Geophysical & Astrophysical Fluid Dynamics

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ggaf20

Diagnosing the causes of bias in climate models – why is it so hard?

T. N. Palmer [a] & Antje Weisheimer [a]

[a] University of Oxford, Atmospheric, Oceanic and Planetary Physics, Clarendon Laboratory, Oxford, OX1 3PU, UK

[b] ECMWF, Shinfield Park, Reading, RG2 9AX, UK

Available online: 11 Mar 2011

To cite this article: T. N. Palmer & Antje Weisheimer (2011): Diagnosing the causes of bias in climate models – why is it so hard?, Geophysical & Astrophysical Fluid Dynamics, 105:2-3, 351-365

To link to this article: http://dx.doi.org/10.1080/03091929.2010.547194

Geophysical and Astrophysical Fluid Dynamics
Vol. 105, Nos. 2–3, April–June 2011, 351–365

Diagnosing the causes of bias in climate models – why is it so hard?

T. N. PALMER* and ANTJE WEISHEIMER

University of Oxford, Atmospheric, Oceanic and Planetary Physics, Clarendon Laboratory, Oxford, OX1 3PU, UK
ECMWF, Shinfield Park, Reading, RG2 9AX, UK

(*Received 3 March 2010; in final form 24 August 2010; first published online 11 March 2011*)

The equations of climate are, in principle, known. Why then is it so hard to formulate a bias-free model of climate? Here, some ideas in nonlinear dynamics are explored to try to answer this question. Specifically it is suggested that the climatic response to physically different forcings shows a tendency to project onto structures corresponding to the systems natural internal modes of variability. This is shown using results from complex climate models and from the relatively simple Lorenz three-component model. It is suggested that this behaviour is consistent with what might be expected from the fluctuation–dissipation theorem. Based on this, it is easy to see how climate models can easily suffer from having errors in the representation of two or more different physical processes, whose responses compensate one another and hence make individual error diagnosis difficult. A proposal is made to try to overcome these problems and advance the science needed to develop a bias-free climate model. The proposal utilises powerful diagnostics from data assimilation. The key point here is that these diagnostics derive from short-range forecast tendencies, estimated long before the model has asymptotically settled down to its (biased) climate attractor. However, it is shown that these diagnostics will not identify all sources of model error, and a so-called "bias of the second kind" is discussed. This latter bias may be alleviated by recently developed stochastic parametrisations.

Keywords: Model bias; Analysis increments; Nonlinear dynamics; Stochastic parametrisations

1. Introduction

Climate prediction models play a crucial role in today's society, governing, for example, decisions on preparing regionally for climate-related malaria epidemics, or the extent to which the global community commits to radical reductions in greenhouse gas emissions.

However, such models are not perfectly faithful representations of reality. Take any climate model, integrate it for a period of years so that it has asymptoted to its climatology, and it will be relatively easy to find specific meteorological variables which are biased against corresponding observations. Examples might include seasonal-mean

*Corresponding author. Email: tim.palmer@ecmwf.int

Geophysical and Astrophysical Fluid Dynamics
ISSN 0309-1929 print/ISSN 1029-0419 online © 2011 Taylor & Francis
DOI: 10.1080/03091929.2010.547194

upper tropospheric temperature in high latitudes, near-surface winds over the equatorial Pacific, or rainfall in the Asian monsoon region.

Whilst it is easy to identify these biases, it is another matter to determine the causes of these biases. In section 2, we discuss a specific example where what appears to be a climate bias associated with excessive drag over the ocean is in fact associated with insufficient drag over land. In section 3, this class of problem is illustrated in a much simpler setting, using the Lorenz (1963) system. The results from this study point us to a rather generic perspective on the notion of model bias, based on two rather generic features of dynamical systems: the fluctuation–dissipation theorem and the non-self adjointness of linearised dynamical operators. This leads us to a novel perspective on why it has proven so hard to eliminate model bias in climate models. In section 4, a potential solution to the problem is discussed based on the method of analysis increments. For this method to be viable in practice, the model in question should have a data assimilation capability. Currently not many climate models have such capability.

This proposal is not, however, a panacea. In section 5, a type of climate bias is discussed which cannot be diagnosed from biases in analysis increments – one related to the functional representation of sub-grid processes as deterministic. This "bias of the second kind" is very much associated with the issue of structural uncertainty in weather and climate models, and has led to the development of stochastic rather than deterministic parametrisations.

This article is dedicated to Raymond Hide in celebration of his 80th birthday. Raymond was the single most influential person in the first author's decision to move fields from his doctoral topic in general relativity theory. Raymond's influence continued during TNP's early years in the Met Office. During this period TNP spent many a morning tea break – often extended to lunch time – learning about rotating fluids and dynamo theory. This article touches on many of Raymond's interests – the angular momentum balance of the atmosphere, on predictability, turbulence and chaos (many of Ed Lorenz's formative ideas on chaos were motivated by Raymond's rotating annulus experiments), and on ideas in non-equilibrium thermodynamics. This latter topic has a special resonance for TNP: in the last year of his doctoral training, TNP's supervisor, the cosmologist Dennis Sciama, began a programme of work aimed at deriving Hawking's famous black-hole radiation formula from the fluctuation–dissipation theorem and the principle of maximum entropy production. TNP was astonished to learn from Raymond that the same ideas in non-equilibrium thermodynamics were also being exploited to understand climate (e.g. Paltridge 1975). As mentioned below, ideas in non-equilibrium thermodynamics may also be relevant in the discussion about model diagnosis of climate model bias.

2. Climate models and Jeffrey's theory of the westerly flow

As mentioned in section 1, comprehensive weather and climate prediction models play a crucial role in society today. One would like to incorporate as much as possible the (partial differential equation) laws of physics into these models. But high-resolution models are computationally expensive: a doubling of resolution can result in an increase in computing time of up to 2^4 (the exponent representing the dimension of space-time). What is a minimal resolution needed if we are only interested in simulating horizontal

scales of, let us say, a thousand kilometres or more? Perhaps a model with grid point spacing of a few hundred kilometres might be sufficient.

