An applied mathematics perspective on stochastic modelling for climate

BY ANDREW J. MAJDA 1,*, CHRISTIAN FRANZKE 2 AND BOUALEM KHOUIDER 3

1 Department of Mathematics and Center for Atmosphere-Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, 10012 NY, USA
2 British Antarctic Survey, Natural Environment Research Council, Cambridge CB3 0ET, UK
3 Department of Mathematics and Statistics, University of Victoria, Victoria, BC, Canada V8W 3P4

Systematic strategies from applied mathematics for stochastic modelling in climate are reviewed here. One of the topics discussed is the stochastic modelling of mid-latitude low-frequency variability through a few teleconnection patterns, including the central role and physical mechanisms responsible for multiplicative noise. A new low-dimensional stochastic model is developed here, which mimics key features of atmospheric general circulation models, to test the fidelity of stochastic mode reduction procedures. The second topic discussed here is the systematic design of stochastic lattice models to capture irregular and highly intermittent features that are not resolved by a deterministic parametrization. A recent applied mathematics design principle for stochastic column modelling with intermittency is illustrated in an idealized setting for deep tropical convection; the practical effect of this stochastic model in both slowing down convectively coupled waves and increasing their fluctuations is presented here.

Keywords: low-frequency variability; tropical convection; multiplicative noise; intermittency

1. Introduction

Stochastic modelling for climate is important for understanding the intrinsic variability of dominant low-frequency teleconnection patterns in climate, to provide cheap low-dimensional computational models for the coupled atmosphere–ocean system and to reduce model error in standard deterministic computer models for extended-range prediction through appropriate stochastic noise (Palmer 2001).

The present contribution is a research-expository paper on systematic strategies for stochastic climate modelling from the perspective of modern applied mathematics. In the modern applied mathematics, ‘modus operandi’ rigorous mathematical analysis, qualitative, asymptotic and numerical modelling are all

* Author for correspondence (majda@cims.nyu.edu).

One contribution of 12 to a Theme Issue ‘Stochastic physics and climate modelling’.

2429 This journal is © 2008 The Royal Society
blended together in a multidisciplinary fashion to provide systematic guidelines to address real-world problems (Majda 2000). For stochastic modelling in climate, the modern applied mathematics tool kit includes stochastic differential equations and discontinuous Markov jump processes (Gardiner 1985), systematic asymptotic reduction techniques, nonlinear dynamical systems theory and ideas from both statistical physics (Majda & Wang 2006) and mathematical statistics (Kravtsov et al. 2005; Majda et al. 2006a); mathematical rigour provides unambiguous guidelines in idealized models. Another facet of the modern applied mathematics philosophy is the development of qualitative models that represent a Platonic ideal for central issues simultaneously in diverse scientific disciplines such as material science, biomolecular dynamics and climate science.

In §2, we illustrate and apply this modern applied mathematics philosophy to stochastic modelling of the low-frequency variability of the atmosphere. The systematic mathematical theory (Majda et al. 1999, 2001, 2002, 2003, 2006b; collectively referred to as MTV hereafter; Franzke et al. 2005; Franzke & Majda 2006) for these problems is briefly reviewed including the central role and physical mechanisms responsible for multiplicative noise in the low-frequency dynamics. In this context, the Platonic ideal from applied mathematics is the truncated Burgers–Hopf model (Majda & Timofeyev 2000). A new simplified low-dimensional stochastic model that reproduces key features of atmospheric general circulation models (GCMs) is used there to test the fidelity of stochastic model reduction techniques. A recent diagnostic statistical test with firm mathematical underpinning for understanding and interpreting the dynamical sources of the small departures from Gaussianity in low-frequency variables (Franzke et al. 2007) is also developed there.

While §2 deals with applied mathematical modelling through stochastic differential equations and §3 is devoted to the systematic development of stochastic lattice models to capture unresolved features that are highly intermittent in space and time such as deep convective clouds, cloud cover in subtropical boundary layers, sub-mesoscale eddies in the ocean and mesoscale sea ice cover. Here the mathematical tools involve a family of discontinuous Markov jump processes with multi-scale behaviour in space–time called stochastic spin-flip models. The key mathematical development involves systematic strategies to coarse grain such stochastic spin-flip models to achieve computational efficiency while retaining crucial features of the microscale interactions (Katsoulakis & Vlachos 2003; Katsoulakis et al. 2003a,b). The use of such stochastic lattice models to parametrize key features of tropical convection is briefly reviewed (Majda & Khouider 2002; Khouider et al. 2003). For the coupling of continuum models like a GCM to a stochastic lattice model as well as in many diverse applications, an applied mathematics Platonic ideal model has recently been introduced and analysed by Katsoulakis et al. (2004, 2005, 2006, 2007; hereafter KMS). This model consists of a system of ordinary differential equations (ODEs) for continuum variables $X$, 
\[ \frac{dX}{dt} = F(X, \sigma), \]  
(1.1)

two-way coupled to a stochastic spin-flip model written abstractly here as 
\[ \frac{d}{dt} \mathbb{E}f(\sigma) = \mathbb{E}Lf(\sigma), \]  
(1.2)
where $\bar{\sigma}$ denotes the spatial coverage; $L$ is the generator; $f$ is a test function; and $\mathbb{E}$ denotes the expected value. This idealized class of models has been used to systematically analyse the effects of various coarse-graining procedures on processes with intermittency, large-scale bifurcations and microscale phase transitions (KMS 2004, 2005, 2006, 2007). A concrete example for tropical convection in climate is given in §3. A new application of these stochastic lattice models to capture intermittent features and improve the fidelity of deterministic parametrizations of convection with clear deficiencies is developed in §3. First, the systematic design principles for (1.1) and (1.2) (KMS 2006, 2007) are used to calibrate a stochastic column model for tropical convection with intermittency and then the new results are presented on the practical effect of slowing down convectively coupled waves and increasing their fluctuations through the stochastic lattice models.

2. Systematic low-dimensional stochastic mode reduction and atmospheric low-frequency variability

A remarkable fact of Northern Hemisphere low-frequency variability is that it can be efficiently described by only a few teleconnection patterns that explain most of the total variance (e.g. Wallace & Gutzler 1981). These few teleconnection patterns not only exert a strong influence on regional climate and weather but are also related to climate change (Hurrell 1995). These properties of teleconnection patterns make them an attractive choice as basis functions for climate models with a highly reduced number of d.f. The development of such reduced climate models involves the solution of two major issues: (i) how to properly account for the unresolved modes that are also known as the closure problem, and (ii) how to define a small set of basis functions that optimally represent the dynamics of the major teleconnection patterns. This section addresses primarily the first issue and presents a rigorous strategy of how to systematically account for the unresolved d.f.

The simplest approach to derive highly truncated models of teleconnection patterns is to empirically fit simple stochastic models (e.g. autoregressive models and fractionally differenced models) to individual scalar teleconnection indices (Feldstein 2000; Stephenson et al. 2000; Percival et al. 2001). Statistical tests usually cannot distinguish if short- or long-memory models provide the better fit. A more complex approach, which also tries to capture deterministic interactions between different teleconnection patterns, is to linearize the equations of motion around a climatological mean state. Such models can be determined empirically from data or by using the linearized equations of motion. These models can be forced either by a random forcing (Branstator 1990; Newman et al. 1997; Whitaker & Sardeshmukh 1998; Zhang & Held 1999) or by an external forcing representing tropical heating (Branstator & Haupt 1998). To ensure stability of these linear models, damping is added according to various ad hoc principles. There is a recent survey of such modelling strategies (Delsole 2004).

A more powerful method is to empirically fit nonlinear stochastic models with possibly multiplicative (state dependent) noise by using the Fokker–Planck equation (Gardiner 1985; Sura 2003; Berner 2005). To reliably estimate the drift and diffusion coefficients in the Fokker–Planck equation is a subtle inverse
problem that requires very long time series, and is further complicated by the need to retain the leading-order eigenvalue structure of the Fokker–Planck operator in order to keep the autocorrelation time scales of the original model (Crommelin & Vanden-Eijnden 2006); the fitting procedure in most of the recent work is the most attractive current regression strategy for low-frequency behaviour. Recently, Kravtsov et al. (2005) have developed a simplified nonlinear regression strategy that produces very good results for a three-layer quasi-geostrophic model with a realistic climate. However, order 2000 regression coefficients need to be fitted in a model with order 1000 state variables to achieve these results. Some inherent limitations of this approach in describing the correct physics are discussed briefly below in a simplified model.

