On recent trends in climate dynamics

Georg A. Gottwald*

1. Introduction

The dynamics of the atmosphere and the oceans are inherently complex. There are active entangled processes running on spatial scales from millimetres to thousands of kilometres, and temporal scales from seconds to millennia. Capturing the whole range of spatial and temporal scales is impossible given current computer power. A numerical forecaster must decide, depending on their specific objectives, what scales they want to resolve. A corollary of this decision is that each numerical scheme inevitably fails to resolve so-called *unresolved scales* or *subgrid scales*. Usually, the interesting information is carried by the slow and large scales. For example, for weather forecasts we want to resolve large-scale high- and low-pressure fields rather than small-scale fast oscillations of the stratification surfaces, or for climate prediction in a coupled ocean-atmosphere model we want to learn about the slow dynamics of the ocean that are constantly kicked by fast-evolving weather systems swirling above.

In recent years, interest in stochastic dynamics has increased across disciplines. The reason for this in the field of climate dynamics is that stochasticity may be used to parametrise subgrid-scale phenomena. In climate modelling, the idea of modelling fast chaotic dynamics by stochastic processes and thereby reducing the effective dimension of the full system goes back to the seminal work by Hasselmann [11] and Leith [15]. In their work, Hasselmann [11] and Leith [15] have suggested studying climatic regime switches by introducing, in an ad hoc way, a stochastic driver for the slow dynamics. Such an approximation describes the deviations from an averaged climatological system. Of course, it is natural to expect such behaviour only if the fast variables (e.g. weather in a coupled climatic ocean-atmosphere model) are sufficiently chaotic and approximately random.

This approach of modelling fast small-scale chaotic processes by a stochastic process is intuitive: provided the fast processes decorrelate rapidly enough, the slow variables experience, during one slow time unit, the sum of uncorrelated events of the fast dynamics, which according to the (weak) central limit theorem corresponds to approximate Gaussian noise. A method whereby many fast degrees of freedom are replaced by a stochastic process is called *stochastic model reduction*.

Invited technical paper, communicated by Kerry Landman.

*School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia.
Email: georg.gottwald@sydney.edu.au

These ideas have been used to simulate, for example, coupled ocean-atmosphere models [18] and urban air pollution [2], and have also been employed in different fields, such as macromolecular systems [9]. Scientists have recently realised that these methods can be applied to many complex systems [6], [10], [13], [16], [20], [21], [23]. The effective dimension reduction achieved if a large number of fast equations are replaced by only a few stochastic processes, and the associated computational advantage of such a reduction, is a major driving force behind this research.

The 'Hasselmann project' [1] of stochastic model reduction, which has received renewed attention in the past few years, has not yet been finished and poses a fascinating challenge for mathematicians. In particular, how can the transition from a purely deterministic system to a stochastic system be made in a controllable way? In the following we introduce a formalism which allows us to rewrite a deterministic system in such a way that it 'looks like' a stochastic system in the form of generalised Langevin equations, and may be a formal starting point for controlled stochastic model reduction.

2. The Mori-Zwanzig projection operator formalism

Given a dynamical system

$$\dot{\boldsymbol{z}} = \boldsymbol{f}(\boldsymbol{z}), \tag{1}$$

with initial condition $\boldsymbol{z}(0) = \boldsymbol{z}_0$ and $\boldsymbol{z} \in \mathbb{R}^d$, suppose we are not interested in the full solution $\boldsymbol{z}(t)$, but rather only in a few $n \leq d$ observables $\Phi(\boldsymbol{z}) = (\Phi_1(\boldsymbol{z}), \Phi_2(\boldsymbol{z}), \ldots, \Phi_n(\boldsymbol{z}))$. This includes the case $\Phi(\boldsymbol{z}) = (z_1, \ldots, z_n)$, when the state space is decomposed as $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y})$ into 'interesting' variables, $\boldsymbol{x} = (z_1, \ldots, z_n) \in \mathbb{R}^n$, and 'uninteresting' variables, $\boldsymbol{y} = (z_{n+1}, \ldots, z_d) \in \mathbb{R}^{d-n}$. Now let us ask the following question: what are the effective dynamics of the interesting observables for an ensemble of initial conditions $\boldsymbol{z}(0)$, where $\Phi(\boldsymbol{z}(0))$ is known and the uninteresting subspace is equipped with a known distribution?