Figure 1(a) shows the climatological boreal winter surface pressure in Northern and Southern Hemispheres (based on ECMWF analyses) whilst figure 1(b) shows the equivalent fields simulated by a (Met Office) climate model (c. 1980s) run with a grid with 5° spacing in the longitudinal direction and 7.5° in the latitudinal direction.

If one was to make a very broad-brush diagnosis of this simulation, one might say that the pressure distribution was not too unrealistic in the Northern Hemisphere, but was very poor in the Southern Hemisphere: by geostrophy, one could deduce from figure 1(b) that instead of "roaring forties", the model has "whispering forties". Since most of the Southern Hemisphere surface is ocean, perhaps one might conclude that the cause of the model bias lies in an overestimation of the model's drag coefficient over sea.

However, instead of pursuing this idea further, suppose a further set of simulations is performed at the higher horizontal resolution with grid point spacing half that in figure 1(b). The results are shown in figure 1(c). Without any change to the oceanic drag coefficient, the Southern Hemisphere surface pressure distribution is now simulated quite well. However, a price is paid for this improvement: the Northern Hemisphere surface pressure simulation deteriorates and the corresponding surface flow is excessively westerly.

What is going on? First, let us recall Jeffrey's (1926) seminal contribution to the theory of the atmospheric general circulation: the surface winds in midlatitudes are maintained against friction by a poleward flux of angular momentum from lower latitudes, generated by the extratropical weather disturbances. Although at the time this idea seemed contrary to the notion that turbulent fluxes should be downgradient, development of the theory of baroclinic instability confirmed Jeffrey's proposal. In particular, the flux of angular momentum needed to maintain the midlatitude westerlies against frictional drag can be generated provided that the baroclinic eddy troughs and ridges have a tilted north east/south west orientation. Resolving such tilted structures in a numerical model may require significantly more meridional resolution than correspondingly structures without tilt. This was indeed found in the idealised baroclinic lifecycle experiments of Simmons and Hoskins (1976).

Hence, by increasing model resolution, the "whispering forties" bias has been cured. But why then, has Northern Hemisphere bias been worsened by increasing resolution? The answer is that in the Northern Hemisphere, diagnosis of the causes of bias in the low resolution model is compounded by a compensation of errors. The model not only had an inadequate representation of angular momentum flux associated with the baroclinic eddies, but also had an inadequate representation of frictional coupling to the land surface. In particular the versions of the model shown in figure 1(b) and (c) had no parametrised representation of momentum coupling of the atmosphere to the solid earth associated with unresolved orography. Since part of this momentum coupling is in the form of orographically forced gravity waves, a parametrisation of the vertical propagation and breaking of such waves in the lower stratosphere was developed once the biases of $2° \times 3°$ (or T42) and higher resolution models became apparent (Palmer *et al.* 1986).

Hence, what began as a suggestion that the Met Office model had too much drag over the ocean, ended by concluding that the real problem was insufficient drag over land! Along the way, it was found that inadequate representation of horizontal baroclinic