All the work presented above derives reduced models by regression fitting of the resolved modes. Another approach is to take advantage of the basis function property of teleconnection patterns. Schubert (1985), Selten (1995), Achatz & Branstator (1999) and Achatz & Opsteegh (2003a,b) developed low-order models with empirical orthogonal functions (EOFs) as basis functions. Truncated EOF models experience climate drift due to the neglected interactions with the unresolved modes. Selten (1995) and Achatz & Branstator (1999) parametrize these neglected interactions by a linear damping, whose strength is determined empirically. A possibly more powerful tool to represent the dynamics of a system is principal interaction patterns (PIPs; Hasselmann 1988; Kwasniok 1996, 2004). The calculation of PIPs takes into account the dynamics of the model for which one tries to find an optimal basis and also often involves ad hoc closure through linear damping and an ansatz for nonlinear interactions. Crommelin & Majda (2004) compare different optimal bases. They find that the models based on PIPs are superior to models based on EOFs. On the other hand, they also point out that the determination of PIPs can show sensitivities regarding the calculation procedure, at least for some low-order atmospheric dynamical systems with regime transitions. Furthermore, PIPs have two more disadvantages: (i) much higher computational cost than EOFs, and (ii) one needs not only data as for EOFs but also the dynamical equations to calculate PIPs. These features can make PIPs possibly a less attractive basis.

Majda et al. (1999, 2001, 2002, 2003, 2005, 2006a,b) provide a systematic framework for how to account for the effect of the fast d.f. on the slow modes in combination with using the dominant teleconnection patterns as basis functions. In contrast to the empirical fitting procedures applied in the studies discussed above, the stochastic mode reduction strategy put forward in MTV predicts the functional form of all deterministic and stochastic correction terms and provides a minimal regression fitting procedure of only the fast modes (Franzke et al. 2005; Franzke & Majda 2006). In general, only an estimate for the variance and eddy turnover time for each fast mode is needed. It has been applied and tested on a wide variety of simplified models and examples.

(a) Overview of the MTV strategy

We illustrate the ideas for stochastic climate modelling by considering the following prototype equation for geophysical flow:

\[
\frac{\partial u}{\partial t} = F + Lu + B(u, u). \tag{2.1}
\]
An applied mathematics perspective

The above functional form (2.1) is typical of dry dynamical cores of climate models, but the MTV strategy is easily extended to include non-quadratic nonlinearities such as associated with boundary fluxes and moist processes. In stochastic climate modelling, the variable $u$ is decomposed into an orthogonal decomposition through the variables $\tilde{u}$ and $u'$, which are characterized by strongly differing time scales (MTV 1999; Majda et al. 2005). The variable $\tilde{u}$ denotes a slow low-frequency mode (also referred to as climate mode) of the system, which evolves slowly in time compared with the $u'$ variables (also referred to as fast mode). By decomposing $u = \tilde{u} + u'$ in terms of some optimal energy norm basis, we can write them as

$$u = \sum_{i=1}^{N} a_i e_i = \sum_{i=1}^{R} \alpha_i e_i + \sum_{j=R+1}^{N} \beta_j e_j,$$

with $\tilde{u} = \sum_{i=1}^{R} \alpha_i e_i$ and $u' = \sum_{j=R+1}^{N} \beta_j e_j$, where $R$ is the number of climate modes; $a_i$ denotes the expansion coefficients; and $\alpha_i$ and $\beta_j$ are the expansion coefficients of the slow (fast) modes. The use of the energy norm ensures the conservation of energy by the nonlinear operator (Selten 1995). By properly projecting the energy norm basis, derived from the geophysical model, onto equation (2.1), we get two sets of equations for slow $\alpha_i$ and fast $\beta_j$ modes

$$\dot{\alpha}_i(t) = \varepsilon H^\alpha_i + \sum_j L_{ij}^\alpha \alpha_j(t) + \frac{1}{\varepsilon} \sum_j L_{ij}^\beta \beta_j(t) + \sum_{jk} B_{ijk}^\alpha \alpha_j(t) \alpha_k(t),$$

$$\quad + \frac{2}{\varepsilon} \sum_{jk} B_{ijk}^{\alpha \beta} \alpha_j(t) \beta_k(t),$$

$$\dot{\beta}_i(t) = \varepsilon H^\beta_i + \frac{1}{\varepsilon} \sum_j L_{ij}^\beta \alpha_j(t) + \frac{1}{\varepsilon} \sum_j L_{ij}^\beta \beta_j(t) + \frac{1}{\varepsilon} \sum_{jk} B_{ijk}^\beta \alpha_j(t) \alpha_k(t),$$

$$\quad + \frac{2}{\varepsilon} \sum_{jk} B_{ijk}^{\alpha \beta} \alpha_j(t) \beta_k(t),$$

where the nonlinear operators have been symmetrized, that is, $B_{ijk} = B_{ikj}$ in (2.3) and (2.4). The upper indices $\alpha$ and $\beta$ indicate the respective subsets of the full operators in (2.1). Here $\varepsilon$ is a small positive parameter that controls the separation of time scale between the slow and fast modes and measures the ratio of the correlation time of the slowest non-climate mode $u'$ to the fastest climate mode $\tilde{u}$. In placing the parameter in front of particular terms, we tacitly assume that they evolve on a faster time scale than the terms involving the climate modes alone (see MTV (2001) and Franzke et al. (2005) for more details). Ultimately, $\varepsilon$ is set to the value $\varepsilon = 1$ in developing all the final results (MTV 2002, 2003), that is, introducing $\varepsilon$ is only a technical step in order to carry out the MTV mode reduction strategy. Such a use of $\varepsilon$ has been checked on a wide variety of idealized examples where the actual value of $\varepsilon$ ranges from quite small to order one (MTV 2002, 2003, 2006b; Majda & Timofeyev 2004). Following MTV (1999, 2001, 2002, 2003, 2006b) and Franzke et al. (2005), the mode...
The elimination procedure is based on the assumption that the dynamics of the fast modes alone in equation (2.4), that is, the dynamical system

$$\dot{c}_i = \sum_{jk} D_{ijk} \beta \beta_k c_j c_k$$  \hspace{1cm} (2.5)

is ergodic and mixing with integrable decay of correlation. In other words, we assume that for almost all initial conditions, and suitable functions \(f\) and \(g\), we have

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T f(c(t)) \, dt = \langle f \rangle,$$  \hspace{1cm} (2.6)

where \(\langle \cdot \rangle\) denotes expectation with respect to some appropriate invariant distribution and

$$G(s) = \lim_{T \to \infty} \frac{1}{T} \int_0^T g(c(t + s), c(t)) \, dt - \lim_{T \to \infty} \frac{1}{T^2} \int_0^T \int_0^T g(c(t), c(t')) \, dt \, dt'$$  \hspace{1cm} (2.7)

is an integrable function of \(s\), that is, \(|\int_0^s G(s) \, ds| < \infty\). Furthermore, we assume that the low-order statistics for the fast modes in (2.5) are Gaussian. Under the above assumptions, it can be shown in the limit \(\epsilon \to 0\) (Kurtz 1973; MTV 2001) that the dynamics of the slow modes \(\alpha_i\) in (2.3) can be written as the following Itô stochastic equation for the slow modes alone:

$$\begin{align*}
\, \text{d} \alpha_i(t) &= \lambda_B \left( H_i^\alpha \, \text{d} t + \sum_j L_{ij}^{\alpha \alpha} \alpha_j(t) \, \text{d} t + \sum_{jk} D_{ijk}^{\alpha \alpha} \alpha_j(t) \alpha_k(t) \, \text{d} t \right) \\
&\quad + \lambda_A \sum_j \tilde{L}_{ij}^{(2)} \alpha_j(t) \, \text{d} t + \lambda_A \sqrt{2} \sum_j \sigma_{ij}^{(2)} \, \text{d} W_j^{(2)} \\
&\quad + \lambda_M \left( \sum_j \tilde{L}_{ij}^{(3)} \alpha_j(t) \, \text{d} t + \sum_{jk} \tilde{M}_{ijk} \alpha_j(t) \alpha_k(t) \alpha_l(t) \, \text{d} t \right) \\
&\quad + \lambda_L \left( \sum_j \tilde{L}_{ij}^{(1)} \alpha_j(t) \, \text{d} t \right) + \lambda_M \lambda_L \left( \tilde{H}_j^{(1)} \, \text{d} t + \sum_{jk} \tilde{B}_{ijk} \alpha_j(t) \alpha_k(t) \, \text{d} t \right) \\
&\quad + \lambda_A \lambda_L \tilde{H}_j^{(2)} \, \text{d} t + \sqrt{2} \sum_j \sigma_{ij}^{(1)}(\alpha(t)) \, \text{d} W_j^{(1)},
\end{align*}$$  \hspace{1cm} (2.8)

where the nonlinear noise matrix \(\sigma^{(1)}\) satisfies

$$\lambda_L^2 Q_{ij}^{(1)} + \lambda_L \lambda_M \sum_k U_{ijk} \alpha_k(t) + \lambda_M^2 \sum_{kl} V_{ijkl} \alpha_k(t) \alpha_l(t) = \sum_k \sigma_{jk}^{(1)}(\alpha(t)) \sigma_{jk}^{(1)}(\alpha(t)).$$  \hspace{1cm} (2.9)

It is guaranteed (MTV 2001) that the operator on the l.h.s. of (2.9) is always positive definite ensuring the existence of the nonlinear noise matrix on the r.h.s. All coefficients are defined explicitly in MTV (2001) and Franzke et al. (2005). A comprehensive mathematical theory of the stochastic mode reduction strategy for geophysical applications is developed in MTV (2001) with many new mathematical phenomena in the resulting equations explored there.

