

From reliable weather forecasts to skilful climate response: A dynamical systems approach

Hannah M. Christensen  | Judith Berner

National Center for Atmospheric Research,
Boulder, CO, USA

Correspondence

Hannah M. Christensen, Atmospheric, Oceanic
and Planetary Physics, Clarendon Laboratory,
Oxford OX1 3PU, UK.
Email: hannah.christensen@physics.ox.ac.uk

While weather forecasting models can be tested by performing and evaluating many hindcasts, the limited observational record restricts the degree to which climate projections can be evaluated. Therefore a question of interest is: to what degree can we evaluate the potential skill of a climate model's response to forcing by assessing the reliability of short-range weather and seasonal forecasts produced by the same model? We address this question using a dynamical systems framework. We use linear response theory to provide the mean climate response of a general dynamical system to a small external forcing. We relate this response to the reliability of initial value forecasts. We find that, in order to capture the mean climate response, the forecast model must correctly represent the slowest evolving modes of variability in the system. The reliability of forecasts on seasonal and longer time-scales, which is sensitive to the representation of these slow modes, could therefore indicate if the forecast model has the correct climate sensitivity and so will respond correctly to an applied external forcing. In this way, the skill of initialized forecasts could act as an 'emergent constraint' on climate sensitivity. However, we also highlight that unreliable seasonal forecasts do not necessarily indicate an incorrect climate projection. This is because correctly representing rapidly evolving modes is also necessary for reliable seasonal forecasts.

KEYWORDS

climate prediction, climate sensitivity, dynamical systems theory, emergent constraint, initial value forecasts, reliability, response theory, seamless prediction

1 | INTRODUCTION

The "seamless prediction" paradigm has gained much momentum in recent years. Since accurate weather and climate prediction both depend on the representation of fast physical processes in the climate system, it has been proposed that the weather and climate communities should work in collaboration to produce and evaluate Earth-system models (Brunet *et al.*, 2010). The nonlinear nature of the atmospheric equations of motion results in upscale transfer of errors from smaller to larger scales (Lorenz, 1969; Palmer, 2001; Tribbia and Baumhefner, 2004). However slower evolving modes can also provide predictability on shorter time-scales (Hoskins,

2013; Vannitsem and Lucarini, 2016). A unified approach to evaluating weather and climate models can therefore provide insight for both communities by evaluating models across a wide range of time-scales (Phillips *et al.*, 2004; Hurrell *et al.*, 2009). This development has been made possible by advances in computational resources, which has allowed the resolution of global models to approach that of the numerical weather prediction community (Haarsma *et al.*, 2016; Roberts *et al.*, 2018), and which has enabled weather forecast models to be integrated out towards the long lead times of interest to the climate community.

Following the seamless prediction paradigm, several authors have initialized Earth-system models for production

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

of short-range forecasts, where verification is possible, before using these models for climate applications (Phillips *et al.*, 2004; Hurrell *et al.*, 2009). This approach has proven useful in diagnosing fast physics biases (e.g. Hannay *et al.*, 2009; Medeiros *et al.*, 2012; Williams *et al.*, 2013), as models are observed to develop biases over days to weeks that are similar to their climatic biases. This indicates that the subgrid-scale physics schemes are the source of the biases in the climate model, and allows attribution of the error to a particular parametrization scheme. Palmer *et al.* (2008) and Weisheimer and Palmer (2014) go a step further and argue that the *reliability* of a model on shorter time-scales is a necessary criterion for skilful climate *projections*. The authors argue that, due to the nonlinearity of the climate system, certain processes (such as radiative feedbacks) influence both seasonal and climate time-scales. The fidelity of these processes should be evaluated by considering short-range seasonal forecasts, before using these forecasts to calibrate climate change projections (Palmer *et al.*, 2008; Weisheimer and Palmer, 2014; Palmer and Weisheimer, 2018). This is an attractive proposal: instead of waiting decades to verify climate projections, models can be evaluated today using initialized forecasts. This information can be used to constrain Earth's climate sensitivity by excluding models which perform poorly with respect to this metric, providing information about likely anthropogenic climate change. In other words, it is proposed that the skill of initialized forecasts could be used as an 'emergent constraint' (Klein and Hall, 2015), that is, an empirical relationship between characteristics of a model's behaviour in today's climate and that model's climate sensitivity. However, no rigorous theory has been presented to justify linking the reliability of short-range forecasts with the accuracy of climate change projections. In this paper, we seek to address this problem.

A natural framework for considering the impact of a small forcing perturbation on the state of a system is linear response theory. For example, the fluctuation–dissipation theorem (FDT) recognises the intimate relationship between natural fluctuations that excite the system of interest, and the damping effect of these fluctuations on the system's response to an external forcing (Kubo, 1966). Through this relationship, the properties of the fluctuations in the system can be used to predict the system's response to a forcing. Leith (1975) was the first to apply these ideas to the climate system, expressing the climatic response of a system in equilibrium to a constant external forcing. Response theory and FDT have been applied to the climate system with some success (Cionni *et al.*, 2004; Gritsun and Branstator, 2007; Cooper and Haynes, 2011; Cooper *et al.*, 2013; Achatz *et al.*, 2013). Further work has considered a general linear response theory for systems far from equilibrium (Ruelle, 1998; 1999; Lucarini and Sarno, 2011; Lucarini *et al.*, 2014; Gritsun and Lucarini, 2017), demonstrating that many of the classical results at equilibrium still hold.

On the other hand, a natural framework for considering predictability of initial value forecasts on short time-scales is

dynamical systems theory (e.g. Smith *et al.*, 1999 give a summary). There is a wealth of literature describing the use of dynamical systems theory to understand the atmospheric predictability problem (e.g. Lorenz, 1996; Palmer, 1999; Franzke *et al.*, 2015), with simple dynamical systems such as those of Lorenz (1963) and Lorenz (1996) becoming widely used (e.g. Arnold *et al.*, 2013; Trevisan and Legnani, 1995; Smith *et al.*, 1999; James, 2002; Wilks, 2005; Vissio and Lucarini, 2018). The framework is particularly useful for understanding how small perturbations grow in time, and project most strongly in preferred directions. In this manuscript, we use the phrase *mode* to refer to a preferred pattern of growth (or variability) in the system in question. Such a mode will have an associated time-scale of growth. Examples could include baroclinic instability or the El Niño–Southern Oscillation. On the other hand, *processes* are physical phenomena, such as cloud microphysics or percolation of water through soil. While processes typically occur on a particular time-scale and in a particular location, they may impact the evolution of modes with a wide range of space- and time-scales.

In this paper, we seek to draw together the existing body of literature on response theory and dynamical systems to explore whether there is a link between the reliability of weather forecasts and the response to external forcing. This will enable us to explore the possibility of using initial value forecasts as an emergent constraint for assessing the climate characteristics of a model in the way described by Palmer *et al.* (2008). In Section 2, we review the existing literature concerning error growth in initialized forecasts, and thereby define the properties of a forecast model that will enable it to produce reliable weather forecasts. In Section 3, the response of a dynamical system to an applied forcing is derived as an analogy for the climate change problem, and the resultant response equations validated in the Lorenz (1963) dynamical system. In Section 4 we draw on results from the previous sections to discuss the relationship between forecast reliability and climate sensitivity: in Section 5 the more complex two-scale Lorenz (1996) dynamical system is introduced as a testbed to evaluate this proposed relationship. Finally, the results are discussed and conclusions are drawn in Section 6.

2 | THE DYNAMICAL SYSTEMS FRAMEWORK

Our discussion is largely framed using dynamical systems theory. A dynamical system is described by a set of nonlinear ordinary differential equations:

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}), \quad (1)$$

where the system's m -dimensional state vector is denoted in bold \mathbf{x} . The state vector characterises the state of the dynamical system: given an initial condition, \mathbf{x}_0 , and the equations of motion, \mathbf{F} , the future trajectory of the system is completely determined.

Error growth in dynamical systems, with a view to the atmospheric forecasting problem, is expounded in Smith *et al.* (1999). Here we provide only a brief summary. First we define the Jacobian, \mathbf{J} , of the dynamical system in Equation 1:

$$(J_{ij})_{\mathbf{x}} = \left. \frac{\partial F_i}{\partial x_j} \right|_{\mathbf{x}},$$

which will be written

$$\mathbf{J}_{\mathbf{x}} = \left. \frac{\partial \mathbf{F}}{\partial \mathbf{x}} \right|_{\mathbf{x}}. \quad (2)$$

The Jacobian gives the linearized equations of motion for the system. It summarizes the instantaneous local growth of infinitesimal uncertainties in the system, and thereby governs how small perturbations to the dynamical system, such as errors in the initial conditions, $\epsilon = \mathbf{x} - \mathbf{x}_0$, will evolve in time:

$$\dot{\epsilon} = \mathbf{J}_{\mathbf{x}} \epsilon. \quad (3)$$

Correctly representing the Jacobian at all locations in phase space will ensure correct initial error growth in the forecast model.

Note that, since \mathbf{J} is non-normal in general, it has non-orthogonal eigenvectors. Non-normality means that, even if \mathbf{J} has only *negative* eigenvalues, the system can still display instantaneous error *growth* (Smith *et al.*, 1999). Furthermore, the fastest growing mode will not necessarily be excited by a perturbation along the leading eigenvector. The instantaneous growth rate will evolve in time, and asymptote towards the leading eigenvalue of \mathbf{J} . For this reason, it is also helpful to introduce the *tangent propagator*, which maps a perturbation at t_0 to that observed at a later time, t (Molteni and Palmer, 1993; Smith *et al.*, 1999):

$$\epsilon_t = \mathbf{M}(t, t_0, \mathbf{x}_0) \epsilon_0. \quad (4)$$

Note that we have made no linearity assumptions in defining \mathbf{M} and that, following this definition, \mathbf{M} is defined for all t , without requiring ϵ_t to remain small. However, while ϵ_t does remain small, \mathbf{M} can be approximated by:

$$\mathbf{M}(t, t_0, \mathbf{x}_0) = \exp \left[\int_{t_0}^t \mathbf{J}_{t'} dt' \right]. \quad (5)$$

The magnitude of the perturbation, ϵ_t , is given by

$$\|\epsilon_t\| = \sqrt{\epsilon_t^T \epsilon_t} = \sqrt{\epsilon_0^T \mathbf{M}(t, t_0, \mathbf{x}_0)^T \mathbf{M}(t, t_0, \mathbf{x}_0) \epsilon_0}. \quad (6)$$

While \mathbf{M} is not necessarily symmetric, $\mathbf{M}^T \mathbf{M}$ is symmetric by construction, such that the eigenvalues of $\mathbf{M}^T \mathbf{M}$ form an orthogonal basis set. The singular value decomposition of \mathbf{M} is $\mathbf{M} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$. The matrix \mathbf{V} contains the right singular vectors of \mathbf{M} , which describe the perturbation directions which will maximize the rate of growth of $\|\epsilon_t\|$. The leading left singular vectors contained in \mathbf{U} define the direction of this optimal growth. The diagonal matrix, $\mathbf{\Sigma}$, contains the singular values, σ_i . The rate of growth of $\|\epsilon_t\|$ is governed by the eigenvalues of $\mathbf{M}^T \mathbf{M}$, given by σ_i^2 , also known as the local finite-time *Lyapunov exponents* (Smith *et al.*, 1999). In the

limit $t \rightarrow \infty$, the finite time Lyapunov exponents converge to the global Lyapunov exponents.

2.1 | Reliability in dynamical systems

Let us now consider *reliability* in the dynamical systems framework, as this was highlighted by Palmer *et al.* (2008) as a potential predictor of the climate prediction skill of Earth-system models. A probabilistic forecast is reliable if the forecast probability distribution function (pdf) is a good indicator of the uncertainty in the forecast. In other words, given a particular forecast pdf, the verification (i.e. observation) will behave as if it is drawn from that predicted distribution (Wilks, 2006) at a range of lead times. To assess reliability, statistics are aggregated from many initial start dates to test whether the verification follows the distribution predicted by the ensemble. However, this sampling also samples over the system's climate distribution. Stratification of forecasts by predicted ensemble spread (Leutbecher and Palmer, 2008; Christensen *et al.*, 2015a) is a first step towards removing this averaging over the attractor, as a partial test of state-dependent reliability.

A probabilistic forecast is generated by creating an ensemble (or set) of forecasts, where individual members of this ensemble explore the sources of uncertainty present in the forecast (Slingo and Palmer, 2011). Dynamical systems theory can be used to consider the rate of growth of small perturbations between different ensemble members, and is therefore a natural framework for discussing reliability. For a reliable forecast, the expected divergence of two randomly selected ensemble members must match the expected rate at which the observed truth diverges from a randomly selected ensemble member. Figure 1 shows a schematic of the scenario of interest, indicating the evolution of the ensemble forecast and observed truth, given a degree of uncertainty in the initial conditions. Note that, since the truth and model forecast live in different spaces, one must first perform an operation of projection to compare the two in this way. We assume for simplicity that the initial condition uncertainty is known, and is correctly accounted for in the ensemble forecast. While this is a major assumption, it allows us to focus on the dynamical properties of the forecast system. The reliability of the estimated initial condition uncertainty can be assessed separately (Rodwell *et al.*, 2016).