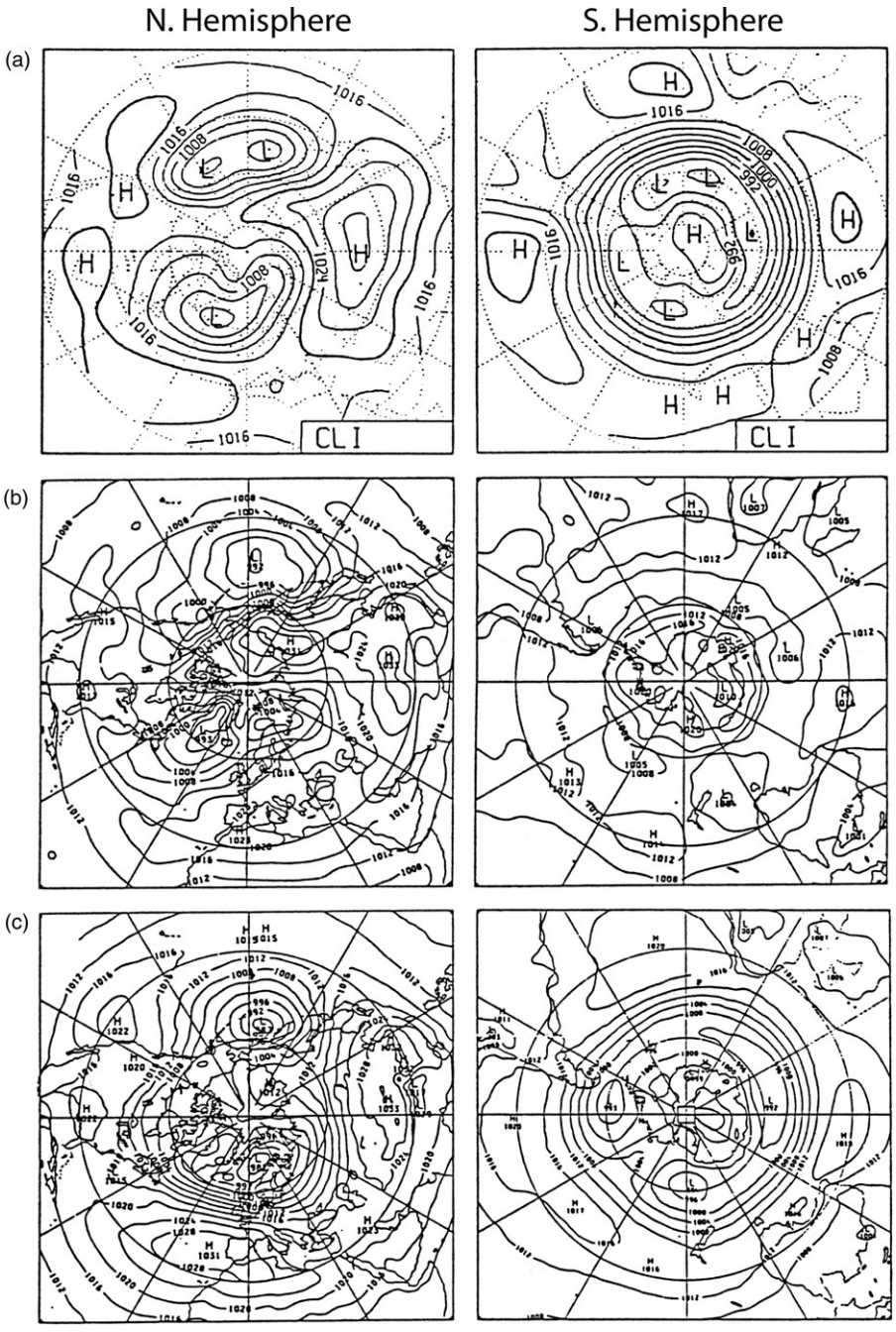

Figure 1. Boreal winter (December–February) surface pressure for Northern Hemisphere (left) and Southern Hemisphere (right). (a) From observations, (b) from a low resolution model simulation, and (c) from a model whose resolution is double that in (b).

wave fluxes was compensating for insufficient representation of vertical orographic gravity-wave fluxes. How could two completely different processes have such compensating effects on the climate of the model?

3. The forced Lorenz model and the fluctuation-dissipation theorem

To pursue this question further, consider a much simpler system, the "forced" Lorenz (1963) model:

$$\dot{X} = -\sigma X + \sigma Y + f\cos\theta, \tag{1a}$$

$$\dot{Y} = -XZ + rX - Y + f\sin\theta, \tag{1b}$$

$$\dot{Z} = XY - bZ, \tag{1c}$$

where $\sigma = 10, r = 28, b = 4/3$, and a vector $\mathbf{F} = (F_X, F_Y, F_Z) = (f\cos\theta, f\sin\theta, 0)$ has been added to the canonical Lorenzian equations. Like the weather and climate models discussed above, the system (1a–c) is chaotic for sufficiently small f and therefore has limited predictability.

Let us suppose that the "true" system is the unforced Lorenz system (with $f = 0$), and our "model" for the true system has a systematic error, represented by $f = 10$ and some $\theta = \theta_0$. In the discussion below, we will fix $\theta_0 = 3\pi/4$. However, suppose this value is "hidden" from us. Our job is to diagnose the systematic error of the "model" to determine the value of θ_0.

Figure 2(a) shows a time series of the X-component of the unforced Lorenz model – "truth". Figure 2(b) shows a time series of the X-component of the state vector of the "model". As can be seen, the probability of finding the state in the regime with positive X is greater than what it would be from "observations" of the "true" system (figure 2(a)). Hence, not knowing θ_0, a reasonable guess would be that $\theta_0 = 0$. The time series of the X-component generated with $\theta_0 = 0$ is shown in figure 2(c), and indeed it does resemble, in some statistical sense at least, the time series in figure 2(b). However, the X-component of the forcing error F_X with $\theta_0 = 0$ is in fact opposite in sign to the value associated with the actual value $\theta_0 = 3\pi/4$ used for our erroneous model.

Figure 3 summarises the relationship between the angle θ and the direction ψ associated with the time-mean response of the model to the forcing $\mathbf{F}$, i.e. the angle to the X-axis of the line which joins the time-mean state of the forced model and the time-mean state of the unforced model – the origin in the X–Y plane. Figure 3 shows that, for a selection of values θ, there is a tendency for the response to point along the diagonal in the X–Y plane where $\psi = \pi/4$. Generally, the response does not lie exactly along this line, but to a first approximation it does. What is special about this line? It can be shown (Selten 1995) that this line corresponds to the dominant empirical orthogonal function (EOF) of the unforced Lorenz model; that is, it corresponds to the leading eigenvector of the lag-zero covariance matrix of the (three-dimensional) state vector of the unforced system.

This opens a possible link to the example in the previous section. It is well-known that the so-called annular modes correspond to dominant EOFs for the atmosphere. The Northern Annular Mode, or Arctic Oscillation, (Thompson and Wallace 1998)

