# Improving Complex Models Through Stochastic Parameterization and Information Theory

**Andrew J. Majda, Michał Branicki and Yevgeniy Frenkel**

*Department of Mathematics and Center for Atmosphere-Ocean Science*
*Courant Institute, New York University, New York, New York*

**ABSTRACT**

This paper has a brief discussion of three topics which are central themes of this ECMWF-Thorpex-Clivar workshop and also are major research themes of the first author and several collaborators. The three topics discussed briefly here are the following: 1) Improving tropical convective parameterization through the stochastic multi-cloud model ; 2) Improving model fidelity and long-range forecasting skill through empirical information theory and stochastic parameterization; 3) Judicious model errors in filtering turbulent dynamical systems: Stochastic Parameterization Extended Kalman Filter (SPEKF)

## 1 Improving tropical convective parameterization through the stochastic multi-cloud model

Despite recent advances in supercomputing, current general circulation models (GCMs) poorly represent the variability associated with organized tropical convection. A stochastic multicloud convective parameterization based on three cloud types (congestus, deep, and stratiform), introduced in 2010 by Khouider, Biello and Majda (hereafter KBM10) in the context of a single column model, has been used recently to study flows above the equator without rotation effects (Frenkel et al., 2011). The stochastic model dramatically improves the variability of tropical convection compared to the conventional moderate and coarse resolution paradigm GCM parameterizations. This increase in variability comes from intermittent coherent structures such as synoptic and mesoscale convective systems, analogs of squall lines and convectively coupled waves seen in nature whose representation is improved by the stochastic parameterization. Furthermore, simulations with sea surface temperature (SST) gradient yield realistic mean Walker-cell circulation with plausible high variability. An additional feature of the new stochastic parameterization is a natural scaling of the model from moderate to coarse grids which preserves the variability and statistical structure of the coherent features. These results systematically illustrate, in a paradigm model, the benefits of using the stochastic multi-cloud framework to improve deterministic parameterizations with clear deficiencies.

The search for methods of adequately addressing the interactions across temporal and spatial scales between the large scale circulation and organized cloud systems, from individual clouds to large-scale clusters and superclusters to planetary-scale disturbance, has not been fruitless. Cloud-resolving models on fine computational grids and high-resolution numerical weather prediction models with improved convective parametrizations succeed in representing some aspects of organized convection (ECMWF, 2003; Moncrieff et al., 2007). However, due to their extremely high computational cost, these methods cannot be applied to large ensemble-size weather prediction or climate simulations. The complexity of the problem has motivated the development of an approach that directly addresses the multiscale nature of the problem. Superparameterization (SP) methods (Grabowski and Smolarkiewicz, 1999; Grabowski, 2001, 2004; Randall et al., 2003; Majda, 2007) use a cloud resolving model (CRM) in each column of the large scale GCM to explicitly represent small scale and mesoscale processes and interactions among them. The computational cost can be further reduced by techniques such as sparse space-time SP (Xing et al., 2009). Nonetheless computationally inexpensive GCM parameterizations that capture the variability and coherent structure of deep convection have remained a central unsolved problem in the atmospheric community.

The most common conventional cumulus parameterizations are based on the quasi-equilibrium (QE) assumption first postulated by Arakawa and Schubert (1974), the moist convective adjustment idea of Manabe et al. (1965), or the large scale moisture convergence closure of Kuo (1974) type. As such, the mean response of unresolved modes on large/resolved scale variables is formulated according to a prescribed deterministic closure. While many recipes for the closures have been created (Kain and Fritsch, 1990; Betts and Miller, 1986; Zhang and McFarlane, 1995), these purely deterministic parameterizations were found to be inadequate for the representation of the highly intermittent and organized tropical convection (Palmer, 2001). Many of the improvements in GCMs of the last decade came from the relaxation of the QE assumption, for example, through the addition of a stochastic perturbation. Buizza et al. (1999) used a stochastic backscattering model to represent the model uncertainties in a GCM while Lin and Neelin (2003) used a stochastic parametrization to randomize the way in which deep convection responds to large fluctuations via a prescribed probability distribution function for the convective time scale. Majda and Khouider (2002) were the first to propose a stochastic model for convective inhibition (CIN), that allows both internal interactions between convective elements and two-way interactions between the convective elements and the large scale/resolved variables. Their model is based on an Ising-type spin-flip model used as a model for phase transitions in material science (Katsoulakis et al., 2003). When coupled to a toy GCM, this stochastic parameterization produced eastward propagating convectively coupled waves that qualitatively resemble observations (Khouider et al., 2003; Majda et al., 2008) despite the extreme simplicity of the model and deficiency of the underlying convective parameterization. Stochastic processes have been used to parameterize convective momentum transport (Majda and Stechmann, 2008), to improve conceptual understanding of the transition to deep convection through critical values of column water vapor (Stechmann and Neelin, 2010), as well as for the analysis of cloud cover data in the tropics and the extratropics (Horenko, 2010).

The stochastic multicloud model for tropical convection introduced by Khouider et al. (2010) (hereafter KBM10) is a novel approach to the problem of missing tropical variability in GCMs. The stochastic parameterization is based on a Markov chain lattice model where each lattice site is either occupied by a cloud of a certain type (congestus, deep or stratiform) or it is a clear sky site. The convective elements interact with the large scale environment and with each other through convectively available potential energy (CAPE) and middle troposphere dryness. When local interactions between the lattice sites are ignored, a coarse grained stochastic process that is intermediate between the microscopic dynamics and the mean field equations (Katsoulakis et al., 2003; Khouider et al., 2003; Majda et al., 2008) is derived for the dynamical evolution of the cloud area fractions. Besides deep convection, the stochastic multicloud model includes both low-level moisture preconditioning through congestus clouds and the direct effect of stratiform clouds including downdrafts which cool and dry the boundary layer. The design principles of the multicloud parameterization framework are extensively explored in the deterministic version of the model developed by Khouider and Majda (2006a; 2006b; 2007; 2008a; 2008b, hereafter KM06a, KM06b,KM07, KM08a, KM08b, respectively).

### 1.1 Dynamical core of the multicloud model

The multicloud parametrization framework assumes three heating profiles associated with the main cloud types that characterize organized tropical convective systems (Johnson et al., 1999): cumulus congestus clouds that heat the lower troposphere and cool the upper troposphere, through radiation and detrainment, deep convective towers that heat the whole tropospheric depth, and the associated lagging-stratiform anvils that heat the upper troposphere and cool the lower troposphere, due to evaporation of stratiform rain. Accordingly, Khouider and Majda (e.g. KM06a, KM08a) used the momentum and potential temperature equations for the first and second baroclinic modes of vertical structure, that are directly forced by deep convection and both congestus and stratiform clouds, respectively, as a minimal dynamical core that captures the main (linear response) effects of these three cloud types. Versions of this simple modeling framework that include effects of convective momentum transport (CMT) are found in Majda and Stechmann (2008, 2009) and Khouider et al. (2011). The multicloud model also carries equations for the vertically averaged moisture (water vapor mixing ratio), over the tropospheric depth, and bulk boundary layer dynamics averaged over the atmospheric boundary layer (ABL).

