Predictability of extreme events in a nonlinear stochastic-dynamical model

Christian Franzke*
British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, United Kingdom
(Received 1 November 2011; revised manuscript received 13 February 2012; published 22 March 2012)

The objective of this work is to evaluate the potential of reduced order models to reproduce the extreme event and predictability characteristics of higher dimensional dynamical systems. A nonlinear toy model is used which contains key features of comprehensive climate models. First, we demonstrate that the systematic stochastic mode reduction strategy leads to a reduced order model with the same extreme value characteristics as the full dynamical models for a wide range of time-scale separations. Second, we find that extreme events in this model follow a generalized Pareto distribution with a negative shape parameter; thus extreme events are bounded in this model. Third, we show that a precursor approach has good forecast skill for extreme events. We then find that the reduced stochastic models capture the predictive skill of extreme events of the full dynamical models well. Consistent with previous studies we also find that the larger the extreme events, the better predictable they are. Our results suggest that systematically derived reduced order models have the potential to be used for the modeling and statistical prediction of weather- and climate-related extreme events and, possibly, in other areas of science and engineering too.

DOI: 10.1103/PhysRevE.85.031134 PACS number(s): 02.50.Ey

I. INTRODUCTION

The accurate modeling and probabilistic prediction of natural extreme and catastrophic events are of utmost importance for the insurance and reinsurance industry, catastrophe modeling companies, policy makers, and the wider society. Natural catastrophes include earthquakes, hurricanes, winter storms, and flooding, and the corresponding financial losses can reach billions of dollars.

The Intergovernmental Panel on Climate Change (IPCC) has stated that it is likely that anthropogenic climate change leads to changes in the frequency and intensity of weather and climatic extreme events [1]. The first 6 months of 2011 incurred insurance losses of about US$60 billion, which is about five times the average for the first 6 months of the year in the period 2001–2010 [2]. This illustrates the challenge that societies are facing in mitigating the effects of natural catastrophes.

The natural starting point in characterizing extreme events is extreme value statistics. This is a very mature statistical framework but it is based on the assumption that extreme events are independent and identically distributed (iid). However, there is mounting evidence that many natural extreme events are clustered; i.e., extreme events come in bunches. A prime example is European wind storms [3]. This shows that the dependence structure of the underlying dynamics has to be taken into account for the estimation of the occurrence frequencies of extremes. It has been shown that a strong serial dependence in time series leads to clustering of extreme events [4]. This illustrates the need to better understand extreme events in dynamical systems which have a temporal dependence structure [5–10] and also to investigate the predictability of extreme events [11–13]. If extreme events were indeed iid, their distribution function would also provide the mean return periods of extreme events of a given magnitude. The serial dependence of extreme events leads to clustering and thus the mean return periods can be misleading, because large events can recur sooner than expected. This issue is particularly important for the insurance and reinsurance industry.

The most common approach to predicting weather- and climate-related extreme events is the use of high-dimensional numerical weather and climate prediction models. This is typically done by use of ensemble predictions. These ensemble predictions start from different initial conditions. However, current numerical weather and climate prediction systems tend to be underdispersive. This means that the observed state being forecasted is too often outside of the forecast ensemble [14], which leads, among other things, to an underestimation of extreme events. Furthermore, the computational expense of state-of-the-art weather and climate prediction models limits the number of ensembles and simulation length of climate simulations.

In many applications, in particular, insurance and reinsurance, one is not interested in whether an extreme event will occur on a particular day in a particular location. Rather, one is interested in the expected number of extreme events exceeding a certain threshold over the next year. For this purpose one needs very long time series in order to estimate the waiting-time distribution between consecutive extreme events and the propensity of clustering of extreme events. This calls for alternative methods of predicting extreme events to complement the above computationally expensive prediction systems.

An attractive alternative to high-dimensional models is reduced order models. In a series of papers, Majda *et al.* [15–18] introduced a stochastic mode reduction strategy to systematically derive stochastic climate models. These models are nonlinear and have both additive and multiplicative noise components and are thus different from Langevin equations. This mode reduction strategy is strictly valid only in the limit of infinite time scale separation but has been shown also to perform well in systems with only small time scale separations. The stochastic mode reduction strategy has been successfully applied to realistic complex atmospheric circulation models

*chan1@bas.ac.uk

[20,21]. These studies verified the ability of the reduced order models to accurately reproduce average statistics like autocorrelation functions and probability density functions (PDFs) with as few as four modes.

The objective of this study is to evaluate the potential of the reduced order models to accurately reproduce extreme value statistics and predictive skill. Because the mode reduction strategy used is only strictly valid in systems with widely different time scales, which is seldom the case in realistic systems, we have to prove empirically the success of the reduced models *a posteriori*, and in this article we provide evidence that reduced order models indeed can be used for the prediction of extreme events in high-dimensional systems with small time scale separation. Of course, for practical real-world applications the reduced models need always to be carefully calibrated and extensively tested.

In Sec. II we introduce a simple nonlinear climate model and derive its reduced order version by applying the stochastic mode reduction strategy. Then we discuss the extreme value characteristics of the models in Sec. III, their predictability in Sec. IV, and their likely dynamical origin in Sec. V. Conclusions are given in Sec. VI.

II. NONLINEAR STOCHASTIC-DYNAMICAL MODEL

A. Stochastic mode reduction strategy

We use a four-mode stochastic climate model to examine the potential of the systematic stochastic mode reduction strategy to derive reduced order models which reproduce extreme events of the full dynamic models and to examine how well these extreme events can be predicted [17,22,23]. This simple stochastic model contains many of the important dynamical features of comprehensive climate models, though it is of much lower dimensionality.