Phil. Trans. R. Soc. A (2008)
To see which of these correction terms play a vital role in the integrations of the low-order stochastic model (2.8), we grouped the interaction terms between the slow and fast modes according to their physical origin and set a parameter \( \lambda_i \) in front of the corresponding interaction coefficient (see Franzke et al. 2005; Franzke & Majda 2006 for more details). The bare truncation is indicated by a \( \lambda_B \) and describes the interaction between the slow modes. The interaction between the triads \( B^{\alpha \beta \gamma} \) and \( B^{\beta \alpha \delta} \) gives rise to additive noise and a linear correction term and arises from the advection of the fast modes by the slow ones; we name these triads ‘additive’ triads and set a \( \lambda_A \) in front of them (MTV 1999, 2001, 2002, 2003). The other type of triad interaction is between \( B^{\alpha \beta \gamma} \) and \( B^{\beta \alpha \delta} \). These interactions create multiplicative noises and cubic nonlinear correction terms (MTV 1999, 2001, 2002, 2003); we call them ‘multiplicative’ triads hereafter and indicate them by a \( \lambda_M \). These triad interactions describe the advection of the slow modes by the fast modes that induce tendencies in the slow modes. The linear coupling between the slow and fast modes \( L^{\alpha \beta} \) and \( L^{\beta \alpha} \) gives rise to additive noise and a linear correction term (MTV 2001; Franzke et al. 2005), which is called the augmented linearity here and is indicated by a \( \lambda_L \). The augmented linearity describes the effect of the linear interaction between the fast (slow) modes and the climatological mean state onto the slow (fast) modes and is the main interaction captured in the linear stochastic modelling strategy (Delsole 2004).

We set a \( \lambda_F \) in front of the last remaining interaction term \( L^{\beta \gamma} \), the linear coupling of the fast modes. The quadratic nonlinear corrections, a forcing term and a further multiplicative noise contribution are caused by the interaction between the linear coupling terms and the multiplicative triads. Another forcing correction term comes from the interaction between additive triads and the linear coupling of the fast modes.

In ch. 3 of Majda et al. (2005), a simplified three-mode elementary ‘toy climate model’ is discussed and the MTV procedure is applied explicitly to that example. The origin of all the terms in (2.8) is developed in a transparent fashion in these examples. Once the low-order stochastic model has been developed from the above procedure, one can assess the importance of the various deterministic and stochastic processes systematically by varying the coefficients \( \lambda_B, \lambda_A, \lambda_L, \lambda_M \) and \( \lambda_F \) systematically in (2.8) (Franzke et al. 2005) and even develop simple physically motivated regression fitting strategies (Franzke & Majda 2006). In interesting recent work, Sura & Sardeshmukh (2008) have used scalar linear stochastic models with multiplicative and additive noise to explain non-Gaussian sea surface temperature (SST) variability. If the reduced stochastic models in (2.10) from the MTV procedure are linearized at the climate mean state, they automatically produce vector systems of linear stochastic equations with both multiplicative and additive noise with the same structure, with clear sources for the underlying physical contributions to this equation.

(b) Idealized models for stochastic mode reduction

The idealized models, where the procedure has been tested, have order 100 d.f. and include those with trivial climates (MTV 2002), periodic orbits or multiple equilibria (MTV 2003) and heteroclinic chaotic orbits coupled to a deterministic bath of modes satisfying the truncated Burgers–Hopf equation (Majda & Timofeyev 2000, 2004). The truncated Burgers–Hopf equation is a toy model.
with some remarkable features mimicking behaviour in the real atmosphere; it has a well-defined equipartition spectrum and a simple scaling theory for correlations with the large scales decorrelating more slowly than the small scales, that is, low-frequency variability. Furthermore, these predictions are confirmed with very high precision by numerical simulations (Majda & Timofeyev 2000; Majda & Wang 2006). The MTV procedure has been validated in these examples even when there is little separation of time scales between the slow and fast modes. In the example of a four-dimensional resonant system with chaotic dynamics coupled to the truncated Burgers system, only one empirical regression fitting coefficient is used and complex bifurcation diagrams and probability distribution functions (pdfs) in a climate change scenario are reproduced by the four-dimensional stochastic mode reduction resulting from the MTV procedure applied to the 104 d.f. deterministic system. An especially stringent recent test is the application of this procedure to the first few large-scale modes of the truncated Burgers equations in the turbulent cascade (MTV 2006).

(c) A priori stochastic modelling for mountain torque

The ideal barotropic quasi-geostrophic equations with a large-scale zonal mean flow $U$ on a $2\pi \times 2\pi$ periodic domain (Carnevale & Frederiksen 1987) are given by

$$\frac{\partial q}{\partial t} + \nabla \cdot \psi q + U \frac{\partial q}{\partial x} + \beta \frac{\partial \psi}{\partial x} = 0,$$

$$q = \Delta \psi + h,$$

$$\frac{dU}{dt} = \frac{1}{4\pi^2} \int h \frac{\partial \psi}{\partial x} \, dx \, dy,$$

where $q$ is the potential vorticity; $U$ is the large-scale zonal mean flow; $\psi$ is the stream function; and $h$ is the topography. In (2.10), the mean flow changes in time through the topographic stress; this effect is the direct analogue for periodic geometry of the change in time of angular momentum due to mountain torque in spherical geometry (Frederiksen et al. 1996; Majda & Wang 2006). Here the a priori stochastic modelling strategy (MTV 1999, 2001) is applied to the stochastic modelling of the topographic stress terms in (2.10) as an analogue for mountain torque; thus, the variable $U$ is the slow variable while all the modes $\psi_k$ are fast variables for the MTV procedure.

In this example, the systematic stochastic modelling procedure (MTV 2003) results in the predicted nonlinear reduced equation for $U$,

$$\frac{dU}{dt} = -\gamma(U)U + \frac{\gamma'(U)}{\alpha \mu} + \sqrt{\frac{2\gamma(U)}{\alpha \mu}} \dot{W},$$

where $\gamma'(U) = d\gamma/dU$ and

$$\gamma(U) = 2 \sum_k \frac{\mu k_x^2 |H_k|^2 \gamma_k}{\gamma_k^2 + (\Omega_k - k_x U)^2}.$$
Here $\Omega_k = k_x \beta / |k|^2 - g k_x$ is the Rossby wave frequency Doppler shifted by the mean flow. Under the additional assumption that $|k_x U|^2 \ll \gamma_k^2 + (\Omega_k)^2$, a standard predicted linear stochastic model for $U$ emerges from (2.11) with $\gamma = \gamma(0)$ from (2.12) and $\gamma'(0) = 0$ (MTV 1999, 2001; see Majda et al. (2003) for definitions of $\mu$ and $\gamma_k$). It is shown in MTV (2003) that this nonlinear stochastic equation is superior to the linear one for large amplitude topography, where $\varepsilon \equiv 0.7$. This is the simplest example with multiplicative noise. Egger (2005) has used the systematic strategies from MTV (1999, 2003) to improve regression strategies for analyzing observational data for angular momentum. In Majda et al. (2006a), the recent systematic regression fitting strategy mentioned earlier (Crommelin & Vanden-Eijnden 2006) is applied to (2.10) and independently confirms the predictions in (2.11) and (2.12).