## 1.2 The stochastic multicloud model

The stochastic multicloud parameterization is designed to capture the dynamical interactions between the three cloud types that characterize organized tropical convection and the environment using a coarse grained lattice model (KBM10). To mimic the behavior within a typical GCM grid box, a rectangular $n$ x $n$ lattice is considered, where each element can be either occupied by a congestus, deep or a stratiform cloud or is clear sky, through an order parameter that takes values of 0,1,2 or 3 on each lattice site. A continuous time stochastic process is then defined by allowing the transitions, for individual cloud sites, from one state to another according to intuitive probability transition rates, which depend on the large scale-resolved variables. These large scale variables are the convectively available potential energy integrated over the whole troposphere (CAPE), the convectively available energy integrated over the lower troposphere CAPE$_l$, and the dryness of the mid troposphere, which is a function of the difference between the atmospheric boundary layer (ABL) temperature $\theta_{eb}$ and the middle tropospheric potential temperature $\theta_{em} = q + (2\sqrt{2}/\pi)(\theta_1 + \alpha_2\theta_2)$. The inclusion of dryness of the middle troposphere accounts for mixing of the convective parcels with dry environmental air (KM06a, KM06b,KM07, KM08a, KM08b,KBM10).

The probability rates are constrained by a set of intuitive rules which are based on observations of cloud dynamics in the tropics (e.g, Johnson et al., 1999; Mapes, 2000, Khouider and Majda 2006, and references therein). Following KBM10, a clear site turns into a congestus site with high probability if low level CAPE is positive and the middle troposphere is dry. A congestus or clear sky site turns into a deep convective site with high probability if CAPE is positive and the middle troposphere is moist. A deep convective site turns into a stratiform site with high probability. Finally, all three cloud types decay naturally to clear sky at some fixed rate. All other transitions are assumed to have negligible probability. These rules are formalized in (Frenkel et al., 2011)(hereafter FMK2011) in terms of the transition rates $R_{ik}$ and the associated time scales $\tau_{ik}$. Notice that the assumption that the transition rates depend on the large scale variables accounts for the feedback of the large scales on the stochastic model, while ignoring the interactions between the lattice sites all together implies that the stochastic processes associated with the different sites are identical. The latter simplification makes it easy to derive the stochastic dynamics for the GCM grid box cloud coverages alone, which can be evolved without the detailed knowledge of the micro-state configuration, by using a coarse-graining technique (?Katsoulakis et al., 2003b, Khouider et al. 2003) that yields here a system of three birth-death-like processes, corresponding to the three cloud types. The resulting birth-death Markov system is easily evolved in time using Gillespies exact algorithm (Gillespie, 1975, 1977). Thus given the large scale thermodynamic quantities the stochastic process yields the dynamical evolution for the congestus, deep and stratiform cloud fractions $\sigma_c, \sigma_d$ and $\sigma_s$ respectively.

Figure 1 shows the performance of the multicloud model in a column test mode (FMK2011). Note the random intermittent bursts of convection often with phasing of the congestus to deep to stratiform area fractions associated with tropical convection. Figure 2 compares the precipitation profiles produced by the stochastic multi-cloud model with a Walker simulation on 40,000 km domain (FMK2011) with the high resolution CRM of (Grabowski et al., 2000) in similar set-up on a 4,000 km domain(right). The qualitative reproduction of the variability by the much cheaper stochastic multi-cloud model is striking! In addition, the stochastic multi-cloud model produces much clearer intense rain events, associated with convectively coupled waves, outside the warm pool, as observed in nature.

## 2 Improving model fidelity and sensitivity, and long range forecasting skill through empirical information theory and stochastic parameterization

Predicting the long range behavior of complex systems in nature in diverse disciplines ranging from climate change science (Randall et al. (2007); Hwang and Frierson (2010)) to materials (Chatterjee and Vlachos (2007)) and neuroscience (Rangan et al. (2009)) is an issue of central importance in contemporary engineering and science. Accurate predictions are hampered by the fact that the true dynamics of the system in nature are actually unknown due to inadequate scientific understanding or inadequate spatio-temporal resolution in the imperfect computer models used for these predictions; in other words, there are significant model errors compared to the true signal from nature. Recently, information theory has been utilized in different ways to systematically improve model fidelity and sensitivity (Majda and Gershgorin (2010, 2011a)), to quantify the role of coarse-grained initial states in long range forecasting (Giannakis and Majda (2011a,b)), and to make an empirical link between model fidelity and forecasting skill (DelSole (2005); DelSole and Shukla (2010)). Imperfect models for complex systems are constrained by their capability to reproduce certain statistics in a training phase where the natural system has been observed; for example, this training phase in climate science is roughly the sixty year data set of extensive observations of the Earth's climate system. For long range forecasting, it is natural to guarantee statistical equilibrium fidelity for an imperfect model and a framework using information theory is a natural way to achieve this in an unbiased fashion (Majda and Gershgorin (2010, 2011a); DelSole and Shukla (2010); Giannakis and Majda (2011a,b)). First, equilibrium statistical fidelity for an imperfect model depends on the choice of coarse-grained variables utilized (Majda and Gershgorin (2010, 2011a)); secondly, equilibrium model fidelity is a necessary but not sufficient condition to guarantee long range forecasting skill (Giannakis and Majda (2011b)). For example, Section 2.6 of Majda et al. (2005) extensively discusses three very different strongly mixing chaotic dynamical models with forty variables and with the same Gaussian equilibrium measure, the TBH, K-Z, and IL96 models, so that all three models have the same climate equilibrium fidelity but have completely different forecasting skill. On the other hand, there are notable examples where improving equilibrium fidelity results in improved model sensitivity (Majda and Gershgorin (2010)) or intermediate range forecasting skill (DelSole and Shukla (2010)). The central issue addressed recently in Majda and Gershgorin (2011b) is the following one: Is there a systematic way to improve long range forecasting skill of imperfect models satisfying equilibrium fidelity? Are there a systematic set of statistical prediction tests in the training phase beyond equilibrium fidelity which guarantee improved long range forecasting skill for an imperfect model?

The main goal of Majda and Gershgorin (2011a) was to provide such a direct link by utilizing fluctuation dissipation theorems (FDT) for complex dynamical systems (DekerH1975 (1975); Majda et al. (2005); Marconi et al. (2008)) together with the framework of empirical information theory for improving imperfect models developed recently (Majda and Gershgorin (2010, 2011a)). The main link utilizing FDT is developed in Majda and Gershgorin (2011b). This is followed by demonstration of this approach on a suite of mathematical test models which despite their simplicity and mathematical tractability, nevertheless, mimic crucial statistical features of complex systems such as Earth's climate.

A simple example illustrating the fundamental difficulties in climate science in trying to use imperfect models to predict the sensitivity for the perfect model is presented next (Majda and Gershgorin (2011a)).