The main idea is simple, and essentially boils down to the method of variation of constants (see, for example, the beautiful book by Zwanzig [26]). Consider the very simple coupled linear system

$$\begin{aligned} \dot{x} &= L_{11}x + L_{12}y \\ \dot{y} &= L_{21}x + L_{22}y. \end{aligned}$$

Suppose we are only interested in the dynamics of x, and have only some climatic knowledge of the initial conditions of the variables y, that is the mean and variance. We can then solve for y to obtain

$$y(t) = \mathrm{e}^{L_{22}t}y(0) + \int_0^t \mathrm{e}^{L_{22}(t-s)}L_{21}x(s)\,\mathrm{d}s,$$

which we may use to express the dynamics of the interesting variable as

$$\dot{x} = L_{11}x + L_{12}\int_0^t \mathrm{e}^{L_{22}(t-s)}L_{21}x(s)\,\mathrm{d}s + L_{12}\mathrm{e}^{L_{22}t}y(0).$$

This is of the form of a generalised Langevin equation, where the first term is Markovian, the second is a memory term, and the last can be interpreted as a noise term, provided that the initial conditions $y(0)$ are randomly distributed. (Paul Langevin [14] studied Brownian motion from a different perspective to Albert's Einstein's seminal 1905 paper [7] in an equally seminal paper, describing the motion of a single Brownian particle as a dynamic process via a stochastic differential equation; to be precise in modern terminology, as an Ornstein–Uhlenbeck process [24].)

Let us now set the scene for the general nonlinear case. Assume that the vector field $\mathbf{f}$ is Lipschitz to assure the existence and uniqueness of a solution of (1). Denote the flow map associated with (1) by $\varphi_t\colon \boldsymbol{z}_0 \to \boldsymbol{z}(t;\boldsymbol{z}_0)$. Rather than investigating the dynamical system (1) directly, one may choose to look at how observables $V(z(t)) \in C^1(\mathbb{R}^d,\mathbb{R})$ evolve in time. Applying the chain rule, one can naturally define the *generator*

$$\mathcal{L} = \mathbf{f}(\boldsymbol{z})\cdot\nabla\,,$$

and write

$$\frac{\mathrm{d}}{\mathrm{d}t}V(\boldsymbol{z}(t)) = \mathcal{L}V(\boldsymbol{z}(t)).$$

Note that $\mathcal{L}$ is the formal L^2-adjoint operator of the Liouville operator $\mathcal{L}^\star$ with $\mathcal{L}^\star\rho = -\nabla\cdot(f(\boldsymbol{z})\rho)$ controlling the evolution of densities of ensembles propagated according to (1).

There is an intimate link between solutions $\varphi_t(\boldsymbol{z}_0)$ of (1), parametrised by the initial conditions (which may be randomly distributed), and the solution $v(\boldsymbol{z},t)$ of the following Cauchy problem

$$\frac{\partial v}{\partial t} = \mathcal{L}v \quad \text{with} \quad v(\boldsymbol{z},0) = \phi(\boldsymbol{z}), \tag{2}$$

where $\boldsymbol{z}$ is an independent variable and denotes initial conditions. Provided that the initial datum $\phi(\boldsymbol{z})$ is sufficiently smooth so that the Cauchy problem (2) has a classical solution (that is, there exists a solution $v(\boldsymbol{z},t)$ such that the Cauchy problem is satisfied pointwise for all $(\boldsymbol{z},t)\in\mathbb{R}^d\times(0,\infty)$), one has

$$v(\boldsymbol{z},t) = \phi(\varphi_t(\boldsymbol{z})) \tag{3}$$

for all $t\in\mathbb{R}^+$ and $\boldsymbol{z}\in\mathbb{R}^d$. Hence, instead of analysing the possibly nonlinear ODE (1), one can employ theory developed for linear hyperbolic PDEs. The solution of (2) can be formally written as

$$v(\boldsymbol{z},t) = \mathrm{e}^{\mathcal{L}t}\phi(\boldsymbol{z}). \tag{4}$$

For proofs and more details the reader is referred to the monograph [19].