\((d)\) Geophysical and climate models

Franzke et al. (2005) put the above systematic stochastic mode reduction strategy in a form that makes the practical implementation of the MTV procedure in the complex geophysical models simpler with the same reduced stochastic equations for the fast modes. EOFs in an appropriate metric are used to distinguish between the slow and fast modes in high-dimensional geophysical systems, such as climate models. Even though EOFs do not strictly decompose the modes according to their autocorrelation time scale, the leading EOFs usually evolve on a slower time scale than the higher EOFs. In the study by Franzke et al. (2005), a T21-truncated barotropic model on the sphere with a realistic climate was used to derive low-order stochastic models based on kinetic energy norm EOFs by the MTV strategy. Low-order models with as little as two slow modes succeed in capturing the geographical distributions of the climatological mean field, the variance and the eddy forcing. Furthermore, the envelope of the autocorrelation functions is captured reasonably well.

Recently, Franzke & Majda (2006) applied the systematic stochastic mode reduction strategy to a baroclinic three-layer quasi-geostrophic model on the sphere (Marshall & Molteni 1993), which mimics the climatology of the European Centre for Medium-Range Weather Forecasts reanalysis data. The low-order stochastic climate model consists of climate modes as slow modes defined as the leading total energy norm EOFs and the stochastic mode reduction procedure predicts all forcing, linear, quadratic and cubic correction terms as well as additive and multiplicative noises; these correction terms and noises account for the interaction of the climate modes with the neglected non-climate modes and the self-interaction among the non-climate modes. For the three-layer quasi-geostrophic model low-order stochastic models with 10 or less climate modes reproduce the geographical distributions of the standard deviation and eddy forcing well. They underestimate the standard deviations by at most a factor of approximately 1.5. Furthermore, they reproduce the autocorrelation functions reasonably well. A budget analysis shows that both linear and nonlinear correction terms as well as both additive and multiplicative noises are important. The physical intuition behind the noises as derived from the MTV procedure is as follows: the additive noise stems from the linear interaction between the fast modes and the climatological mean state and the multiplicative noise comes from the advection of the slow modes by the fast modes. All these deterministic correction
terms and noises (both additive and multiplicative) are predicted by the systematic stochastic mode reduction strategy, whereas previous studies a priori approximate the nonlinear part of the equations by a linear operator and additive noise. This noise is typically white in time but may be spatially correlated. In other words, these studies truncate the dynamics on both the slow and fast modes and add ad hoc damping in order to stabilize the linear model (Whitaker & Sardeshmukh 1998; Zhang & Held 1999). The systematic MTV approach summarized briefly above truncates the dynamics only on the fast modes and predicts the functional form of all necessary nonlinear correction terms and noises; therefore, it also predicts the necessary damping.

The MTV stochastic climate models for this application experience some climate drift. A minimal empirical MTV model without climate drift can be constructed through three-parameter regression fitting by downscaling the bare truncation terms and upscaling the two important MTV processes (augmented linearity and multiplicative triads). These empirical MTV stochastic climate models with minimal regression fitting still capture the geographical distribution of the standard deviation and eddy forcing and the autocorrelation functions reasonably well, while not experiencing climate drift. This surprising result can be interpreted as the fact that the climate modes are predominantly driven by the fast modes and the self-interactions among the slow modes are less important, as can already be seen from the bare truncation models, which do not capture any feature of the actual dynamics. Furthermore, these empirical MTV stochastic climate models suggest that the bare truncation is probably the cause of the climate drift. Integrations of bare truncation models (without any MTV correction terms) already produce a big climate drift (Franzke & Majda 2006). The MTV mode reduction procedure is able to reduce the climate drift in most of the slow modes, but is not able to overcome it completely. Previous results with a variety of simplified models show no climate drift in an MTV framework (MTV 1999, 2001, 2002, 2003; Majda & Timofeyev 2004). This is probably because these simplified models are constructed in such a way that they have an optimal basis that captures the dynamics of the climate modes. This gives evidence that total energy norm EOFs are not an adequate dynamical basis in capturing the dynamics of the slow modes. Further details of this application can be found in Franzke & Majda (2006).

(e) A simple stochastic model with key features of atmospheric low-frequency variability

In this section, we present a four-mode stochastic climate model of the kind put forward in Majda et al. (2005, ch. 3). This simple stochastic climate model is set-up in such a way that it features many of the important dynamical features of comprehensive GCMs but with many fewer d.f. Such simple toy models allow the efficient exploration of the whole parameter space that is impossible to conduct with GCMs. Thus, we are able to test the predictions of the above framework with direct model experiments by switching on and off certain terms rather than relying on diagnostic methods.

While this model is not rigorously derived from a geophysical flow model (e.g. barotropic vorticity equation), it has the same functional form one would end up with when deriving a reduced stochastic model from a geophysical model.
Thus, consistent with geophysical flow models, the toy model has a quadratically nonlinear part that conserves energy, a linear operator and a constant forcing, which in a geophysical model would represent the effects of external forcing such as solar insulation and sea surface temperature. The linear operator has two contributions: one is a skew-symmetric part formally similar to the Coriolis effect and topographic Rossby wave propagation. The other is a negative definite symmetric part formally similar to dissipative processes such as surface drag and radiative damping.

The model is constructed in such a way that there are two modes, denoted by \( x \), that evolve more slowly than the other two modes, \( y \). In the realistic models, there would be very many additional fast modes representing, for example, synoptic weather systems or convection. To mimic their combined effect, we include damping and stochastic forcing \(- (\gamma / \epsilon) y + (\sigma / \sqrt{\epsilon}) W\) in the equations for \( y \), where \( W \) denotes a Wiener process. The motivation for this approximation is that these fast modes are associated with turbulent energy transfers and strong mixing and that we do not require a more detailed description since we are only interested in their effect on the slow modes. The two fast modes carry most of the variance in this model but as noted earlier, these two modes are surrogates in the model for the entire bath of fast modes so this is very natural. Note that this stochastic Ansatz is different from most of the previous studies (e.g. Newman et al. 1997; Whitaker & Sardeshmukh 1998) that truncated the model on both the resolved and unresolved modes and replaced all nonlinear terms by a stochastic process. In this study, only the nonlinear interactions among the fast modes are modelled by this stochastic Ansatz. The parameter \( \epsilon \) controls the time-scale separation between the slow and fast variables. For testing the predictions of the general framework that is derived in the previous section, we will treat the two slow modes \( x \) as the climate modes and the two fast modes as the non-climate modes \( y \). Therefore, our toy model has the following form:

\[
\begin{align*}
\frac{dx_1}{dt} &= (-x_2(L_{12} + a_1 x_1 + a_2 x_2) + d_1 x_1 + F_1) + L_{13} y_1 + b_{123} x_2 y_1, \\
\frac{dx_2}{dt} &= (+x_1(L_{21} + a_1 x_1 + a_2 x_2) + d_2 x_2 + F_2) + L_{24} y_2 + b_{213} x_1 y_1, \\
\frac{dy_1}{dt} &= (-L_{13} x_1 + b_{312} x_1 x_2 + F_3 - \frac{\gamma_1}{\epsilon} y_1) + \frac{\sigma_1}{\sqrt{\epsilon}} dW_1, \\
\frac{dy_2}{dt} &= (-L_{24} x_2 + F_4 - \frac{\gamma_2}{\epsilon} y_2) + \frac{\sigma_2}{\sqrt{\epsilon}} dW_2.
\end{align*}
\tag{2.13}
\]

Note that in this case we use a different scaling in powers of \( \epsilon \) in order to derive easily a solution by working directly on the stochastic differential equation instead of the Fokker–Planck equation as for equation (2.8) (MTV 1999, 2001). To ensure energy conservation of the nonlinear operator, the coefficients have to satisfy the following relation: \( b_{123} + b_{213} + b_{312} = 0 \), while the nonlinear bare truncation terms also conserve energy. In this particular set-up, the slow and the fast modes are coupled through two mechanisms: one is a skew-symmetric linear coupling and the other is nonlinear triad interaction. The nonlinear coupling involving \( b_{ijk} \) produces multiplicative noise in the MTV framework (1999, 2003; see also equation (2.15a)–(2.15d)) so we refer to it as a multiplicative triad. One advantage of the model in (2.13a)–(2.13d) is that the stochastic mode reduction