The stochastic climate model we are using in this study has two variables, denoted (x_1,x_2), which we call climate modes. These two modes evolve more slowly than the other two modes, (y_1,y_2), which represent turbulent eddies and convective systems. The latter modes evolve much more rapidly than the climate modes and are not fully resolved in climate models. In realistic systems there would be many fast modes, and so in order to mimic their combined effect on the two slow climate modes, we include damping and stochastic forcing in the form of $-\frac{\gamma}{\varepsilon}\mathbf{y}+\frac{\sigma}{\sqrt{\varepsilon}}d\mathbf{W}$ in the equations for $\mathbf{y}$, where $\mathbf{W}$ denotes a Wiener process. This parametrization of the combined effect of the fast modes includes nondiagonal and nonlinear interactions of the fast modes [16,20]. This approximation is motivated by the fact that these fast modes are associated with turbulent energy transfers and strong mixing. In this study we do not require a detailed description of these processes because we are only interested in their effect on the slow resolved modes and not in their detailed evolution. The stochastic climate model is given by

$$dx_1 = \mu\big((-x_2(L_{12}+a_1x_1+a_2x_2)+d_1x_1+F_1) + L_{13}y_1+B^1_{123}x_2y_1+\big(B^2_{131}+B^2_{113}\big)x_1y_1\big)dt, \quad (1a)$$

$$dx_2 = \mu\big((+x_1(L_{21}+a_1x_1+a_2x_2)+d_2x_2+F_2) + L_{24}y_2+B^1_{213}x_1y_1+\big(B^3_{242}+B^3_{224}\big)x_2y_2\big)dt, \quad (1b)$$

$$dy_1 = \left(-L_{13}x_1+B^1_{312}x_1x_2+B^2_{311}x_1x_1+F_3-\frac{\gamma_1}{\varepsilon}y_1\right)dt + \frac{\sigma_1}{\sqrt{\varepsilon}}dW_1, \quad (1c)$$

$$dy_2 = \left(-L_{24}x_2+B^3_{422}x_2x_2+F_4-\frac{\gamma_2}{\varepsilon}y_2\right)dt+\frac{\sigma_2}{\sqrt{\varepsilon}}dW_2. \quad (1d)$$

The parameter ε controls the time-scale separation between the slow and the fast variables. Energy conservation of the nonlinear operator requires that $B^1_{123}+B^1_{213}+B^1_{312}=0$, $B^2_{131}+B^2_{113}+B^2_{311}=0$, and $B^3_{242}+B^3_{224}+B^3_{422}=0$, and energy is defined as $E=\frac{1}{2}|\mathbf{x}|^2$, where $\mathbf{x}=\binom{x_1}{x_2}$. The linear operator matrix L is skew symmetric. The climate and fast modes are nonlinearly coupled through triad and dyad interactions.

The deterministic part of the above model, (1), has the same functional form as a comprehensive climate model and their structural form is given by [16,20,21]

$$d\mathbf{z} = \mathbf{F}dt + L\mathbf{z}dt + B(\mathbf{z},\mathbf{z})dt, \quad \mathbf{z}=\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}, \quad (2)$$

where $\mathbf{x}$ denotes the slow and $\mathbf{y}$ the fast components. The structural form of the fast dynamics can be written as [16]

$$d\mathbf{y} = \mathbf{F}_y dt + L_y\mathbf{y}dt + L_x\mathbf{x}dt + B_{yxx}(\mathbf{x},\mathbf{x})dt + B_{yxy}(\mathbf{x},\mathbf{y})dt + B_{yyy}(\mathbf{y},\mathbf{y})dt \quad (3)$$

The stochastic mode reduction strategy now assumes that the explicit nonlinear self-interaction through $B_{yyy}(\mathbf{y},\mathbf{y})$ of the fast variables $\mathbf{y}$ can be represented by a linear stochastic process [16]:

$$B_{yyy}(\mathbf{y},\mathbf{y})dt \sim -\frac{\gamma}{\varepsilon}\mathbf{y}+\frac{\sigma}{\sqrt{\varepsilon}}d\mathbf{W}. \quad (4)$$

The validity and success of this assumption for climate models have been shown in previous studies [15–17,19–23]. Furthermore, the approximation in Eq. (4) has been relaxed in the seamless reduction procedure in Ref. [20]. But for the purposes of this paper we prefer to derive the reduced order model directly on the level of the stochastic differential equations, which allows a more intuitive understanding of the procedure.

Consistent with comprehensive climate models, the stochastic climate model has a quadratically nonlinear part that conserves energy [denoted B in Eq. (2)], a linear operator [denoted L in Eq. (2)], and a forcing [denoted F in Eq. (2)]. The linear operator has two contributions: one is a skew-symmetric part formally similar to the Coriolis effect and Rossby wave propagation; the other is a negative definite symmetric part formally similar to dissipative processes, such as surface drag and radiative damping. Explicit inclusion of linear nondiagonal components (e.g., topography) would be straightforward but is not done here in order to keep the model relatively simple.

We now apply the systematic stochastic mode reduction procedure [15–17,20–23] to the model, (1), to obtain an explicit reduced stochastic equation for the slow variables $\mathbf{x}$.

See [15–17] for a detailed derivation of the reduced order equations. In the limit that $\varepsilon \to 0$ the reduced order stochastic model is given by

$$
\begin{aligned}
dx_1(t) = {} & (-x_2(t)(L_{12} + a_1 x_1(t) + a_2 x_2(t)) + d_1 x_1(t) + F_1)dt \\
& + \frac{\varepsilon}{\gamma_1}\big(L_{13}F_3 - L_{13}L_{13}x_1(t) + B^1_{123}F_3 x_2(t) \\
& + L_{13}B^1_{312}x_1(t)x_2(t) - L_{13}B^1_{123}x_1(t)x_2(t) \\
& + L_{13}B^2_{311}x_1^2(t) + B^1_{312}B^1_{123}x_1(t)x_2^2(t) \\
& + B^1_{123}B^2_{311}x_2(t)x_1^2(t) + \big(B^2_{131} + B^2_{113}\big)\big(B^2_{311}x_1^3(t) \\
& - L_{13}x_1^2(t) + B^1_{312}x_1^2(t)x_2(t) + F_3x_1(t)\big)\big)dt \\
& + \varepsilon\frac{1}{2}\frac{\sigma_1^2}{\gamma_1^2}\big(B^1_{213}B^1_{123}x_1(t) + \big(L_{13} + B^1_{123}x_2(t) \\
& + \big(B^2_{131} + B^2_{113}\big)x_1(t)\big)\big(B^2_{131} + B^2_{113}\big)\big)dt \\
& \times \sqrt{\varepsilon}\frac{\sigma_1}{\gamma_1}\big(L_{13} + B^1_{123}x_2(t) + \big(B^2_{131} + B^2_{113}\big)x_1(t)\big) \\
& \times dW_1(t),
\end{aligned}
$$