A typical situation with model error for complex systems arises when the true system has additional

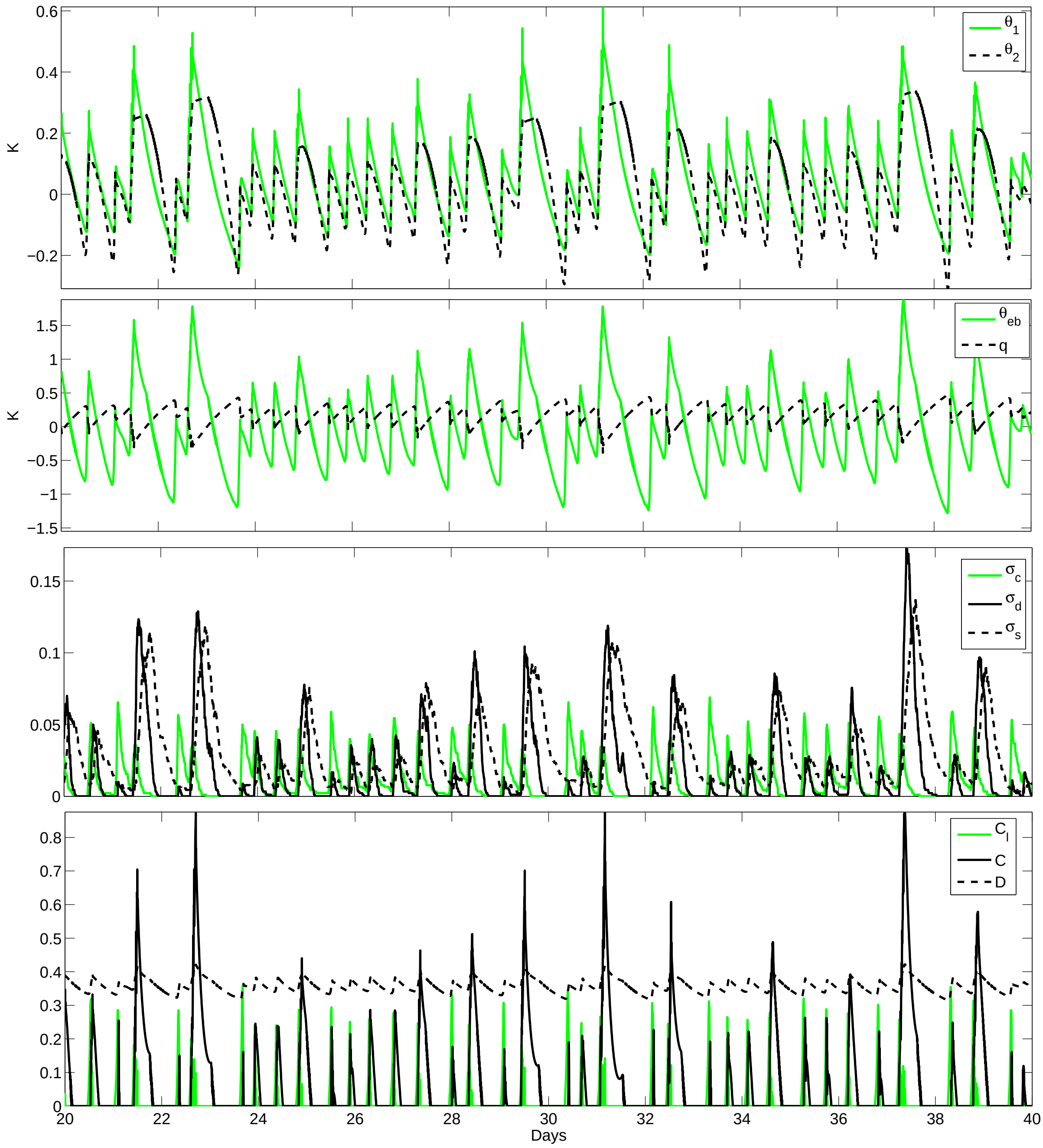


*Figure 1: Time series of the large-scale variables, the (stochastic) cloud area fractions, CAPE and Dryness for the standard parameter of FMK 2011.*

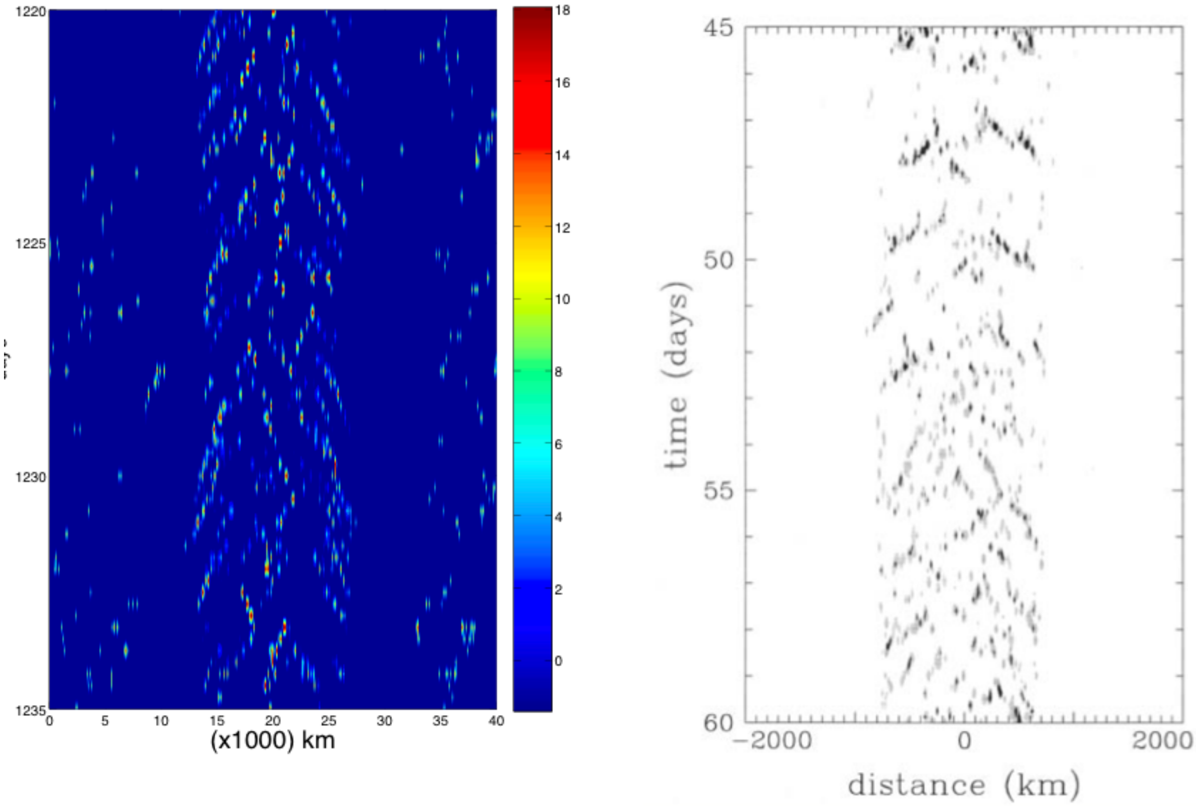


*Figure 2: Precipitation profiles produced by the stochastic multi-cloud model (left) with a Walker simulation on 40,000 km domain (FMK2011) and the high resolution CRM of (Grabowski et al., 2000) in similar set-up on a 4,000 km domain (right).*

degrees of freedom that are hidden from the family of imperfect models utilized to study this system either through lack of scientific understanding or the practical lack of computational resolution. The simplest example with these features is to consider the true system as given by the two linear stochastic equations

$$\begin{aligned}\frac{du}{dt} &= au + v + F,\\ \frac{dv}{dt} &= qu + Av + \sigma\dot{W},\end{aligned} \tag{1}$$

where $\dot{W}$ is white noise; the system of equation in 2.5 has a smooth Gaussian statistical steady state provided that

$$a + A < 0, \quad aA - q > 0. \tag{2}$$

Assume that the variable $v$ in 2.5 is hidden from the modeling process where all imperfect models are given by the scalar stochastic equation

$$\frac{du_M}{dt} = -\gamma_M u_M + F_M + \sigma_M \dot{W}_M. \tag{3}$$

The natural requirement $\gamma_M > 0$ is needed for 2.7 to have a Gaussian statistical steady state. Now consider the situation where the model in 2.7 has been tuned to match the single time statistics for $u$ in 2.5 with perfect fidelity by matching the mean and variance of $u_M$ with $u$; elementary calculations show this is true for a one parameter family of models parameterized by $\gamma_M > 0$ provided that $F_M$, $\sigma_M^2$ satisfy the equilibrium mean and variance equations

$$\frac{F_M}{\gamma_M} = -\frac{AF}{aA - q}, \qquad \frac{\sigma_M^2}{2\gamma_M} = -\frac{\sigma^2}{2(a+A)(aA-q)} \equiv E. \tag{4}$$

