Clim Dyn
DOI 10.1007/s00382-014-2239-9

Simulating weather regimes: impact of stochastic and perturbed parameter schemes in a simple atmospheric model

H. M. Christensen · I. M. Moroz · T. N. Palmer

Received: 31 March 2014 / Accepted: 2 July 2014

Abstract Representing model uncertainty is important for both numerical weather and climate prediction. Stochastic parametrisation schemes are commonly used for this purpose in weather prediction, while perturbed parameter approaches are widely used in the climate community. The performance of these two representations of model uncertainty is considered in the context of the idealised Lorenz '96 system, in terms of their ability to capture the observed regime behaviour of the system. These results are applicable to the atmosphere, where evidence points to the existence of persistent weather regimes, and where it is desirable that climate models capture this regime behaviour. The stochastic parametrisation schemes considerably improve the representation of regimes when compared to a deterministic model: both the structure and persistence of the regimes are found to improve. The stochastic parametrisation scheme represents the small scale variability present in the full system, which enables the system to explore a larger portion of the system's attractor, improving the simulated regime behaviour. It is important that temporally correlated noise is used in the stochastic parametrisation—white noise schemes performed similarly to the deterministic model. In contrast, the perturbed parameter ensemble was unable to capture the regime structure of the attractor, with many individual members exploring only one regime. This poor performance was not evident in other climate diagnostics. Finally, a 'climate change' experiment was performed, where a change in external forcing resulted in changes to the regime structure of the attractor. The temporally correlated stochastic schemes captured these changes well.

Keywords Weather regimes · Stochastic physics · Perturbed parameter schemes · Model uncertainty · Lorenz '96 system · Climate change

1 Introduction

The presence of regimes is a characteristic of non-linear, chaotic systems (Lorenz 2006). In the atmosphere, regimes emerge as familiar persistent circulation patterns such as the Pacific/North American (PNA) pattern, the North Atlantic Oscillation (NAO) and Scandinavian Blocking events. More generally, a regime can be defined as "a region of state space that is more populated than neighbouring regions" (Stephenson et al. 2004). Identifying this localised clustering in state space is a non-trivial statistical problem (Stephenson et al. 2004), but can be achieved using a clustering algorithm such as k-means clustering (Dawson et al. 2012; Pohl and Fauchereau 2012; Straus et al. 2007) or by estimating the pdf of the distribution and searching for multiple maxima (Kimoto and Ghil 1993a, b; Corti et al. 1999).

In recent years there has been much interest in the problem of identifying and studying atmospheric regimes (Charney and DeVore 1979; Palmer 1993, 1999; Selten and Branstator 2004). In particular, there is much interest in how the climate system responds nonlinearly to an external forcing (such as anthropogenic greenhouse gas emissions) through a modification of its regime behaviour (Branstator and Selten 2009). An attractor with regime structure could respond to a forcing in two possible ways. Hasselmann

H. M. Christensen (✉) · T. N. Palmer
Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Oxford, UK
e-mail: h.m.christensen@atm.ox.ac.uk

I. M. Moroz
Oxford Centre for Industrial and Applied Mathematics, University of Oxford, Oxford, UK

Published online: 17 July 2014

(1999) discusses the climate attractor in terms of a field with several potential wells, each of which represents a different atmospheric regime. The first possible response to an external forcing would be a change in the relative depths of the potential wells. This would lead to changes in both the relative residence times in the wells, and to the transition frequencies between regimes. Studying 20th Century reanalysis data indicates greenhouse gas forcing leads, in part, to this response in the climate system. Corti et al. (1999) observed changes in the frequency of Northern Hemisphere intraseasonal–interannual regimes between 1949 and 1994, though the structure of the regimes remained relatively unchanged over this time period.

The second possible response to an external forcing is a change in regime properties, such as centroid location and number of regimes (i.e. the position and number of potential wells). In addition to observing changes in frequency of regimes over the time period 1948–2002, Straus et al. (2007) observe this second type of response in the reanalysis data; the structure of the Pacific trough regime is statistically significantly different at the end of the time period than at the beginning.

The importance of regimes in observed trends over the past 50–100 years indicates that in order to predict anthropogenic climate change, our climate models must be able to accurately represent natural circulation regimes, their statistics and variability. Dawson et al. (2012) show that while numerical weather prediction (NWP) models are able to capture the regime behaviour of the climate system with reasonable accuracy, the same model run at climate resolution does not show any statistically significant regime structure. However, the model used in this study has no representation of model uncertainty; a single deterministic integration is made from each starting date.

It is now well established that representing model uncertainty as well as initial condition uncertainty is important for both reliable weather forecasts (Ehrendorfer 1997) and for indicating uncertainty in climate projections. Many possible methods for representing model uncertainty have been proposed, such as multi-model ensembles (e.g. DEMETER: Palmer et al. 2004) and multiparametrisation schemes (e.g. Houtekamer et al. 1996). Additionally, stochastic parametrisation schemes have also been developed and shown to improve the skill of weather forecast models (Wilks 2005; Crommelin and Vanden-Eijnden 2008; Berner et al. 2009; Palmer et al. 2009; Frenkel et al. 2012; Kwasniok 2012; Arnold et al. 2013). It is possible that including stochastic physics as a representation of model uncertainty could enable a climate simulator to explore larger regions of the climate attractor, including other flow regimes. An alternative but commonly used representation of model uncertainty is a perturbed parameter scheme, whereby physical parameters in subgrid parametrisation schemes are perturbed about their optimal value (Murphy et al. 2004; Stainforth et al. 2005; Rougier et al. 2009). Perturbing parameters gives a greater control over the ensemble than multi-model or multiparametrisation ensembles, and has been used as a representation of model uncertainty in climate prediction (Stainforth et al. 2005; Rougier et al. 2009). The impact of perturbed parameter representations of model uncertainty on the ability of a climate model to represent regime behaviour has not previously been studied.

This paper seeks to investigate the effect of including representations of model uncertainty on the regime behaviour of a simulator. A deterministic parametrisation scheme will be compared to stochastic parametrisation approaches and a perturbed parameter ensemble. An idealised chaotic model of the atmosphere, the Lorenz (1996) system, will be used to study regime behaviour (Lorenz 2006).

2 Data and methods

2.1 The Lorenz '96 system

The Lorenz '96 system is often used as a testbed for the development and testing of parametrisation schemes (Wilks 2005; Crommelin and Vanden-Eijnden 2008; Kwasniok 2012). There are many benefits of performing proof of concept experiments in a simple chaotic system before moving to a global circulation model (GCM): such systems are computationally cheap to run, allow for the robust definition of "truth" required for forecast verification, and yet are able to mimic certain properties of the atmosphere. The results from these studies can then be used to inform future developments in the atmospheric modelling community (Dorrestijn et al. 2012).

This paper uses the second model proposed in Lorenz (1996), henceforth the L96 system, which consists of a set of coupled equations in two variables; the large scale, low frequency X_k variables are each coupled to many small scale, high frequency Y_j variables. The governing equations for these variables are given by (1),

$$\begin{aligned}\frac{\mathrm{d}X_k}{\mathrm{d}t} &= -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F \cdots \\ &\quad - \frac{hc}{b}\sum_{j=J(k-1)+1}^{kJ} Y_j;\ k = 1, ..., K \\ \frac{\mathrm{d}Y_j}{\mathrm{d}t} &= -cbY_{j+1}(Y_{j+2} - Y_{j-1}) - cY_j \cdots \\ &\quad + \frac{hc}{b}X_{int[(j-1)/J]+1};\ j = 1, ..., JK,\end{aligned} \tag{1}$$

where the variables have cyclic boundary conditions; $X_{k+K} = X_k$ and $Y_{j+JK} = Y_j$. The interpretation of the

Table 1 Parameter settings used for the L96 system

Parameter	Symbol	Value
# X variables	K	8
# Y variables per X variable	J	32
Coupling constant	h	1
Forcing term	F	20
Spatial scale ratio	b	10
Timescale ratio	c	10

Table 2 Measured parameters defining the cubic polynomial, $(b_0^{meas}, b_1^{meas}, b_2^{meas}, b_3^{meas})$, and the variability of these parameters, $\sigma(b_i^{samp})$, calculated by sampling from the truth time series

	b_0	b_1	b_2	b_3
μ	0.341	1.30	−0.136	−0.00235
σ	0.146	0.0381	0.00901	0.000650

parameters in these equations, and the values used in this study, are shown in Table 1. The scaling of the variables is such that one model time unit is equal to five atmospheric days, deduced by comparing the error doubling time of the model to that observed in the atmosphere (Lorenz 1996).

A simulator of the L96 system was constructed by assuming that only the X variables are resolved. The effect of the sub-gridscale Y variables on the X variables must be parametrised in terms of the resolved X variables:

$$\frac{\mathrm{d}X_k}{\mathrm{d}t} = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F - U_p(X_k); \quad k = 1, ..., K, \tag{2}$$

where U_p is the parametrised subgrid tendency. The simplest parametrisation scheme assumes a deterministic relationship between the unresolved tendency and the resolved grid-scale variables:

$$U_p(X_k) = U_{det}(X_k) = b_0 + b_1 X_k + b_2 X_k^2 + b_3 X_k^3, \tag{3}$$

where the parameter values, (b_0, b_1, b_2, b_3), are determined by a least squares fit to the (X, U) truth data. The fitted values are shown in Table 2.

However, in the full set of governing equations, the state of the X variables does not uniquely determine the behaviour of the Y variables. The forecasts made using Eq. (3) will include an error due to the truncation and parametrisation of the L96 equations. A probabilistic forecast model, including a representation of model uncertainty, can give an indication of how large the forecaster predicts this error could be.

Two fundamentally different representations of model uncertainty are considered and are outlined below. For more details on the methodology, and for experiments using more complicated stochastic simulators, refer to Arnold et al. (2013).

Table 3 Stochastic parametrisations of the sub-grid tendency, U, used in this experiment, and the values of the model parameters fitted from the truth time series

Parametrisation	Measured parameters
Additive (A)	$\phi = 0.986$
	$\sigma = 1.99$
Multiplicative (M)	$\phi = 0.940$
	$\sigma = 0.469$

2.1.1 Stochastic parametrisation schemes

Two different stochastic parametrisation schemes will be discussed here, summarised in Table 3. Both include a stochastic term into the deterministic parametrisation scheme, Eq. (3), to represent the subgrid-scale variability due to the truncated Y variables.

The first scheme considered is an additive noise parametrisation scheme:

$$U_p(X_k) = U_{det}(X_k) + e_k(t), \tag{4}$$

where the stochastic additive term, $e_k(t)$, for each of the X_k is modelled as a first order autoregressive (AR(1)) process,

$$e_k(t) = \phi e_k(t - \Delta t) + \sigma(1 - \phi^2)^{\frac{1}{2}} z_k(t), \tag{5}$$

where z_k is a normally distributed random variable with unit variance. The optimal magnitude, σ and the temporal autocorrelation, ϕ of the stochastic term can be measured from the "truth" time series $(\sigma_{meas}, \phi_{meas})$: the residual, $r(t)$, between the true tendency, $U(X)$, and the parametrised tendency is calculated as

$$r_k(t) = U(X_k) - U_{det}(X_k), \tag{6}$$

and the characteristics of the stochastic term $e_k(t)$ are designed to represent the residuals $r_k(t)$. The measured values of the parameters are shown in Table 3.