$$
\begin{aligned}
dx_2(t) = {} & (x_1(t)(L_{21} + a_1x_1(t) + a_2x_2(t)) + d_2x_2(t) \\
& + F_2)\,dt + \frac{\varepsilon}{\gamma_2}\big((L_{24}F_4 - L_{24}L_{24}x_2(t) \\
& + L_{24}B^3_{422}x_2^2(t) + \big(B^3_{242} + B^3_{224}\big)\big(B^3_{422}x_2^3(t) \\
& - L_{24}x_2^2(t) + F_4x_2(t)\big)\big)dt \\
& + \frac{\varepsilon}{\gamma_1}\big(-B^1_{213}L_{13}x_1(t)x_1(t) + B^1_{213}B^2_{311}x_1^3(t) \\
& + B^1_{213}B^1_{312}x_1(t)x_1(t)x_2(t) + B^1_{213}F_3x_1(t)\big)\,dt \\
& + \varepsilon\frac{1}{2}\frac{\sigma_1^2}{\gamma_1^2}\big(B^1_{213}B^1_{123}x_2(t) + L_{13}B^1_{213} \\
& + \big(B^2_{131} + B^2_{113}\big)B^1_{213}x_1(t)\big)\,dt \\
& + \varepsilon\frac{1}{2}\frac{\sigma_2^2}{\gamma_2^2}\big(L_{24} + \big(B^3_{242} + B^3_{224}\big)x_2(t)\big) \\
& \times \big(B^3_{242} + B^3_{224}\big)\,dt + \sqrt{\varepsilon}\frac{\sigma_1}{\gamma_1}B^1_{213}x_1(t)dW_1(t) \\
& + \sqrt{\varepsilon}\frac{\sigma_2}{\gamma_2}\big(L_{24} + \big(B^3_{242} + B^3_{224}\big)x_2(t)\big)\,dW_2(t). \qquad (5)
\end{aligned}
$$

Equation (5) is in Ito form.

To highlight the new structural form of the reduced model we rewrite Eq. (5) in compact functional form:

$$
\begin{aligned}
d\mathbf{x} = {} & \tilde{F}dt + \tilde{L}\mathbf{x}dt + \tilde{B}(\mathbf{x},\mathbf{x})dt + \tilde{M}(\mathbf{x},\mathbf{x},\mathbf{x})dt \\
& + \tilde{\sigma}_1 dW_1 + \tilde{\sigma}_2(\mathbf{x})dW_2. \qquad (6)
\end{aligned}
$$

The reduced model has two structurally new terms: (i) a cubic term $\tilde{M}$ and (ii) a multiplicative noise term $\tilde{\sigma}_2(\mathbf{x})dW_2$. Both of these terms stem from the nonlinear coupling between the climate and fast modes [18]. The reduced model, (5), has so-called correlated additive and multiplicative (CAM) noise [18,24],

$$
\tilde{\sigma}_2(\mathbf{x})dW_2 = (a + bx_2)dW_2, \qquad (7)
$$

where $a = \frac{\sigma_2}{\gamma_2}L_{24}$ and $b = \frac{\sigma_2}{\gamma_2}(B^3_{242} + B^3_{224})$. In CAM noise the same noise realization acts in both an additive and a multiplicative fashion at the same time. CAM noise produces non-Gaussian statistics with heavy tails and has been shown to be important in climate [18,24]. CAM noise is one possible mechanism for producing extreme events. But CAM noise only effectively models the deterministic nonlinear interactions of the full dynamics.

B. Numerical simulations

A 10-member ensemble is created by integrating the model for 10^5 time units starting from different initial conditions, which were chosen randomly. To integrate the model in time we use a fourth-order Runge-Kutta scheme for the deterministic part and a Euler forward scheme for the stochastic

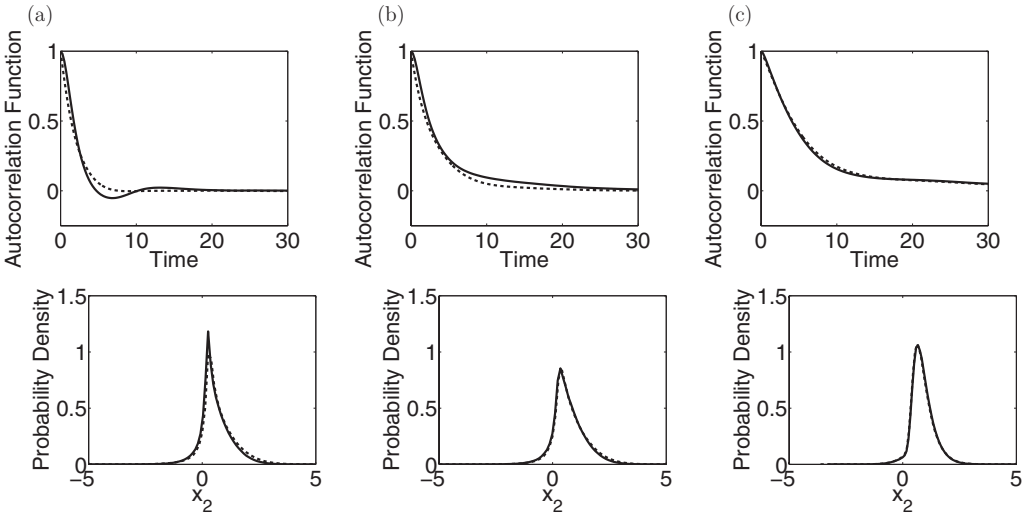

FIG. 1. Autocorrelation function (upper row) and marginal PDFs of mode x_2. Solid line: full dynamics, Eq. (1). Dashed line: reduced dynamics, Eq. (5). (a) $\varepsilon = 1.0$, (b) $\varepsilon = 0.5$, and (c) $\varepsilon = 0.1$.