Thus, the conditions in 2.8 for $F_M$ and $\sigma_M$ guarantee perfect model fidelity for any $\gamma_M > 0$. In many practical situations such as actual experiments or climate science, it is important to understand the response of the natural system to external forcing, $\delta F$, and to hope that the response of the imperfect model captures the features of this response. The natural system response for 2.5 occurs by replacing $F$ in 2.5 by $F + \delta F$ while the same experiment in the model for 2.7 involves replacing $F_M$ by $F_M + \delta F$. For both the natural system in 2.5 and the model system in 2.7, the only change in the equilibrium response is through the change in mean

$$\delta u = -\frac{A}{aA - q}\delta F, \qquad \delta u_M = \frac{1}{\gamma_M}\delta F, \tag{5}$$

while the variance of $u$ for the perfect model and $u_M$ for the imperfect model stays constant at the same value $E$ determined through the second equality in 2.8. Now assume that the natural system satisfies the stability conditions in 2.6 with $A > 0$. We claim that no model from 2.7, even with perfect fidelity in 2.8 for any $\gamma_M > 0$, can match the sensitivity of the natural system correctly; this is easy to see from 2.9 since for $A > 0$, $\text{sign}(\delta u) = -\text{sign}(\delta F)$ but for all models from 2.7, $\text{sign}(\delta u_M) = \text{sign}(\delta F)$ and the perfect and model sensitivity are always anti-correlated!

Thus, even though the climate models satisfying 2.7, 2.8 are tuned to exactly match the true climate, these imperfect models are intrinsically deficient in calculating the crucial climate sensitivity for $A > 0$.

## 2.1 Systematically improving climate models through empirical information theory

With a subset of variables $u \in \mathbb{R}^N$ and a family of measurement functionals $E_L(u) = (E_j(u))$, $1 \le j \le L$, for the perfect system, empirical information theory (Jaynes (1957); Majda and Wang (2006)) builds the least biased probability measure $\pi_L(u)$ consistent with the $L$ measurements of the present climate,

$\overline{E}_L$. There is a unique functional on probability densities (Jaynes (1957); Majda and Wang (2006)) to measure this given by the entropy

$$\mathscr{S} = -\int \pi \log \pi, \tag{6}$$

and $\pi_L(u)$ is the unique probability so that $\mathscr{S}(\pi_L(u))$ has the largest value among those probability densities consistent with the measured information, $\overline{E}_L$. All integrals as in 2.10 are over the phase space $\mathbb{R}^N$ unless otherwise noted. For example, measurements of the mean and second moments of the perfect system necessarily lead to a Gaussian approximation (Majda et al. (2002); Majda and Wang (2006)) to the perfect system from measurements, $\pi_L(u) = \pi_G(u)$. Any model of the perfect system produces a probability density, $\pi^M(u)$. The natural way (Kullback and Leibler (1951); Majda and Wang (2006)) to measure the lack of information in one probability density, $q(u)$, compared with the true probability density, $p(u)$, is through the relative entropy, $\mathscr{P}(p,q)$, given by

$$\mathscr{P}(p,q) = \int p \log\left(\frac{p}{q}\right). \tag{7}$$

This asymmetric functional on probability densities, $\mathscr{P}(p,q)$, has two attractive features (Kullback and Leibler (1951); Majda et al. (2002); Majda and Wang (2006)) as a metric for model fidelity: (i) $\mathscr{P}(p,q) \geq 0$ with equality if and only if $p = q$; (ii) $\mathscr{P}(p,q)$ is invariant under general nonlinear changes of variables. The first issue to contend with is the fact that $\pi_L(u)$ is not the actual perfect model density but only reflects the best unbiased estimate of the perfect model given the $L$ measurements, $\overline{E}_L$. Let $\pi(u)$ denote the probability density of the perfect model, which is not actually known. Nevertheless, $\mathscr{P}(\pi, \pi_L)$ precisely quantifies the intrinsic error in using the $L$ measurements of the perfect model, $\overline{E}_L$. Consider an imperfect model with its associated probability density, $\pi^M(u)$; then the intrinsic model error in the climate statistics is given by $\mathscr{P}(\pi, \pi^M)$. In practice, $\pi^M(u)$ is determined by no more information than that available in the perfect model.

Consider a class of imperfect models, $\mathscr{M}$. The best imperfect model for the coarse-grained variable $u$ is the $M_* \in \mathscr{M}$ so that the perfect model has the smallest additional information beyond the imperfect model distribution $\pi^{M_*}(u)$, i.e.,

$$\mathscr{P}(\pi, \pi^{M_*}) = \min_{M \in \mathscr{M}} \mathscr{P}(\pi, \pi^M). \tag{8}$$

Also, actual improvements in a given imperfect model with distribution $\pi^M(u)$ resulting in a new $\pi^M_{post}(u)$ should result in improved information for the perfect model, so that $\mathscr{P}(\pi, \pi^M_{post}) \leq \mathscr{P}(\pi, \pi^M)$. Otherwise, objectively, the model has not been improved compared with the original perfect model. The following general principle (Majda and Gershgorin (2010); Majda et al. (2005)) facilitates the practical calculation of 2.12

$$\begin{aligned} \mathscr{P}(\pi, \pi^M_{L'}) &= \mathscr{P}(\pi, \pi_L) + \mathscr{P}(\pi_L, \pi^M_{L'}) \\ &= (\mathscr{S}(\pi_L) - \mathscr{S}(\pi)) + \mathscr{P}(\pi_L, \pi^M_{L'}) \qquad \text{for } L' \leq L. \end{aligned} \tag{9}$$

The entropy difference, $\mathscr{S}(\pi_L) - \mathscr{S}(\pi)$ in 2.13 precisely measures an intrinsic error from the $L$ measurements of the perfect system. With 2.13 and a fixed family of $L$ measurements of the actual climate, the optimization principle in 2.12 can be computed explicitly by replacing the unknown density $\pi$ by the hypothetically known $\pi_L$ in these formulas so that, for example, $\pi^{M_*}$ is calculated by

$$\mathscr{P}(\pi_L, \pi^{M_*}_{L'}) = \min_{M \in \mathscr{M}} \mathscr{P}(\pi_L, \pi^M_{L'}). \tag{10}$$

The most practical setup for applying the framework of empirical information theory developed above arises when both the perfect system measurements and the model measurements involve only the mean and covariance of the variables $u$ so that $\pi_L$ is Gaussian with climate mean $\bar{u}$ and covariance $R$ while $\pi^M$

is Gaussian with model mean $\bar{u}_M$ and covariance $R_M$. In this case, $\mathscr{P}(\pi_L, \pi^M)$ has the explicit formula (Majda and Wang (2006); Kleeman (2002a))

$$\mathscr{P}(\pi_L, \pi^M) = \left[\frac{1}{2}(\bar{u}-\bar{u}_M)^*(R_M)^{-1}(\bar{u}-\bar{u}_M)\right] + \left[-\frac{1}{2}\log\det(RR_M^{-1}) + \frac{1}{2}(\mathrm{tr}(RR_M^{-1}) - N)\right]. \tag{11}$$

Note that the first term in brackets in 2.15 is the signal, reflecting the model error in the mean but weighted by the inverse of the model covariance, $R_M^{-1}$ while the second term in brackets, the dispersion, involves only the model error covariance ratio, $RR_M^{-1}$. The intrinsic metric in 2.15 is invariant under any (linear) change of variables which maps Gaussian distributions to Gaussians and the signal and dispersion terms are individually invariant under these transformations; this property is very important.