The second scheme considered here is a multiplicative noise scheme. This scheme was motivated by the multiplicative Stochastically Perturbed Parametrisation Tendencies (SPPT) scheme operationally in use at the European Centre for Medium-Range Weather Forecasts (ECMWF), and which has been shown to be skilful (Palmer et al. 2009). The parametrisation proposed is

$$U_p(X_k) = (1 + e_k(t))U_{det}(X_k), \tag{7}$$

where $e_k(t)$ is modelled as an AR(1) process. In a similar way to the additive scheme, the parameters in the model, (σ, ϕ) can be estimated from the truth time series to give $(\sigma_{meas}, \phi_{meas})$, shown in Table 3.

Table 4 Chosen permutations for the perturbed parameter experiment

Selected permutations																			
#	b_0	b_1	b_2	b_3	#	b_0	b_1	b_2	b_3	#	b_0	b_1	b_2	b_3	#	b_0	b_1	b_2	b_3
1	H	H	H	H	11	H	L	H	L	21	M	M	L	H	31	L	H	L	L
2	H	H	H	L	12	H	L	M	M	22	M	M	L	L	32	L	M	H	M
3	H	H	M	M	13	H	L	L	H	23	M	L	H	M	33	L	M	M	H
4	H	H	L	H	14	H	L	L	L	24	M	L	M	H	34	L	M	M	L
5	H	H	L	L	15	M	H	H	M	25	M	L	M	L	35	L	M	L	M
6	H	M	H	M	16	M	H	M	H	26	M	L	L	M	36	L	L	H	H
7	H	M	M	H	17	M	H	M	L	27	L	H	H	H	37	L	L	H	L
8	H	M	M	L	18	M	H	L	M	28	L	H	H	L	38	L	L	M	M
9	H	M	L	M	19	M	M	H	H	29	L	H	M	M	39	L	L	L	H
10	H	L	H	H	20	M	M	H	L	30	L	H	L	H	40	L	L	L	L

H, M and L represent the High, Medium and Low settings respectively

Three parameter settings are of particular interest and will be considered in the following experiments:

1. The deterministic parametrisation (DET): $\sigma = 0$.
2. The white additive (WA) or white multiplicative (WM) stochastic parametrisations: $\sigma = \sigma_{meas}$, $\phi = 0$ for each case.
3. The measured AR(1) red additive or multiplicative (RA and RM respectively) stochastic parametrisations: $\sigma = \sigma_{meas}$, $\phi = \phi_{meas}$ for each case.

2.1.2 *Perturbed parameter scheme*

An alternative representation of model uncertainty is a perturbed parameter scheme. This explores the uncertainty due to assumptions built into the physical parametrisation schemes by varying the values of uncertain parameters within their physical range, generating an ensemble of deterministic parametrisations. A perturbed parameter ensemble is also tested as a representation of model uncertainty in the Lorenz '96 system. The skill of this representation of model uncertainty is evaluated as for the stochastic parametrisations.

Following Stainforth et al. (2005), each of the four parameters defining the cubic polynomial, $(b_0^{meas}, b_1^{meas}, b_2^{meas}, b_3^{meas})$, is set to one of three values: low (L), medium (M) or high (H). The degree to which the parameters should be varied was estimated from the truth time series: the measured $U(X)$ was split into sections 3 model time units (MTU) long,[1] and a cubic polynomial fitted to each section. The measured variability in each of the parameters was then calculated as the standard deviation of the parameters fitted to each section $\sigma(b_i^{samp})$. The measured standard deviations are shown in Table 2. The low, medium and high values of the parameters are given by:

$$\begin{aligned} L &= b_i^{meas} - \sigma(b_i^{samp}) \\ M &= b_i^{meas} \\ H &= b_i^{meas} + \sigma(b_i^{samp}), \end{aligned} \tag{8}$$

There are $3^4 = 81$ possible permutations of the parameter settings, from which a subset of 40 permutations was selected to sample the uncertainty. The selected permutations are shown in Table 4.

The same "truth" model is used as for the stochastic parametrisations, and the simulator is constructed in an analogous way: only the X variables are assumed resolved, and the effects of the unresolved sub-gridscale Y variables are represented by an ensemble of deterministic parametrisations:

$$U_{pp}(X_k) = b_0^p + b_1^p X_k + b_2^p X_k^2 + b_3^p X_k^3, \tag{9}$$

where the values of the perturbed parameters, b^p, vary between ensemble members.

2.2 Defining regimes in simple systems

Firstly, it needs to be established whether the L96 two-scale model exhibits regime behaviour. Lorenz (2006) carried out a series of experiments using the one-scale Lorenz '96 system (hereafter L96 1D), which describes the evolution of the L96 X_k variables, without the influence of the smaller scale Y_j variables (Lorenz 1996):

$$\frac{\mathrm{d}X_k}{\mathrm{d}t} = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F; \quad k = 1, \cdots, K \tag{10}$$

Lorenz (2006) defines a dynamical system as having regime behaviour if:

1. The phase space of the dynamical system has two separate regions, A and B.
2. Both transitions A–B and B–A are observed.
3. For both modes, the average length of time between transitions must be long compared to some other significant oscillation of the system.

[1] One model time unit corresponds to approximately 5 atmospheric days (Lorenz 1996).

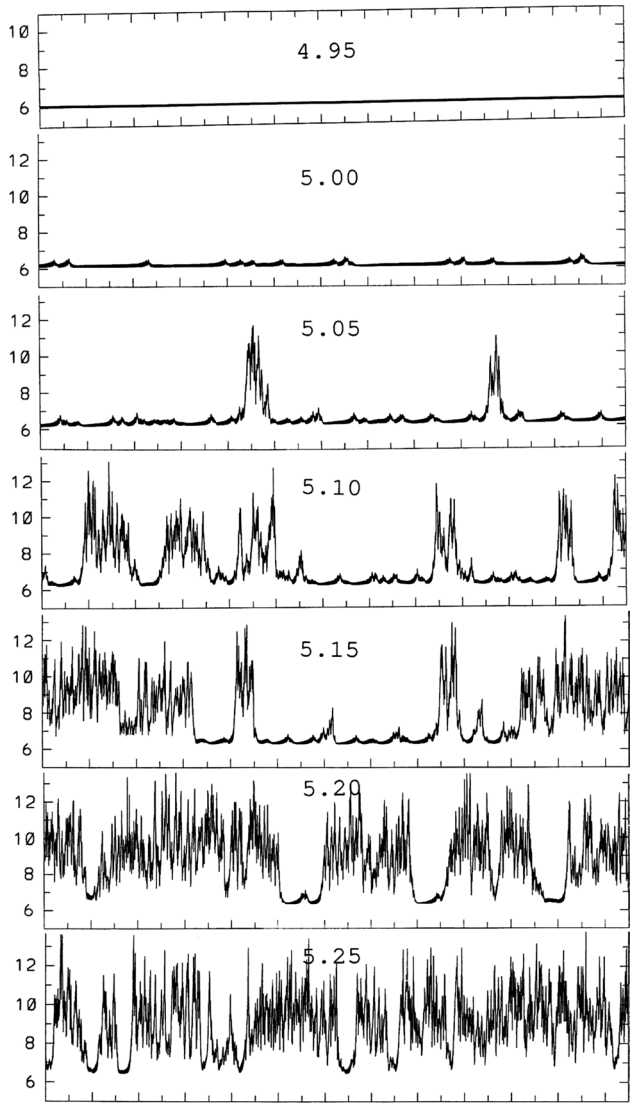

Fig. 1 The time series of total energy for the L96 1D system, where total energy is defined as $E = \frac{1}{2}\sum X_k^2$. The labels [4.95–5.25] indicate the value of the forcing, F, in (10). Taken from Lorenz (2006) (Fig. 3). ©American Meteorological Society. Used with permission

Lorenz performed extended numerical runs and examined the resultant time series of total energy.

$$E = \frac{1}{2}\sum_{k=1}^{K} X_k^2 \tag{11}$$

Figure 1 shows the time series for the L96 1D system for different values of the forcing parameter, F. From his criteria, Lorenz identified regimes in the time series for $F =$ 5.05–5.25.

Figure 2a shows the time series of total energy for the L96 System. The time series shows drops in the total energy of the system which persist for a few model time units, which could indicate the presence of regimes: it looks

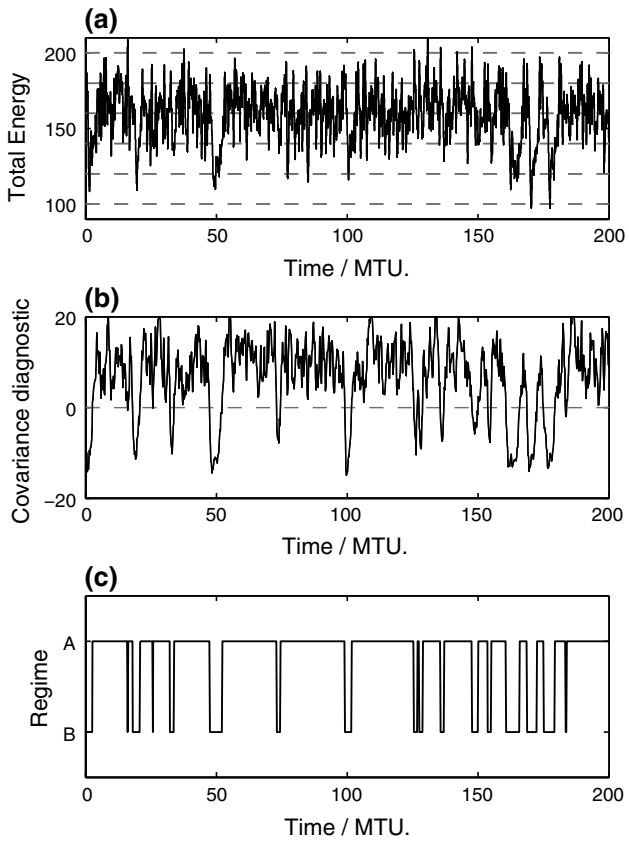

Fig. 2 **a** The time series of total energy for the L96 System, where total energy is defined as $E = \frac{1}{2}\sum X_k^2$. Total energy is not conserved as the system is forced and dissipative. **b** The covariance diagnostic evaluated for the data shown in **a**. If the diagnostic is positive, a wave-2 pattern dominates the behaviour of the X variables, whereas if it is negative, a wave-1 pattern is dominant. See text for details. **c** The same data set, interpreted in terms of two regimes—A and B—defined using the covariance diagnostic. If the diagnostic is positive, the system is interpreted as being in Regime A

qualitatively similar to the series for $F = 5.25$ in Fig. 1, for the L96 1D system.

Lorenz (2006) also considers the spatial distribution of the X variables as a test for the presence of regimes. Figure 3 a shows profiles of the L96 system X_k variables at 6 h intervals taken from 60 MTU after the start of the dataset in Fig. 2, when the total energy of the system tends to be higher with large oscillations. Figure 3 b shows profiles from 50 MTU after the start of the dataset, when the total energy of the system has dropped to a lower, more quiescent state. The two sets of profiles are different: since these two samples are characterised by physically different states, it is reasonable to interpret them as coming from two different regimes.