Using the superscript t for the “true” system, and f for the forecast model, consider the difference between the truth and one ensemble member:

$$\begin{aligned} \delta \mathbf{x}(t) &= \mathbf{e}^f(t) + \mathbf{b}(\mathbf{x}, t) - \mathbf{e}^t(t) \\ &= \mathbf{M}^f(t, 0, \mathbf{x}_0) \epsilon^f(0) + \mathbf{b}(\mathbf{x}, t) - \mathbf{M}^t(t, 0, \mathbf{x}_0) \epsilon^t(0), \quad (7) \end{aligned}$$

which includes the impact of a potential state-dependent model error, $\mathbf{b}(\mathbf{x}, t)$ (i.e. the difference between the dark red and pale blue dashed lines in Figure 1), as well as the random component of error growth due to the uncertainty in the

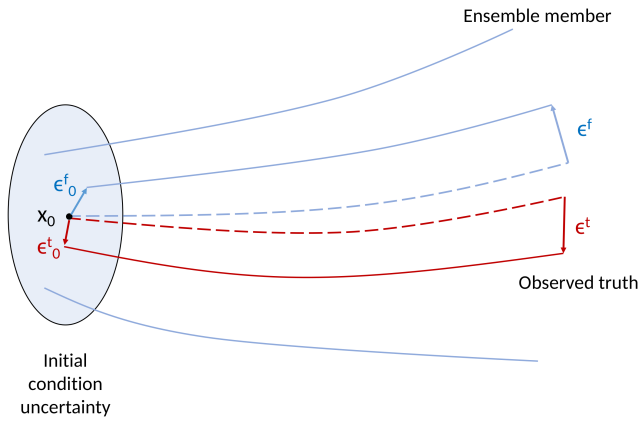


FIGURE 1 Schematic of reliability in a dynamical system framework. The blue oval indicates initial condition uncertainty, the centre of which is marked by the reference point \mathbf{x}_0 . Pale blue lines represent ensemble forecast members, while the dark red line represents the observed truth, with time increasing towards the right. The dashed pale blue and dashed dark red lines represent the time evolution of the reference point \mathbf{x}_0 using the forecast model and “true” equations respectively: the difference between the dashed lines is due to model error [Colour figure can be viewed at wileyonlinelibrary.com]

initial conditions. In contrast, two randomly selected ensemble members will diverge following

$$\delta\epsilon^f(t) = \mathbf{M}^f(t, 0, \mathbf{x}_0)\delta\epsilon^f(0), \quad (8)$$

where $\delta\epsilon^f(0)$ was their initial separation.

All methods for evaluating reliability assess whether the verification behaves similarly to the members of the ensemble forecast. From Equations 7 and 8, it is evident that assessing reliability, that is, by comparing the rate of divergence of ensemble members with the rate of divergence of the verification from the ensemble, will indicate whether the forecast tangent propagator, $\mathbf{M}^f(t, 0, \mathbf{x}_0)$ is adequately representing the “true” tangent propagator, $\mathbf{M}^t(t, 0, \mathbf{x}_0)$, as long as the model error term $\mathbf{b}(\mathbf{x}, t)$ is small.

3 | PREDICTING THE FORCED RESPONSE OF A DYNAMICAL SYSTEM

We next consider the climate change problem: the goal is to predict how the climate of a dynamical system will respond to an imposed forcing. Consider applying a small forcing perturbation $\delta\mathbf{F}$ to the dynamical system in Equation 1.

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}) + \delta\mathbf{F}. \quad (9)$$

The unperturbed system could represent the Earth’s climate system before 1850, while the perturbed system would include the additional forcing from anthropogenic emissions of greenhouse gases. In general the forcing perturbation, $\delta\mathbf{F}$, is allowed to be a function of the state of the system, $\delta\mathbf{F}(\mathbf{x})$, but this need not be the case.

The FDT seeks a linear relationship between a small applied forcing and the mean climate response (Gritsun and Branstator, 2007). The theory predicts that such a response of

a system to a forcing can be written, following the notation in Cooper and Haynes (2011):

$$\delta\mathbf{x}(t) = - \int_0^t \left\langle \mathbf{x}(\tau) \left[\frac{\nabla_{\mathbf{x}}\rho(\mathbf{x}(\mathbf{0}))}{\rho(\mathbf{x}(\mathbf{0}))} \right]^T \right\rangle \delta\mathbf{F}(t - \tau) d\tau, \quad (10)$$

where the system’s pdf is written $\rho(\mathbf{x})$, and we assume that the pdf is differentiable. The angle brackets indicate the expectation taken over the attractor, for example, through taking an ensemble mean over a large number of initial conditions. The response at time t , $\delta\mathbf{x}(t)$, is determined by the integrated effect of forcing perturbations at all earlier times, $t - \tau$. Note that the “linearity” of the response theory stems from the assumption that the forcing perturbation is sufficiently small for the response to be linearly proportional to the forcing. It makes no assumptions about the dynamics of the system. Leith (1975) made the assumption that $\rho(\mathbf{x})$ can be approximated by a Gaussian, and so expressed the climatic response of a system in equilibrium to a constant external forcing as:

$$\delta\mathbf{x} = \Theta\delta\mathbf{F}, \quad (11)$$

where

$$\Theta = \int_0^\infty \mathbf{C}_\tau \mathbf{C}_0^{-1} d\tau, \quad (12)$$

where \mathbf{C}_τ and \mathbf{C}_0 are the covariance matrices at lag τ and 0 respectively.

3.1 | Ruelle’s response theory

For some systems the unforced climatological pdf is undefined. For example, a climatological pdf does not exist for any chaotic system which explores a strange attractor. For these systems, Ruelle (1998) derived a generalized response theory which showed that many of the classical results still hold. Here we will re-derive the linear response theory of Ruelle, making some simplifications, and using the notation already defined in this paper to highlight the links to dynamical systems theory.

As before, let a dynamical system be represented $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$. Let a small forcing perturbation be turned on at time t_0 . The system can now be written:

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}) + \zeta(t)\mathbf{f}(\mathbf{x}), \quad (13)$$

where the small forcing perturbation has explicitly been split into a time-dependent and space-dependent component. Following Ruelle (1998), we can write the change in the system’s trajectory, $\delta\mathbf{x}$, over the time window (t_0, t) due to the forcing as

$$\delta\mathbf{x}(t_0, t) = \int_{t_0}^t dt' \mathbf{M}(t, t', \mathbf{x}_{t'}) \zeta(t') \mathbf{f}(\mathbf{x}_{t'}), \quad (14)$$

which includes the familiar tangent propagator (equivalently, linear resolvent operator), $\mathbf{M}(t, t', \mathbf{x}_{t'})$ from Equation 4. We recognise that $\mathbf{M}(t, t', \mathbf{x}_{t'})$ acts as a Green’s function, mapping the influence of a small perturbation at time t' onto the

state of the system at some later time, t . Next, defining the expectation value of an observable of interest, Φ , as¹

$$\langle \Phi \rangle = \int dx \rho(\mathbf{x}) \Phi(\mathbf{x}), \quad (15)$$

and, following Ruelle (1998), we can write the change in the expected observable due to the forcing perturbation as a function of time, t , as

$$\delta \langle \Phi \rangle_t = \int_{-\infty}^t dt' \int dx \rho(\mathbf{x}) \mathbf{M}(t, t', \mathbf{x}_{t'}) \zeta(t') \mathbf{f}(\mathbf{x}_{t'}) \cdot \nabla_{\mathbf{x}(t-t')} \Phi,$$

where we have also extended the integral in time to consider the response due to all past perturbations, and the expectation is calculated over the climate distribution of the unperturbed system, $\rho(\mathbf{x})$. Instead of proceeding for a general observable, consider only the response of the mean, $\Phi = \mathbf{x}$. While this is a simplification, even constraining the mean response of the Earth system (the climate sensitivity) is of substantial interest. The equation simplifies to

$$\delta \langle \mathbf{x} \rangle_t = \int_{-\infty}^t dt' \int dx \rho(\mathbf{x}) \mathbf{M}(t, t', \mathbf{x}_{t'}) \zeta(t') \mathbf{f}(\mathbf{x}_{t'}). \quad (16)$$

If we make the assumption that the function $\mathbf{F}(\mathbf{x})$ in Equation 1 has no time dependence, we can replace the pair of time indices (t, t') in \mathbf{M} with a single time lag, $\tau = t - t'$. Changing the variable of integration, we find

$$\delta \langle \mathbf{x} \rangle_t = \int_0^{\infty} d\tau \int dx \rho(\mathbf{x}) \mathbf{M}(\tau, \mathbf{x}) \zeta(t - \tau) \mathbf{f}(\mathbf{x}). \quad (17)$$

Increasing the new index of integration, τ , represents looking further into the past, mapping the impact of past perturbations on to the present state of the system at a time lag $\tau = 0$.

We will ultimately consider the case of a time-independent forcing. However it is instructive to first consider the more general case of time-varying forcing. We proceed by taking the Fourier transform of Equation 17

$$\begin{aligned} & \int_{-\infty}^{\infty} dt \delta \langle \mathbf{x} \rangle_t e^{i\omega t} \\ &= \int_{-\infty}^{\infty} dt \int_0^{\infty} d\tau \int dx \rho(\mathbf{x}) \mathbf{M}(\tau, \mathbf{x}) \zeta(t - \tau) \mathbf{f}(\mathbf{x}) e^{i\omega\tau} e^{i\omega(t-\tau)} \\ &= \int_0^{\infty} d\tau \int dx \rho(\mathbf{x}) \mathbf{M}(\tau, \mathbf{x}) \mathbf{f}(\mathbf{x}) e^{i\omega\tau} \cdot \tilde{\zeta}(\omega), \end{aligned} \quad (18)$$

where we have written the Fourier transform of $\zeta(t)$ as $\tilde{\zeta}(\omega)$. We have arrived at a key result of response theory, namely that the response of a dynamical system to a small forcing perturbation is local in frequency space, such that

$$\tilde{\delta} \langle \mathbf{x} \rangle_{\omega} = \tilde{\kappa}(\omega) \tilde{\zeta}(\omega), \quad (19)$$

where the linear susceptibility, $\tilde{\kappa}(\omega)$, is

$$\tilde{\kappa}(\omega) = \int_0^{\infty} d\tau \int dx \rho(\mathbf{x}) \mathbf{M}(\tau, \mathbf{x}) \mathbf{f}(\mathbf{x}) e^{i\omega\tau}. \quad (20)$$

As a thought experiment, if the system were linear, we note that the tangent propagator could be written $\mathbf{M}(\tau, \mathbf{x}) = \exp[\mathbf{L}\tau]$, where \mathbf{L} is independent of \mathbf{x} . Only the forcing, $\mathbf{f}(\mathbf{x})$, depends on \mathbf{x} . Writing the forcing averaged over the attractor as $\langle \mathbf{f} \rangle$ and integrating,

$$\tilde{\kappa}(\omega) = \left[\frac{\exp[(i\omega + \mathbf{L})\tau] \langle \mathbf{f} \rangle}{i\omega + \mathbf{L}} \right]_0^{\infty} \quad (21)$$

$$= -\frac{\langle \mathbf{f} \rangle}{i\omega + \mathbf{L}}, \quad (22)$$

where we have evaluated the upper limit using the fact that the unperturbed system is in a steady state: perturbations must be damped over long time lags.

Returning to the nonlinear case, it is difficult to proceed more generally from Equation 20. However, we can first consider the case of a state independent forcing perturbation, $\mathbf{f}(\mathbf{x}) = \mathbf{f}$. Defining $\overline{\mathbf{M}}(\tau)$ as the integral of $\mathbf{M}(\tau, \mathbf{x})$ over the attractor, the linear susceptibility reduces to

$$\tilde{\kappa}(\omega) = \int_0^{\infty} d\tau \overline{\mathbf{M}}(\tau) \mathbf{f} e^{i\omega\tau}. \quad (23)$$

The response to forcing of a system is given by the integral under the $\overline{\mathbf{M}}(\tau)$ curve, where $\overline{\mathbf{M}}(\tau)$ expresses how small perturbations to the system affect the average evolution of the system. For a forcing that is constant in time, the linear susceptibility can be written

$$\tilde{\kappa}(0) = \int_0^{\infty} d\tau \overline{\mathbf{M}}(\tau) \mathbf{f}. \quad (24)$$

For a forcing that is constant in time, the response is also constant in time, as predicted by Equation 19.