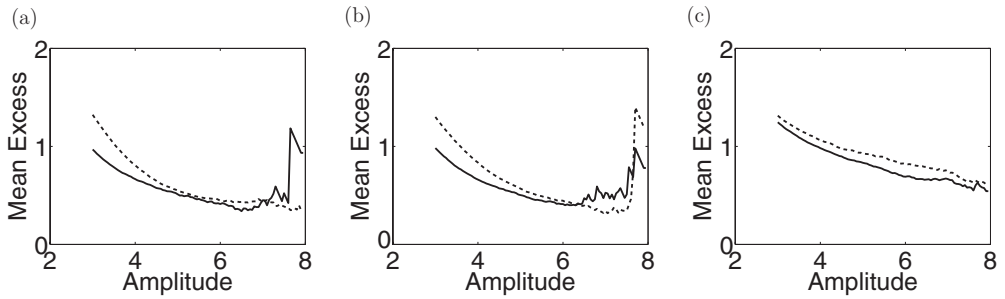

FIG. 2. Mean excess curves: full dynamics (solid line) and reduced dynamics (dashed line). (a) $\varepsilon = 1.0$, (b) $\varepsilon = 0.5$, and (c) $\varepsilon = 0.1$.

part. We use a time step of 10^{-5} time unit and data are saved every 0.0005 time unit. The following parameter values have been used: $\mu = 1$, $B^1_{123} = 4.0$, $B^1_{231} = 4.0$, $B^1_{312} = -8.0$, $B^2_{131} = 0.25$, $B^2_{113} = 0.25$, $B^2_{311} = -0.5$, $B^3_{242} = -0.3$, $B^3_{224} = -0.4$, $B^3_{422} = 0.7$, $L_{13} = -L_{24} = -0.2$, $\omega = 1.0$, $a_1 = 1.0$, $a_2 = -1.0$, $d_1 = -0.2$, $d_2 = -0.1$, $\gamma_1 = \gamma_2 = 1.0$, $\sigma_1 = 3.0$, and $\sigma_2 = 1.0$. The forcing vector (F_1, F_2, F_3, F_4) is given by $(-0.25, 0, 0, 0)$ and the bare truncation has two stable fixed equilibrium points [17,23]. The bare truncation is the part of Eqs. (1a) and (1b) which involves only the slow variables (x_1, x_2) and the forcing F. An extensive discussion of the bifurcation structure of the bare truncation is given in Ref. [17]. This set of parameter values was chosen because the model then produces extreme values as shown below. Our main results are fairly insensitive to the exact model parameter values.

To demonstrate the ability of the reduced model to reproduce key characteristics of the full model we compute the autocorrelation function (Fig. 1) and marginal PDF (Fig. 1) for mode x_2 (mode x_1 performs very similarly). The reduced model reproduces the full dynamics very accurately for $\varepsilon = 0.1$ and increasingly less acurately for $\varepsilon = 0.5$ and $\varepsilon = 1.0$.

For all three ε values the autocorrelation time scale is very well captured, and also the highly non-Gaussian features of the marginal PDFs are very well reproduced by the reduced models. The reduced models also reproduce third- and fourth-order two-time statistics [19,20] very well (not shown). The largest discrepancy in the statistics is for $\varepsilon = 1.0$. This is to be expected because the stochastic mode reduction strategy is only valid in the asymptotic limit that $\varepsilon \to 0$. Similar results of the stochastic mode reduction strategy have been reported in Refs. [17,22,23].

A visual inspection of Fig. 1 reveals that the PDFs are highly non-Gaussian. The PDFs are highly skewed and kurtotic and have heavy tails. This suggests that the toy model produces extreme events. Whether the extreme values follow an extreme value distribution is investigated next.

III. EXTREME EVENT CHARACTERISTICS

There are two common approaches in extreme value statistics: block maxima and threshold exceedances. In the block maxima approach one examines the maximum value over a fixed period; in climate applications this is usually 1 year. This approach has the disadvantage that many high-amplitude events get discarded. Here we use the threshold exceedance approach. In this approach extreme events are defined as events which are larger than a given fixed threshold. This is also the more relevant approach for many practical applications; e.g., one is usually interested in the prediction of the number of hurricanes and European windstorms exceeding a certain threshold which will occur in the next season.

A. Generalized Pareto distribution

In order to examine the extreme value characteristics of the stochastic climate models we use a threshold exceedance approach and fit a generalized Pareto distribution (GPD [25]) whose PDF is given by

$$f_{(\xi,\mu,\sigma)}(x) = \frac{1}{\sigma}\left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{\left(-\frac{1}{\xi}-1\right)}, \qquad (8)$$

where ξ denotes the shape parameter, μ the threshold (or location parameter), and σ the scale parameter. The shape and scale parameters are fitted with a standard maximum likelihood approach [25]. The GPD is generalized in the sense that it contains three special cases: (i) when $\xi > 0$ the GPD is equivalent to an ordinary Pareto distribution, (ii) when $\xi = 0$ the GPD becomes an exponential distribution, and (iii) for $\xi < 0$ the GPD is a short-tailed Pareto type II distribution.

TABLE I. GPD parameter estimates, with 95% confidence bounds given in parentheses.