The difference in structure between the two regimes is most clearly revealed by consideration of the covariance matrix of the X variables, shown in Fig. 4. It is convenient to define Regimes A and B in terms of this covariance

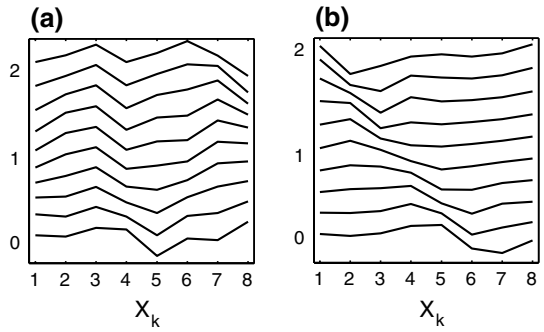

Fig. 3 Profiles of the $K = 8$ X_k variables for the L96 system. The profiles are taken from **a** 60 MTU and **b** 50 MTU after the start of the time series shown in Fig. 2. The labelling on the Y-axis indicates the number of "atmospheric days" since the first profile. The profiles from Regime A show a wave-2 type behaviour, while those from Regime B show a dominant wave-1 pattern

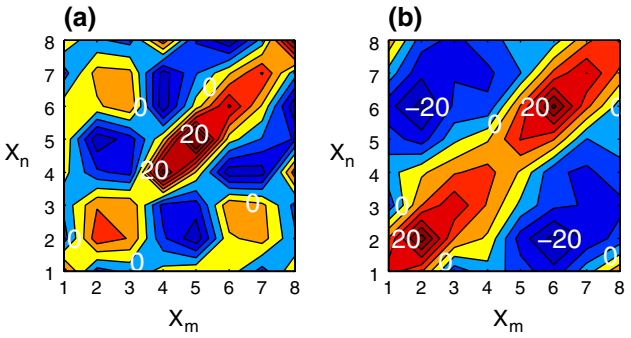

Fig. 4 The covariance matrix, $C(m, n)$, for the covariance between X_m and X_n calculated from a 1 MTU sample, **a** 60 MTU and **b** 50 MTU after the start of the time series shown in Fig. 2. **a** The dominant feature is a wave-2 pattern, with the 'opposite' X variables in phase with each other. **b** The dominant feature is a wave-1 pattern, with the 'opposite' X variables out of phase with each other

matrix C, calculated using samples of the time series 1 MTU long, where $C(m, n)$ represents the covariance between X_m and X_n:

$$\begin{aligned} \text{Regime} = A &: (C(1,5) + C(2,6) + C(3,7) + C(4,8)) > 0 \\ \text{Regime} = B &: (C(1,5) + C(2,6) + C(3,7) + C(4,8)) < 0 \end{aligned} \tag{12}$$

In other words, the system is defined to be in Regime A if opposite X variables are in phase (for $K = 8$), and in Regime B if opposite X variables are out of phase. The time series of this "covariance diagnostic" and the resultant identified regimes are shown in Fig. 2b, c respectively.

Lorenz's third criterion for regimes requires their duration to be longer than some other significant oscillation of the system. In the L96 system for the chosen parameter settings, the dominant oscillation of the X variables has a time period of approximately 0.5 MTU. Figure 5 shows a pdf of the duration of each regime. The mean duration of Regime

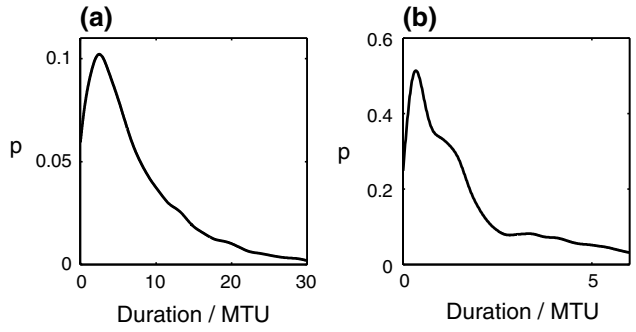

Fig. 5 The probability distribution function (pdf) for the duration of **a** Regime A and **b** Regime B. Regime A events are observed to be longer lasting on average than Regime B events

A is 7.12 MTU, and the mean duration of Regime B is 1.55 MTU. The average duration of both regimes is greater than 0.5 MTU, so we can conclude that, for the case where the ratio of time scales equals ten, the L96 system does indeed exhibit regime behaviour, and is a suitable model for use in this investigation.

The predictability of the regime behaviour of the L96 system will be studied using the same techniques used for atmospheric data. Firstly, it has been suggested that the time series should be temporally smoothed to help identify the regimes (Straus et al. 2007; Stephenson et al. 2004). For example, consider the well-known Lorenz (1963) system (the "butterfly attractor"): the system clearly has two regimes corresponding to the two lobes of the attractor, but these regimes are only apparent in a pdf of the system if the time series is first temporally averaged (Corti et al. 1999; Stephenson et al. 2004).[2] In the L96 system, the modal residence time in Regime B is ∼0.4 MTU (Fig. 5), so a running time average over 0.4 MTU will be used to smooth the time series.

When studying atmospheric data sets, the dimensionality of the problem is usually reduced using an empirical orthogonal function (EOF) analysis on the temporally smoothed data series (Selten and Branstator 2004; Straus et al. 2007). Due to the symmetry of the L96 system, the leading EOFs are simply the dominant harmonics, as seen in Fig. 6. The first two EOFs are degenerate, and are $\frac{\pi}{4}$ out of phase wavenumber two oscillations, i.e. are in phase quadrature. The third and fourth EOFs are similarly in phase quadrature, and are $\frac{\pi}{2}$ out of phase wavenumber one oscillations. Consideration of Fig. 6 shows that EOF1 and EOF2 are likely to dominate in Regime A, whereas EOF3 and EOF4 dominate in Regime B. The principal components (PCs) were calculated for each EOF. Due to the

[2] The time series must not be too heavily smoothed as this will cause the pdf to tend towards a Gaussian distribution (following the central limit theorem).

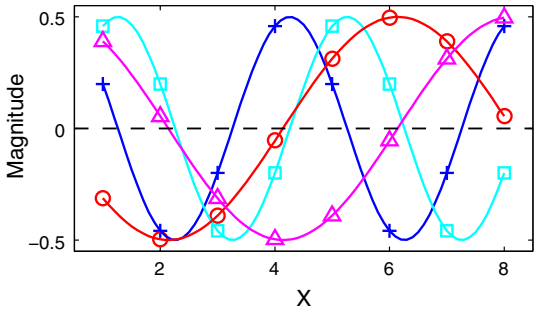

Fig. 6 The first four Empirical Orthogonal Functions (EOFs) calculated from the truth time series: EOF1 (*crosses*), EOF2 (*squares*), EOF3 (*circles*), EOF4 (*triangles*). Due to the symmetry of the system, the EOFs correspond to the leading harmonics

degeneracies of the EOFs, the magnitude of the principal component vectors, $||PC1, PC2]||$ and $||[PC3, PC4]||$, will be considered and the pdf of the system plotted in this space.[3] The corresponding eigenvalues show that EOF1 and EOF2 account for 68.7 % of the variance, while EOF3 and EOF4 account for a further 14.4 %.

3 The true attractor

Starting from initial conditions on the attractor, the full set of L96 equations is integrated forward for 50,000 MTU $\sim$ 700 "atmospheric years" to ensure the attractor is fully sampled. This full system is defined to be the "true" system which the simulators seek to represent. The truth time series is temporally smoothed with a running average over 0.4 MTU. For comparison, the raw, unsmoothed time series is also considered. An EOF decomposition is carried out on both the raw and smoothed time series, and the PCs calculated. The dimensionality of the space is further reduced by considering only the magnitude of the PC vectors $||[PC1, PC2]||$ and $||[PC3, PC4]||$ as a function of time. The state vector pdf for the full truth model is shown in Fig. 7 for (a) the unsmoothed time series and (b) the smoothed time series. Temporally smoothing the time series helps to identify the two regimes. The maximum of the pdf in both (a) and (b) is located at large $||[PC1, PC2]||$ and small $||[PC3, PC4]||$, corresponding to the more common wave-2 "Regime A" state of the system. However, in (b) the pdf is elongated away from this maximum towards large $||[PC3, PC4]||$ and small $||[PC1, PC2]||$, where there is a small but distinct subsidiary peak; this corresponds to the less common "Regime B". Figure 7a is also elongated, but does not have a distinct second peak, so does not indicate the presence of regimes.

Figure 7c and d show the mean residence time of trajectories within local areas of phase space. For each point in PC space, a circular region with radius R is defined, and the average residence time of trajectories within that region is calculated, following Frame et al. (2013) [(equivalent to the technique presented in Branstator and Berner (2005)]. Here $R = 2$, and the displayed circle indicates the size of the region for comparison. For both (c) the unsmoothed and (d) the smoothed time series, two regions of high residence time can be identified. The longest residence times occur at large $||[PC1, PC2]||$ and small $||[PC3, PC4]||$, corresponding to Regime A. There is a further peak in residence time at large $||[PC3, PC4]||$ and small $||[PC1, PC2]||$, corresponding to Regime B. These two distinct peaks provide further evidence for the regime nature of the L96 system: there are two regions in phase space which the system preferentially occupies for extended periods of time, and transitions between these regions are more rapid. This diagnostic confirms that Regime A is more persistent than Regime B, as expected from Fig. 5.

Figure 7e and f show the mean velocity of the system's motion through phase space, following Frame et al. (2013). The colour indicates mean speed, and arrows indicate mean direction. A region with radius 0.5 is defined centred on each point in phase space, and the net displacement of trajectories starting within this region is calculated over 0.05 MTU, before the average magnitude and direction is calculated. Both figures (e) and (f) show two centres of rotation in phase space corresponding to the two regimes. The centres of rotation approximately coincide with the maxima of the pdfs shown in (c) and (d), as expected (Berner and Branstator 2007). On average, trajectories circle these centres, resulting in persistent conditions (cf. the Lorenz '63 'butterfly attractor'). The structure of the flow field is different for the smoothed time series—the second centre is less clearly defined, but coincides with a maximum in the average magnitude of displacement: trajectories oscillate vertically about this centre during a persistent Regime B phase, resulting in a large average magnitude, but small average displacement vector. Both Fig. 7e and f provide conclusive evidence of the existence of regimes in the L96 system, which can be detected in both the raw and smoothed data.

4 Simulating the attractor

Section 3 demonstrated that the full L96 system exhibits regime behaviour. In this Section, the skill of the truncated

[3] This is equivalent to considering complex EOFs, which are used to capture propagating modes. In this study, EOF1 and EOF2 together represent the first complex EOF, and EOF3 and EOF4 represent the second complex EOF, and therefore $||[PC1, PC2]|| = ||CPC1||$, where CPC1 is the PC of the first complex EOF, etc.

