (2012)

A basic pitfall of Kubo's approach in terms of physical applicability is the impossibility of dealing with perturbations resulting from non-conservative forces. In fact, Kubo's theory does not allow for a consistent treatment of the energy budget of the perturbed system: in general, the external field will inject or subtract energy, so that in order to reach a well-defined steady state it is necessary to add a thermostat (Gallavotti 1997, Cohen and Rondoni 1998, Ruelle 2000). The natural question is then whether a specific choice of the thermostat alters the computed linear response. Fortunately, as shown in Evans and Morriss (2008), in the thermodynamic limit of a system with an infinite number of particles, the choice of the thermostat does not alter the predictions of linear response theory: the sensitivity of macroscopic observables does not depend on the details of the microscopic dynamics.

What is also unsatisfactory about the Kubo response theory is that mathematical rigour has been missing in establishing whether the many limits involved in constructing the response formulae are well defined. Additionally, no provision is given for computing the response of nonequilibrium systems to perturbations.

Ruelle (1997, 1998) and (2009) showed that it is possible to establish a rigorous response theory for Axiom A maps and flows, which possess invariant Sinai–Ruelle–Bowen (SRB) measures. In other terms, Ruelle showed that in the case of Axiom A systems the invariant measure is differentiable with respect to the parameters controlling small modifications to the flow of the system, and provided explicit expressions for the linear and higher order contributions to the response.

Axiom A systems are indeed far from being typical dynamical systems, but, according to the *chaotic hypothesis* formulated by Gallavotti and Cohen (1995) and Gallavotti (1996), they can be taken as effective models for chaotic dynamical systems with many degrees of freedom. Specifically, this means that when looking at macroscopic observables in *sufficiently* chaotic (to be intended in a qualitative sense) high-dimensional systems, it is expected that it is extremely hard to distinguish their properties from those of an Axiom A system, including some degree of structural stability. Note that the chaotic hypothesis can be seen as the natural extension of the ergodic hypothesis, which is the fundamental heuristic step needed to apply results of equilibrium statistical mechanics to interpret and predict the properties of real systems at equilibrium. Linear response is therefore expected to hold in practice for very general dynamical systems, while the known counter-examples are currently limited to low-dimensional non-uniformly expanding maps (Baladi and Smania 2008, Gottwald *et al* 2016).

Axiom A systems corresponding to equilibrium physical systems possess an invariant measure that is absolutely continuous with respect to the Lebesgue measure because the phase space does not contract nor expand, as the flow is nondivergent. Axiom A systems featuring—on the average—a contraction in the phase space provide excellent mathematical models for nonequilibrium systems (Gallavotti 2006). In this case, the invariant measure lives on a set with a Hausdorff dimension lower than the number of degrees of freedom of the system and is singular with respect to the Lebesgue measure, as a result of the contraction taking place in the stable manifold (Eckmann and Ruelle 1985). Despite the geometrical complexity associated to the attractors of nonequilibrium systems, the Ruelle response theory, somewhat surprisingly, ensures that differentiability can be established also in this case.

In the case of an equilibrium system, the Ruelle response theory allows for deriving the FDT. In nonequilibrium systems, instead, there is no one-to-one correspondence between forced and free fluctuations, as already suggested by Lorenz (1979): Ruelle (1997, 1998) and (2009) provides a mathematical explanation of this property, while a physical interpretation is given in, e.g. Lucarini (2008, 2009) and Lucarini and Sarno (2011). The basic idea is that while the natural fluctuations are able to mimick the effect of the components of the forcing along the unstable manifold of the system, the impact of the components of the forcing along the stable manifold have no counterpart in the unperturbed system.

Interestingly, while on one side there have been positive examples of applications of the FDT in nonequilibrium systems, like the climate, it is clear that, for a given class of forcing, the quality of the obtained response operator depends substantially on the chosen observable (Gritsun and Branstator 2007, Gritsun *et al* 2008). In a recent paper, Gritsun and Lucarini (2017) have provided examples in a system of geophysical relevance of various scenarios supporting or not the applicability of the FDT to reconstruct the response of the system to perturbations. They have clearly shown that, indeed, when the applied forcing has a relevant projection on the stable manifold of the unperturbed system, the forced variability can have little resemblance to the natural one. In particular, the forcing can in some cases excite resonances corresponding to special dynamical features that are virtually unexplored by the unperturbed system, so that one can observe so called *climatic surprises*.

The difficulties in constructing *ab initio* the response operator using Ruelle's formulae come from the extremely different behaviour of the contribution coming from the unstable and

stable manifold (Abramov and Majda 2007). The computation of the contributions coming from the stable directions give neither numerical nor conceptual problems. When the unstable directions are considered, problems emerge from the fact that contributions to the response come from integrals over time of exponentially growing functions, resulting from the presence of sensitive dependence on initial conditions. The ill–posedness of this operation is at the core of the van Kampen (1971) criticism mentioned above. On the other side, response operators, as described in the next section, are constructed by integrating over the statistical ensemble of the (unpertubed) system. Such an operation—under suitable conditions—regularises the previous divergences and explains why linear response is indeed well-posed. Nonetheless, obtaining in practice a stable estimate of the response operators from a finite number of ensemble members and from finite numerical simulations is far from obvious. We note that algorithms based on adjoint methods seem to partially ease these issues (Eyink *et al* 2004, Wang 2013).

Convincingly good results in terms of climate prediction performed using the linear response theory have instead been obtained through bypassing the problem of constructing the response operator and using, instead, the formal properties of the Green's function (Lucarini and Sarno 2011, Lucarini *et al* 2014, Ragone *et al* 2016, Lucarini *et al* 2017). Tests in simple models have emphasized that also the nonlinear response theory is extremely solid and amenable to numerical verification (Lucarini 2009).

Modern methods of spectral theory have provided different and elegant proofs and further generalisations of Ruelle's results. The response theory can be developed by comparing the Perron-Frobenius transfer operator (Baladi 2000) of the unperturbed and of perturbed system, thus focussing on the evolution of ensembles rather than of individual trajectories—see e.g. Liverani and Gouëzel (2006), Baladi and Smania (2008) and Baladi *et al* (2014). This approach has allowed the extension of Ruelle's results to systems more general than the Axiom A case, by constructing suitable Banach space of anisotropic distributions. The practical applicability of transfer operator-based methods for studying the response in high dimensional systems is still not entirely clear, as a result of the *curse of dimensionality*, even if some optimism comes from the overall positive results obtained when severely reduced order models are considered (Tantet *et al* 2015b, 2015a). Additionally, ideas borrowed from the theory of the transfer operator have proved extremely useful for studying the behaviour of geophysical systems in the vicinity of critical transitions, where the response theory breaks apart, decorrelation times become very long, and the presence of Ruelle–Pollicott resonances lead to the appearance of rough dependence of the system properties on the perturbation parameter (Chekroun *et al* 2014). Recently, explicit formulae based on simple matrix algebra have been proposed for computing the response of a finite state Markov chain to perturbations, thus providing a model for studying finer and finer partitions of actual phase spaces (Lucarini 2016).

A different way to approach the problem of constructing a response theory can be followed by taking the point of view of stochastic dynamics, as proposed initially by Hänggi and Thomas (1975) and Hänggi and Thomas (1977); see a recent review by Baiesi and Maes (2013). Adding (suitably chosen, typically gaussian white) noise on top of the deterministic dynamics allows to deal with invariant measures that are absolutely continuous with respect to Lebesgue and for making sure that the decay of correlations in the system is fast. As a result, some of the mathematical issues discussed above are automatically sorted out and, in particular, the FDT holds in all cases. Thanks to the presence of noise it is possible to set a general framework for linear response theory in much greater regularity, including the case of infinite dimensional systems; see Hairer and Majda (2010) for a mathematically accurate study of linear response for stochastic system, where many subtleties are sorted out. One needs to note, though, that while the presence of noise smoothens the invariant measure of the system, the weaker the noise, the harder it is for a numerical model to appreciate such

smoothness given the finite length numerical simulations and the finite size of the ensemble of performed simulations.

1.2. Parameterisation of a coarse grained model: stochasticity and memory effects

Adding stochastic forcings on top of the deterministic dynamics should be justified on physical grounds and not used just as an ad hoc assumption. A convincing way to motivate the introduction of a random component to the dynamics comes from the need of taking into account the effect of *microscopic*, unresolved scales; see a mathematically rigorous and complete treatment in Chekroun *et al* (2015a) and (2015b). Along the lines of the early results by Zwanzig (1961) and Mori (1965), Chekroun *et al* (2015a) and (2015b) also clearly show that the construction of reduced order models unavoidably leads also to introducing non-Markovian terns in the surrogate dynamics of the variables of interest.

The problem of constructing accurate and robust parameterisations for degrees of freedom that are hard to simulate explicitly is a crucial problem in a variety of scientific fields, and most notably in condensed matter physics (Bhalla *et al* 2016), molecular dynamics (Shinoda *et al* 2007, Baron *et al* 2007, Kmiecik *et al* 2016), and in geophysical fluid dynamics (Franzke *et al* 2015, Berner *et al* 2016).

The situation in the case of atmospheric, ocean, and climate models is particularly complex because there is no clear gap (in terms of temporal and spatial scales) in variability of the fluid motions (Ghil and Childress 1987, Peixoto and Oort 1992, Lucarini *et al* 2014). As a result, first, the approximation of infinite time separation between resolved and unresolved scales is unsatisfactory, so that the standard homogenisation theory (Pavliotis and Stuart 2008) cannot be safely applied in this case. As a result, on one side the stochastic terms in the parameterisation cannot be represented as white noise, and the presence of memory effects leads additionally to the need to incorporate, in principle, non-Markovian terms in the dynamics.

