Towards the Probabilistic Earth-System Simulator: A Vision for the Future of Climate and Weather Prediction

By

T.N.Palmer

Atmospheric, Oceanic and Planetary Physics, University of Oxford

European Centre for Medium-Range Weather Forecasts, Reading

Based on the 2011 Royal Meteorological Society Presidential Address

To appear Quart.J.Roy.Meteor. Soc

"You can thank your lucky stars that you are not economists. Those poor souls don't even know their equations!" Sir John Mason, Director-General Meteorological Office, to his 1977 graduate intake.

"I believe that the ultimate climate models..will be stochastic, ie random numbers will appear somewhere in the time derivatives" Lorenz (1975).

Abstract

There is no more challenging problem in computational science than that of estimating, as accurately as science and technology allows, the future evolution of Earth's climate; nor indeed is there a problem whose solution has such importance and urgency. Historically, the simulation tools needed to predict climate have been developed, somewhat independently, at a number of weather and climate institutes around the world. Whilst these simulators are individually deterministic, it is often assumed that the resulting diversity provides a useful quantification of uncertainty in global or regional predictions. However, this notion is not well founded theoretically and corresponding "multi-simulator" estimates of uncertainty can be prone to systemic failure. Separate to this, individual institutes are now facing considerable challenges in finding the human and computational resources needed to develop more accurate weather and climate simulators with higher resolution and full Earthsystem complexity. A new approach, originally designed to improve reliability in ensemble-based numerical weather prediction, is introduced to help solve these two rather different problems. Using stochastic mathematics, this approach recognises uncertainty explicitly in the parametrised representation of unresolved climatic processes. Stochastic parametrisation is shown to be more consistent with the underlying equations of motion and, moreover, provides more skilful estimates of uncertainty when compared with estimates from traditional multi-simulator ensembles, on timescales where verification data exists. Stochastic parametrisation can also help reduce long-term biases which have bedevilled numerical simulations of climate from the earliest days to the present. As a result, it is suggested that the need to maintain a large "gene pool" of quasi-independent deterministic simulators may be obviated by the development of probabilistic Earth-system simulators. Consistent with the conclusions of the World Summit on Climate Modelling, this in turn implies that individual institutes will be able to pool human and computational resources in developing future-generation simulators, thus benefitting from economies of scale; the establishment of the Airbus consortium provides a useful analogy here. As a further stimulus for such evolution, discussion is given to a potential new synergy between the development of dynamical cores, and stochastic processing hardware. However, it is concluded that the traditional challenge in numerical weather prediction, of reducing deterministic measures of forecast error, may increasingly become an obstacle to the seamless development of probabilistic weather and climate simulators, paradoxical as that may appear at first sight. Indeed, going further, it is argued that it may be time to consider focussing operational weather forecast development entirely on high-resolution ensemble prediction systems. Finally, by considering the exceptionally challenging problem of quantifying cloud feedback in climate change, it is argued that the development of the probabilistic Earth-system simulator may actually provide a route to reducing uncertainty in climate prediction.

1. Introduction

The problem of understanding and predicting climate is fundamentally a scientific one, but with extraordinary relevance for society. However, our understanding and ability to predict climate is still rudimentary. For example, due to profound uncertainties, primarily with the hydrological cycle, we are still unable to rule out the possibility that anthropogenic climate change will be catastrophic for humanity over the coming century, or something to which we can adapt relatively easily. Hence, whilst climate policy on mitigation or adaptation is rightly based on risk assessment, the risks cover a very broad range of potential outcomes, presenting a barrier to clear-cut policy and decision making. How well do we understand these uncertainties? Are they irreducible? Could the climate science community do better in reducing uncertainty? Key conclusions of this paper are that whilst there indeed are irreducible uncertainties in predicting climate, and our understanding of these uncertainties is poor, new techniques promise not only to improve our ability to quantify climate prediction uncertainties more reliably, these techniques may actually help reduce uncertainty.

To take this further, the analysis presented in this paper suggests that development of new scientific tools to quantify uncertainty in predictions of climate more reliably, have implications for the way in which weather and climate institutes are themselves organised, both internally, and with respect to one another. For example, it could be argued that the existence of a substantial "gene pool" of quasi-independent climate simulators[1] not only allows an assessment of uncertainty in climate predictions (through the internal spread of the corresponding multi-simulator ensembles), it also engenders a spirit of competition between institutes thereby fostering creativity. Whilst these arguments have merits, there are counterarguments to be discussed in this paper: firstly that multi-simulator ensembles may be prone to systemic failure due to shortcomings in the basic numerical ansatz used to formulate all contemporary simulators, and secondly that the limited human and computational resources available at the institutional level are major obstacles to the development of more accurate climate simulators.

---

[1] Throughout this paper, the word "simulator" is used instead of the more conventional word "model" (cf Goldstein and Rougier, 2004). This may irritate some readers within the weather and climate community. However, for the public and many policy makers too, use of the word "model" has a tendency to conjure up a picture of a child's toy. Some so-called climate "sceptics" take advantage of this word association in portraying climate models merely as glorified computer games and not as the sophisticated mathematical representations of basic laws of physics that they are. When communicating with the public we have a tendency to use our own jargon, often subconsciously; hence we use the word "model" in public because that's what we use amongst ourselves, unaware of these pejorative word associations. Perhaps using the word "simulator" will engender more respect for these numerical representations. Modern commercial pilots are trained almost exclusively on simulators; that apparently does not deter the public from flying. (If instead the pilots were trained merely on "models" perhaps the public would be deterred!) As such, it may be time to start using the word "simulator" in place of "model" even within scientific discourse.