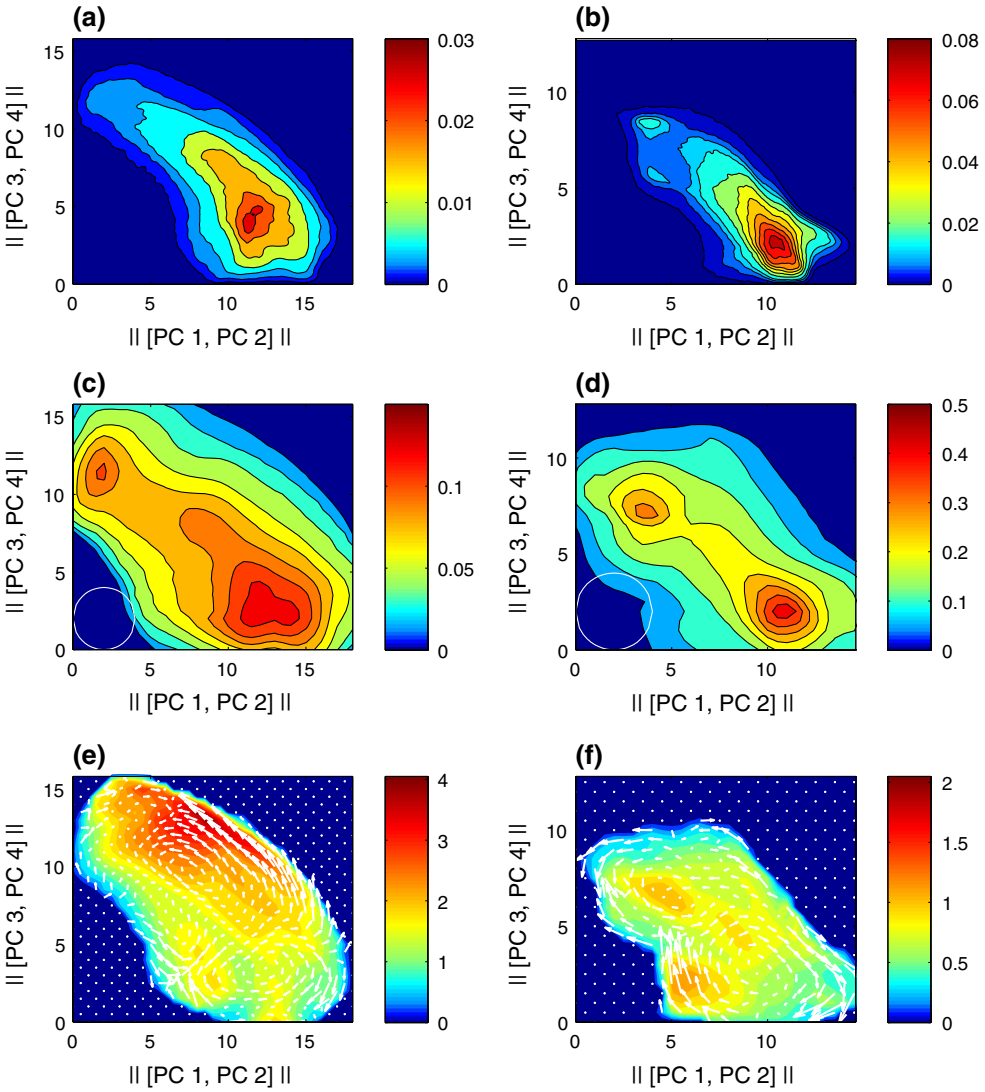

Fig. 7 Regime characteristics of the full L96 system. Both the raw and temporally smoothed time series are considered, where the smoothing is a running average over 0.4 MTU. Each diagnostic is shown in the space of pairs of leading PCs, $|||[PC1, PC2]|||$ and $|||[PC3, PC4]|||$. See text for details. **a** Raw and **b** smoothed pdfs. **c** Raw and **d** smoothed mean residence times (MTU): the mean length of time a trajectory remains within 2 units of each location. A circle of radius two is indicated. **e** Raw and **f** smoothed magnitude (*colour*) and orientation (*arrows*) of the average displacement in phase space over 0.05 MTU, averaged over trajectories passing within 0.5 units of each position in phase space

L96 simulators at reproducing this regime behaviour will be evaluated. Each simulator is integrated for 50,000 MTU as for the full L96 system. For the perturbed parameter experiment, each of the 40 deterministic models are integrated for 50,000 MTU. While regime behaviour can be detected in both the raw and smoothed truth time series, the results in Sect. 3 indicate that it is easier to detect the presence of regimes in the smoothed time series pdf, as was suggested by Straus et al. (2007). Therefore, the time series are temporally smoothed with a running average over 0.4 MTU. The four leading order truth EOFs are used to calculate the PCs of the simulated time series to ensure a fair comparison (Corti et al. 1999). The dimensionality of the space is further reduced by considering only the magnitude of the PC vectors $[PC1, PC2]$ and $[PC3, PC4]$ as a function of time. The state vector pdfs for each of the simulators considered are shown in Fig. 8; the pdf for the full "truth" model has also been reproduced for ease of comparison.

Panel (a) in Fig. 8 corresponds to the full "truth" Eq. (1), reproduced from Fig. 7, and shows two distinct peaks corresponding to the two regimes, A and B. The simulators are unable to capture accurately the subsidiary peak corresponding to Regime B. The DET, WA and WM models all assign too little weight to this area of phase space. However, the AR(1) stochastic models show a large improvement—Regime B is explored more frequently than by the deterministic or white stochastic models, though not as frequently as by the full truth system. The attractor of the perturbed parameter model, Fig. 8c, shows a distinct peak for Regime B, unlike the other simulators. However, the attractor has a very different structure to that for the truth time series—Regime B is visited too frequently.

In fact, the PP attractor in Fig. 8c consists of the average of 40 constituent members, shown in Fig. 9. The contour colours are consistent with Fig. 8. Many of the 40 different perturbed parameter ensemble members

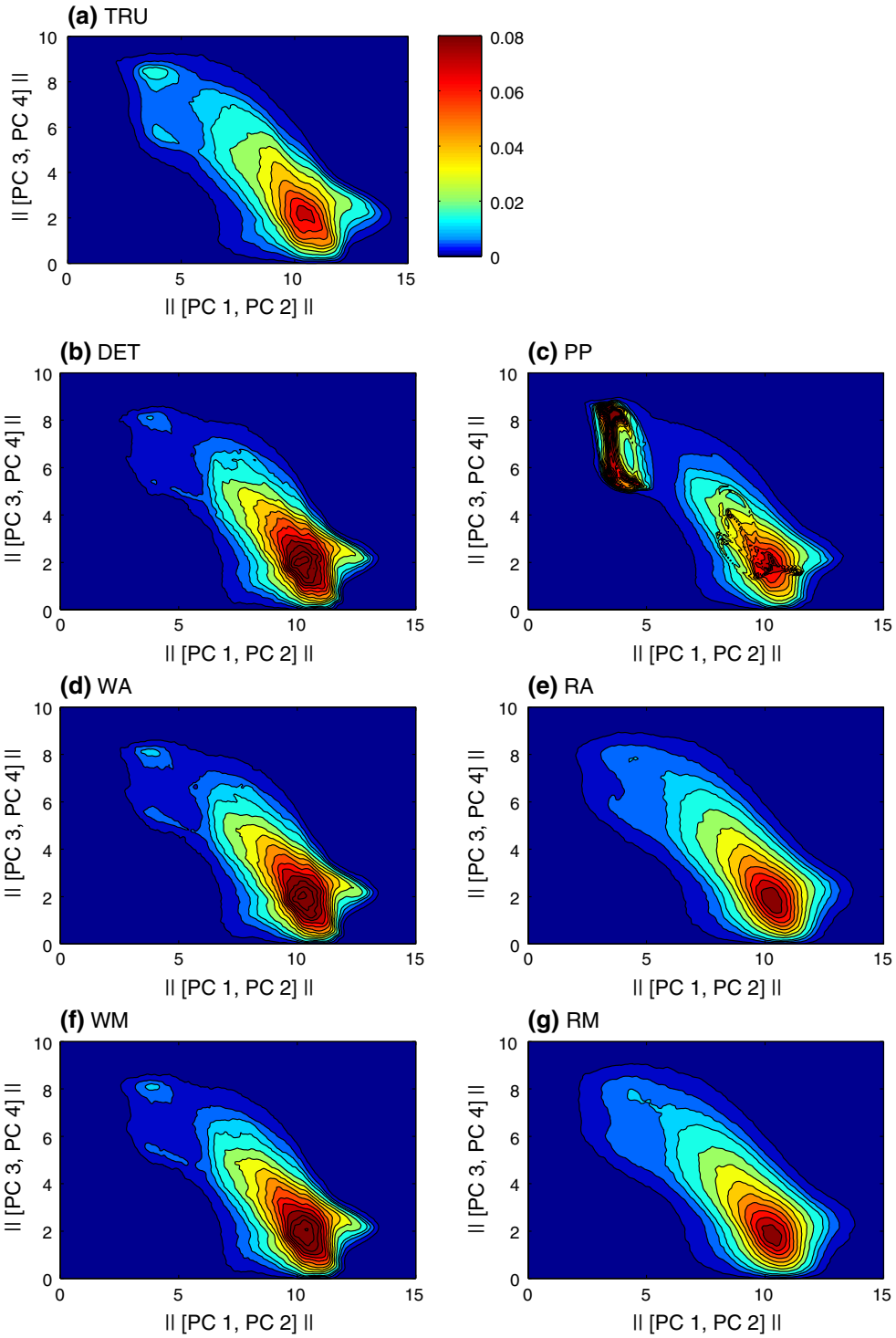

Fig. 8 Ability of different parametrisation models to reproduce the true attractor (**a**). The pdf of the state vector for the L96 system is plotted in the space of pairs of leading EOFs. See text for details. Six different forecasting models are shown. **b** Deterministic parametrisation scheme; **c** perturbed parameter scheme; additive stochastic parametrisation with **d** white and **e** red noise; multiplicative stochastic parametrisation with **f** white and **g** red noise. The degree of perturbation of the perturbed parameters, and the standard deviation and autocorrelation in the stochastic parametrisations have been estimated from the truth time series. The same EOFs determined from the full truth data set are used in each panel, and the colour of the contours is also consistent

show very different regime behaviour compared to the true attractor. While some ensemble members (e.g. numbers 11, 17, 18 and 21) look reasonable, many remain in Regime A and do not explore Regime B (e.g. numbers 6, 15, 16 and 27), while some predominantly explore Regime B (e.g. numbers 5, 37–39). Perturbed parameter ensembles are often used for climate prediction. However, if individual members only explore one region of the true climate attractor, how can the effect of forcing on the frequency of occurance of different regimes be established?

The perturbed parameter ensemble also displays interesting behaviour which was not observed in the 'true' system, and which can be interpreted in terms of 'attractor