Additionally, given the available numerical resolution at hand, one always faces the problem of dealing with the so-called *grey zone*, a range of scales where physical processes are only partially resolved (Gerard 2007). Further, the parameterisation depends on where one defines the cutoff between resolved and unresolved scales of motion (practically often determined by the computational facilities at hand and/or the required length or number of the model runs), so that a painstaking process of tuning is in principle necessary each time the resolution of the model needs to be changed. As a result, the quest for self-adaptive parameterisation has been recently emphasized in the literature, see e.g. Arakawa *et al* (2011), Park (2014) and Sakradzija *et al* (2016). Self-adaptivity is crucial for the goal of constructing models able to perform seamless prediction, i.e. to be used for weather forecast, seasonal prediction, and climate modelling (Palmer *et al* 2008).

As for the scopes of this paper, it is relevant to note that one can use the Ruelle response theory to compute explicitly the effect of small scale, fast degrees on freedom on the macroscopic ones. In this case, the perturbation one studies using the results by Ruelle is exactly the coupling between the dynamics occurring at the different scales. One discovers that it is possible to derive an explicit parameterisation providing a deterministic, a stochastic, and a non-Markovian contribution to the dynamics of the variables of interest, thus obtaining a perturbative yet self-consistent closure to the problem (Wouters and Lucarini 2012, Wouters and Lucarini 2013, Wouters and Lucarini 2016). The various terms are constructed in terms of specific response operators at first and second order. Some first promising examples of applications of the theory and investigation of the skills of the parameterisation schemes have been recently presented in models of various degrees of complexity (Wouters *et al* 2016, Vissio and Lucarini2016, Demaeyer and Vannitsem 2017).

1.3. This paper

In this paper we set ourselves in the context of (possibly high-dimensional) chaotic deterministic dynamical systems, assume the chaotic hypothesis and, consequently, the applicability of the Ruelle response theory. We expect, nonetheless, that our results should apply also in the case of stochastic dynamics, apart from obvious changes in the notation. This paper has a twofold purpose and addresses an interdisciplinary audience.

We first take a rather general point of view and note that most of the theoretical results presented in the literature focus on assessing the response of the system to perturbations in terms of changes of the expectation values of suitably defined observables, or, equivalently, of the invariant measure. This statement applies to both more heuristic and more rigorous studies, and both to approaches based on the framework of deterministic or stochastic dynamics. The *elephant in the room* is, in our view, the lack (at least up to the authors' knowledge) of general explicit formulae predicting how the time-lagged correlations of observables change as a result of perturbations to the dynamics. Therefore, in this paper we provide explicit linear response formulae for $n-$point time correlations of observables. As discussed below, in the general case treated here the response formulae become more involved than in the usual case of observables and one derives new terms that cannot be framed, even in the case of unperturbed systems possessing smooth invariant measure, in terms of the FDT. The possibility of having formulae for studying the response of higher order moments is quite attractive because it paves the way to asking how the statistical properties of the fluctuations of the system (or, equivalently, its spectral properties) change as a result of the applied perturbation. In the specific case of climate dynamics, which is an application of special interest for the authors, this amounts to being able to address the question of how the climate variability changes in response to climate forcing (Ghil 2015). This is a major and indeed open problem in the climate literature.

We then discuss a—seemingly unrelated—problem of interdisciplinary relevance, which was, in fact, the original driver of the investigation presented in this paper. We look into the problem of constructing reduced order models for multiscale systems and take advantage of the fact that, as mentioned above, it can be framed as an indeed nontrivial exercise that can be studied using response theory. Finding an accurate and efficient way to perform coarse graining in multiscale systems amounts to constructing a parameterised dynamics for the variables of interest (usually the large scale, slow ones) and is key to supporting the development of practically usable numerical models. A much desired quality of a parameterisation is its adaptivity with respect to changes in the properties of the system. In previous publications (Wouters and Lucarini 2012, Wouters and Lucarini 2013, Wouters and Lucarini 2016) we have introduced a general method for constructing parameterisations whose main advantage is its adaptivity to the parameters describing the coupling and/or the time scale separation between the slow and fast scale of motion, whose lack is, instead, a key drawback of many other methods, and especially of the empirical ones. A basic issue, both at practical and at theoretical level, is to assess the robustness of a parameterisation with respect to small changes in the dynamics of the system. In this paper, using the general results mentioned above, we are able to construct a *response theory for the reduced order, coarse grained model*, and derive explicit formulae for the change of the various terms composing the parameterisation. This has relevance for the goal of constructing parameterisations able to adjust to small changes in the dynamics of the *full* system. Note that such perturbations can also be considered as a representation of the model error: in this case, our results address the problem of understanding how the model error translates in the formulation of the reduced order model.

Being the numerical implementation and analysis of the response-based parameterisation a topic that is in full development, the current extension of the theory consists mostly of formal calculations, at this stage. Numerical studies on specific systems of interest will be the subject of future investigations.

The paper is organised as follows. In section 2 we show how the response formulae are changed when the observable we are considering is also a function of the small parameter controlling the intensity of the forcing. We then use such a result to present the extension of the response theory for the case of $n-$point correlations. We show in detail the calculations needed to reach general formulae that include, as special case, the usual response formulae for observables. The results contained in section 2 might be of interest for experts in dynamical systems and statistical mechanics. In section 3 we recapitulate how to construct parameterisations allowing one to perform consistently coarse graining on multiscale systems and we show how the theory developed in section 2 allows one to find explicit formulae for the corrections to the parameterisations due to a perturbation applied to the full system. The results contained in section 3 might be additionally of interest for scientists interested in specific applications of coarse graining methods, such as those working on the development of parameterisations for describing the coarse grained dynamics of systems of interest for, e.g. molecular dynamics or geophysical fluid dynamics. In section 4 we discuss our results and present our conclusive remarks. In the appendix we present some ideas possibly relevant for the study of thermostatted systems.

2. A simple extension of the standard response theory

Let's consider a continuous time Axiom A dynamical system (Eckmann and Ruelle 1985, Ruelle 1989) defined on a compact n-dimensional manifold $\mathcal{M}$ of the form

$$\dot{\vec{x}} = \vec{F}(\vec{x}) \tag{1}$$

possessing a physical invariant measure ρ_0. We frame our results below in the setting of deterministic dynamical systems but we stress that equivalent equations will hold for stochastic differential equations.

The expectation value of a general observable $\Phi_0(\vec{x})$ on such a measure can be written as $\int_{\mathcal{M}} \rho_0(\mathrm{d}\vec{x})\Phi_0(\vec{x})$. We can also write the expectation value in a more compact form as $\rho_0(\Phi)$ or as $\langle \rho_0, \Phi_0 \rangle$, where we stress that the expectation value is the result of applying a linear functional (the measure ρ_0) to the measurable function Φ_0.

Let $\vec{x}(t, \vec{x}_0)$ be the flow from an initial condition $\vec{x}_0$, i.e. $\vec{x}(0, \vec{x}_0) = x_0$ and $\vec{x}(t, \vec{x}_0)$ satisfies (1). Then the Koopman operator Π_0 is the composition of an observable with the flow: $(\Pi_0(t)\Phi)(\vec{x}_0) = \Phi(x(t, x_0))$. Under suitable conditions, one can express the Koopman operator as $\Pi_0(t) = \exp(\mathcal{L}_{(0)}t)$, where $\mathcal{L}_{(0)} = \vec{F} \cdot \vec{\nabla}$ is such that $\dot{\Psi} = \mathcal{L}_{(0)}\Psi$ for all differentiable functions $\Psi = \Psi(\vec{x})$. The Perron–Frobenius–Ruelle operator is the adjoint of the Koopman operator $\Pi_0^\top(t)$ and defines the push-forward of an initial measure ρ^* so that $\rho(t, \rho^*) = \Pi_0^\top(t)\rho^*$, defined as follows:

$$\int_{\mathcal{M}} \rho^*(\mathrm{d}\vec{x}_0)\Phi_0(\vec{x}(t, x_0)) = \langle \rho^*, \Pi_0(t)\Phi_0 \rangle = \langle \Pi_0^\top(t)\rho^*, \Phi_0 \rangle = \langle \rho(t, \rho^*), \Phi_0 \rangle = \int_{\mathcal{M}} \Pi_0^\top(t)\rho^*(\mathrm{d}\vec{x}_0)\Phi_0(\vec{x}_0). \tag{2}$$

Note that we have $\Pi_0^\top(t) = \exp(\mathcal{L}_{(0)}^\top t)$, with $\mathcal{L}_{(0)}^\top \rho^* = \vec{\nabla} \cdot (\vec{F}\rho^*)$. Additionally, by definition, we have $\Pi_0^\top(t)\rho_0 = \rho_0$ and, correspondingly, $\mathcal{L}_{(0)}^\top \rho_0 = 0$.

Let's now consider a small $\epsilon-$perturbation to the vector flow of the form

$$\dot{\vec{x}} = \vec{F}(\vec{x}) + \epsilon\vec{G}(\vec{x}) \tag{3}$$

so that the perturbed flow possesses an invariant measure ρ_ϵ, and one can define the perturbed Liouville operator as $\mathcal{L}_\epsilon = \mathcal{L}_{(0)} + \epsilon\mathcal{L}_{(1)}$, where $\mathcal{L}_{(1)} = \vec{G}\cdot\vec{\nabla}$. We also define the perturbed evolution and Perron-Frobenius-Ruelle operators as $\Pi_\epsilon(t) = \exp(\mathcal{L}_\epsilon t)$ and $\Pi_\epsilon^\top(t) = \exp(\mathcal{L}_\epsilon^\top t)$, respectively.