3.2 | Response to finite forcing in a nonlinear system: the Lorenz '63 System

The extent to which the response Equation 24 is true for a non-infinitesimal external forcing is evaluated in an example dynamical system, the Lorenz (1963; L63) system:

$$\begin{aligned} \dot{X} &= -\sigma X + \sigma Y, \\ \dot{Y} &= -XZ + rX - Y, \\ \dot{Z} &= XY - bZ. \end{aligned} \quad (25)$$

The system is a simplified model of Rayleigh–Bénard convection, where X describes the intensity of the convection, Y is proportional to the difference in temperature between ascending and descending plumes, and Z describes the deviation of the vertical temperature profile from its unperturbed state. The parameters σ, r and b are the Prandtl number, the Rayleigh number and a parameter characterizing the size of the convective cell. The standard values of $\sigma = 10, r = 28$ and $b = 8/3$ produce the familiar Lorenz “butterfly” attractor, shown in Figure 2a. The L63 system is commonly used in atmospheric science as a testbed both for atmospheric predictability studies, and for testing new mathematical techniques, including data assimilation and sensitivity analysis (e.g. Miller *et al.*, 1994; Trevisan and Legnani, 1995;

¹For simplicity of notation, we represent a measure on the state space as $\rho(\mathbf{x})d\mathbf{x}$. Note that this assumes the existence of a probability density function $\rho(\mathbf{x})$, which is not generally true, for example in the case of the strange attractor of a chaotic system. In that case, the notation $\rho(\mathbf{x})d\mathbf{x}$ can be interpreted as shorthand for the mathematically correct $\rho(d\mathbf{x})$ (Eyink *et al.*, 2004).

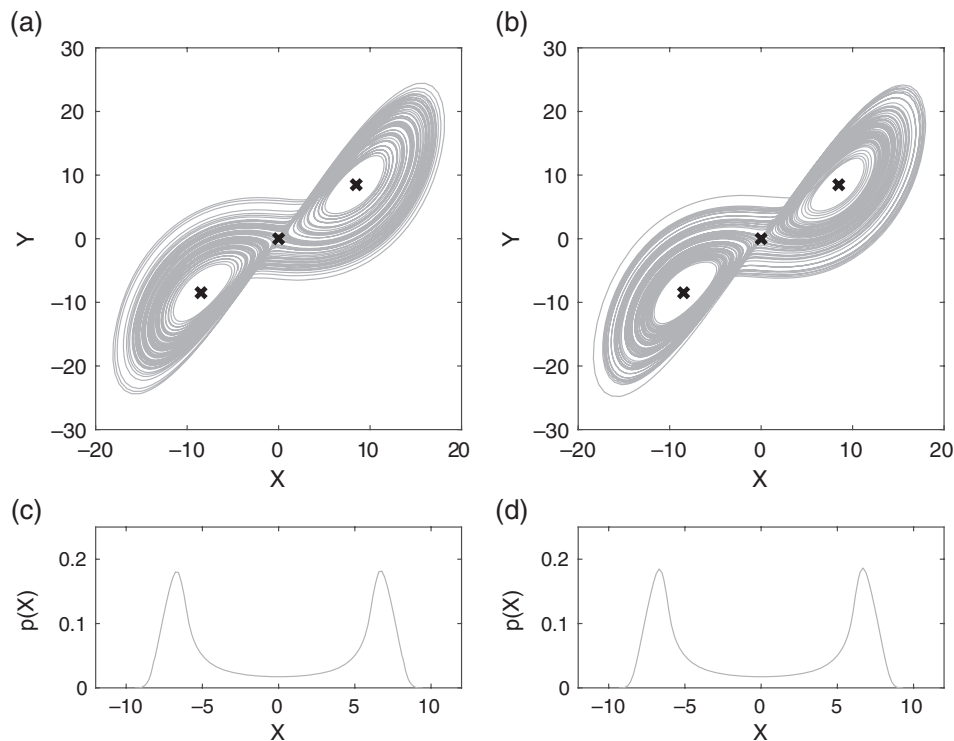


FIGURE 2 The response of the Lorenz system to a forcing perturbation. A sample of the trajectory of the (a) unforced and (b) forced system (grey lines) and their respective fixed points (black crosses) in the X - Y plane. The corresponding marginal probability density functions of (c) the unforced system and (d) the forced system in the X direction. The Lorenz system is forced following Equation 26 using $f = 10$ and $\theta = 3\pi/4$

Smith *et al.*, 1999; Lucarini, 2009; Van Leeuwen, 2010; Marzban, 2013).

Consider the impact of a small forcing perturbation to the evolution of the L63 system. Following Palmer (1999), a possible small forcing perturbation to that system can be written:

$$\begin{aligned}\dot{X} &= -\sigma X + \sigma Y + f \cos \theta, \\ \dot{Y} &= -XZ + rX - Y + f \sin \theta, \\ \dot{Z} &= XY - bZ.\end{aligned}\quad (26)$$

The L63 system is integrated for one million model time units (MTU) using a fourth-order Runge–Kutta scheme with a timestep of 0.01. Figure 2a shows a sample of the trajectory from an unforced simulation with $f = 0$. The response of the Lorenz system to a forcing of $f = 10$ and $\theta = 3\pi/4$ is shown in Figure 2b. The time series are smoothed using a 0.7 model time unit running average for both the forced and unforced systems, and the pdfs of the smoothed time series are shown in Figure 2c,d. While (b) indicates that the overall “shape” of the attractor is unchanged, (d) shows that the system now spends more time in one wing of the attractor than the other.

We test the linearity of the true response to increasing forcing by producing a set of forced simulations, each with a different value of f . Figure 3 shows the observed response in each of the X , Y and Z directions as a function of f . For X and Y , the climate response is remarkably linear with magnitude of the imposed forcing, even up to large values of $f = 10$. For Z , the climate response is not linear: for small f , δZ is zero, while for larger f , the response is small but positive. Note that,

interestingly, the forcing on X is negative, but the response is positive (Palmer and Weisheimer, 2011).

To test whether the response Equation 24 can predict the mean climate response of the Lorenz system, we must estimate the tangent operator averaged over the attractor for the *unforced* system, $\overline{\mathbf{M}}(\tau)$, as a function of lag, τ . A long unforced simulation is produced spanning 10^8 MTU. Every 2 MTU, a set of three simulations are branched from the first, each of duration 10 MTU. The starting conditions of these three simulations are generated from the parent simulation, with an additional initial condition perturbation of $(\epsilon, 0, 0)$, $(0, \epsilon, 0)$, and $(0, 0, \epsilon)$ respectively, where ϵ is set to 0.01. The growth of the initial condition perturbation is used to estimate $\mathbf{M}(\tau, \mathbf{x})$ following Equation 4. The mean is calculated over the samples as an estimate of the average over the attractor, $\overline{\mathbf{M}}(\tau)$. We also calculate the mean for ten sub-samples, to indicate the uncertainty in our estimate.

Having estimated $\overline{\mathbf{M}}(\tau)$, the predicted climate response is calculated following Equation 24. Figure 4a–c shows the projection of the forcing through the tangent propagator, $\overline{\mathbf{M}}(\tau)\mathbf{f}$, as a function of time lag, in each of the X , Y and Z directions. The integration of this quantity over τ is shown in (d)–(f), which will tend towards the solution of the response equation as the sample size increases, and as the upper limit, α , is increased towards ∞ .

We compare this predicted response to the observed response in the L63 system shown in Figure 3. In each panel, the circles with error bars at i, j, and k indicate the predicted response for $f = 10$, estimated at $\alpha = 4, 6$ and 9 MTU

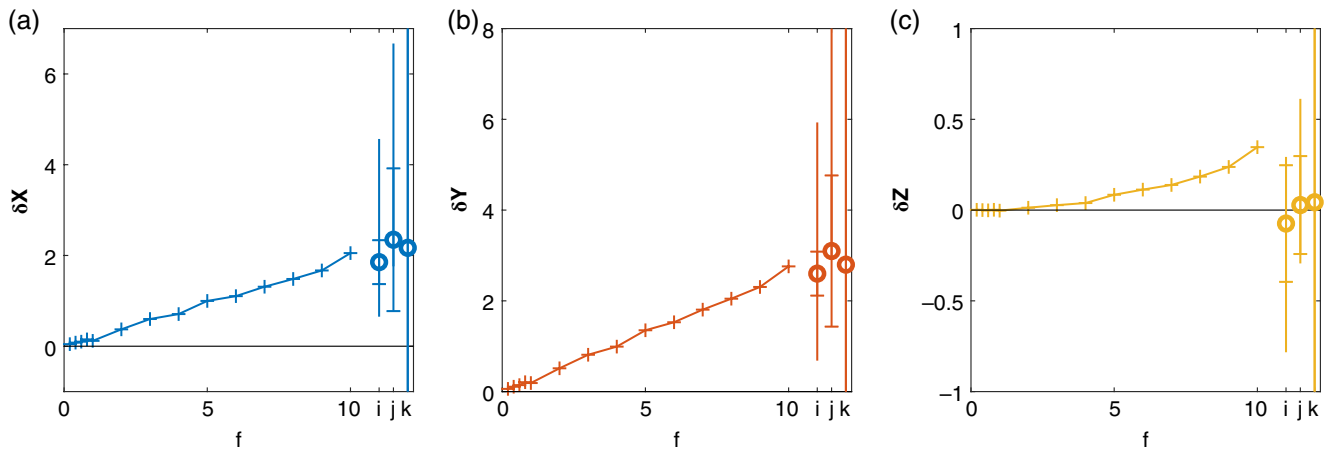


FIGURE 3 The simulated response of the mean of the L63 system as a function of the forcing perturbation, f : (a) δx ; (b) δy ; (c) δz . For each panel, the circles with error bars at i , j , and k indicate the response predicted by the response equation for $f = 10$, estimated at α of 4, 6 and 9 MTU respectively. Error bars show the uncertainty in the predicted response as indicated by the sub-samples in Figure 4 [Colour figure can be viewed at [wileyonlinelibrary.com](#)]

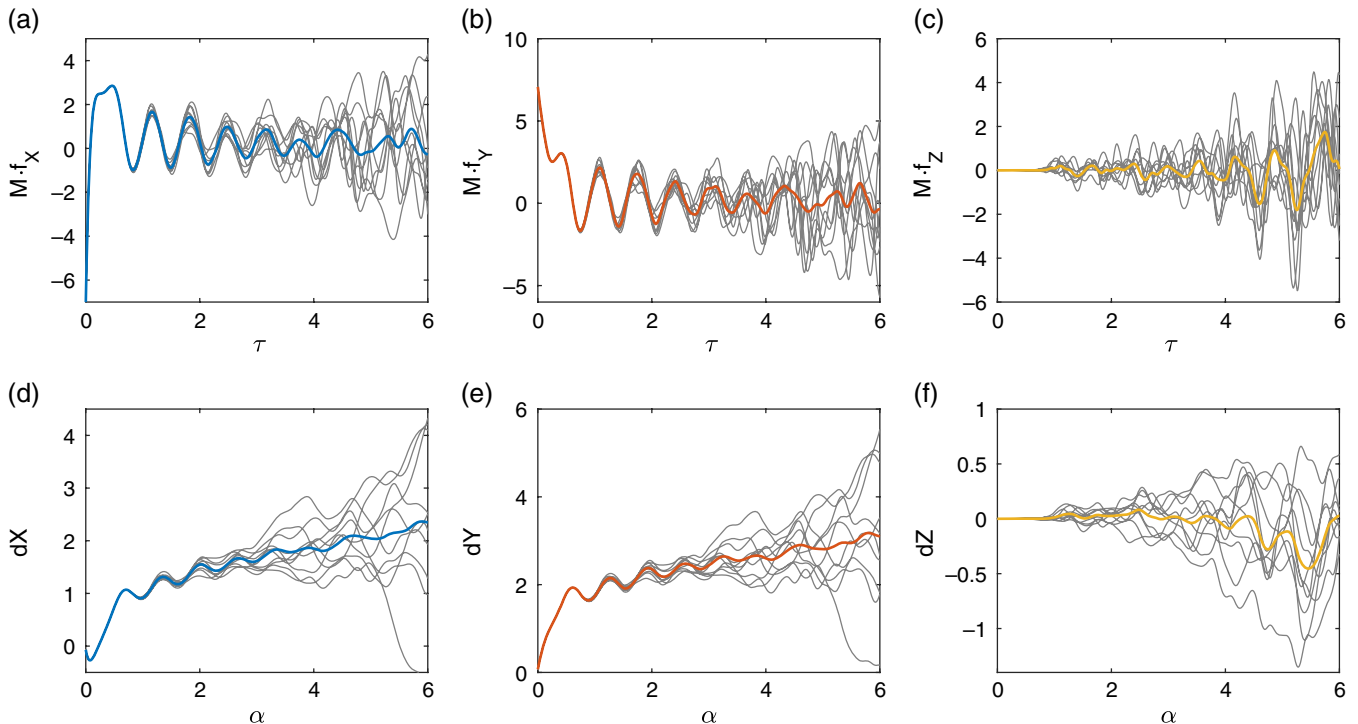


FIGURE 4 The projection of the forcing perturbation through the attractor-averaged tangent propagator for the L63 system, $\overline{\mathbf{M}}(\tau) \cdot \mathbf{f}$, in the (a) X , (b) Y and (c) Z directions, as a function of lag τ . The predicted response of the L63 system estimated as $\int_0^\alpha d\tau \overline{\mathbf{M}}(\tau) \cdot \mathbf{f}$, as a function of the upper limit of integration, α , in the (d) X , (e) Y and (f) Z directions. As α increases, the predicted response tends to the true value, but the uncertainty in the response increases. The thin grey lines indicate the quantities estimated from ten sub-samples of the simulations, while the thick coloured lines indicate the average over all samples [Colour figure can be viewed at [wileyonlinelibrary.com](#)]

respectively. The error bars show uncertainty in the predicted response: the bold capped error bars indicate the $\pm 1\sigma$ interval, while the thin uncapped bars indicate the maximum and minimum estimates from the subsample. The response equation gives a good prediction of the climatic response to an imposed forcing in the Lorenz system in the X and Y directions, even for forcings as large as $f = 10$ (note for comparison that the standard deviations of X , Y , and Z are 9.0, 9.0 and 8.6 respectively). In the Z direction, the response equation predicts a climatic response of $\delta Z = 0$, which is

observed for small f . However by $f = 10$, the response in Z has become nonlinear.