The new scientific element introduced into this discussion hinges on a developing programme to reformulate stochastically our weather and climate prediction simulators. This “stochastic” programme has emerged from the numerical weather prediction (NWP) community (eg Buizza et al, 1999; Palmer, 1997, 2001), and its relevance to the climate problem can be seen as exemplifying the “seamless prediction” philosophy (Palmer and Webster, 1993; WCRP, 2005; Slingo and Palmer, 2011) whereby the insights and constraints of NWP are brought to the climate table. The outline of the paper is as follows. In Section 2, a number of reasons are given as to why incremental developments in the *status quo* for climate simulation science may not be able to provide the needed improvements in coming years. Section 3 discusses a programme to reformulate our comprehensive weather and climate simulators stochastically. Results are presented indicating how ensembles based on a single simulator with stochastic representations of simulator uncertainty can outperform the more conventional multi-simulator approach to uncertainty. Discussion of the need to integrate this stochastic approach into programmes of basic simulator development are discussed in Section 4, using standard arguments familiar in other areas of physics. Section 5 discusses, briefly, a potential synergy between the development of probabilistic weather and climate simulators, and an emerging computer hardware design where exact bit-reproducibility is sacrificed in order to improve energy efficiency. Section 6 presents an analysis of one obstacle to progress, indeed it is suggested that it may be time to stop production of a separate deterministic weather forecast, and to focus entirely on the development of probabilistic prediction systems – this may also require some evolution of practices in weather forecast offices too. Section 7 presents a vision for the development of future generation probabilistic weather and climate simulators, using the establishment of the successful Airbus consortium as an analogy. It is argued, focussing on the thorny issue of cloud feedback in climate change prediction, that the development of the probabilistic Earth-system simulator may actually help reduce uncertainty in the magnitude (and indeed sign) of this feedback. Conclusions are given in Section 8.

A key aspect of this paper is that it provides new scientific arguments to support the conclusions of the World Summit on Climate Modelling (Shukla et al, 2010) that the community worldwide should be evolving towards a small number of high-resolution Earth-system simulators, possibly based the major geopolitical groupings.

Regarding the quotes at the beginning of the paper, the author was very lucky to be one of Sir John Mason’s new graduate intake in 1977, and has enjoyed the most marvellous career as a result, at the Met Office, at the European Centre for Medium-Range Weather Forecasts (ECMWF), and now at Oxford University. The author agrees with Sir John’s quote at the beginning of the paper, but only up to a point! And the point, as with so many other points of foundational importance on prediction and predictability, was first made by Ed Lorenz, with whom the author has had the privilege to interact during Ed’s many visits to ECMWF.

In the discussion below, the importance and urgency of developing reliable climate simulators – to inform global policy on climate mitigation, to help society adapt to climate change, and to assess the impacts of proposals to actively geoengineer climate - will be assumed.

2. A critique of the traditional deterministic weather and climate simulator

a) The gene pool of "ab initio" climate simulators

Arrhenius (1896) developed the first mathematical simulator to quantify the effects of anthropogenic climate change. Based on the notion of energy balance in one dimension, the simulator incorporated both the direct greenhouse effect from increased carbon dioxide, and the positive amplifying effect of water vapour, the latter through an assumption that as the atmosphere warms, its relative humidity will remain constant.

The key problem with this approach is that water, unlike carbon dioxide, is not well-mixed in the atmosphere, and water's three dimensional distribution, in all its phases, is sensitive to dynamical effects. The development of "*ab initio*" climate simulators, where dynamical effects are represented using the Navier-Stokes equations and notions such as constant relative humidity are not assumed, began with the work of Phillips (1956), who was able to adapt the simulators emerging in the rapidly developing field of numerical weather prediction (NWP). The first projections of anthropogenic climate change using such *ab initio* climate simulators were given by Manabe and Wetherald (1975).

Over the years, a diversity of *ab initio* climate simulators has been produced, as individual institutes around the world sought to replicate and extend the work of these pioneers. This diversity (sometimes referred to as a "gene pool") can be seen as a virtue. By not putting "all our eggs in one basket", the diversity of predictions provides an estimate of prediction uncertainty. For example, results in the IPCC Fourth Assessment Report (Solomon et al, 2007) are based on a pool of coordinated projections made by some 24 climate simulators developed in different climate institutes (CMIP3: Meehl et al, 2007). A similar set is currently being made for the IPCC Fifth Assessment Report.

In addition, the development of such a diversity of simulators engenders a degree of rivalry and competition between institutes, that may be considered necessary to foster creativity. For example, there is kudos for the institute whose climate simulator is perceived by the community as "being the best", and having a "world leading" climate simulator can be considered a matter of national and institutional pride.

Maintaining such a diversity means there are relatively few opportunities to pool resources internationally, and thus to benefit from "economies of scale" when trying to improve these simulators. Hence, the funding needed to improve an Earth-system simulator must largely be found at the national level. As such, even if the investment for the supercomputing needed to make global climate projections at high spatial resolution is small compared with the global costs of mitigation and adaptation, the investment may indeed be significant compared with other national funding priorities, especially in (these) times of economic difficulty.