Let us now conduct the full program of deriving a generalised Langevin equation. To filter out the dynamics of the interesting variables we require a projection operator $\boldsymbol{P}$ that maps functions of $\boldsymbol{z}$ to functions of $\Phi(\boldsymbol{z})$. For simplicity of exposition let us restrict our attention to the case where $\boldsymbol{z}=(\boldsymbol{x},\boldsymbol{y})$ and $\Phi(\boldsymbol{z})=\boldsymbol{x}$. A suitable projector for the situation when the initial conditions of the interesting variables $\boldsymbol{x}$

are known exactly but only statistical information is available for the uninteresting variables $\boldsymbol{y}$, is the conditional expectation of a function $\omega(\boldsymbol{x},\boldsymbol{y})$, given by

$$(\boldsymbol{P}\omega)(\boldsymbol{x}) = \mathrm{E}[\omega(\boldsymbol{x},\boldsymbol{y}) \mid \boldsymbol{x}] = \frac{1}{\Omega(\boldsymbol{x})}\int_{\mathbb{R}^{d-n}} \rho(\boldsymbol{\xi},\boldsymbol{\eta})\delta(\boldsymbol{\xi}-\boldsymbol{x})\omega(\boldsymbol{\xi},\boldsymbol{\eta})\,\mathrm{d}\boldsymbol{\xi}\,\mathrm{d}\boldsymbol{\eta},$$

where $\rho(\boldsymbol{x},\boldsymbol{y})$ denotes the joint probability function of the initial conditions for the full system (1) and δ denotes the Dirac function. The normalisation

$$\Omega(\boldsymbol{x}) = \int_{\mathbb{R}^{d-n}} \rho(\boldsymbol{\xi},\boldsymbol{\eta})\delta(\boldsymbol{\xi}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{\xi}\,\mathrm{d}\boldsymbol{\eta}$$

is the probability density of $\boldsymbol{x}$. It measures in the language of statistical mechanics the number of *microstates* that give rise to the *macrostate* $\boldsymbol{x}$. A simple calculation shows that $(\boldsymbol{P}\omega)(\boldsymbol{x})$ is the best approximation of the function $\omega(\boldsymbol{x},\boldsymbol{y})$ by a function of $\boldsymbol{x}$ in the L^2-sense. (In the context of PDEs one may use Galerkin approximations and (1) would be defined on a suitable Hilbert space. In this case, a perfectly valid projector would be to simply truncate the Galerkin series at some specified high wave number cut-off [22], [23].) We also define the orthogonal projector $\boldsymbol{Q}$ that projects onto $\boldsymbol{y}$, with $\boldsymbol{Q} = \boldsymbol{1} - \boldsymbol{P}$. Now, the derivation of the Mori-Zwanzig equation is a two-liner: given the Cauchy problem (2) and its formal solution (4) we write, using $\boldsymbol{P} + \boldsymbol{Q} = \boldsymbol{1}$,

$$\frac{\partial v}{\partial t}(\boldsymbol{z},t) = \mathcal{L}\mathrm{e}^{\mathcal{L}t}\Phi(\boldsymbol{z}) = \mathrm{e}^{\mathcal{L}t}\boldsymbol{P}\mathcal{L}\Phi(\boldsymbol{z}) + \mathrm{e}^{\mathcal{L}t}\boldsymbol{Q}\mathcal{L}\Phi(\boldsymbol{z}),$$

which, upon using the Duhamel–Dyson formula [8] for operators $\boldsymbol{A}$ and $\boldsymbol{B}$, yields

$$\mathrm{e}^{t(\boldsymbol{A}+\boldsymbol{B})} = \mathrm{e}^{t\boldsymbol{A}} + \int_0^t \mathrm{e}^{(t-s)(\boldsymbol{A}+\boldsymbol{B})}\,\boldsymbol{B}\,\mathrm{e}^{s\boldsymbol{A}}\,\mathrm{d}s.$$