It is of clear relevance to be able to say under which conditions for small values of ϵ it is possible to expand $\langle\rho_\epsilon, \Phi_0\rangle_\epsilon$ as follows:

$$\langle\rho_\epsilon, \Phi_0\rangle = \langle\rho_0, \Phi_0\rangle + \epsilon\frac{\mathrm{d}}{\mathrm{d}\epsilon}\langle\rho_\epsilon, \Phi_0\rangle|_{\epsilon=0} + \text{h.o.t.} \tag{4}$$

where h.o.t. indicates higher order terms, and to find an explicit expression for the key quantity $\frac{\mathrm{d}}{\mathrm{d}\epsilon}\langle\rho_\epsilon, \Phi_0\rangle|_{\epsilon=0}$, which controls the first order correction of the expectation value. The Ruelle response theory says that if the unperturbed dynamical system $\dot{\vec{x}} = \vec{F}(\vec{x})$ is Axiom A and we consider a C^3 observable $\Phi_0(\vec{x})$, one can write

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\langle\rho_\epsilon, \Phi_0\rangle|_{\epsilon=0} = \int_0^\infty \mathrm{d}\tau\langle\rho_0, \mathcal{L}_{(1)}\exp\left(\mathcal{L}_{(0)}\tau\right)\Phi_0\rangle, \tag{5}$$

so that one can alternatively write $\rho_\epsilon = \rho_0 + \epsilon\frac{\mathrm{d}}{\mathrm{d}\epsilon}\rho_\epsilon|_{\epsilon=0} + \text{h.o.t.}$ where

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\rho_\epsilon|_{\epsilon=0} = \int_0^\infty \mathrm{d}\tau\Pi_0^\top(t)\mathcal{L}_{(1)}^\top\rho_0; \tag{6}$$

we write in this case $\frac{\mathrm{d}}{\mathrm{d}\epsilon}\langle\rho_\epsilon, \Phi_0\rangle|_{\epsilon=0} = \langle\frac{\mathrm{d}}{\mathrm{d}\epsilon}\rho_\epsilon|_{\epsilon=0}, \Phi_0\rangle$.

Note that if $\mathcal{L}_{(1)} = \mathcal{L}_{(0)}$, so that the perturbation is just a linear change in the time variable $t \to t(1+\epsilon)$, we have that $\frac{\mathrm{d}}{\mathrm{d}\epsilon}\rho_\epsilon|_{\epsilon=0} = 0$ because $\mathcal{L}_{(1)}^\top\rho_0 = \mathcal{L}_{(0)}^\top\rho_0 = 0$, from the definition of ρ_0. Note that rescaling time does not affect the expectation value of any observable at all orders of perturbations.

It is easy to generalise the problem to the case where the observable is a C^1 function of ϵ so that one can write the following expansion for small values of ϵ: $\Phi_\epsilon = \Phi_0 + \epsilon\frac{\mathrm{d}}{\mathrm{d}\epsilon}\Phi_\epsilon|_{\epsilon=0} + \text{h.o.t.}$. In this case, we have that

$$\langle\rho_\epsilon, \Phi_\epsilon\rangle = \langle\rho_0, \Phi_0\rangle + \epsilon\frac{\mathrm{d}}{\mathrm{d}\epsilon}\langle\rho_\epsilon, \Phi_\epsilon\rangle|_{\epsilon=0} + \text{h.o.t.} \tag{7}$$

where the linear sensitivity can be expressed as:

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\langle\rho_\epsilon, \Phi_\epsilon\rangle|_{\epsilon=0} = \langle\frac{\mathrm{d}}{\mathrm{d}\epsilon}\rho_\epsilon|_{\epsilon=0}, \Phi_0\rangle + \langle\rho_0, \frac{\mathrm{d}}{\mathrm{d}\epsilon}\Phi_\epsilon|_{\epsilon=0}\rangle, \tag{8}$$

where the first term corresponds to the usual response theory, and comes from the change of the dynamics of the system, while second term comes from the change of the definition of the observable as a function of ϵ.

Let's take a first simple and relevant example to illustrate the meaningfulness of this result. We consider as observable the divergence of the flow $\Phi_\epsilon = \vec{\nabla}\cdot(\vec{F} + \epsilon\vec{G})$ in equation (3). The expectation value of this observable is equal to the sum of the Lyapunov exponents of the system and can be interpreted as the opposite of its entropy production (Ruelle 1989, Gallavotti 2014). We have that

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\langle\rho_\epsilon, \Phi_\epsilon\rangle|_{\epsilon=0} = \int_0^\infty \mathrm{d}\tau\langle\rho_0, \mathcal{L}_{(1)}\Pi_0(\tau)(\vec{\nabla}\cdot\vec{F})\rangle + \langle\rho_0, \vec{\nabla}\cdot\vec{G}\rangle. \tag{9}$$

If the expectation value on the unperturbed measure of the divergence of perturbation flow is zero (or *a fortiori* if the perturbation flow is divergence-free), the second term vanishes. See the appendix for a discussion on the physical interpretation of equation (9).

2.1. Derivation of response formulae for n-point correlations

2.1.1. Two-point correlations We now consider the product of the value two observables Ψ_a and Ψ_b taken as different times, i.e. without loss of generality $c_{\Psi_a,\Psi_b}(t) := \Psi_a(\vec{x})\Psi_b(\vec{x}(t))$. The expectation value of $c_{\Psi_a,\Psi_b}(t)$, is $C_{\Psi_a,\Psi_b}(t) = \langle \rho_0, c_{\Psi_a,\Psi_b}(t)\rangle$, the $t-$lagged correlation between Ψ_a and Ψ_b. The local quantity $c_{\Psi_a,\Psi_b}(t)$ measures the joint fluctuations of the two observables Ψ_a and Ψ_b at different times but along the same orbit.

We consider the perturbed flow given in equation (3). The product $\Psi_a(\vec{x})\Psi_b(\vec{x}(t))$ can be written as $\Psi_a(\vec{x})\Pi_\epsilon(t)\Psi_b(\vec{x})$, so that we must add a lower index ϵ to the expressions $c_{\Psi_a,\Psi_b,\epsilon}(t)$ and to $C_{\Psi_a,\Psi_b,\epsilon}(t)$.

In order to obtain an expression for $\frac{\mathrm{d}}{\mathrm{d}\epsilon}c_{\Psi_a,\Psi_b,\epsilon}(t)|_{\epsilon=0}$, we need to expand the Koopman operator for small values of ϵ. Using the Dyson formalism, we have:

$$\Pi_\epsilon(t) = \exp\left(\mathcal{L}_{(0)}t + \epsilon\mathcal{L}_{(1)}t\right) = \Pi_{(0)}(t) + \epsilon\int_0^t \mathrm{d}\tau_2 \Pi_{(0)}(t-\tau_2)\mathcal{L}_{(1)}\Pi_{(0)}(\tau_2) + \text{h.o.t.}, \tag{10}$$

where h.o.t. indicates terms featuring higher powers of the parameter ϵ. Note that the term proportional to ϵ in the right hand side of the previous equation is instrumental for deriving the desired result. We then have that the linear response of the $t-$lagged time correlation between the two observables Ψ_a and Ψ_b can be written as:

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}C_{\Psi_a,\Psi_b,\epsilon}(t)|_{\epsilon=0} = \frac{\mathrm{d}}{\mathrm{d}\epsilon}\langle\rho_\epsilon, c_{\Psi_a,\Psi_b,\epsilon}(t)\rangle|_{\epsilon=0} = \langle\frac{\mathrm{d}}{\mathrm{d}\epsilon}\rho_\epsilon|_{\epsilon=0}, c_{\Psi_a,\Psi_b,0}(t)\rangle + \langle\rho_0, \Psi_a\frac{\mathrm{d}}{\mathrm{d}\epsilon}\Pi_\epsilon(t)\Psi_b|_{\epsilon=0}\rangle. \tag{11}$$

The first term on the right hand side gives to the correction of the local (in phase space) fluctuations computed according to the unperturbed dynamics due to the fact that the perturbation flow modifies the invariant measure, and corresponds to what one would obtain with a naive application of the response theory for studying the change in the correlations of the system. The second term corresponds to the expectation value on the unperturbed dynamics of the change in the evolution law due to the presence of the $\epsilon-$perturbation.

In particular, we can write the first term as:

$$\begin{aligned}\langle\frac{\mathrm{d}}{\mathrm{d}\epsilon}\rho_\epsilon|_{\epsilon=0}, c_{\Psi_a,\Psi_b,0}(t)\rangle &= \langle\rho_0, \int_0^\infty \mathrm{d}\tau_1 \mathcal{L}_{(1)}\Pi_{(0)}(\tau_1)\Psi_a\Pi_{(0)}(t)\Psi_b\rangle \\ &= \int_0^\infty \mathrm{d}\tau_1 \int_{\mathcal{M}} \rho_0(\mathrm{d}\vec{x})\vec{G}(\vec{x})\cdot\vec{\nabla}_{\vec{x}}(\Psi_a(\vec{x}(\tau_1))\Psi_b(\vec{x}(t+\tau_1))).\end{aligned} \tag{12}$$

Comparing with Colangeli and Lucarini (2014), we observe that this expression resembles a second order response term for regular observables, but, thanks to the presence of a slightly simpler functional form, can be brought to a FDT-like form by applying the operator $\mathcal{L}^\top_{(1)}$ to the unperturbed invariant measure ρ_0:

$$\langle\frac{\mathrm{d}}{\mathrm{d}\epsilon}\rho_\epsilon|_{\epsilon=0}, c_{\Psi_a,\Psi_b,0}(t)\rangle = \int_0^\infty \mathrm{d}\tau_1 \langle\mathcal{L}^\top_{(1)}\rho_0, \Pi_{(0)}(\tau_1)\Psi_a\Pi_{(0)}(t)\Psi_b\rangle, \tag{13}$$

where we have an integral over one time variable of a three-point correlation.