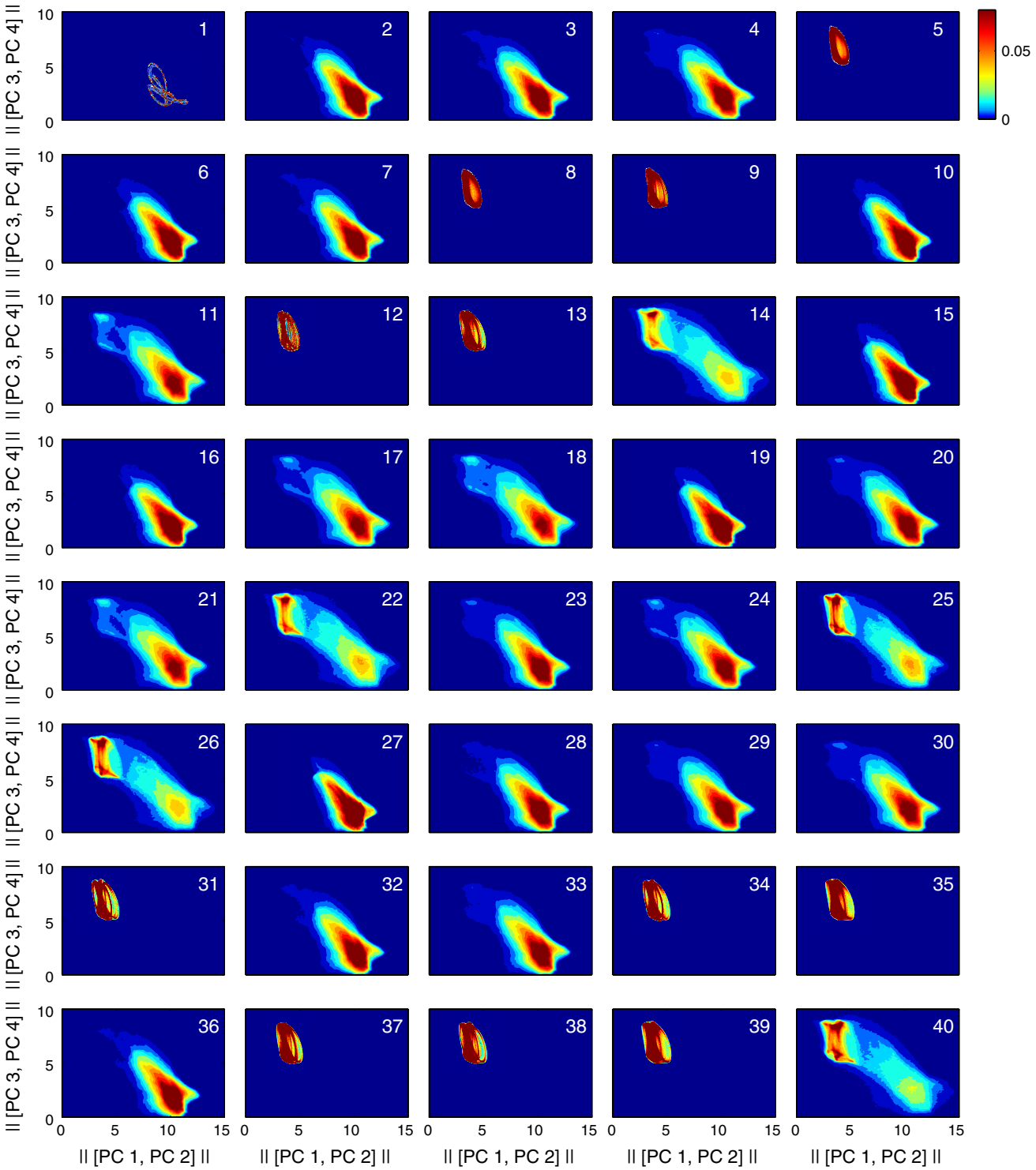

Fig. 9 Ability of the perturbed parameter scheme to reproduce the true attractor. The pdf of the state vector for the L96 system is plotted in the space of pairs of leading EOFs, $[EOF1, EOF2]$ and $[EOF3, EOF4]$ for each of the forty perturbed parameter ensemble members (numbered). The attractors of individual ensemble members appear very different to the true attractor, with some members only exploring one of the two regimes present in the full system. The figure numbers correspond to the numbering of the ensemble members in Table 4. The same *colour bar* is used as in Fig. 8. Note that many of the ensemble members saturate this colour bar

ruins' (Itoh and Kimoto 1996, 1997). Figure 9 shows that PP ensemble member 1 is not chaotic, and instead follows a stable periodic orbit. Similarly, ensemble members 5, 8, 9 etc. also follow a quasi-stable orbit on a neighbouring attractor. On varying the parameters, the system begins to move among the ruins of these attractors, exploring both regimes. The 'true' system can now be interpreted similarly as moving between the ruins of two attractors, each of which represents a distinct observed regime.

To ease comparison, the 2D Fig. 8 is decomposed into two, 1D pdfs for each of the simulators considered (Fig. 10a, b). The Kolmogorov–Smirnov Statistic, D_{ks}, and Hellinger distance, D_{Hell}, are calculated as quantitative measures of the difference between the true and simulated pdfs for each case:

$$D_{ks} = \max_x |P(x) - Q(x)|, \tag{13}$$

$$(D_{Hell})^2 = \frac{1}{2}\int \left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 \mathrm{d}x, \tag{14}$$

where P and Q are the simulated and truth cumulative PDFs respectively, and $p(x)$ and $q(x)$ are the simulated and truth PDFs respectively. For both D_{ks} and D_{Hell}, the smaller the measure, the closer the simulated pdf is to the true pdf. The results are shown in Table 5. The RM simulator is the most skilful for each case according to the Hellinger distance. The RA scheme also scores well for both cases, and is the most skilful representation of $||[PC1, PC2]||$ according to D_{ks}. The PP ensemble performs similarly to the DET, WA and WM simulations for the $||[PC3, PC4]||$ case, but is considerably poorer than the DET scheme for the $||[PC1, PC2]||$ pdf.

However, as illustrated above, a PP ensemble must be interpreted carefully. Since each member of the ensemble is

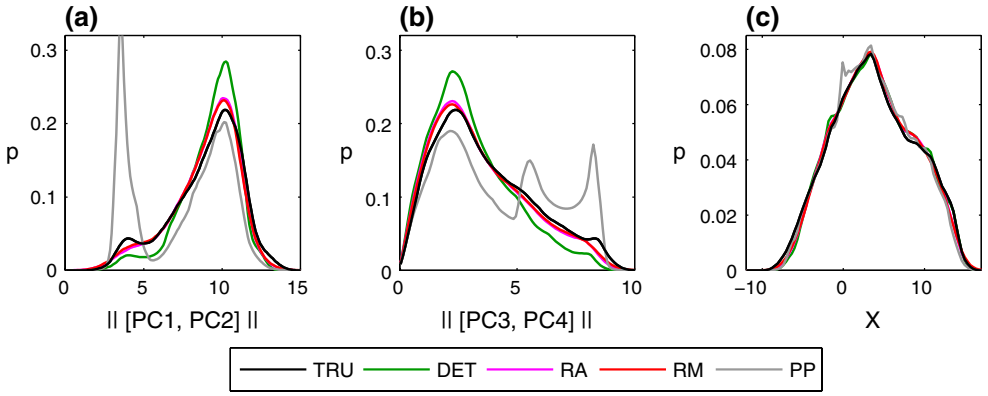

Fig. 10 Ability of different parametrisation models to reproduce the true attractor. The pdf of the state vector for the L96 system is plotted for **a** the magnitude of $[PC1, PC2]$, **b** the magnitude of $[PC3, PC4]$, and **c** the pdf in X space. Four different forecasting models are shown on each panel (*coloured lines*) together with that for the truth data. White noise, whether additive or multiplicative, is indistinguishable from the deterministic case, so has not been shown. It is easier to distinguish between the parametrisation schemes if the regime behaviour is considered

Table 5 The skill of different parametrisation schemes at reproducing the structure of the Truth attractor along each of two directions defined by the dominant EOFs for $F = 20$, and for a changed climate, where $F = 23$

Parametrisation	$F = 20$				$F = 23$			
	[PC1,PC2]		[PC3,PC4]		[PC1,PC2]		[PC3,PC4]	
	D_{ks}	D_{hell}	D_{ks}	D_{hell}	D_{ks}	D_{hell}	D_{ks}	D_{hell}
Deterministic	0.082	0.120	0.106	0.110	0.102	0.102	0.047	0.074
WA Stochastic	0.091	0.128	0.113	0.117	0.053	0.064	0.025	0.058
RA Stochastic	**0.030**	0.057	0.048	0.054	0.025	0.021	0.035	0.047
WM Stochastic	0.086	0.122	0.108	0.112	0.066	0.071	**0.025**	0.061
RM Stochastic	0.034	**0.053**	**0.037**	**0.043**	**0.019**	**0.018**	0.029	**0.038**
Pert. Parameters	0.245	0.261	0.151	0.142	0.149	0.146	0.074	0.090

The Kolmogorov–Smirnov distance D_{ks}, and Hellinger distance D_{hell}, are used to measure the similarity between the true and simulated pdfs. The smaller each of these measures, the closer the simulated pdf is to the true pdf. The best simulator according to each measure is shown in bold

a physically distinct model of the system, the simulated climatology of each member should be assessed individually. The 1D pdfs of $|||[PC1, PC2]|||$ and $|||[PC3, PC4]|||$ were calculated for each ensemble member, and D_{Hell} between the simulated and true pdfs evaluated for each case. For comparison, the deterministic and stochastic simulations were also split into 40 sections, each 1,250 *MTU* long, and D_{Hell} evaluated for each section as for the PP ensemble. This allows the effect of sampling error to be considered. The distribution of D_{Hell} for each case is shown in Fig. 11. The spread of skill is largest for the PP ensemble, with some members showing very poor climatologies, while others are more skilful. The WA and WM schemes do not show a significant difference from the DET simulation, while the RA and RM schemes significantly outperform the other schemes.

Figure 10c shows the pdf of the X variables for each parametrisation scheme tested (the WA and WM schemes performed very similarly to the DET scheme, so have not been shown). This is a conventional definition of climate, including information about both the mean and expected variability of the variable of interest. This diagnostic indicates that all parametrisation schemes perform well—it is hard to distinguish between the different schemes if only the distribution of the X variables is taken into account. However, Fig. 10a, b show clearly how poorly the PP scheme reproduces the attractor of the system. Defining the climate of the system in terms of its regime behaviour has unveiled

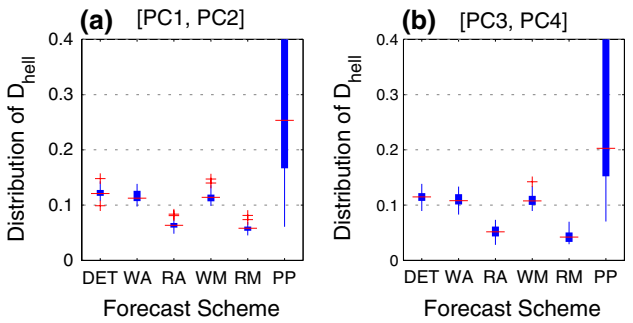

Fig. 11 The distribution of Hellinger distance calculated for the difference between forecast and observed EOF climatologies. The pdf for the magnitude of **a** $[PC1, PC2]$ and **b** $[PC3, PC4]$ is calculated. For the deterministic and stochastic models, the time series is split into 40 sections, 250 MTU long, and the pdfs calculated for each. For the perturbed parameter ensemble, the pdfs are calculated for each ensemble member separately. The Hellinger distance between each forecast pdf and the true pdf is evaluated, and the distribution of Hellinger distance represented by a *box* and *whisker plot*. The median value is marked by a *horizontal red line*. The 25th and 75th percentiles are indicated by the edges of the *box*, and the *whiskers* extend to the minimum and maximum value in each case, unless there are outliers, which are marked by a *red cross*. An outlier is defined as a value smaller than 1.5 times the inter-quartile range (IQR) below the lower quartile, or greater than 1.5 IQR above the upper quartile

more information about the system, and is a harsher test of the performance of the different simulators.

5 Simulating regime statistics

While reproducing the pdf of the true system is important for capturing regime behaviour, it is also necessary for a climate simulator to represent the temporal characteristics of the regimes. This is evaluated using the distribution of persistence of each regime (Dawson et al. 2012; Pohl and Fauchereau 2012; Frame et al. 2013), and will be considered using two techniques.

Firstly, the behaviour of the system in PC space is used to examine the temporal characteristics of the system, following Frame et al. (2013). The mean residence time of trajectories in phase space is calculated. For each point in PC space, a circular region with radius R is defined, and the average residence time of trajectories within that region is calculated, as for Fig. 7c, d. Figure 12 shows the mean residence time of trajectories when $R = 2$ PC units. Two regions of high residence time can be identified in the truth simulation shown in panel (a). The longest residence times occur at large $|||[PC1, PC2]|||$ and small $|||[PC3, PC4]|||$, corresponding to Regime A. There is a smaller peak in residence time at large $|||[PC3, PC4]|||$ and small $|||[PC1, PC2]|||$, corresponding to Regime B.

The simulators are able to capture the temporal characteristics of the true system. They show two distinct peaks in residence time of the correct magnitude. However, there are subtle differences between the different simulators. In Fig. 12, the DET, PP, WA and WM simulators have regimes that are too persistent—the two peaks in residence time are too high, particularly for Regime A. The RA and RM schemes perform better, and capture the average residence time for Regime B particularly well.

It is important to recall that the PP ensemble consists of 40 physically distinct representations of the system. The residence time pdfs are plotted for each perturbed parameter ensemble member in Fig. 13 for the $R = 2$ case. The individual ensemble members have vastly different temporal characteristics. Many ensemble members have persistent regimes with few transitions, including some which had realistic pdfs in Fig. 8 (e.g. numbers 11, 21). The same colour scale is used as for Fig. 12, but has saturated for several panels. For example, the maximum residence time of Regime A for ensemble member 6 is 0.72 MTU, which is almost double the maximum residence time observed in the full system. Considered individually, the PP ensemble members are a poor representation of the regime behaviour of the true system.