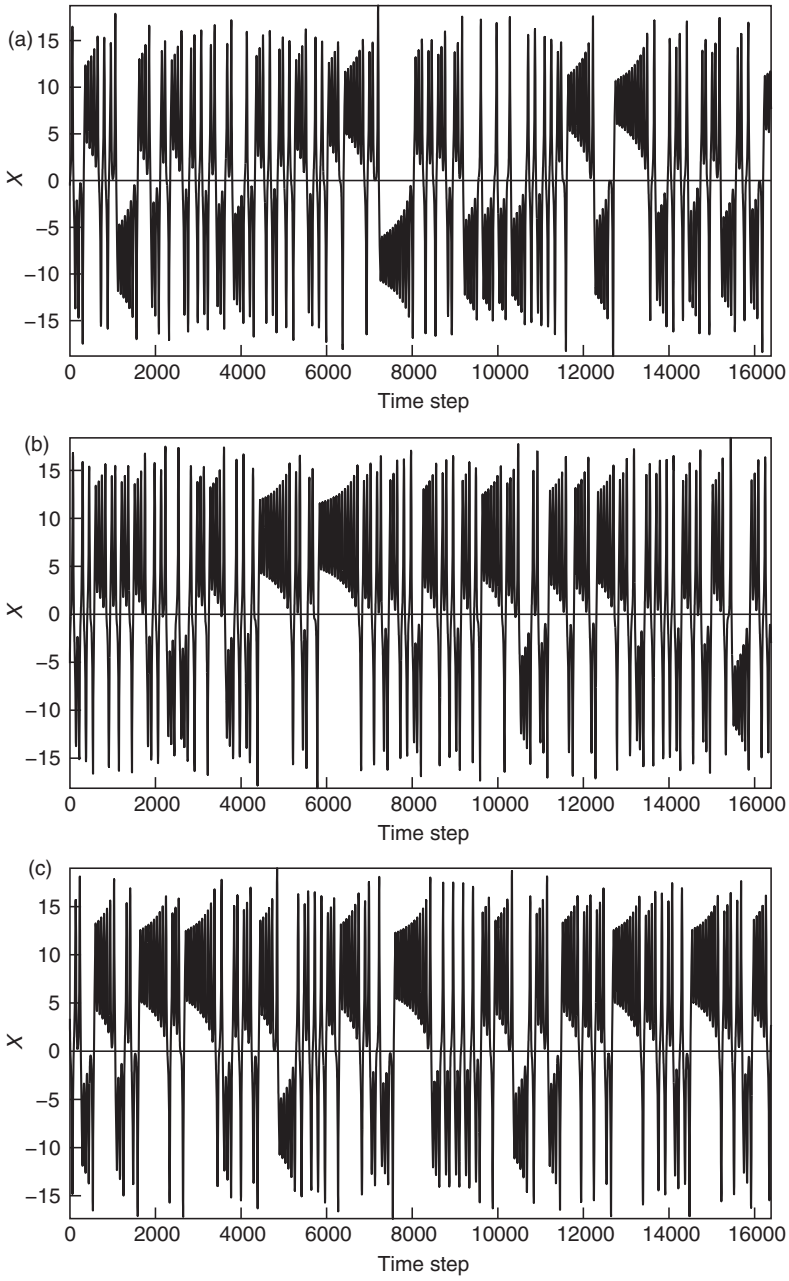

Figure 2. Time series of the X-component: (a) of the unforced Lorenz system, (b) and (c) of the forced Lorenz system (1a–c) for $\theta = 3\pi/4$ and $\theta = 0$, respectively.

corresponds largely to fluctuations in the zonal wind in the Northern Hemisphere, with the Southern Annular Mode playing a corresponding role in the Southern Hemisphere. As we saw, there was a strong response of the model to changes in resolution in terms of zonal modes, and in the low resolution model, there appeared to be a compensation in terms of these zonal structures between insufficient sub-grid vertical frictional

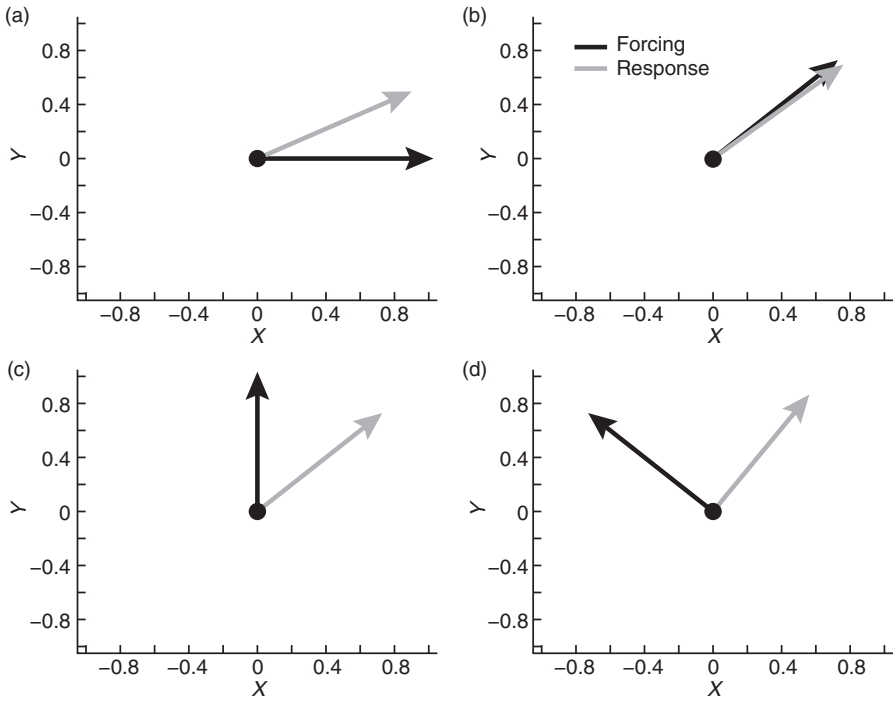

Figure 3. Time-mean response to forced Lorenz system (grey arrow) in relation to direction of forcing (black arrow) in the X–Y plane: (a) $\theta = 0$, (b) $\theta = \pi/4$, (c) $\theta = \pi/2$ and (d) $\theta = 3\pi/4$.

coupling over land and insufficient horizontal momentum transport (primarily over the ocean). But why should the dominant EOF of a dynamical system appear to play a key role in determining the response to some imposed forcing?

One of Einstein's papers, published in his "Annus Mirabilis" of 1905 was on the theory of Brownian motion (Einstein 1905). In this article, Einstein established that the same random forces which cause the erratic movement of a particle in Brownian motion would also cause drag if the particle were pulled through the fluid. This result in turn became developed and generalised to the so-called fluctuation–dissipation theorem in statistical thermodynamics, quantifying the relation between the fluctuations of a system in thermal equilibrium and the response of the system to applied perturbations. Leith (1975) has applied the fluctuation–dissipation theorem to understand the forced response of the atmosphere. Let

$$\dot{\mathbf{X}} = \mathbf{F}[\mathbf{X}], \quad \dot{\mathbf{X}}' = \mathbf{F}[\mathbf{X}'] + \delta\mathbf{f}, \tag{2a,b}$$

and $\delta\mathbf{X} = \mathbf{X}' - \mathbf{X}$, then (Leith's version of) the fluctuation dissipation theorem states

$$\delta\bar{\mathbf{X}} = \mathbf{L}\delta\mathbf{f}, \tag{3}$$

where the overbar represents a long-time average and

$$\mathbf{L} = \int_0^\infty \mathbf{C}(\tau)\mathbf{C}^{-1}(0)\mathrm{d}\tau, \tag{4}$$

where $\mathbf{C}$ is the lag-τ covariance matrix of $\mathbf{X}$.