Hence, one is therefore forced to ask two questions. Notwithstanding the benefits discussed above, is this institutional-based framework unquestionably a good thing, and are the merits of the "gene pool" incontrovertible? If not, is there an alternative?

b) Determinism, Parametrisation and Scaling Symmetry

All climate simulators used in CMIP3 (and indeed CMIP5) have inherited a basic feature from early NWP code: determinism. At one level, this is hardly surprising: the underlying partial differential equations on which the simulators are based (eg the Navier-Stokes equations) are deterministic. However, the assumption of determinism in the computational code implies that representations of unresolved processes in such simulators are themselves deterministic. For example, in his recent essay on the need for improved parametrisation in atmospheric simulators, Jakob (2010) notes that, since many important processes in the atmosphere remain unresolved, "it is therefore necessary to represent those subgrid-scale processes as a function of the grid-scale variables." In mathematics, a function associates one quantity- the argument- with another quantity- the value- in the sense that exactly one value is associated with each argument. This characterises perfectly the conventional approach to parametrisation: the grid-scale variables determine precisely the grid-box tendency associated with the sub-grid processes.

The basis for determinism appears superficially solid. Since, unlike those poor economists, we mostly know our equations at the level of partial differential equations (though see comments about Earth-system complexity near the end of Section 4), we should therefore know them at the computational level too, at least at sufficiently high resolution. On top of this, improvements in deterministic parametrisations have increased the realism of comprehensive climate simulators enormously since the early days of Manabe and Wetherald, and this increase in realism has also lead to substantial gains in conventional deterministic skill in weather prediction (Simmons and Hollingsworth, 2002). Is there any reason to doubt that similar improvements lie just around the corner?

However, is the argument for determinism unassailable, and is it possible that the assumption of determinism at the computational level is actually holding back progress in the development of climate and weather simulators? Let us start by going back to basics. Although the atmosphere is a compressible multi-phase fluid and indeed a considerable part of its complexity arises from this, consider for simplicity an incompressible homogeneous fluid for which the Navier-Stokes equations can be written:

$$\rho\left(\frac{\partial u}{\partial t} + u.\nabla \mathrm{u}\right) = -\nabla p + \mu\nabla^2 u \qquad (3.1)$$

where u is fluid velocity, p pressure, ρ density and μ viscosity. These *ab initio* equations are solved numerically by truncating the equations using some finite grid, or other finite (eg spherical harmonic) basis. If we write $u(x,t) = \bar{u}(x,t) + u'(x,t)$ where the overbar denotes some Reynolds-average operator, which we assume here to be a gridbox mean, then the "Reynolds-averaged" form for the Navier Stokes equations above can be written (schematically) as:

$$\rho\left(\frac{\partial \bar{u}}{\partial t} + \bar{u}.\nabla \bar{\mathrm{u}}\right) = -\nabla \bar{p} + \mu\nabla^2 \bar{u} + E$$

The effect of unresolved sub-grid processes on the resolved scales are represented by the quadratic "Reynolds stresses" E written in component form

$$E_i = -\rho\nabla_{\mathrm{j}}\left(\overline{\mathrm{u}_{\mathrm{i}}'\mathrm{u}_{\mathrm{j}}'}\right)$$

Jakob's definition of parametrisation, applied to these Reynolds stresses, follows a long tradition in fluid dynamics, including luminaries such as Boussinesq, Prandtl, Smagorinsky (and many others), in trying to close the Reynolds-averaged equations by representing E as a deterministic function of the resolved scale variables:

$$E = P(\bar{u};\alpha)$$

and where α denotes a number of parameters which can be determined, in principle at least, by observations and/or theory.

However, a key symmetry of equation (3.1) is associated with scale invariance: if $u(x,t), p(x,t)$ is a solution to the Navier Stokes equations, so also is

$$u_\tau(x,t) = \tau^{-1/2} u\left(\frac{x}{\tau^{1/2}}, \frac{t}{\tau}\right) \quad p_\tau(x,t) = \tau^{-1} p\left(\frac{x}{\tau^{1/2}}, \frac{t}{\tau}\right)$$

for any $\tau > 0$ (Majda and Bertozzi, 2001)

Whilst we would not expect precise scale invariance of this sort to apply to the real atmosphere (not least because of latent-heating and other diabatic sources), the existence of such scaling symmetries in the underlying equations is consistent with observations of power-law structure in the atmosphere. Fig 1 reproduces the celebrated result of Nastrom and Gage (1985) showing an observational analysis of atmospheric kinetic energy as a function of horizontal scale (shown in terms of horizontal wavenumber k). This analysis draws attention to two separate power-law slopes, a *"-3"* slope at large scales and a *"-5/3*" slope" at smaller scales. The truncation scale of all weather forecast simulators, and a number of contemporary climate simulators, lies within the "*-5/3*" range. Similar power-law behavior has been seen in cloud data (Rossow and Cairns, 1995). Whilst there is some disagreement concerning the physical interpretation of these power laws (see eg Lindborg, 2007), broadly speaking it appears that the "*-3*" slope is indicative of quasi-two-dimensional flow dominated by rotation, whilst the "*-5/3*" slope is indicative of three-dimensional flow with substantial divergent motion (enhanced by latent heat release in cloud systems, associated with the compressible multi-phase nature of the atmosphere).