	Full dynamics		Reduced dynamics	
ε	Shape parameter	Scale parameter	Shape parameter	Scale parameter
0.1	−0.16 (−0.14, −0.17)	1.15 (1.11, 1.18)	−0.14 (−0.12, −0.16)	1.22 (1.19, 1.26)
0.5	−0.14 (−0.12, −0.15)	0.76 (0.74, 0.79)	−0.16 (−0.15, −0.17)	0.97 (0.95, 0.99)
1.0	−0.14 (−0.13, −0.16)	0.77 (0.75, 0.79)	−0.17 (−0.17, −0.18)	0.95 (0.94, 0.97)

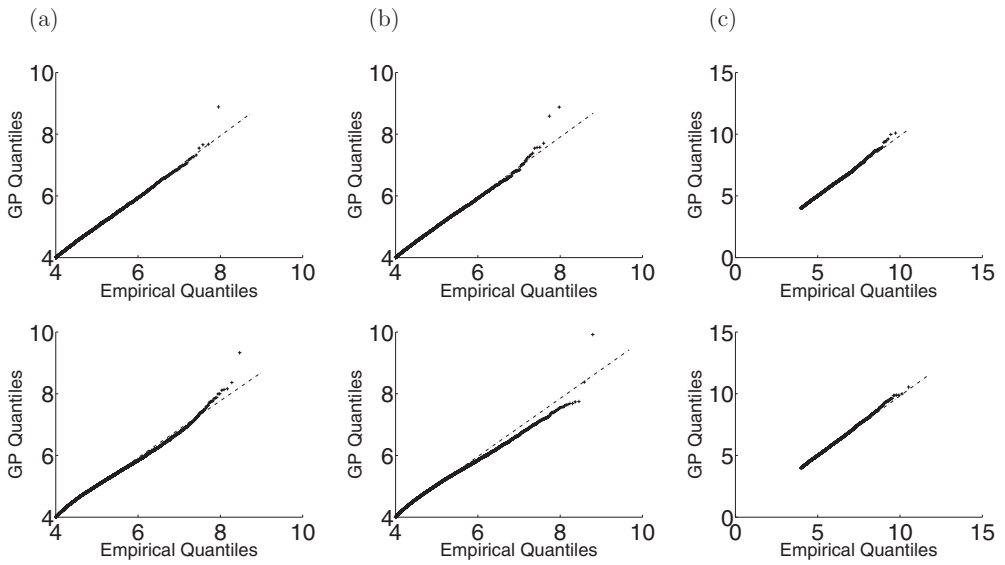

FIG. 3. Quantile-quantile plots: full dynamics (upper row) and reduced dynamics (lower row). (a) $\varepsilon = 1.0$, (b) $\varepsilon = 0.5$, and (c) $\varepsilon = 0.1$.

In order to identify the threshold above which the GPD applies, we compute the mean excess function [26]:

$$M(\mu) = \frac{\sum_{i=1}^{n}(x_i - \mu)I_{[x_i \geq \mu]}}{\sum_{i=1}^{n} I_{[x_i \geq \mu]}}. \tag{9}$$

If the data x_i follow a GPD above the threshold μ, then $M(\mu)$ is a linear function. Figure 2 shows that for $\varepsilon = 0.1$ both the full and the reduced models have approximately linear mean excess functions for values larger than $\mu = 4$. The mean excess functions are also approximately linear above $\mu = 4$ for $\varepsilon = 0.5$ and $\varepsilon = 1.0$ for both the full and the reduced dynamics. In order to ensure that the data points used are independent (or at least well decorrelated), we identified the maxima of the time series above the threshold μ and consider only maxima which are at least 100 sample points apart from each other (our results are not sensitive to this particular value).

The parameter values for the GPD model are given in Table I. For $\varepsilon = 0.1$ the parameter estimates for the full dynamics are inside the error bounds of the reduced model. For $\varepsilon = 0.5$ and $\varepsilon = 1.0$ the shape parameters are just outside the error bounds. However, as we show below, the reduced models nonetheless reproduce the extreme values of the full dynamics well.

Using the GPD parameter values we generate variates from the extreme value distribution and compare them with the model variable x_2 in the form of quantile-quantile (q-q) plots. A q-q plot is a graphical tool to decide if two data sets stem from the same distribution. This is done by plotting the quantiles which denote the point below which a given percentage of points lies against the quantiles of the second data set. If both data sets come from the same distribution, the points will follow the 45° reference line. As Fig. 3 shows, for all three ε values the functions are very well approximated by the straight 45° reference line. For large values deviations from the straight-line fit are visible but this is to be expected because the large values are also very rare and the statistic suffers from sampling issues. This indicates that $\mu = 4$ is indeed a good threshold value and that x_2 follows a GPD well. We get qualitatively very similar results for x_1 and minima. This shows that the extreme events of both the full and the reduced dynamics follow a GPD.

We use also q-q plots to verify that the reduced order models reproduce well the extreme value characteristics of the full models. As Fig. 4 shows, the reduced models reproduce the extremes very well for $\varepsilon = 0.1$ and $\varepsilon = 1.0$ and reasonably well for $\varepsilon = 0.5$. This shows that reduced order models are able to reproduce the extreme value characteristics of higher dimensional models, even though for moderate or no time-scale separation the GPD parameter estimates are outside the error bounds. This is an important result

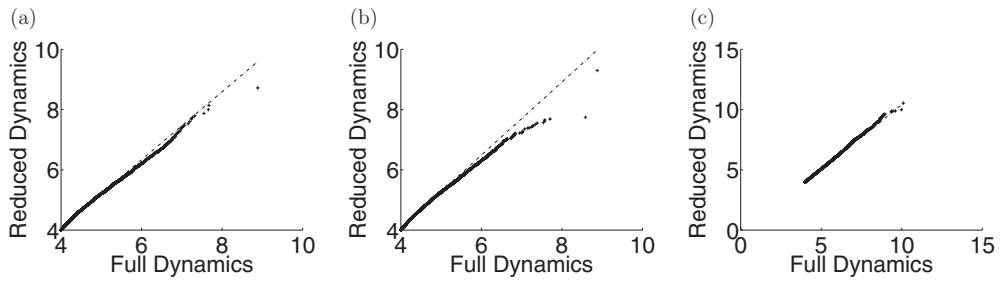

FIG. 4. Quantile-quantile plots comparing the extreme value distributions of the full and reduced dynamics. (a) $\varepsilon = 1.0$, (b) $\varepsilon = 0.5$, and (c) $\varepsilon = 0.1$.