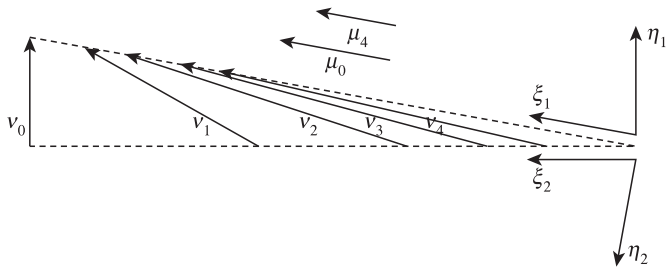

Figure 4. Schematic illustration of perturbation growth in a highly non-self adjoint system. The perturbation μ_0 projects onto the dominant eigenvector ξ_1. The optimal perturbation v_0 projects onto the adjoint η_1 of ξ_1. From Palmer (1996).

One can question whether the fluctuation–dissipation theorem holds quantitatively for a system like the atmosphere, far from equilibrium. Also, the theorem will not hold quantitatively for the dissipative Lorenz system either. Nevertheless, qualitatively we see in the fluctuation–dissipation theorem the notion that the response of a system to some prescribed forcing can be strongly conditioned by that system's internal modes of variability, i.e. the response to the forcing will be conditioned by the projection of $\delta\mathbf{f}$ in the direction of the leading eigenvectors of $\mathbf{L}$. Additionally, given the non-self adjoint nature of $\mathbf{L}$ (Farrell and Ioannou 1996), perturbations which optimally excite the leading eigenvector of $\mathbf{L}$ need not point in the direction of this eigenvector (figure 4)

What to do? The results above relate to problems of diagnosing the cause of model bias from integrations where the model has asymptoted towards its climatology. In principle, this suggests a relatively simple solution: perform the diagnosis well before the model integrations have asymptoted to climatology. We study this in the following section.

4. Can a 6-h weather forecast help determine Earth's climate 100 years from now?

Climate change is the defining issue of our age, yet predictions of climate for the end of this century remain remarkably uncertain. In large part this is because the amplification of increases in greenhouse gases by cloud-radiative interactions remains uncertain. This is turn arises because the parametrised representation of clouds themselves is especially uncertain.

The issue of uncertainty was put into sharp focus by analysis of the http://climateprediction.net ensemble of climate change projections. According to Stainforth *et al.* (2005), climate sensitivity is predicted to be as large as 12 K or more. As discussed in Rodwell and Palmer (2007), many of the models producing such strong global warming signals had convective parametrisations with anomalously small values of the convective entrainment parameter.

Although the amplification of the effect of enhanced CO_2 by convective cloud systems will occur on timescales of decades, the intrinsic timescale associated with a deep convective system itself is typically on the order of hours. Hence it should in principle be possible to assess whether the anomalously small values of convective entrainment are realistic or not, by studying the performance of such models in short-range weather prediction mode.

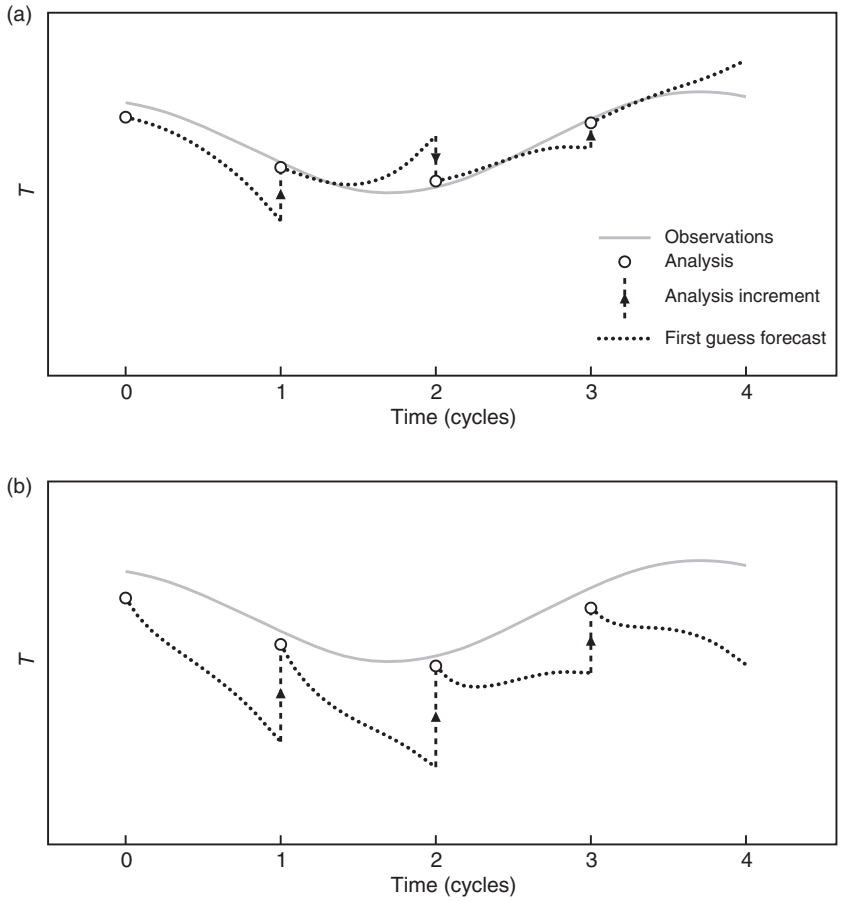

Figure 5. (a) Schematic diagram of data assimilation for an unbiased model and (b) as (a) but for a biased model.