As first clearly pointed out by Schertzer and Lovejoy (1993), the "deterministic truncation/parametrisation ansatz" outlined above, is inconsistent with the existence of scaling symmetries and associated power-law behaviour – for the simple reason that such power laws preclude any meaningful separation between "resolved" and "unresolved" scales, and hence between "resolved" and "unresolved" processes. Possibly consistent with this, it can be noted that some simulators, eg that of ECMWF have difficulty simulating the "-5/3" spectrum, even at relatively high truncation scales of 10km (Straus, 2011, using data from integrations performed as part of the Athena project; Jung et al, 2011: Kinter et al, 2011).

It can be argued that the failure of deterministic parametrisations to represent this observed power law structure is the fundamental cause of systematic model error. For example, in the IPCC AR4 it is concluded (Solomon et al, 2007; Chapter 8):

"…models still show significant errors. Although these are generally greater at smaller scales, important large-scale problems also remain. ……The ultimate source of most such errors is that many important small-scale processes cannot be represented explicitly in models, and so must be included in approximate form as they interact with larger-scale features. …consequently models continue to display a substantial range of global temperature change in response to specified greenhouse gas forcing. "

Perhaps one could argue that with fine-enough simulator resolution (eg T2047, much higher than any contemporary climate simulator), large-scale errors associated with any violation of power-law behavior can be made arbitrarily small. A simple scaling argument (Lilly, 1973, see also Palmer, 2001) indicates that this is not a reliable conclusion. Let $E(k)$ denote atmospheric kinetic energy per unit wavenumber, at wavenumber k. We can define a timescale $\tau(k)$ in terms of a length divided by a velocity ie $\tau(k) \sim k^{-\frac{3}{2}} E^{-\frac{1}{2}}(k)$. Let us suppose $\tau(k)$ characterises the time it takes errors at wavenumber k to grow and infect nonlinearly the accuracy of simulations at wavenumber $k/2$. As above, suppose we are only interested in large-scale aspects of the flow, ie wavenumbers less than some k_L. We can ask how long it will take before truncation errors at large wavenumbers $2^N k_L,\ N \gg 1$ will affect large-scale simulations of the flow. A plausible estimate of this is given by:

$$\Omega(N) = \tau(2^N k_L) + \tau(2^{N-1} k_L) + \ldots\ \tau(2^0 k_L) = \sum_{n=0}^{N} \tau(2^n k_L)$$

Now if $E(k) \sim k^{-3}$ then $\tau(k)$ is independent of k and $\Omega(N)$ diverges as $N \to \infty$. This suggests that if the atmosphere was quasi two-dimensional all the way down to very small scales, errors at small scales could be "shielded" from the large scales, by increasing the simulator resolution sufficiently. However, if $E(k) \sim k^{-5/3}$ then $\tau(k) \sim k^{-2/3}$ and $\Omega(N) \sim 2.7\tau(k_L)$. There is nothing especially significant about the precise value 2.7. Hence, let us say that with a -5/3 power law, the series $\Omega(N)$ converges to a value less than a few "eddy turn-over times" of k_L, as $N \to \infty$. Hence, with a "-5/3" power law, it may be impossible to shield the large scales from truncation-scale errors, by increasing sufficiently the resolution of the simulator. This analysis is consistent with the study of Lorenz (1969), see also the more robust analysis of Rotunno and Snyder (2008) using the surface quasi-geostrophic equations, but has not been proven rigorously from the underlying 3D Navier Stokes equations. (It is not literally true in the limit where $2^N k_L \sim k_V$

and k_V lies in the viscous range of scales; however, it appears to be an open question asymptotically in the range $k_L \ll 2^N k_L \ll k_V$.) It is worth commenting that the predictability estimates above do not depend on the mechanism by which the -5/3 power law is established.

Despite this, there are very good reasons for attempting to increase the resolution of atmospheric simulators as much as possible. Firstly, the higher a simulator's resolution the better Earth's topography and land/sea boundary can be represented. Secondly, high resolution ensures that Rossby wave breaking, important for the maintenance of blocking anticyclones and other nonlinear weather-regime phenomena (see Section 7), can be simulated properly. Thirdly, the higher the resolution, the better the simulator can utilise high-resolution observations, eg from satellite instruments with small pixel size. Finally, at some stage, high-resolution simulators will be capable of representing the key atmospheric phenomenon of deep convection (which, along with baroclinic instability, can be considered one of the core dynamical modes of atmospheric instability and hence variability). Similar arguments apply to the oceans too. In addition to these theoretical considerations, regional predictions of climate change, particularly for precipitation change, have been shown to be sensitive to changes in resolution, horizontal and vertical (Matsueda and Palmer, 2011; Scaife et al, 2011).

However, a plausible consequence of the analysis above is that as the truncation scale of a climate simulator moves into *"-5/3"* range, the effects of the inconsistency of using deterministic parametrisation cannot be reduced to zero by increasing resolution sufficiently (building a comprehensive climate simulator whose truncation scale lies in the viscous range is utterly impracticable in the foreseeable future). By this, it is not to be inferred that the effect of misrepresenting the small scales will damage the larger scales uniformly in time; that very pessimistic scenario is inconsistent with the fact that conventional weather prediction simulators can, from time to time at least, predict large scales very accurately, well beyond the limit $\Omega(N) \sim 2.7\tau(k_L)$. That is to say, experience suggests that the rapid upscale error propagation associated with the "-5/3" spectrum will occur somewhat intermittently (for example the source of some especially erroneous medium-range weather forecasts over Europe have been traced to short-range forecast errors associated with intense mesoscale convection over the US Mid-West). This raises the fundamental question: How can we ensure that the advantages of integrating simulators at higher and higher resolution will not be somehow be destroyed by rapid intermittent upscale propagation of error?