To achieve a good estimate of $\kappa(0)$, we must estimate $\overline{\mathbf{M}}(\tau)$ out to large lags τ . Unfortunately, as the lag increases, the sample size needed to adequately sample the attractor becomes prohibitively large, even for a system as simple as the Lorenz system. This problem is well known (Eyink *et al.*, 2004). Several approaches have been proposed to estimate $\overline{\mathbf{M}}(\tau)$ at long lags, and thereby calculate the response, though each approach is limited by necessary assumptions

and approximations (Eyink *et al.*, 2004; Thuburn, 2005; Abramov and Majda, 2007; 2008; Cooper and Haynes, 2011).

4 | THE PROPOSED LINK BETWEEN WEATHER FORECASTS AND CLIMATE PREDICTION

Equation 24 provides a useful framework for considering the mean response of a dynamical system to a forcing perturbation. It indicates that a forecast model which correctly represents the tangent propagator for each state on the attractor, and therefore also its integral $\bar{\mathbf{M}}(\tau)$, will respond to forcing in a similar manner to the “true” system. Section 2.1 discussed how error growth statistics in initial value forecasts, assessed by measuring forecast reliability, are also dependent on correct representation of the tangent propagator. Does this indicate that the reliability of existing simulations performed by weather forecasting centres could be used to evaluate if that forecast model would produce the correct forced climate response?

For a short-range initial value forecast on time-scales of hours to days, the divergence of ensemble members will be dominated by the fastest growing modes (the leading singular vectors of $\mathbf{M}(\tau)$ for a τ of hours to days), so capturing these modes correctly is sufficient for reliable short-range forecasts. Slower growing modes will also result in ensemble divergence, but on longer time-scales. Short-range weather forecasts are not long enough to detect the fidelity of these slower growing modes in a forecast model. However, correctly predicting the response to forcing requires an accurate representation of $\mathbf{M}(\tau)$ for these longer time-scales. Furthermore, for the L63 system, Figure 4 shows that the behaviour of $\mathbf{M}(\tau)$ at the shortest time-scales (< 0.5 MTU) is very different to the behaviour at longer time-scales, so in that system at least, short-range reliability does not necessarily indicate long-range reliability.

Following an alternative line of reasoning, we have seen that for a linear system the response Equation 22 is given by $\delta\langle\mathbf{x}\rangle_t = -\mathbf{L}^{-1}\mathbf{f}$ whereas small perturbations grow following Equation 3; $\dot{\epsilon} = \mathbf{L}\epsilon$. Returning to the equation for linear susceptibility (20), let us write $\bar{\mathbf{M}}(\tau) \approx \exp[\bar{\mathbf{J}}_{\text{eff}}\tau]$ (following Equation 5), where we have introduced the τ -independent effective average Jacobian, $\bar{\mathbf{J}}_{\text{eff}}$. Substituting this into Equation 20, assuming constant forcing and integrating gives $\delta\langle\mathbf{x}\rangle_t = -\bar{\mathbf{J}}_{\text{eff}}^{-1}\mathbf{f}$, analogously to the linear system. These equations suggest that it is correctly representing the *inverse* of the (effective) Jacobian that is important for a forecast model to capture the response of a system to an external forcing. While the eigenvectors of $\bar{\mathbf{J}}_{\text{eff}}^{-1}$ are the same as those of $\bar{\mathbf{J}}_{\text{eff}}$, the eigenvalues are the inverse. Therefore the larger magnitude eigenvalues of $\bar{\mathbf{J}}_{\text{eff}}$, which contribute to the most rapid growth of perturbations in initial value forecasts, will have little influence on the response of a system to a forcing,

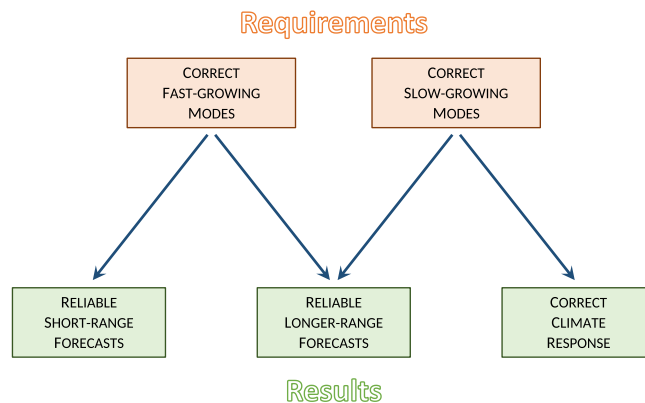


FIGURE 5 A schematic suggesting the required characteristics of a forecast model for reliable forecasts and for skilful climate prediction. While reliable longer-range forecasts could indicate that slow growing modes are correct, as needed for skilful climate prediction, unreliable longer-range forecasts could be due to either incorrect fast or slow growing modes. The precise time-scales corresponding to “shorter”- and “longer”-range forecasts will be system-dependent [Colour figure can be viewed at wileyonlinelibrary.com]

as $1/\lambda$ will be very small. Instead, the response to forcing will in general be dominated by the eigenvalues corresponding to the slowest evolving modes. Therefore, in order to judge how well a forecast model will predict the response to an external forcing, it is these slow modes that must be verified through initial value forecasts, not the fastest modes. This is a much harder task (Vannitsem and Lucarini, 2016).

We therefore propose a more nuanced link between the reliability of initial value forecasts and the forecast model’s response to an imposed forcing. While the reliability of short-range forecasts (on the scale of hours to days) could provide some information on the fidelity of the forecast model’s fastest growing modes, it is the fidelity of the slower growing modes that are important for a correct response to forcing: the integral over $\mathbf{M}(\tau)$ extends to many decades or centuries, and it is this long tail that will dominate the response. It is possible that the reliability of forecasts on longer time-scales (e.g. seasonal or longer) could provide information about the evolution of these slower modes. This would provide information relevant to the forced response of the system. However, unreliable longer-range forecasts would not necessarily indicate that the slow-growing modes have not been captured, as incorrect representation of fast growing modes would infect the forecast at longer time-scales. It is therefore important to also validate forecasts at shorter time-scales to assess sources of unreliability at longer time-scales. Figure 5 summarizes the necessary characteristics of the forecast model for a particular result.

A specific exception to this argument comes if the forcing projects directly on a growing mode of the system (a singular vector of $\mathbf{M}(\tau)$ for the τ of interest). In that case, the response of the forecast model to forcing depends solely on correctly capturing the corresponding mode. Since singular vectors evolve in time, it is possible this situation will not arise in the case of the Earth system.

5 | TESTING THE THEORY IN THE LORENZ '96 SYSTEM

To test the proposed link between reliability of shorter and longer time-scale forecasts and the response of a system to forcing, a second dynamical system is considered for which we can also develop a forecasting model. The system of choice is the second system proposed in Lorenz (1996), which we call the L96 system. This system describes a set of coupled equations for two types of variables arranged in a ring:

$$\begin{aligned} \frac{dX_k}{dt} &= -X_{k-1}(X_{k-2} - X_{k+1}) - X_k \\ &\quad + F + \delta F_k - \frac{hc}{b} \sum_{j=J(k-1)+1}^{kJ} Y_j; \quad k = 1, \dots, K, \\ \frac{dY_j}{dt} &= -cbY_{j+1}(Y_{j+2} - Y_{j-1}) - cY_j \\ &\quad + \frac{hc}{b} X_{\text{int}[(j-1)/J]+1}; \quad j = 1, \dots, JK, \end{aligned} \quad (27)$$

where the variables have cyclic boundary conditions; $X_{k+K} = X_k$ and $Y_{j+JK} = Y_j$. The X_k variables are large-amplitude, low-frequency variables, each of which is coupled to many small-amplitude, high-frequency Y_j variables. The parameter values used in this study are shown in Table 1; for these values, the system undergoes chaotic behaviour. The variables are scaled such that one model time unit is approximately equal to five atmospheric days (Lorenz, 1996). Lucarini and Sarno (2011) have demonstrated that the simplified one-scale L96 system is a suitable testbed for considering linear response theories, providing some justification for the use of the more sophisticated, albeit still simple, two-scale version to validate this theory.

The L96 equations are integrated for 10,000 MTU with a time step of 0.001 MTU and $\delta \mathbf{F}_0$ using a fourth-order Runge–Kutta scheme. Note that 1 MTU is equivalent to approximately 5 “atmospheric days,” deduced by comparing error-doubling times in the real atmosphere and the L96 system. This control simulation is used to derive the forecast models described below.

Two types of forecast model are developed for the L96 system: a linear inverse model, and a dynamical (nonlinear) stochastically parametrized forecast model. Both forecast

TABLE 1 Parameter settings used for the L96 system

Parameter	Symbol	Setting
Number of X	K	8
Number of Y per X	J	32
Coupling constant	h	1
Forcing term	F	20
Change in forcing	$\delta \mathbf{F}_0$	(0,0,0,0,0,0,0,0)
	$\delta \mathbf{F}_s$	(3,3,3,3,3,3,3,3)
	$\delta \mathbf{F}_a$	(3,0,0,0,0,0,0,0)
Spatial scale ratio	b	10
Time-scale ratio	c	10

models are constructed assuming only the large-scale X variables are predicted, while the Y variables are unresolved.

5.1 | The statistical linear inverse model

A linear inverse model (LIM) is a statistical model that is constructed to capture the covariance properties of a system. It is empirically derived from an observed time series of a dynamical system, and leads to an accurate linear representation of the system, which cannot necessarily be achieved by linearizing the governing equations of motion (Branstator and Haupt, 1998). This linear approximation to the full nonlinear equations of motion can give significant insight into the system’s dynamics. For example, LIMs are used to great effect in a climate system context to model the sea surface temperatures in the Pacific (Penland and Sardeshmukh, 1995; Newman, 2007) and Atlantic Oceans (Hawkins and Sutton, 2009; Zanna, 2012). To construct a LIM, following Penland and Sardeshmukh (1995), we rewrite Equation 1 as:

$$\dot{\mathbf{x}} = \mathcal{L}\mathbf{x} + \mathbf{n}(\mathbf{x}), \quad (28)$$

where $\mathbf{F}(\mathbf{x})$ has been decomposed into a linear component, $\mathcal{L}\mathbf{x}$ and a nonlinear component $\mathbf{n}(\mathbf{x})$, and \mathbf{x} is the state vector of the system, $\mathbf{x} = [X; Y]$. The aim is to construct a forecast model of L96 given observations of the large-scale X variables. The simplest trial model is a linear system driven by Gaussian white noise. We therefore approximate Equation 28 as

$$\dot{X} = \mathbf{L}X + \xi, \quad (29)$$

where now only the “resolved” X variables are considered, and the linear feedback matrix, \mathbf{L} is the appropriate subset of \mathcal{L} . The nonlinear terms have been replaced with white noise forcing, ξ . The most likely solution to Equation 29 is given by the propagator, \mathbf{G} :

$$\hat{X}(t + \tau) = \mathbf{G}(\tau)X(t) = \exp(\mathbf{L}\tau)X(t). \quad (30)$$

The linear feedback matrix, \mathbf{L} , can be estimated from the time series of X as

$$\mathbf{L} = \frac{1}{\tau} \ln(\mathbf{C}_\tau \mathbf{C}_0^{-1}), \quad (31)$$

where \mathbf{C}_τ and \mathbf{C}_0 are the covariance matrices at lag τ and 0 respectively. The properties of the white noise forcing can be calculated through the fluctuation–dissipation relation (Penland and Sardeshmukh, 1995):

$$\begin{aligned} \mathbf{L}\mathbf{C}_0 + \mathbf{C}_0\mathbf{L}^T + \mathbf{Q} &= \mathbf{0}, \\ \mathbf{Q} &= \langle \xi \xi^T \rangle. \end{aligned} \quad (32)$$

Penland and Sardeshmukh (1995) present a series of tests to validate whether a LIM is a good model of the chosen system. In particular, if a LIM is a good model of the system, the matrix \mathbf{L} should be independent of the lag used to estimate it. This is tested for the L96 model in Figure 6, which shows the 2-norm of \mathbf{L} as a function of the lag, τ , used to estimate it (Penland and Sardeshmukh, 1995). Below a critical value of