As a simple illustration of these concepts, let's assume the elementary perfect and imperfect climate models discussed in (1) and (3) above, where as shown below, empirical information theory reveals an intrinsic barrier for the imperfect models to prediction of the sensitivity for $A > 0$.

The formula in 2.15 applies exactly to these models with perfect fidelity with

$$\mathscr{P}(\pi_\delta, \pi_\delta^M) = \frac{1}{2}E^{-1}\left|-\frac{A}{aA-q} - \frac{1}{\gamma_M}\right|^2 |\delta F|^2. \tag{12}$$

In this situation with $A > 0$, the attempt to minimize the information theoretic model error in the sensitivity through the general principle in 2.12 is futile because no finite minimum over $\gamma_M$ of 2.16 is achieved and necessarily $\gamma_M \to \infty$ in the approach to this minimum value; in other words, there is an intrinsic barrier to skill in sensitivity which cannot be overcome with the imperfect models in 2.7 even though they satisfy perfect model fidelity in 2.8. In this situation, information theory predicts that one needs to enlarge the class of models beyond 2.7 by introducing more degrees of freedom in the model. On the other hand, if the natural system satisfies 2.6 with $A < 0$, then using 2.16 to minimize the lack of information in the sensitivity in the models which satisfy perfect fidelity in 2.8 results in the unique model with

$$\gamma_M^* = -A^{-1}(aA-q), \quad A < 0, \tag{13}$$

and this model captures both the model fidelity and model sensitivity to this forcing parameter exactly.

The relative entropy in (7) occupies a central role in statistics (Bernardo and Smith (2000); Berger (1985); Williams (2001)) and large deviation theory in the limit of large sample sizes (Varadhan (1984, 1985)). The empirical point of view presented here is useful for developing unbiased empirical statistical/physics-based models and has been utilized to predict the location and structure of Jupiter's Red Spot from observations of the Galileo mission, as well as the behavior of large-scale quantities in statistical fluid dynamics (Majda and Wang (2006)). Kleeman (2002a) first applied these ideas to the prediction skill for long range forecasting in perfect models and these concepts have been developed extensively (Kleeman (2011); Majda et al. (2002); Kleeman et al. (2002b); Abramov et.al (2005); Haven et al. (2005)) in the context of perfect models. Recent research utilizing empirical information theory has focussed on important coarse-grained descriptions of perfect and imperfect models and improving the long range forecasting and sensitivity of imperfect models (Majda and Gershgorin (2010); Giannakis and Majda (2011a,b); Majda and Gershgorin (2011a,b)).

## 2.2 Statistically exactly solvable test models capturing crucial features of climate change science

An important role of mathematics in applied sciences is to develop simpler exactly or easily solvable test models with unambiguous mathematical features which nevertheless capture crucial features of vastly more complex systems in science and engineering. Such models provide firm underpinning for advancing scientific understanding and developing new numerical or statistical algorithms. With all of the difficult issues in climate science, such unambiguous test models assume a crucial role.

We introduce a family of test models for climate change science which have direct qualitative relevance for actual observed features for tracers in the atmosphere (Neelin et al. (2009); Bourlioux and Majda (2002)) with the additional attractive feature of exactly solvable statistics for the mean and covariance with many degrees of freedom despite the inherent statistical nonlinearity. Thus, they are physically relevant unambiguous test models for uncertainty in climate change science (Majda and Gershgorin (2010, 2011a,b)). The models have a zonal (east-west) mean jet, $U(t)$, a family of planetary and synoptic scale waves with north-south velocity $v(x,t)$ with $x$, a spatially periodic variable representing a fixed midlatitude circle in the east-west direction, and tracer gas $T(x,t)$ with a north-south environmental mean gradient $\alpha$ and molecular diffusivity $\kappa$ (Majda and Gershgorin (2011c); Bourlioux and Majda (2002)). The dynamical equations for these variables are

$$\begin{aligned} &A) \qquad \frac{dU}{dt} = -\gamma U + f(t) + \sigma \dot{W}, \\ &B) \qquad \frac{dv}{dt} = P\left(\frac{\partial}{\partial x}\right) v + \sigma_v(x)\dot{W}_v + f_v(x,t), \\ &C) \qquad \frac{\partial T}{\partial t} + U(t)\frac{\partial T}{\partial x} = -\alpha v(x,t) + \kappa \frac{\partial^2 T}{\partial x^2} - d_T T. \end{aligned} \tag{14}$$

The functions $f(t)$, $f_v(x,t)$ are known time-periodic functions with period of 1 yr reflecting the changing external forcing of the seasonal cycle, while $\dot{W}$, $\dot{W}_v$, represent random white noise fluctuations in forcing arising from hidden nonlinear interactions and other processes (Majda et al. (2010); Majda and Grote (2007)). The equation in (2.2B) for the turbulent planetary waves is solved by Fourier series with independent scalar complex variable versions of the equation in (2.2A) for each different wave number $k$ (Majda et al. (2010); Majda and Grote (2007)); in Fourier space the operator $\hat{P}_k$ has the form $\hat{P}_k = -\gamma_k + i\omega_k$ with frequency $\omega_k = \beta k/(k^2 + F_s)$ corresponding to the dispersion relation of baroclinic Rossby waves and dissipation $\gamma_k = \nu(k^2 + F_s)$ where $\beta$ is the north-south gradient of rotation, $F_s$ is the stratification, and $\nu$ is a damping coefficient; the white noise forcing for (2.2)B is chosen to vary with each spatial wave number k to generate an equipartition energy spectrum for planetary scale wave numbers $1|k|10$ and a $|k|^{-5/3}$ turbulent cascade spectrum for $11|k|52$ (see Majda et al. (2010); Majda and Grote (2007)). The zonal jet $U(t) = \bar{U}(t) + U'(t)$, where $\bar{U}(t)$ is the climatological periodic mean with $\gamma$, and $\sigma$ chosen so that this jet is strongly eastward while the random fluctuations, $U'(t)$, have a standard deviation consistent with such eastward dynamical behavior. While $U(t), v(x,t)$ have exactly solvable Gaussian statistics mimicking features of the atmosphere, the tracer $T(x,t)$ has non-Gaussian behavior due to the nonlinear tracer flux term $U'(t)\frac{\partial T}{\partial x}$ in (2.2C) with intermittent fat tails like realistic tracers in the atmosphere (Majda and Gershgorin (2011c); Neelin et al. (2009)); nevertheless, $T(x,t)$ has exactly solvable mean and covariance climate statistics with explicit formulas. These procedures define the exactly solvable statistics for the perfect climate. Actual AOS models utilized in climate change science typically have too much additional damping and one can mimic this here in the representative AOS models by increasing the two parameters $\gamma$, $\nu$ for (2.2A,B) to $\gamma_M$, $\nu_M$ to define the AOS model velocity fields $U(t)_M = \bar{U}_M(t) + U'_M(t), v_M(x,t)$, with model error. The turbulent tracer in an AOS model is usually calculated roughly by an eddy diffusivity (Emanuel et al. (2005); Neelin et al. (2003); NRC (1979); Randall et al. (2007)), $\overline{U'_M(t)\frac{\partial T}{\partial x}} = -\kappa^*_M T_{xx}$, and in the present models there is an exact explicit formula for $\kappa^*_M$. Thus, the AOS model tracer satisfies