The reader may verify by differentiation that this becomes the celebrated Mori-Zwanzig equation [17], [25]

$$\frac{\partial v}{\partial t}(\boldsymbol{z},t) = \mathrm{e}^{\mathcal{L}t}\boldsymbol{P}\mathcal{L}\Phi(\boldsymbol{z}) + \int_0^t \mathrm{e}^{(t-s)\mathcal{L}}\,\boldsymbol{P}\mathcal{L}\,\mathrm{e}^{s\boldsymbol{Q}\mathcal{L}}\boldsymbol{Q}\mathcal{L}\Phi(\boldsymbol{z})\,\mathrm{d}s + \mathrm{e}^{t\boldsymbol{Q}\mathcal{L}}\boldsymbol{Q}\mathcal{L}\Phi(\boldsymbol{z}). \tag{5}$$

Note that the Mori-Zwanzig equation (5) is not an approximation but is exact and constitutes an equivalent formulation of the full problem (1). The reader is referred to [3], [4], [8], [10] and [26] for more details. As in the simple example, the Mori-Zwanzig equation (5) is in the form of a generalised Langevin equation. The first term on the right-hand side $\mathrm{e}^{\mathcal{L}t}\boldsymbol{P}\mathcal{L}\Phi(\boldsymbol{z}) = (\boldsymbol{P}\mathcal{L}\Phi(\boldsymbol{z}))(v(t,\boldsymbol{z}))$ is Markovian, the second term is a memory term, and the last term $n(\boldsymbol{z},t) = \mathrm{e}^{t\boldsymbol{Q}\mathcal{L}}\boldsymbol{Q}\mathcal{L}\Phi(\boldsymbol{z})$, which lives in the uninteresting orthogonal subspace, is labelled a noise term.

To illustrate the connection between this formulation of a deterministic system and generalised stochastic Langevin equations, we restrict our attention in the following to Hamiltonian systems with Hamiltonian $H(\boldsymbol{x},\boldsymbol{y})$. We choose the joint probability $\rho(\boldsymbol{x},\boldsymbol{y})$ to be the microcanonical equilibrium density $\rho^{\mathrm{eq}}(\boldsymbol{x},\boldsymbol{y}) = \delta(H(\boldsymbol{x},\boldsymbol{y}) - E)$, where E is the conserved energy of the system determined by the initial conditions [12]. Then, for $\Phi(\boldsymbol{z}) = \boldsymbol{x}$ (we assume here for simplicity that $\boldsymbol{x}$ and $\boldsymbol{y}$ are

conjugate variables), we can recast the memory term, evaluating the ith component of

$$\begin{aligned}\boldsymbol{P}\mathcal{L}n_i(\boldsymbol{z},s) &= \frac{1}{\Omega(\boldsymbol{x})}\int \rho^{\mathrm{eq}}(\boldsymbol{\xi},\boldsymbol{\eta})\delta(\boldsymbol{\xi}-\boldsymbol{x})\mathcal{L}\mathrm{e}^{s\boldsymbol{Q}\mathcal{L}}\boldsymbol{Q}\mathcal{L}\xi_i\,\mathrm{d}\boldsymbol{\xi}\,\mathrm{d}\boldsymbol{\eta}\\ &= -\frac{1}{\Omega(\boldsymbol{x})}\int \rho^{\mathrm{eq}}(\boldsymbol{\xi},\boldsymbol{\eta})\mathrm{e}^{s\boldsymbol{Q}\mathcal{L}}\boldsymbol{Q}\mathcal{L}\xi_i[\mathcal{L}\delta(\boldsymbol{\xi}-\boldsymbol{x})]\,\mathrm{d}\boldsymbol{\xi}\,\mathrm{d}\boldsymbol{\eta},\end{aligned}$$

as ρ^{eq} satisfies the stationary Liouville equation $\mathcal{L}^\star\rho^{\mathrm{eq}}=0$. We evaluate further

$$\begin{aligned}\boldsymbol{P}\mathcal{L}n_i(\boldsymbol{z},s) &= \frac{1}{\Omega(\boldsymbol{x})}\int \rho^{\mathrm{eq}}(\boldsymbol{\xi},\boldsymbol{\eta})\mathrm{e}^{s\boldsymbol{Q}\mathcal{L}}\boldsymbol{Q}\mathcal{L}\xi_i\left[\mathcal{L}\xi_j\frac{\partial}{\partial x_j}\delta(\boldsymbol{\xi}-\boldsymbol{x})\right]\mathrm{d}\boldsymbol{\xi}\,\mathrm{d}\boldsymbol{\eta}\\ &= \boldsymbol{M}_{ij}(\boldsymbol{x},s)\frac{\partial S}{\partial x_j}+k_B\frac{\partial \boldsymbol{M}_{ij}(\boldsymbol{x},s)}{\partial x_j},\end{aligned}$$

