nature reviews physics

https://doi.org/10.1038/s42254-023-00650-8

Perspective

Check for updates

Theoretical tools for understanding the climate crisis from Hasselmann's programme and beyond

Valerio Lucarini [1,2] & Mickaël D. Chekroun [3,4]

Abstract

Klaus Hasselmann's revolutionary intuition in climate science was to use the stochasticity associated with fast weather processes to probe the slow dynamics of the climate system. Doing so led to fundamentally new ways to study the response of climate models to perturbations, and to perform detection and attribution for climate change signals. Hasselmann's programme has been extremely influential in climate science and beyond. In this Perspective, we first summarize the main aspects of such a programme using modern concepts and tools of statistical physics and applied mathematics. We then provide an overview of some promising scientific perspectives that might clarify the science behind the climate crisis and that stem from Hasselmann's ideas. We show how to perform rigorous and data-driven model reduction by constructing parameterizations in systems that do not necessarily feature a timescale separation between unresolved and resolved processes. We outline a general theoretical framework for explaining the relationship between climate variability and climate change, and for performing climate change projections. This framework enables us seamlessly to explain some key general aspects of climatic tipping points. Finally, we show that response theory provides a solid framework supporting optimal fingerprinting methods for detection and attribution.

Sections

[1]Department of Mathematics and Statistics, University of Reading, Reading, UK. [2]Centre for the Mathematics of Planet Earth, University of Reading, Reading, UK. [3]Department of Earth and Planetary Sciences, Weizmann Institute of Science, Rehovot, Israel. [4]Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, CA, USA. e-mail: v.lucarini@reading.ac.uk; mchekroun@atmos.ucla.edu

Introduction

The climate system is multiscale because it features variability over a wide range of scales, as a result of the interplay of a diverse array of forcings, instabilities and feedback mechanisms[1,2]. At different temporal and spatial ranges, the dominant role is played by different subsystems[3] (Fig. 1), of which there are five: the atmosphere, the hydrosphere, the cryosphere, the biosphere and the land surface. These subsystems differ in physical-chemical features, dominant dynamical processes and characteristic timescales, and they are coupled through a complex array of processes that exchange mass, momentum and energy[4,5].

Although the climate system is multiscale, models and phenomenological theories are typically developed by focusing on specific scales of motion and specific processes[6]. This is because it is challenging to construct satisfactory theories of climate dynamics and is virtually impossible to develop numerical models that accurately describe climate processes over all scales. These challenges arise because there are major knowledge gaps on the climate system that come from the lack of homogeneous, high-resolution and coherent observations, owing to several factors: the sheer size and practical inaccessibility of the climate subdomains; changes in the technology of data collection in the industrial era; and the need to resort to proxy (hence, indirect) data for the preindustrial epoch.

Thus, there is a need for a comprehensive framework for understanding the multiscale nature of climate variability and climate response to forcing. In the latter part of the twentieth century, Klaus Hasselmann proposed a coherent scientific angle on the climate system, with the goal of understanding climate variability, of detecting and interpreting the climate change signal, and of characterizing the behaviour of climate models. Our goal in this Perspective is to give a critical appraisal of the Hasselmann programme based on ideas and concepts from statistical mechanics and functional analysis that have gained traction, for the most part, in the past two decades. We also propose a comprehensive framework for understanding the multiscale nature of climate variability and climate response to forcing, and for fundamentally advancing understanding of the ongoing climate crisis.

Separating climate variability from any climate change signal is hard[7], in part because of the presence of unsteady external forcings and non-trivial internal multiscale dynamics. The natural history of Earth is characterized by continuously changing conditions. There is an interplay of rapid, irreversible transitions and gradual variations[8], with alternating dominance of positive and negative feedback depending on the timescale of interest[9]. Even so, as documented in successive assessment reports of the Intergovernmental Panel for Climate Change, the scientific community has painstakingly come to an agreement regarding the presence of a statistically significant climate change signal with respect to the conditions prevailing in the nineteenth century (climate change is real); and the possibility of attributing such a signal to anthropogenic causes (humans are responsible for it).

Attribution of the climate change signal requires being able to make a stringent case for a causal link between forcings (such as land-use change, change in the atmospheric compositions due to human activities, effects of vulcanism and solar variability) and the observed climate response. According to Judea Pearl's causality framework[10], doing so requires, in turn, comparing the observed state of the climate system with the counterfactual realities in which the forcings are selectively switched off. Clearly, we do not have access to counterfactual realities. Instead, we can create reasonable approximations of them by performing numerical simulations with climate models that have suitably defined protocols for the forcings. Hence, making statements on climate change attribution takes into account the unavoidable uncertainty due to natural variability and model error[11–15].

Over the years, the research focus of climate change science has shifted from making statements on globally averaged climate quantities to assessing climate change at regional level, to studying the changes in the higher statistical moments and in extreme events, and to investigating how climate change can manifest itself in the form of critical transitions. The current focus on the study of extremes[16] and critical phenomena (often referred to as tipping points[17,18]) has led the scientific community to use expressions like climate crisis or climate emergency instead of climate change[19].

This Perspective is structured as follows. We first give an overview of Hasselmann's programme. We then discuss rigorous theory-informed and effective data-driven model reduction strategies to highlight connections and integration between the top-down and bottom-up approaches. We then show how response theory for non-equilibrium systems[20–26] allows one to find explicit formulas and to devise experimental protocols aimed at performing climate change projections using climate models of different levels of complexity[27–30]. Finally, we show how that linear response formalism provides the mathematical and physical backbone behind the optimal fingerprinting

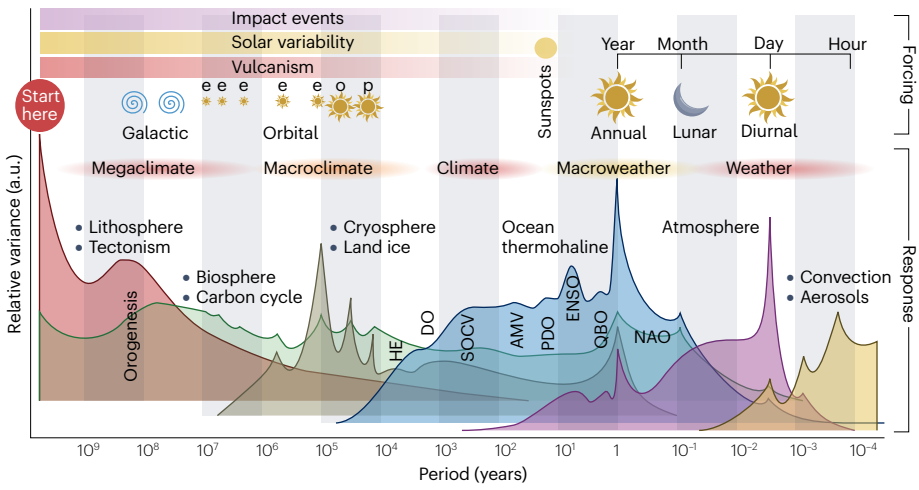


Fig. 1 | A Mitchell's diagram depicting a qualitative representation of the climate variability across a range of scales. The relative role of each climatic subcomponent of the acting forcings (on top) is indicated. The lower axis shows timescale in years. AMV, Atlantic Multidecadal Variability; DO, Dansgaard–Oeschger event; e, eccentricity; ENSO, El Niño–Southern Oscillation; HE, Heinrich event; NAO, North Atlantic Oscillation; o, obliquity; p, precession; PDO, Pacific Decadal Oscillation; QBO, Quasi-biennial Oscillation; SOCV, Southern Ocean centennial variability. Adapted with permission from ref. 3, Elsevier.

method for detection and attribution of the climate change signal. Ultimately, we demonstrate the strong link between the three main axes of Hasselmann's research programme.

A brief summary of Hasselmann's programme

The Hasselmann programme has had great influence on the modern development of climate science, both in its everyday practice and its more foundational aspects. The fundamental idea is to treat noise not like a nuisance one needs to filter out to gather useful information, but rather as a key aspect of the climate system one must fully explore and appreciate to further its understanding and to link model output and observational data. This is key for making progress in detecting and attributing the climate change signals at different spatial and temporal scales. We discuss three main axes of Hasselmann's programme. An informative summary of the programme, of some of its key developments in climate science, and of some of its implications for mathematics and physics at large is given elsewhere[31,32]. Useful accounts of Hasselmann's work on the occasion of his being awarded the 2021 Nobel prize in physics can be found in refs. 33,34.

Stochastic climate models

A key part of Hasselmann's programme is the separation of the multiscale system into slow dynamics (the climate) and fast dynamics (weather); the fast dynamics are incorporated into models for the slow dynamics as noise. The starting point of this journey is Hasselmann's proposal[35] to improve modelling of slow climate variables by parameterizing the influence of fast weather variables by means of an appropriate stochastic forcing. Thus at the core of Hasselmann's stochastic programme lies the derivation of an effective model with improved capability to capture the dynamics and statistical properties of the slow variables. Following previous work[36], we assume that our climate model can be written as the following large-dimensional system of ordinary differential equations

$$\begin{aligned} \dot{\mathbf{x}} &= f(\mathbf{x},\mathbf{y}), \quad \mathbf{x} \text{ slow climate variables}, \\ \dot{\mathbf{y}} &= \frac{1}{\delta} g(\mathbf{x},\mathbf{y}), \quad \mathbf{y} \text{ fast weather variables}, \end{aligned} \tag{1}$$

where $\mathbf{x}$ lies in $\mathbb{R}^d$, $\mathbf{y}$ lies in $\mathbb{R}^D$, f is a smooth mapping from $\mathbb{R}^{d+D}$ into $\mathbb{R}^d$ and g is a smooth mapping from $\mathbb{R}^{d+D}$ into $\mathbb{R}^D$.

The parameter δ is aimed at controlling the degree of timescale separation between the dynamics of the two sets of variables, and one typically assumes $0 < \delta \ll 1$. In mathematical language, Hasselmann's programme consists of deriving in the $\delta \to 0$ limit an effective reduced equation of equation (1) to approximate the statistical behaviour of the climate variables. When $\mathbf{y}$ follows some fast chaotic dynamics as commonly assumed[37,38], the reduced equation takes the form of the following system of Itô stochastic differential equations (SDEs),

$$\mathrm{d}\mathbf{x} = \mathbf{F}(\mathbf{x})\,\mathrm{d}t + \Sigma(\mathbf{x})\,\mathrm{d}\mathbf{W}_t. \tag{2}$$

The terms $\mathbf{F}$ and Σ in equation (2) have intuitive interpretations. Typically, the deterministic component $\mathbf{F}$ (also called the drift term) includes the average contribution of $f(\mathbf{x},\mathbf{y})$ to the dynamics of the slow variables, after averaging out the fast variables[39], but potentially may also include other correction terms[37]. The second term is aimed at parameterizing the effects of the fluctuations that remain after averaging and takes the form of a state-dependent noise, in which the $(d \times p)$-matrix Σ with (possible) nonlinear entries in $\mathbf{x}$ is driven by a vector of increments $\mathrm{d}\mathbf{W}_t$ (white noise) of a p-dimensional Wiener process $\mathbf{W}_t$.

In the case of climate dynamics, the chaoticity of the dynamics of the $\mathbf{y}$ variables is usually assumed to derive from the fast fluid dynamical instabilities occurring in the atmosphere and in the ocean[2,6,40]. Such instabilities are usually associated with conversion of energy between different forms (such as potential to kinetic) and between spatially symmetric and eddy components[4,41]. The derivation of such a limiting SDE in the presence of infinite timescale separation has a long history[39,42–45] and may be obtained through diverse routes pertaining to homogenization[44], averaging[46] or singular perturbation techniques[47,48]. The Majda, Timofeyev and Vanden-Eijnden (MTV) approach[43,49] builds on such techniques and provides a modern treatment of the topic in the context of climate dynamics, including rigorous results when nonlinear self-interactions between the fast waves, for example, can be modelled by means of Ornstein–Uhlenbeck processes.

A key message from equation (2) is that the impact of unresolved scales of motion on the scales of interest cannot be reproduced by using bulk formulas (contributing to the drift term) only. This has practical relevance: it is the fundamental motivation behind the development of stochastic parameterizations for weather and climate models, some of which are used in operational situations[50,51].

Physically, getting to the limit (2) allows for interpreting the fast weather dynamics as inducing a diffusion process. As such, the understanding of the complex nonlinear interactions of, for example, unstable modes with (possibly very) stable ones[52] — at the core of the chaotic nature of many geophysical flows[40,53,54] — is replaced by the understanding of the interactions between noise and nonlinear effects[55–57].

As appealing as such attributes are, the infinite timescale-separation assumption underlying equation (2) is often challenged in climate applications. An extension of the MTV approach has been proposed for the stochastic model reduction valid for slow–fast systems with a moderate timescale separation[58,59]. In the section below on model reduction, we discuss how to amend Hasselmann's programme in such situations, either by seeking natural extensions to equation (2) or stochastic alternatives. In the climate system, there is actually a multiplicity of spatiotemporal scales interacting across a wealth of processes (Fig. 1), and the Hasselmann ansatz, although inspiring, calls for revision.

An important implication of Hasselmann's approach is the provision of a probabilistic interpretation of climate dynamics, going beyond the study of individual trajectories. Indeed, the SDE (2) can then be translated into a Fokker–Planck equation (FPE)[60] that describes the evolution of the probability distribution of the climate's states, $\rho(\mathbf{x},t)$, as

$$\partial_t \rho = \mathcal{L}_0 \rho \tag{3}$$

where the operator $\mathcal{L}_0$ is defined as

$$\mathcal{L}_0 \rho = -\nabla \cdot (\mathbf{F}(\mathbf{x})\rho) + \frac{1}{2}\nabla \cdot \nabla (\Sigma\Sigma^T(\mathbf{x})\rho). \tag{4}$$

In practice, $\rho(\mathbf{x},t)$ is constructed using an ensemble of trajectories; see ref. 61 for a recent summary of the use of ensembles in climate modelling. $\Sigma\Sigma^T$ is the non-negative definite noise covariance matrix. The unperturbed climate is then given by the stationary solution $\rho_0(\mathbf{x})$ to the FPE defined by $\mathcal{L}_0\rho_0 = 0$. Defining stationarity in a multiscale system is non-trivial, as the stationary state can critically depend on the timescale of interest (Box 1).

When the noise term in equation (2) is sufficiently non-degenerate, that is, when, roughly speaking, the noise propagates out in the whole phase space through interactions with the nonlinear terms, the probability density ρ_0 is smooth[62]. In contrast, when $\Sigma = 0$, equation (3)

Box 1

Stationarity and multiscale behaviour

The very notion of a stationary reference climate is intimately related to considering autonomous dynamics as a starting point of our analysis; see equation (1). But boundary conditions and external forcings relevant for the climate system do change on a variety of timescales[3,9]; see a careful discussion of the notion of statistical balance and stationary state for the climate system in a biogeochemical perspective in ref. 280. The conundrum of defining stationarity in a multiscale system can be pragmatically dealt with by taking advantage of the viewpoint of ref. 295. One can extend equation (1) by including, instead of only ($\mathbf{x}$,$\mathbf{y}$), a larger set of qualitatively different variables ($\mathbf{x}_1, ..., \mathbf{x}_p$) ordered by their characteristic timescale, ranging from very slow to very fast behaviour. Depending on the timescale of interest (decadal, multicentennial or multimillennial climate variability, for example), one ends up choosing the relevant set of active climate variables of interest, $\mathbf{x}_k$ (for example, atmosphere, ocean and ice caps), with $\mathbf{x}_\ell$, $1 \le \ell < k$ determining the boundary conditions, and $\mathbf{x}_m$, $m > k$ contributing to defining the modified drift and noise law in equation (2). Time-dependent forcing terms associated with geological or astrophysical processes can be considered as constant (or corresponding to stochastic forcing, in cases such as fast solar variability) when focusing on climate processes of multimillennial or shorter timescales. Indeed, the definition of a climate system and of its evolution law depends on the timescale of interest; see ref. 6 for a related discussion of the so-called hierarchy of climate models. Hence, when referring to a reference stationary ρ_0, we are implicitly imposing a cutoff on the slow variability of the system.

becomes the Liouville equation and, in the case of dissipative chaotic systems, the probability distribution ρ_0 is typically singular with respect to the Lebesgue volume in $\mathbb{R}^d$ (ref. 63).

Assuming that $\int d\mathbf{x}\rho_0(\mathbf{x}) = 1$, the reference climatological mean for the observable Ψ is $\langle\Psi\rangle_0 = \int d\mathbf{x}\rho_0(\mathbf{x})\Psi(\mathbf{x})$. The function Ψ could be in principle any quantity of climatic interest, corresponding to local properties or spatially averaged ones. It makes sense to associate possible Ψs with essential climate variables (ECVs) — key physical, chemical or biological variables that critically contribute to the characterization of Earth's climate and are targeted for observations[64] — or the quantities used for defining performance metrics for Earth system models (ESMs)[65].

Climate response to forcings

A second landmark of the Hasselmann programme deals with the study of the response of climate models to perturbations. Aiming at studying how a climate model relaxes to steady-state conditions or responds to perturbations, such as a sudden CO_2 increase, Hasselmann and collaborators heuristically proposed a methodology inspired by the dynamics of linear systems[66,67]. They showed that one can express the variation $\delta\Psi(t)$ describing the departure of the model from steady-state conditions (or convergence to it) by performing the convolution of a suitably defined causal Green's function $G_\psi(t)$ ($G_\psi(t) = 0$ if $t < 0$) with the time modulation of the acting perturbation $f(s)$:

$$\delta\Psi(t) = \int_{-\infty}^{\infty} ds G_\Psi(t-s) f(s), \tag{5}$$

where Ψ describes a climate variable of interest. This approach amounts to treating the problem of climate change using response theory. An early approach along these lines proposed using the fluctuation–dissipation theorem (FDT)[68] to express the Green's functions in terms of readily accessible correlations of climate observables in the unperturbed state[69]. Additionally, Hasselmann and collaborators expressed the Green's function G_ψ as a sum of exponential terms:

$$G_\Psi(t) = \sum_k \alpha_\Psi^k \exp(\lambda_k^\Psi t), \tag{6}$$

where each of the λ_k encodes a feedback, with α_Ψ^k being the corresponding weighting factor. The Green's-function-based method was shown to have good skill both in performing climate change projections for individual model runs after filtering the natural variability[67], and in studying the carbon cycle in a climate model[66]. In the section on the climate response, we critically revise this approach by discussing how a formalism based on use of Green functions allows one to cast the problem of climate change as the response of the system's probability distribution or of the statistics of the Ψs to (possibly time-dependent) perturbations applied to equation (2)[6,70].

Detection and attribution of climate change

Using stochastic climate models, one can make statements on climate variability and climate change in terms of probability distributions, by computing averages and higher statistical moments. Separating climate variability and climate change in the course of our single realization of the climate evolution relies on statistically testing departure from stationarity on observational time series. Performing causal attribution (in Pearl's sense) of the climate change signal to specific forcings is a much harder task because it requires, as mentioned above, comparing information from observations and from climate model runs. In the 1990s, Hasselmann and collaborators proposed the basic conceptual framework for performing attribution studies of climate change[11–13]. This framework played a major role in clarifying that we are currently experiencing a statistically relevant and physically attributable shift from previous climatic conditions[14,15]. Following ref. 71, the problem of attribution can be cast as

$$Y_k = \sum_{p=1}^{F} \bar{X}_k^p \beta_p + \mathcal{R}_k \quad \bar{X}_k^p = X_k^p + \mathcal{Q}_k^p, k = 1, ..., N. \tag{7}$$

Here Y_k is the kth component of an N-component vector $\mathbf{Y}$ describing the observed climate change. As an example, $\mathbf{Y}$ could represent the 2011–2020 average temperature measured at N different locations. The goal is to reconstruct it as a linear combination of F regressors or fingerprints, that is, externally forced signals $\bar{X}_k^p$, $p = 1, ..., F$ associated with F different forcings, plus a vector describing the natural variability of the system $\mathcal{R}_k$ (refs. 11–13,72,73). The pth fingerprint $\bar{X}_k^p$ is obtained as an ensemble average of the climate change signal over (possibly many) simulations that are realized by perturbing the reference climate by applying exclusively the pth considered forcing. Practically speaking, we can have access to the approximations X_k^p for such fingerprints, whereby the difference $\mathcal{Q}_k^p$ with respect to the true value is associated with our incomplete sampling of the model response and with model

error. In many applications, information coming from different climate models is bundled together[74].