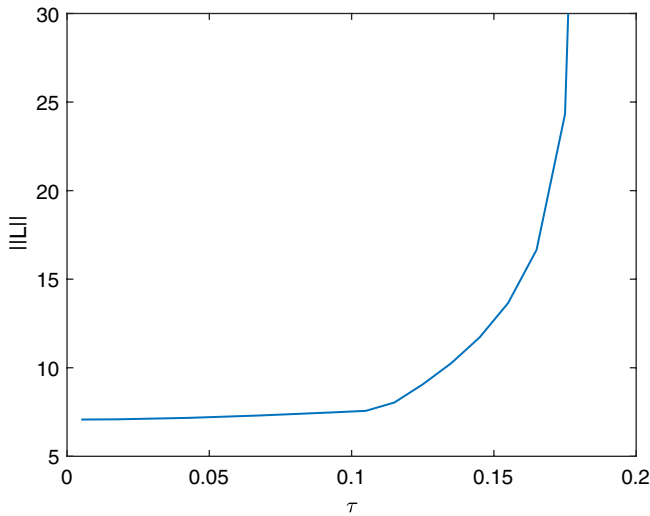


FIGURE 6 Dependence of \mathbf{L} in the LIM representing the L96 system on the lag, τ , used to estimate it. The 2-norm of the linear matrix, \mathbf{L} is shown for each τ , where \mathbf{L} is evaluated every $d\tau = 0.005$ MTU [Colour figure can be viewed at wileyonlinelibrary.com]

$\tau_{\text{crit}} = 0.105$ MTU, the norm of \mathbf{L} does not vary significantly, providing support for the use of a LIM. A further requirement is that the fast processes need to be sufficiently mixed (Penland, 2003), such that the rapidly interacting nonlinear terms can be represented as noise. If a LIM does not provide good forecasts, it is an indication that this assumption might not be valid for a particular system.

Because of the linearity of the LIM, the response equation can be written:

$$\delta(\mathbf{x}) = -\mathbf{L}^{-1} \mathbf{f}, \quad (33)$$

where the right-hand side has been simplified by recognising that the forcing perturbation is independent of the state, X . Note that this is also consistent with the result from classical FDT, since within the LIM framework, we can rewrite Equation 12 as

$$\begin{aligned} \Theta &= \int_0^{\infty} \exp(\mathbf{L}\tau) d\tau \\ &= \frac{1}{\mathbf{L}} [\exp(\mathbf{L}\tau)]_0^{\infty} \\ &= -\mathbf{L}^{-1}, \end{aligned} \quad (34)$$

since the real part of the eigenvalues of \mathbf{L} are negative. Equation 33 can be used to compute the response of the LIM to forcing.

The control simulation of the unperturbed L96 system with $\delta\mathbf{F}_0$ was coarsened by selecting every fifth data point (every 0.005 MTU) to represent the discreteness of observations of the system. Following the results of the τ sensitivity testing shown in Figure 6, we use $\tau = 0.1$ MTU to fit \mathbf{L} and \mathbf{Q} . Table 2 indicates the properties of the fitted \mathbf{L} , summarized by its eigenvalues, $\lambda = \sigma \pm i\omega$. There are two purely damped modes, and three oscillatory modes (each consisting of a complex conjugate pair of eigenvalues), such that the total number of modes is $K = 8$.

TABLE 2 The decay time ($-1/\sigma$) and period ($\omega/2\pi$) corresponding to the eigenvalues of \mathbf{L} for a LIM representation of the L96 system, found using $\tau = 0.1$ MTU

Mode	Decay time (MTU)	Period (MTU)
1/2	1.6	1.5
3/4	1.0	0.84
5	0.23	—
6/7	0.21	2.5
8	0.15	—

The modes are displayed in order of decreasing decay time.

5.2 | The stochastically parametrized forecast model

A second type of forecast model is constructed, more similar to the type of model used to make modern weather and climate predictions. This second model assumes knowledge of the underlying dynamical equations that describe the L96 system, which can be used to make a forecast. However, due to limited computer resources, a compromise is necessary and the effect of the small scale, Y variables on the X variables must be parametrized. This is equivalent to a weather or climate model where the equations of motion (the Navier–Stokes equations) are represented in the atmospheric simulator, but the sub-grid processes such as convection, gravity waves, and turbulence, must be parametrized due to computational constraints.

The stochastically parametrized forecast model used here is described in Arnold *et al.* (2013). In short, the impact of the Y variables on the X variables are parametrized using a cubic function of X . The parameters in this cubic function are fitted using observations taken from the “true” L96 system. In the full system, there is considerable variability in the sub-grid Y forcing which is not captured by this simple cubic equation. The uncertainty in the parametrization is represented by using a multiplicative stochastic perturbation. The reduced equations of motion for the forecast $X^*(t)$ of $X(t)$ can be written:

$$\begin{aligned} \frac{dX_k^*}{dt} &= -X_{k-1}^* (X_{k-2}^* - X_{k+1}^*) - X_k^* + F + \delta F_k - U_p(X_k^*), \\ U_p(X_k^*) &= \{1 + e_k(t)\} U_{\text{det}}(X_k^*), \\ U_{\text{det}}(X_k^*) &= b_0 + b_1 X_k^* + b_2 (X_k^*)^2 + b_3 (X_k^*)^3, \end{aligned} \quad (35)$$

where the stochastic perturbation $e_k(t)$ is independently generated for each X_k , and follows a zero-mean first-order autoregressive process:

$$e_k(t) = \phi e_k(t - \Delta t) + \sigma_e (1 - \phi^2)^{1/2} z_k(t), \quad (36)$$

where ϕ is the first order autoregressive parameter, σ_e^2 is the variance of the stochastic perturbation, and $z_k(t)$ is unit variance Gaussian distributed white noise. The cubic fit parameters, b_i , and the magnitude and autocorrelation of the stochastic perturbation are fitted using data from the L96 control integration (Arnold *et al.*, 2013). Two stochastic models are considered. Both models use the magnitude of the stochastic perturbation fitted from the true system. A white noise forecast model ($\phi = 0$) is compared to a red

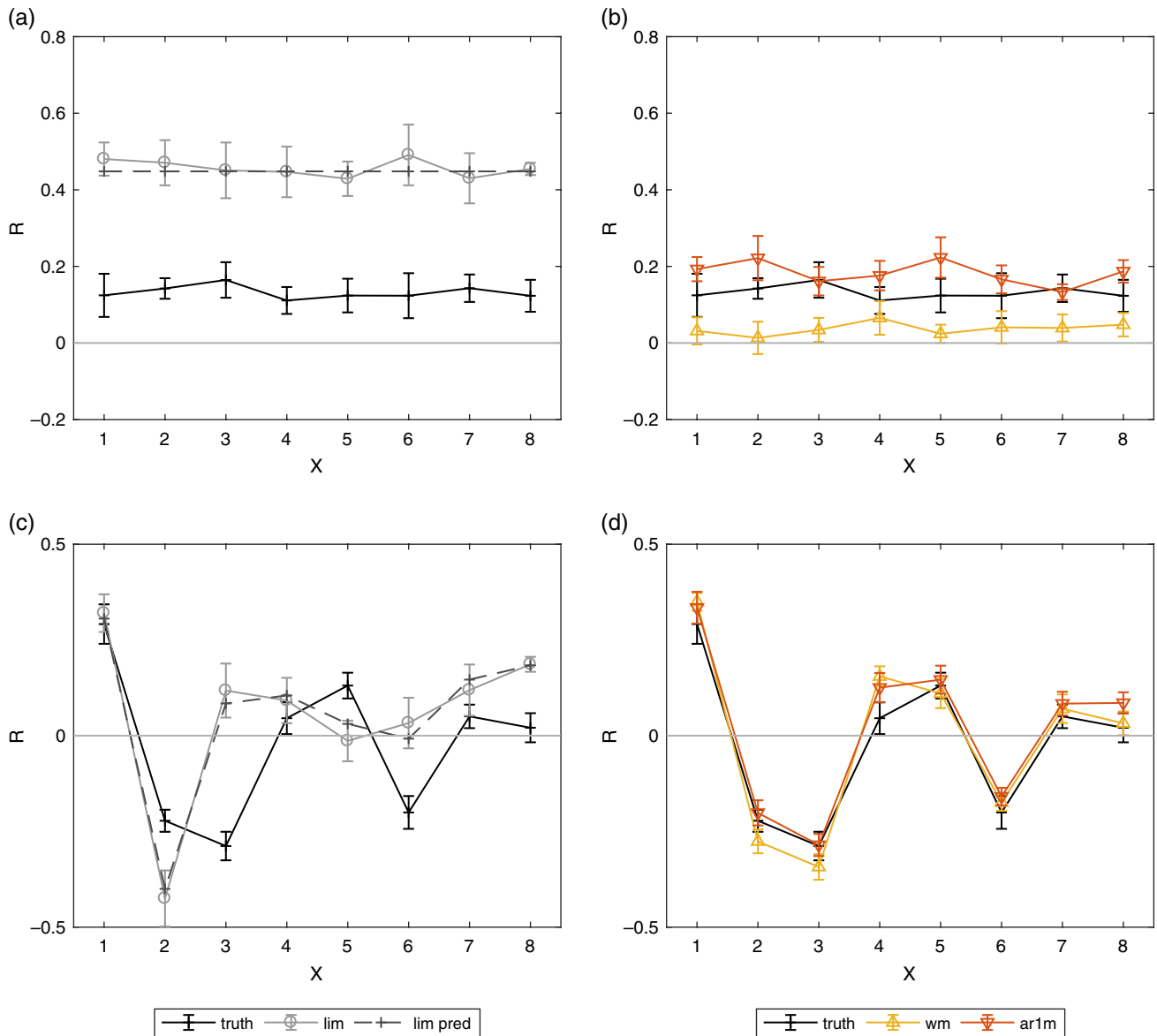


FIGURE 7 The response of the L96 system to a forcing perturbation; the observed response (black lines) to (a, b) a symmetric forcing perturbation, $\delta\mathbf{F} = [3, 3, 3, 3, 3, 3, 3, 3]$, and to (c, d) an asymmetric forcing perturbation, $\delta\mathbf{F} = [3, 0, 0, 0, 0, 0, 0, 0]$. In (a) and (c), the response is obtained from a Linear Inverse Model (LIM). The pale grey line with circles indicates the climatological response from a long integration of the LIM. The dark grey dashed line with crosses indicates the theoretical response indicated by Equation 33. As expected, there is excellent agreement between the theoretical and measured response. In (b) and (d), a nonlinear dynamical forecast model with stochastic parametrization of the fast scales is used with white multiplicative noise (yellow lines with upward triangles) or red multiplicative noise (red lines with downward triangles) [Colour figure can be viewed at wileyonlinelibrary.com]

noise forecast model, for which ϕ is fitted from the “true” tendencies.

5.3 | Climate projection

The impact of a change in forcing on the mean climate of the L96 system is considered, following Christensen *et al.* (2015b). Two state-independent perturbations are introduced, both constant in time. The first is constant across the X variables (symmetric: $\delta\mathbf{F}_s$), while the second is applied only to X_1 (asymmetric: $\delta\mathbf{F}_a$), as shown in Table 1. In addition to the control simulation with $\delta\mathbf{F}_0$, two further 10,000 MTU integrations are produced for each forcing perturbation in turn. The response of the mean of the L96 system to the change

in forcing is shown in Figure 7 (black line), while the error bars indicate the variability in response due to the finite sample size². The symmetric forcing produces a symmetric response, while the asymmetric forcing breaks the symmetry of the L96 equations, resulting in an asymmetric response as shown.

Each forecast model is integrated for 10,000 MTU with $\delta\mathbf{F}_0$ to produce estimates of the unperturbed ‘climate’ of the L96 system. To use the LIM as a forward model, Equation 29 is integrated using the first-order Stratonovich

²The error bars indicate the variation in response of the mean when 2,000 MTU segments of the time series are considered instead of the full time series.

scheme described in Penland and Matrosova (1994)³ with a time step of 0.005 MTU. For the stochastic dynamical models, a second-order Runge–Kutta time-stepping scheme is used with the same time step of 0.005 MTU. Two further integrations are produced with each of the forecast models for the two forcing perturbations respectively, $\delta\mathbf{F}_s$ and $\delta\mathbf{F}_a$. The forecast models are not changed or re-tuned prior to making this “climate change” prediction: analogously to practice in the climate modelling community, the models are tuned and verified using the present-day observed climate, before that same model is used to produce climate change projections. The differences in mean between the unperturbed and perturbed climates for each L96 forecast model are shown in Figure 7.

Figure 7 shows the response to forcing predicted by the LIM compared to the true response (i.e. the response of the mean of the full L96 system) for both (a) symmetric and (c) asymmetric forcing perturbations. Figure 7 (a) shows that, while the LIM can predict that the response of the true L96 system to a symmetric forcing will be symmetric, the predicted magnitude is too large. The LIM also fails to predict the response to the asymmetric forcing $\delta\mathbf{F}_a$.