with the entropy $S(\boldsymbol{x})=k_B\log\Omega(\boldsymbol{x})$ and the memory matrix

$$\boldsymbol{M}_{ij}(\boldsymbol{x},t)=\frac{1}{k_B}\boldsymbol{P}([\mathcal{L}\xi_j][\mathrm{e}^{t\boldsymbol{Q}\mathcal{L}}\boldsymbol{Q}\mathcal{L}\xi_i])=\frac{1}{k_B}\boldsymbol{P}(n_i(t,\boldsymbol{z})n_j(0,\boldsymbol{z})). \tag{6}$$

Here we see that, as in generalised Langevin equations, the memory term contains information on the autocorrelation of the noise. The Mori-Zwanzig equation (5) may then be written for $v(\boldsymbol{z},t)=\boldsymbol{x}(t)$ — keeping in mind that $\boldsymbol{x}(t)$ depends on the initial conditions $(\boldsymbol{x}(0),\boldsymbol{y}(0))$ — as

$$\begin{aligned}\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} &= (\boldsymbol{P}\mathcal{L}\boldsymbol{x})(\boldsymbol{x}(t))+\int_0^t \boldsymbol{M}(\boldsymbol{x}(t-s),s)\frac{\partial S}{\partial \boldsymbol{x}}(\boldsymbol{x}(t-s))\,\mathrm{d}s\\ &\quad + k_B\int_0^t \frac{\partial \boldsymbol{M}}{\partial \boldsymbol{x}}(\boldsymbol{x}(t-s),s)\,\mathrm{d}s+n(t).\end{aligned} \tag{7}$$

Since the Mori-Zwanzig equation involves projected dynamics rather than the full dynamics, it is usually difficult to find explicit expressions that may be used in calculations and simulations.

3. Approximations of the Mori-Zwanzig equation

The Mori-Zwanzig formalism provides a conceptual framework for the study of dimension reduction and the parametrisation of uninteresting variables by a stochastic process. Ideally one would like to approximate the noise term $n(\boldsymbol{z},t)$ by white noise. Heuristically this should be possible in the case of time-scale separation or of weak coupling.

In a time-scale separated system, during one slow-time unit the fast uninteresting variables $\boldsymbol{y}$ perform many 'uncorrelated' events (provided that the fast dynamics are sufficiently chaotic). The contribution of the uncorrelated events to the dynamics of the slow interesting variables $\boldsymbol{x}$ is as a sum of independent random variables. By the weak central limit theorem this can be expressed by a normally distributed variable.

Similarly, if a large number of uninteresting variables $\boldsymbol{y}$ are weakly coupled to the resolved interesting variables $\boldsymbol{x}$, it takes many uncorrelated events of the unresolved variables to have a significant effect on the dynamics of the resolved variables. The resolved variables $\boldsymbol{x}$ experience a cumulative contribution of those events, which

again by the central limit theorem allows us to parametrise the unresolved 'heat bath' $\boldsymbol{y}$ by a random process. Here the randomness is not mediated by chaotic dynamics and time-scale separation, but by a large number of weakly coupled variables with random initial conditions drawn typically from some thermodynamic equilibrium density.

Obviously this program can only be exercised within approximations. In particular one would like to approximate the complicated memory term and render the system as Markovian. An approximation that allows for a complete analytical treatment is to approximate $\mathrm{e}^{s\boldsymbol{Q}\mathcal{L}} = \mathrm{e}^{s\mathcal{L}}$ [5]. Loosely speaking this assumption states that the resolved and the unresolved subspaces do not couple, and one may use the full dynamics in order to propagate the elements of the orthogonal subspace. This may be a good approximation for short time scales only. In this case, the noise term becomes

$$n(\boldsymbol{z},t) = \mathrm{e}^{t\boldsymbol{Q}\mathcal{L}}\boldsymbol{Q}\mathcal{L}\Phi(\boldsymbol{z}) = \mathcal{L}\boldsymbol{x}(t) - (\boldsymbol{P}\mathcal{L}\boldsymbol{x})(\boldsymbol{x}(t)),$$