The goal is to perform an optimal inference of the coefficients β_p given the uncertainties described by $\mathcal{R}_k$ and $\mathcal{Q}_k^p$, hence the expression optimal fingerprinting. Usually, the vector column errors $\mathcal{R}_k$ and $\mathcal{Q}_k^p$, are modelled as independent, normally distributed, stochastic vectors with zero mean and covariance matrices C and Ω_p, respectively. The matrix C is constructed taking into account the correlations of the climate variables in the unperturbed climate. A simpler version of the theory above assumes $\Omega_p = 0$ for all p (ref. 72). In this case, via linear algebra, one derives a relatively simple expression for the best estimates for the βs and their uncertainties. As a next step, if one assumes that the natural variability simulated by the climate models matches that of the observations, and one estimates $\overline{X}_k^p$ as the average of L runs from a single climate model under the pth forcing, then, under the approximation above, one gets $\Omega_p = C/L$ (ref. 73).

One speaks of attribution of climate change for the pth fingerprint if the confidence interval of β_p does not intersect zero and includes the value 1. The procedure is truly successful if the confidence intervals for the coefficients β_p are not too spread out. Optimal fingerprinting relies critically on the assumption of linear dependence of the climate response to the applied forcings. Additionally, the method, despite its great success, has recently been criticized, as it has been suggested that uncertainties in the inference are sometimes underestimated[75], or, more radically, on the basis that the statistical foundations of the procedure are not very solid[76]; see also discussion in ref. 77. In the section on detection and attribution of climate change, we describe how the equations describing Hasselmann's optimal fingerprinting method can gain new insights when reframed within the response theory of dynamical systems, and, in particular, can be derived seamlessly from linear response formalism.

Theory-guided and data-driven model reduction

We present in this section the quest for stochastic model reduction, in a more general setting than for the slow–fast systems discussed above, as framed here within the closure problem from the point of view of the Mori–Zwanzig (MZ) expansion. Reduced-order modelling is inextricably associated with performing partial observations: that is, gaining only a partial, imperfect knowledge of the properties of a system. As clarified by the MZ formalism[78,79], the effective dynamics defining the evolution on the projected space of the variables of interest has three groups of terms — Markovian deterministic, stochastic and non-Markovian components — even if the dynamics of the whole system is purely deterministic[80,81].

In the first two subsections, we review the mathematics behind this expansion and the different approaches adopted in the literature to approximate its elements, in particular regarding the memory and stochastic terms. However, we argue that, contrary to common belief, there are physically relevant regimes for which the three groups of terms do not share the same predominance, even in the presence of no timescale separation. Thus, although the timescale separation discussed above is assumed to be infinite, that assumption can be relaxed.

We then clarify physical situations in which the role of memory effects in the reduced model is secondary, whereas that of stochastic parameterization is key to recover the multiscale dynamics. In the subsection on the Lorenz 1980 (L80) model[82], we give an important example tied to primitive equations, the fundamental equations of atmosphere–ocean dynamics, in regimes for which the absence of timescale separation is manifested in each of the model's variables by bursts of fast dynamics, popping out irregularly in time, on top of a slow trend motion. There, we review how the proper knowledge of the manifold capturing this slow trend motion (tied to Rossby waves) enables us to figure out that the dynamics outside of this manifold is tied to inertia-gravity waves (IGWs) and that the latter can be efficiently parameterized by means of coupled stochastic oscillators, without the need of memory terms. Another example is reviewed in the subsection on neural turbulence, for the closure problem of forced 2D turbulence. There, we discuss how the recent advances on machine-learned parameterizations gain access to accurate coarse-grained closures without the need for memory or even noise terms. Such results are nevertheless subject to the choice of the cutoff scale, as lowering this cutoff is inevitably prone at some point to the emergence of such terms. In the final subsection, we review, in the context of a data-driven modelling problem of coarse-grained oceanic turbulence, the importance of the choice of latent variables for simplifying the equation discovery problem (in the manner of the MZ method) from time-sequential data that have large spatial dimensions.

Mori–Zwanzig decomposition

We consider a dynamical system of the form

$$\dot{\mathbf{U}} = \mathbf{F}(\mathbf{U}),\ \mathbf{U} \in \mathbb{R}^N. \tag{8}$$

The state vector $\mathbf{U}$ can be assumed here to be high-dimensional, describing a the state of a group — or subgroup — of climate variables (atmospheric variables, streamfunctions and so on). Equation (8) can be thought as resulting from discretization of a system of partial differential equations (PDEs) describing the motion of geophysical fluids or the evolution of other climate variables[40]. As such, equation (8) does not necessarily belong to the class of slow–fast systems such as equation (1), although the latter can be recast into the general abstract setting presented here. We denote by $S(t;\mathbf{U})$ the solution to equation (8) emanating from $\mathbf{U}$ in $\mathbb{R}^N$ at some initial time.

We assume that long-term statistics such as power spectral densities or correlations are well defined from equation (8), namely that equation (8) possesses an invariant probability distribution μ, also known as stationary statistical equilibrium, that satisfies

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T \varphi(S(t;\mathbf{U}))\,\mathrm{d}t = \int \varphi(\mathbf{U})\,\mathrm{d}\mu(\mathbf{U}), \tag{9}$$

for almost every $\mathbf{U}$ (in the Lebesgue sense) that lies in the basin of attraction of μ, for any sufficiently smooth observable φ. In general, φ is a field quantity that represents perturbations of density, pressure, electrostatic potential and so on. Note that when a global attractor exists, a statistical equilibrium μ satisfying equation (9) is supported by the global attractor[83] and gives the asymptotic 'mother' distribution of the dynamics over this attractor from which any probability density functions (PDFs) are derived (through marginals, for instance). We remark that such statistical equilibrium should not be confused with the classical notion of thermodynamical equilibrium in statistical mechanics.

Given an observable $\varphi : \mathbb{R}^N \to \mathbb{R}$, recall that the evolution of this observable along the flow associated with equation (8) is given by the Koopman semigroup, K_t, defined as[84]

$$K_t\varphi(\mathbf{U}) = \varphi(S(t;\mathbf{U})), \tag{10}$$

that satisfies the Liouville equation $\partial_t K_t\varphi = \mathcal{L}K_t\varphi$ with the operator $\mathcal{L}$

$$\mathcal{L}\varphi = \left(\sum_{j=1}^{N} F^j(\mathbf{U})\partial_j\right)\varphi, \tag{11}$$

denoting the Lie derivative along the vector field $\mathbf{F}$, while the F^j denote the components of the latter.

We are now given a decomposition $\mathbf{U} = (\mathbf{u},\mathbf{v})$ in which $\mathbf{u}$ denotes a vector of p relevant variables, typically those resolved, and $\mathbf{v}$ denotes the vector collecting the $q = N - p$ neglected ones. We are interested in describing the evolution of any sufficiently smooth observable ψ of the variable $\mathbf{u}$ only, without having to resolve the $\mathbf{v}$ equation in equation (8). In climate dynamics, this closure problem is motivated, for instance, by predicting or simulating a scalar field of interest (such as temperature or pressure) over a coarse grid without having to resolve the subgrid processes. Within this context, one may think of $\mathbf{u}$ as a collection of coarse-scale variables, and $\mathbf{v}$ as a collection of small-scale ones. In this case, equation (8) is an abstract formulation of a slow–fast system; in what follows, the variable $\mathbf{u}$ is referred to as slow and corresponds to the $\mathbf{x}$ variables in equation (1), and $\mathbf{v}$ is referred to as fast and corresponds to the $\mathbf{y}$ variables in equation (1).

In operator form, given an observable ψ of the coarse-scale variable $\mathbf{u}$ only, the problem consists of finding a parameterization in the following transport equation of the coefficients depending on the unresolved variables $\mathbf{v}$,

$$\partial_t\psi = \left(\sum_{j=1}^{p} F^j(\mathbf{u}(t),\mathbf{v}(t))\partial_j\right)\psi, \tag{12}$$

in order to describe the time evolution of $\psi(\mathbf{u}(t))$ without having to resolve the evolution of the $\mathbf{v}$ variables. This PDE, describing how any observable ψ of the reduced state space is advected by the flow of equation (8), is obtained by observing that $\partial_t\psi(\mathbf{u}(t)) = \nabla_{\mathbf{u}}\psi\cdot\partial_t\mathbf{u}$. Thus, the key issue we aim at solving is an effective parameterization of the interactions between the coarse-scale and subgrid variables $\mathbf{v}$ (via the terms F^j) for an accurate description of the advection of ψ in terms of the resolved variable $\mathbf{u}$ only.

To address this closure problem, associated with the cutoff scale consisting of retaining the $\mathbf{u}$ variable, we introduce the conditional expectation operator acting on observables of the full state vector $\mathbf{U} = (\mathbf{u},\mathbf{v})$

$$[P\varphi](\mathbf{u}) = \int \varphi(\mathbf{u},\mathbf{v})\,\mathrm{d}\mu_{\mathbf{u}}(\mathbf{v}) = \overline{\varphi}, \tag{13}$$

in which $\mu_{\mathbf{u}}$ denotes the disintegration of the invariant probability measure μ (ref. 85); roughly speaking, it gives the distribution of the $\mathbf{v}$ variable on the attractor when the coarse-scale variable is frozen to $\mathbf{u}$. The operator P thus corresponds to an averaging with respect to the neglected variable $\mathbf{v}$ as conditioned on $\mathbf{u}$. Note that $\mu_{\mathbf{u}} \neq \mu_{\mathbf{u}'}$ for $\mathbf{u} \neq \mathbf{u}'$ for complex systems, causing the variable $\mathbf{v}$ not to be identically and independently distributed (that is, not i.i.d.).

Now, by rewriting the transport equation (12) as $\partial_t(K_t\psi) = K_t\mathcal{M}\psi$ where $\mathcal{M} = \sum_{j=1}^{p} F^j\partial_j$, we have that

$$\partial_t(K_t\psi) = K_tP\mathcal{M}\psi + K_t(\mathcal{M}\psi - P\mathcal{M}\psi). \tag{14}$$

Here, $P\mathcal{M} = \sum_{j=1}^{p}\overline{F^j}(\mathbf{u})\partial_j$ denotes the averaged advection operator with respect to the neglected, unresolved variable $\mathbf{v}$.

By performing the change of variable $s \leftarrow t - s$ in the integral term of equation (31) in Box 2, we arrive finally at the following equivalent formulation of equation (31),

$$\partial_t\psi = P\mathcal{M}\psi(\mathbf{u}(t)) + \int_0^t \Gamma(t-s)\psi(\mathbf{u}(s))\mathrm{d}s + \eta(t)\psi, \tag{15}$$

which gives the desired closure of equation (12). The (possibly nonlinear) kernel operator $\Gamma(t-s)$ is typically a time-lagged damping

Box 2

Derivation of the generalized Langevin equation

In equation (14), the operator

$$\mathcal{M} - P\mathcal{M} = \sum_{j=1}^{p}(F^j(\mathbf{u},\mathbf{v}) - \overline{F^j}(\mathbf{u}))\partial_j$$

accounts for the fluctuations with respect to the conditional average and $K_t(\mathcal{M}\psi - P\mathcal{M}\psi)$ informs on how these fluctuations are transported by the flow of equation (8) for any observable ψ of the reduced state space. Hence, the operator $\boldsymbol{\delta}f = (\mathrm{Id} - P)\mathcal{M}f = Q\mathcal{M}f$, where Id is the identity, encodes the fluctuation terms. It defines the fluctuation semigroup $\mathrm{e}^{t\boldsymbol{\delta}}$ that constitutes a key element in the closure of equation (12). This closure is obtained by application of the perturbation theory of semigroups in the Miyadera–Voigt variation-of-constants formulation[249,296,297] to this fluctuation semigroup and the Koopman operator given by equation (10). The Miyadera–Voigt formula (29) below is also known as the Dyson's formula in the MZ literature[80].

The Miyadera–Voigt perturbation theorem[249] gives then that

$$K_tf = \mathrm{e}^{t\boldsymbol{\delta}}f + \int_0^t K_{t-s}P\mathcal{M}\mathrm{e}^{s\boldsymbol{\delta}}f\,\mathrm{d}s, \tag{29}$$

for any observable f for which $\boldsymbol{\delta}f$ is well defined. Note that $P\mathrm{e}^{t\boldsymbol{\delta}}f = 0$ if $Pf = 0$, that is, the orthogonal complement of P is invariant under $\mathrm{e}^{t\boldsymbol{\delta}}$. The fluctuation semigroup $\mathrm{e}^{t\boldsymbol{\delta}}$ gives the solution of the orthogonal dynamics equation

$$\partial_t\mathrm{e}^{t\boldsymbol{\delta}}f = Q\mathcal{M}\mathrm{e}^{t\boldsymbol{\delta}}f. \tag{30}$$

We refer to ref. 298 for the study of the existence of solutions to equation (30).

Now let us take $f = \mathcal{M}\psi - P\mathcal{M}\psi = Q\mathcal{M}\psi$ in equation (29) and observe that $Pf = 0$. Then equation (14) becomes

$$\begin{aligned}\partial_tK_t\psi &= K_tP\mathcal{M}\psi + \mathrm{e}^{t\boldsymbol{\delta}}f + \int_0^t K_{t-s}P\mathcal{M}\mathrm{e}^{s\boldsymbol{\delta}}f\,\mathrm{d}s\\ &= K_tP\mathcal{M}\psi + \int_0^t K_{t-s}\Gamma(s)\psi\,\mathrm{d}s + \eta(t)\psi,\end{aligned} \tag{31}$$

with $\eta(t)\psi = \mathrm{e}^{t\boldsymbol{\delta}}Q\mathcal{M}\psi$ denoting the orthogonal element of the MZ decomposition (since $PQ\mathcal{M}\psi = 0$ implying that $\eta(t)\psi$ lies in $\ker(P)$), while $\Gamma(s)$ defines the operator $\Gamma(s) = P\mathcal{M}\eta(s)$.

kernel, and $\eta(t)$ is interpreted as an effective random forcing uncorrelated with the time evolution of the resolved variable, $\mathbf{u}(t)$, but can be strongly correlated in time, as discussed below. In the slow–fast system metaphor, the Markovian term $P\mathcal{M}$ provides the slow component of the dynamics, $\eta(t)$ is void of slow oscillations, and Γ is intended to account for the disparate interactions between the timescales.

Calculating the terms in the MZ decomposition

Equation (15) is called the generalized Langevin equation (GLE)[60] or the MZ decomposition. Thus, the effective dynamics defining the evolution of any observable of the reduced state space can be achieved by determining the Markovian, stochastic and non-Markovian components appearing in equation (15), making the GLE the fundamental equation to determine in the MZ approach to closure[78–81,86]. We clarify below a subtle point for applications, namely that not all the terms are necessarily needed in this triptych decomposition to derive accurate closures for multiscale dynamics even when no timescale separation is apparent. In the special case of one-way coupling between the subgrid and resolved variables, the non-Markovian component mentioned above disappears[87]. We discuss below a fully coupled model from atmospheric dynamics in which the non-Markovian terms are negligible but the stochastic ones are essential to recover the multiscale nature of the dynamics from closure.

As elegant it may be, the MZ decomposition is a technically challenging solution to the closure problem of disparate-scale interactions, and various assumptions about the memory kernel Γ are typically made to propose approximations to the GLE.

The memory kernel Γ and the 'noise' operator $\eta(t)$ involve the implicit knowledge of the fluctuation semigroup $e^{s\delta}$, accounting for the effects of the neglected variables on the fluctuations with respect to the average motion. This operator is difficult to resolve as it boils down to solving the orthogonal dynamics equation (equation (30))[88].

The noise and memory terms can be extremely complicated to calculate, especially in cases with weak or no obvious timescale separation between the resolved and unresolved variables. The approximation of these terms thus constitutes the main theme of most research on the MZ decomposition.

Many techniques have been proposed to address this problem in practice, and they can be grouped into two categories: data-driven methods, and methods based on analytical insights tied to the derivation of the MZ decomposition. Data-driven methods aim to recover the MZ memory integral and fluctuation terms based on data, by exploiting sample trajectories of the full system. Data-driven methods can yield accurate results, but they often require many sample trajectories to faithfully capture memory effects[89–92]. Typical examples include the NARMAX (nonlinear auto-regression moving average with exogenous input) technique[93–95], the rational function approximation proposed in ref. 90, the conditional expectation techniques of ref. 92, methods based on Markovian approximations by means of surrogate hidden variables[90,96–99], or kernel-based linear estimators in delay-coordinate[100].

Methods based on analytical considerations aim at approximating the MZ memory integral and fluctuation terms based on the original model's equations, without using any simulation data. The first effective method developed within this class can be traced back to the continued fraction expansion[101], which can be conveniently formulated in terms of recurrence relations[102,103]; see also ref. 104. Other theoretically guided methods to compute the memory and fluctuation terms in the MZ decomposition include optimal prediction methods[80,105,106], mode coupling techniques[107,108], methods based on approximations of the orthogonal equations[109], matrix function methods[110], series expansion methods[111–115], perturbation methods[116] and methods based on Ruelle's response theory[117,118]. These analytically grounded methods can lead to effective calculations of non-Markovian effects in various applications such as coarse-grained particle simulations[119,120] or some fluid problems[112,113], including intermediate-complexity climate models[121]. However, these calculations are often involved, and they do not generalize well to systems with no scale separation[81]; see, instead, an example of scale adaptivity in ref. 122.

To better appreciate the difficulty posed by the lack of timescale separation, it is useful to recall that, for instance, a long-range memory approximation consisting of keeping the zeroth-order term in a Taylor expansion of the memory operator in equation (15) simplifies the memory term calculation, but at the price of restrictive conditions. Indeed, such a long-range approximation shows relevance if the unresolved modes exhibit sufficiently slow decay of correlations (t-model[105,123]), essentially by assuming information about the initial value to be sufficient to make predictions. Assuming the unresolved modes to have fast decay of correlations, one is left with short-range memory approximation schemes. The two cases, of extreme or very weak non-locality in time, are two sides of the same coin[105]. Most of the challenging cases for closure lie thus in the intermediate cases[106], for which there is no neat separation of timescales such as exists in climate science[6].

Keeping higher-order terms in the Taylor expansion of the memory operator is a natural way to handle cases of weak timescale separation. It is illuminating in many ways, including to design data-driven methods, as explained below. This higher-order approximation approach of the memory operator has been retained in ref. 88 and further developed and analysed in ref. 114. The approach consists of breaking down the memory approximation problem into a hierarchy of auxiliary Markovian equations.

Denoting by $\mathfrak{m}_0(t)$ the integral term in equation (15), such a Markovian approximation is accomplished by observing that $\mathfrak{m}_0(t)=\int_0^t K_s P\mathcal{M}e^{(t-s)Q\mathcal{M}}Q\mathcal{M}\,\mathrm{d}s$ satisfies the following infinite-dimensional system of PDEs[88,114]

$$\frac{\mathrm{d}\mathfrak{m}_{n-1}}{\mathrm{d}t}=K_t P\mathcal{M}(Q\mathcal{M})^n v+\mathfrak{m}_n(t),\quad n=1,\ldots. \tag{16}$$

where $Q\mathcal{M}$ is the generator of the orthogonal equation (equation (30) in Box 2). Integrating equation (16) backward, that is, from the 'last' equation to the first one, yields a Dyson series representation of $\mathfrak{m}_0(t)$ involving repeated integrals[114]. In practice, one performs a truncation of equation (16), which consists of keeping the first n equations, while closing the last equation by using an ansatz in place of $\mathfrak{m}_n(t)$, such as $\mathfrak{m}_n(t)=0$ (ref. 88) or $\mathfrak{m}_n(t)$ given by Chorin's t-model, among other choices[114]. Depending on the order of truncation retained and the corresponding choice of the ansatz for $\mathfrak{m}_n$, error estimates with respect to the genuine memory integral in equation (15) are available[114]. However, the implementation of such Markovian schemes is not trivial, as it requires computing $(Q\mathcal{M})^n$ to a high order in n, a delicate operation especially when the original system is large.