As suggested by this analysis, contemporary simulators may have common failings due the universal use of the deterministic truncation/parametrisation ansatz. This implies that multi-simulator ensembles may be blind to the consequences of such systemic failings, so that ensemble agreement cannot be assumed a reliable measure of forecast confidence. Is there any evidence for this?

There is some evidence from the poorness of the "attributes curve" in reliability diagrams (Wilks 2006) from seasonal forecasts of regional precipitation based on DEMETER multi-simulator ensembles (Palmer et al, 2008). An attributes curve can assess whether, for a particular forecast event E, forecast probabilities of E are well calibrated against observed frequencies of E – the technical definition of "reliability". The attributes curve for a reliable forecast system should lie on the diagonal. Fig 2 shows an update of such seasonal-forecast reliability diagrams but based on the more recent ENSEMBLES multi-simulator ensemble (Weisheimer et al, 2009). Fig 2 shows examples (for seasonal-mean Sahel and North European rainfall) where the flatness of the attributes curves indicates that the ensemble is extremely overconfident and hence highly unreliable. The origin of such unreliability is, most likely, an inadequate representation of simulator error in the multi-simulator ensemble (the author is unaware of any systematic misrepresentation of observational uncertainty that would lead to such unreliability).

As discussed in Palmer et al (2008), some of the unreliability of seasonal forecasts arises from difficulties which climate simulators have in simulating the statistics of weather regimes (Straus et al, 2007). For example, ability to simulate anticyclonic blocking accurately is a well-known problem amongst low-resolution climate simulators. However, recent results from the Athena project (Kinter et al, 2011; Jung et al, 2011) suggest even at higher resolutions, climate simulators may have difficulty replicating the multi-modal probability distributions of regional weather regimes (Andrew Dawson, personal communication 2011) even though such multi-modality is highly significant when diagnosed from reanalysis datasets. As discussed in Section 7, it is suggested that an ability to simulate regional weather regimes accurately will be key to reducing uncertainty in the cloud feedback problem for predicting global climate change.

In a recent paper, Doblas-Reyes et al (2011) concluded that the dominance of simulator bias in state-of-the-art coupled ocean-atmosphere simulators is a major impediment to the investigation of decadal timescale predictability in particular in assessing whether useful decadal predictions can be made, given our current ability to observe the sub-surface ocean. Two key points can be made here. Firstly, one of the goals of the emerging programme of "climate services", that of providing reliable near-term climate forecast information to a range of customers, is not likely to be met by current-generation simulators. Secondly, the value of investment in ocean (and other) observations is not being fully realised because of simulator bias. This in turn raises the following point. There have been many discussions in the community about the relative importance of funding Earth observations, *vis a vis* climate simulator development. However, this is a false dichotomy; in truth, we will only realise the full value of investment in Earth observations when climate and weather simulators are of sufficient quality to be able to ingest and utilise these observations fully (either in analysis/reanalysis mode, or in predictive mode). If the information content in an observation is being lost prematurely due to simulator bias, then the investment in producing this observation will not have been fully realised.

c) *True diversity of the "gene pool" of climate simulators*

Given the problems above, it is worth asking just how diverse is our "gene pool" of climate simulators really is. Many climate institutes use the same basic closures in their simulators' parametrisations, indeed some share the same parametrisations. Estimating the effective size, M_{eff} , of the CMIP3 multi-simulator ensemble has recently been studied by Pennell and Reichler (2011) who note that "for the full [CMIP3] 24-member ensemble, this leads to an M_{eff} that…lies only between 7.5 and 9." They conclude: "The strong similarities in model error structures found in our study indicate a considerable lack of model diversity. It is reasonable to suspect that such model similarities translate into a limited range of climate change projections."

Hence, possibly related to the systemic problems discussed above, the effective size of the gene-pool is rather small: many of the institutional simulators whose integrations are submitted to CMIP, are relatively minor modifications of a small number of core simulators.

There are techniques to expand ensemble size by perturbing the parameters α within a given simulator, according to expert opinion about inherent uncertainty in fixing the values of these parameters (Murphy et al, 2004: Stainforth et al, 2005; see also Section 4 below). Whilst there is certainly merit in treating these parameters as uncertain and representing this uncertainty in "perturbed-parameter" ensembles, evidence to date suggests that adding perturbed-parameter integrations to a multi-simulator ensemble does not change M_{eff} by much (Masson and Knutti, 2011).

d) *Climate complexity*

Notwithstanding the remarks above, there are two fundamental problems that all climate institutes acknowledge as obstacles to the development of accurate climate simulators: insufficient human resources and insufficient computing resources. These problems are especially acute in (current) economically challenged times.

Since the days of Phillips, and Manabe and Wetherald, climate simulators have become more and more complex. In terms of parametrisations, the sub-grid representations for deep convection, clear-sky and cloud radiative effects, sub-grid orography, boundary-layer turbulence, aerosols, cloud microphysics, and so on and so forth, have become immeasurably more sophisticated (and computationally demanding) since the early days. Moreover, what in the 1970s were essentially atmosphere-only simulators (eg with simple "slab" oceans and "bucket" land-surface hydrology) have in the 2010s become fully coupled representations of the atmosphere, oceans, cryosphere and land surface with a range of biogeochemical processes

("Earth-system complexity"). The need to ensure that chemical tracers are properly represented during simulations, yet at the same time allowing the simulators to run efficiently on massively parallel computers, means that the numerics of the dynamical cores of weather and climate simulators have to be extremely sophisticated.