Instead, the second term in equation (11) can be written as:

$$\langle \rho_0, \Psi_a \frac{\mathrm{d}}{\mathrm{d}\epsilon}\Pi_\epsilon(t)\Psi_b|_{\epsilon=0}\rangle = \langle \rho_0, \Psi_a \int_0^t \mathrm{d}\tau_2 \Pi_{(0)}(t-\tau_2)\mathcal{L}_{(1)}\Pi_{(0)}(\tau_2)\Psi_b\rangle$$
$$= \int_0^t \mathrm{d}\tau_2 \int_{\mathcal{M}} \rho_0(\mathrm{d}\vec{x})\Psi_a(\vec{x})\vec{G}(\vec{x}(t-\tau_2))\cdot\vec{\nabla}_{\vec{x}(t-\tau_2)}(\Psi_b(\vec{x}(t))). \quad (14)$$

Note that this term vanishes if $t = 0$ because in this case the function $c_{\Psi_a,\Psi_b,\epsilon}(t=0)$ is not anymore a function of ϵ, and the usual response theory formulae for simple observables apply. Due to the presence of a different time ordering in the operators, we cannot reframe equation (14) in a FDT-like form.

We also wish to note that if the system is mixing and has rapid decay of correlations, both terms given in the right hand side of equations (12)–(14) will tend to zero for large values of t.

In order to have a simple consistency test of our results, let's also take the special case seen above where $\mathcal{L}_{(1)} = \mathcal{L}_{(0)}$, i.e. we rescale the time variable $t \to t(1+\epsilon)$. In this case, the first term given in equation (12) vanishes, because $\mathcal{L}^\top_{(0)}\rho_0 = 0$. This corresponds to what discussed before when looking at the response theory for observables.

Instead, the second term reads $t\int_{\mathcal{M}} \rho_0(\mathrm{d}\vec{x})\Psi_a(\vec{x})\vec{F}(x(\vec{t}))\cdot\vec{\nabla}_{\vec{x}(t)}(\Psi_b(\vec{x}(t)))$. The (trivial) fact that rescaling time leads to a change in the correlations functions can be immediately derived by observing that

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\langle\rho_0, \Psi_a(\vec{x})\Psi_b(\vec{x}(t(1+\epsilon)))\rangle|_{\epsilon=0} = t\langle\rho_0, \Psi_a(\vec{x})\dot{\vec{x}}(\vec{x}(t))\cdot\vec{\nabla}\Psi_b(\vec{x}(t))\rangle = t\langle\rho_0, \Psi_a(\vec{x})\vec{F}(\vec{x}(t))\cdot\vec{\nabla}\Psi_b(\vec{x}(t))\rangle. \quad (15)$$

just as obtained above.

2.1.2. The general case of n-point correlations We now consider the case of general correlation functions. Take

$$c_{\Psi_0,\Psi_1,\ldots,\Psi_{n-1}}(s_1,s_2,\ldots,s_{n_1}) = \Psi_0(\vec{x})\Psi_1(\vec{x}(s_1))\ldots\Psi_{n-1}(\vec{x}(s_1+\ldots+s_{n-1})) \quad (16)$$

and define the n-point correlation function for the perturbed system as:

$$C_{\Psi_0,\Psi_1,\ldots,\Psi_{n-1};\epsilon}(s_1,s_2,\ldots,s_{n-1}) = \langle\rho_\epsilon, c_{\Psi_0,\Psi_1,\ldots,\Psi_{n-1}}(s_1,s_2,\ldots,s_{n-1})\rangle$$
$$= \langle\rho_\epsilon, \Psi_0(\vec{x})\Pi_\epsilon(s_1)\Psi_1(\vec{x})\Pi_\epsilon(s_2)\Psi_2(\vec{x})\ldots\Pi_\epsilon(s_{n-1})\Psi_{n-1}(\vec{x})\rangle\ . \quad (17)$$

We can then construct the following first order expansion for the n-point correlation as follows:

$$C_{\Psi_0,\Psi_1,\ldots,\Psi_{n-1};\epsilon}(s_1,s_2,\ldots,s_{n-1}) = C_{\Psi_0,\Psi_1,\ldots,\Psi_{n-1};0}(s_1,s_2,\ldots,s_{n-1})$$
$$+\epsilon\frac{\mathrm{d}}{\mathrm{d}\epsilon}C_{\Psi_0,\Psi_1,\ldots,\Psi_{n-1};\epsilon}(s_1,s_2,\ldots,s_{n-1})|_{\epsilon=0} + \text{h.o.t.}$$

The term proportional to ϵ is given by the sum of n terms, the first one resulting from the linear correction to the measure, which corresponds to what one would naively obtain by applying the standard response theory, and the other $n-1$ terms resulting from the linear correction to each of the $n-1$ Koopman operators appearing in the definition of the n-point correlation function. We have:

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}C_{\Psi_0,\Psi_1,\ldots,\Psi_{n-1};\epsilon}(s_1,s_2,\ldots,s_{n-1})|_{\epsilon=0} = \int_0^\infty \mathrm{d}\tau\langle\rho_0, \mathcal{L}_1\Pi_0(\tau)\Psi_0(\vec{x})\ldots\Pi_0(s_{n-1})\Psi_{n-1}(\vec{x})\rangle$$
$$+\sum_{k=1}^{n-1}\int_0^{s_k}\mathrm{d}\tau\langle\rho_0, \Psi_0(\vec{x})\ldots\Pi_0(s_k-\tau)\mathcal{L}_1\Pi_0(\tau)\Psi_k(\vec{x})\ldots\Pi_0(s_{n-1})\Psi_{n-1}(\vec{x})\rangle. \quad (18)$$

As seen in the case of two-point correlations, the first term can be brought to a FDT-like form by applying the operator $\mathcal{L}_{(1)}^{\top}$ to the unperturbed invariant measure ρ_0, while the other terms have a more convolute structure.

2.1.3. Change in the spectral properties of the system We can use the results presented before to draw interesting conclusions on how the spectral properties of the system under investigation change as a result of the $\epsilon-$ perturbation. Under suitable conditions of integrability, we have that $\mathcal{F}[C_{\Psi,\Phi}(t)] = P(\Psi,\Phi) = \mathcal{F}[(\Psi)]^*\mathcal{F}[(\Phi)]$, where $\mathcal{F}[g] = \hat{g}$ is the Fourier transform of g and f^* is the complex conjugate of f. With $P(\Psi,\Phi) = P(\Phi,\Psi)^*$ we indicate the cross-spectrum of the two functions Ψ and Φ (note the effect of the time lag). In particular, we have that if $\Psi = \Phi$, $\mathcal{F}[C_{\Psi,\Psi}(t)] = |\mathcal{F}[(\Psi)]|^2 = |\hat{\Psi}|^2 = P(\Psi,\Psi)$, which corresponds to the Khinchin-Wiener theorem. Thanks to the linearity of the Fourier transform, we can then derive the following expression from equation (11):

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}P_\epsilon(\Psi_a,\Psi_b)|_{\epsilon=0} = \mathcal{F}\left[\langle\frac{\mathrm{d}}{\mathrm{d}\epsilon}\rho_\epsilon|_{\epsilon=0}, c_{\Psi_a,\Psi_b,0}(t)\rangle\right] + \mathcal{F}\left[\langle\rho_0,\Psi_a\frac{\mathrm{d}}{\mathrm{d}\epsilon}\Pi_\epsilon(t)\Psi_b|_{\epsilon=0}\rangle\right], \tag{19}$$

where we have added a lower index to the cross-spectrum P in order to keep track of the presence of the ϵ-perturbation to the dynamics. Equation (19) provides the answer to the quite relevant question of how the spectral properties of the system change as a result of the presence of perturbations. Note that the first term on the right hand-side of equation (19) can be interpreted as cross-spectrum of the same observables Ψ_a and Ψ_b where the time statistics is computed according to the measure $\mathrm{d}\rho_\epsilon/\mathrm{d}\epsilon|_{\epsilon=0}$ (instead of the original invariant measure ρ_0). A simple dynamical-statistical interpretation for the second term is harder to provide, as the time-dependent operator appearing between the two observables leads to computing correlations (with respect to the unperturbed invariant measure ρ_0) between points in the phase space having no obvious dynamical link. See also the previous discussion around equations (12) and (13).

Note also that the linear response of higher order spectral properties of the system to the $\epsilon-$ perturbation can be derived by applying the $n-1$ dimensional Fourier transform in equation (18). This shows that our results allow for a more comprehensive understanding of the response of the system to perturbations than the usual response theory.

We note that in Lucarini (2012) the problem of looking at the change of the spectral properties of a system had been approached from a different angle, studying the effect of stochastic perturbations applied on top of deterministic chaotic dynamics. The main result obtained there is that one can establish a simple algebraic link between the change of the power spectrum of an observable (corresponding to the specific choice $\Psi_a = \Psi_b$ in terms of what presented here) and the squared modulus of the susceptibility referred to the same observable.

3. Response formulae for reduced order models

We find a useful application of the results detailed above in the special case of constructing parameterisations for reduced order models, along the lines of Wouters and Lucarini (2012, 2013) and (2016). Let's first recapitulate the main results obtained there and we shall then see how to apply the extended response theory described above to obtain some new results. The idea is to derive formulae able to describe how the parameterisation changes as a result of perturbations applied to the full system, or, in other terms, how applying a perturbation changes the properties of the Mori–Zwanzig projection operator.