$$\frac{\partial T_M}{\partial t} + \bar{U}_M(t)\frac{\partial T_M}{\partial x} = -\alpha v_M(x,t) + (\kappa + \kappa^*_M)\frac{\partial^2 T_M}{\partial x^2} - d_T T_M + \sigma_T \dot{W}(x,t), \tag{15}$$

where $\dot{W}(x,t)$ denotes space-time white noise forcing with variance $\sigma_T$ to overcome deterministic model error. With (15) the AOS model with $(U_M, v_M, T_M)$ has Gaussian statistics.

Note that the above perfect and imperfect climate models do not have positive Lyapunov exponents but nevertheless exhibit non-normal transient growth through the non-zero mean gradient, $\alpha > 0$, for the tracer. These models have been utilized as unambiguous test models for all the issues of climate change science, information theory, prediction, and FDT described earlier in this section (Majda and Gershgorin (2010, 2011a,b)). These are also important test models for the real-time recovery of turbulent tracer fields from partial observations, an important topic with much practical interest in climate science, as well as other disciplines (Gershgorin et al. (2011)). A complete development of the turbulent statistics of such test models is presented in Majda and Gershgorin (2011c). Similar exactly solvable test models with intermittent positive Lyapunov exponents are developed elsewhere (Gershgorin et al. (2010a,b); Branicki et al. (2011)) and mentioned briefly in section 3 in the context of filtering.

## 3 Judicious model errors in filtering or data assimilation

Can simple models incorporate non-Gaussian features of turbulent dynamical systems yet have the advantage of cheap computational overhead for filtering turbulent dynamical systems from sparse observations? A key feature of turbulence is bursts of energy across multiple scales with intermittent instability and random forcing. Stochastic Parameterization Extended Kalman Filters (SPEKF) have been introduced and analyzed recently (Gershgorin et al. (2010a,b); Majda et al. (2010); Majda and Harlim (2011)) as computationally cheap algorithms which make judicious model errors which retain high filtering skill for complex turbulent signals (Harlim and Majda (2010); Keating et al. (2011); Branicki et al. (2011); Majda et al. (2010)). For example, aliasing is usually viewed as a bad feature of numerical algorithms; in the present context, judicious use of aliasing yields stochastic superresolution (Majda et al. (2010); Keating et al. (2011); Majda and Harlim (2011)).

The basis for the SPEKF algorithms is the following system for the complex scalar partially observed turbulent signal $u$ (the reader can think of a Fourier amplitude of turbulence at a given spatial wavenumber) coupled with stochastic additive forcing and multiplicative damping/instability coefficients, $b, \gamma$, which are learned "on the fly" from the observed turbulent signal

$$\begin{aligned}
&(a) \qquad \mathrm{d}u(t) = \big[(-\gamma(t) + i\omega)u(t) + b(t) + f(t)\big]\mathrm{d}t + \sigma_u \mathrm{d}W_u(t),\\
&(b) \qquad \mathrm{d}b(t) = \big[(-\gamma_b + i\omega_b)(b(t) - \hat{b})\big]\mathrm{d}t + \sigma_b \mathrm{d}W_b(t),\\
&(c) \qquad \mathrm{d}\gamma(t) = -d_\gamma(\gamma(t) - \hat{\gamma})\mathrm{d}t + \sigma_\gamma \mathrm{d}W_\gamma(t),
\end{aligned} \tag{16}$$

where $W_u, W_b$ are independent complex Wiener processes with independent components and $W_\gamma$ is a real Wiener process. There are nine parameters in the system (16): two damping parameters $\gamma_b$, $d_\gamma$, two oscillation frequencies $\omega$ and $\omega_b$, two stationary mean terms $\hat{b}$ and $\hat{\gamma}$ and noise amplitudes $\sigma_u, \sigma_b, \sigma_\gamma$; $f$ is a deterministic forcing. The advantage of the equations in (16) is that they have non-Gaussian dynamics but nevertheless exactly solvable first and second-order statistics, so they are readily implemented practically in a filtering algorithm and avoid linear tangent approximations which are dangerous when filtering (hidden) instabilities (Branicki et al. (2011)). The equations in (16) have rich statistical behavior in a variety of regimes and this complex behavior can be utilized to test the filter performance of a wide variety of Gaussian filter approximations (Branicki et al. (2011)). Such models are also useful as an unambiguous test bed for all of the issues of prediction and model error discussed in section 2.

The systematic approach of making judicious model errors in filtering or data assimilation for turbulent dynamical systems in order to avoid the "curse of dimension" and "curse of ensemble size" requires the blending of concepts from stochastic turbulence theory, numerical analysis of PDE's, and Kalman filtering. A survey of the approach is found in Majda et al. (2010) and there is a recent introductory text

Majda and Harlim (2011) suitable for first and second year graduate students, as well as researchers, which develops and applies these concepts systematically to turbulent dynamical systems; this text also includes extensive discussion, numerical examples, and comparisons with finite ensemble Kalman filtering as well as recent attempts to do particle filtering. The recent paper Harlim and Majda (2010) reports high skill for the SPEKF algorithms with low computational overhead for the recovery of turbulence in baroclinic instability from sparse observations for both atmospheric and oceanic regimes of the two-layer model; in contrast Harlim and Majda (2010), there is very high filtering skill comparable to SPEKF of the finite ensemble transform methods at much higher computational cost for the atmospheric regime but extremely poor skill of the finite ensemble methods in the ocean regime where the underlying dynamics are stiff due to the small Rossby deformation radius. The challenging problem of recovering turbulent heat fluxes in the ocean from coarse satellite altimetry measurements (Keating et al. (2011)) is studied through computationally inexpensive "stochastic superparameterization" SPEKF algorithms Majda et al. (2010); Majda and Harlim (2011) with significant filtering skill in prototype two-layer test problems for both low latitude and high latitude regimes. The filtering skill of the SPEKF algorithms for recovery of turbulent tracers and their statistics from partial observations is discussed completely in Gershgorin et al. (2011) for the test models from section 2.

Acknowledgements

The first author thanks his collaborator in tropical convection, Boualem Khouider, and his collaborators for information theory and filtering, Boris Gershgorin and John Harlim for their explicit and implicit contributions. The research of A. Majda is partially supported by NFS grant DMS-0456713, NSF CMG grant DMS-1025468, and ONR grants ONR-DRI N00014-10-1-0554 and N00014-11-1-0306. M. Branicki and Y. Frenkel are supported as postdoctoral fellows on the last two grants.

## References

Abramov R., A.J. Majda, and R. Kleeman, 2005: Information theory and predictability for low frequency variability. *J. Atmos. Sci.*, **62**, 65–87.

Arakawa, A., and W. H. Schubert, 1974: Interaction of a cumulus cloud ensemble with the large-scale environment, Part I. *J. Atmos. Sci.*, **31**, 674–701.