Figure 7 also compares the observed response of the white and red stochastically parametrized forecast models to that of the true system for both (b) symmetric and (d) asymmetric forcing perturbations. For both the symmetric and asymmetric cases, both stochastic dynamical models do a reasonable job at predicting the shape of the response of the mean. For the symmetric forcing, the red noise model (red line with downward triangles) predicts the correct response to within sampling uncertainty, while the white noise model (yellow line with upward triangles) predicts too weak a response compared to that observed (black line). For the asymmetric forcing perturbation shown in (d), both parametrized models significantly outperform the LIM. The error bars indicate no significant difference between the response of the red noise forecast model and the true system, such that the red noise model slightly outperforms the white noise case.

5.4 | Reliability

Initialized forecasts in the L96 system are now considered. Given access to such “weather” forecasts, would it have been possible to predict whether the forecast models would correctly capture the response of the “true” L96 system to forcing?

We first use the LIM operator \mathbf{L} to provide an estimate of the key time-scales and modes in the L96 system. From Table 2, the fastest time-scale modes in the linearized model have a time-scale of order a few “atmospheric days,” while the slow modes have time-scales on the order of a few “atmospheric weeks.” The symmetric forcing perturbation, $\delta\mathbf{F}_s$ is

aligned with one of the eigenvectors of \mathbf{L} (not shown).⁴ Since the symmetric forcing projects onto one of the modes of the system, we anticipate that (to leading order) we need capture only that single mode to predict the response to $\delta\mathbf{F}_s$. In contrast, the asymmetric forcing perturbation, $\delta\mathbf{F}_a$ is not aligned with any eigenvectors of \mathbf{L} . The response of the forecast model to this perturbation will therefore be dominated by the slowest evolving modes, as argued above.

The three forecast models were used to produce short-range ensemble weather forecasts. Initial conditions were selected every 10 MTU from the full “true” L96 simulation to give a sample of 300 start dates. A 40-member ensemble forecast out to 3 MTU was produced from each start date for each model. The ensemble members for a given start date differ only in their stochastic forcing: it is assumed that the initial state is known perfectly, and so the initial conditions are not perturbed. Figure 8 shows the reliability of the three forecast models as a function of lead time. Figure 8a shows the standard deviation (“spread”) of the ensemble forecasts compared to the root mean square error (RMSE) in the ensemble mean as a function of lead time for each forecast model. For a reliable forecast, the ensemble spread should match the RMSE in the ensemble mean (Wilks, 2006; Leutbecher and Palmer, 2008). From this summary diagnostic, both the LIM and the red noise parametrized forecast model are reliable at a range of lead times.

It is clear that variability in the L96 system has a narrow range of time-scales when compared to the Earth system, where processes in the stratosphere, ocean, ice sheets, and land surface introduce a wide spectrum of time-scales into tropospheric motions. Nevertheless, we will consider the reliability of comparatively “fast” and “slow” time-scales within the L96 system, to evaluate the predictions from Section 4. The “fast” (“slow”) time-scales in the L96 system will be dominated by the dynamics of the Y (X) variables respectively. We select two lead times of particular interest. The eigenvalue of \mathbf{L} that corresponds to $\delta\mathbf{F}_s$, which we denote λ_s , has a real part with magnitude 6.69, corresponding to an e-folding time-scale of 0.15 MTU.⁵ We observe that error saturation in the parametrized stochastic models occurs as late as 3 MTU (Figure 8a), corresponding to the time-scale of the slowest modes in the system. To consider how well the L96 forecast models will respond to the forcing perturbations $\delta\mathbf{F}_s$ and $\delta\mathbf{F}_a$, we therefore consider the reliability of forecasts at 0.15 MTU and 2.8 MTU (i.e. just before saturation) respectively. Considering the reliability at these time-scales will probe the relevant growing modes in the forecast model, compared to the full L96 system.

³Note that the Ito and Stratonovich schemes are equivalent for this LIM as the noise term is independent of the state of the system.

⁴Due to the symmetries of the L96 system, the symmetric forcing perturbation is also the leading singular vector of the tangent propagator, $\exp(\mathbf{L}\tau)$, with an associated Lyapunov vector that is parallel to the singular vector.

⁵The singular value of the tangent propagator associated with $\delta\mathbf{F}_s$ is $1/e$ when $\tau = 0.15$ MTU, as expected.

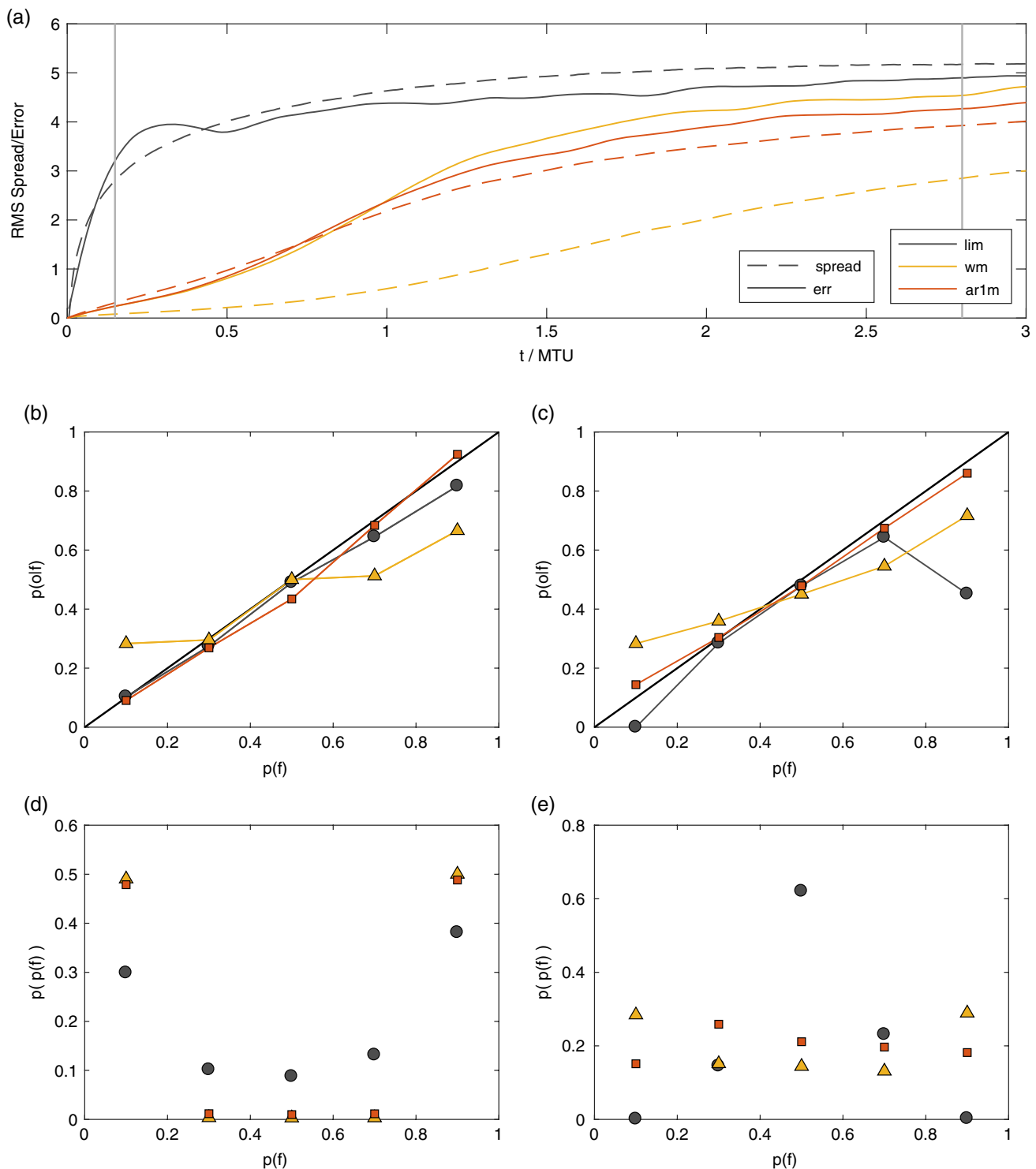


FIGURE 8 (a) Root mean square (RMS) spread (dashed lines) and error (solid lines) for the LIM and nonlinear dynamical stochastic forecast models, the reliability diagram (b, c) calibration function, showing predicted probability against observed frequency of an event, and (d, e) refinement distribution, showing the frequency of occurrence of different forecast probabilities at (b, d) 0.15 model time units (MTU) and (c, e) 2.8 MTU. Lead times of 0.15 and 2.8 MTU are also indicated in (a) by the vertical lines. For all panels, grey (circles) denote the LIM, yellow (triangles) the white multiplicative dynamical model, and red (squares) the red multiplicative dynamical model [Colour figure can be viewed at wileyonlinelibrary.com]

The vertical lines on Figure 8a indicate the lead times of particular interest: 0.15 and 2.8 MTU. At each of these lead times, a reliability diagram is generated for forecasts of the event “ $X > \bar{X}$,” where \bar{X} is the climatological mean of the full L96 system, and statistics are aggregated from across all X_k : the calibration functions are shown in (b) and (c)

while the refinement distributions are shown in (d) and (e) (Wilks, 2006). For a reliable forecast, the calibration function, which shows the observed frequency of an event against predicted probability, should lie on the one-to-one line. The refinement distribution provides additional information about the confidence of the forecast model, indicating how often

the forecast diverges from climatology (Wilks, 2006). At a lead time of 0.15 MTU ((b) and (d)), the LIM and red noise parametrized models both produce reliable forecasts: the conditional observed probability closely matches the forecast probability. Both forecast models show high confidence, with forecast probabilities that often deviate from the mean. In contrast, at a lead time of 2.8 MTU ((c) and (e)), while both the LIM and red noise parametrized models still show high reliability, the LIM shows low confidence, rarely producing forecasts that deviate from the climatological forecast: the LIM has saturated to climatology. The white noise parametrized model is unreliable at both short and long lead times, producing systematically under-dispersive forecasts.

These results show that the LIM produces reliable forecasts at a range of lead times. However, the LIM response to forcing, as shown in Figure 7, was poor. This implies that reliability is not an indicator of a good climate response. To interpret these results, we return to Equations 7 and 8. In Section 2.1, we saw that evaluating reliability could only indicate whether the forecast tangent propagator, $\mathbf{M}^f(t, 0, \mathbf{x}_0)$ adequately represents the ‘true’ tangent propagator if the model error term $\mathbf{b}(\mathbf{x}, t)$ is small. However, Figure 8 shows that error growth rates in the LIM are too high, since the two dynamical models show predictability out to substantially longer lead times than the LIM. In fact, the dominant modes in the LIM have time-scales of 1.0 and 1.5 MTU, after which the LIM ensemble forecast approaches the climatological spread (the slower 2.5 MTU mode rapidly decays). The model error term in the LIM must be significant. The rapid growth of perturbations in the LIM compensates for the large model error, to produce reliable forecasts. Therefore, reliability is not always a good indicator of climatological response. Only if the model error is relatively small, or equivalently, if the error growth rates in the forecast model are similar to those in the true system, can reliability provide useful information about that forecast model’s climatic response to forcing, when evaluated at a suitable lead time.

At 0.15 MTU, the red noise dynamical model is more reliable than the white noise dynamical model, indicating a better representation of λ_s (the eigenvalue of \mathbf{L} that corresponds to $\delta\mathbf{F}_s$). The red noise model is also more skilful at predicting the climatic response of the corresponding symmetric forcing perturbation, $\delta\mathbf{F}_s$. The asymmetric forcing perturbation did not project onto any one eigenmode of the system, so we expect the slowest growing modes to be of most importance for predicting the correct climate response. At longer lead times, the red noise forecast maintained its reliability, indicating a good representation of the slowest growing modes of the system. This forecast model also predicted a good response to the asymmetric forcing. We find that a reliable forecast model has also responded well to an imposed forcing, as predicted by the theory.

At longer lead times, the white noise parametrized model produced unreliable forecasts. However, it is not possible to discern the cause of this unreliability. Forecasts from the

white noise model were unreliable at short lead times, indicating a poor representation of the fastest growing modes in the L96 system. Errors in the fast-growing modes would infect the forecast at longer time-scales, making it difficult to assess the representation of the slower growing modes in that forecast system. From Figure 7, the white noise parametrized model was in fact able to capture the response of the L96 system to the asymmetric forcing, indicating that the model captures the slowest growing modes in the L96 system. This is as suggested by the theory, and indicated in Figure 5.

Physically, in the L96 system, the long time-scale predictability comes from the correct representation of the low-frequency X variables. These are correctly represented in both the red and white noise dynamical models. In contrast, the LIM provides a statistical representation of both the large- and small-scale dynamics: it is a linearized representation of the system, with the effect of the nonlinear dynamics incorporated into the noise. While the LIM captures the impact of the nonlinear processes in a statistical sense, it seems that the details of the nonlinear processes are important for setting the slowest evolving modes in the L96 system. The fastest evolving modes come from the unresolved sub-grid Y variables. The increased forecast reliability indicates that the correlated, red noise dynamical model captures the statistical effect of these small-scale variables on the slow scales better than the white noise model.