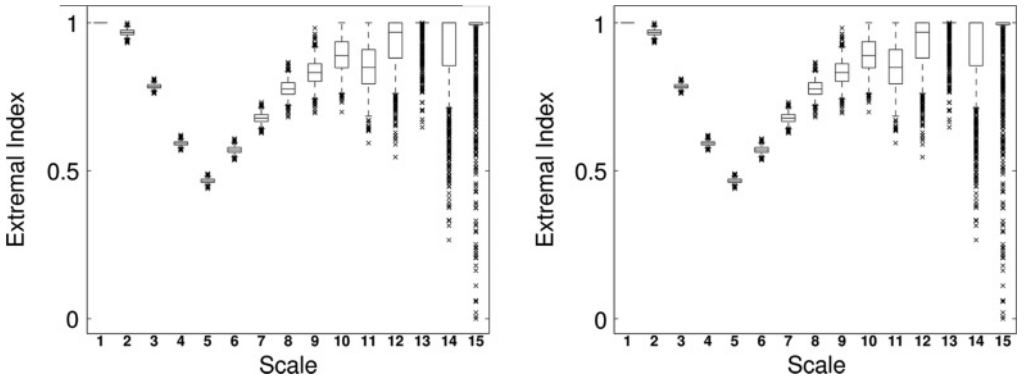

FIG. 5. Extremal index of x_2 for $\varepsilon = 0.1$: full dynamics (left) and reduced dynamics (right). In each box the central mark is the median, the edges of the boxes are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered to be outliers, and outliers are marked individually.

because the investigation of extreme values requires very long time series and most high-dimensional complex models are computationally very expensive. These results offer the potential of improved extreme predictions with reduced order models.

As Table I reveals, all models have a negative shape parameter. This indicates that the distribution of excesses has an upper bound [25]. This is consistent with recent studies of temperature [27] and windspeed (Franzke, 2012, in preparation). This is also qualitatively consistent with the study by Majda *et al*. [18], which shows that while the normal form of stochastic climate models allows for a power-law-like decay of the PDF tail over some range of values, the ultimate decay is exponential; thus very large values have a vanishing probability. This is due to energy constraints of the nonlinear operator which are obeyed by both the full dynamics and the reduced models.

B. Clustering of extreme events

The stochastic model, Eq. (1), also exhibits clustering of extreme events as revealed by the extremal index θ [28]. θ characterizes the extend of temporal dependency of extreme events. θ is inversely proportional to the average cluster size. The approach of [28] is based on the asymptotic scaling properties of block maxima and resampling. The maxima of blocks of size m scale as $m^{\frac{1}{\alpha}}$, where α is the tail exponent. Thus, by examining a sequence of dyadic block sizes $m(j) = 2^j$ and resampling, one can estimate the extremal index $\theta(j)$ and the corresponding uncertainty bounds (see [28] for more details). Evidence for clustering of extremes is given if θ turns out to be stable over a range of scales. An extremal index value close to 1 indicates almost-independent extremes.

Figure 5 shows the extremal index $\theta(j)$ for different block sizes. For the model simulation with $\varepsilon = 0.1$, it can be seen that for block sizes $j = 9$ through $j = 12$ one gets a stable estimate of about 0.88, which is significantly different from 1. Thus, the model exhibits clustering of extreme events which is reproduced by the reduced model. The θ values of both models lie inside the error bounds. The simulations with larger ε have extremal indices very close to 1 and thus show no evidence of clustering of extreme events. This lack of clustering is also reproduced by the reduced models. This difference between the case $\varepsilon = 0.1$ and the case $\varepsilon = 0.5, 1.0$ indicates that, for moderate time-scale separations, the dynamical system, Eq. (1), has not reached its limiting dynamics. Our results show that for this system only the limiting dynamics exhibit clustering of extreme events.

IV. PREDICTABILITY

In this section we examine the predictability of extreme events in our models. We also examine whether the reduced models have the same predictive power as the full model. In order to quantify predictability we use the receiver-operator characteristic (ROC) curve [12,29]. This curve displays the true-positive forecast rates (hit rate) against the false-positive rate (false-alarm rate). The optimal prediction model will be

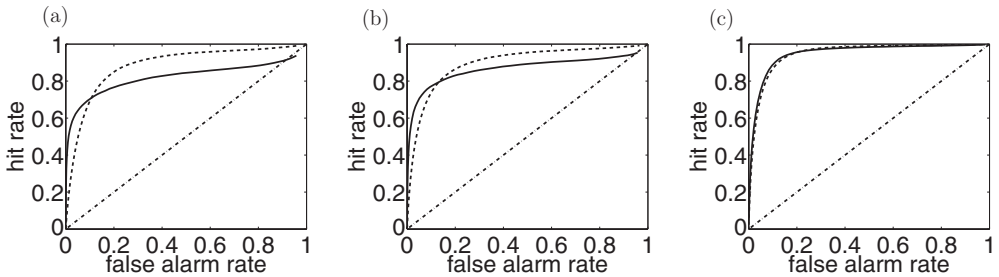

FIG. 6. Receiver-operator characteristic curves for the precursors. Full dynamics (solid line) and reduced dynamics (dashed line). The threshold is 4. (a) $\varepsilon = 1.0$, (b) $\varepsilon = 0.5$, and (c) $\varepsilon = 0.1$.