Betts, A. K., and M. J. Miller, 1986: A new convective adjustment scheme. Part II: Single column tests using GATE wave, BOMEX, atex and arctic air-mass data sets. *Quart. J. Roy. Meteor. Soc.*, **112 (473)**, 693–709.

Bernardo J.M., and A.F.M. Smith, 2000: *Bayesian theory*, Wiley series in probability and mathematical statistics, John Wiley & Sons, Hoboken NJ, p. 586.

Berger J.O., 1985: *Statistical decision theory and Bayesian analysis*, Springer, New York, p. 624.

Bourlioux A., and A.J. Majda, 2002: Elementary models with probability distribution function intermittency for passive scalars with a mean gradient. *Phys. Fluids*, **14**, 881–897

Branicki M., B. Gershgorin, and A.J. Majda, 2011: Filtering skill for turbulent signals for a suite of nonlinear and linear Kalman filters. *J. Comp. Phys.*, submitted.

Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908.

Chatterjee A., and D. Vlachos, 2007: An overview of spatial microscopic and accelerated kinetic Monte Carlo methods. *J. of Computer-Aided Materials Design*, **14**, 253–308.

Deker U., and F. Haake, 1975: Fluctuation-dissipation theorems for classical processes. *Phys. Rev. A*, **11**(6), 2043–2056.

DelSole T., 2005: Predictability and information theory. Part 2: Imperfect models. *J. Atmos. Sci.*,**62**, 3368–3381.

DelSole T., and J. Shukla, 2010: Model fidelity versus skill in seasonal forecasting. *J. Climate*, **23**, 4794–4806.

ECMWF, 2003: *Proceedings ECMWF/CLIVAR Workshop on Simulation and Prediction of Intraseasonal variability with Emphasis on the MJO*, 3–6 November 2003.

Emanuel K.A., J.C. Wyngaard, J.C. McWilliams, D.A. Randall, and Y.L. Yung, 2005: *Improving the scientific foundation for atmosphere-land ocean simulations*, Natl. Acad. Press, Washington DC, pp. 72.

Frenkel, Y., A. Majda, and B. Khouider, 2011: Using the stochastic multicloud model to improve tropical convective parameterization: A paradigm example. *J. Atmos. Sci.*, submitted

Gershgorin B., J. Harlim, and A.J. Majda, 2010: Test models for improving filtering with model errors through stochastic parameter estimation. *J. Comp. Phys*, **229**(1), 1–31.

Gershgorin B., J. Harlim, and A.J. Majda, 2010: Improving filtering and prediction of spatially extended turbulent systems with model errors through stochastic parameter estimation. *J. Comp. Phys*, **229**(1), 32–57.

Gershgorin B., and A.J. Majda, 2011: Filtering a statistically exactly solvable test model for turbulent tracers from partial observations. *J. Comp. Phys.*, **230**, 1602–1638.

Giannakis D., and A.J. Majda, 2011: Quantifying the Predictive Skill in Long-Range Forecasting. Part I: Coarse-grained Predictions in a Simple Ocean Model. *J. Climate*, submitted.

Giannakis, D., and A.J. Majda, 2011: Quantifying the Predictive Skill in Long-Range Forecasting. Part II: Model Error in Coarse-Grained Markov Models with Application to Ocean-Circulation Regimes. *J. Climate*, submitted.

Gillespie, D. T., 1975: An exact method for numerically simulating the stochastic coalescence process in a cloud. *J. Atmos. Sci.*, **32**, 1977–1989.

Gillespie, D. T., 1977: Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, **81 (25)**, 2340–2361.

Grabowski, W. W., 2001: Coupling cloud processes with the large-scale dynamics using the cloud-resolving convection parameterization (CRCP). *J. Atmos. Sci.*, **58 (9)**, 978–997.

Grabowski, W. W., 2004: An improved framework for superparameterization. *J. Atmos. Sci.*, **61 (15)**, 1940–1952.

Grabowski, W. W., and P. K. Smolarkiewicz, 1999: CRCP: a Cloud Resolving Convection Parameterization for modeling the tropical convecting atmosphere. *Physica D Nonlinear Phenomena*, **133**, 171–178.

Grabowski, W. W., J.-I. Yano, and M. W. Moncrieff, 2000: Cloud resolving modeling of tropical circulations driven by large-scale sst gradients. *J. Atmos. Sci.*, **57 (13)**, 2022–2040.

Harlim, J., and A.J. Majda, 2010: Filtering Turbulent Sparsely Observed Geophysical Flows. *Mon. Wea. Rev.*, **138**(4), 1050–1083.

Haven, K., A.J. Majda and R. Abramov, 2005: Quantifying predictability through information theory: small sample estimation in non-Gaussian framework. *J. Comp. Phys.*, **206**, 334–362.

Horenko, I., 2010: Nonstationarity in multifactor models of discrete jump processes, memory and application to cloud modeling. *J. Atmos. Sci.*, doi: 10.1175/2011JAS3692.1

Hwang, Y., and D. Frierson, 2010: Increasing atmospheric poleward energy transport with global warming. *Geophys. Res. Letters*, **37**, L24807

Johnson, R. H., T. M. Rickenbach, S. A. Rutledge, P. E. Ciesielski, and W. H. Schubert, 1999: Trimodal characteristics of tropical convection. *J. Climate*, **12 (8)**, 2397–2418.

Kain, J. S., and J. M. Fritsch, 1990: A one-dimensional entraining/detraining plume model and its application in convective parameterization. *J. Atmos. Sci.*, **47**.

Katsoulakis, M. A., A. J. Majda, and D. G. Vlachos, 2003: Coarse-grained stochastic processes for microscopic lattice systems. *Proc. Natl. Acad. Sci. USA*, **100**, 782–787.

Katsoulakis, M. A., A. J. Majda, and D. G. Vlachos, 2003b: Coarse-grained stochastic processes and monte carlo simulations in lattice systems. *J. Comp. Phys.*, **186 (1)**, 250 – 278.

Keating S., A.J. Majda, and K.S. Smith, 2011: New methods for estimating poleward eddy heat transport using satellite altimetry. *Mon. Wea. Rev.*, submitted.

Khouider, B., J. Biello, and A. J. Majda, 2010: A stochastic multicloud model for tropical convection. *Comm. Math. Sci.*, **8 (1)**, 187–216.

Khouider, B., Y. Han, and J. Biello, 2011: Convective momentum transport in a simple multicloud model. *J. Atmos. Sci.*, submitted.

Khouider, B., and A. J. Majda, 2006a: Multicloud convective parametrizations with crude vertical structure. *Theor. Comp. Fluid Dyn.*, **20**, 351–375.

Khouider, B., and A. J. Majda, 2006b: A simple multicloud parametrization for convectively coupled tropical waves. Part I: Linear analysis. *J. Atmos. Sci.*, **63**, 1308–1323.

Khouider, B., and A. J. Majda, 2007: A simple multicloud parametrization for convectively coupled tropical waves. Part II: Nonlinear simulations. *J. Atmos. Sci.*, **64**, 381–400.

Khouider, B., and A. J. Majda, 2008a: Equatorial convectively coupled waves in a simple multicloud model. *J. Atmos. Sci.*, **65**, 3376–3397.

Khouider, B., and A. J. Majda, 2008b: Multicloud models for organized tropical convection: Enhanced congestus heating. *J. Atmos. Sci.*, **65**, 897–914.