\cite{Phil. Trans. R. Soc. A (2008)}
as $\varepsilon \to 0$ is done explicitly through elementary manipulations (MTV 1999, 2001; Majda et al. 2005). The corresponding reduced Itô stochastic differential equation (SDE) for the climate variables alone is given by

$$
\begin{aligned}
\text{d}x_1(t) &= (-x_2(t)(L_{12} + a_1 x_1(t) + a_2 x_2(t)) + d_1 x_1(t) + F_1) \, \text{d}t \\
&\quad + \frac{\varepsilon}{\gamma_1} (L_{13} F_3 - L_{13} L_{13} x_1(t) + b_{123} F_3 x_2(t) + L_{13} b_{312} x_1(t) x_2(t) \\
&\quad - L_{13} b_{123} x_1(t) x_2(t) + b_{312} b_{123} x_1(t) x_2^2(t)) \, \text{d}t \\
&\quad + \varepsilon \frac{1}{2} \frac{\sigma_1^2}{\gamma_1} b_{213} b_{123} x_1(t) \, \text{d}t \sqrt{\varepsilon / \gamma_1} (L_{13} + b_{123} x_2(t)) \, \text{d}W_1(t), \quad (2.14a)
\end{aligned}
$$

$$
\begin{aligned}
\text{d}x_2(t) &= (x_1(t)(L_{21} + a_1 x_1(t) + a_2 x_2(t)) + d_2 x_2(t) + F_2) \, \text{d}t \\
&\quad + \frac{\varepsilon}{\gamma_2} (L_{24} F_4 - L_{24} L_{24} x_2(t)) \, \text{d}t + \frac{\varepsilon}{\gamma_1} (-b_{213} L_{13} x_1^2(t) \\
&\quad + b_{213} b_{312} x_1^2(t) x_2(t) + b_{213} F_3 x_1(t)) \, \text{d}t + \varepsilon \frac{1}{2} \frac{\sigma_1^2}{\gamma_1} (b_{213} b_{123} x_2(t) \\
&\quad + L_{13} b_{213}) \, \text{d}t + \sqrt{\varepsilon / \gamma_1} b_{213} x_1(t) \, \text{d}W_1(t) + \sqrt{\varepsilon / \gamma_2} L_{24} \, \text{d}W_2(t). \quad (2.14b)
\end{aligned}
$$

Note that coarse-graining time as $t \to t/\varepsilon$ amounts to setting $\varepsilon = 1$, because by coarse-graining time the factors of $\varepsilon$ disappear (see Majda et al. (2005) for more details).

To evaluate the performance of the reduced dynamics, we calculate the autocorrelation function $\rho(s) = \langle x'(t + s) x'(t) \rangle / \langle x'(t) x'(t) \rangle$ and the third-order two-time moment $K(s) = \langle x'^2(t + s) x'(t) \rangle / \langle x'(t) x'^2(t) \rangle^{3/2}$, which is a measure of deviations from Gaussianity (MTV 2002; Majda et al. 2005). The comparison of the reduced model (2.14a) and (2.14b) with the full model (2.13a)–(2.13d) results shows excellent agreement for moderate values of $\varepsilon=0.1, 0.5$ and still good agreement for $\varepsilon=1.0$ (figures 1 and 2). Especially, the non-Gaussian features are reproduced with high accuracy for $\varepsilon=0.1$ and also $\varepsilon=0.5$, even though slightly less well. For large values of $\varepsilon$, the reduced dynamics get the sign of the non-Gaussian right. All of the reduced models in (2.11)–(2.14a) and (2.14b) cannot be approximated by the interesting regression strategy of Kravtsov et al. (2005) because there is nonlinear multiplicative noise in (2.13a)–(2.13d), (2.14a) and (2.14b), nonlinear triad interaction in (2.13a)–(2.13d) and augmented cubic nonlinearity in (2.14a) and (2.14b). Thus, these regression strategies necessarily have large model error in this example.

(f) A mathematical framework for the mean tendency equation as a dynamical diagnostic

In this section, we provide a general framework to estimate the origin of nonlinear signatures of planetary wave dynamics and how important the observed deviations from Gaussianity are for the planetary waves (Franzke et al. 2007). This general framework diagnoses contributions to the mean values of state-dependent tendencies in a low-dimensional subspace of complex geophysical systems. The mean phase-space tendencies in some GCMs show distinct nonlinear signatures in certain planes, which are spanned by its leading EOFs (Selten & Branstator 2004; Branstator & Berner 2005; Franzke et al. 2007) that also show
only weak deviations from Gaussianity; mostly in the form of weak skewness and kurtosis and in the case of joint pdfs multiple radial ridges of enhanced density (Franzke & Majda 2006; Berner & Branstator 2007). This mathematical framework can be applied to complex geophysical systems and reveals how much the self-interaction among the modes spanning those planes (climate modes) and how much the unresolved modes contribute to these mean phase-space tendencies and also the effect of the observed small deviations from Gaussianity.

To derive the mean tendency equation for the resolved modes, we split the state vector of a quadratic dynamical system into resolved modes $\alpha$ and unresolved modes $\beta$ and use conditional mean probability density relations. Note that while there is a similarity in motivation in decomposing the system into the resolved and unresolved modes compared with the MTV strategy, where the

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{acorr.pdf}
\caption{(a–c(i)) Autocorrelation function and (a–c(ii)) third-order moment for different values of $\epsilon$. (a) $\epsilon = 1.0$, (b) $\epsilon = 0.5$, and (c) $\epsilon = 0.1$. Solid lines, full dynamics (2.13a)–(2.13d) and dashed lines, reduced dynamics (2.14a) and (2.14b).}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{jpdf.pdf}
\caption{Joint pdf of the slow modes $x_1$ and $x_2$ for $\epsilon = 0.1$. (a) Full dynamics (2.13a)–(2.13d) and (b) reduced dynamics (2.14a) and (2.14b).}
\end{figure}
system is decomposed into the slow and fast modes, the mean tendency equation is more general such that it applies also to systems without time-scale separation. The general formula for the conditional mean tendencies for the dynamics in the resolved variables is derived in Franzke et al. (2007) and is given by

\[
\left\langle \frac{\partial \alpha_i}{\partial t} | \alpha \right\rangle = F_i + \sum_{j \in I_l} L_{ij} \alpha_j + \sum_{j, k \in I_l} B_{ijk} \alpha_j \alpha_k, \tag{2.15a}
\]

\[
+ \sum_{j \in I_u} L_{ij} \langle \beta_j | \alpha \rangle, \tag{2.15b}
\]

\[
+ 2 \sum_{j \in I_l \cap k \in I_l} B_{ijk} \alpha_j \langle \beta_k | \alpha \rangle, \tag{2.15c}
\]

\[
+ \sum_{j, k \in I_l} B_{ijk} \langle \beta_j \beta_k | \alpha \rangle. \tag{2.15d}
\]

Note that the r.h.s. of (2.15a) is the bare truncation restricted to interactions among the resolved modes, while (2.15b) and (2.15c) involve all of the conditional mean statistics \( \langle \beta_j | \alpha \rangle \); the terms of (2.15c) are associated with multiplicative triad interactions (leading to multiplicative noise in an MTV (1999, 2003) framework of \( \beta_j \)); the terms in (2.15d) involve all of the conditional interaction statistics \( \langle \beta_j \beta_k | \alpha \rangle \) of second moments and include all of the additive triad interactions (leading to additive noise in an MTV framework) of \( \beta_j \) and \( \beta_k \). Thus, equation (2.15a)–(2.15d) provides a general framework to investigate the mean phase-space tendencies, which allows the decomposition into contributions from interactions among the resolved planetary waves themselves and various contributions involving unresolved d.f.

Now we simplify the above conditional mean tendency equation for purely Gaussian EOF modes. The pdfs of EOFs of GCM data are nearly Gaussian (Hsu & Zwiers 2001; Franzke et al. 2005; Franzke & Majda 2006; Majda et al. 2006a; Berner & Branstator 2007) and there are many geophysical models without damping and forcing that exactly satisfy these assumptions such as barotropic flow on the sphere with topography (Carnevale & Frederiksen 1987; Salmon 1998; Majda & Wang 2006); thus, it is reasonable as a starting point to assume that the pdf is exactly Gaussian. Thus, in the EOF basis, the pdf factors like

\[
p(\alpha, \beta) = \left( \Pi p_i^G(\alpha_i) \right) \left( \Pi p_j^G(\beta_j) \right), \tag{2.16}
\]

where \( p_i^G(\alpha_i) \) and \( p_j^G(\beta_j) \) are Gaussian distributions with mean zero. Thus, in the Gaussian case, the conditional mean tendency equation simplifies to (see Franzke et al. (2007) for more details)

\[
\left\langle \frac{\partial \alpha_i}{\partial t} | \alpha \right\rangle = F_i + \sum_j L_{ij} \alpha_j + \sum_{j, k} B_{ijk} \alpha_j \alpha_k \tag{2.17a}
\]

\[
+ \sum_j B_{ij} \lambda_j. \tag{2.17b}
\]

Note that contributions from (2.15b) and (2.15c), that is, linear coupling and multiplicative triad interactions, are identically zero. These terms vanish because the modes are uncorrelated in the EOF basis. Thus, in this Gaussian case, the conditional mean tendency equation recognizes bare truncation (2.15a) and a constant forcing.
from additive triad interactions $(2.15d)$. Since the leading EOFs of GCM data have 
pdfs with only small, but significant departures from Gaussianity the behaviour in 
$(2.16)$, $(2.17a)$ and $(2.17b)$ serves as a ‘null hypothesis’ for these deviations from 
Gaussianity in the climate models.