Nevertheless, the layered structure of equation (16) and related error estimates provide a strong basis for the design of data-driven methods based on Markovianization ideas to approximate the memory integral term. Such ideas are commonly used for the mathematical analysis of physical models involving integro-differential equations; see, for example, refs. 83,124 and references therein.

The data-driven approach proposed in ref. 125 is intimately related to such Markovianization ideas for the MZ decomposition[97]. The class of data-driven models of[125] involves also multilayered SDEs of a structure very similar to that of equation (16) that has been generalized[97] to handle the approximation of more complex memory kernels, from a data-driven perspective. The usage of such multilayered SDEs to provide approximation of the GLE is further discussed in the next subsection.

Efforts to approximate the memory and noise terms should not, however, make us lose sight of another key problem, that of approximating the conditional expectation, namely the Markovian terms in equation (15). This is where recent hybrid approaches exploiting the original model's equations and simulated data have shown relevance and have indicated that a blind application of data-driven methods can lead to uncertain outcomes or incorrect interpretations even when the latter are successful[126,127].

In that respect, the data-informed and theory-guided variational approach introduced in ref. 126 allows for computing approximations of the conditional expectation term, $P\mathcal{M}$, by relying on the concept of the optimal parameterizing manifold (OPM)[126]. The OPM is the manifold that averages out optimally the neglected variables as conditioned on the resolved ones[126]. The approach to determine OPMs introduced in ref. 126 in practice consists of first deriving analytic parametric formulas that match rigorous leading approximations of unstable/centre manifolds or slow manifolds near, for example, the onset of instability, and then to perform a data-informed minimization of a least-square parameterization defect to recalibrate the manifold formulas' parameters to handle regimes further away from that instability onset[126]. There, the optimization stage alleviates the small denominator problems[128] rooted in small spectral gaps[129], and makes it possible to derive accurate parameterizations in regimes where constraining spectral-gap or timescale-separation conditions are responsible for the well-known failure of standard invariant/inertial or slow manifolds[130–132]; see refs. 126,129 for examples. In more physical terms, this problem is also tied to deficient or excessive parameterization of the small-scale energy but dynamically important variables, leading to an incorrect reproduction of the backscatter transfer of energy to the large scales[133–135], and to inverse cascade errors[135–137].

For multiscale dynamics, failure to accurately resolve the conditional expectation results typically in a residual that contains too many spurious frequencies to be efficiently resolved by data-driven methods based on, for example, the aforementioned multilayered SDEs. Such multilayered SDEs exploit either polynomial libraries of functions or other specified interaction laws[97] between the resolved and unresolved variables.

In recent years, machine-learning (ML) techniques have been used to learn memory terms in MZ decompositions[98,99,138–140]. These go beyond prior efforts involving polynomial libraries of specific interaction laws between the slow and fast variables[97,118,125]. Thus, one may be tempted to use complex neural architectures to learn the MZ terms, but this should not be done at the price of physical interpretations and understanding. Careful studies in that regard include refs. 98,99. The examples discussed below provide other elements for caution.

Variational approach to closure

In the context of subgrid parameterizations, non-locality in time in equation (15) means that the subgrid variables exert both reactive and resistive forces on the resolved variables; and such attributes of the subgrid variables may play an important role in reproducing finite-amplitude instabilities and other properties of the resolved variables[141]. In the absence of timescale separation, the subgrid variables exert fluctuating driving forces on the resolved variables, which are conceptually distinct from eddy viscosity (or even negative eddy viscosity)[142].

We assume thus that F in equation (8) proceeds from a forced fluid model, that is, that $F(\mathbf{U}) = L\mathbf{U} + B(\mathbf{U},\mathbf{U}) + \boldsymbol{f}$ where B is a bilinear operator, L is a linear dissipative operator, $\boldsymbol{f}$ is a time-independent force (to simplify), and $\mathbf{U} = (\mathbf{u},\mathbf{v})$ as above. We are interested in finding an accurate closure in the slow and coarse-scale $\mathbf{u}$ variable. To achieve this goal, the key issue is the parameterization of the $\mathbf{u}$–$\mathbf{v}$ and $\mathbf{v}$–$\mathbf{v}$ interaction terms in the original $\mathbf{u}$ equation, that is, the terms accounting for the disparate-scale and fast-scale interactions. Denoting by $\boldsymbol{\tau}_{\mathrm{int}}$ the grouping of these interaction terms, a convenient way to address this problem is by seeking the optimal parameterization, $\boldsymbol{\tau}_{\mathrm{opt}}$, which is the nonlinear vector field of the $\mathbf{u}$ variable that solves the minimization problem

$$\min_{\boldsymbol{\tau}} \int_0^T \|\boldsymbol{\tau}(\mathbf{u}(t)) - \boldsymbol{\tau}_{\mathrm{int}}(\mathbf{u}(t),\mathbf{v}(t))\|^2 \,\mathrm{d}t,\ T \gg 1. \tag{17}$$

$\boldsymbol{\tau}_{\mathrm{opt}}$ relates naturally to the conditional expectation (13) $\boldsymbol{F}$ because $\boldsymbol{\tau}_{\mathrm{opt}}(\mathbf{u}) = \overline{\boldsymbol{F}}$ minus the linear and $\mathbf{u}$–$\mathbf{u}$ interaction terms that project onto the coarse-scale variables.

The aforementioned OPM, Φ_{opt}, which provides the best approximation in a least-square sense of $\mathbf{v}$ as a mapping of $\mathbf{u}$, satisfies $\boldsymbol{\tau}_{\mathrm{int}}(\mathbf{u},\Phi_{\mathrm{opt}}(\mathbf{u})) \approx \boldsymbol{\tau}_{\mathrm{opt}}(\mathbf{u})$ (ref. 126), with a small residual error when the $\mathbf{v}$–$\mathbf{v}$ interaction terms are negligible after averaging in the original $\mathbf{u}$ equation. At this stage, knowing $\boldsymbol{\tau}_{\mathrm{opt}}$ or Φ_{opt} allows us thus to approximate the average motion of $\mathbf{u}(t)$ when averaging is performed over the unresolved variable $\mathbf{v}(t)$.

The fluctuations carried by $\mathbf{v}(t)$ onto the dynamics of $\mathbf{u}(t)$ have effects that are beyond the averaging approximation. To take them into account, the MZ formalism introduced above leads us to revise the minimization problem (17) as

$$\min_{\boldsymbol{\tau},\Gamma} \int_0^T \|\boldsymbol{\tau}(\mathbf{u}(t)) + \int_0^t \Gamma(t-s)\mathbf{u}(s)\mathrm{d}s - \boldsymbol{\tau}_{\mathrm{int}}(\mathbf{u}(t),\mathbf{v}(t))\|^2 \mathrm{d}t,\ T \gg 1. \tag{18}$$

Solving this second minimization problem thus consists of decomposing the nonlinear interaction term to account for a memory function and a fluctuating force, namely

$$\boldsymbol{\tau}_{\mathrm{int}}(\mathbf{u}(t),\mathbf{v}(t)) \approx \boldsymbol{\tau}(\mathbf{u}(t)) + \int_0^t \Gamma(t-s)\mathbf{u}(s)\,\mathrm{d}s + \boldsymbol{r}(t), \tag{19}$$

where $\boldsymbol{r}(t)$ is the residual obtained after minimization of equation (18). This minimization can be addressed by means of recurrent neural networks that allow functional dependence of the 'past', such as long short-term memory (LSTM) networks[138–140]. Neural networks enable learning of possibly complex functional dependences in $\boldsymbol{\tau}$ and of the operator Γ, but the resulting learned elements suffer from issues of interpretability.

Another approach to going beyond averaging consists of minimizing equation (18) via Markovianization. This method consists of breaking down the memory terms and noise terms by means of SDEs that have a multilayer structure (similar to equation (16)). The coefficients of the SDEs are learned successively via recursive regressions using surrogate stochastic variables that account for the residual errors produced by the successive regressions, until a white noise limit is reached[96,97]. This data-driven approach[125] has yielded striking results in many fields of applications, such as the modelling of El Niño–Southern Oscillation

(ENSO)[143–145], extratropical atmospheric dynamics[146], palaeoclimate[147] and the Madden–Julian Oscillation[148–150], to name a few.

These regression-based multilayered SDEs that approximate the MZ decomposition (15) benefit furthermore from theoretical insights. Intimate connections with the multilayered SDEs derived in refs. 117,118, based on Ruelle's response theory[151], were shown[151] to hold for a subclass of multilayered SDEs considered in refs. 96,97. These connections clarify the circumstances of success for multilayered SDEs with linear coupling terms between the layers that correspond to approximating the memory operator Γ in equation (15) by repeated convolutions of exponentially decaying kernels[97]. The multilayered SDEs of this form are particularly relevant for weakly coupled slow–fast systems, and the corresponding memory and noise terms relate naturally to the Koopman eigen-elements of the 'unperturbed weather' Koopman semigroup K_t^w whose generator is $\sum G^j(\mathbf{v})\partial_j$ when $\mathbf{g}(\mathbf{u},\mathbf{v}) = \mathbf{G}(\mathbf{v}) + \varepsilon C(\mathbf{u},\mathbf{v})$ in equation (1) with ε small; see ref. 151. The approximation of the MZ decomposition (15) via Markovianization thus sheds light on Koopman modes[84] and the related dynamic mode decomposition[152–154], as well as on the principal oscillation pattern method proposed earlier in atmospheric sciences by Hasselmann[155,156].

As mentioned above, however, depending on the problem, it is not always necessary to determine the memory and/or noise terms. One should thus always look first at the virtue of solving the minimization problem (17), instead of solving the more challenging minimization problem (18), which may involve memory or noise terms of negligible importance for closure depending, for example, on the cutoff scale retained for a given dynamical regime, as discussed in the subsection on neural turbulent closures.

An emblematic example is also found in the context of the primitive equations of the atmosphere. It is known that at low Rossby number, the conditional expectation that coincides with the balance equation (BE) suffices for an accurate closure[157]. However, once a critical Rossby number is crossed, the BE must be seriously amended to capture the complex interactions between the Rossby waves and IGWs; the latter become non-negligible at large Rossby number, as discussed below.

Closure of the atmospheric Lorenz 1980 model

On a timescale of days, atmospheric and oceanic flows constrained by Earth's rotation satisfy an approximately geostrophic momentum balance on larger scales, associated with slow evolution, but the flows also exhibit fast IGWs dynamics. Central to geophysical fluid dynamics are the slow manifold problem[158], namely that of identifying the slow component (for weather forecast initialization[159–162], for example), and the problem of characterizing slow–fast interactions. The Lorenz 63 model[163], famous for its chaotic strange attractor, is a paradigm for the geostrophic component, whereas the L80 model[82] is its paradigmatic successor, both for the generalization of slow balance and for slow–fast coupling. The closure of the atmospheric L80 model has Markovian and noise terms but is memoryless, as discussed below.

Unlike other slow–fast systems, this physically based model exhibits regimes with energetic bursts of fast oscillations superimposed on slow ones in each variable of the model; these fast oscillations greatly complicate the parameterization of the regimes[164] (Box 3). Regimes beyond exponential smallness of the fast oscillations are not only part of the L80 model. They have been observed in other primitive equation models as conspicuously generated by fronts and jets[165,166], and in cloud-resolving models in which large-scale convectively coupled gravity waves spontaneously develop[167]. Regions of organized convective activity in the tropics also generate gravity waves that lead to a spectrum that contains notable contributions from horizontal wavelengths of scales from 10 km to beyond 1,000 km (ref. 168). Such IGWs have also been identified from satellite observation of continental shallow convective cumulus, which forms organized mesoscale patterns over forests and vegetated areas[169].

The L80 model provides a remarkable paradigm of such regimes with a lack of timescale separation at large Rossby numbers, in which the solutions have slow and fast components (mixture of high and low frequencies, HLF). In these regimes, fast variables are a function of the slow variables at the same time instant, thus calling for a revision of slow manifold methods[158] and the like. The generic elements for solving such hard closure problems have been identified only in recent years[170]. Key to the solution is the balance equation (BE) manifold[171,172] as rooted in earlier works[159,173–175]. The BE manifold has been shown to provide, even for large Rossby number, the slow trend motion of HLF solutions to the L80 model as it optimally averages out the fast oscillations; thus it approximates the OPM, Φ_{opt}, to a high precision[164].

For such regimes, the L80 dynamics evolves on this manifold and experiences excursions away from it, corresponding to bursts of fast oscillations caused by IGWs (Box 3 and Fig. 2a). The residual off the BE manifold is mainly orthogonal to it, causing the memory terms to be negligible[170] and making the stochastic modelling of the $\eta(t)$ term central in the MZ decomposition (15). An inspection of this residual in the time domain shows that it is strongly correlated in time, narrowband in frequency and modulated in amplitude (Fig. 2b). Progress in characterizing the spectral signature in terms of Ruelle–Pollicott resonances and Koopman eigenvalues (and the like[176,177]) of such time series[62,178] indicates that such residuals can be efficiently modelled by means of a network of Stuart–Landau oscillators (SLOs) of the form

$$\begin{aligned}\dot{z}_j = (\mu_j + \mathrm{i}\omega_j)z_j - (\alpha_j + \mathrm{i}\beta_j)z_j|z_j|^2 + \text{coupling terms}\\ + \text{white noise terms}\end{aligned} \tag{20}$$

in which the μ_j and ω_j denote rates of growth and fundamental frequencies, while the α_j and β_j are positive parameters; see Fig. 2c and ref. 170 for more details. Remarkably, the BE manifold provides a nonlinear separation of variables that makes it possible to decompose the mixed HLF dynamics of the L80 model into a slow component captured by the BE, and a fast one modelled by a network of SLOs.

For the L80 model, the resulting OPM–SLO closure is written for the streamfunction amplitude (the $\boldsymbol{y}$ variable; Box 3) and takes the form

$$\dot{\mathbf{y}} = \Pi_{\mathbf{y}}\left(L\mathbf{y} + B\left(\mathbf{y} + \Phi_{\text{opt}}(\mathbf{y}) + \xi_t\right) + \boldsymbol{f}\right), \tag{21}$$

in which $B(u)$ denotes the bilinear term $B(u,u)$ and $\boldsymbol{f}$ the forcing term from the original L80 model, $\Pi_{\mathbf{y}}$ denotes the projector onto the resolved variable $\mathbf{y}$, and the stochastic vector ξ_t is modelled by the auxiliary networks of SLOs in equation (20). Through the nonlinear interactions of the network of SLOs with the parameterization $\Phi_{\text{opt}}(\mathbf{y})$ of the slow motion and the $\mathbf{y}$ variable in equation (21), one can accurately recover the multiscale dynamics of the L80 model along with its complex bursts of fast oscillations caused by IGWs[170]. Qualitatively, one can say that although diagnostically the balanced flow captures most of the variance of the L80 model, its prognostics requires a careful representation of the fast components associated with the IGWs.

In terms of Hasselmann's programme, the L80 model shows that efficiently modelling regimes with a lack of timescale separation characterized by a mixture of intertwined slow and fast motions requires both a good approximation of the OPM capturing the slow motion

Box 3

The L80 model and bursts of inertia-gravity waves

The Lorenz 1980 (L80) model, obtained as a nine-dimensional truncation of the primitive equations onto three Fourier modes with low wavenumbers[82], can be written as

$$a_i\frac{dx_i}{dt} = \varepsilon a_i b_i x_i x_k - c(a_i - a_k)x_j y_k - c^2 y_j y_k + c(a_i - a_j)y_j x_k - N_0 a_i^2 x_i + \varepsilon^{-2} a_i (y_i - z_i),$$

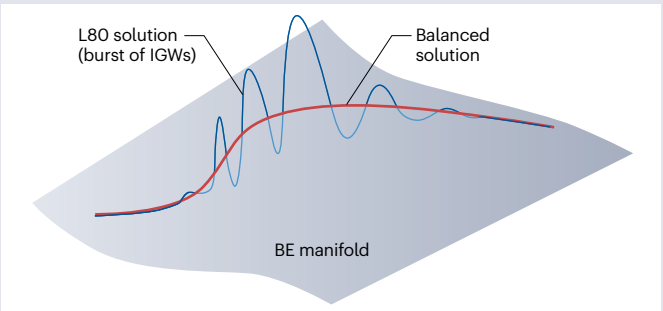
$$a_i\frac{dy_i}{dt} = -\varepsilon a_k b_k x_j y_k - \varepsilon a_j b_j y_j x_k + c(a_k - a_j)y_j y_k - a_i x_i - N_0 a_i^2 y_i,$$

$$\frac{dx_i}{dt} = -\varepsilon b_k x_j (z_k - H_k) - \varepsilon b_j (z_j - H_j)x_k + c y_j (z_k - H_k) - c(z_j - H_j)y_k + g_0 a_i x_i - K_0 a_i z_i + \mathcal{F}_i,$$

whose model's parameters are described in refs. 82,164.

The above equations are written for each cyclic permutation of the set of indices (1,2,3), namely, for (i,j,k) in $\{(1,2,3),(2,3,1),(3,1,2)\}$. The variables $(\mathbf{x},\mathbf{y},\mathbf{z})$ are amplitudes for the three Fourier modes of the divergent velocity potential, streamfunction and dynamic height, respectively. Transition to chaos occurs as the Rossby number ε is increased[164,172].

At small ε, the solutions to the L80 model remain entirely slow for all time (that is, dominated by Rossby waves) whereas fast oscillations emerge spontaneously and are superimposed on such slow solutions as the Rossby number is further increased (see the figure). In such regimes, the balance equation (BE) manifold on which balanced solutions lie[171,172] is no longer able to encode the dynamics (see the figure), as the dynamics associated inertia–gravity waves (IGWs) become transverse to the BE manifold[126]. These regimes with energetic bursts of IGWs lie beyond the parameter range explored in ref. 82 (see ref. 299) and beyond other regimes with exponential smallness of IGW amplitudes as encountered in the subsequent Lorenz 86 model[300–302] and the full primitive equations[131] at smaller Rossby numbers[303].



and to go beyond stochastic homogenization and the like[43] to model the noise, the use of network of SLOs showing a great deal of promises in that respect.

Note that by thinking of the bilinear terms B in equation (21) as proceeding from advective terms in the L80 model, one may interpret the nonlinear terms involving ξ_i in the stochastic OPM–SLO closure (21) as stochastic advective terms. Other approaches have shown the relevance of such terms for deriving stochastic formulations of fluid flows and for emulating the coarse-grained dynamics[179–182]. In particular, homogenization theory can be used to rigorously derive effective slow stochastic particle dynamics for the mean part, under the assumption of mildly chaotic fast small-scale dynamics[38]. Interestingly, the route taken for deriving equation (21) differs from homogenization techniques. Owing to specificities of the L80 dynamics, the approach indicates that even for regimes in which the timescale separation is violated, closure involving stochastic advection terms may still be relevant.

From a practical viewpoint, the interest of having an accurate stochastic closure of the stochastic advective form equation (21) lies in its ability to simulate key features of the multiscale dynamics, offline, in an uncoupled way to mimic the effects of IGWs. As a result, by simply running offline the network of SLOs (20) and plugging its output into equation (21) as a random forcing input, one recovers, by integrating online equation (21), the multiscale nature of the L80 dynamics through interactions with the nonlinear terms[170].

The OPM–SLO approach is thus promising for further application to the closure of other, more complex, slow–fast systems, in strongly coupled regimes. In particular, regimes that have a mixture of fast oscillations superimposed on slower timescales, as displayed by the L80 model, provide a challenging ground for closure in more sophisticated fluid problems. Such regimes are known to arise in multilayer shallow water models[183]. In certain regions of the oceans, IGWs account for approximately half of the near-surface kinetic energy at scales of 10–40 km (ref. 184), making IGWs energetic on surprisingly large scales. Thus, geophysical kinetic energy spectra can exhibit a band of wavenumbers within which waves and turbulence are equally energetic[185]. We believe in the ability of the OPM–SLO approach to show closure skills for such problems. There, the approximation of the OPM or conditional expectation should benefit from recent progress in neural turbulent closures, and the fast component of the motion should also benefit from the wealth of dynamics that networks of SLOs can embody, as discussed below.

Neural turbulent closures

Lately, much effort has been devoted to learning successful neural subgrid parameterizations, that is, subgrid parameterizations built upon neural networks, for the closure of fluid models in turbulent regimes such as the forced Navier–Stokes equations or quasi-geostrophic (QG) flow models on a β-plane[186–191]. Neural turbulent closures have no memory and no noise, but spatially non-local Markovian terms, as we discuss.