3.1. Constructing the projected evolution equations for coarse grained variables

We consider a high-dimensional chaotic dynamical system $\dot{\vec{z}} = \vec{F}_z(\vec{z})$ where $\vec{z}$ belongs to a compact manifold $\mathcal{Z}$, and then rewrite the dynamics by separating $\vec{z}$ into two subsets of variables, with $\vec{z} = [\vec{x}; \vec{y}]$. Such a separation typically comes from the fact that we are interested in studying the properties of the $\vec{x}$-variables only, corresponding to the coarse grained quantities of interest. Usually, the number of $\vec{y}$-variables is much larger than the number of $\vec{x}$-variables, and one would like to have a time-scale separation (or spectral gap) between the two sets of variables. Without loss of generality one can write:

$$\dot{\vec{x}} = \vec{F}_x(\vec{x}) + \delta\vec{\Psi}_x(\vec{x}, \vec{y}) \tag{20}$$

$$\dot{\vec{y}} = \vec{F}_y(\vec{y}) + \delta\vec{\Psi}_y(\vec{x}, \vec{y}) \tag{21}$$

where we have separated the part of the vector field ($\vec{\Psi}$) coupling the $\vec{x}$- and the $\vec{y}$-variables from the part of the vector field ($\vec{F}$) that drives independently the two groups of variables. We have also introduced the bookkeeping parameter δ, which measures the strength of the coupling between the $\vec{x}$- and $\vec{y}$-variables. We wish to derive a reduced model for the $\vec{x}$-variables able to reproduce accurately (in some sense to be defined later) its statistical properties resulting from the full dynamics given in equations (20) and (21). The Mori–Zwanzig theory allows for an exact and powerful yet implicit solution to this problem, obtained by formally removing the evolution of the $\vec{y}$-variables. As a result, one obtains that it is possible to write the projected dynamics of the $\vec{x}$-variables as follows:

$$\dot{\vec{x}} = \vec{F}_x(\vec{x}) + \vec{M}_\delta(\{\vec{x}\}) \tag{22}$$

where $\vec{M}$ contains both Markovian and non-Markovian components and provides the so-called parameterisation of the effect of the $\vec{y}$-variables on the $\vec{x}$-variables. The vector field $\vec{M}$ contains information on the average effect of the coupling between the $\vec{x}$- and $\vec{y}$-variables, on the impact of the fluctuations of the $\vec{y}$-variables, and on the memory effects due to nonlinear cross-correlations between the two groups of variables.

Unfortunately, the explicit form of $\vec{M}$ is not in general available. In the limit of infinite time scale separation between the $\vec{x}$- and $\vec{y}$-variables, such that the $\vec{y}$-variables fluctuate infinitely faster than the $\vec{x}$-variables, it is instead possible to derive explicit results using the homogenisation technique (Pavliotis and Stuart 2008). One obtains that the $\vec{M}_\delta(\{\vec{x}\})$ term is given by the sum of a deterministic term, corresponding to the intuitive mean field effect, plus a white noise stochastic term, which describes the effect of the fluctuations, while the memory term disappears. Following Pavliotis and Stuart (2008), one has that in physical systems the white noise should be interpreted in the sense of Stratonovich, as it should be considered as limiting case of a red noise having vanishing decorrelation time.

This approach is extremely powerful and physically appropriate in all the situations where a substantial time-scale separation can be found between the two sets of variables. In situations, like in the case of climate dynamics, where there is no real spectral gap, the assumption of infinite time scale separation is risky.

In Wouters and Lucarini (2012, 2013) and (2016) we have shown that, assuming that δ is small (weak coupling hypothesis), it is possible to find an explicit expression of the Mori–Zwanzig corrections to the dynamics by performing a formal expansion of the Koopman operator in powers of δ and retaining the first two orders. The idea is to treat the coupling as a perturbation to the otherwise uncoupled dynamics of the $\vec{x}$- and $\vec{y}$-variables. One obtains that the surrogate dynamics of the $\vec{x}$-variables can be written as follows:

$$\dot{\vec{x}} = \vec{F}_x(\vec{x}) + \delta\vec{M}_1(\vec{x}) + \delta\vec{M}_2(\{\vec{x}\}) + \delta^2\vec{M}_3(\{\vec{x}\}) \tag{23}$$

where $\vec{M}_1(\vec{x})$ is a determistic vector field, $\vec{M}_2(\{\vec{x}\})$ is a stochastic term constructed from the statistics of the fluctuations of the $\vec{y}$ variables, and $\vec{M}_3(\{\vec{x}\})$ is a non-Markovian term describing the fact that in the fully coupled dynamics the current state of the $\vec{y}$-variables contains information on the state of the $\vec{x}$-variables at previous times. This result is in agreement with the general theory on model reduction proposed by Chekroun *et al* (2015a, 2015b).

The explicit expressions for the terms on the right hand side of equation (23) are obtained as follows. We start by defining $\rho_{u,y}$ as the invariant measure of the dynamical system $\dot{\vec{y}} = \vec{F}_y(\vec{y})$, where u in the lower index refers to the fact the dynamics of $\vec{y}$ is uncoupled from the dynamics of $\vec{x}$, so that $\langle\rho_{u,y}, \xi(\vec{y})\rangle$ the expectation value of a $\rho_{u,y}$−measurable observable $\xi(\vec{y})$.

We then take the simplifying assumption that $\vec{\Psi}_x(\vec{x},\vec{y}) = \vec{\Psi}^1_x(\vec{x})\vec{\Psi}^2_x(\vec{y})$ and $\vec{\Psi}_y(\vec{x},\vec{y}) = \vec{\Psi}^1_y(\vec{x})\vec{\Psi}^2_y(\vec{y})$. As discussed in Wouters and Lucarini (2013) and (2016), such an assumption leads to simpler and easier to interpret formulae; yet, it does not really lead to a loss of generality, if one takes into account the possibility of expanding a function of both $\vec{x}$- and $\vec{y}$-variables as a sum of products of functions of separately $\vec{x}$- and $\vec{y}$-variables only, using a Schauder decomposition (Lindenstrauss and Tzafriri 1996).

The deterministic mean field term is given by:

$$\vec{M}_1(\vec{x}) = \vec{\Psi}^1_x(\vec{x})\langle\rho_{u,y}, \vec{\Psi}^2_x(\vec{y})\rangle. \tag{24}$$

We introduce now the anomalies $\vec{\Psi'}^j_q(\vec{q}) = \vec{\Psi}^j_q(\vec{q}) - \langle\rho_{u,q}, \vec{\Psi}^j_q(\vec{q})\rangle$ for $j = 1,2$ and $q = x,y$. We have that the second term of the parameterisation can be written as:

$$\vec{M}_2(\{\vec{x}\}) = \vec{\Psi}^1_x(\vec{x})\vec{\eta} \tag{25}$$

where $\vec{\eta}$ is a centered random process with time correlation given by

$$C(\vec{\eta}(0), \vec{\eta}(t)) = \langle\rho_{u,y}, \vec{\Psi'}^2_x(\vec{y})\Pi_{0,x}(t)\vec{\Psi'}^2_x(\vec{y})\rangle, \tag{26}$$

where $\Pi_{0,q}(t)$ indicates the Koopman operator of the $\vec{q} = \vec{x},\vec{y}$-variables in the uncoupled case with $\delta = 0$, such $\Pi_{0,q}(t)A(\vec{q}) = A(\vec{q}(t))$ for any function of the phase space A. Note that the random process $\vec{\eta}$ is not unique, as, at the desired level of precision in terms of δ, we only require that the noise is centered and with the above mentioned correlation properties. Finally, the third term in the parameterisation provides the non-Markovian contribution to the reduced model and is given by

$$\vec{M}_3(\{\vec{x}\}) = \int_0^\infty \mathrm{d}\tau\vec{h}(t-\tau, \Pi_0(t-\tau)\vec{x}) \tag{27}$$

where the integration kernel $\vec{h}$ is written as

$$\vec{h}(\sigma, \Pi_0(\sigma)\vec{x}) = \vec{\Psi}^1_y(\vec{x})\Pi_{0,x}(\sigma)\vec{\Psi}^1_x(\vec{x})\langle\rho_{u,y}, \vec{\Psi}^1_y(\vec{y})\vec{\nabla}_{\vec{y}}\Pi_{0,y}(\sigma)\vec{\Psi}^2_x(\vec{y})\rangle. \tag{28}$$

A thorough interpretation of the three terms is reported in Wouters and Lucarini (2012, 2013, 2016).

We note that, using the Ruelle response theory, one also proves that up to second order in δ the invariant measure of the dynamical system given in equation (23) is the same as the $\vec{x}$−projection of the measure of the full dynamics given in equations (20) and (21). Therefore, the parameterisation given in equation (23) is effective in reproducing both the dynamical and the statistical properties of the full system.

Furthermore, as opposed to more common heuristic approaches, it performs—in the limit of small δ—consistently well no matter which observable Φ we are considering; it is, in this sense, universal and not targetted to a specific measure of skill. In Wouters *et al* (2016), Vissio and Lucarini (2016) and Demaeyer and Vannitsem (2017) the properties of parameterisations of models of different level of complexity obtained following this strategy are studied in detail. Note that in the limit of infinite time-scale separation between the $\vec{x}$- and $\vec{y}$-variables, the homogenisation theory results are recovered and the non-Markovian term drops out.

3.2. Impact of the perturbations on the parameterisation

A basic problem often encountered when constructing parameterisations for unresolved processes is assessing the robustness of the reduced model with respect to small changes of the dynamics of the full system. When the dynamics of the full system is weakly perturbed with respect to reference conditions, one expects that also the reduced model undergoes small changes. In what follows, we define a set of response formulae able to predict how the various terms in equations (24)–(27) defining the parameterisation change as a result of such a perturbation. One needs to note that the presence of a small perturbation to the dynamics is usually interpreted as resulting from changes in the applied forcing applied or from changes in the value of some internal parameters. Alternatively, the small perturbation can be interpreted as caused by model error due to our incomplete knowledge of the system. We then consider the following system:

$$\dot{\vec{x}} = \vec{F}_x(\vec{x}) + \delta\vec{\Psi}_x(\vec{x},\vec{y}) + \epsilon G_x(\vec{x}) \tag{29}$$

$$\dot{\vec{y}} = \vec{F}_y(\vec{y}) + \delta\vec{\Psi}_y(\vec{x},\vec{y}) + \epsilon G_y(\vec{y}) \tag{30}$$

where we have included on the right hand side of the evolution equations a (small) perturbation vector field, whose intensity is controlled by the bookkeeping parameter ϵ, while leaving the coupling unaltered with respect to the original system shown in equations (20) and (21). In this case, the uncoupled model reads as