This general framework predicts that in the case of purely Gaussian modes the 
nonlinear signatures are stemming from the bare truncation (i.e. the self-
interaction among the planetary waves resolved in the low-dimensional plane). In 
Franzke et al. (2007), these diagnostics were applied to a plane of two low-
frequency EOFs in a three-layer climate model with a nonlinear double swirl (see 
figure 5a in Franzke et al. 2007); the origin of this double swirl is primarily the 
three contributions in $(2.15b)$–$(2.15d)$ from the unresolved modes and not 
nonlinear effects from the bare truncation. The EOFs of the three-layer model 
have small but significant departures from Gaussianity (see figure 9 in Franzke 
et al. 2007). Therefore, these deviations account for the contribution of the 
unresolved onto the resolved modes in the mean phase-space tendencies. Thus, 
these results show that small deviations from Gaussianity have a substantial 
impact on the dynamics of the leading EOFs.

3. Coarse-grained stochastic lattice models for climate: 
tropical convection

The current practical models for prediction of both weather and climate involve 
GCMs where the physical equations for these extremely complex flows are 
discretized in space and time, and the effects of unresolved processes are 
parametrized according to various recipes. With the current generation of 
supercomputers, the smallest possible mesh spacings are approximately 
10–50 km for short-term weather simulations and of order 100 km for short-
term climate simulations. There are many important physical processes that are 
unresolved in such simulations such as the mesoscale sea ice cover, the cloud 
cover in subtropical boundary layers and deep convective clouds in the tropics. 
Most of these features are highly intermittent in space and time. An appealing 
way to represent these unresolved features is through a suitable coarse-grained 
stochastic model that simultaneously retains crucial physical features of the 
interaction between the unresolved and resolved scales in a GCM. In work from 
2002 and 2003, two of the authors have developed a new systematic stochastic 
strategy (Majda & Khouider 2002; Khouider et al. 2003) to parametrize key 
features of deep convection in the tropics involving suitable stochastic spin-flip 
models and also a systematic mathematical strategy to coarse grain such 
microscopic stochastic models to practical mesoscopic meshes in a computa-
tionally efficient manner while retaining crucial physical properties of the interaction.

As regards tropical convection, crucial scientific issues involve the fashion in 
which a stochastic model effects the climate mean state and the strength and 
nature of fluctuations about the climate mean. Here the strategy to develop a 
new family of coarse-grained stochastic models for tropical deep convection is 
briefly reviewed (Majda & Khouider 2002; Khouider et al. 2003) as an illustrative 
example of the potential use of stochastic lattice models. In Khouider et al. 
(2003), it has been established that in suitable regimes of parameters, the coarse-
grained stochastic parametrizations can significantly alter the climatology as well.
as increase wave fluctuations about the climatology. This was established in Khouider et al. (2003) in the simplest scenario for tropical climate involving the Walker circulation, the east–west climatological state that arises from local region of enhanced surface heat flux, mimicking the Indonesian marine continent. Convectively coupled waves in the tropics such as the Madden–Julian oscillation play an important role in medium-range forecasts, yet the current generation of computer models fails to represent such waves adequately (Lin et al. 2006). Palmer (2001) has emphasized the potential of stochastic parametrization to reduce the model error in a deterministic computer model. Here in an idealized setting, we show how to develop a stochastic parametrization to modify and improve the behaviour of convectively coupled waves in a reasonable prototype GCM; this is achieved by following a path guided by the systematic design principle for the idealized model in (1.1) (KMS 2006, 2007) to build in suitable intermittency effects.

(a) The microscopic stochastic model for convective inhibition

In a typical GCM, the fluid dynamical and thermodynamical variables denoted here by the generic vector $u$ are regarded as known only over a discrete horizontal mesh with $u_j = u(j\Delta x, t)$ denoting these discrete values. Throughout the discussion, one horizontal spatial dimension along the equator in the east–west direction is assumed for simplicity in notation and explanation. As mentioned above, the typical mesh spacing in a GCM is coarse with $\Delta x$ ranging from 50 to 250 km depending on the time duration of the simulation. The stochastic variable used to illustrate the approach is convective inhibition (CIN). Observationally, CIN is known to have significant fluctuations on a horizontal spatial scale on the order of a kilometre, the microscopic scale here, with changes in CIN attributed to different mechanisms in the turbulent boundary layer, such as gust fronts, gravity waves and turbulent fluctuations in equivalent potential temperature (Mapes 2000). In Majda & Khouider (2002) and Khouider et al. (2003), it was proposed that all of these different microscopic physical mechanisms and their interaction that increase and decrease CIN are too complex to model in detail in a coarse-mesh GCM parametrization and instead, as in statistical mechanics, should be modelled by a simple order parameter $\sigma_1$, taking only two discrete values,

\[
\begin{align*}
\sigma_1 &= 1 \quad \text{at a site if convection is inhibited (a CIN site)} \\
\sigma_1 &= 0 \quad \text{at a site if there is potential for deep convection (a PAC site).}
\end{align*}
\]

(3.1)

The value of CIN at a given coarse-mesh point is determined by the averaging of CIN over the microscopic states in the vicinity of the given mesh point, i.e.

\[
\bar{\sigma}_1(j\Delta x, t) = \frac{1}{\Delta x} \int_{(j-1/2)\Delta x}^{(j+1/2)\Delta x} \sigma_1(x, t) \, dx.
\]

(3.2)

Note that the mesh size $\Delta x$ is mesoscopic, that is, between the microscale $O(1 \, \text{km})$ and the macroscale $O(10,000 \, \text{km})$, and that $\bar{\sigma}_1$ can have any value in the range $0 \leq \bar{\sigma}_1 \leq 1$. Discrete sums over microscopic mesh values (of order 1 km) and continuous integrals are used interchangeably for notational convenience (Majda & Khouider 2002).
(b) The simplest coarse-grained stochastic model

In practical parametrization, it is desirable for computational feasibility to replace the microscopic dynamics by a process on the coarse mesh, which retains critical dynamical features of the interaction. Following the general procedure developed and tested in Katsoulakis & Vlachos (2003) and Katsoulakis et al. (2003a,b), the simplest local version of the systematic coarse-grained stochastic process is developed in Khouider et al. (2003) and summarized here.

Each coarse cell $\Delta x_k$, $k=1, \ldots, m$, of the coarse-grained lattice is divided into $q$ microscopic cells, such that $\Delta x_k \equiv 1/q \{1, 2, \ldots, q\}$, $k=1, \ldots, m$. In the coarse-grained procedure, given the coarse-grained sequence of random variables

$$
\eta_i(k) = \sum_{y \in \Delta x_k} \sigma_{1,i}(y),
$$

so that the average in (3.2) verifies $\bar{\sigma}_i(j \Delta x) = \eta(k)/q$, for $j=k$ in some sense, the microscopic dynamics is replaced by a birth/death Markov process defined on the variables $\{0, 1, \ldots, q\}$ for each $k$, such that $\eta_i(k)$ evolves according to the following probability law:

\[
\begin{align*}
\text{Prob} \{ \eta_{t+\Delta t}(k) = n+1 | \eta_t(k) = n \} &= C_a(k,n) \Delta t + o(\Delta t) \\
\text{Prob} \{ \eta_{t+\Delta t}(k) = n-1 | \eta_t(k) = n \} &= C_d(k,n) \Delta t + o(\Delta t) \\
\text{Prob} \{ \eta_{t+\Delta t}(k) = n | \eta_t(k) = n \} &= 1 - (C_a(k,n) + C_d(k,n)) \Delta t + o(\Delta t) \\
\text{Prob} \{ \eta_{t+\Delta t}(k) \neq n, n-1, n+1 | \eta_t(k) = n \} &= o(\Delta t).
\end{align*}
\]