These neural closure results are typically obtained with convolutional neural networks (CNNs)[192] that are by definition non-local in space and aim at parameterizing the subgrid scale (SGS) stress tensor in terms of coarse-grained variables. The achievements of these neural closures include their ability to provide accurate closures for cut-offs within the inertial range and for high Reynolds numbers, outperforming more standard schemes such as those based on Smagorinsky parameterizations[191].

Obtaining accurate closures is known to be challenging as small errors made on the parameterization of the SGS are typically backscattered, owing to the inverse cascade, into amplified errors at large scales[193,194]. Disposing of SGS parameterizations at low cutoff levels for such turbulent flows with a controlled error is thus of prime importance for the closure problem.

The accuracy and stability of closure results reported in[186–191] have an interesting mathematical interpretation. These results hint at the existence of a nonlinear functional relationship of the form (after removal of transient dynamics),

$$\boldsymbol{\tau} = \boldsymbol{\tau}_{\mathrm{CNN}}(\bar{u}, \bar{v}) + \varepsilon, \tag{22}$$

in which $\boldsymbol{\tau}$ denotes the SGS, $\boldsymbol{\tau}_{\mathrm{CNN}}$ denotes the parameterized SGS, $\bar{u}$ and $\bar{v}$ denote the coarse-grained velocity variables (not to be confused with **u** and **v** used above), and the spatiotemporal residual ε is small in a mean-square sense.

In equation (22), $\boldsymbol{\tau}_{\mathrm{CNN}}$ denotes the function found by means of CNNs trained by minimizing a loss function reminiscent of that involved in equation (18). The relation (22) based on the quality of the closure results reported in refs. 186–191 thus suggests that $\boldsymbol{\tau}_{\mathrm{CNN}}$, in the respective cases, is close to the conditional expectation $\overline{\boldsymbol{\tau}}$ (ref. 126), namely the best nonlinear functional that averages out the unresolved variables as conditioned on the coarse variables. Thus, finding a good approximation of $\overline{\boldsymbol{\tau}}$ is sufficient for an accurate closure of forced 2D turbulence problems at high Re, at least for a range of physically and computationally interesting cutoffs within the inertial range.

Accordingly, for such turbulent flows and choice of cutoffs, these neural closure results seem to rule out the use of memory terms in the MZ expansion, thus questioning the need for memory terms in other closure studies for similar problems[112,127,195]. For instance, memory terms have been advocated for the closure of Kuramoto–Sivashinsky (KS) turbulence in the reduced state space spanned by the unstable modes in, for example, refs. 94,127, whereas learning the conditional expectation suffices for high-skill closure that retains only the unstable modes, and for even more turbulent regimes[126].

In a certain sense, one can thus argue that the neural turbulent closure results of refs. 186–191 restore some credentials to old ideas of the late 1980s[196,197] where turbulent solutions to the 2D Navier–Stokes equations were pursued as trajectories evolving in the phase space, within some thin neighbourhood of a finite-dimensional manifold parameterizing the small scales in terms of the large ones[126,196]. Attempts to construct these manifolds based on analytical formulas derived from the model's equations were successful only for cutoff wavenumbers within or close to the dissipation range[198]. The use of neural networks thus opens up new perspectives on this old problem. The aforementioned results show indeed that data-driven formulas of such manifolds are accessible for much lower and computationally relevant cutoffs within the inertial range. However, note that lowering the cutoff scale is inevitably prone to the emergence of memory and/or stochastic terms at some point. Such is the case, for instance, in closing KS turbulence when the cutoff scale is chosen such that unstable modes are present in the space of scales to parameterize.

Choice of latent variables in a baroclinic ocean model

The MZ framework is conditioned to the choice of resolved and neglected variables inherent to that of the reduced state space in which a closure is sought. This aspect is actually a key step when dealing with the equation discovery problem from time-sequential data that have large spatial dimensions, such as in climate applications. Then, one typically uses dimensional reduction to compress the original set of variables into a few variables aimed at simplifying the computational burden of finding the governing equations.

The most common method for dimensionality reduction is principal component analysis (PCA), also known as empirical orthogonal function (EOF) decomposition[199,200]. In it, the principal components constitute the latent variables. PCA is commonly used to infer, out of various climate fields, stochastic differential models that are either linear[201–204] or include nonlinear terms and various degrees of approximations of memory terms[143–146,205,206]. Many other methods of dimensionality reduction could be used at this stage, such as those based on

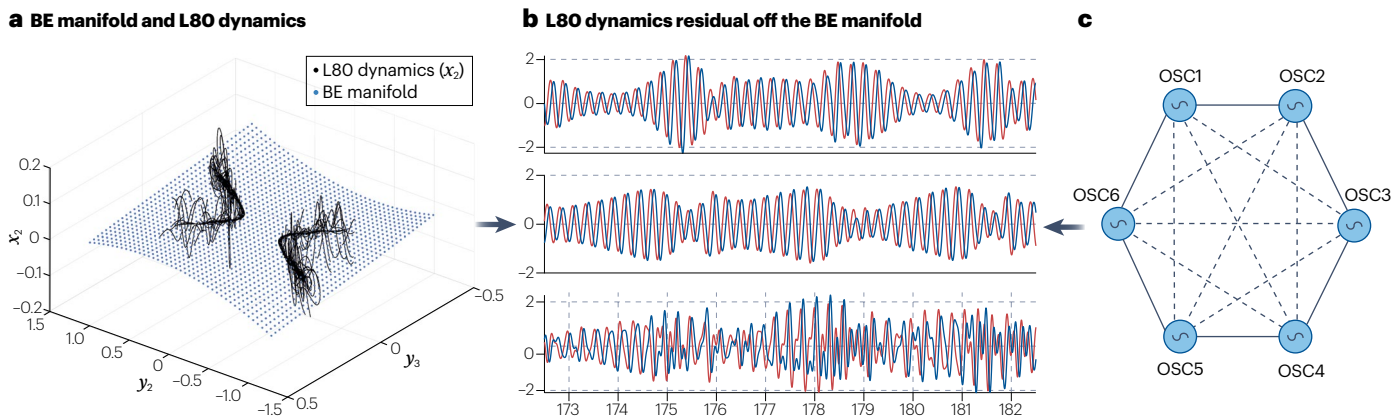


Fig. 2 | Example: Mori–Zwanzig decomposition without memory but with a noise term. a, The optimal parameterizing manifold (OPM) that allows expression of the $\mathbf{x} = (x_1, x_2, x_3)$ and $\mathbf{y} = (y_1, y_2, y_3)$ variables in terms of the **y** variable is the balance equation (BE) manifold (blue dots). It provides the slow motion of the Lorenz 1980 (L80) dynamics. The L80 dynamics (black curve) evolves onto this manifold and experiences excursions away from it, corresponding to bursts of fast oscillations caused by inertia-gravity waves (IGWs) (Box 3). The residual off the BE manifold is mainly orthogonal to it, causing memory terms to be negligible, and making their stochastic modelling central as the $\eta(t)$ term in the MZ decomposition (15). **b**, This residual in the time domain is strongly correlated in time and can be grouped in pairs that are narrowband in frequency and modulated in amplitude with a possible combination of 'tones' (lower panel). **c**, Networks of stochastic oscillators (OSCs) are well suited to model such properties. Part **b** adapted with permission from ref. 170, National Academy of Sciences.

nonlinear[207,208] or probabilistic[209] versions of EOFs, spectral versions of EOFs[210] and the like[176,211], transfer or Koopman operators[212,213], Laplacian and diffusion maps[214–216], or variational encoders[217], to name a few.

Whatever the dimensionality reduction method used, the latent variables may display a frequency mixture issue as encountered above for the L80 model, affecting the timescale one wishes to resolve via a reduced model. This issue is encountered, for example, when aiming at reduced models on decadal timescales of fully resolved wind-driven baroclinic QG models of the ocean. The ocean circulation of an eddy-resolving simulation with ~10^6 spatial degrees of freedom at reference model parameters[218] is characterized by a robust large-scale decadal low-frequency variability with a dominant 17-year cycle, involving coherent meridional shifts of the eastward jet extension separating the gyres (Fig. 3a). On this decadal variability is superimposed an interannual variability caused by the eddy dynamics[219]. Owing to the highly turbulent and multiscale nature of the flow, the capture of the eddies' dynamics on a coarse grid by a reduced model is challenging.

Within the reduced state space of (the first few dominant) principal components, this challenge is manifested by the multiscale nature of the temporal evolution of principal components: a slow evolution (decadal) contaminated by 'fast' interannual oscillations due to the eddy dynamics (Fig. 3f). Such multiscale features are the main cause of, for example, the failure of multilayered SDEs (such as those of refs. 96,125) in approximating the memory and noise terms in the MZ decomposition, in spite of their successes in other geophysical problems as recalled in the subsection on the variational approach to closure. The underlying reason is in the set of predictor functions used to learn the multilayered SDE ingredients, either responsible for an explanatory deficit, or subject to a spectral bias if conventional neural networks such as multilayer perceptrons are used[220].

This is a case where multivariate signal decomposition methods[176,221] may offer an alternative, by extracting from data frequency-ranked coherent modes of variability. Indeed, such methods, when effective in separating the slow and fast temporal components of the PCs (or analogues), provide a natural ground for modelling these temporal components by means of stochastic SLOs such as in equation (20), ranked by frequency to be resolved. They are also known as multiscale Stuart–Landau models (MSLMs)[176]. MSLMs have demonstrated skills in modelling challenging Arctic sea-ice datasets that have nonlinear trends[222,223]. The MSLMs provide coarse-grained models

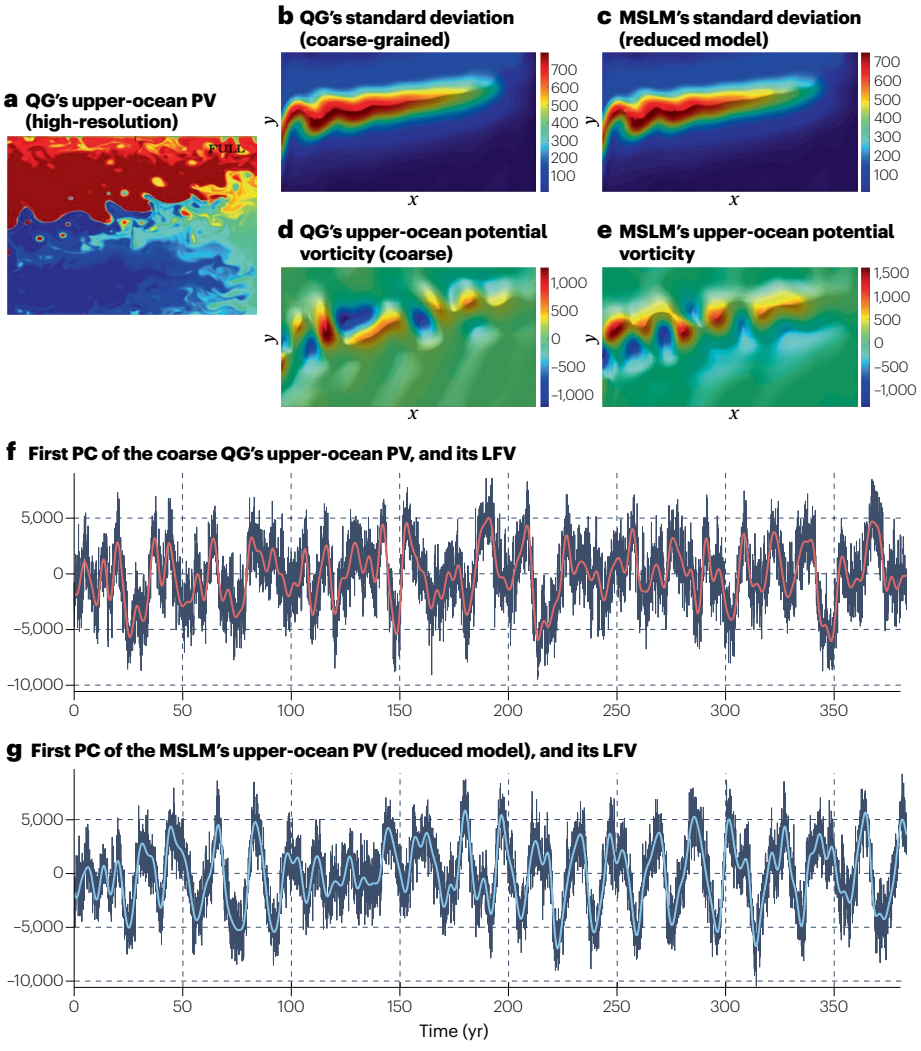


Fig. 3 | The time evolution and standard deviation of upper-ocean potential vorticity anomalies in quasi-geostrophic turbulence and a multiscale Stuart–Landau model. a, Instantaneous upper-ocean potential vorticity (PV) anomaly field from a high-resolution simulation (HRS) of a baroclinic quasi-geostrophic turbulent model (QG). **b**, Standard deviation of this HRS over a 64 × 26 coarse grid. **c**, Standard deviation as simulated by the multiscale Stuart–Landau model (MSLM) reduced model over the same coarse grid. **d**,**e**, Instantaneous upper-ocean PV anomalies from the coarse-grained QG model (part **d**) and its MSLM reduced model (part **e**). **f**,**g**, First principal component (PC) (red) of the coarse QG (part **f**) and MSLM (part **g**) upper-ocean PV, and its decadal low-frequency variability (LFV) content (blue). The MSLM is able to reproduce the multiscale temporal variability of the QG coarse-grained dynamics remarkably well. Part **a** adapted with permission from ref. 293 under a Creative Commons licence CC BY 4.0. Parts **b**–**g** adapted with permission from ref. 219 under a Creative Commons licence CC BY 4.0.

with high closure skill for QG turbulent flows[219] (Fig. 3). MSLMs and signal decomposition methods are promising for tackling the closure of more realistic primitive equation models, and for providing more understanding. Furthermore, because MSLMs are made of stochastic oscillators, the question arises whether the quasiperiodic Landau view of turbulence[224], despite having been replaced by the Ruelle and Takens vision based on chaos theory[225], may be suitable to describe turbulent motions if stochasticity is included.

Describing the climate crisis via response theory

We can address the problem of quantifying the climate response to forcings by considering how perturbations affect the statistical properties of ensembles of trajectories that evolve according to a given SDE. Hence, we take Hasselmann's ansatz for a stochastic climate model and consider the following d-dimensional Itô SDE

$$\begin{aligned} d\mathbf{x} = [\mathbf{F}(\mathbf{x}) + \sum_{u=1}^{U} \varepsilon_1^u g_1^u(t)\mathbf{G}_u(\mathbf{x})] \\ dt + [\Sigma(\mathbf{x}) + \sum_{v=1}^{V} \varepsilon_2^v g_2^v(t)\Gamma_v(\mathbf{x})]\, d\mathbf{W}_t, \end{aligned} \tag{23}$$

where the unperturbed dynamics is given by equation (2). Hence, $\mathbf{x}$ here indicates a vector of slow climatic variables. We consider the case of general time-dependent perturbations acting on either the deterministic component (a parametric modulation of the system) or in the stochastic component (a perturbation to the noise law) associated with, for example, changes in the properties of the unresolved degrees of freedom. The U independent perturbations to the drift term are embodied by the vector fields $\mathbf{G}_u$, each modulated by a (scalar) amplitude function $g_1^u(t)$ and a small parameter ε^u. The V independent perturbations to the noise term are embodied by the $d \times p$ matrices Γ_v, whose amplitude are controlled by the functions $g_2^v(t)$ and the small parameters ε^v.

Note that if one considers a background deterministic dynamics ($\Sigma(\mathbf{x}) = 0$), and the time-dependent forcing to the drift term is non-vanishing, finding the solution $\rho_\varepsilon(\mathbf{x},t)$ of the FPE that corresponds to equation (23) amounts to studying the properties of the statistical equilibrium supported by the system's pullback attractor[226–229]. In practical terms, such a measure can be constructed by initializing an ensemble in the infinitely distant past and letting it evolve according to the time-dependent dynamics[28]. Following ref. 26, let us assume that $\rho_\varepsilon(\mathbf{x},t)$ can be written as:

$$\rho_\varepsilon(\mathbf{x},t) = \rho_0(\mathbf{x}) + \sum_{u=1}^{U} \varepsilon_1^u \rho_{1,\mathrm{d}}^u(\mathbf{x},t) + \sum_{v=1}^{V} \varepsilon_2^v \rho_{1,\mathrm{s}}^v(\mathbf{x},t) + \text{h.o.t.}$$

where h.o.t. denotes higher-order terms. Such an asymptotic expansion is the starting point of virtually all linear response formulas for statistical mechanical systems; see refs. 230,231 for a discussion of the radius of convergence of the expansion above.

The expected value of Ψ at time t is

$$\langle \Psi \rangle_{\rho_\varepsilon^t} = \int d\rho_\varepsilon(\mathbf{x},t)\Psi(\mathbf{x}) = \langle \Psi \rangle_0 + \delta^{(1)}[\Psi](t) + \text{h.o.t.}$$

where the linear response is

$$\delta^{(1)}[\Psi](t) = \sum_{u=1}^{U} \varepsilon_1^u (g_1^u \bullet \mathcal{G}_{\mathrm{d},\Psi}^u)(t) + \sum_{v=1}^{V} \varepsilon_2^v (g_2^v \bullet \mathcal{G}_{\mathrm{s},\Psi}^v)(t) \tag{24}$$

where '•' indicates the convolution product between the forcing amplitudes $g_{1/2}^{u/v}$ and the Green's functions $\mathcal{G}_{\mathrm{d/s},\Psi}^{u/v}$, where the combination of superscript u and subscript d refers to the impact of perturbations acting on the deterministic components of the flow, and superscript v and subscript s refer to the impact of perturbations acting on the stochastic components (ref. 26) (Box 4). In what follows we consider, without loss of generality, observables that have vanishing expectation value in the unperturbed state. This condition can be achieved by redefining Ψ as $\Psi - \langle \Psi \rangle_0$. In physical terms, this amounts to considering anomalies with respect to the reference climatology.

As discussed in ref. 232, having explicit formulas for the linear response of a climate model to perturbation would entail having an exact theory for the eddy–mean flow feedback. Obtaining explicit formulas would require a fully coherent theory of climate dynamics, which is still far from being achieved. Hence, we need to find ways to estimate the response operators. The Green's functions shown in Box 4 (equation (32)) can be interpreted as lagged correlations between the observables $\Phi = \mathcal{L}_{1,\mathrm{d/s}}^{u/v}(\log \rho_0)$ and Ψ. This indicates a generalization of the classical FDT[60,68,233]. The FDT has been applied to the output of climate models to predict the climate response to changes in the solar irradiance[234] and greenhouse gas concentration[235,236], and to study the impact of localized heating anomalies[237].

Nonetheless, the use of Gaussian or quasi-Gaussian approximations for ρ_0, which leads to using Green–Kubo formulas, leads to potentially large errors in the estimate of the response in a non-equilibrium system. Previous work[238] contains a rather detailed analysis of the reasons that classical FDT methods fail in reproducing the response operators: features associated with weak modes of natural variability (which are possibly filtered out in data preprocessing targeted for near-equilibrium systems) can have an important role in determining the response; see also discussion in ref. 239.

A possible way forward is to estimate the Green's functions for the observable(s) of interest from a set of suitably defined simulations. For the case of CO_2 forcing, it is convenient to perform an ensemble of N simulations where the CO_2 concentration is instantaneously doubled, and the runs continue until the new steady state is obtained[27,28,30]. The Green's functions are estimated by taking the time derivative of the ensemble average of the response of the model to such a forcing, and can then be used for performing projections of climate response to arbitrary protocols of CO_2 increase. See Box 4 for clarifications of how this formalism sheds light on various classical notions of sensitivity for the climate system.

Figure 4 portrays the application of response theory to an ESM where accurate projections are obtained for the globally averaged surface temperature and for the strength of the Atlantic Meridional Overturning Circulation (AMOC)[240,241] for a 1% annual increase of the CO_2 concentration from preindustrial conditions up to doubling. Note the very pronounced weakening of the AMOC and the slow recovery after the applied forcing stabilizes, which is discussed below in the subsection on tipping points.