$$\dot{\vec{x}} = \vec{F}_x(\vec{x}) + \epsilon G_x(\vec{x}) \tag{31}$$

$$\dot{\vec{y}} = \vec{F}_y(\vec{y}) + \epsilon G_y(\vec{y}). \tag{32}$$

The reduced model, following equation (23), can be written as:

$$\dot{\vec{x}} = \vec{F}_x(\vec{x}) + \epsilon G_x(\vec{x}) + \delta\vec{M}_{1,\epsilon}(\vec{x}) + \delta\vec{M}_{2,\epsilon}(\{\vec{x}\}) + \delta^2\vec{M}_{3,\epsilon}(\{\vec{x}\}) \tag{33}$$

where the dependence on ϵ is implicit for all terms except the trivial one. We now wish to expand the terms $\vec{M}_{1,\epsilon}(\vec{x})$, $\vec{M}_{2,\epsilon}\{\vec{x}\}$, and $\vec{M}_{3,\epsilon}\{\vec{x}\}$ in powers of ϵ and retain the 0^{th} and 1^{st} terms. This will lead us to the response formulae for the reduced order model. In order to do so, we define $\rho_{u,\epsilon,y}$ the invariant measure of the dynamical system in equation (32), so that clearly $\rho_{u,\epsilon=0,y} = \rho_{u,y}$, and take advantage of the results contained in section 2 in order to compute the linear response of expectation values of observables and correlations to the perturbation proportional to ϵ. Let's first look at the deterministic term introduced in equation (24). We use equations (4) and (5) to derive:

$$\begin{aligned}\delta\vec{M}_{1,\epsilon}(\vec{x}) &= \delta\vec{\Psi}^1_x(\vec{x})\langle\rho_{u,\epsilon,y}, \vec{\Psi}^2_x(\vec{y})\rangle \\ &= \delta\vec{M}_{1,\epsilon=0}(\vec{x}) \\ &\quad + \delta\epsilon\vec{\Psi}^1_x(\vec{x})\int_0^\infty \mathrm{d}\tau\langle\rho_{u,y}, \mathcal{L}_{(1),y}\exp\left(\mathcal{L}_{(0),y}\tau\right)\vec{\Psi}^2_x(\vec{y})\rangle + \text{h.o.t.}\end{aligned} \tag{34}$$

where $\vec{M}_{1,\epsilon=0}(\vec{x})$ is given in equation (24), $\mathcal{L}_{(0),y} = \vec{F}_y \cdot \vec{\nabla}$, and $\mathcal{L}_{(1),y} = \vec{G}_y \cdot \vec{\nabla}$. Note that the correction term is proportional to ϵ so that, when we insert it in equation (33), it brings a contribution proportional to the product of the two perturbation parameters δ and ϵ.

When looking at the modifications of the stochastic term given in equation (25), we have that the ϵ-correction to the dynamics of the $\vec{y}$-variables leads to a change in the correlation properties of the random process η_ϵ. We obtain that

$$\delta\vec{M}_{2,\epsilon}(\{\vec{x}\}) = \delta\vec{\Psi}^1_x(\vec{x})\vec{\eta_\epsilon} \tag{35}$$

where we have that:

$$C(\vec{\eta_\epsilon}(0),\vec{\eta_\epsilon}(t)) = C(\vec{\eta}_{\epsilon=0}(0),\vec{\eta}_{\epsilon=0}(t)) + \epsilon\frac{\mathrm{d}}{\mathrm{d}\epsilon}\langle\rho_{u,\epsilon,y},\vec{\Psi'}^2_x(\vec{y})\Pi_{\epsilon,x}(t)\vec{\Psi'}^2_x(\vec{y})\rangle|_{\epsilon=0} + \text{h.o.t.}, \tag{36}$$

with $C(\vec{\eta}_{\epsilon=0}(0),\vec{\eta}_{\epsilon=0}(t))$ given in equation (26). Using equations (12)–(14) we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\epsilon}\langle\rho_{u,\epsilon,y},\vec{\Psi'}^2_x(\vec{y})\Pi_{\epsilon,x}(t)\vec{\Psi'}^2_x(\vec{y})\rangle|_{\epsilon=0} &= \int_0^\infty \mathrm{d}\tau_1\langle\rho_{u,y},\mathcal{L}_{(1),y}\Pi_{0,y}(\tau_1)\vec{\Psi'}^2_x(\vec{y})\Pi_{0,y}(t)\vec{\Psi'}^2_x(\vec{y})\rangle \\ &+ \int_0^t \mathrm{d}\tau_2\langle\rho_{u,y},\vec{\Psi'}^2_x(\vec{y})\Pi_{0,y}(t-\tau_2)\mathcal{L}_{(1),y}\Pi_{0,y}(\tau_2)\vec{\Psi'}^2_x(\vec{y})\rangle.\end{aligned} \tag{37}$$

The previous formula shows that the changes in the correlation of the noise due to the ϵ-perturbation of the dynamics are non-trivial. In the limit of infinite time separation between the $\vec{x}$- and the $\vec{y}$-variables, such that the noise correlation is proportional to a Dirac's delta in both the unperturbed and perturbed system, the correction above results into a change of the constant in front of the Dirac's delta by a factor proportional to ϵ.

Finally, in order to construct the response formula for the term responsible for the non-Markovian part of the parameterisation, we need to evaluate the first order ϵ–correction to the memory kernel $\vec{h}_\epsilon$, where

$$\vec{M}_{3,\epsilon}\{\vec{x}\} = \int_0^\infty \mathrm{d}\tau\vec{h}_\epsilon(t-\tau,\Pi_\epsilon(t-\tau)\vec{x}). \tag{38}$$

By definition we have:

$$\vec{h}_\epsilon(\sigma,\Pi_\epsilon(\sigma)\vec{x}) = \vec{\Psi}^1_y(\vec{x})\Pi_{\epsilon,x}(\sigma)\vec{\Psi}^1_x(\vec{x})\langle\rho_{u,\epsilon,y},\vec{\Psi}^1_y(\vec{y})\vec{\nabla}_{\vec{y}}\Pi_{\epsilon,y}(\sigma)\vec{\Psi}^2_x(\vec{y})\rangle, \tag{39}$$

and we wish to construct the following expansion:

$$\vec{h}_\epsilon(\sigma,\Pi_\epsilon(\sigma)\vec{x}) = \vec{h}_{\epsilon=0}(\sigma,\Pi_0(\sigma)\vec{x}) + \epsilon\frac{\mathrm{d}}{\mathrm{d}\epsilon}\vec{h}_\epsilon(\sigma,\Pi_\epsilon(\sigma)\vec{x})|_{\epsilon=0} + \text{h.o.t.} \tag{40}$$

where $\vec{h}_{\epsilon=0}(\sigma,\Pi_0(\sigma)\vec{x})$ is given in equation (28). On the r.h.s. of equation (39) the parameter ϵ appears, reading from left to right, in the Koopman operator of the $\vec{x}$-variables, in the definition of the invariant measure for the unperturbed dynamics of the $\vec{y}$-variables, and in the Koopman operator of the $\vec{y}$-variables, thus implying that the term proportional ϵ in equation (40) includes the sum of three separate corresponding contributions. The three terms are reported below in equations (41)–(43), respectively:

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\vec{h}_\epsilon(\sigma,\Pi_\epsilon(\sigma)\vec{x})|_{\epsilon=0} = \vec{\Psi}^1_y(\vec{x})\int_0^\sigma \mathrm{d}\tau_0\Pi_{0,x}(\sigma-\tau_0)\mathcal{L}_{(1),x}\Pi_{0,x}(\tau_0)\vec{\Psi}^1_x(\vec{x})\langle\rho_{u,y},\vec{\Psi}^1_y(\vec{y})\vec{\nabla}_{\vec{y}}\Pi_{0,y}(\sigma)\vec{\Psi}^2_x(\vec{y})\rangle \tag{41}$$

$$+\vec{\Psi}^1_y(\vec{x})\Pi_{0,x}(\sigma)\vec{\Psi}^1_x(\vec{x})\int_0^\infty \mathrm{d}\tau_1\langle\rho_{u,y},\mathcal{L}_{(1),y}\Pi_{0,y}(\tau_1)\vec{\Psi}^1_y(\vec{y})\Pi_{0,y}(t)\vec{\Psi}^2_x(\vec{y})\rangle \tag{42}$$

$$+\vec{\Psi}^1_y(\vec{x})\Pi_{0,x}(\sigma)\vec{\Psi}^1_x(\vec{x})+\int_0^\sigma \mathrm{d}\tau_2\langle\rho_{u,y},\vec{\Psi}^1_y(\vec{y})\Pi_{0,y}(\sigma-\tau_2)\mathcal{L}_{(1),y}\Pi_{0,y}(\tau_2)\vec{\Psi}^2_x(\vec{y})\rangle. \tag{43}$$

It is interesting to note that the first contribution above in equation (41) is the only one involving the perturbation to the Liouville operator for the $\vec{x}$-variables $\mathcal{L}_{(1),x}$. Correspondingly, it leads to a memory term in the definition of the kernel, which makes the overall non-Markovian term of the parameterisation more cumbersome; compare with equation (38).

The results presented here, albeit admittedly convoluted, show how it is in principle possible to construct the response theory for a reduced order model resulting from the coarse graining of higher dimensional system. In other terms, we find how one can construct a flexible parameterisation that can be explicitly adapted when the background system is altered, as a result of perturbations to the dynamics or taking into account the model error.

4. Summary and conclusions

Response formulae are extremely useful tools for predicting how the properties of statistical mechanical systems change as a result of perturbations. In practice, such perturbation can result from changes in the forcing applied to the system or to the internal parameters. Mathematically solid response theories can be constructed both taking the point of view of chaotic deterministic dynamical systems—see e.g. Ruelle (2009) and Liverani and Gouëzel (2006)—and of stochastic dynamical systems—see e.g. Hairer and Majda (2010). The deterministic point of view faces the difficulty of requiring relatively stringent conditions of the nature of the flow, while the stochastic point of view permits deriving the desired results under more general conditions. The unavoidable price we pay in this latter case is that we should be able to justify the nature of the noise we use in our mathematical construction. For any practical use, the deterministic and the stochastic formulation of the problem are virtually equivalent.