Problems of algorithmic complexity do not stop there. For climate-service applications, shorter-range decadal predictions require that simulators are initialised with contemporary observations, implying the need for sophisticated data assimilation schemes for the atmosphere, oceans and land surface.

Finally, the dynamical cores themselves are increasingly complex as quasi-geostrophic equations have given rise to the hydrostatic primitive equations, and now to the non-hydrostatic dynamical cores, needed to be able to probe kilometre truncation scales where deep convection is at least partially resolved. At these high resolutions, it is a highly nontrivial problem to ensure that numerical code can run efficiently over the very large numbers of processors of modern supercomputers (the scalability problem).

Not surprisingly then, climate institutes struggle to find the human resources needed to develop these manifold elements. On top of this, the computational demands of a contemporary climate simulator means it is impossible for an institute to develop simulators both with full Earth-system complexity and with the resolution of a contemporary NWP simulator, and at the same time run large ensemble integrations from states initialised with contemporary observations. This extremely important issue has been discussed at length elsewhere, being a key topic of the major World Summit on Climate Modelling (Shukla et al,2010; Palmer, 2011).

In the next sections, we discuss a relatively new approach to the representation of unresolved processes in weather and climate simulators, which may provide a solution to the complex and challenging problems outlined in this Section.

3. Stochastic Representation of Unresolved Processes

Let us begin by considering a generalisation of the definition of what we mean by "parametrisation" and frame it, not in terms of functional relationships, but as a constraint on some prior (eg climatological) probability distribution of sub-grid tendency based on a knowledge of contemporaneous values of grid-scale variables. An explicit example will be given below. This automatically suggests we treat the notion of parametrisation as an inherently probabilistic problem, to be tackled by explicitly stochastic techniques (Palmer, 2001).

There is nothing new in the use of stochastic mathematics to describe climate simulators; the idea can be traced to Hasselman (1976) who developed an idealised coupled ocean-atmosphere simulator in which the entire atmosphere was represented by a simple Markov process. Using this simulator, Hasselman showed how ocean-atmosphere coupling would redden the spectrum of atmospheric variability. However, the use of stochastic mathematics in such earlier approaches, is conceptually different to the concept being explored here: Hassleman's simulator is (deliberately) a simplified idealised representation of climate, and the use of stochastic mathematics made the representation of internal atmospheric variability in the simulator equations mathematically tractable. Here, we are not interested in mathematical tractability *per se.* Rather it is being argued that stochastic mathematics also has an inherent role to play in comprehensive *ab initio* weather and climate simulators.

A key conceptual difference between deterministic and stochastic parametrisation is illustrated in Fig 3. Whilst deterministic parametrisation represents the bulk-average effect of some putative large ensemble of sub-grid processes occurring on scales smaller than the grid scale, stochastic parametrisation attempts to represent actual realisations of the sub-grid flow when no scale separation exists. Fig 3 indicates that the stochastic parametrisations must necessarily impact directly on scales larger than the truncation scale. This is because, as discussed above, with power-law behavior uncertainty in sub-grid processes will propagate upscale by nonlinear dynamical effects (Thuburn et al, 2011). Hence part of the (stochastic) parametrisation process requires one to represent the effect of uncertainty in the sub-grid processes on the resolved grid.

In order to quantify the potential benefits of this stochastic approach to parametrisation, it is useful to consider a reasonably tractable example where we know precisely the "true" system, which we will attempt to simulate approximately using parametrisations, both deterministic and stochastic. Consider, then, the set of linked nonlinear ordinary differential equations put forward by Lorenz (1996):

$$\frac{dX_k}{dt} = -X_{k-1}\left(X_{k-2}-X_{k+1}\right) - X_k + F - \frac{hc}{b}\sum_{j=J(k-1)+1}^{kJ} Y_j \qquad (4.1)$$

$$\frac{dY_j}{dt} = -cbY_{j+1}\left(Y_{j+2}-Y_{j-1}\right) - cY_j + \frac{hc}{b}X_{\mathrm{int}[(j-1)/J]+1} \qquad (4.\ 2)$$

Here the X_k represent the large, slow scales (analogous to wavenumbers $\leq k_L$) that we are interested in, and the Y_j represent the small, fast scales (analagous to wavenumbers $\geq 2^N k_L$) that we wish to parametrise. Here $1 \leq j \leq 32$, and k is cyclic mod 8. The last term of the first equation couples the small scales to the large scales; we will call this "the small-scale tendency". Below we consider two values of the *c* parameter: *c=10* and *c=4;* the *h, b* and F parameters are held fixed. When *c=10*, the *Y* variables typically evolve over substantially faster timescales than do the *X* variables, ie there is clear temporal scale separation between these variables. It will turn out that parametrising the small-scales deterministically will work reasonably well for this parameter setting. By contrast when *c=4*, this scale separation is weaker and the parametrisation problem becomes inherently less deterministic. By way of analogy, then, we use the values *c=10* to mimic the relatively steep "-3" energy spectrum, and *c*=4 to represent the relatively shallow "-5/3" energy spectrum of the real atmosphere.