Response theory can also be used for other problems of practical relevance in climate science, such as estimating the point of no return for climate action[29] and explaining whether, along the lines of defining causal links, one can use a climate observable as a surrogate forcing acting on another observable of interest[242,243]. Previous work[232] contains a useful discussion of additional potential uses of response theory in a climate context. Response theory can also be used to determine the forcing (of given norm) producing the largest response[244,245] and to solve inverse problems such as determining the forcing needed to achieve

a given response. This viewpoint relates to optimal control ideas[246]. A related application pertains to the critical appraisal of geoengineering strategies related to the stratospheric injection of aerosols[247].

Response, feedback and Koopman modes

Response theory makes it possible to investigate the key features of the climate feedback acting on different timescales. We provide a rewriting of the Green's functions in terms of individual components specifically associated with the eigenmodes of the unperturbed Kolmogorov operator $\mathcal{L}_0^*$; see equation (33a) in Box 4. We assume that $\mathcal{L}_0^*$ has a vanishing eigenvalue $\lambda_0 = 0$ with unitary multiplicity, which is tantamount to assuming a unique invariant measure.

Let $\{\lambda_j\}_{j=1}^M$ denote the M non-zero eigenvalues of finite algebraic multiplicity m_j, ordered by their real part in decreasing order, that is, with λ_1 having the largest real part. The λ_j are either real or come in complex conjugate pairs. Namely, if λ_j is an eigenvalue of the unperturbed Kolmogorov operator $\mathcal{L}_0^*$ with eigenfunction ψ_j^*, so is $e^{\lambda_j t}$ of $K_t = e^{t\mathcal{L}_0^*}$ relative to the same eigenfunction. With an abuse of language, we shall refer to λ_j and ψ_j^* as Koopman eigenvalues and Koopman eigenmodes (or simply modes), respectively. These have been introduced in the discussion of MZ decompositions for the case of deterministic evolution, and we adopt here the same terminology, the Kolmogorov operator being the analogue to the Koopman generator in the stochastic setting[62]. The eigenfunctions ψ_j^* encode the stochastic system's natural variability, decay of correlations and (temporal) power spectra[62]. Koopman analysis and related methods have demonstrated great promise in the past decade in capturing modes of climate variability from high-dimensional model and observational data[213].

Following refs. 62,248, it was shown[26] that, using the Koopman mode formalism, it is possible to express the Green's functions $\mathcal{G}^{u/v}_{\mathrm{d/s},\Psi}$ introduced in equation (32) as a sum of exponential functions (possibly multiplied by polynomials):

$$\mathcal{G}^{u/v}_{\mathrm{d/s},\Psi}(t) = \Theta(t)\sum_{j=1}^{M}\sum_{\ell=0}^{m_j-1}\alpha_j^{\ell,u/v,\mathrm{s/d}}(\Psi)\frac{1}{\ell!}e^{\lambda_j t}t^{\ell}, \tag{25}$$

where the coefficients $\alpha_j^{\ell,u/v,\mathrm{s/d}}(\Psi)$ are discussed in Box 5. In equation (25) we neglect the contribution to the response from the essential component of the spectrum of $\mathcal{L}_0^*$ (ref. 26). Making this important simplifying assumption amounts, by and large, to assuming that the operator $\mathcal{L}_0^*$ is quasi-compact[249]. This is often implicitly assumed when performing extended dynamic mode decomposition[250]. This derivation explains why the formula presented in equation (6) allowed for a correct interpretation of the climate feedback in refs. 11,66. We stress that the λ_js do not depend on either the observable or the forcing considered, but are instead a fundamental property of the reference unperturbed system's dynamics.

Tipping points

Because any Green's function for the observable Ψ can be interpreted as a correlation function C between Ψ and a suitably defined observable Φ (Box 4), it should not come as a surprise that, considering equation (25), one has

$$C_{\Psi_1,\Psi_2}(t) = \sum_{j=1}^{M}\sum_{\ell=0}^{m_j-1}\beta_j^{\ell}(\Psi_1,\Psi_2)\frac{1}{\ell!}e^{\lambda_j t}t^{\ell}, \tag{26}$$

Box 4

Green's functions, equilibrium climate sensitivity and transient climate response

The Green's functions $\mathcal{G}^{u/v}_{\mathrm{d/s},\Psi}$ are key tools for computing a system's linear response to perturbations. They can be written as

$$\mathcal{G}^{u/v}_{\mathrm{d/s},\Psi}(t) = \Theta(t)\int \mathrm{d}\rho_0(\mathbf{x})e^{t\mathcal{L}_0^*}\Psi(\mathbf{x})\mathcal{L}^{u/v}_{1,\mathrm{d/s}}(\log(\rho_0(\mathbf{x}))),$$

that is,

$$\mathcal{G}^{u/v}_{\mathrm{d/s},\Psi}(t) = \Theta(t)\int \mathrm{d}\mathbf{x}\rho_0(\mathbf{x})\Psi(\mathbf{x}(t))\Phi(x) = \Theta(t)C_{\Psi,\Phi}(t), \tag{32}$$

where $\Theta(t)$ is the Heaviside distribution, which ensures causality[21,28,242]. The operators $\mathcal{L}_0^*$, $\mathcal{L}^u_{1,\mathrm{d}}$ and $\mathcal{L}^v_{1,\mathrm{s}}$ are

$$\mathcal{L}_0^*\Psi = \mathbf{F}\cdot\nabla\Psi + \frac{1}{2}\Sigma\Sigma^T : \nabla^2\Psi, \tag{33a}$$

$$\mathcal{L}^u_{1,\mathrm{d}}\rho = -\nabla\cdot(\mathbf{G}_u\rho),\ 1\le u\le U, \tag{33b}$$

$$\mathcal{L}^v_{1,\mathrm{s}}\rho = \frac{1}{2}\nabla^2 : ((\Sigma_v\Gamma^T + \Gamma\Sigma_v^T)\rho),\ 1\le v\le V. \tag{33c}$$

In equation (33a), $\mathcal{L}_0^*$ is the Kolmogorov operator, the dual of the Fokker–Planck operator $\mathcal{L}_0$ associated with the unperturbed stochastic differential equation given in equation (2), and ':' denotes the Hadamard product.

The sensitivity of the system as measured by the observable Ψ with respect to the forcing encoded by $\mathcal{G}^{u/v}_{\mathrm{d/s},\Psi}(t)$ measures the long-term impact of switching on the forcing and keeping it at a constant value, which corresponds to choosing a constant (unitary time modulation). Hence, such sensitivity can be written as $\mathrm{S}^{u/v}_{\mathrm{d/s},\Psi} = \int_0^\infty \mathrm{d}t\mathcal{G}^{u/v}_{\mathrm{d/s},\Psi}(t)$.

Indeed, such a relationship allows one to write the equilibrium climate sensitivity (ECS) — that is, the long-term globally averaged surface air temperature increase due to a doubling of the CO_2 concentration[304] — in terms of the corresponding Green's function[27,28]. Response theory also allows writing an explicit formula[27] linking ECS to the transient climate response — the increase in globally averaged surface air temperature recorded at the time at which CO_2 has doubled as a result of 1% annual increase rate, that is, after approximately 70 years[305]. Intuitively, one has that ECS is larger than the transient climate response because of the thermal inertia of the climate system, namely the fact that following the forcing due to increased CO_2 concentration, the system needs some time to adjust to its final, steady-state temperature.

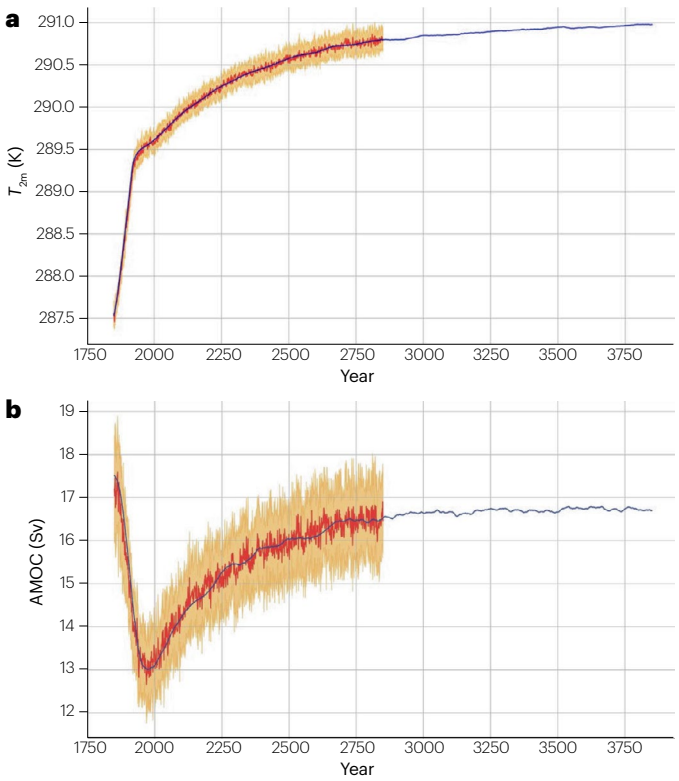


Fig. 4 | Application of response theory to an Earth system model. a,**b**, Projection of globally averaged surface temperature (part **a**) and Atlantic Meridional Overturning Circulation (AMOC) strength (part **b**) as a result of an annual 1% increase of the CO_2 concentration from preindustrial conditions up to doubling. Blue curves indicate the prediction by application of response theory; thick red curves show the ensemble mean of the Earth system model runs (yellow curves). Figure adapted with permission from ref. 30 under a Creative Commons licence CC BY 4.0.

for any pair of observables Ψ_1, Ψ_2, as proved in ref. 62. It is clear from equations (25) and (26) that to have a non-explosive behaviour, one must have $\Re\mathrm{e}(\lambda_j) < 0, 1 \le j \le M$. From equations (25) and (26) it is also clear that the rate of decay of any Green's function and that of any lagged correlation is dominated for large times by the real part of the subdominant pair of the $\mathcal{L}_0^*$ spectrum. We refer to $\gamma = \Re\mathrm{e}(\lambda_1)$ as the corresponding spectral gap. The eigenfunctions corresponding to the slowly decaying eigenvalues are the rigorous counterpart of the 'neutral climate modes'[238,251–253].

If we assume that one parameter η of our system (not to be confused with η in equation (15)) is such that the spectral gap $\gamma = \Re\mathrm{e}(\lambda_1)$ vanishes as $\eta \to \eta_c$, we have that as η nears its critical value η_c

- Any Green function and any lagged correlation decay subexponentially unless the corresponding factors (αs and βs, respectively) vanish;
- Owing to the sensitivity formula given in Box 4, one immediately derives that the sensitivity of the system increases, consistent with earlier results linking the smallness of the spectral gap γ to sensitivity of the statistics[254].

These two phenomena – critical slowing down and diverging sensitivity – are key manifestations of the proximity of tipping points[17]. If the dynamics of the system can be approximated, in a coarse-grained sense, by an Ornstein–Uhlenbeck process such as in ref. 204, another manifestation of getting to close to a tipping point is the increase in the signal's variance[255–257]. In more general terms, this feature corresponds to an increased sensitivity of the system to background noise[26,258].

The AMOC has long been seen as a climatic subsystem with potential tipping behaviour as a result of changing climate conditions, and, specifically, of changes in the hydrological cycle and cryosphere in the Atlantic basin[240,241,259]. Figure 5 shows signs from observations that support a nearing of the AMOC to a highly plausible critical transition. One finds an increase of the sensitivity and in the variance of an AMOC index tied to the sea surface temperature (SST), as well as by a decrease of the rate of decay of the index[260]. Large sensitivity and slow recovery of the same large-scale climate driver have already been discussed in a modelling context in Fig. 4.

We can learn more about such critical behaviour by taking the Fourier transform of the Green's function given in equation (25):

$$\mathfrak{F}[\mathcal{G}^{u/v}_{\mathrm{d/s},\Psi}](\omega) = \sum_{j=1}^{M} \sum_{\ell=0}^{m_j-1} \frac{\alpha_j^{\ell,u/v,\mathrm{s/d}}(\Psi)}{(\mathrm{i}\omega - \lambda_j)^{\ell+1}}. \tag{27}$$

The poles of the susceptibility are located at $\omega = -\mathrm{i}\lambda_j$. Equation (27) implies the existence of resonances in the response for real frequencies $\omega = \Im\mathrm{m}(\lambda_j)$. Neglecting the possible existence of non-unitary algebraic multiplicities, the susceptibility at the resonance j is proportional to $1/\Re\mathrm{e}(\lambda_j)$. Hence, as $\eta \to \eta_c$, the susceptibility for $\omega = \Im\mathrm{m}(\lambda_1)$ diverges, thus implying a breakdown of the response operator. If at criticality $\Im\mathrm{m}(\lambda_1) = 0$, the static response of the system diverges, indicating a saddle-node-like bifurcation phenomenon (turning point). If, instead, $\Im\mathrm{m}(\lambda_1) \neq 0$, we face an oscillatory unstable phenomenon that is reminiscent of a Hopf bifurcation[178,261]. The viewpoint proposed here allows for linking the fundamental features of the tipping phenomenology within a coherent framework.

Detection and attribution of climate change

The statistical mechanical tools discussed above allow for performing climate change projections for ensembles of trajectories: statements

Box 5

Non-equilibrium oscillator strengths

Equation (27) indicates that the coefficients $\alpha_j^{\ell,u/v,\mathrm{s/d}}(\Psi)$ weight the contributions to the frequency-dependent response coming from the eigenmode(s) corresponding to the eigenvalue λ_j (which has in general multiplicity m_j) for a given combination of observable and forcing. Note that there is a total of $M = M_1 + 2M_2$ eigenvalues. Of these, M_1 are real, and M_2 are complex conjugate pairs. Hence, the αs are the non-equilibrium, classical equivalent of the well-known oscillator strengths discussed in spectroscopy, which weight the contributions to the optical susceptibility from each of the allowed quantum transitions from the ground states to the accessible excited states[306,307]. Thus, equation (27) provides the basis for a spectroscopy of general non-equilibrium systems, and, specifically, of the climate system. Resonant terms are associated with tipping phenomena.

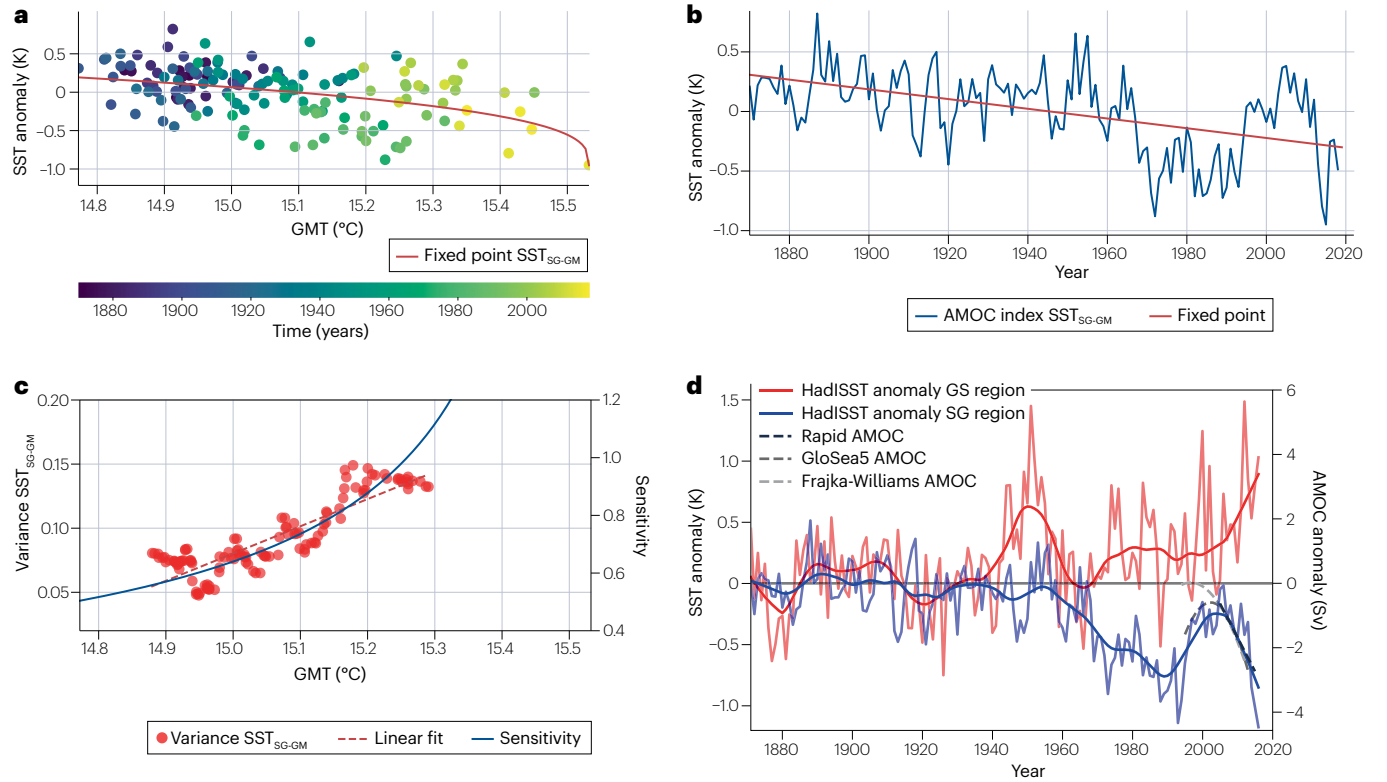


Fig. 5 | Observation evidence of the AMOC getting closer to a tipping point. **a**, Atlantic Meridional Overturning Circulation (AMOC) index SST_{SG-GM}[294] as a function of global mean temperature (GMT) and least-squares fit of the fixed point of a conceptual AMOC model. **b**, SST_{SG-GM} constructed by subtracting the global mean sea surface temperature (SST) from the average SSTs of the subpolar gyre region, supplemented by the same least-squares fit (red). **c**, Variance of fluctuations of the AMOC index around the fixed point (red) and corresponding sensitivity of the model, with control parameter T given by the global mean SSTs. These variances are estimated over a sliding temporal window, and the results are plotted at the centre point of that window. **d**, Observational evidence of the nearing of the AMOC tipping point during the past century. Shown are time series of SST anomalies with respect to the global mean SST in the subpolar gyre (SG) and the Gulf Stream (GS) regions (HadISST data). Parts **a**–**c** adapted with permission from ref. 260, Springer Nature Ltd. Part **d** adapted with permission from ref. 294, Springer Nature Ltd.

are made in terms of (changes of) the expectation value of general observables. Clearly this is a mathematical construction that does not fully comply with the requirements of climate science, because, as mentioned above, we wish to make statements on the statistical properties of the climate signal we are experiencing (we do not have access to the hypothetical multiverse comprising other statistically equivalent realizations) and possibly link it to acting forcing. Nonetheless, linear response theory applied to climate provides solid physical and dynamical foundations for detection and attribution via statistical techniques.

Let Ψ_k, $k = 1, \ldots, N$ be a collection of climatic variables. Consider also the possibility that the forcing to the climate system comes from F different sources, whether anthropogenic or natural. We rewrite equation (24) by combining all the $U + V = F$ acting forcings as $\delta^{(1)}[\Psi_k](t) = \sum_{p=1}^{F} \varepsilon_p \, (g_p \bullet \mathcal{G}_{p,\Psi_k})(t)$, where we have correspondingly simplified the notation for the Green's functions by removing the distinction between those referring to deterministic and stochastic perturbations to the dynamics. Hence we have that, at first order,

$$Y_k(t) = \Psi_k(t) - \langle \Psi_k \rangle_0 = \sum_{p=1}^{F} \bar{X}_k^p(t) + \mathcal{R}_k(t), \qquad (28)$$

where $\Psi_k(t)$ indicates the actual value of the climate variable Ψ_k at time t, the terms $X_k^p(t) = \varepsilon_p(g_p \bullet \mathcal{G}_{p,\Psi_k})(t)$ account for the forced variability, and $\mathcal{R}_k(t) = \Psi_k(t) - \langle \Psi \rangle_{\rho_\varepsilon^t}$ is a random vector whose correlations are governed by the probability distribution ρ_ε^t solving the FPE associated with equation (23). If the variables Ψ_k correspond to anomalies with respect to the reference climatology, we have $\langle \Psi_k \rangle_0 = 0$.