5.5 | Implications for the development of stochastic forecasting models

Recent years have seen an increase in the adoption of stochastic parametrization schemes by weather forecasting centres (Palmer *et al.*, 2009; Sanchez *et al.*, 2016). The stochastic approach is designed to represent uncertainty in forecasts due to the limitations of the forecast model. Without stochastic parametrizations, ensemble forecasts are unreliable. The use of stochastic schemes improves the growth of perturbations in the model, substantially improving the reliability of weather forecasts (Berner *et al.*, 2008; Palmer *et al.*, 2009; Christensen *et al.*, 2015c; Leutbecher *et al.*, 2017). The improved reliability of weather forecasts indicates an improved representation of fast-evolving modes in models. However, improving the statistics of fast processes through stochastic parametrization also systematically impacts the evolution of slowly evolving modes in climate models. For example, stochastic parametrization schemes have been shown to transform the representation of the slowly evolving El Niño–Southern Oscillation (ENSO) in climate models (Christensen *et al.*, 2017). In a follow-on study from Christensen *et al.* (2017), Berner *et al.* (2018) use LIM as a tool to study tropical Pacific variability in the Community Climate System Model (CCSM) version 4, and find that a simulation including a stochastic parametrization scheme had a better representation of the time-scales associated with the slowest evolving modes than a deterministic (control) simulation. Given this improvement in

low-frequency variability, it is possible that including stochastic schemes in a climate model will improve the response of that model to an applied external forcing. In other words, stochastic parametrizations could lead to improved estimates of climate sensitivity.

6 | DISCUSSION AND CONCLUSIONS

In this paper, linear response theory was reviewed and presented using the notation of dynamical systems theory. Assuming a forcing constant in time and space, and considering only the mean response of a dynamical system to an imposed forcing, we show that

$$\delta\langle\mathbf{x}\rangle = \int_0^\infty d\tau \overline{\mathbf{M}(\tau)}\mathbf{f}, \quad (37)$$

where $\delta\langle\mathbf{x}\rangle$ is the mean response of the system averaged over the attractor, $\overline{\mathbf{M}(\tau)}$ is the tangent propagator, averaged over the attractor, which maps a small forcing perturbation at a given time onto the response at time τ later, and \mathbf{f} is the applied forcing. For a climate model to correctly represent the climate sensitivity of the Earth system, it must have a good representation of $\overline{\mathbf{M}(\tau)}$.

In an ensemble forecast, the tangent propagator, \mathbf{M} , maps small initial condition perturbations forward in time, embodying how the ensemble will evolve. *Reliability* is the property of an ensemble forecast which evaluates whether the verification behaves as if drawn from the predicted distribution, that is, whether the ensemble members are diverging at the correct rate. It is therefore intuitive to link reliability with response to forcing, as proposed by Palmer *et al.* (2008).

However, we note that evaluating the reliability of a short-range weather forecast can only evaluate $\mathbf{M}(\tau)$ for $\tau \leq$ the length of the forecast. In contrast, the response equation requires correct representation of $\mathbf{M}(\tau)$ out to $\tau = \infty$. Furthermore, we show evidence that the dominant response of a system to an applied forcing is governed by the system's *slowest* evolving modes (i.e. $\mathbf{M}(\tau)$ for large τ).

Therefore, instead of looking to short-range weather forecasts, we propose that the reliability of forecasts at the time-scale of the slowest modes in the system will be of particular interest. In the atmosphere, slowly evolving modes are associated with large scales, while rapidly evolving modes are associated with smaller scales. It is possible that considering seasonal and longer time-scale ensemble forecasts, averaged over large spatial and temporal scales, could be used to constrain projections of climate change. A reliable seasonal forecast indicates that slow modes in the atmosphere–ocean–land system have been well-captured by the forecast model on the annual time-scale; in order to evaluate lower-frequency modes present in the ocean, even longer time-scales would need to be considered.

This theory therefore provides some justification for using the reliability of initialized forecasts as an emergent constraint

on climate sensitivity, which until now had been postulated but not proved (Palmer *et al.*, 2008). However, we highlight some important caveats, which we demonstrate by producing forecasts for the simple Lorenz (1996) system. We note that an unreliable seasonal (or longer time-scale) forecast is not sufficient to deduce that the slow modes in the system are poorly captured by the forecast model. This is because a reliable seasonal forecast also requires the rapidly evolving modes to be captured. Consideration of weather forecasts on scales of hours to days could provide some insight into the representation of these fast modes. Furthermore, we demonstrate that it is possible for a forecast system to produce reliable forecasts and yet be unable to capture the response of the system to an external forcing. This is the case if the forecast model has a large model error, and yet correctly accounts for forecast uncertainty introduced by this error.

There are several limitations of the approach presented above which are worth further discussion. Firstly, the approach has considered the linear climate response to a small perturbation. It is not clear how this assumption affects the applicability of this theory to the climate system, where the applied external forcing (anthropogenic greenhouse gas emissions) far exceeds natural variation as indicated by the palaeoclimate records. However, Lucarini *et al.* (2017) show that linear response theory is a useful framework for the climate system, despite the large magnitude of the applied forcing. Secondly, this discussion has restricted itself to the response of the mean of the climate system. It is also of great interest how the variability of the climate system will respond to an external forcing, with implications for changes in climate extremes such as droughts and floods or extremes of temperature. No evidence has been presented to indicate that the reliability of short-range forecasts is a good indicator of the response of higher-order modes of the climate system to greenhouse gas forcing. Our discussion has assumed stationarity of the system of interest, when in reality the non-stationarity of the climate system should be accounted for (e.g. Lucarini and Sarno, 2011). It is also evident that, if the short-range forecast model only includes an atmosphere and ocean, then assessing the reliability of forecasts will only evaluate the response of those aspects of the model to forcing. Longer time-scales present in the cryo- or biosphere will not be evaluated, but are likely very important for predicting the correct climate sensitivity of the Earth-system (Hurrell *et al.*, 2009). Finally, we note that, while model bias appears explicitly in evaluating reliability, it also appears implicitly in the response equation, as the integral is performed over the biased model attractor, which will naturally affect the predicted response.

The results presented here suggest a focus in climate model evaluation should be on the fidelity of the slowest modes in the model. However, changes to parametrization of a particular fast *process* made during model development will almost certainly impact the model across a range of time-scales (e.g. Neale *et al.*, 2008; Christensen *et al.*, 2017) due to the interaction of different scales of motion in

the climate system. In Section 5.5 we discuss evidence that stochastic parametrization of fast, small-scale processes can improve the representation of both fast and slow modes in forecast models. It is precisely this multi-scale interaction which motivates the seamless evaluation of weather and climate models. While other studies have used very short time-scale forecasts to evaluate parametrization schemes and thereby constrain climate sensitivity (Rodwell and Palmer, 2007), we propose that this apparent relationship is primarily because changing a parametrization scheme can adversely impact both the faster and the slower evolving modes in the forecast model (Berner *et al.*, 2018).

In conclusion, linear response theory provides insight into the link between the reliability of weather and seasonal forecasts, and the climatic response of a forecast model. The purpose of this manuscript is to discuss the implications that this theory has for predicting *a priori* whether a given climate model will be able to predict climate sensitivity, or in other words, the correct response of the climate system to an applied external forcing. We conclude that, by framing the problem in terms of a general dynamical system, there is some justification for the use of initialized forecasts to validate climate models.

ACKNOWLEDGEMENTS

Many thanks to Joe Tribbia and Fenwick Cooper for numerous helpful discussions. Thanks to Valerio Lucarini and an anonymous reviewer for their constructive comments, which substantially improved this manuscript.

ORCID

Hannah M. Christensen  <https://orcid.org/0000-0001-8244-0218>

REFERENCES

- Abramov, R.V. and Majda, A.J. (2007) Blended response algorithms for linear fluctuation–dissipation for complex nonlinear dynamical systems. *Nonlinearity*, 20, 2793–2821.
- Abramov, R.V. and Majda, A.J. (2008) New approximations and tests of linear fluctuation response for chaotic nonlinear forced-dissipative dynamical systems. *Journal of Nonlinear Science*, 18(3), 303–341. <https://doi.org/10.1007/s00332-007-9011-9>.
- Achatz, U., Löbl, U., Dolaptchiev, S.I. and Gritsun, A. (2013) Fluctuation-dissipation supplemented by nonlinearity: a climate-dependent subgrid-scale parameterization in low-order climate models. *Journal of the Atmospheric Sciences*, 70(6), 1833–1846. <https://doi.org/10.1175/jas-d-12-0229.1>.
- Arnold, H.M., Moroz, I.M. and Palmer, T.N. (2013) Stochastic parametrizations and model uncertainty in the Lorenz '96 system. *Philosophical Transactions of the Royal Society of London, Series A*, 371, 2013. <https://doi.org/10.1098/rsta.2011.0479>.
- Berner, J., Doblas-Reyes, F.J., Palmer, T.N., Shutts, G.J. and Weisheimer, A. (2008) Impact of a quasi-stochastic cellular automaton backscatter scheme on the systematic error and seasonal prediction skill of a global climate model. *Philosophical Transactions of the Royal Society of London, Series A*, 366, 2561–2579. <https://doi.org/10.1098/rsta.2008.0033>.
- Berner, J., Sardeshmukh, P.D. and Christensen, H.M. (2018) On the dynamical mechanisms governing El Niño–Southern Oscillation irregularity. *Journal of Climate*, 31, 8401–8419.
- Branstator, G. and Haupt, S.E. (1998) An empirical model of barotropic atmospheric dynamics and its response to tropical forcing. *Journal of Climate*, 11(10), 2645–2667.
- Brunet, G., Shapiro, M., Hoskins, B.J., Moncrieff, M.J., Dole, R., Kiladis, G.N., Kirtman, B., Lorenc, A.C., Mills, B., Morss, R., Polavarapu, S., Rogers, D., Schaake, J. and Shukla, J. (2010) Collaboration of the weather and climate communities to advance subseasonal-to-seasonal prediction. *Bulletin of the American Meteorological Society*, 91(10), 1397–1406. <https://doi.org/10.1175/2010BAMS3013.1>.
- Christensen, H.M., Moroz, I.M. and Palmer, T.N. (2015a) Evaluation of ensemble forecast uncertainty using a new proper score: application to medium-range and seasonal forecasts. *Quarterly Journal of the Royal Meteorological Society*, 141, 538–549.
- Christensen, H.M., Moroz, I.M. and Palmer, T.N. (2015b) Simulating weather regimes: impact of stochastic and perturbed parameter schemes in a simple atmospheric model. *Climate Dynamics*, 44(7), 2195–2214.
- Christensen, H.M., Moroz, I.M. and Palmer, T.N. (2015c) Stochastic and perturbed parameter representations of model uncertainty in convection parameterization. *Journal of the Atmospheric Sciences*, 72(6), 2525–2544.
- Christensen, H.M., Berner, J., Coleman, D. and Palmer, T.N. (2017) Stochastic parameterisation and the El Niño–Southern Oscillation. *Journal of Climate*, 30(1), 17–38.
- Cionni, I., Visconti, G. and Sassi, F. (2004) Fluctuation dissipation theorem in a general circulation model. *Geophysical Research Letters*, 31(9), L09206. <https://doi.org/10.1029/2004GL019739>.
- Cooper, F.C. and Haynes, P.H. (2011) Climate sensitivity via a nonparametric fluctuation–dissipation theorem. *Journal of the Atmospheric Sciences*, 68(5), 937–953. <https://doi.org/10.1175/2010JAS3633.1>.
- Cooper, F.C., Esler, J.G. and Haynes, P.H. (2013) Estimation of the local response to a forcing in a high dimensional system using the fluctuation–dissipation theorem. *Nonlinear Processes in Geophysics*, 20(2), 239–248. <https://doi.org/10.5194/npg-20-239-2013>.
- Eyink, G.L., Haine, T.W.N. and Lea, D.J. (2004) Ruelle's linear response formula, ensemble adjoint schemes and Lévy flights. *Nonlinearity*, 17(5), 1867–1889. <https://doi.org/10.1088/0951-7715/17/5/016>.
- Franzke, C.L.E., Osprey, S.M., Davini, P. and Watkins, N.W. (2015) A dynamical systems explanation of the Hurst effect and atmospheric low-frequency variability. *Scientific Reports*, 5(9068). <https://doi.org/10.1038/srep09068>.
- Gritsun, A. and Branstator, G. (2007) Climate response using a three-dimensional operator based on the fluctuation–dissipation theorem. *Journal of the Atmospheric Sciences*, 64(7), 2558–2575. <https://doi.org/10.1175/JAS3943.1>.
- Gritsun, A. and Lucarini, V. (2017) Fluctuations, response, and resonances in a simple atmospheric model. *Physica D*, 349, 62–76. <https://doi.org/10.1016/j.physd.2017.02.015>.
- Haarsma, R.J., Roberts, M.J., Vidale, P.L., Senior, C.A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N.S., Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L.R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M.S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J. and von Storch, J.-S. (2016) High resolution model intercomparison project (highresmip v1.0) for CMIP6. *Geoscientific Model Development*, 9(11), 4185–4208. <https://doi.org/10.5194/gmd-9-4185-2016>.
- Hannay, C., Williamson, D.L., Hack, J.J., Kiehl, J.T., Olson, J.G., Klein, S.A., Bretherton, C.S. and Köhler, M. (2009) Evaluation of forecasted southeast Pacific stratocumulus in the NCAR, GFDL, and ECMWF models. *Journal of Climate*, 22(11), 2871–2889. <https://doi.org/10.1175/2008JCL12479.1>.
- Hawkins, E. and Sutton, R. (2009) Decadal predictability of the Atlantic Ocean in a coupled GCM: Forecast skill and optimal perturbations using linear inverse modeling. *Journal of Climate*, 22(14), 3960–3978. <https://doi.org/10.1175/2009JCL12720.1>.
- Hoskins, B.J. (2013) The potential for skill across the range of the seamless weather–climate prediction problem: a stimulus for our science. *Quarterly Journal of the Royal Meteorological Society*, 139, 573–584. <https://doi.org/10.1002/qj.1991>.
- Hurrell, J., Meehl, G.A., Bader, D., Delworth, T.L., Kirtman, B. and Wielicki, B. (2009) A unified modeling approach to climate system prediction. *Bulletin of the American Meteorological Society*, 90(12), 1819–1832. <https://doi.org/10.1175/2009BAMS2752.1>.