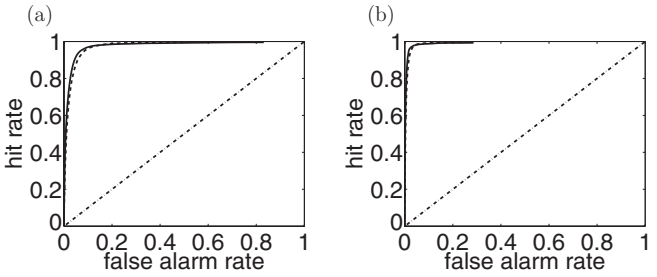

FIG. 7. Receiver-operator characteristic curves for $\varepsilon = 0.1$ and threshold (a) $\mu = 5$ and (b) $\mu = 6$. Full dynamics (solid line) and reduced dynamics (dashed line).

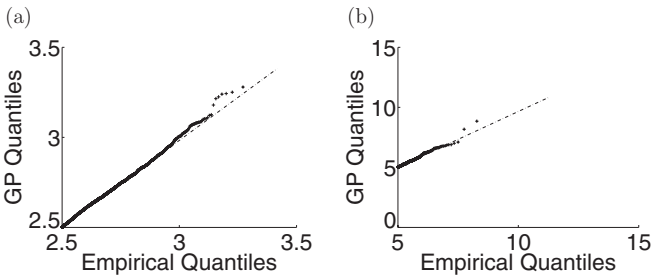

FIG. 8. Quantile-quantile plots: (a) model simulation with no multiplicative triads active and (b) model simulation with no bare truncation active (case with $\varepsilon = 0.1$).

located in the upper left corner (0,1) of the ROC space, i.e., 100% true-positive forecasts and 0% false-positive forecasts. Points along the diagonal line from the lower left to the upper right corner indicate random forecast guesses. This diagonal line divides the ROC space. Points above this line indicate good forecasts, while points below the line represent poor forecasts. ROC curves are particularly suitable for validating the predictive skill of rare, extreme events of models [12], provided that there are sufficiently many extreme events to reliably compute the ROC statistics. ROC curves do not explicitly depend on the frequency of occurrence of extreme events as do some other skill scores.

For the forecast experiments we first generate independent model time-series realizations by starting the models with random initial values. We first identify the probability distribution of precursors which precede events above a given threshold μ, similarly to [11] and [12] in the control runs. For this purpose we bin all states which precede a threshold crossing. We consider only univariate extremes and focus on x_2 but the precursor is two-dimensional, containing both x_1 and x_2. Then we compute the maximum of the joint precursor probability function $x_i^{\max}$. The precursor PDF is unimodal (not shown). To compute the ROC curves we follow the approach used in Ref. [12]. We vary the allowed distance between the actual state of the model and the precursor PDF maximum,

$$D = \sqrt{\left(x_1^{\text{pre}} - x_1^{\max}\right)^2 + \left(x_2^{\text{pre}} - x_2^{\max}\right)^2}, \qquad (10)$$

from 0 in increments of 0.1 for 50 increments starting from 0.1, where x_i^{pre} is the current model state. Whenever D is smaller than the threshold we anticipate that an extreme event will occur.

As Fig. 6 reveals, extreme events are predictable. For all three cases the ROC curves are well above the diagonal line, indicating good predictive skill. Furthermore, the reduced model does an excellent job of predicting extreme events for $\varepsilon = 0.1$. The predictive skill of the reduced model is still good for $\varepsilon = 0.5$ and reasonable for $\varepsilon = 1.0$. Figure 7 shows that the predictive skill is higher for larger extreme events. This is consistent with the findings of [12] and [13], where they examined, among other things, observed wind gusts.

V. DYNAMICS OF EXTREME EVENTS

Are the extreme events due to nonlinear interactions of the slow modes alone or due to coupling with the fast modes and hence CAM noise in the reduced models? To address this question we carried out sensitivity experiments where either the bare truncation is set to 0 ($\mu = 0$) or the nonlinear coupling between the slow and the fast modes is set to 0 ($B^1 = B^2 = B^3 = 0$). In the former case extreme events are caused by nonlinear interactions in the full dynamics or CAM noise in the reduced dynamics. In the latter case extreme events are due to the deterministic nonlinearity of the slow x modes, which is driven by additive white noise.

As Fig. 8 shows, both of these experiments also exhibit extreme value behavior which again follows a GPD. This shows that both the deterministic nonlinearity of the bare truncation and the nonlinear interaction between the slow and the fast modes can cause extreme events on their own. However, the magnitude of the extreme events is larger in the experiments with the bare truncation set to 0 ($\mu = 0$).

Furthermore, predictability experiments show that both models as well as their reduced dynamics again show good predictive skill for forecasting extreme events (Fig. 9). This suggests that both the bare truncation and the nonlinear interactions between slow and fast modes (or CAM noise for the reduced dynamics) are important for the extreme event behavior of this model.

VI. SUMMARY AND DISCUSSION

We have investigated the extreme value and predictability characteristics of a nonlinear stochastic-dynamical model which is a toy version of a climate model and its reduced order version. We find that the extreme values of the model used follow an extreme value distribution, a GPD. The shape parameter of the GPD is negative; thus extreme values are bounded in the used model. This extreme value

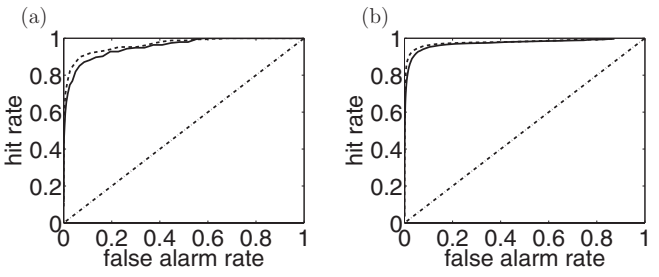

FIG. 9. Receiver-operator characteristic curves: (a) model simulation with no multiplicative triads active and (b) model simulation with no bare truncation active (case with $\varepsilon = 0.1$).

behavior was predicted for stochastic climate models by Majda *et al*. [18] based on the normal forms of stochastic climate models.