Equation (28) is already cast into a form that is close to the usual mathematical formulation of optimal fingerprinting for climate change given in equation (7). Response theory indicates that if we use the forced run of a model to perform detection and attribution of climate change as simulated by the same model, and if we are in the linear regime of response, all the β factors in equation (7) should be unitary, apart from uncertainty.

We stress that the linear response theory indicates that the optimal fingerprinting procedure could be applied seamlessly for different time horizons of the climate change signal and for suitably linearly filtered signals (by considering time averages, for example). This suggests that one should perform the optimal fingerprinting for different time horizons at the same time and check the consistency of the obtained results (in terms of confidence intervals for the β factors) across the time of the hindcast. Additionally, the fact that linear response theory

applies for a large class of forcings and can even be adapted for studying extremes[262] explains why optimal fingerprinting finds such a broad range of applications, including the analysis of climatic extremes[263,264].

Our derivation emphasizes that optimal fingerprinting relies strongly on assuming linearity in the response, with values of the β factors different from unity describing amplified ($\beta > 1$) and damped ($\beta < 1$) response due to nonlinear effects. A more accurate treatment of nonlinearities could be achieved by adapting the 'factor separation' method, which is used in meteorology to disentangle linear and nonlinear components of the sensitivity of a system to various forcings[265]. Such a method has been used to evaluate the factors affecting past climate conditions[266]. The consideration of higher-order terms in the response operators[267] could aid the development of fingerprinting methods that go beyond the linear approximation. Additionally, it is clear that performing optimal fingerprinting without considering the whole portfolio of acting forcings can lead to misattribution of the causes of detected climate change. The systematic cross-check proposed above for different time horizons might partly take care of these possible criticalities.

The term $\mathcal{R}_k(t)$ in equation (28) is not associated with the variability of the unperturbed climate — compare with equation (7) — but rather is tied to the system's variability encoded by the probability distribution ρ_ϵ^t; see also discussions on time-dependent probability measures and pullback attractors in refs. 70,226–228,268,269. This allows us to point out that the classical formulation of optimal fingerprinting lacks the framework to account for changes in variability due to climate change.

The expression for the Green's function (25) provides useful information for better understanding the robustness of the optimal fingerprinting method. The model error manifests itself in the difference between the spectrum of eigenvalues (and eigenfunctions) of the Koopman operator of the 'real' climate and that of the model used for constructing the fingerprints. Additionally, different climate models generally feature different Koopman modes and associated eigenvalues. Hence, constructing fingerprints by bundling together information derived from different models seems not so promising in terms of reducing model error.

It is also clear that if only one of the actual climate system or the model used for optimal fingerprinting is close to a tipping point, one expects major uncertainties and biases because of the qualitative mismatch between the leading $j = 1$ term of all the Green's functions involved, and, hence, between the model fingerprints and the actual climate response to the considered forcings. This becomes even more critical if more models are used for estimating the fingerprints, because heavily spurious information could be added.

Outlook

The emphasis on the role of noise in creating — somewhat counterintuitively — meaningful signal at lower frequencies has led to fundamental discoveries such as the mechanism of stochastic resonance, which stems from direct application of the Hasselmann paradigm in a climate context[270–273] and has since had success in many other research areas[274]. Another important area of application of the Hasselmann paradigm deals with an age-old problem of dynamical meteorology. The fast atmospheric processes due to baroclinic disturbances have been interpreted as stochastic forcing impacting the low-frequency variability of the mid-latitudes that arises from the transitions between competing regimes of circulations, mainly associated with zonal flow and blockings[275–279].

The joint analysis of climate variability and climate response presented here assumes that we are considering forced or free fluctuations around a stable, balanced state, where overall negative feedbacks dominate and control the relaxation processes. Yet the very concept of balance might need a critical appraisal when considering ultra-long timescales[280], because the internal feedback mechanisms of the system can change sign when different characteristic timescales are considered[9]. The prevalence of positive feedback indicates processes associated with transitions between possibly radically different climate regimes[6,8]. In this case too, the stochastic climate modelling angle shows great potential. Taking advantage of Fredilin–Wentzell theory of noise-induced escapes from attractors[281] and of large deviation theory[282], it is possible to develop a theory of metastability for geophysical flows[283,284] and for the climate system[285–288]. Although Gaussian noise is often used in investigations for reasons of practical and theoretical convenience, more general classes of noise laws — specifically α-stable Lévy processes — are sometimes invoked for studying climate transitions in which sudden jumps between competing states are observed[289–292]. For reasons of space and internal coherence, we have chosen not to cover these important research areas in this Perspective.

The Hasselmann programme, by construction, omits the problem of finding the root causes of the noise that affects the slow climatic variables. In this sense, it can be seen as proposing a phenomenological theory of climate. As is well known, the noise comes from the fast fluid dynamical instabilities occurring at different scales in the atmosphere and in the ocean, leading to chaotic behaviour[2,6,40]. The details of the fast processes do in fact matter, precisely because there is no time separation one can use to separate the variables of interest from those one wants to parameterize, so that the kind of parsimony one would derive from the use of homogenization theory cannot be recovered[60]. Conversely, there is no dichotomy between deterministic behaviour and stochastic representation, because chaotic dynamics generates stochastic processes. To advance towards constructing a theory of climate able to account for variability and response to forcing, and able to provide useful and usable information to address the climate crisis, there is a need to inform the stochastic angle on climate with the key details obtained by a multiscale analysis of the dynamical processes. This Perspective is a preliminary attempt in this direction.

Published online: 2 November 2023

References

1. Mitchell, J. An overview of climatic variability and its causal mechanisms. *Quat. Res.* **6**, 481–493 (1976).
2. Ghil, M. A century of nonlinearity in the geosciences. *Earth Space Sci.* **6**, 1007–1042 (2019).
3. von der Heydt, A. S. et al. Quantification and interpretation of the climate variability record. *Glob. Planet. Change* **197**, 103399 (2021).
4. Peixoto, J. P. & Oort, A. H. *Physics of Climate* (AIP, 1992).
5. Lucarini, V. et al. Mathematical and physical ideas for climate science. *Rev. Geophys.* **52**, 809–859 (2014).
6. Ghil, M. & Lucarini, V. The physics of climate variability and climate change. *Rev. Mod. Phys.* **92**, 035002 (2020).
7. Ghil, M. in *Climate Change: Multidecadal and Beyond* (eds Chang, C. P. et al.) 31–51 (World Scientific/Imperial College Press, 2015).
8. Rothman, D. H. Thresholds of catastrophe in the Earth system. *Sci. Adv.* **3**, e1700906 (2017).
9. Arnscheidt, C. W. & Rothman, D. H. Presence or absence of stabilizing Earth system feedbacks on different time scales. *Sci. Adv.* **8**, eadc9241 (2022).
10. Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann, 1988).
11. Hasselmann, K. Optimal fingerprints for the detection of time-dependent climate change. *J. Clim.* **6**, 1957–1971 (1993).
12. Hegerl, G. C. et al. Detecting greenhouse-gas-induced climate change with an optimal fingerprint method. *J. Clim.* **9**, 2281–2306 (1996).
13. Hasselmann, K. Multi-pattern fingerprint method for detection and attribution of climate change. *Clim. Dyn.* **13**, 601–611 (1997).

14. Bindoff, N. L. et al. *Detection and Attribution of Climate Change: From Global to Regional*, 867–952 (Cambridge Univ. Press, 2013).
15. IPCC. *Climate Change 2021: The Physical Science Basis* (eds Masson-Delmotte, V. et al.) (Cambridge Univ. Press, 2021).
16. IPCC. *Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (eds Field, C. B. et al.) (Cambridge Univ. Press, 2012).
17. Lenton, T. et al. Tipping elements in the Earth's climate system. *Proc. Natl Acad. Sci. USA* **105**, 1786–1793 (2008).
18. Ashwin, P., Wieczorek, S., Vitolo, R. & Cox, P. Tipping points in open systems: bifurcation, noise-induced and rate-dependent examples in the climate system. *Phil. Trans. Royal Soc. A* **370**, 1166–1184 (2012).
19. Ripple, W. J. et al. World Scientists' warning of a climate emergency 2021. *BioScience* **71**, 894–898 (2021).
20. Ruelle, D. General linear response formula in statistical mechanics, and the fluctuation-dissipation theorem far from equilibrium. *Phys. Lett. A* **245**, 220–224 (1998).
21. Ruelle, D. A review of linear response theory for general differentiable dynamical systems. *Nonlinearity* **22**, 855–870 (2009).
22. Hairer, M. & Majda, A. J. A simple framework to justify linear response theory. *Nonlinearity* **23**, 909–922 (2010).
23. Baiesi, M. & Maes, C. An update on the nonequilibrium linear response. *New J. Phys.* **15**, 013004 (2013).
24. Sarracino, A. & Vulpiani, A. On the fluctuation-dissipation relation in non-equilibrium and non-Hamiltonian systems. *Chaos* **29**, 083132 (2019).
25. Gottwald, G. A. Introduction to focus issue: linear response theory: potentials and limits. *Chaos Interdiscip. J. Nonlinear Sci.* **30**, 20401 (2020).
26. Santos Gutiérrez, M. & Lucarini, V. On some aspects of the response to stochastic and deterministic forcings. *J. Phys. A* **55**, 425002 (2022).
27. Ragone, F., Lucarini, V. & Lunkeit, F. A new framework for climate sensitivity and prediction: a modelling perspective. *Clim. Dyn.* **46**, 1459–1471 (2016).
28. Lucarini, V., Ragone, F. & Lunkeit, F. Predicting climate change using response theory: global averages and spatial patterns. *J. Stat. Phys.* **166**, 1036–1064 (2017).
29. Aengenheyster, M., Feng, Q. Y., van der Ploeg, F. & Dijkstra, H. A. The point of no return for climate action: effects of climate uncertainty and risk tolerance. *Earth Syst. Dyn.* **9**, 1085–1095 (2018).
30. Lembo, V., Lucarini, V. & Ragone, F. Beyond forcing scenarios: predicting climate change through response operators in a coupled general circulation model. *Sci. Rep.* **10**, 8668 (2020).
31. Imkeller, P. & von Storch, J. S. *Stochastic Climate Models* (Birkhauser, 2001).
32. von Storch, H. *From Decoding Turbulence to Unveiling the Fingerprint of Climate Change* (Springer, 2022).
33. Gupta, S., Mastrantonas, N., Masoller, C. & Kurths, J. Perspectives on the importance of complex systems in understanding our climate and climate change — the Nobel Prize in Physics 2021. *Chaos Interdiscip. J. Nonlinear Sci.* https://doi.org/10.1063/5.0090222 (2022).
34. Hegerl, G. C. Climate change is physics. *Commun. Earth Environ.* **3**, 14 (2022).
35. Hasselmann, K. Stochastic climate models Part I. Theory. *Tellus* **28**, 473–485 (1976).
36. Arnold, L. in *Stochastic Climate Models* (eds Imkeller, P. & von Storch, J.-S.) 141–157 (Birkhäuser, 2001).
37. Kelly, D. & Melbourne, I. Deterministic homogenization for fast–slow systems with chaotic noise. *J. Funct. Anal.* **272**, 4063–4102 (2017).
38. Cotter, C. J., Gottwald, G. A. & Holm, D. D. Stochastic partial differential fluid equations as a diffusive limit of deterministic Lagrangian multi-time dynamics. *Proc. R. Soc. A* **473**, 20170388 (2017).
39. Just, W., Kantz, H., Rödenbeck, C. & Helm, M. Stochastic modelling: replacing fast degrees of freedom by noise. *J. Phys. A* **34**, 3199–3213 (2001).
40. Ghil, M. & Childress, S. *Topics in Geophysical Fluid Dynamics: Atmospheric Dynamics, Dynamo Theory, and Climate Dynamics* (Springer, 1987).
41. Lorenz, E. *The Nature and Theory of the General Circulation of the Atmosphere* (World Meteorological Organization, 1967).
42. Beck, C. Brownian motion from deterministic dynamics. *Phys. A* **169**, 324–336 (1990).
43. Majda, A. J., Timofeyev, I. & Vanden-Eijnden, E. A mathematical framework for stochastic climate models. *Comm. Pure Appl. Math* **54**, 891–974 (2001).
44. Pavliotis, G. A. & Stuart, A. M. *Multiscale Methods* (Springer, 2008).
45. Gottwald, G. A. & Melbourne, I. Homogenization for deterministic maps and multiplicative noise. *Proc. R. Soc. A* **469**, 20130201 (2013).
46. Khasminsky, R. Principle of averaging for parabolic and elliptic differential equations and for Markov processes with small diffusion. *Theor. Probab. Appl.* **8**, 1–21 (1963).
47. Kurtz, T. A limit theorem for perturbed operator semigroups with applications to random evolutions. *J. Funct. Anal.* **12**, 55–67 (1973).
48. Papanicolaou, G. C. & Kohler, W. Asymptotic theory of mixing stochastic ordinary differential equations. *Commun. Pure Appl. Math.* **27**, 641–668 (1974).
49. Majda, A. J., Timofeyev, I. & Vanden-Eijnden, E. Systematic strategies for stochastic mode reduction in climate. *J. Atmos. Sci.* **60**, 1705–1722 (2003).
50. Palmer, T. N. & Williams, P. (eds) *Stochastic Physics and Climate Modelling* (Cambridge Univ. Press, 2009).
51. Berner, J. et al. Stochastic parameterization: toward a new view of weather and climate models. *Bull. Am. Meteorol. Soc.* **98**, 565–588 (2017).
52. Chekroun, M., Dijkstra, H., Şengül, T. & Wang, S. Transitions of zonal flows in a two-layer quasi-geostrophic ocean model. *Nonlinear Dynamics* **109**, 1887–1904 (2022).
53. Dijkstra, H. *Nonlinear Physical Oceanography: A Dynamical Systems Approach to the Large-scale Ocean Circulation and El Niño* (Springer, 2005).
54. Dijkstra, H. A. & Ghil, M. Low-frequency variability of the large-scale ocean circulation: a dynamical systems approach. *Reviews of Geophysics* **43**, RG3002 (2005).
55. Chekroun, M. D., Liu, H. & Wang, S. *Approximation of Stochastic Invariant Manifolds: Stochastic Manifolds for Nonlinear SPDEs I* (Springer Briefs in Mathematics, 2015).
56. Chekroun, M. D., Liu, H. & Wang, S. *Stochastic Parameterizing Manifolds and Non-Markovian Reduced Equations: Stochastic Manifolds for Nonlinear SPDEs II* (Springer Briefs in Mathematics, 2015).
57. Chekroun, M., Liu, H., McWilliams, J. & Wang, S. Transitions in stochastic non-equilibrium systems: efficient reduction and analysis. *J. Differ. Equ.* **346**, 145–204 (2023).
58. Wouters, J. & Gottwald, G. A. Edgeworth expansions for slow–fast systems with finite time-scale separation. *Proc. R. Soc. A* **475**, 20180358 (2019).
59. Wouters, J. & Gottwald, G. A. Stochastic model reduction for slow–fast systems with moderate time scale separation. *Multiscale Model. Simul.* **17**, 1172–1188 (2019).
60. Pavliotis, G. A. *Stochastic Processes and Applications* (Springer, 2014).
61. Maher, N., Milinski, S. & Ludwig, R. Large ensemble climate model simulations: introduction, overview, and future prospects for utilising multiple types of large ensemble. *Earth Syst. Dyn.* **12**, 401–418 (2021).
62. Chekroun, M., Tantet, A., Dijkstra, H. & Neelin, J. D. Ruelle–Pollicott resonances of stochastic systems in reduced state space. Part I: Theory. *J. Stat. Phys.* **179**, 1366–1402 (2020).
63. Eckmann, J.-P. & Ruelle, D. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.* **57**, 617–656 (1985).
64. Bojinski, S. et al. The concept of essential climate variables in support of climate research, applications, and policy. *Bull. Am. Meteorol. Soc.* **95**, 1431–1443 (2014).
65. Eyring, V. et al. Earth System Model Evaluation Tool (ESMValTool) v2.0 — an extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of Earth system models in CMIP. *Geosci. Model Dev.* **13**, 3383–3438 (2020).
66. Maier-Reimer, E. & Hasselmann, K. Transport and storage of CO_2 in the ocean — an inorganic ocean-circulation carbon cycle model. *Clim. Dyn.* **2**, 63–90 (1987).
67. Hasselmann, K., Sausen, R., Maier-Reimer, E. & Voss, R. On the cold start problem in transient simulations with coupled atmosphere–ocean models. *Clim. Dyn.* **9**, 53–61 (1993).
68. Kubo, R. The fluctuation-dissipation theorem. *Rep. Prog. Phys.* **29**, 255–284 (1966).
69. Leith, C. Climate response and fluctuation dissipation. *J. Atmos. Sci.* **32**, 2022–2026 (1975).
70. Tél, T. et al. The theory of parallel climate realizations. *J. Stat. Phys.* **179**, 1496–1530 (2020).
71. Hannart, A., Ribes, A. & Naveau, P. Optimal fingerprinting under multiple sources of uncertainty. *Geophys. Res. Lett.* **41**, 1261–1268 (2014).
72. Allen, M. & Tett, S. Checking for model consistency in optimal fingerprinting. *Clim. Dyn.* **15**, 419–434 (1999).
73. Allen, M. & Tett, S. Estimating signal amplitudes in optimal fingerprinting, part I: theory. *Clim. Dyn.* **21**, 477–491 (2003).
74. Hegerl, G. & Zwiers, F. Use of models in detection and attribution of climate change. *Wiley Interdiscip. Rev. Clim. Change* **2**, 570–591 (2011).
75. Li, Y., Chen, K., Yan, J. & Zhang, X. Uncertainty in optimal fingerprinting is underestimated. *Environ. Res. Lett.* **16**, 084043 (2021).
76. McKitrick, R. Checking for model consistency in optimal fingerprinting: a comment. *Clim. Dyn.* **58**, 405–411 (2022).
77. Chen, H., Chen, S. X. & Mu, M. A review on the optimal fingerprinting approach in climate change studies. Preprint at https://arxiv.org/abs/2205.10508 (2022).
78. Mori, H. Transport, collective motion, and Brownian motion. *Prog. Theor. Phys.* **33**, 423–455 (1965).
79. Zwanzig, R. Memory effects in irreversible thermodynamics. *Phys. Rev.* **124**, 983–992 (1961).
80. Chorin, A. J., Hald, O. H. & Kupferman, R. Optimal prediction with memory. *Phys. D* **166**, 239–257 (2002).
81. Givon, D., Kupferman, R. & Stuart, A. Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity* **17**, R55–R127 (2004).
82. Lorenz, E. N. Attractor sets and quasi-geostrophic equilibrium. *J. Atmos. Sci.* **37**, 1685–1699 (1980).
83. Chekroun, M. D. & Glatt-Holtz, N. E. Invariant measures for dissipative dynamical systems: abstract results and applications. *Commun. Math. Phys.* **316**, 723–761 (2012).
84. Budišić, M., Mohr, R. & Mezić, I. Applied Koopmanism. *Chaos* **22**, 047510 (2012).
85. Ambrosio, L., Gigli, N. & Savaré, G. *Gradient Flows in Metric Spaces and in the Space of Probability Measures* (Springer, 2008).
86. Chorin, A. & Hald, O. *Stochastic Tools in Mathematics and Science* (Springer, 2006).
87. Vissio, G. & Lucarini, V. Evaluating a stochastic parametrization for a fast–slow system using the Wasserstein distance. *Nonlinear Process. Geophys.* **25**, 413–427 (2018).
88. Stinis, P. Higher order Mori–Zwanzig models for the Euler equations. *Multiscale Model. Simul.* **6**, 741–760 (2007).
89. Li, Z., Bian, X., Li, X. & Karniadakis, G. Incorporation of memory effects in coarse-grained modeling via the Mori–Zwanzig formalism. *J. Chem. Phys.* **143**, 243128 (2015).
90. Lei, H., Baker, N. & Li, X. Data-driven parameterization of the generalized Langevin equation. *Proc. Natl Acad. Sci. USA* **113**, 14183–14188 (2016).
91. Li, Z., Lee, H., Darve, E. & Karniadakis, G. Computing the non-Markovian coarse-grained interactions derived from the Mori–Zwanzig formalism in molecular systems: application to polymer melts. *J. Chem. Phys.* **146**, 014104 (2017).