In this paper we have extended the usual results of linear response theory by computing how the n-point correlations at different times of general smooth observables of the system under investigation change as a result of adding a weak perturbation to the vector flow. The obtained response formulae entail exactly n different terms. The first term results results from the change in the invariant measure of the system, and is what one would guess from a naive use of response theory. The additional $n-1$ terms result from the linear correction to the Koopman operator of the system evaluated at all the $n-1$ consecutive intervals defining the ordering of the time variables in the argument of the correlation function. Such terms cannot be framed in any form similar to the FTD, as opposed to the first term. By taking advantage of the linearity of the Fourier transform, we are able to derive expressions describing how the spectral properties of the system are altered as a result of the presence of the perturbation. Formulae for second or higher order response to perturbations can also be obtained but are not presented here as they are rather complicated and do not add much for the scopes of this paper.

We have then applied the general findings above to a problem of specific interest in the theory of coarse graining of multi-scale dynamical systems. From a truncation of the Mori–Zwanzig projection operator we can derive a parameterisation of the neglected degrees of freedom such that the resulting invariant measure of the surrogate system is identical to the projected measure of the full system up to second order in the parameter controlling the

intensity of the coupling between the degrees of freedom of interest and the ones we want to neglect (Wouters and Lucarini 2012, 2013, 2016). One obtains that the parameterisation contains a deterministic component, a stochastic component, and a non-Markovian component, in agreement with the general theory of Chekroun *et al* (2015a, 2015b), and derives explicit expressions for the three terms. In this paper we have derived explicit expressions describing how the parameterisation changes as a result of a perturbation applied to the full system, or, in other terms, we have computed how the additional forcing projects in the reduced order model. Alternatively, one can see our results as a way to predict how the model error in the full system is translated as error in the reduced order model.

One has to note that all the terms in (34)–(43) are expectation values w.r.t. $\rho_{u,y}$, the uncoupled y measure. Therefore, if we have access to such statistical properties, it is possible not only to construct a reduced model, but also to adapt it to account for small perturbations. Therefore, our results provide a basis for constructing general parameterisations for reduced order models that can be modified in order to account for changes in the dynamics of the full system. We suggest that this might be of relevance for fields such as condensed matter, molecular dynamics, and geophysical fluid dynamics, where the construction of accurate, flexible, and adaptive coarse graining procedures is of the uttermost relevance and urgency. In particular, in the case of geophysical fluid dynamics, our results might be useful for the construction of robust scale aware parameterisations, i.e. parameterisations that can be automatically or easily adapted to a changing grid resolution of the numerical model, which determines which physical processes can be explicitly resolved.

We will delve into the problem of implementing these results in specific numerical models and testing their accuracy in future investigations.

The formulae presented provide an overarching framework for understanding how higher order statistical moments of the systems are impacted by changes in the dynamics, and appear to be of general interest. In previous papers we showed that the Ruelle response theory is a tool of practical utility for approaching the problem of predicting climate change (Ragone *et al* 2016, Lucarini *et al* 2017). Among the many possible applications of the results presented in this paper, we would to emphasise that the generalised response formulae introduced here allow for framing the question of how the climate variability responds to anthropogenic and natural forcings. This is a major and indeed open problem in the climate literature (Ghil 2015) and we will try to approach it in future studies.

An application of possible interest in the area of statistical mechanics deals with the study of the equivalence of perturbed Hamiltonian systems that are allowed to reach a steady state thanks to the coupling with thermostats described by different microscopic dynamics. In appendix we briefly describe the motivations behind the introduction of thermostats in physical systems. The formulae presented here allow for computing explicitly the linear response of the correlations of the macroscopic physical variables of differently thermostatted perturbed Hamiltonian systems and then for checking whether an equivalence in the thermodynamic limit of such corrections exists, and if, so, how fast in terms of N, thus extending the results obtained by Evans and Morriss (2008) for the case of the linear response for physical observables.

Acknowledgments

The results contained in this paper have been drafted during the Workshop *Transport in Unsteady Flows: from Deterministic Structures to Stochastic Models and Back Again* held on January 16–20 2017 at the Banff International Research Station, Banff, Canada. We thank the

organizers and the institution for having made this possible. The authors wish to thank Judith Berner, Tamas Bodai, Mickael Chekroun, Matteo Colangeli, Giovanni Gallavotti, Michael Ghil, Georg Gottwald, Cecile Penland, David Ruelle, Stephane Vannitsem, and Gabriele Vissio for many stimulating conversations on the topics discussed in this paper. VL wishes to thank the DFG TRR181 *Energy Transfers in the Atmosphere and the Ocean* for partial financial support. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement No. PIOF-GA-2013-626210.

Appendix. Thermostatted systems

A short note should be added in the case we are studying the response to perturbations of an N-particle system described by a Hamiltonian $H_0 = K_0 + V_0$ where $K_0 = \sum_{j=1}^{N} \vec{p}_i^2/2m$ and $V_0 = \sum_{j\neq i=1}^{N} V(|\vec{q}_i - \vec{q}_j|)$, where $\{\vec{q}_1, \ldots, \vec{q}_n, \vec{p}_1, \ldots, \vec{p}_n\}$ are the canonical variables, m is the mass of the particles, and V is the internal potential describing the interaction between the particles. The unperturbed system obeys the following equation of motions, for $i = 1, \ldots, N$:

$$\begin{aligned} \dot{\vec{q}}_i &= \vec{p}_i/m \\ \dot{\vec{p}}_i &= -\vec{\nabla}_{\vec{q}_i} V_0. \end{aligned} \tag{A.1}$$

If we want to study the problem of deviations from equilibrium due to the application of an external (in general, non conservative) force $\epsilon\vec{B}(\vec{q}_i)$ acting on each particle, in order to keep physical well-posedness, we need to alter the vector flow as follows:

$$\begin{aligned} \dot{\vec{q}}_i &= \vec{p}_i/m \\ \dot{\vec{p}}_i &= -\vec{\nabla}_{\vec{q}_i} V_0 + \epsilon\vec{B}(\vec{q}_i) - \alpha\vec{p}_i, \end{aligned} \tag{A.2}$$

where α is a nontrivial friction coefficient describing the action of a thermostat (Gallavotti 1997, Cohen and Rondoni 1998, Ruelle 2000) that avoids the long-term accumulation or depletion of energy in the system and allows for the set up of a well-defined steady state. We consider here the case of deterministic thermostats.

As as example, choosing $\alpha = \epsilon \sum_{i=1}^{N} \vec{B}(\vec{q}_i) \cdot \vec{p}_i / \sum_{i=1}^{N} /(\, p_i^2/m)$, one obtains that the function H_0 is an invariant of the system given in equation (A.2). Using such thermostatted equations of motions and considering as perturbation flow in equation (3) $\epsilon\vec{G}(\vec{x}) = (0, \ldots, 0, \epsilon\vec{B}(\vec{q}_1) - \alpha\vec{p}_1, \ldots, \epsilon\vec{B}(\vec{q}_N) - \alpha\vec{p}_N)$ where the perturbation affects only the evolution equations for the momentum variables, one recovers in equation (9) the correspondence between change in the phase space contraction rate and entropy production of the system mentioned above (Cohen and Rondoni 1998). Instead, neglecting the term responsible for the thermostatting, one instead derives from equation (9) the physically wrong result that the entropy production of an equilibrium system driven out of equilibrium by an external field vanishes.

Many functional forms can be given for α, describing different ways of realising microscopically such long term balance. The equivalence of the thermostats means that in the thermodynamic limit $N \to \infty$ the expectation values of macroscopic physical observables does not depend on the choice of α, with differences between the results obtained using different thermostats typically going to zero typically as $1/N$ (Gallavotti 1997, Cohen and Rondoni 1998, Ruelle 2000, Evans and Morriss 2008, Gallavotti 2014, Gallavotti and Lucarini 2014). This property persists also when the sensitivity of the system is considered: in the thermodynamic limit the linear response of observables to perturbations is also independent of the choice of α (Evans and Morriss 2008).

References

Abramov R V and Majda A J 2007 Blended response algorithms for linear fluctuation-dissipation for complex nonlinear dynamical systems *Nonlinearity* **20** 2793–821

Arakawa A, Jung J-H and Wu C-M 2011 Toward unification of the multiscale modeling of the atmosphere *Atmos. Chem. Phys.* **11** 3731–42

Baiesi M and Maes C 2013 An update on the nonequilibrium linear response *New J. Phys.* **15** 013004

Baladi V 2000 *Positive Transfer Operators and Decay of Correlations* (Singapore: World Scientific)

Baladi V, Benedicks M and Schnellmann D 2014 Linear response, or else ICM Seoul 2014 talk

Baladi V and Smania D 2008 Linear response formula for piecewise expanding unimodal maps *Nonlinearity* **21** 677

Baron R, Trzesniak D, de Vries? A H, Elsener A, Marrink S J and van Gunsteren W F 2007 Comparison of thermodynamic properties of coarse-grained and atomic-level simulation models *Chem. Phys. Chem.* **8** 452–61

Berner J *et al* 2016 Stochastic parameterization: towards a new view of weather and climate models *Bull. Am. Meteorol. Soc.*

Bhalla P, Das N and Singh N 2016 Moment expansion to the memory function for generalized drude scattering rate *Phys. Lett.* A **380** 2000–7

Chekroun M D, Neelin D J, Kondrashov D, McWilliams J C and Ghil M 2014 Rough parameter dependence in climate models and the role of Ruelle–Pollicott resonances *Proc. Natl Acad. Sci.* **111** 1684–90

Chekroun M D, Liu H and Wang S 2015a *Approximation of Stochastic Invariant Manifolds: Stochastic Manifolds for Nonlinear SPDEs* I (*Springer Briefs in Mathematics*) (Berlin: Springer)