which is an expression for the deviations between the full dynamics and the projected dynamics. Under the approximation $\mathrm{e}^{s\boldsymbol{Q}\mathcal{L}} = \mathrm{e}^{s\mathcal{L}}$ the memory term simplifies to

$$\begin{aligned}\int_0^t \mathrm{e}^{(t-s)\mathcal{L}}\,\boldsymbol{P}\mathcal{L}\,\mathrm{e}^{s\boldsymbol{Q}\mathcal{L}}\boldsymbol{Q}\mathcal{L}\Phi(\boldsymbol{z})\,\mathrm{d}s &= \int_0^t \mathcal{L}\mathrm{e}^{(t-s)\mathcal{L}}\mathrm{e}^{s\boldsymbol{Q}\mathcal{L}}\boldsymbol{Q}\mathcal{L}\Phi(\boldsymbol{z})\,\mathrm{d}s \\ &\quad - \int_0^t \mathrm{e}^{(t-s)\mathcal{L}}\,\mathrm{e}^{s\boldsymbol{Q}\mathcal{L}}\,\boldsymbol{Q}\mathcal{L}\boldsymbol{Q}\mathcal{L}\Phi(\boldsymbol{z})\,\mathrm{d}s \\ &= t\mathrm{e}^{t\mathcal{L}}\boldsymbol{P}\mathcal{L}\boldsymbol{Q}\mathcal{L}\Phi(\boldsymbol{z}),\end{aligned}$$

which clearly can only be valid on short time scales. However there are, to date, no rigorous estimates on the temporal range of validity of this approximation.

A commonly used approximation is the 'short memory approximation', which assumes that $n(\boldsymbol{z},t)$ is white noise, yielding

$$\boldsymbol{M}(\boldsymbol{x},t) \approx M(\boldsymbol{x})\delta(t) \quad \text{with} \quad M(\boldsymbol{x}) = \int_0^\infty \mathrm{d}t\boldsymbol{M}(\boldsymbol{x},t) = \frac{1}{k_B}\int_0^\infty \mathrm{d}tP(n(t)n(0)),$$

which renders the Mori-Zwanzig equation (7) Markovian. However to make the equation self-contained, the orthogonal subspace needs to be propagated with the full dynamics, without proper justification, and with the unwanted *plateau problem* that then $M \to 0$ (see for example [12]). In recent work [12] homogenisation methods were used to establish a more controlled approximation in the limit of infinite time scale separation between a slow variable $\boldsymbol{x}$ and a fast variable $\boldsymbol{y}$, leading to a Markovianisation that avoids the plateau problem.

4. Looking forward

The Mori-Zwanzig operator technique was developed in the 1960s and 1970s in the context of nonequilibrium statistical mechanics. It has since then mostly been seen as a mere reformulation of the dynamical system (1), and has served so far only as a philosophical motivation for the general possibility of dimension reduction of deterministic dynamical systems to stochastic Langevin equations. It seems

an interesting avenue to pursue whether this framework can actually be used in a constructive fashion to achieve controlled dimension reduction. Applications would most certainly go beyond those to climate dynamics. In particular the biological sciences and molecular dynamics community would obviously benefit from advances in this direction. To follow up the questions addressed in this brief overview requires serious advances in the applied aspect of modelling as well as in the pure aspect of proving convergence results and providing rigorous estimates.

The interested reader should start with [3], [10], [12] and [26]. In [5] an explicit example is elaborated and one can see the operator formalism 'in action' (although therein the short-memory approximation is used with the above-mentioned caveats).

An explicit word of caution for the interested reader seems appropriate. The success of the suggested research direction is by no means guaranteed, and it is not known whether the complexity of possibly non-ergodic dynamical systems allows for the development and analysis of controlled approximations within the Mori-Zwanzig framework. Worth a try though.

Acknowledgments

Thanks to Daniel Daners, Holger Dullin, Charlie Macaskill and Martin Wechselberger for valuable comments on the manuscript. The author is supported by the Australian Research Council.

References

[1] Arnold, L. (2001). Hasselmann's program revisited: the analysis of stochasticity in deterministic climate models. In: Imkeller, P., von Storch, J.-S. (eds) *Stochastic Climate Models*. Birkhäuser, Boston.