92. Brennan, C. & Venturi, D. Data-driven closures for stochastic dynamical systems. *J. Comput. Phys.* **372**, 281–298 (2018).
93. Chorin, A. J. & Lu, F. Discrete approach to stochastic parametrization and dimension reduction in nonlinear dynamics. *Proc. Natl Acad. Sci. USA* **112**, 9804–9809 (2015).
94. Lu, F., Lin, K. K. & Chorin, A. J. Data-based stochastic model reduction for the Kuramoto–Sivashinsky equation. *Phys. D* **340**, 46–57 (2017).
95. Lin, K. K. & Lu, F. Data-driven model reduction, Wiener projections, and the Koopman–Mori–Zwanzig formalism. *J. Comput. Phys.* **424**, 109864 (2021).
96. Majda, A. J. & Harlim, J. Physics constrained nonlinear regression models for time series. *Nonlinearity* **26**, 201–217 (2013).
97. Kondrashov, D., Chekroun, M. D. & Ghil, M. Data-driven non-Markovian closure models. *Phys. D* **297**, 33–55 (2015).
98. Harlim, J., Jiang, S., Liang, S. & Yang, H. Machine learning for prediction with missing dynamics. *J. Comput. Phys.* **428**, 109922 (2021).
99. Qi, D. & Harlim, J. A data-driven statistical-stochastic surrogate modeling strategy for complex nonlinear non-stationary dynamics. *J. Comput. Phys.* **485**, 112085 (2023).
100. Gilani, F., Giannakis, D. & Harlim, J. Kernel-based prediction of non-Markovian time series. *Phys. D* **418**, 132829 (2021).
101. Mori, H. A continued-fraction representation of the time-correlation functions. *Prog. Theor. Phys.* **34**, 399–416 (1965).
102. Lee, M. Solutions of the generalized Langevin equation by a method of recurrence relations. *Phys. Rev. B* **26**, 2547 (1982).
103. Florencio Jr, J. & Lee, M. H. Exact time evolution of a classical harmonic-oscillator chain. *Phys. Rev. A* **31**, 3231 (1985).
104. Kupferman, R. Fractional kinetics in Kac–Zwanzig heat bath models. *J. Stat. Phys.* **114**, 291–326 (2004).
105. Chorin, A. & Stinis, P. Problem reduction, renormalization, and memory. *Commun. Appl. Math. Comp. Sci.* **1**, 1–27 (2007).
106. Stinis, P. A comparative study of two stochastic mode reduction methods. *Phys. D* **213**, 197–213 (2006).
107. Götze, W. Recent tests of the mode-coupling theory for glassy dynamics. *J. Phys. Condens. Matter* **11**, A1 (1999).
108. Reichman, D. & Charbonneau, P. Mode-coupling theory. *J. Stat. Mech. Theory Exp.* **2005**, P05013 (2005).
109. Darve, E., Solomon, J. & Kia, A. Computing generalized Langevin equations and generalized Fokker–Planck equations. *Proc. Natl Acad. Sci. USA* **106**, 10884–10889 (2009).
110. Chen, M., Li, X. & Liu, C. Computation of the memory functions in the generalized Langevin models for collective dynamics of macromolecules. *J. Chem. Phys.* **141**, 064112 (2014).
111. Stinis, P. Renormalized Mori–Zwanzig-reduced models for systems without scale separation. *Proc. R. Soc. A* **471**, 20140446 (2015).
112. Parish, E. & Duraisamy, K. Non-Markovian closure models for large eddy simulations using the Mori–Zwanzig formalism. *Phys. Rev. Fluids* **2**, 014604 (2017).
113. Parish, E. J. & Duraisamy, K. A dynamic subgrid scale model for large eddy simulations based on the Mori–Zwanzig formalism. *J. Comput. Phys.* **349**, 154–175 (2017).
114. Zhu, Y., Dominy, J. & Venturi, D. On the estimation of the Mori–Zwanzig memory integral. *J. Math. Phys.* **59**, 103501 (2018).
115. Zhu, Y. & Venturi, D. Faber approximation of the Mori–Zwanzig equation. *J. Comput. Phys.* **372**, 694–718 (2018).
116. Venturi, D. & Karniadakis, G. Convolutionless Nakajima–Zwanzig equations for stochastic analysis in nonlinear dynamical systems. *Proc. R. Soc. A* **470**, 20130754 (2014).
117. Wouters, J. & Lucarini, V. Disentangling multi-level systems: averaging, correlations and memory. *J. Stat. Mech.* **2012**, P03003 (2012).
118. Wouters, J. & Lucarini, V. Multi-level dynamical systems: connecting the Ruelle response theory and the Mori–Zwanzig approach. *J. Stat. Phys.* **151**, 850–860 (2013).
119. Yoshimoto, Y. et al. Bottom-up construction of interaction models of non-Markovian dissipative particle dynamics. *Phys. Rev. E* **88**, 043305 (2013).
120. Hijón, C., Español, P., Vanden-Eijnden, E. & Delgado-Buscalioni, R. Mori–Zwanzig formalism as a practical computational tool. *Faraday Discuss.* **144**, 301–322 (2010).
121. Demaeyer, J. & Vannitsem, S. Comparison of stochastic parameterizations in the framework of a coupled ocean–atmosphere model. *Nonlinear Process. Geophys.* **25**, 605–631 (2018).
122. Vissio, G. & Lucarini, V. A proof of concept for scale-adaptive parametrizations: the case of the Lorenz '96 model. *Q. J. R. Meteorol. Soc.* **144**, 63–75 (2018).
123. Hald, O. H. & Stinis, P. Optimal prediction and the rate of decay for solutions of the Euler equations in two and three dimensions. *Proc. Natl Acad. Sci. USA* **104**, 6527–6532 (2007).
124. Chekroun, M. D., Di Plinio, F., Glatt-Holtz, N. E. & Pata, V. Asymptotics of the Coleman–Gurtin model. *Discrete Contin. Dyn. Syst. Ser. S* **4**, 351–369 (2011).
125. Kravtsov, S., Kondrashov, D. & Ghil, M. Multilevel regression modeling of nonlinear processes: derivation and applications to climatic variability. *J. Clim.* **18**, 4404–4424 (2005).
126. Chekroun, M. D., Liu, H. & McWilliams, J. C. Variational approach to closure of nonlinear dynamical systems: autonomous case. *J. Stat. Phys.* **179**, 1073–1160 (2020).
127. Ma, C., Wang, J. & E, W. Model reduction with memory and the machine learning of dynamical systems. *Commun. Comput. Phys.* **25**, 947–962 (2019).
128. Arnold, V. I. *Geometrical Methods in the Theory of Ordinary Differential Equations* 2nd edn (Springer, 1988).
129. Chekroun, M. D., Liu, H. & McWilliams, J. C. Optimal parameterizing manifolds for anticipating tipping points and higher-order critical transitions. Preprint at https://arxiv.org/abs/2307.06537 (2023).
130. Debussche, A. & Temam, R. Inertial manifolds and the slow manifolds in meteorology. *Differ. Integral Equ.* **4**, 897–931 (1991).
131. Temam, R. & Wirosoetisno, D. Slow manifolds and invariant sets of the primitive equations. *J. Atmos. Sci.* **68**, 675–682 (2011).
132. Zelik, S. Inertial manifolds and finite-dimensional reduction for dissipative PDEs. *Proc. R. Soc. Edinb. A* **144**, 1245–1327 (2014).
133. Kraichnan, R. H. Eddy viscosity in two and three dimensions. *J. Atmos. Sci.* **33**, 1521–1536 (1976).
134. Leith, C. E. Stochastic backscatter in a subgrid-scale model: plane shear mixing layer. *Phys. Fluids A* **2**, 297–299 (1990).
135. Debussche, A., Dubois, T. & Temam, R. The nonlinear Galerkin method: a multiscale method applied to the simulation of homogeneous turbulent flows. *Theor. Comput. Fluid Dyn.* **7**, 279–315 (1995).
136. Dubois, T., Jauberteau, F. & Temam, R. Incremental unknowns, multilevel methods and the numerical simulation of turbulence. *Comput. Methods Appl. Mech. Eng.* **159**, 123–189 (1998).
137. Dubois, T. & Jauberteau, F. A dynamic multilevel model for the simulation of the small structures in homogeneous isotropic turbulence. *J. Sci. Comput.* **13**, 323–367 (1998).
138. Fu, X., Chang, L.-B. & Xiu, D. Learning reduced systems via deep neural networks with memory. *J. Machine Learn. Model. Comput.* **1**, 97–118 (2020).
139. Wang, Q., Ripamonti, N. & Hesthaven, J. S. Recurrent neural network closure of parametric POD-Galerkin reduced-order models based on the Mori–Zwanzig formalism. *J. Comput. Phys.* **410**, 109402 (2020).
140. Gupta, A. & Lermusiaux, P. F. J. Neural closure models for dynamical systems. *Proc. R. Soc. A* **477**, 20201004 (2021).
141. Kraichnan, R. H. Eddy viscosity and diffusivity: exact formulas and approximations. *Complex Syst.* **1**, 805–820 (1987).
142. Rose, H. A. Eddy diffusivity, eddy noise and subgrid-scale modelling. *J. Fluid Mech.* **81**, 719–734 (1977).
143. Kondrashov, D., Kravtsov, S., Robertson, A. W. & Ghil, M. A hierarchy of data-based ENSO models. *J. Clim.* **18**, 4425–4444 (2005).
144. Chekroun, M. D., Kondrashov, D. & Ghil, M. Predicting stochastic systems by noise sampling, and application to the El Niño–Southern Oscillation. *Proc. Natl Acad. Sci. USA* **108**, 11766–11771 (2011).
145. Chen, C. et al. Diversity, nonlinearity, seasonality, and memory effect in ENSO simulation and prediction using empirical model reduction. *J. Clim.* **29**, 1809–1830 (2016).
146. Kondrashov, D., Kravtsov, S. & Ghil, M. Empirical mode reduction in a model of extratropical low-frequency variability. *J. Atmos. Sci.* **63**, 1859–1877 (2006).
147. Boers, N. et al. Inverse stochastic-dynamic models for high-resolution Greenland ice-core records. *Earth Syst. Dyn.* **8**, 1171–1190 (2017).
148. Kondrashov, D., Chekroun, M. D., Robertson, A. W. & Ghil, M. Low-order stochastic model and 'past-noise forecasting' of the Madden–Julian oscillation. *Geophys. Res. Lett.* **40**, 5305–5310 (2013).
149. Chen, N., Majda, A. J. & Giannakis, D. Predicting the cloud patterns of the Madden–Julian oscillation through a low-order nonlinear stochastic model. *Geophys. Res. Lett.* **41**, 5612–5619 (2014).
150. Chen, N. & Majda, A. J. Predicting the real-time multivariate Madden–Julian oscillation index through a low-order nonlinear stochastic model. *Mon. Weather Rev.* **143**, 2148–2169 (2015).
151. Santos Gutiérrez, M., Lucarini, V., Chekroun, M. D. & Ghil, M. Reduced-order models for coupled dynamical systems: data-driven methods and the Koopman operator. *Chaos* **31**, 053116 (2021).
152. Rowley, C. W., Mezić, I., Bagheri, S., Schlatter, P. & Henningson, D. S. Spectral analysis of nonlinear flows. *J. Fluid Mech.* **641**, 115–127 (2009).
153. Schmid, P. J. Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.* **656**, 5–28 (2010).
154. Kutz, J., Brunton, S., Brunton, B. & Proctor, J. *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems* (Society for Industrial and Applied Mathematics, 2016).
155. Hasselmann, K. PIPs and POPs: the reduction of complex dynamical systems using principal interaction and oscillation patterns. *J. Geophys. Res.* **93**, 11015–11021 (1988).
156. Tu, J. H., Rowley, C. W., Luchtenburg, D. M., Brunton, S. L. & Kutz, J. N. On dynamic mode decomposition: theory and applications. *J. Comput. Dyn.* **1**, 391–421 (2014).
157. Chekroun, M. D., Liu, H. & McWilliams, J. C. The emergence of fast oscillations in a reduced primitive equation model and its implications for closure theories. *Comput. Fluids* **151**, 3–22 (2017).
158. Leith, C. Nonlinear normal mode initialization and quasi-geostrophic theory. *J. Atmos. Sci.* **37**, 958–968 (1980).
159. Bolin, B. Numerical forecasting with the barotropic model. *Tellus* **7**, 27–49 (1955).
160. Baer, F. & Tribbia, J. J. On complete filtering of gravity modes through nonlinear initialization. *Mon. Weather Rev.* **105**, 1536–1539 (1977).
161. Machenhauer, B. On the dynamics of gravity oscillations in a shallow water model with applications to normal mode initialization. *Beitr. Phys. Atmos* **50**, 253–271 (1977).
162. Daley, R. Normal mode initialization. *Rev. Geophys.* **19**, 450–468 (1981).
163. Lorenz, E. N. Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141 (1963).
164. Chekroun, M., Liu, H. & McWilliams, J. C. The emergence of fast oscillations in a reduced primitive equation model and its implications for closure theories. *Comput. Fluids* **151**, 3–22 (2017).

165. Plougonven, R. & Snyder, C. Inertia–gravity waves spontaneously generated by jets and fronts. Part I: Different baroclinic life cycles. *J. Atmos. Sci.* **64**, 2502–2520 (2007).
166. Polichtchouk, I. & Scott, R. Spontaneous inertia-gravity wave emission from a nonlinear critical layer in the stratosphere. *Q. J. R. Meteorol. Soc.* **146**, 1516–1528 (2020).
167. Tulich, S., Randall, D. & Mapes, B. Vertical-mode and cloud decomposition of large-scale convectively coupled gravity waves in a two-dimensional cloud-resolving model. *J. Atmos. Sci.* **64**, 1210–1229 (2007).
168. Lane, T. P. Convectively generated gravity waves. In *Encyclopedia of Atmospheric Sciences* 2nd edition, 171–179 (Elsevier, 2015).
169. Dror, T., Chekroun, M. D., Altaratz, O. & Koren, I. Deciphering organization of GOES-16 green cumulus through the empirical orthogonal function (EOF) lens. *Atmos. Chem. Phys.* **21**, 12261–12272 (2021).
170. Chekroun, M., Liu, H. & McWilliams, J. Stochastic rectification of fast oscillations on slow manifold closures. *Proc. Natl Acad. Sci. USA* **118**, e2113650118 (2021).
171. McWilliams, J. & Gent, P. Intermediate models of planetary circulations in the atmosphere and ocean. *J. Atmos. Sci.* **37**, 1657–1678 (1980).
172. Gent, P. R. & McWilliams, J. C. Intermediate model solutions to the Lorenz equations: strange attractors and other phenomena. *J. Atmos. Sci.* **39**, 3–13 (1982).
173. Monin, A. Change of pressure in a barotropic atmosphere. *Akad. Nauk. Izv. Ser. Geofiz.* **4**, 76–85 (1952).
174. Charney, J. The use of the primitive equations of motion in numerical prediction. *Tellus* **7**, 22–26 (1955).
175. Lorenz, E. Energy and numerical weather prediction. *Tellus* **12**, 364–373 (1960).
176. Chekroun, M. D. & Kondrashov, D. Data-adaptive harmonic spectra and multilayer Stuart–Landau models. *Chaos* **27**, 093110 (2017).
177. Zhen, Y., Chapron, B., Mémin, E. & Peng, L. Eigenvalues of autocovariance matrix: a practical method to identify the Koopman eigenfrequencies. *Phys. Rev. E* **105**, 034205 (2022).
178. Tantet, A., Chekroun, M., Dijkstra, H. & Neelin, J. D. Ruelle–Pollicott resonances of stochastic systems in reduced state space. Part II: stochastic Hopf bifurcation. *J. Stat. Phys.* **179**, 1403–1448 (2020).
179. Mémin, E. Fluid flow dynamics under location uncertainty. *Geophys. Astrophys. Fluid Dyn.* **108**, 119–146 (2014).
180. Holm, D. D. Variational principles for stochastic fluid dynamics. *Proc. R. Soc. A* **471**, 20140963 (2015).
181. Cotter, C., Crisan, D., Holm, D. D., Pan, W. & Shevchenko, I. Numerically modeling stochastic lie transport in fluid dynamics. *Multiscale Model. Simul.* **17**, 192–232 (2019).
182. Resseguier, V., Mémin, E. & Chapron, B. Geophysical flows under location uncertainty, Part I: random transport and general models. *Geophys. Astrophys. Fluid Dyn.* **111**, 149–176 (2017).
183. Simonnet, E., Ghil, M., Ide, K., Temam, R. & Wang, S. Low-frequency variability in shallow-water models of the wind-driven ocean circulation. Part II: time-dependent solutions. *J. Phys. Oceanogr.* **33**, 729–752 (2003).
184. Rocha, C. B., Chereskin, T. K., Gille, S. T. & Menemenlis, D. Mesoscale to submesoscale wavenumber spectra in Drake Passage. *J. Phys. Oceanogr.* **46**, 601–620 (2016).
185. Young, W. R. Inertia-gravity waves and geostrophic turbulence. *J. Fluid Mech.* **920**, F1 (2021).
186. Bolton, T. & Zanna, L. Applications of deep learning to ocean data inference and subgrid parameterization. *J. Adv. Model. Earth Syst.* **11**, 376–399 (2019).
187. Maulik, R., San, O., Rasheed, A. & Vedula, P. Subgrid modelling for two-dimensional turbulence using neural networks. *J. Fluid Mech.* **858**, 122–144 (2019).
188. Kochkov, D. et al. Machine learning–accelerated computational fluid dynamics. *Proc. Natl Acad. Sci. USA* **118**, e2101784118 (2021).
189. Zanna, L. & Bolton, T. Data-driven equation discovery of ocean mesoscale closures. *Geophys. Res. Lett.* **47**, e2020GL088376 (2020).
190. Subel, A., Guan, Y., Chattopadhyay, A. & Hassanzadeh, P. Explaining the physics of transfer learning a data-driven subgrid-scale closure to a different turbulent flow. Preprint at https://doi.org/10.48550/arXiv.2206.03198 (2022).
191. Srinivasan, K., Chekroun, M. D. & McWilliams, J. C. Turbulence closure with small, local neural networks: Forced two-dimensional and β-plane flows. Preprint at https://doi.org/10.48550/arXiv.2304.05029 (2023).
192. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (MIT Press, 2016).
193. Piomelli, U., Cabot, W. H., Moin, P. & Lee, S. Subgrid-scale backscatter in turbulent and transitional flows. *Phys. Fluids A* **3**, 1766–1771 (1991).
194. Jansen, M. F. & Held, I. M. Parameterizing subgrid-scale eddy effects using energetically consistent backscatter. *Ocean Model.* **80**, 36–48 (2014).
195. Miyanawala, T. P. & Jaiman, R. K. An efficient deep learning technique for the Navier–Stokes equations: application to unsteady wake flow dynamics. Preprint at https://arxiv.org/abs/1710.09099 (2017).
196. Foias, C., Manley, O. & Temam, R. Modeling of the interaction of small and large eddies in two-dimensional turbulent flows. *RAIRO Modél. Math. Anal. Numér.* **22**, 93–118 (1988).
197. Foias, C., Manley, O. P. & Temam, R. Approximate inertial manifolds and effective viscosity in turbulent flows. *Phys. Fluids A* **3**, 898–911 (1991).
198. Pascal, F. & Basdevant, C. Nonlinear Galerkin method and subgrid-scale model for two-dimensional turbulent flows. *Theor. Comput. Fluid Dyn.* **3**, 267–284 (1992).
199. Lorenz, E. N. *Empirical Orthogonal Functions and Statistical Weather Prediction*. Scientific Report no. 1, Statistical Forecasting Project (1956).
200. Jolliffe, I. *Principal Component Analysis* (Wiley Online Library, 2002).
201. Penland, C. Random forcing and forecasting using principal oscillation pattern analysis. *Mon. Weath. Rev.* **117**, 2165–2185 (1989).
202. Penland, C. & Magorian, T. Prediction of Niño-3 sea surface temperatures using iinear inverse modeling **6**, 1067–1076 (1993).
203. Penland, C. & Ghil, M. Forecasting Northern Hemisphere 700-mb geopotential height anomalies using empirical normal modes. *Mon. Weather Rev.* **121**, 2355–2372 (1993).
204. Penland, C. & Sardeshmukh, P. D. The optimal growth of tropical sea surface temperature anomalies. *J. Clim.* **8**, 1999–2024 (1995).
205. Franzke, C., Majda, A. J. & Vanden-Eijnden, E. Low-order stochastic mode reduction for a realistic barotropic model climate. *J. Atmos. Sci.* **62**, 1722–1745 (2005).
206. Franzke, C. & Majda, A. J. Low-order stochastic mode reduction for a prototype atmospheric GCM. *J. Atmos. Sci.* **63**, 457–479 (2006).
207. Schölkopf, B., Smola, A. & Müller, K.-R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural comput.* **10**, 1299–1319 (1998).
208. Mukhin, D., Gavrilov, A., Feigin, A., Loskutov, E. & Kurths, J. Principal nonlinear dynamical modes of climate variability. *Sci. Rep.***5** (2015).
209. Tipping, M. E. & Bishop, C. M. Probabilistic principal component analysis. *J. R. Stat. Soc. B* **61**, 611–622 (1999).
210. Schmidt, O., Mengaldo, G., Balsamo, G. & Wedi, N. Spectral empirical orthogonal function analysis of weather and climate data. *Mon. Weather Rev.* **147**, 2979–2995 (2019).
211. Zerenner, T., Goodfellow, M. & Ashwin, P. Harmonic cross-correlation decomposition for multivariate time series. *Phys. Rev. E* **103**, 062213 (2021).
212. Das, S. & Giannakis, D. Delay-coordinate maps and the spectra of Koopman operators. *J. Stat. Phys.* **175**, 1107–1145 (2019).
213. Froyland, G., Giannakis, D., Lintner, B. R., Pike, M. & Slawinska, J. Spectral analysis of climate dynamics with operator-theoretic approaches. *Nat. Commun.* **12**, 6570 (2021).
214. Belkin, M. & Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **15**, 1373–1396 (2003).
215. Coifman, R. & Lafon, S. Diffusion maps. *Appl. Comput. Harmon. Anal.* **21**, 5–30 (2006).
216. Giannakis, D. & Majda, A. J. Nonlinear Laplacian spectral analysis: capturing intermittent and low-frequency spatiotemporal patterns in high-dimensional data. *Stat. Anal. Data Min. ASA Data Sci. J.* **6**, 180–194 (2013).
217. Kingma, D. & Welling, M. Auto-encoding variational Bayes. Preprint at https://arxiv.org/abs/1312.6114 (2013).
218. Berloff, P. Dynamically consistent parameterization of mesoscale eddies. Part I: Simple model. *Ocean Model.* **87**, 1–19 (2015).
219. Kondrashov, D., Chekroun, M. & Berloff, P. Multiscale Stuart–Landau emulators: application to wind-driven ocean gyres. *Fluids* **3**, 21 (2018).
220. Rahaman, N. et al. On the spectral bias of neural networks. In *International Conference on Machine Learning*, 5301–5310 (PMLR, 2019).
221. Ghil, M. et al. Advanced spectral methods for climatic time series. *Rev. Geophys.* **40**, 1003 (2002).
222. Kondrashov, D., Chekroun, M. D., Yuan, X. & Ghil, M. in *Advances in Nonlinear Geosciences* (ed. Tsonis, A.) 179–205 (Springer, 2018).
223. Kondrashov, D., Chekroun, M. D. & Ghil, M. Data-adaptive harmonic decomposition and prediction of Arctic sea ice extent. *Dyn. Stat. Clim. Syst.* https://doi.org/10.1093/climsys/dzy001 (2018).
224. Landau, L. & Lifshitz, E. M. *Fluid Mechanics: Landau and Lifshitz: Course of Theoretical Physics*, Vol. 6 (Elsevier, 2013).
225. Ruelle, D. & Takens, F. On the nature of turbulence. *Commun. Math. Phys* **20**, 167–192 (1971).
226. Chekroun, M. D., Simonnet, E. & Ghil, M. Stochastic climate dynamics: random attractors and time-dependent invariant measures. *Phys. D* **240**, 1685–1700 (2011).
227. Carvalho, A. N., Langa, J. A. & Robinson, J. C. *Attractors for Infinite-Dimensional Non-autonomous Dynamical Systems* (Springer, 2013).
228. Tél, T. et al. The theory of parallel climate realizations. *J. Stat. Phys.* https://doi.org/10.1007/s10955-019-02445-7 (2019).
229. Pierini, S. Statistical significance of small ensembles of simulations and detection of the internal climate variability: an excitable ocean system case study. *J. Stat. Phys.* **179**, 1475–1495 (2020).
230. Lucarini, V. 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**, 312–333 (2016).
231. Santos Gutiérrez, M. & Lucarini, V. Response and sensitivity using Markov chains. *J. Stat. Phys.* **179**, 1572–1593 (2020).
232. Hassanzadeh, P. & Kuang, Z. The linear response function of an idealized atmosphere. Part I: Construction using Green's functions and applications. *J. Atmos. Sci.* **73**, 3423–3439 (2016).
233. Abramov, R. V. & Majda, A. J. Blended response algorithms for linear fluctuation-dissipation for complex nonlinear dynamical systems. *Nonlinearity* **20**, 2793 (2007).
234. North, G. R., Bell, R. E. & Hardin, J. W. Fluctuation dissipation in a general circulation model. *Clim. Dyn.* **8**, 259–264 (1993).
235. Cionni, I., Visconti, G. & Sassi, F. Fluctuation dissipation theorem in a general circulation model. *Geophys. Res. Lett.***31** (2004).