Khouider, B., A. J. Majda, and M. A. Katsoulakis, 2003: Coarse-grained stochastic models for tropical convection and climate. *Proc. Natl. Acad. Sci. USA*, **100**, 11 941–11 946.

Kleeman R., 2002: Measuring dynamical prediction utility using relative entropy. *J. Atmos. Sci.*, **59**, 2057–2072.

Kleeman R., Majda A.J., Timofeyev I., 2002: Quantifying predictability in a model with statistical features of the atmosphere. *Proc. Natl. Acad. Sci. USA*, **99**, 15291–15296.

Kleeman R., 2011: Information theory and dynamical system predictability. *Entropy*, **13**, 612–649.

Kullback S., Leibler R., 1951: On information and sufficiency. *Ann. Math. Stat.*, **22**, 79–86.

Kuo, H. L., 1974: Further studies of the parameterization of the influence of cumulus convection on large-scale flow. *J. Atmos. Sci.*, **31**, 1232–1240.

Jaynes E.T., 1957: Information Theory and Statistical Mechanics. *Phys. Rev.*, **106**, 620–630.

Lin, J., and J. D. Neelin, 2003: Toward stochastic deep convective parameterization in general circulation models. *Geophys. Res. Letters*, **30 (4)**, 1162. doi:10.1029/2002GL016203

Majda A.J., R. Kleeman, and D. Cai, 2002: A framework for predictability through relative entropy. *Meth. Appl. Anal.*, **9**, 425–444.

Majda, A. J., and B. Khouider, 2002: Stochastic and mesoscopic models for tropical convection. *Proc. Natl. Acad. Sci. USA*, **99**, 1123–1128.

Majda A.J., R. Abramov, and M. Grote, 2005: *Information Theory and Stochastics for Multiscale Nonlinear Systems*, CRM Monogr. Series, American Mathematical Society, 133 pp.

Majda A.J., and X. Wang, 2006: *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows*, Cambridge University Press, 564 pp.

Majda, A. J., 2007: Multiscale models with moisture and systematic strategies for superparameterization. *J. Atmos. Sci.*, **64 (7)**, 2726–2734.

Majda A.J., and M. Grote, 2007: Explicit off-line criteria for stable accurate time filtering of strongly unstable spatially extended systems. *Proc. Natl. Acad. Sci. USA*, **104**, 1124–1129.

Majda, A. J., C. Franzke, and B. Khouider, 2008: An applied mathematics perspective on stochastic modelling for climate. *Philosophical Trans. Roy. Soc. A: Mathematical, Physical and Engineering Sciences*, **366 (1875)**, 2427–2453.

Majda, A. J., and S. N. Stechmann, 2008: Stochastic models for convective momentum transport. *Proc. Natl. Acad. Sci. USA*, **105 (46)**, 17 614–17 619.

Majda, A. J., and S. N. Stechmann, 2009: A simple dynamical model with features of convective momentum transport. *J. Atmos. Sci.*, **66**, 373392.

Majda A.J., and B. Gershgorin, 2010: Quantifying Uncertainty in Climate Change Science Through Empirical Information Theory. *Proc. Natl. Acad. Sci. USA*, **107**(34), 14958–14963.

Majda A.J., J. Harlim, and B. Gershgorin, 2010: Mathematical strategies for filtering turbulent dynamical systems. *Discrete Contin. Dyn. Sys.*, **27**, 441–486.

Majda A.J., and J. Harlim, 2011: *Systematic Strategies for real time filtering of turbulent signals in complex systems*, Cambridge Universtiy Press, UK.

Majda A.J., and B. Gershgorin, 2011: Improving model fidelity and sensitivity for complex systems through empirical information theory, in press, *Proc. Natl. Acad. Sci. USA*.

Majda A.J. and B. Gershgorin, 2011: The link between statistical equilibrium fidelity and forecasting skill for complex systems with model error. In press, *Proc. Natl. Acad. Sci. USA*.

Majda A.J., and B. Gershgorin, 2011: Elementary models for turbulent diffusion with complex physical features: eddy diffusivity, spectrum, and intermittency. *Phil. Trans. Roy. Soc*, submitted.

Manabe, S., J. Smagorinsky, and R. F. Strickler, 1965: Simulated climatology of a general circulation model with a hydrologic cycle1. *Mon. Wea. Rev.*, **93 (12)**, 769–798.

Mapes, B. E., 2000: Convective inhibition, subgrid-scale triggering energy, and stratiform instability in a toy tropical wave model. *J. Atmos. Sci.*, **57 (10)**, 1515–1535.

Marconi U., A. Puglisi, L. Rondoni, and A. Vulpani, 2008: Fluctuation-dissipation: Response theory in statistical physics. *Physics Reports*, **461**, 111–195.

Moncrieff, M., M. Shapiro, J. Slingo, and F. Molteni, 2007: Collaborative research at the intersection of weather and climate. *WMO Bulletin*, **56**, 204–211.

Neelin J.D., M. Munnich, H. Su, J.E. Meyerson, and C.E. Holloway, 2003: Tropical drying trends in global warming models and observations. *Proc. Natl. Acad. Sci. USA*, **103**, 6110–6115.

Neelin, J.D., B.R. Lintner, B. Tian, Q. Li, L. Zhang, P.K. Patra, M.T. Chahine, and S.N.Stechmann, 2009: Long tails in deep columns of natural and anthropogenic tropospheric tracers. *Geophys. Res. Lett.*, **37**, L05804.

National Research Council/Ad Hoc Study Group on Carbon Dioxide and Climate, 1979: *Carbon dioxide and climate: A scientific assessment: Report of an ad hoc study group on carbon dioxide and climate. Woods Hole, Massachusetts, July 23–27, 1979, to the Climate Research Board, Assembly of Mathematical and Physical Sciences*. Natl. Acad. Press, Washington DC, pp 22.

Palmer, T. N., 2001: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Q. J. Roy. Meteor. Soc.*, **127**, 279–304.

Randall, D., M. Khairoutdinov, A. Arakawa, and W. Grabowski, 2003: Breaking the cloud parameterization deadlock. *Bull. Americ. Meteor. Soc.*, **84 (11)**, 1547–1564.

Randall D.A., et al., 2007: Climate models and their evaluation. *Climate change 2007: The physical science basis, contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*, eds S. Solomon et al., Cambridge Univ. Press, Cambridge, UK, 589–662.

Rangan A., L. Tao, G. Kovacic, and D. Cai, 2009: Multiscale modeling of the primary visual cortex, *IEEE Engineering in Medicine and Biology Magazine*, **28**(3), 19–24.

Stechmann, S. N. and J. D. Neelin, 2011: A stochastic model for the transition to strong convection. *J. Atmos. Sci.* doi: 10.1175/JAS-D-11-028.1

Varadhan S.R.S., 1984: *Large Deviations and Applications*. CBMS-NSF Regional Conference Series in Applied Mathematics 46, SIAM.

Varadhan S.R.S., 1985: Large Deviations and Applications. *Expo. Math.*, **3**, 251–272.

Williams D., 2001: *Weighing the odds: A course in probability and statistics*. Cambridge University Press, UK, pp. 566.

Xing, Y., A. J. Majda, and W. W. Grabowski, 2009: New efficient sparse space-time algorithms for superparameterization on mesoscales. *Mon. Wea. Rev.*, **137 (12)**, 4307–4324.

Zhang, G. J. and N. A. McFarlane, 1995: Role of convective-scale momentum transport in climate simulation. *J. Geophys. Res.*, **100**, 1417–1426.