[2] Brasseur, G.P., Orlando, J.J. and Tyndall, G.S. (eds) (1999). *Atmospheric Chemistry and Global Change*. Oxford University Press.

[3] Chorin, A.J. and Hald, O.H. (2006). *Stochastic Tools in Mathematics and Science*, 2nd edn (Surveys and Tutorials in the Applied Mathematical Sciences **1**). Springer, New York.

[4] Chorin, A.J., Hald, O.H. and Kupferman, R. (2000). Optimal prediction and the Mori-Zwanzig representation of irreversible processes. *PNAS* **97,** 2968–2973.

[5] Chorin, A.J., Hald, O.H. and Kupferman, R. (2002). Optimal prediction with memory. *Physica D* **166,** 239–257.

[6] Duan, J. and Nadiga, B.T. (2006). Stochastic parameterization for large eddy simulation of geophysical flows. *Proc. American Math. Soc.* **135,** 1187–1196.

[7] Einstein, A. (1905). Investigations on the theory of Brownian movement. *Annalen der Physik* **17,** 549–560.

[8] Evans, D.J. and Morriss, G. (2008). *Statistical Mechanics of Nonequilibrium Liquids*, 2nd edn. Cambridge University Press.

[9] Frenkel, D. and Smit, B. (1996). *Understanding Molecular Simulations: from Algorithms to Applications*. Academic Press, San Diego.

[10] Givon, D., Kupferman, R., and Stuart, A. (2004). Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity* **17,** R55–R127.

[11] Hasselmann, K. (1976). Stochastic climate models: Part I. Theory. *Tellus* **28,** 473–485.

[12] Hijón, C., Español, P., Vanden-Eijnden, E. and Delgado-Buscalioni, R. (2010). Mori-Zwanzig formulation as a practical computational tool. *Faraday Discussions* **144,** 301–322.

[13] Just, W., Kantz, H., Rödenbeck, C. and Helm, M. (2001). Stochastic modelling: replacing fast degrees of freedom by noise. *J. Phys. A* **34,** 3199–3213.
[14] Langevin, P. (1908). On the theory of Brownian motion. *Comptes Rendues* **146,** 530–533.
[15] Leith, C.E. (1975). Climate response and fluctuation dissipation. *J. Atmos. Sci.* **32,** 2022–2025.
[16] Majda, A., Timofeyev, I. and Vanden-Eijnden, E. (2002). A priori tests of a stochastic mode reduction strategy. *Physica D* **170,** 206–252.
[17] Mori, H., Fujisaka, H. and Shigematsu, H. (1974). A new expansion of the master equation. *Prog. Theor. Phys.* **51,** 109–122.
[18] Nadiga, B.T. and Luce, B. (2001). Global bifurcation of Shilnikov type in a double-gyre ocean model. *J. Phys. Ocean.* **31,** 2669–2690.
[19] Pavliotis, G.A. and Stuart, A.M. (2008). *Multiscale Methods: Averaging and Homogenization* (Texts in Applied Mathematics **53**). Springer, New York.
[20] Riegert, A., Baba, N., Gelfert, K., Just, W. and Kantz, H. (2005). Hamiltonian chaos acts like a finite energy reservoir: accuracy of the Fokker–Planck approximation. *Phys. Rev. Lett.* **94,** 054103.
[21] Rödenbeck, C., Beck, C. and Kantz, H. (2001). Dynamical systems with time scale separation: averaging, stochastic modelling, and central limit theorems. In: Imkeller, P. and von Storch, J.-S. (eds) *Stochastic Climate Models*. Birkhäuser, Boston.
[22] Stinis, P. (2006). A comparative study of two stochastic mode reduction methods. *Physica D* **213,** 197–213.
[23] Stinis, P. (2007). Higher order Mori-Zwanzig models for the Euler equations. *Multiscale Model. Simul.* **6,** 741–760.
[24] Uhlenbeck, G.E. and Ornstein, L.S. (1930). On the theory of Brownian motion. *Phys. Rev.* **36,** 823–841.
[25] Zwanzig, R. (1960). Ensemble method in the theory of irreversibility. *J. Chem. Phys.* **33,** 1338–1341.
[26] Zwanzig, R. (2001). *Nonequilibrium statistical mechanics*. Oxford University Press.