236. Langen, P. L. & Alexeev, V. A. Estimating 2×CO_2 warming in an aquaplanet GCM using the fluctuation-dissipation theorem. *Geophys. Res. Lett.* https://doi.org/10.1029/2005GL024136 (2005).
237. Gritsun, A. & Branstator, G. Climate response using a three-dimensional operator based on the fluctuation–dissipation theorem. *J. Atmos. Sci.* **64**, 2558–2575 (2007).
238. Hassanzadeh, P. & Kuang, Z. The linear response function of an idealized atmosphere. Part II: Implications for the practical use of the fluctuation–dissipation theorem and the role of operator's nonnormality. *J. Atmos. Sci.* **73**, 3441–3452 (2016).
239. Gritsun, A. & Lucarini, V. Fluctuations, response, and resonances in a simple atmospheric model. *Phys. D* **349**, 62–76 (2017).
240. Dijkstra, H. A. & Ghil, M. Low-frequency variability of the large-scale ocean circulation: a dynamical systems approach. *Rev. Geophys.* https://doi.org/10.1029/2002RG000122 (2005).
241. Kuhlbrodt, T. et al. On the driving processes of the Atlantic Meridional Overturning Circulation. *Rev. Geophys.* https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004RG000166 (2007).
242. Lucarini, V. Revising and extending the linear response theory for statistical mechanical systems: evaluating observables as predictors and predictands. *J. Stat. Phys.* **173**, 1698–1721 (2018).
243. Tomasini, U. M. & Lucarini, V. Predictors and predictands of linear response in spatially extended systems. *Eur. Phys. J.: Spec. Top.* **230**, 2813–2832 (2021).
244. Antown, F., Dragičević, D. & Froyland, G. Optimal linear responses for Markov chains and stochastically perturbed dynamical systems. *J. Stat. Phys.* **170**, 1051–1087 (2018).
245. Antown, F., Froyland, G. & Galatolo, S. Optimal linear response for Markov Hilbert–Schmidt integral operators and stochastic dynamical systems. *J. Nonlinear Sci.* **32**, 79 (2022).
246. Chekroun, M. D., Kröner, A. & Liu, H. Galerkin approximations of nonlinear optimal control problems in Hilbert spaces. *Electron. J. Differ. Equ.* **189**, 1–40 (2017).
247. Bódai, T., Lucarini, V. & Lunkeit, F. Can we use linear response theory to assess geoengineering strategies? *Chaos: Interdiscip. J. Nonlinear Sci.* **30**, 023124 (2020).
248. Tantet, A., Lucarini, V. & Dijkstra, H. A. Resonances in a chaotic attractor crisis of the Lorenz flow. *J. Stat. Phys.* **170**, 584–616 (2018).
249. Engel, K.-J. & Nagel, R. *One-Parameter Semigroups for Linear Evolution Equations* (Springer, 2000).
250. Williams, M., Kevrekidis, I. & Rowley, C. A data–driven approximation of the Koopman operator: extending dynamic mode decomposition. *J. Nonlinear Sci.* **25**, 1307–1346 (2015).
251. Navarra, A. A new set of orthonormal modes for linearized meteorological problems. *J. Atmos. Sci.* **50**, 2569–2583 (1993).
252. Palmer, T. N. A nonlinear dynamical perspective on climate prediction. *J. Clim.* **12**, 575–591 (1999).
253. Lu, J., Liu, F., Leung, L. R. & Lei, H. Neutral modes of surface temperature and the optimal ocean thermal forcing for global cooling. *NPJ Clim. Atmos. Sci.* **3**, 9 (2020).
254. Chekroun, M. D., Neelin, J. D., Kondrashov, D., McWilliams, J. C. & Ghil, M. Rough parameter dependence in climate models: the role of Ruelle-Pollicott resonances. *Proc. Natl Acad. Sci. USA* **111**, 1684–1690 (2014).
255. Held, H. & Kleinen, T. Detection of climate system bifurcations by degenerate fingerprinting. *Geophys. Res. Lett.* https://doi.org/10.1029/2004GL020972 (2004).
256. Scheffer, M. et al. Anticipating critical transitions. *Science* **338**, 344–348 (2012).
257. Boettner, C. & Boers, N. Critical slowing down in dynamical systems driven by nonstationary correlated noise. *Phys. Rev. Res.* **4**, 013230 (2022).
258. Lucarini, V. Stochastic perturbations to dynamical systems: a response theory approach. *J. Stat. Phys.* **146**, 774–786 (2012).
259. Rahmstorf, S. Bifurcations of the Atlantic thermohaline circulation in response to changes in the hydrological cycle. *Nature* **378**, 145–149 (1995).
260. Boers, N. Observation-based early-warning signals for a collapse of the Atlantic Meridional Overturning Circulation. *Nat. Clim. Change* **11**, 680–688 (2021).
261. Tantet, A., Chekroun, M., Neelin, J. & Dijkstra, H. Ruelle–Pollicott resonances of stochastic systems in reduced state space. Part III: Application to the Cane–Zebiak model of the El Niño–Southern Oscillation. *J. Stat. Phys.* **179**, 1449–1474 (2020).
262. Lucarini, V., Kuna, T., Faranda, D. & Wouters, J. Towards a general theory of extremes for observables of chaotic dynamical systems. *J. Stat. Phys.* **154**, 723–750 (2014).
263. Naveau, P., Hannart, A. & Ribes, A. Statistical methods for extreme event attribution in climate science. *Annu. Rev. Stat. Appl.* **7**, 89–110 (2020).
264. Wang, Z., Jiang, Y., Wan, H., Yan, J. & Zhang, X. Toward optimal fingerprinting in detection and attribution of changes in climate extremes. *J. Am. Stat. Assoc.* **116**, 1–13 (2021).
265. Stein, U. & Alpert, P. Factor separation in numerical simulations. *J. Atmos. Sci.* **50**, 2107–2115 (1993).
266. Hossain, A. et al. The impact of different atmospheric CO_2 concentrations on large scale Miocene temperature signatures. *Paleoceanogr. Paleoclimatol.* **38**, e2022PA004438 (2023).
267. Ruelle, D. Nonequilibrium statistical mechanics near equilibrium: computing higher-order terms. *Nonlinearity* **11**, 5–18 (1998).
268. Chekroun, M. D., Ghil, M. & Neelin, J. D. in *Advances in Nonlinear Geosciences* (ed. Tsonis, A.), 1–33 (Springer, 2018).
269. Chekroun, M. D., Koren, I., Liu, H. & Liu, H. Generic generation of noise-driven chaos in stochastic time delay systems: bridging the gap with high-end simulations. *Sci. Adv.* **8**, eabq7137 (2022).
270. Benzi, R., Sutera, A. & Vulpiani, A. The mechanism of stochastic resonance. *J. Phys. A* **14**, L453–L457 (1981).
271. Nicolis, C. Solar variability and stochastic effects on climate. *Sol. Phys.* **74**, 473–478 (1981).
272. Benzi, R., Parisi, G., Sutera, A. & Vulpiani, A. Stochastic resonance in climatic change. *Tellus* **34**, 10–16 (1982).
273. Nicolis, C. Stochastic aspects of climatic transitions — response to a periodic forcing. *Tellus* **34**, 308–308 (1982).
274. Gammaitoni, L., Hänggi, P., Jung, P. & Marchesoni, F. Stochastic resonance. *Rev. Mod. Phys.* **70**, 223–287 (1998).
275. Charney, J. G. & DeVore, J. G. Multiple flow equilibria in the atmosphere and blocking. *J. Atmos. Sci.* **36**, 1205–1216 (1979).
276. Benzi, R., Malguzzi, P., Speranza, A. & Sutera, A. The statistical properties of general atmospheric circulation: observational evidence and a minimal theory of bimodality. *Q. J. R. Meteorol. Soc.* **112**, 661–674 (1986).
277. Benzi, R. & Speranza, A. Statistical properties of low-frequency variability in the Northern Hemisphere. *J. Clim.* **2**, 367–379 (1989).
278. Kimoto, M. & Ghil, M. Multiple flow regimes in the Northern Hemisphere winter. Part I: Methodology and hemispheric regimes. *J. Atmos. Sci.* **50**, 2625–2643 (1993a).
279. Itoh, H. & Kimoto, M. Multiple attractors and chaotic itinerancy in a quasigeostrophic model with realistic topography: implications for weather regimes and low-frequency variability. *J. Atmos. Sci.* **53**, 2217–2231 (1996).
280. Arnscheidt, C. W. & Rothman, D. H. The balance of nature: a global marine perspective. *Ann. Rev. Mar. Sci.* **14**, 49–73 (2022).
281. Freidlin, M. I. & Wentzell, A. D. *Random Perturbations of Dynamical Systems* (Springer, 1998).
282. Touchette, H. The large deviation approach to statistical mechanics. *Phys. Rep.* **478**, 1–69 (2009).
283. Bouchet, F. & Venaille, A. Statistical mechanics of two-dimensional and geophysical flows. *Phys. Rep.* **515**, 227–295 (2012).
284. Herbert, C. *An Introduction to Large Deviations and Equilibrium Statistical Mechanics for Turbulent Flows*, 53–84 (Springer International Publishing, 2015).
285. Lucarini, V. & Bódai, T. Transitions across melancholia states in a climate model: reconciling the deterministic and stochastic points of view. *Phys. Rev. Lett.* **122**, 158701 (2019).
286. Lucarini, V. & Bódai, T. Global stability properties of the climate: melancholia states, invariant measures, and phase transitions. *Nonlinearity* **33**, R59–R92 (2020).
287. Margazoglou, G., Grafke, T., Laio, A. & Lucarini, V. Dynamical landscape and multistability of a climate model. *Proc. R. Soc. A* **477**, 20210019 (2021).
288. Rousseau, D.-D., Bagniewski, W. & Lucarini, V. A punctuated equilibrium analysis of the climate evolution of Cenozoic exhibits a hierarchy of abrupt transitions. *Sci. Rep.* **13**, 11290 (2023).
289. Ditlevsen, P. D. Observation of α-stable noise induced millennial climate changes from an ice-core record. *Geophys. Res. Lett.* **26**, 1441–1444 (1999).
290. Penland, C. & Ewald, B. D. On modelling physical systems with stochastic models: diffusion versus Lévy processes. *Phil. Trans. Roy. Soc. A* **366**, 2455–2474 (2008).
291. Gottwald, G. A. A model for Dansgaard–Oeschger events and millennial-scale abrupt climate change without external forcing. *Clim. Dyn.* **56**, 227–243 (2021).
292. Lucarini, V., Serdukova, L. & Margazoglou, G. Lévy noise versus Gaussian-noise-induced transitions in the Ghil–Sellers energy balance model. *Nonlinear Process. Geophys.* **29**, 183–205 (2022).
293. Berloff, P. Dynamically consistent parameterization of mesoscale eddies — Part II: Eddy fluxes and diffusivity from transient impulses. *Fluids* https://doi.org/10.3390/fluids1030022 (2016).
294. Caesar, L., Rahmstorf, S., Robinson, A., Feulner, G. & Saba, V. Observed fingerprint of a weakening Atlantic Ocean overturning circulation. *Nature* **556**, 191–196 (2018).
295. Saltzman, B. *Dynamical Paleoclimatology: Generalized Theory of Global Climate Change* (Academic, 2001).
296. Miyadera, I. On perturbation theory for semi-groups of operators. *Tohoku Math. J. Second Ser.* **18**, 299–310 (1966).
297. Voigt, J. On the perturbation theory for strongly continuous semigroups. *Math. Ann.* **229**, 163–171 (1977).
298. Givon, D., Kupferman, R. & Hald, O. Existence proof for orthogonal dynamics and the Mori–Zwanzig formalism. *Isr. J. Math.* **145**, 221–241 (2005).
299. McWilliams, J. C. A perspective on the legacy of Edward Lorenz. *Earth Space Sci.* **6**, 336–350 (2019).
300. Lorenz, E. On the existence of a slow manifold. *J. Atmos. Sci.* **43**, 1547–1558 (1986).
301. Lorenz, E. N. & Krishnamurthy, V. On the nonexistence of a slow manifold. *J. Atmos. Sci.* **44**, 2940–2950 (1987).
302. Vanneste, J. Exponential smallness of inertia–gravity wave generation at small Rossby number. *J. Atmos. Sci.* **65**, 1622–1637 (2008).
303. Vanneste, J. Balance and spontaneous wave generation in geophysical flows. *Ann. Rev. Fluid Mech.* **45**, 147–172 (2013).
304. IPCC. *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. et al.) (Cambridge Univ. Press, 2014).
305. Otto, A. et al. Energy budget constraints on climate response. *Nat. Geosci.* **6**, 415–416 (2013).
306. Hilborn, R. C. Einstein coefficients, cross sections, *f* values, dipole moments, and all that. *Am. J. Phys.* **50**, 982–986 (1982).
307. Lucarini, V., Saarinen, J. J., Peiponen, K.-E. & Vartiainen, E. M. *Kramers–Kronig Relations in Optical Materials Research* (Springer, 2005).

Acknowledgements
V.L. acknowledges support received from the European Union (EU) Horizon 2020 research and innovation programme through the projects TiPES (grant agreement no. 820970) and CriticalEarth (grant agreement no. 956170) and by the EPSRC through grant EP/T018178/1. M.D.C. acknowledges the European Research Council under the EU Horizon 2020 research and innovation programme (grant no. 810370) and the Ben May Center grant for theoretical and/or computational research. This work has been also partially supported by the Office of Naval Research (ONR) Multidisciplinary University Research Initiative (MURI) grant N00014-20-1-2023. Finally, the authors thank many close collaborators over the years without whom this review would have not been possible: O. Altaratz, P. Ashwin, P. Berloff, R. Blender, T. Bódai, N. Boers, H. Dijkstra, T. Dror, B. Dubrulle, D. Faranda, K. Fraedrich, V. M. Gálfi, G. Gallavotti, N. Glatt-Holtz, G. Gottwald, A. Gritsun, A. von der Heydt, D. Kondrashov, I. Koren, S. Kravtsov, T. Kuna, J. Kurths, H. Liu, F. Lunkeit, D. Neelin, G. Pavliotis, C. Penland, F. Ragone, L. Roques, J. Roux, M. Santos Gutiérrez, S. Schubert, E. Simonnet, A. Speranza, K. Srinivisan, A. Tantet, T. Tél, S. Vannitsem, S. Wang, J. Wouters, N. Zagli, and I. Zaliapin, with special gratitude to A. Chorin, M. Ghil, J. C. McWilliams, D. Ruelle and R. Temam for their guidance and inspirational works.

Author contributions
Both authors contributed to all aspects of this work.

Competing interests
The authors declare no competing interests.

Additional information
Peer review information *Nature Reviews Physics* thanks the anonymous referees for their contribution to the peer review of this work.