Here $C_a$ and $C_d$ are the coarse-grained adsorption and desorption rates representing the rates at which CIN sites are being created and destroyed, respectively, within one coarse-grained cell, that is, the birth/death rates for the Markov process is

\[
\begin{align*}
C_a(k, \eta) &= \frac{1}{\tau_1} [q - \eta(k)], \\
C_d(k, \eta) &= \frac{1}{\tau_1} \eta(k) e^{-\beta \bar{V}(k)},
\end{align*}
\]

where

\[
\bar{V}(k) = \bar{J}(0,0)(\eta(k) - 1) + h_{\text{ext}},
\]

with the coarse-grained interaction potential within the coarse cell given by $\bar{J}(0,0) = 2U_0/(q-1)$, where $U_0$ is the mean strength of the microscopic interaction potential $J$ between neighbouring sites (Katsoulakis et al. 2003a,b). The coarse-grained energy content for CIN is given by the coarse-grained Hamiltonian

\[
\bar{H}(\eta) = \frac{U_0}{q-1} \sum_k \eta(k)(\eta(k) - 1) + h_{\text{ext}} \sum_k \eta(k),
\]

where $h_{\text{ext}}$ is the external potential that modifies the energy for CIN depending on the large-scale flow variables. The canonical invariant Gibbs measure for the coarse-grained stochastic process is a product measure given by

\[
G_{m,q,\beta}(\eta) = (Z_{m,q,\beta})^{-1} e^{\beta \bar{H}(\eta)} P_{m,q}(d\eta),
\]

where $P_{m,q}(d\eta)$ is a given explicit/prior distribution (Katsoulakis et al. 2003b). As shown in Katsoulakis et al. (2003b), the coarse-grained birth/death process above satisfies detailed balance with respect to the Gibbs measure in (3.8) as well as a
number of other attractive theoretical features. The simplest coarse-grained approximation given above assumes that the effect of the microscopic interactions on the mesoscopic scales occurs within the mesoscopic coarse-mesh scale $\Delta x$, otherwise systematic non-local couplings are needed (Katsoulakis et al. 2003b). The accuracy of these approximations is tested for diverse examples from material science elsewhere (Katsoulakis & Vlachos 2003; Katsoulakis et al. 2003a,b) and for the instructive idealized coupled models in (1.1) (KMS 2004, 2005, 2006, 2007).

The practical implementation of the coarse-grained birth/death process in (3.3)–(3.6) requires specification of the parameters, $\tau_l$, $U_0$, $q$ and the external potential $h_{\text{ext}}(u_j)$ as well as the statistical parameter $\beta$. The advantages of such a stochastic lattice model are as follows:

— retains systematically the energetics of unresolved features through the coarse-grained Gibbs’s measure;
— has minimal computational overhead since there are rapid algorithms for updating birth–death processes;
— incorporates feedbacks of the resolved modes on the unresolved modes and their energetics through an external field; and
— includes dynamical coupling through not only sampling the probability distributions of unresolved variables but also their evolving behaviour in time is constrained by the large-scale dynamics.

(c) The model deterministic convective parametrization

A prototype mass flux parametrization with crude vertical resolution (Majda & Shefter 2001; Khouider & Majda 2006b) is used to illustrate the fashion in which the coarse-grained stochastic model for CIN can be coupled to a deterministic convective mass flux parametrization. The prognostic variables $(u, \theta, \theta_{\text{eb}}, \theta_{\text{em}})$ are the $x$-component of the fluid velocity, $u$, the potential temperature in the middle troposphere, $\theta$, the equivalent potential temperatures, $\theta_{\text{eb}}$ and $\theta_{\text{em}}$, measuring the potential temperatures plus moisture content of the boundary layer and middle troposphere, respectively. The vertical structure is determined by projection on a first baroclinic heating mode (Majda & Shefter 2001; Khouider & Majda 2006b). The dynamic equations for these variables in the parametrization are given by

\[
\begin{align*}
\frac{\partial u}{\partial t} - \bar{\alpha} \frac{\partial \theta}{\partial x} &= - \left( C_D^0 \frac{1}{h} \sqrt{\frac{u_0^2 + u^2}{}} \right) u - \frac{1}{\tau_D} u, \\
\frac{\partial \theta}{\partial t} - \bar{\alpha} \frac{\partial u}{\partial x} &= S - Q_R^0 - \frac{\theta}{\tau_R}, \\
\tau \frac{\partial \theta_{\text{eb}}}{\partial t} &= -D(\theta_{\text{eb}} - \theta_{\text{em}}) + \left( C_\theta \sqrt{\frac{u_0^2 + u^2}{}} \right) (\theta_{\text{eb}}^* - \theta_{\text{eb}}), \\
H \frac{\partial \theta_{\text{em}}}{\partial t} &= D(\theta_{\text{eb}} - \theta_{\text{em}}) - HQ_R^0 - H \frac{\theta_{\text{em}}}{\tau_R},
\end{align*}
\]

(3.9)

Phil. Trans. R. Soc. A (2008)
while the constants $Q^0_R$ and $\theta^c_{eb}$ are externally imposed and represent the radiative cooling at equilibrium in the upper troposphere and saturation equivalent potential temperature in the boundary layer. The constants $h$ and $H$ measure the depths of the boundary layer and the troposphere above the boundary layer, respectively. The typical values used here are $h = 500 \text{ m}$ and $H = 16 \text{ km}$, while $u_0 = 2 \text{ m s}^{-1}$ and $C^0_D = 0.001$, corresponding to a momentum spin up time of approximately 3 days. The explicit values for the other constants used in (3.9) and elsewhere in this section can be found in Majda & Shefter (2001) and Khouider et al. (2003).

The crucial quantities in the prototype mass flux parametrization are the terms $S$ and $D$, where $S$ represents the middle troposphere heating due to deep convection, while $D$ represents the downward mass flux on the boundary layer. The heating term $S$ is given by

\begin{equation}
S = M \sigma_c (\text{CAPE})^{1/2},
\end{equation}

where $M$ is a fixed constant; $\sigma_c$ is the area fraction for deep convective mass flux; and \( \text{CAPE} = R(\theta_{eb} - \gamma \theta) \) is the convectively available potential energy. Here $R$ is a dimensional constant (Majda & Shefter 2001; Khouider et al. 2003). The downward mass flux on the boundary layer $D$ includes the environmental downdrafts $m_e$ and the downward mass flux due to convection $m_K$, which are non-negative quantities with explicit formulae described elsewhere (Majda & Shefter 2001; Khouider et al. 2003). The notation $(X)^+$ denotes the positive part of $X$. This parametrization respects conservation of vertically integrated moist static energy.

The equations (3.9) and (3.10) represent an idealized GCM with crude vertical resolution based on a reasonable design principle for deep convection including basic conservation principles. However, like the current generation of GCMs, the model has major deficiencies as regards convectively coupled waves. First, instabilities for the full model in (3.9) arise only through the nonlinearity from $\sqrt{u_0^2 + v^2}$ and are called the wind-induced surface heat exchange (WISHE) instability (Majda & Shefter 2001, and references therein). Second, if the bulk aerodynamic flux coefficients in (3.9), $\sqrt{u_0^2 + v^2}$ are replaced by the constant value $|u_0|$, that is no WISHE, all waves in the system are stable and no fluctuations emerge in an aquaplanet simulation (Majda & Shefter 2001). There is no observational evidence supporting the role of WISHE in driving the large-scale convectively coupled waves in nature (Lin et al. 2006 and references therein); here the WISHE term is regarded as a deterministic fix in a GCM parametrization to generate instabilities. The simulation results reported in figure 5a show that in an aquaplanet model above the equator, two regular periodic convectively coupled waves moving eastward and westward at roughly equal strength are generated by the idealized GCM. Since GCMs often have convectively coupled waves that move too fast and are too regular (Lin et al. 2006), the goal here is to see whether the stochastic lattice coupling will slow down the waves in the deterministic parametrization and simultaneously increase the spatio-temporal fluctuations in those waves. It is important to note here that there are recent deterministic multicloud models (Khouider & Majda 2006a, 2007, 2008) for convectively coupled waves involving the three cloud...
types in observations above the boundary layer, congestus, deep and stratiform, and their heating structure that reproduces key features of the observational record for convectively coupled waves (Lin et al. 2006). The mechanism of instability in these models (Khouider & Majda 2006a) is completely different from WISHE that is not active in the multicloud models. For the idealized setting of flow above the equator, the multicloud models can produce packets of convectively coupled waves moving in one direction at 15–20 m s\(^{-1}\) with their low-frequency envelopes moving at 4–7 m s\(^{-1}\) in the opposite direction across the warm pool in a fashion like the Madden–Julian oscillation (Khouider & Majda 2007, 2008).