With the true system represented by equations (4.1) and (4.2), we now consider a simulator

$$\frac{dX_k}{dt} = -X_{k-1}\left(X_{k-2} - X_{k+1}\right) - X_k + F - P_k$$

$$P_k = (1 + r_k^{mult})\, P_k^{det}(X_k;\alpha) + r_k^{add}$$

of the "true" Lorenz (1996) system, where the small-scale tendency is parametrised by the formulae P_k (first discussed by Wilks, 2005). Here we have generalized the conventional deterministic formula $P_k = P_k^{det}(X_k;\alpha)$ using stochastic variables r_k^{add} and r_k^{mult}. A number of parametrisations are considered: "Deterministic" denotes a deterministic parametrisation ($r_k^{add} = r_k^{mult} = 0$) based on fitting a cubic polynomial in X_k to points in a scatter diagram of instantaneous small-scale tendency against X_k; "White Additive" denotes a simple white-noise term added to the deterministic parametrisation ($r_k^{add} \neq 0; r_k^{mult} = 0$); "Red Additive" denotes a red-noise AR1 process added to the deterministic parametrisation; "Multiplicative" denotes a red-noise AR1 process multiplying the tendencies from the deterministic parametrisation ($r_k^{add} = 0; r_k^{mult} \neq 0$).

We can use this system to illustrate the utility of the probabilistic notion of parametrisation as defined earlier in this section. Fig 4 shows (solid curve) the unconstrained (ie climatological) probability distribution of the small-scale tendency term, the last term on the right hand side of equation (4.1) when *c*=4. On this figure is plotted the probability distribution of this tendency when the X_k variable is constrained to lie in $-6 \leq X_k \leq -5$ (dotted line) and $13 \leq X_k \leq 14$ (dashed line). It can be seen that the constrained probability distributions are quite different from the climatological distribution. That is, knowledge of the large-scale variable is important in constraining the prior distribution. However, this knowledge does not constrain the distribution so much that it collapses to a Dirac delta function – which would be the case if deterministic parametrisation were accurate. Corresponding hat functions for the putative deterministic parametrisation, for $-6 \leq X_k \leq -5$ and $13 \leq X_k \leq 14$, are shown in Fig 4 for *c*=4; compared with the constrained probability distributions, these hat functions are quite obviously too sharp. As such, it can be expected that the simulator with deterministic parametrisation will perform relatively poorly. Fig 4 also shows that the probability distributions are sharper for small deterministic tendency suggesting that the simulator with multiplicative noise parametrisation may be especially skilful.

Fig 5 shows skill score results for a large number of initial-value ensemble predictions (Fig 5a) and one long climate integration (Fig 5b). Full details are given Arnold (2011). In the initial-value ensembles, evaluated at t=0.6 time units (perhaps equivalent to about 3 days for weather forecasting), the initial conditions $X_k(t=0)$ are known perfectly, hence there is no initial uncertainty, only simulator uncertainty. The solid line denotes the results with c=10, the dashed line gives results with c=4.For the initial-value problem, we use the Ranked Probability Skill Score (RPSS: Wilks, 2006) to assess the probabilistic skill in forecasting X_k. For the climate integrations, we use the Hellinger Distance (related to the more familiar Kolmogorov-Smirnov Distance; Pollard, 2002) between the "true" and simulated probability distribution of X_k values. Note that the larger the RPSS, the more skilful the forecast, whereas the smaller the Hellinger Distance, the closer is the simulated probability distribution to the probability distribution of truth. Again, see Arnold et al (2011) for details. Additionally, for the initial-value ensembles (Fig 5a for c=4) we also show the traditional deterministic score, root-mean-square (RMS) error, averaged over all the individual forecasts.

A number of interesting results can be concluded from Fig 5:

1) Based on RPSS and Hellinger Distance, and as expected, the *c=10* system is "easier" to parametrise than the *c=4* system, and whilst stochastic parametrisation improves forecast skill for both values of *c*, the improvement is relatively small when *c=10*. By analogy, we would expect comprehensive weather simulators to be harder to parametrise deterministically, if their truncation scales probe the -5/3 part of the spectrum. As discussed above, there is an inherent tension (perhaps one would even say incompatibility) between high-resolution simulation and deterministic parametrisation.

2) Based on RPSS and Hellinger Distance, there is an overall strong correlation between simulator performance in initial value mode, and in climate mode, consistent with the philosophy underpinning the notion of seamless prediction. That is to say, the performance of the simulator in climate mode can be gauged by its success in initial-value mode. Of course, in the real world, one would not expect a one-to-one correspondence between weather and climate skill, because there are many slow climate process which are not important for weather prediction. Nevertheless, the results here hint that skill on the weather timescale should be considered a necessary step for reliable climate prediction.

3) The link between initial-value skill and climate accuracy is only apparent when probabilistic measures of skill are used to assess the initial value ensembles. If the more traditional deterministic RMS error metric is used to assess initial value skill, there is no correlation between initial value skill and climate skill; indeed the simulator with deterministic parametrisation appears most "skilful". As discussed in Section 6, the conclusion to draw from this result is not that the link between weather and climate skill is metric dependent, but rather that the RMS error may actually be an inappropriate metric of weather forecast skill. The physical reason for this is discussed in Section 6 where it is concluded that assessing simulators based on weather-forecast RMS error may in fact be detrimental to the development of reliable climate forecast systems.