The second technique used to study the regime statistics uses the definition of Regimes A and B given by (12). The

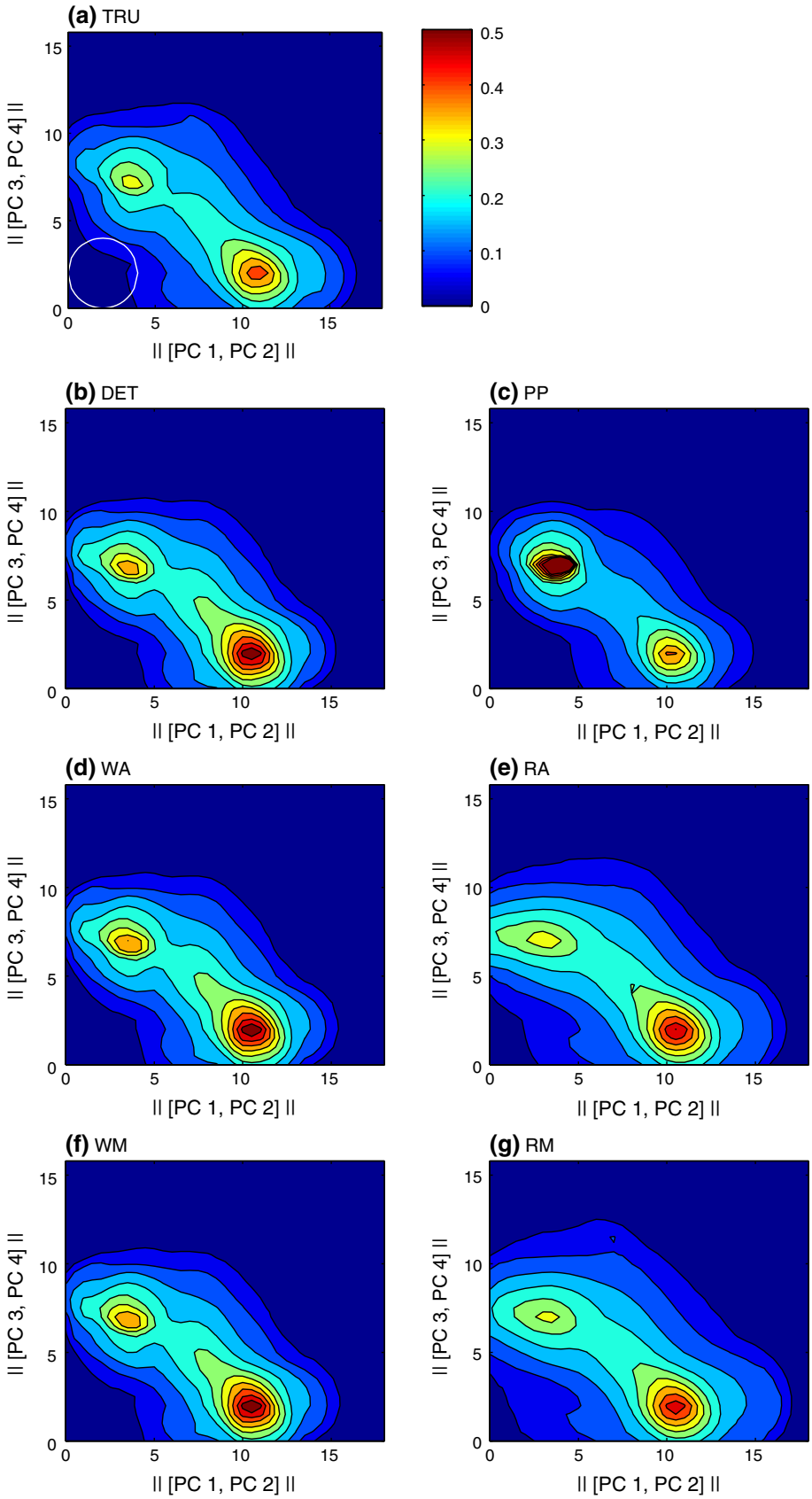

Fig. 12 Mean residence time in model time units. The mean length of time trajectories remain within 2 units of each position in PC space. A circle of radius 2 units is shown for comparison in **a**

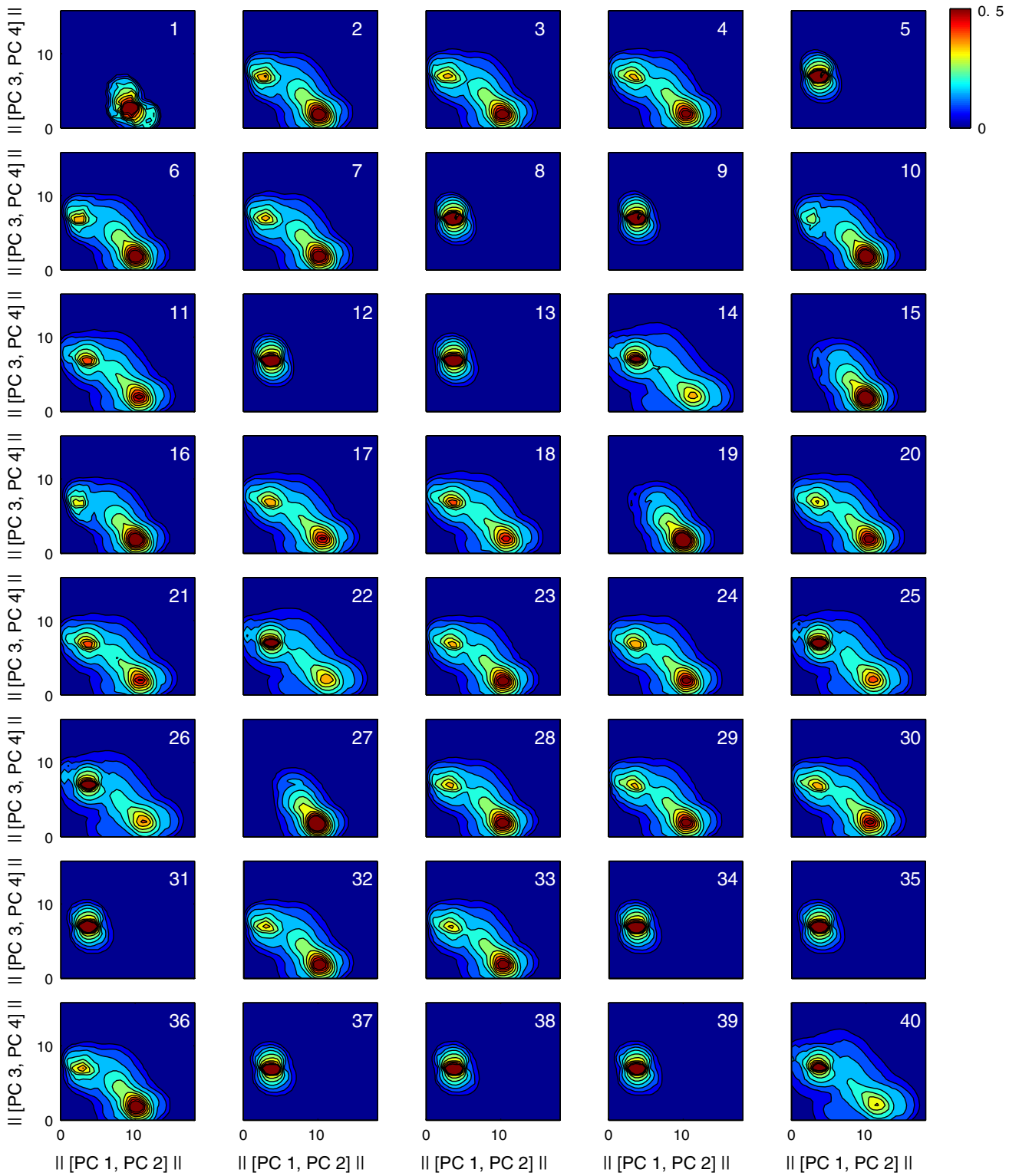

Fig. 13 Mean residence time in model time units for each member of the perturbed parameter ensemble. The mean length of time trajectories remain within 2 units of each position in PC space. The figure numbers correspond to the numbering of the ensemble members in Table 4

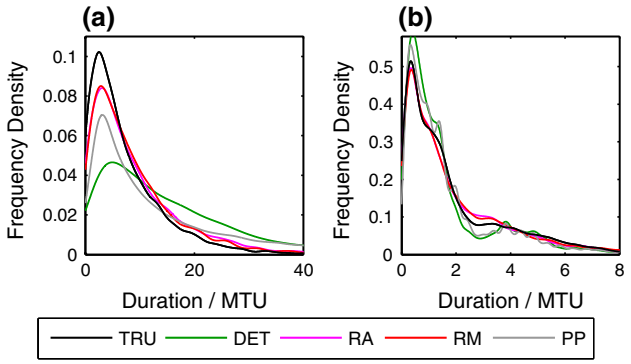

Fig. 14 Predicting the distribution of persistence of **a** Regime A and **b** Regime B. The true distribution is shown in *black*, and the six different simulators shown as *coloured lines*. The results for the white additive and multiplicative schemes are very similar to the deterministic scheme, so have not been shown

definition is used to determine the regime at each time step, and the pdf of persistence of each regime calculated as for Fig. 5. Figure 14 compares the persistence pdfs for the full L96 system with that of the different simulators. The AR(1) stochastic parametrisation schemes (red and magenta lines) improve significantly over the white stochastic schemes and the deterministic scheme (green)—the distribution of regime durations more closely matches the true distribution for the AR(1) noise cases.

As before, it is helpful to consider the climatology of each perturbed parameter ensemble member separately. The Hellinger distance between the truth and each perturbed parameter persistence pdf was calculated. For comparison, the deterministic and stochastic simulations were split into 40 sections, and the pdfs of persistence calculated for each section, and the distribution of D_{Hell} considered (not shown). As before, the WA and WM schemes perform very similarly to the DET scheme, with similar median and range in D_{Hell}. The red AR(1) stochastic schemes are significantly more skilful at predicting the regime statistics than the deterministic and white stochastic schemes, especially when considering Regime A: both schemes have a significantly lower median and smaller range of D_{Hell} values. The skill of the PP ensemble shows the greatest variability. While the median score is only slightly greater than the median for the deterministic ensemble, some PP ensemble members score very large Hellinger distances. Twelve PP ensemble members underwent no regime transition, so were excluded from this analysis.

Table 6 shows the overall proportion of time spent in each regime for each simulator. This also improves when red stochastic schemes are considered; the DET, WA and WM schemes visit Regime B too rarely, whereas the proportion of time spent in Regime B by the AR(1) stochastic schemes is close to the truth.

Table 6 Predictability of regime frequencies by different simulators

Parametrisation	p(A)	p(B)
Truth	**0.8199**	**0.1801**
Deterministic	0.9294	0.0706
WA Stochastic	0.8879	0.1121
RA Stochastic	0.8017	0.1983
WM Stochastic	0.9015	0.0985
RM Stochastic	0.7780	0.2220
Perturbed Parameters	0.4197	0.5803

The deterministic and white stochastic schemes all underpredict the proportion of time spent in the rarer Regime B, while the AR(1) stochastic schemes explore this region of phase space with the correct frequency

For reference, the true regime frequencies are shown in bold

6 Climate change experiments

The results presented above indicate that stochastic parametrisation schemes which include a temporally correlated noise term significantly improve the representation of regimes in the L96 system. However the question remains, do they also improve the ability of a simulator to capture changes in regime behaviour due to an external forcing? In the introduction, two possible responses to forcing were outlined, where the climate attractor is considered as a field with several potential wells (Hasselmann 1999).