(d) Coupling of the stochastic CIN model into the parametrization

The equations (3.9) and (3.10) are regarded here as the prototype deterministic GCM parametrization when discretized in a standard fashion using central differences on a coarse mesh \(\Delta x\) with \(\Delta x\) ranging from 50 to 250 km. In the simulations from Khouider et al. (2003), and those presented below, \(\Delta x\)=80 km. The coarse-grained stochastic CIN model is coupled to this basic parametrization. The area fraction for deep convection \(\sigma_c\) governing the upward mass flux strength is allowed to vary on the coarse mesh and is given by

\[
\sigma_c(j\Delta x) = [1-\bar{\sigma}_I(j\Delta x)]\sigma_c^+, \tag{3.11}
\]

where \(\bar{\sigma}_I\) is the average in (3.2) and \(\sigma_c^+\) is a threshold constant and equal to 0.002 (Majda & Shefter 2001; Khouider et al. 2003). When the order parameter \(\sigma_I\) signifies strong CIN locally so that \(\bar{\sigma}_I=1\) the flux of deep convection is diminished to zero while with PAC locally active \(\bar{\sigma}_I=0\), this flux increases to the maximum allowed by the value \(\sigma_c^+\). To complete the coupling of the stochastic CIN model into the parametrization, the coarse-mesh external potential \(h_{\text{ext}}(u_j)\), from (3.6) and (3.7), needs to be specified from the coarse-mesh values \(u_j\). There is no unique choice of the external potential but its form can be dictated by simple physical reasoning. In Khouider et al. (2003), the plausible physical assumption is made that when the convective downward mass flux \(m_-\) decreases the energy for CIN decreases. Since the convective downward mass flux results from the evaporative cooling induced by precipitation falling into dry air, it constitutes a mechanism that carries negatively buoyant cool and dry air from the middle troposphere onto the boundary layer, hence tending to reduce CAPE and deep convection.

Another natural external potential is the boundary-layer equivalent potential, since the flux at the boundary is crucial physically. As such, it yields somewhat stronger and more intermittent fluctuations. Thus, the choice

\[
h_{\text{ext}}(j\Delta x, t) = \tilde{\gamma}\theta_{eb}(j\Delta x, t) \tag{3.12}
\]

is used here with \(\tilde{\gamma}\) a calibration factor. The other parameters in the stochastic lattice model are chosen as \(\tau_I=2\) hours, \(\beta=1\), \(U_0=1\) so that CIN sites are favoured in the equilibrium Gibbs measure.

(e) The stochastic single-column model and intermittency

A central issue is how to calibrate the stochastic lattice model to generate intermittent fluctuations with plausible magnitudes as observed in tropical
convection. A natural design framework is first to achieve such behaviour in the stochastic single-column model given by the following equations:

\[
\begin{align*}
\frac{d\theta}{dt} &= (1 - \sigma_1) M \sqrt{R(\theta_{eb} - \gamma\theta)^+ - Q_R^0} - \frac{1}{\tau_R} \theta, \\
\frac{d\theta_{eb}}{dt} &= -\frac{1}{\tau_{eb}} (1 - \sigma_1) \sqrt{R(\theta_{eb} - \gamma\theta)^+ (\theta_{eb} - \theta_{em})} + \frac{1}{\tau_e} (\theta_{eb} - \theta_{em}), \\
\frac{d\theta_{em}}{dt} &= \frac{1}{\tau_{em}} (1 - \sigma_1) \sqrt{R(\theta_{eb} - \gamma\theta)^+ (\theta_{eb} - \theta_{em})} - \frac{1}{\tau_R} \theta,
\end{align*}
\]

Here \(\sigma_1\) is computed from the birth–death stochastic model in (3.3)–(3.6) with the interaction potential defined through \(\theta_{eb}\) from (3.7) and the definition of \(h_{ext}\) in (3.12) above. The different time scales were computed for a radiative convective equilibrium consistent with the Jordan tropical sounding and the values \(\tau_{eb} \approx 6\) hours, \(\tau_{em} \approx 8\) days, \(\tau_e \approx 8\) hours, \(Q_R^0 = 1\) K d\(^{-1}\), \(\tau_R = 50\) days and \(\tau_D = 75\) days are used here (Khouider et al. 2003). This is a specific example of the prototype models from (1.1) studied extensively in idealized settings (KMS 2004, 2005, 2006). As mentioned earlier, the choices \(\beta = 1, U_0 = 1\) guarantee that the Gibbs measure in (3.8) has a high probability for CIN states; this is a natural requirement for deep convection, where the area fraction of deep convection is much smaller than the area with positive CAPE. In order to generate intermittency and fluctuations in the stochastic column model, it is very natural to require that the ratio of the creation rate of PAC sites to the creation rate of CIN sites, \(C_d / C_a\), becomes large as \(\sigma_1 \gg 1\) (KMS 2006, 2007). In figure 3, it is shown that this property of the stochastic model is satisfied for the coefficient \(\tilde{\gamma} = 0.1\) in the external potential in (3.12) with similar behaviour for \(q = 10\),
100 and 1000. Recall that $q$ is the number of microscopic states in the birth–
death process. A time series of the stochastic column model implemented with
these parameters is displayed in figure 4 with $q=10$. The strongly intermittent
fluctuations in $\theta_{eb}$ over several degrees Kelvin as well as similar intermittency
in the mass flux is evident. They are consistent with SST variations observed in the
Indian Ocean warm pool (e.g. Fasullo & Webster 1999). Also note that the
fluctuations in the mid-troposphere potential temperature are much weaker in
magnitude as occurs in the actual tropics, where middle tropospheric
temperature anomalies are on the order of 0.1–0.5°C (e.g. Wheeler et al. 2000).

\( \text{(f) The effects of the stochastic parametrization on convectively coupled waves} \)

All parameters in the stochastic lattice model have been determined through
systematic design principles in the previous section. Here the results of numerical
simulations of the stochastic model in statistical steady state are reported for

---

flow above the equator in a standard aquaplanet set-up with uniform SST (Majda & Khouider 2002; Khouider et al. 2003). The results of various simulations are reported in figure 5 for comparison with the deterministic parametrization with WISHE reported in figure 5a. As shown in figure 5b, the effect of the fluctuations of the stochastic lattice model is simultaneously to create more realistic intermittency in the convectively coupled waves and to slow down their phase speed from 15 to 11 m s$^{-1}$; as mentioned earlier, these are desirable qualitative features of a stochastic parametrization. In figure 5c, we report the result of running the model parametrization in (3.9) without WISHE but coupled to the stochastic lattice model; recall that this is the situation where the deterministic model without WISHE produces no waves in the statistical steady state. In figure 5c, there is a clear evidence for stochastic-generated convectively coupled waves with the reduced phase speed of roughly 8 m s$^{-1}$ and roughly half the amplitude; these waves have been created by coupling alone to the intermittent stochastic lattice model and without deterministic instability. This is another attractive feature of the present stochastic lattice models in changing the character of model error for convection (Palmer 2001).
4. Concluding remarks

This paper both reviewed and provided new illustrations and examples of the fashion in which modern applied mathematics can provide new perspectives and systematic design principles for stochastic modelling for climate. Section 2 was devoted to systematic stochastic modelling of low-frequency variability including the quantitative sources for multiplicative noise. A new simplified low-dimensional stochastic model with key features of atmospheric low-frequency variability was introduced in §2e in order to test stochastic mode reduction strategies. A recent diagnostic test with firm mathematical underpinning for exploring the subtle departures from Gaussianity and their sources was discussed in §2f. Section 3 was devoted to developing stochastic lattice models to capture intermittent features and improve the fidelity of deterministic parametrizations. Recent systematic design principles (KMS 2006, 2007) were used to calibrate a stochastic column model for tropical convection in §3e; the practical effect of slowing down convectively coupled waves and increasing their fluctuations through these stochastic lattice models was presented in §3f.

The authors thank their collaborators, Eric Vanden-Eijnden, Ilya Timofeyev, Markos Katsoulakis and Alex Sopasakis, for their explicit and implicit contributions to the work presented here. The research of A.J.M. is partially funded by the National Science Foundation grant no. DMS-0456713, the Office of Naval Research grant no. N00014-05-1-0164 and the Defense Advanced Research Project Agency grant no. N00014-07-1-0750. The research of B.K. is partly supported by a grant from the National Sciences and Engineering Research Council of Canada.

References