Chekroun M D, Liu H and Wang S 2015b *Approximation of Stochastic Invariant Manifolds: Stochastic Manifolds for Nonlinear SPDEs* II (*Springer Briefs in Mathematics*) (New York: Springer)

Cohen E G D and Rondoni L 1998 Note on phase space contraction and entropy production in thermostatted hamiltonian systems *Chaos* **8** 357–65

Colangeli M and Lucarini V 2014 Elements of a unified framework for response formulae *J. Stat. Mech.* P01002

Demaeyer J and Vannitsem S 2017 Stochastic parametrization of subgrid-scale processes in coupled oceanatmosphere systems: benefits and limitations of response theory *Q. J. R. Meteorol. Soc.* **143** 881–96

Eckmann J P and Ruelle D 1985 Ergodic theory of chaos and strange attractors *Rev. Mod. Phys.* **57** 617–56

Evans D J and Morriss G P 2008 *Statistical Mechanics of Nonequilibrium Liquids* (Cambridge: Cambridge University Press)

Eyink G L, Haine T W N and Lea D J 2004 Ruelle's linear response formula, ensemble adjoint schemes and Lévy flights *Nonlinearity* **17** 1867

Franzke C L E, O'Kane T J, Berner J, Williams P D and Lucarini V 2015 Stochastic climate theory and modeling *Wiley Interdiscip. Rev.* **6** 63–78

Gallavotti G 1996 Chaotic hypothesis: Onsager reciprocity and fluctuation-dissipation theorem *J. Stat. Phys.* **84** 899–925

Gallavotti G 1997 Dynamical ensembles equivalence in fluid mechanics *Phys. D: Nonlinear Phenom.* **105** 163–84

Gallavotti G 2006 Stationary nonequilibrium statistical mechanics *Encyclopedia of Mathematical Physics* vol 3, ed J P Francoise *et al* (Amsterdam: Elsevier) pp 530–9

Gallavotti G 2014 *Nonequilibrium and Irreversibility* (New York: Springer)

Gallavotti G and Cohen E G D 1995 Dynamical ensembles in stationary states *J. Stat. Phys.* **80** 931–70

Gallavotti G and Lucarini V 2014 Equivalence of non-equilibrium ensembles and representation of friction in turbulent flows: the lorenz 96 model *J. Stat. Phys.* **156** 1027–65

Gerard L 2007 An integrated package for subgrid convection, clouds and precipitation compatible with meso-gamma scales *Q. J. R. Meteorol. Soc.* **133** 711–30

Ghil M 2015 A mathematical theory of climate sensitivity or, How to deal with both anthropogenic forcing and natural variability? *Climate Change: Multidecadal and Beyond* ed C P Chang (Dordrecht: Kluwer) pp 31–51

Ghil M and Childress S 1987 *Topics in Geophysical Fluid Dynamics: Atmospheric Dynamics, Dynamo Theory, and Climate Dynamics* (Berlin: Springer)
Gottwald G A, Wormell J P and Wouters J 2016 On spurious detection of linear response and misuse of the fluctuation-dissipation theorem in finite time series *Phys. D: Nonlinear Phenom.* **331** 89–101
Gritsun A and Branstator G 2007 *J. Atmos. Sci.* 64 2558–75
Gritsun A, Branstator G and Majda A J 2008 Climate response of linear and quadratic functionals using the fluctuation-dissipation theorem *J. Atmos. Sci.* **65** 2824–41
Gritsun A and Lucarini V 2017 Fluctuations, response, and resonances in a simple atmospheric model *Phys. D: Nonlinear Phenom.* **349** 62–76
Hairer M and Majda A J 2010 A simple framework to justify linear response theory *Nonlinearity* **23** 909
Hänggi P and Thomas H 1975 Linear response and fluctuation theorems for nonstationary stochastic processes *Z. Phys.* B **22** 295–300
Hänggi P and Thomas H 1977 Time evolution, correlations, and linear response of non-markov processes *Z. Phys.* B **26** 85–92
van Kampen N 1971 The case against linear response theory *Phys. Nor.* **5** 279
Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid A E and Kolinski A 2016 Coarse-grained protein models and their applications *Chem. Rev.* **116** 7898–936
Kubo R, Toda M and Hashitsume N 1988 *Statistical Physics II: Nonequilibrium Statistical Mechanics* (Berlin: Springer)
Kubo R 1957 Statistical-mechanical theory of irreversible processes. i. General theory and simple applications to magnetic and conduction problems *J. Phys. Soc. Japan* **12** 570–86
Lindenstrauss J and Tzafriri L 1996 *Classical Banach Spaces* (New York: Springer)
Liverani C and Gouëzel S 2006 Banach spaces adapted to Anosov systems *Ergod. Theor. Dynam. Syst.* **26** 189–217
Lorenz E N 1979 Forced and free variations of weather and climate *J. Atmos. Sci.* **36** 1367–76
Lucarini V 2008 Response theory for equilibrium and non-equilibrium statistical mechanics: causality and generalized Kramers–Kronig relations *J. Stat. Phys.* **131** 543–58
Lucarini V 2009 Evidence of dispersion relations for the nonlinear response of the Lorenz 63 system *J. Stat. Phys.* **134** 381–400
Lucarini V, Blender R, Herbert C, Ragone F, Pascale S and Wouters J 2014 Mathematical and physical ideas for climate science *Rev. Geophys.* **52** 809–59
Lucarini V, Saarinen J J, Peiponen K-E and Vartiainen E M 2005 *Kramers–Kronig Relations in Optical Materials Research* (New York: Springer)
Lucarini V and Sarno S 2011 A statistical mechanical approach for the computation of the climatic response to general forcings *Nonlinear Proc. Geophys.* **18** 7–28
Lucarini V 2012 Stochastic perturbations to dynamical systems: a response theory approach *J. Stat. Phys.* **146** 774–86
Lucarini V 2016 Response operators for markov processes in a finite state space: radius of convergence and link to the response theory for axiom a systems *J. Stat. Phys.* **162** 312–33
Lucarini V and Colangeli M 2012 Beyond the linear fluctuation-dissipation theorem: the role of causality *J. Stat. Mech.* **2012** P05013
Lucarini V, Ragone F and Lunkeit F 2017 Predicting climate change using response theory: global averages and spatial patterns *J. Stat. Phys.* **166** 1036–64
Marconi U M B, Puglisi A, Rondoni L and Vulpiani A 2008 Fluctuation-dissipation: response theory in statistical physics *Phys. Rep.* **461** 111
Mori H 1965 Transport, collective motion, and Brownian motion *Prog. Theor. Phys.* **33** 423–55
Palmer T N, Doblas-Reyes F J, Weisheimer A and Rodwell M J 2008 Toward seamless prediction: calibration of climate change projections using seasonal forecasts *Bull. Am. Meteorol. Soc.* **89** 459–70
Park S 2014 A unified convection scheme (unicon). Part I: formulation *J. Atmos. Sci.* **71** 3902–30
Pavliotis G A and Stuart A M 2008 *Multiscale Methods: Averaging and Homogenization* (New York: Springer)
Peixoto J P and Oort A H 1992 *Physics of Climate* (New York: AIP Press)
Ragone F, Lucarini V and Lunkeit F 2016 A new framework for climate sensitivity and prediction: a modelling perspective *Clim. Dyn.* **46** 1459–71
Ruelle D 1989 *Chaotic Evolution and Strange Attractors* (Cambridge: Cambridge University Press)
Ruelle D 1997 Differentiation of SRB states *Commun. Math. Phys.* **187** 227–41

Ruelle D 1998 Nonequilibrium statistical mechanics near equilibrium: computing higher-order terms *Nonlinearity* **11** 5–18
Ruelle D 2009 A review of linear response theory for general differentiable dynamical systems *Nonlinearity* **22** 855–70
Ruelle D 2000 A remark on the equivalence of isokinetic and isoenergetic thermostats in the thermodynamic limit *J. Stat. Phys.* **100** 757–63
Sakradzija M, Seifert A and Dipankar A 2016 A stochastic scale-aware parameterization of shallow cumulus convection across the convective gray zone *J. Adv. Model. Earth Syst.* **8** 786–812
Shinoda W, DeVane R and Klein M L 2007 Multi-property fitting and parameterization of a coarse grained model for aqueous surfactants *Mol. Simul.* **33** 27–36
Tantet A, Lucarini V, Lunkeit F and Dijkstra H A 2015a Crisis of the chaotic attractor of a climate model: a transfer operator approach (arXiv:1507.02228 [nlin.CD])
Tantet A, van der Burgt F R and Dijkstra H A 2015b An early warning indicator for atmospheric blocking events using transfer operators *Chaos* **25** 036406
Vissio G and Lucarini V 2016 A proof of concept for scale-adaptive parameterizations: the case of the lorenz '96 model (arXiv:1612.07223 [cond-mat.stat-mech])
Wang Q 2013 Forward and adjoint sensitivity computation of chaotic dynamical systems *J. Comput. Phys.* **235** 1–13
Wouters J, Dolaptchiev S I, Lucarini V and Achatz U 2016 Parameterization of stochastic multiscale triads *Nonlinear Process. Geophys.* **23** 435–45
Wouters J and Lucarini V 2012 Disentangling multi-level systems: averaging, correlations and memory *J. Stat. Mech.* P03003
Wouters J and Lucarini V 2013 Multi-level dynamical systems: connecting the ruelle response theory and the Mori–Zwanzig approach *J. Stat. Phys.* **151** 850–60
Wouters J and Lucarini V 2016 Parametrization of cross-scale interaction in multiscale systems *Climate Change: Multidecadal and Beyond* ed C-P Chang *et al* (Singapore: World Scientific) pp 67–80
Zwanzig R 1961 Memory effects in irreversible thermodynamics *Phys. Rev.* **124** 983–92