1. A change in the relative depths of the potential wells, leading to changes in both the relative residence times and transition frequencies
2. A change in regime properties, such as centroid location and number of regimes

In the L96 system, the regime structure of the attractor can be altered by changing the constant forcing term, F, in (1). It was found that a moderate change from $F = 20$ to $F = 23$ resulted in significant changes to the regime structure of the system. A "climate change" experiment was performed. The truth model and different simulators were each integrated for 500,000 MTU with $F = 20$, before the constant forcing term was increased linearly from $F = 20$ to $F = 23$ over 500,000 MTU. The integration then continued for 500,000 MTU with the new higher forcing term.

Figure 15a shows the difference between the 2D pdf at the start and end of the climate change experiment for the true system, where the 2D pdf is calculated using a 50,000 MTU sample taken from within each of the $F = 20$ and $F = 23$ portions of the time series. The attractor has responded to the change in forcing in both of the ways outlined above: the centroid location for the rarer regime has shifted towards higher values of $||[PC1, PC2]||$ and

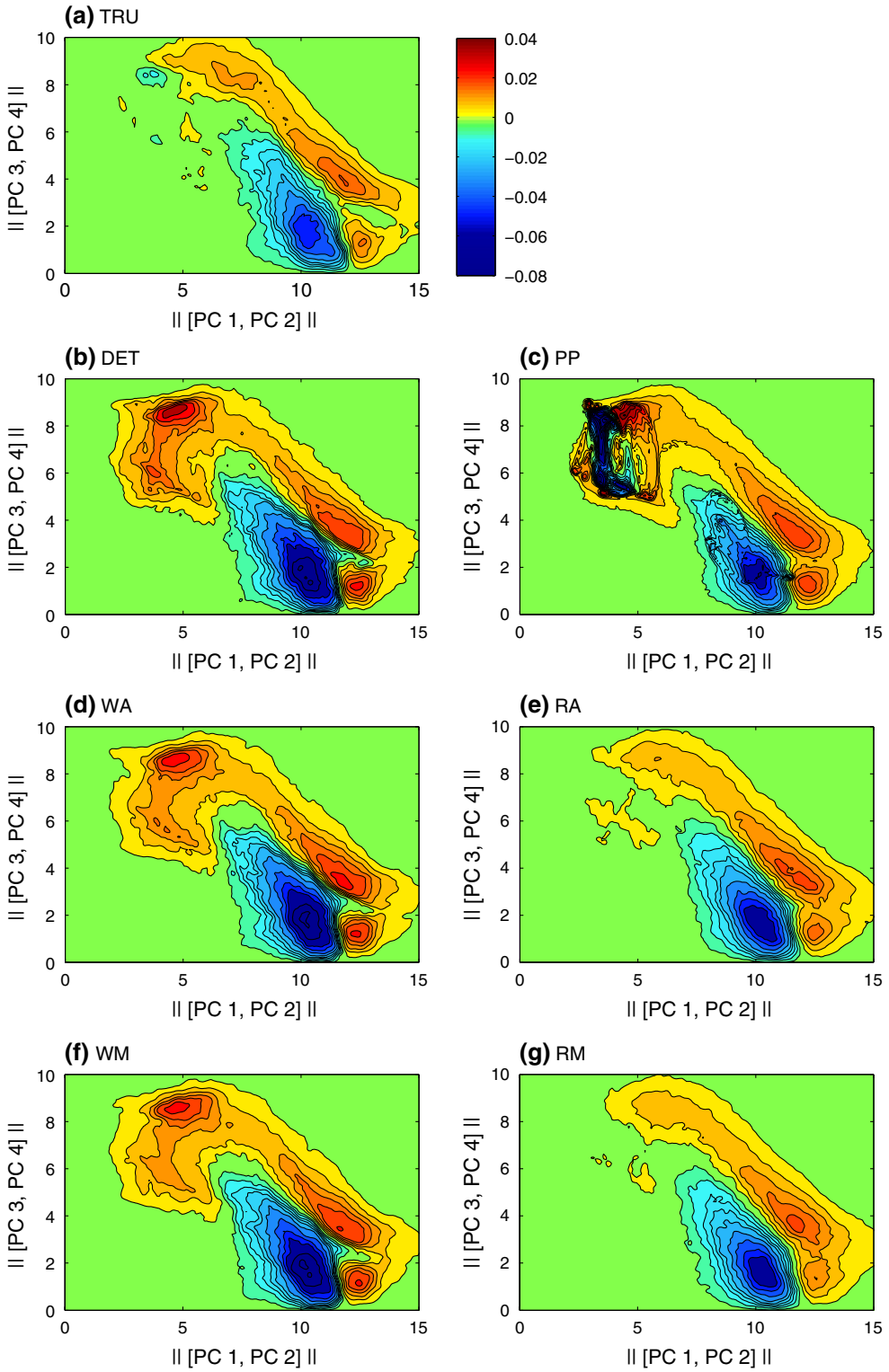

Fig. 15 Ability of different parametrisation models to reproduce the true attractor for a changed climate ($F = 23$) compared to the original system ($F = 20$). The difference between the $F = 23$ and $F = 20$ pdfs of the state vector for the L96 system is plotted in the space of pairs of leading EOFs. The results from six different simulators are shown. **b** Deterministic parametrisation scheme; **c** perturbed parameter scheme; additive stochastic parametrisation with **d** white and **e** red noise; multiplicative stochastic parametrisation with **f** white and **g** red noise. The same EOFs determined from the $F = 20$ full truth data set are used in each panel, and the *colour* of the contours is also consistent

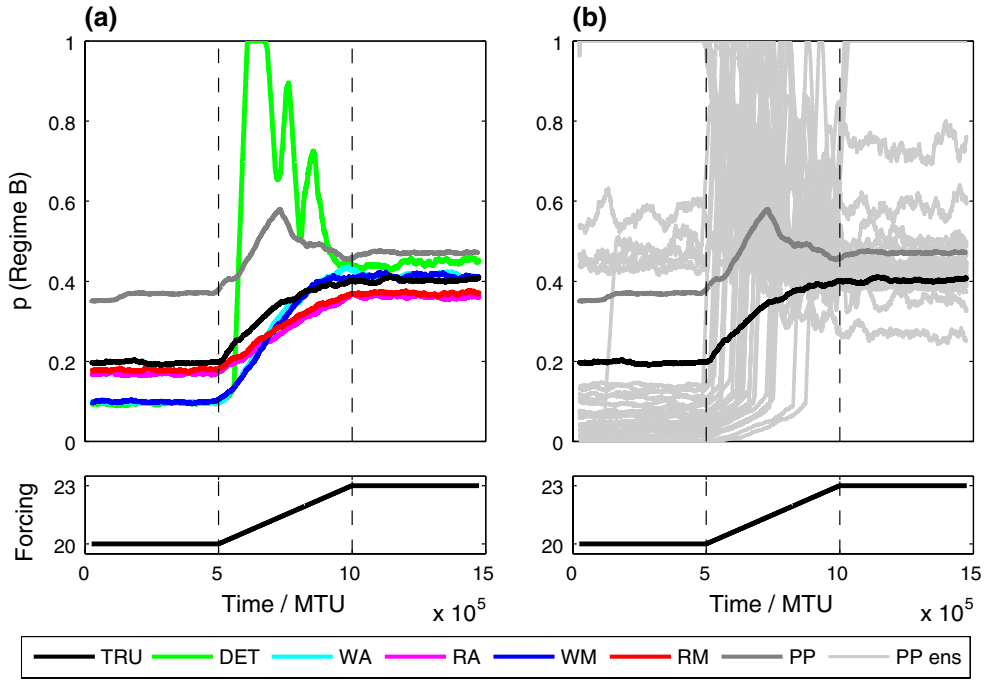

Fig. 16 The proportion of time spent in regime B, $p(B)$, and the applied external forcing as a function of time for the climate change experiment. **a** the results from the full 'truth' model are compared to those from each simulator. **b** the behaviour of individual perturbed parameter ensemble members (*pale grey*) is compared to the ensemble mean (*dark grey*) and the results from the full 'truth' system

$||[PC3, PC4]||$, and the proportion of time spent in the rarer regime (B) has increased.

It is interesting to analyse how successfully the different simulators capture these changes. The change in structure of the attractor can be analysed by considering the difference in 2D pdf for each simulator in the same space spanned by the leading EOFs of the true attractor. The deterministic, white additive and white multiplicative simulators each show a strong increase in the secondary peak in the pdf at (5,8), corresponding to the rarer Regime B, however this peak is in a similar position to that for the $F = 20$ case. In contrast, the secondary peak for the true attractor has shifted to (7,8), and is less distinct. The red additive and multiplicative parametrisation schemes accurately capture the changes to the structure of the pdf, including the new location and depth of the second regime. As for the $F = 20$ case, the perturbed parameter ensemble members do not necessarily explore both regimes, with some occupying only one, resulting in a distorted composite pdf.

As before, for ease of comparison, the 2D pdf can be decomposed into two, 1D pdfs for each of the simulators, and the Kolmogorov–Smirnov Statistic, D_{ks}, and Hellinger distance, D_{Hell}, calculated as measures of the difference between the true pdf and the simulated pdf for each case. Table 5 gives the results for the different simulators tested. The AR(1) multiplicative stochastic simulator is most skilful overall at reproducing the attractor according to these measures, though the white stochastic models also improve on the deterministic attractor.

The proportion of time spent in the rarer regime B, $p(B)$, is calculated following (12), and is a useful scalar summary for considering how well the different simulators capture the change in regime statistics and the changing nature of the climate attractor. Figure 16 shows the proportion of time spent in Regime B over the course of the climate change experiment. The plotted proportion was averaged over 50,000 MTU sections of the time series. The full truth model shows a smooth transition from $p(B) = 0.2$ to $p(B) = 0.4$, where the initial increase is observed to be more rapid than at later times.

No simulator captures this rate of change of $p(B)$ accurately. However, the stochastic simulators do capture the smooth transition in regime properties. In particular, the red stochastic models also give a good indication of the overall change in $p(B)$ between the two forcing scenarios, whereas the white stochastic models indicate a larger change occurs than is observed in the truth run. The deterministic model does not capture the smooth transition in regime properties: for certain values of F the simulated regime properties are completely dissimilar to those observed in the true system, though the final indicated value of $p(B)$ is a good estimate of the truth. The perturbed parameter ensemble members behave similarly to the deterministic model, undergoing sudden changes in behaviour as F is varied.

7 Discussion and conclusions

Regime behaviour, commonly observed in the atmosphere, is also observed in the L96 system. It is argued that the L96 system has two regimes for the parameters considered here ($F = 20$)—the system is in Regime A 82 % of the time, while the less common Regime B occurs 18 % of the time. The regime behaviour of this system makes it a useful testbed for analysing the ability of different simulators to reproduce regime behaviour. Three types of simulator were considered: one which used a deterministic parametrisation scheme, simulators including stochastic parametrisation schemes with additive or multiplicative noise, and a perturbed parameter ensemble. The L96 system mimics certain properties of the atmosphere, including chaotic motion and interaction of scales, so conclusions drawn from these experiments can indicate useful avenues of research in full complexity GCMs.

Each simulator was tested on its ability to reproduce the attractor of the full system, defined in a reduced space based on an EOF decomposition of the truth time series. None of the simulators accurately captures the less common Regime B, though a significant improvement is observed over the deterministic parametrisation when a temporally correlated stochastic parametrisation is used instead—the AR(1) stochastic simulators were able to capture the structure of the attractor well. The deterministic parametrisation scheme represents the mean effects of the unresolved subgrid scale motions, but is insufficient for capturing the regime behaviour of the system. These results suggest that it is the variability of the unresolved motion, represented by the stochastic parametrisation scheme, which plays a leading role in determining the structure of regimes in this system. The varying sub-grid scale forcing enables the system to explore a larger region of the attractor in the same way in which a ball-bearing in a potential well will explore around its equilibrium position when subjected to a random forcing.