Then we have examined whether the corresponding reduced order models of the nonlinear full dynamical model reproduce the extreme value characteristics. We find that the extreme value characteristics are reproduced extremely well in the case of time-scale separation and are also good for moderate and even no time-scale separation. We have also shown that extreme events are better predictable the larger they are. This is consistent with the findings of [11–13] for various simple stochastic processes and wind gusts. While reduced order models usually have problems capturing the correct decay of the autocorrelation function, it seems to be much easier for reduced order models to capture the PDFs even in situations without time-scale separation. Our results show the potential role that systematically derived nonlinear reduced order models can play in practical applications like natural hazard predictions and insurance pricing. Of course, the reduced models need to be carefully calibrated and extensively tested for each new application.

While [24] put forward the idea that the amplitude distribution of atmospheric extreme events follows a power-law decay due to CAM noise, our results suggest that the deterministic nonlinearity will prevent fluctuations from becoming arbitrarily large. This is due to the energy conservation of the nonlinear operator, which makes the models nonlinearly stable. It has been shown that in the reduced models the cubic term acts as a nonlinear damping [18]. A criterion for this nonlinear stability was derived in Ref. [18]. These stability criterion can be used for parameter inference of reduced order models from observed data.

All results in this study are for a simple stochastic toy model. It remains to be seen whether the reported results are also valid for more complex models. We also have considered only stationary processes. Future studies will investigate how well the reduced order models perform for time-varying forcing. This is especially important in the context of global warming, where the IPCC projects an increase in the frequency and intensity of climate extremes.

ACKNOWLEDGMENTS

I would like to thank two anonymous reviewers whose comments improved an early version of the manuscript. This study is part of the British Antarctic Survey Polar Science for Planet Earth Programme. It was funded by The Natural Environment Research Council.

[1] K. E. Trenberth, P. D. Jones, P. Ambenje, R. Bojariu, D. Easterling, A. Klein Tank, D. Parker, F. Rahimzadeh, J. A. Renwick, M. Rusticucci, B. Soden, and P. Zhai, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller (Cambridge University Press, Cambridge, 2007).

[2] Press release by Munich Re; available at: [www.munichre.com/en/media_relations/press_releases/2011/2011_07_12_press_ release.aspx].

[3] P. J. Mailier, D. B. Stephenson, C. A. T. Ferro, and K. I. Hodges, Month. Weather Rev. **134**, 2224 (2006).

[4] A. Bunde, J. F. Eichner, J. W. Kantelhardt, and S. Havlin, Phys. Rev. Lett. **94**, 048701 (2005).

[5] A. C. M. Freitas and J. M. Freitas, Stat. Prob. Lett. **78**, 1088 (2008).

[6] C. Nicolis, V. Balakrishnan, and G. Nicolis, Phys. Rev. Lett. **97**, 210602 (2006).

[7] C. Nicolis and G. Nicolis, Phys. Rev. E **80**, 061119 (2009).

[8] C. Nicolis and S. C. Nicolis, Europhys. Lett. **80**, 40003 (2007).

[9] T. Schweigler and J. Davidsen, Phys. Rev. E **84**, 016202 (2011).

[10] R. Vitolo, M. P. Holland, and C. A. T. Ferro, Chaos **19**, 043127 (2009).

[11] S. Hallerberg, E. G. Altmann, D. Holstein, and H. Kantz, Phys. Rev. E **75**, 016706 (2007).

[12] S. Hallerberg, J. Bröcker, and H. Kantz, *Prediction of Extreme Events. Nonlinear Time Series Analysis in the Geosciences, Lecture Notes in Earth Sciences* (Springer, New York, 2008).

[13] S. Hallerberg and H. Kantz, Phys. Rev. E **77**, 011108 (2008).

[14] R. Buizza, P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, and Y. Zhu, Month. Weather Rev. **133**, 1076 (2005).

[15] A. J. Majda, I. Timofeyev, and E. Vanden-Eijnden, Proc. Natl. Acad. Sci. USA **96**, 14687 (1999).

[16] A. J. Majda, I. Timofeyev, and E. Vanden-Eijnden, Commun. Pure Appl. Math. **54**, 891 (2001).

[17] A. J. Majda, R. Abramov, and M. Grote, *Information Theory and Stochastics for Multiscale Nonlinear Systems. CRM Monograph Series* (American Mathematical Society, Providence, RI, 2005).

[18] A. J. Majda, C. Franzke, and D. Crommelin, Proc. Natl. Acad. Sci. USA **106**, 3649 (2009).

[19] A. J. Majda, I. Timofeyev, and E. Vanden-Eijnden, Physica D **170**, 206 (2002).

[20] C. Franzke, A. J. Majda, and E. Vanden-Eijnden, J. Atmos. Sci. **62**, 1722 (2005).

[21] C. Franzke and A. J. Majda, J. Atmos. Sci. **63**, 457 (2006).

[22] A. J. Majda, C. Franzke, and B. Khouider, Phil. Trans. Roy. Soc. A **366**, 2429 (2008).

[23] C. Franzke, A. J. Majda, and G. Branstator, J. Atmos. Sci. **64**, 3987 (2007).

[24] P. D. Sardeshmukh and P. Sura, J. Climate **22**, 1193 (2009).

[25] S. Coles, *An Introduction to Statistical Modelling of Extreme Values* (Springer, London, 2001).

[26] S. Ghosh and S. Resnick, Stoch. Process. Appl. **120**, 1492 (2010).

[27] C. Franzke, Int. J. Climatol. (in press, 2012).

[28] K. Hamidieh, S. Stoev, and G. Michailidis, J. Comput. Graph. Stat.**18**, 731 (2009).

[29] T. Fawcett, Technical Report HPL-2003-4, Palo Alto, CA: HP Laboratories.