This can indeed be done, as reported in Rodwell and Palmer (2007). The technique is illustrated in figure 5 based on the technique proposed by Klinker and Sardeshmukh (1992). Essentially the idea is to look at the mean "analysis increment" averaged over a month of four-times-a-day atmospheric analyses. Here an analysis increment is defined as the difference between a 6-h forecast and the corresponding objective analysis of the contemporary observations, valid at the same time as the forecast. These objective analyses are used to initialise weather predictions. Over a sufficiently long-time series of analyses, the mean analysis increment should be close to zero. However, if the model is biased against the observations, then the mean analysis increment will be non-zero.

This approach to model diagnosis overcomes the constraints of the fluctuation–dissipation theorem because the approach is based on a diagnosis of model output well before any asymptotic climatological state is reached. In practical terms this means, for example, that mean analysis increments associated with an error in the representation of orographic gravity wave momentum–flux convergence will be largest in the momentum equation and in the lower stratosphere above regions of large sub-grid orography where

such waves tend to break. Indeed this was one of the first applications of this technique (Klinker and Sardeshmukh 1992).

Rodwell and Palmer (2007) found that the mean analysis increment in a model with anomalously small entrainment parameter (as used in the climateprediction.net experiments) was substantially larger than that from a model with more typical values for this parameter.

Analysis increments provide a diagnostic tool that is used to assess routinely biases in the ECMWF system (Rodwell and Jung 2008). We propose here that it could prove an invaluable tool for climate modelling, in reducing bias in models and in reducing uncertainty in projections of climate change. However, in order to implement such a tool, the modelling system must have data assimilation capability. Currently, climate prediction models do not typically have this capability. However, in recent years, the concept of seamless prediction (Palmer *et al.* 2008) is bringing weather forecast and climate prediction models closer together. This will allow this technique to be explored more thoroughly in climate prediction mode. For the climate problem, one particularly relevant extension of this technique will be to address systematic biases in the ocean model component of climate models. Whether or not the method could be used, once carbon dioxide concentrations are properly assimilated into Earth system models, to constrain model representations of the carbon cycle, and so on, remains to be seen.

5. Bias of the second kind

Have we solved the problem of diagnosing the causes of model bias? That is, is it sufficient to focus on the initial tendencies of the model from an ensemble of initial states? Clearly this technique will not work for diagnosing very long-timescale processes, e.g. associated errors in the representation of the carbon cycle. On the other hand, many of the important uncertainties in climate models are associated with rather fast timescale processes linked to clouds, boundary layer turbulence etc.

However, as discussed in this section, there is a second type of model error which will not be revealed by this type of diagnosis.

To see this, let us return to the Lorenz (1963) system, this time written in terms of the three principal components of the model (Selten 1995):

$$\dot{a}_1 = 2.3a_1 - 6.2a_3 - 0.49a_1a_2 - 0.57a_2a_3, \tag{5a}$$

$$\dot{a}_2 = -62 - 2.7a_2 + 0.49a_1^2 - 0.49a_3^2 + 0.14a_1a_3, \tag{5b}$$

$$\dot{a}_3 = -0.63a_1 - 13a_3 + 0.43a_1a_2 + 0.49a_2a_3. \tag{5c}$$

Now, it turns out that the third principal component only explains about 4% of the variance of the total system. We might therefore consider parametrising the equation for the third principal component in the following form:

$$\dot{a}_1 = 2.3a_1 - 6.2a_3 - 0.49a_1a_2 - 0.57a_2a_3, \tag{6a}$$

$$\dot{a}_2 = -62 - 2.7a_2 + 0.49a_1^2 - 0.49a_3^2 + 0.14a_1a_3, \tag{6b}$$

$$\dot{a}_3 = P(a_1, a_2; \alpha, \beta, \ldots), \tag{6c}$$

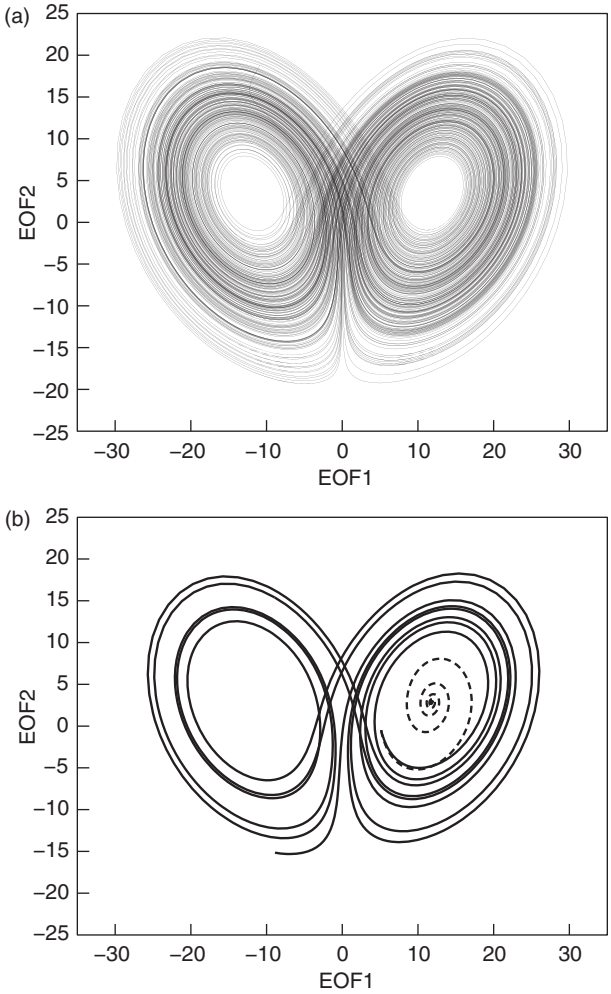

Figure 6. (a) Trajectories of Lorenz model in a principal component basis and (b) solid lines as (a), dashed lines for a 2D model with deterministically parametrised third principal component.

where P is some deterministic formula and $\alpha, \beta, \ldots$ are parameters. Figure 6 shows integrations of the full model (equations (5a–c)) and the parametrised model (equations (6a–c)) where $P = \alpha a_1 + \beta a_2$.