In this study, it was found that the critical beneficial characteristic of the stochastic parametrisation scheme is not the method used to include the stochastic term (since the additive and multiplicative schemes performed very similarly), but instead it is the characteristics of the stochastic noise itself. Namely, in order to recover regime behaviour, the noise term must be temporally correlated. This reinforces this conclusion made by Arnold et al. (2013), where it was concluded that parametrisation schemes must go beyond representing sub-gridscale variability and instead should represent the effects of the unresolved subgrid motions on spatial and temporal scales greater than the truncation level. Due to the coupling of scales in complex systems, such schemes were found to produce the most skilful 'weather' forecasts, and showed improvements in the climatological pdf of the system.

However, unlike in Arnold et al. (2013) where an improvement in skill was observed when first white then red stochastic terms were included in the parametrisation scheme, in this work there is a negligible improvement when white stochastic terms are included into a deterministic parametrisation scheme. This implies that, while it is the unresolved terms which are responsible for determining the regime behaviour of the system, it is the organised nature of these unresolved terms which triggers regime changes and allows the system to explore the whole attractor. The temporally correlated stochastic schemes mimic the temporal organisation of the unresolved forcing, so better captures the regime structure of the system. This was observed as an improvement in both the statistics describing the persistence of the regimes and their frequency of occurrence when the red noise stochastic parametrisations were used instead of the deterministic scheme. The deterministic and white stochastic schemes tended to have regimes which were too persistent. The temporally correlated stochastic parametrisation was able to provide the 'activation energy' required to excite regime transitions, and improved the mean residence time for the two regimes.

Recent work by Dawson and Palmer (2014), described in a companion paper, has demonstrated the improvement in simulated regime behaviour in the GCM operationally used at ECMWF. The introduction of stochastic physics into the model resulted in an improvement in the spatial pattern of North Atlantic weather regimes, as well as on the climatological frequency and occurence of the regimes. This indicates that the results presented here from the Lorenz '96 system are generalisable to a full complexity atmospheric general circulation model.

The "climate change" experiment indicates that stochastic parametrisation schemes improve the response of the system to an external forcing. In particular, the red noise stochastic schemes skilfully indicated the change in proportion of time spent in each regime which occured as a result of the forcing. This kind of experiment can most easily be carried out in a 'toy model' setting, where the true regime behaviour of the system can be calculated before and after the external forcing is applied. The results presented here motivate further testing of stochastic parametrisation schemes in models of intermediate complexity, and indicate that temporally correlated stochastic parametrisation schemes should be incorporated into the next generation of climate models, due to their beneficial impact on the mean state of the modelled climate.

The results from the "climate change" experiment indicate a certain robustness to the results presented for the $F = 20$ case—the red stochastic schemes are skilful at reproducing the climate of related attractors to the L96, $F = 20$ system. Additionally, it is important to recall that the stochastic and perturbed parameter schemes were

derived for the $F = 20$ case, and were not re-tuned to be optimal for the changed climate—the same values of σ, ϕ and b_i are used throughout the integration period. The deterministic parametrisation schemes was able to capture the structure of the attractor for which it was developed, but performed poorly at reproducing the $F = 23$ attractor. In contrast, the stochastic schemes, by representing the possible error in a deterministic scheme, are better suited to be used in situations which they have not been tuned for—they are more robust than deterministic schemes.

The attractor for the perturbed parameter ensemble shows a distinct peak corresponding to Regime B, though this peak is considerably more pronounced than in the truth attractor. When considered individually, the 40 constituent members of the perturbed parameter ensemble differ greatly from the true attractor, with many only showing one dominant regime with very rare transitions. This poor performance is observed despite estimating the degree of perturbation of the parameters from the 'truth' time series, and using fairly moderate perturbations. A perturbed parameter ensemble where the selected parameters vary in time could be more skilful, as it would allow each ensemble member to sample the parameter uncertainty, thus allowing each individual ensemble member to capture the regime behaviour.

The regime behaviour of a climate system provides considerable added information to that available from the pdf. The pdf of the perturbed parameter ensemble in X space, while not as skilful as the red noise stochastic parametrisation schemes, shows considerable skill for the case where the ratio of time scales equals ten, and would not cause concern regarding the climatology of the model. However by performing an EOF decomposition and considering the regime behaviour of the ensemble, the vastly different dynamical behaviour of the perturbed parameter ensemble is revealed. In order to correctly simulate the statistics of the weather (for example, the duration of blocking events over Europe), a climate simulator must accurately represent regime behaviour. It is therefore important that climate models are explicitly tested on this ability. The results presented here indicate that certain members of perturbed parameter ensembles currently used for climate prediction are at risk of failing this test. However, the results presented in this paper also strongly motivate the development of stochastic climate simulators alongside the stochastic numerical weather prediction models currently operationally in use.

Acknowledgments The authors would like to thank Andrew Dawson and Susanna Corti for helpful discussions regarding atmospheric regimes. The research of H.M.C. was supported by a Natural Environment Research Council studentship, and the research of T.N.P. was supported by European Research Council grant number 291406.

References

Arnold HM, Moroz IM, Palmer TN (2013) Stochastic parameterizations and model uncertainty in the Lorenz'96 system. Philos Trans R Soc A 371(1991)

Berner J, Branstator G (2007) Linear and nonlinear signatures in the planetary wave dynamics of an AGCM: probability density functions. J Atmos Sci 64(1):117–136

Berner J, Shutts GJ, Leutbecher M, Palmer TN (2009) A spectral stochastic kinetic energy backscatter scheme and its impact on flow dependent predictability in the ECMWF ensemble prediction system. J Atmos Sci 66(3):603–626

Branstator G, Berner J (2005) Linear and nonlinear signatures in the planetary wave dynamics of an AGCM: phase space tendencies. J Atmos Sci 62(6):1792–1811

Branstator G, Selten F (2009) "Modes of variability" and climate change. J Clim 22(10):2639–2658

Charney J, DeVore J (1979) Multiple flow equilibria in the atmosphere and blocking. J Atmos Sci 36(7):1205–1216

Corti S, Molteni F, Palmer TN (1999) Signature of recent climate change in frequencies of natural atmospheric circulation regimes. Nature 398(6730):799–802

Crommelin D, Vanden-Eijnden E (2008) Subgrid-scale parametrisation with conditional markov chains. J Atmos Sci 65(8):2661–2675

Dawson A, Palmer TN (2014) Simulating weather regimes: impact of model resolution and stochastic parametrisation. Clim Dyn. doi:10.1007/s00382-014-2238-x

Dawson A, Palmer TN, Corti S (2012) Simulating regime structures in weather and climate prediction models. Geophys Res Lett 39(21):?L21805

Dorrestijn J, Crommelin DT, Siebesma AP, Jonker HJJ (2012) Stochastic parameterization of shallow cumulus convection estimated from high-resolution model data. Theor Comp Fluid Dyn 27(1-2):133–148

Ehrendorfer M (1997) Predicting the uncertainty of numerical weather forecasts: a review. Meteorol Z 6(4):147–183

Frame THA, Methven J, Gray SL, Ambaum MHP (2013) Flow-dependent predictability of the North-Atlantic jet. Geophys Res Lett 40(10):2411–2416

Frenkel Y, Majda AJ, Khouider B (2012) Using the stochastic multicloud model to improve tropical convective parametrisation: a paradigm example. J Atmos Sci 69(3):1080–1105

Hasselmann K (1999) Climate change—linear and nonlinear signatures. Nature 398(6730):755–756

Houtekamer PL, Lefaivre L, Derome J (1996) A system simulation approach to ensemble prediction. Mon Weather Rev 124(6):1225–1242

Itoh H, Kimoto M (1996) Multiple attractors and chaotic itinerancy in a quasigeostrophic model with realistic topography: implications for weather regimes and low-frequency variability. J Atmos Sci 53(15):2217–2231

Itoh H, Kimoto M (1997) Chaotic itinerancy with preferred transition routes appearing in an atmospheric model. Physica D 109(3–4):274–292

Kimoto M, Ghil M (1993a) Multiple flow regimes in the northern hemisphere winter. Part I: methodology and hemispheric regimes. J Atmos Sci 50(16):2625–2644

Kimoto M, Ghil M (1993b) Multiple flow regimes in the northern hemisphere winter. Part II: sectorial regimes and preferred transitions. J Atmos Sci 50(16):2645–2673

Kwasniok F (2012) Data-based stochastic subgrid-scale parametrisation: an approach using cluster-weighted modelling. Philos Trans R Soc A 370(1962):1061–1086

Lorenz EN (1963) Deterministic nonperiodic flow. J Atmos Sci 20(2):130–141

Lorenz EN (1996) Predictability—a problem partly solved. In: Proceedings of seminar on predictability, 4–8 September 1995, vol 1. ECMWF, Shinfield Park, Reading, pp 1–18

Lorenz EN (2006) Regimes in simple systems. J Atmos Sci 63(8):2056–2073

Murphy JM, Sexton DMH, Barnett DN, Jones GS, Webb MJ, Collins M, Stainforth DA (2004) Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 430(7001):768–772

Palmer TN (1993) A nonlinear dynamical perpective on climate change. Weather 48(10):314–326

Palmer TN (1999) A nonlinear dynamical perpective on climate prediction. J Clim 12(2):575–591

Palmer TN, Alessandri A, Andersen U, Cantelaube P, Davey M, Délécluse P, Déqué M, Díez E, Doblas-Reyes FJ, Feddersen H, Graham R, Gualdi S, Guérémy J-F, Hagedorn R, Hoshen M, Keenlyside N, Latif M, Lazar A, Maisonnave E, Marletto V, Morse AP, Orfila B, Rogel P, Terres J-M, Thomson MC (2004) Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). Bull Am Meteorol Soc 85(6):853–872

Palmer TN, Buizza R, Doblas-Reyes F, Jung T, Leutbecher M, Shutts GJ, Steinheimer M, Weisheimer A (2009) Stochastic parametrization and model uncertainty, Tech. Rep. 598, European Centre for Medium-Range Weather Forecasts, Shinfield park, Reading

Pohl B, Fauchereau N (2012) The southern annular mode seen through weather regimes. J Clim 25(9):3336–3354

Rougier J, Sexton DMH, Murphy JM, Stainforth D (2009) Analyzing the climate sensitivity of the HadSM3 climate model using ensembles from different but related experiments. J Clim 22:3540–3557

Selten F, Branstator G (2004) Preferred regime transition routes and evidence for an unstable periodic orbit in a baroclinic model. J Atmos Sci 61(18):2267–2282

Stainforth DA, Aina T, Christensen C, Collins M, Faull N, Frame DJ, Kettleborough JA, Knight S, Martin A, Murphy JM, Piani C, Sexton D, Smith LA, Spicer RA, Thorpe AJ, Allen MR (2005) Uncertainty in predictions of the climate response to rising levels of greenhouse gases. Nature 433(7024):403–406

Stephenson DB, Hannachi A, O'Neill A (2004) On the existence of multiple climate regimes. Q J R Meteorol Soc 130(597):583–605

Straus DM, Corti S, Molteni F (2007) Circulation regimes: chaotic variability versus sst-forced predictability. J Clim 20(10):2251–2272

Wilks DS (2005) Effects of stochastic parametrizations in the Lorenz '96 system. Q J R Meteorol Soc 131(606):389–407