- James, I.N. (2002) Models of the predictability of a simple nonlinear dynamical system. *Atmospheric Science Letters*, 3(2–4), 42–51. <https://doi.org/10.1006/asle.2002.0099>.
- Klein, S.A. and Hall, A. (2015) Emergent constraints for cloud feedbacks. *Current Climate Change Reports*, 1(4), 276–287.
- Kubo, R. (1966) The fluctuation–dissipation theorem. *Reports on Progress in Physics*, 29(1), 255–284.
- Leith, C.E. (1975) Climate response and fluctuation dissipation. *Journal of the Atmospheric Sciences*, 32, 2022–2026.
- Leutbecher, M. and Palmer, T.N. (2008) Ensemble forecasting. *Journal of Computational Physics*, 227(7), 3515–3539.
- Leutbecher, M., Lock, S.-J., Ollinaho, P., Lang, S.T.K., Balsamo, G., Bechtold, P., Bonavita, M., Christensen, H.M., Diamantakis, M., Dutra, E., English, S., Fisher, M., Forbes, R.M., Goddard, J., Haiden, T., Hogan, R.J., Juricke, S., Lawrence, H., MacLeod, D., Magnusson, L., Malardel, S., Massart, S., Sandu, I., Smolarkiewicz, P.K., Subramanian, A., Vitart, F., Wedi, N. and Weisheimer, A. (2017) Stochastic representations of model uncertainties at ECMWF: state of the art and future vision. *Quarterly Journal of the Royal Meteorological Society*, 143, 2315–2339. <https://doi.org/10.1002/qj.3094>.
- Lorenz, E.N. (1963) Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2), 130–141.
- Lorenz, E.N. (1969) The predictability of a flow which possesses many scales of motion. *Tellus*, 21(3), 289–307.
- Lorenz, E.N. (1996) Predictability—a problem partly solved. In: *Proceedings, Seminar on Predictability*, 4–8 September 1995, Vol. 1, ECMWF, Reading, UK.
- Lucarini, V. (2009) Evidence of dispersion relations for the nonlinear response of the Lorenz 63 system. *Journal of Statistical Physics*, 134, 381–400.
- Lucarini, V. and Sarno, S. (2011) A statistical mechanical approach for the computation of the climatic response to general forcings. *Nonlinear Processes in Geophysics*, 18(1), 7–28.
- Lucarini, V., Blender, R., Herbert, C., Ragone, F., Pascale, S. and Wouters, J. (2014) Mathematical and physical ideas for climate science. *Reviews of Geophysics*, 52(4), 809–859. <https://doi.org/10.1002/2013RG000446>.
- Lucarini, V., Ragone, F. and Lunkeit, F. (2017) Predicting climate change using response theory: global averages and spatial patterns. *Journal of Statistical Physics*, 166(3–4), 1036–1064.
- Marzban, C. (2013) Variance-based sensitivity analysis: an illustration on the Lorenz '63 model. *Monthly Weather Review*, 141(11), 4069–4079. <https://doi.org/10.1175/MWR-D-13-00032.1>.
- Medeiros, B., Williamson, D.L., Hannay, C. and Olson, J.G. (2012) Southeast Pacific stratocumulus in the Community Atmosphere Model. *Journal of Climate*, 25(18), 6175–6192. <https://doi.org/10.1175/JCLI-D-11-00503.1>.
- Miller, R.N., Ghil, M. and Gauthiez, F. (1994) Advanced data assimilation in strongly nonlinear dynamical systems. *Journal of the Atmospheric Sciences*, 51(8), 1037–1056. [https://doi.org/10.1175/1520-0469\(1994\)051<1037:Adain>2.0.Co;2](https://doi.org/10.1175/1520-0469(1994)051<1037:Adain>2.0.Co;2).
- Molteni, F. and Palmer, T.N. (1993) Predictability and finite-time instability of the northern winter circulation. *Quarterly Journal of the Royal Meteorological Society*, 119, 269–298.
- Neale, R.B., Richter, J.H. and Jochum, M. (2008) The impact of convection on ENSO: from a delayed oscillator to a series of events. *Journal of Climate*, 21(22), 5904–5924.
- Newman, M. (2007) Interannual to decadal predictability of tropical and North Pacific sea surface temperatures. *Journal of Climate*, 20(11), 2333–2356. <https://doi.org/10.1175/JCLI4165.1>.
- Palmer, T.N. (1999) A nonlinear dynamical perspective on climate prediction. *Journal of Climate*, 12(2), 575–591.
- Palmer, T.N. (2001) A nonlinear dynamical perspective on model error: a proposal for non-local stochastic–dynamic parametrization in weather and climate prediction models. *Quarterly Journal of the Royal Meteorological Society*, 127, 279–304.
- Palmer, T.N. and Weisheimer, A. (2011) Diagnosing the causes of bias in climate models – why is it so hard? *Geophysical and Astrophysical Fluid Dynamics*, 105(2–3), 351–365.
- Palmer, T.N. and Weisheimer, A. (2018) A simple pedagogical model linking initial-value reliability with trustworthiness in the forced climate response. *Bulletin of the American Meteorological Society*, 99(3), 606–614. <https://doi.org/10.1175/BAMS-85-12-1903>.
- Palmer, T.N., Doblas-Reyes, F.J., Weisheimer, A. and Rodwell, M.J. (2008) Towards seamless prediction: calibration of climate change projections using seasonal forecasts. *Bulletin of the American Meteorological Society*, 89(4), 459–470.
- Palmer, T.N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G.J., Steinheimer, M. and Weisheimer, A. (2009) *Stochastic Parametrization and Model Uncertainty*. Technical Memorandum 598. ECMWF, Reading, UK.
- Penland, C. (2003) Noise out of chaos and why it won't go away. *Bulletin of the American Meteorological Society*, 84, 921–926.
- Penland, C. and Matrosova, L. (1994) A balance condition for stochastic numerical models with application to the El Niño–Southern Oscillation. *Journal of Climate*, 7(9), 1352–1372. [https://doi.org/10.1175/1520-0442\(1994\)007<1352:ABCFSN>2.0.CO;2](https://doi.org/10.1175/1520-0442(1994)007<1352:ABCFSN>2.0.CO;2).
- Penland, C. and Sardeshmukh, P.D. (1995) The optimal growth of tropical sea surface temperature anomalies. *Journal of Climate*, 8(8), 1999–2024.
- Phillips, T.J., Potter, G.L., Williamson, D.L., Cederwall, R.T., Boyle, J.S., Fiorino, M., Hnilo, J.J., Olson, J.G., Xie, S. and Yio, J.J. (2004) Evaluating parameterizations in general circulation models: climate simulation meets weather prediction. *Bulletin of the American Meteorological Society*, 85(12), 1903–1915. <https://doi.org/10.1175/BAMS-85-12-1903>.
- Roberts, M.J., Vidale, P.L., Senior, C.A., Hewitt, H.T., Bates, C., Berthou, S., Chang, P., Christensen, H.M., Danilov, S., Demory, M.-E., Griffies, S., Haarsma, R., Jung, T., Martin, G., Minobe, S., Ringler, T., Satoh, M., Schiemann, R., Scoccimarro, E., Stephens, G. and Wehner, M. (2018) The benefits of global high-resolution for climate simulation: process understanding and the enabling of stakeholder decisions at the regional scale. *Bulletin of the American Meteorological Society*, 99, 2341–2359.
- Rodwell, M.J. and Palmer, T.N. (2007) Using numerical weather prediction to assess climate models. *Quarterly Journal of the Royal Meteorological Society*, 133(622), 129–146.
- Rodwell, M.J., Lang, S.T.K., Ingleby, B., Bormann, N., Hólm, E.V., Rabier, F., Richardson, D.S. and Yamaguchi, M. (2016) Reliability in ensemble data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 142, 443–454.
- Ruelle, D. (1998) General linear response formula in statistical mechanics, and the fluctuation–dissipation theorem far from equilibrium. *Physics Letters A*, 245(3–4), 220–224.
- Ruelle, D. (1999) Nonequilibrium statistical mechanics near equilibrium: computing higher-order terms. *Nonlinearity*, 11, 5–18.
- Sanchez, C., Williams, K.D. and Collins, M. (2016) Improved stochastic physics schemes for global weather and climate models. *Quarterly Journal of the Royal Meteorological Society*, 142, 147–159.
- Slingo, J.M. and Palmer, T.N. (2011) Uncertainty in weather and climate prediction. *Philosophical Transactions of the Royal Society of London, Series A*, 369(1956). <https://doi.org/10.1098/rsta.2011.0161>.
- Smith, L.A., Ziehmann, C. and Fraedrich, K. (1999) Uncertainty dynamics and prediction in chaotic systems. *Quarterly Journal of the Royal Meteorological Society*, 125, 2855–2886.
- Thuburn, J. (2005) Climate sensitivities via a Fokker–Planck adjoint approach. *Quarterly Journal of the Royal Meteorological Society*, 131, 73–92.
- Trevisan, A. and Legnani, R. (1995) Transient error growth and local predictability: a study in the Lorenz system. *Tellus A*, 47(1), 103–117. <https://doi.org/10.1034/j.1600-0870.1995.00006.x>.
- Tribbia, J.J. and Baumhefner, D.P. (2004) Scale interactions and atmospheric predictability: an updated perspective. *Monthly Weather Review*, 132(3), 703–713.
- Van Leeuwen, P.J. (2010) Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Quarterly Journal of the Royal Meteorological Society*, 136, 1991–1999. <https://doi.org/10.1002/qj.699>.
- Vannitsem, S.S. and Lucarini, V. (2016) Statistical and dynamical properties of covariant Lyapunov vectors in a coupled atmosphere–ocean model – multi-scale effects, geometric degeneracy, and error dynamics. *Journal of Physics A: Mathematical and Theoretical*, 49(22). <https://doi.org/10.1088/1751-8113/49/22/224001>.
- Vissio, G. and Lucarini, V. (2018) A proof of concept for scale-adaptive parametrizations: the case of the Lorenz '96 model. *Quarterly Journal of the Royal Meteorological Society*, 144, 63–75.
- Weisheimer, A. and Palmer, T.N. (2014) On the reliability of seasonal climate forecasts. *Journal of the Royal Society Interface*, 11(20131162). <https://doi.org/10.1098/rsif.2013.1162>.
- Wilks, D.S. (2005) Effects of stochastic parametrizations in the Lorenz '96 system. *Quarterly Journal of the Royal Meteorological Society*, 131, 389–407.

- Wilks, D.S. (2006) *Statistical Methods in the Atmospheric Sciences*, 2nd edition. Academic Press, Cambridge, MA.
- Williams, K.D., Bodas-Salcedo, A., Déqué, M., Fermepin, S., Medeiros, B., Watanabe, M., Jakob, C., Klein, S.A., Senior, C.A. and Williamson, D.L. (2013) The transpose-AMIP II experiment and its application to the understanding of southern ocean cloud biases in climate models. *Journal of Climate*, 26(10), 3258–3274. <https://doi.org/10.1175/JCLI-D-12-00429.1>.
- Zanna, L. (2012) Forecast skill and predictability of observed Atlantic sea surface temperatures. *Journal of Climate*, 25, 5047–5056.

How to cite this article: Christensen HM, Berner J. From reliable weather forecasts to skilful climate response: A dynamical systems approach. *Q J R Meteorol Soc.* 2019;145:1052–1069. <https://doi.org/10.1002/qj.3476>