It can be seen that the long-term climate of the parametrised model is clearly not chaotic. This is a consequence of the Poincaré–Bendixson theorem, whereby the state space of a chaotic system based on autonomous differential equations must have at least 3 dimensions ((6a–c) is an autonomous system with only two degrees of freedom). But on the other hand, it can also be seen that the parametrised model is quite accurate for short-range forecasts. In this case, the analysis increment approach would not diagnose the fault with the parametrised model, since in the short range the parametrised model is clearly skilful.

What is this fault? Here the model deficiency lies in the use of a deterministic function P for the parametrisation: the specific linear form above is irrelevant. If we replace the

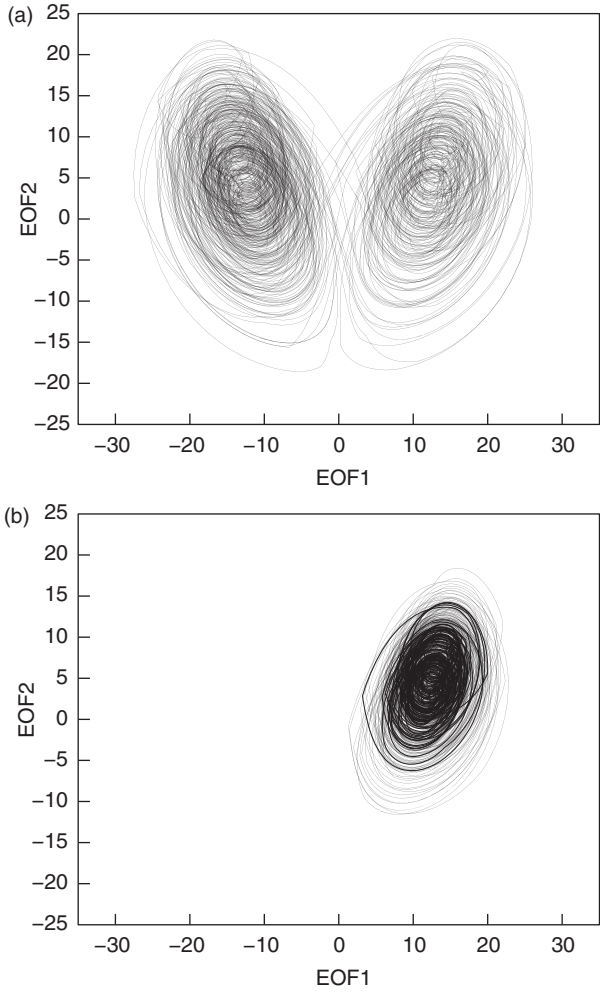

Figure 7. (a) Lorenz model with stochastically parametrised third principal component and (b) as (a) but with smaller amplitude noise.

deterministic parametrisation of equation (6c) with a stochastic parametrisation

$$\dot{a}_1 = 2.3a_1 - 6.2a_3 - 0.49a_1a_2 - 0.57a_2a_3, \tag{7a}$$

$$\dot{a}_2 = -62 - 2.7a_2 + 0.49a_1^2 - 0.49a_3^2 + 0.14a_1a_3, \tag{7b}$$

$$a_3 = \beta, \tag{7c}$$

where β is a stochastic variable, then the broad structure of the Lorenz attractor is simulated reasonably well, as shown in figure 7(a). The amplitude and temporal autocorrelation have to be correctly tuned to give this structure. Figure 7(b) shows a simulation of equations (7a–c) with weaker noise. The simulation shows a bias in both mean state and internal variability.

Hence, we see that there is more to parametrisation error than the determination of the parameters or the functional form of the parametrisation. The very assumption

of a deterministic link between the unresolved scales and the resolved scales is here brought into question. We will refer to a bias resulting from such a structural uncertainty as a "bias of the second kind". The development of stochastic parametrisations in numerical weather prediction models addresses some of these deficiencies (Buizza *et al*. 1999, Lin and Neelin 2002, Palmer 2001, Craig and Cohen 2006, Palmer and Williams 2010). In a nonlinear system, deficiencies in the lack of sub-grid variability can lead to systematic bias in the climate model, but, as suggested in figure 6, the timescale for the development of such biases can be quite slow. Hence, how are we to diagnose model deficiencies associated with the lack of sub-grid variability?

One idea that seems to have considerable potential is by developing stochastic parametrisation through coarse-grain budgets of cloud-resolving models (Shutts and Palmer 2007). For example, the cloud-resolving model may have a resolution of 1 km and one estimates coarse-grain budgets over boxes of size c. 100 km, a typical dimension for a climate model grid box. Here one treats cloud-resolving model output as a surrogate for truth. By treating the cloud-resolving model output as truth, one has exact estimates of the sub-100-km grid tendencies. Based on this one can estimate probability distributions of sub-100-km grid tendencies, conditioned on the 100 km average flow. This method has been used to provide a partial validation for stochastic parametrisations used at ECMWF: both the Stochastically Perturbed Parametrisation Tendency Scheme, and the Stochastic Backscatter Scheme (Palmer *et al*. 2009).

An interesting recent development in this respect has been that of ensemble data assimilation. A relevant diagnostic in ensemble data assimilation is the spread of ensemble at the 6 h lead time. This should be balanced by the ensemble mean analysis increment. The extent to which it is not is a measure of the imperfection of the ensemble of data assimilations in representing analysis error, here a combination of observation error and model error. Recent studies (M. Bonavita, ECMWF) of this diagnostic suggest areas where stochastic parametrisation could be improved. In this way, it is possible that the analysis increment method can be extended to diagnose errors in stochastic parametrisation. However, at present these ideas are at a rudimentary stage of development.

6. Conclusions

Climate prediction models provide the scientific input which underpins climate change mitigation treaties and adaptation strategies. Whilst there has been considerable improvement in climate simulations over recent years, climate models have quantifiable shortcomings and develop biases of magnitude comparable to the climate change signals such models are trying to predict. It is clearly important to try to reduce these biases, and yet diagnosis of model error is difficult not least because of the problem of compensating errors: the response to errors in the representation of two quite different processes in a climate model can partially cancel each other out. This problem of compensating errors can be illustrated in relatively simple nonlinear models, but may be ultimately linked to rather generic properties of nonlinear systems.

A technique is proposed to overcome some of these problems, based on the concept of analysis increments. However, this technique requires the model to come with a data

assimilation system: currently few climate models have this capability. However, with the development of seamless prediction systems, there is a prospect of significant advances in the future.

Finally, a "bias of the second kind" has been discussed. This is neither associated with the functional form of sub-grid parametrisations, nor of the values of the free parameters associated with such parametrisations, but rather with the fact that such parametrisations are deterministic. The development of stochastic parametrisations, aided by coarse-grain budget analyses from cloud-resolving models, may provide the means to reduce such biases.

Climate prediction remains one of the most computationally challenging problems in science and it has proven difficult to reduce uncertainty in estimates of anthropogenic global warming over the years. The analysis increment techniques proposed here, when properly integrated with other, more conventional diagnostic methodologies, provide some grounds for believing that significant reductions of uncertainty may in fact be possible in future years.

Acknowledgement

We thank Dr Mark Rodwell for helpful discussions and for providing Figure 5.

References

Buizza, R., Miller, M.J. and Palmer, T.N., Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.* 1999, **125**, 2887–2908.

Craig, G.C. and Cohen, B.G., Fluctuations in an equilibrium convective ensemble. Part I: theoretical formulation. *J. Atmos. Sci.* 2006, **63**, 1996–2004.

Einstein, A., Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Phys.* 1905, **322**, 549–560.

Farrell, B.F. and Ioannou, P.J., Generalized stability theory part I: autonomous operators. *J. Atmos. Sci.* 1996, **53**, 2025–2040.

Jeffreys, H., On the dynamics of geostrophic winds. *Q. J. R. Meteorol. Soc.* 1926, **52**, 85–101.

Klinker, E. and Sardeshmukh, P.D., The diagnosis of mechanical dissipation in the atmosphere from large-scale balance requirements. *J. Atmos. Sci.* 1992, **49**, 608–627.

Leith, C.E., Climate response and fluctuation dissipation. *J. Atmos. Sci.* 1975, **32**, 2022–2026.

Lin, J.W.-B. and Neelin, J.D., Considerations for stochastic convective parameterization. *J. Atmos. Sci.* 2002, **59**, 959–975.

Lorenz, E.N., Deterministic nonperiodic flow. *J. Atmos. Sci.* 1963, **20**, 130–141.

Palmer, T.N., Predictability of the atmosphere from days to decades. In *Decadal Climate Variability*, edited by D.L.T. Anderson and J. Willebrand, Vol. 44, 493pp., 1996 (*NATO ASI Series I, Global Environmental Change*, Les Houches: NATO Advanced Study Institute).

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

Palmer, T.N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G.J., Steinheimer, M. and Weisheimer, A., Stochastic parametrization and model uncertainty. *ECMWF Tech. Memo* 2009, **598**, 42pp.

Palmer, T.N., Doblas-Reyes, F., Weisheimer, A. and Rodwell, M., Reliability of climate change projections of precipitation: towards "seamless" climate predictions. *Bull. Am. Meteor. Soc.* 2008, **89**, 459–470.

Palmer, T.N., Shutts, G.J. and Swinbank, R., Alleviation of a systematic westerly bias in general circulation and numerical weather prediction models through an orographic gravity wave parametrization. *Q. J. R. Meteorol. Soc.* 1986, **112**, 1001–1039.

Palmer, T.N. and Williams, P. (eds.), *Stochastic Physics and Climate Modelling*, 2010 (Cambridge: Cambridge University Press).

Paltridge, G.W., Global dynamics and climate – A system of minimum entropy exchange. *Q. J. R. Meteor. Soc.* 1975, **101**, 475–484.

Rodwell, M.J. and Jung, T., Understanding the local and global impacts of model physics changes: an aerosol example. *Q. J. R. Meteor. Soc.* 2008, **134**, 1479–1497.

Rodwell, M.J. and Palmer, T.N., Using numerical weather prediction to assess climate models. *Q. J. R. Meteorol. Soc.* 2007, **133**, 129–146.

Selten, F.M., An efficient empirical description of large-scale atmospheric dynamics. Ph.D. Thesis, Vrije Univ. Amsterdam, 1995.

Shutts, G.J. and Palmer, T.N., Convective forcing fluctuations in a cloud resolving model: relevance to the stochastic parametrisation problem. *J. Climate* 2007, **20**, 187–202.

Simmons, A.J. and Hoskins, B.J., Baroclinic instability on the sphere: normal modes of the primitive and quasi-geostrophic equations. *J. Atmos. Sci.* 1976, **33**, 1454–1476.

Stainforth, D.A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D.J., Kettleborough, J.A., Knight, S., Martin, A., Murphy, J.M., Piani, C., Sexton, D., Smith, L.A., Spicer, R.A., Thorpe, A.J. and Allen, M.R., Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature* 2005, **433**, 403–406.

Thomson, D.W.J. and Wallace, J.M., The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.* 1998, **25**, 1297–1300.