4) Based on RPSS and Hellinger distance, there is an overall advantage for the red noise parametrisation over the white noise parametrisation. This is consistent with the discussion above: in stochastic parametrisation, it is necessary to represent the means by which uncertainty in the representation of sub-grid processes affects the large-scale flow, on spatial scales larger than the simulator's truncation scale, and on timescales longer than the simulator's timestep. In Lorenz (1996), correlations between neighbouring X_k variables are small, and, for this particular model, there is not much benefit to the introduction of "spatially-correlated" noise. However, as Fig 5 shows, there is benefit in representing "temporally-correlated" noise. In general, for weather and climate simulators, one would expect the noise to be both spatially and temporally correlated

5) There is an overall advantage for the multiplicative noise parametrisation. This multiplicative noise parametrisation is essentially that developed and tested in the ECMWF simulator by Buizza et al (1999).

In the latest version of the ECMWF multiplicative-noise scheme (or SPPT: Stochastically Perturbed Parametrisation Tendency scheme, see Palmer et al; 2009), the parametrisation is given by

$$\dot{\mathrm{X}}^{\mathrm{stoch}} = (\dot{1} + r^{spec}\mu)\dot{\mathrm{X}}^{\mathrm{det}}$$

where $\dot{X}^{stoch}$ denotes the stochastic tendency, $\dot{X}^{det}$ the total deterministic tendency, r^{spec} denotes a stochastic spectral pattern generator based on an uncorrelated series of red-noise processes, one for each spherical harmonic coefficient. The relative amplitude of these red noise processes in spectral space is such as to produce Gaussian correlations in physical space (see Fig 6). In the results discussed below, there are two sets of such red-noise processes: one with 6 hour decorrelation time, the other with smaller amplitude and 30-day decorrelation time (see Palmer et al, 2009 for details). Finally, μ is an *ad hoc* parameter which clips the stochastic tendencies in the stratosphere and in the boundary layer. We return to this "μ" parameter later.

A more overt example of the need to consider the representation of sub-grid uncertainty on the resolved spatial scales arises in the stochastic backscatter scheme (Shutts, 2005; Berner et al, 2009)

$$F_{\psi} = \left(\frac{b_R D_{tot}}{B_{tot}}\right)^{1/2} P_{\psi_*}$$

Here the streamfunction forcing F_{ψ} is associated with an upscale energy transfer when, for example, divergent kinetic energy associated with deep convection is converted to rotational kinetic energy during mesoscale organization. This forcing is represented by a stochastic pattern generator P_{ψ} (either the spectral generator, cf Fig 6, or an alternative cellular automaton – it can be noted in passing that cellular automata provide computationally cheap means to communicate information at the sub-grid level, between adjacent grid boxes). Here D_{tot} denotes the diagnosed energy dissipation from the corresponding deterministic parametrisations, and B_{tot} and b_R are parameters which ensure dimensional consistency and degree of energy backscatter respectively.

Fig 7 (from Palmer et al, 2009) shows the impact of SPPT on the probabilistic skill of medium-range forecasts of 850hPa temperature in the tropics using the ECMWF Ensemble Prediction System (EPS). The results are dramatic. The skill at day two of the probabilistic forecasts without stochastic parametrisation, is reached at day six with stochastic parametrisation. It is hard to imagine any parametrisation having such an effect on forecast skill.

The introduction of stochastic parametrisation into the ECMWF simulator has fundamentally changed the skill of the EPS in more ways than one. Importantly, it has allowed the estimation of initial uncertainty to be made using ensembles of (4D Var) data assimilations (EDA: Isaksen et al, 2010). Until recently, EPS initial perturbations were made exclusively using singular vector analysis (eg Buizza and Palmer, 1995). The reason for this was that if an EPS was based solely on initial perturbations from ensembles of analyses, these perturbations had to be artificially inflated in order that EPS spread and skill matched in the medium range. Introduction of stochastic parametrisation into the data assimilation process (and the use of higher resolution and hence less damped simulators), has enabled ensemble data assimilation to be used to generate initial EPS perturbations. Indeed Fig 8 shows the performance of EDA in terms of the relationship between ensemble spread at T+6hrs, and ensemble mean error. It can be seen that with representation of observation error only, not only is the EDA underdispersive, but also the EDA spread does not discriminate well between low error and high error short-range forecasts (it is particularly underdispersive for high error forecasts). By contrast, including SPPT and backscatter into EDA, not only is the overall level of spread much closer to that of error, the EDA spread now discriminates well between low and high error short-range forecasts.

Fig 9 shows that EPS-based probabilistic predictions of rainfall over Europe in the medium range are now extremely reliable.

There is no doubt that ensemble forecasts with stochastic parametrisation are skilful. But are they more skilful than forecasts using the more traditional multi-simulator concept? This question, applied to the climate prediction problem, lies at the heart of this paper. Table 1 shows a comparison of probabilistic skill on the monthly timescale (where copious verification data exists), based on three ensemble forecast systems (see Weisheimer et al, 2011 for details). The first system is the single-simulator ECMWF seasonal ensemble forecast system with stochastic (SPPT and backscatter) parametrisation. The second system is a multi-simulator ensemble comprising the climate simulators that contributed to the ENSEMBLES multi-simulator ensemble (Weisheimer et al, 2009). The third ensemble is again based on the single-simulator ECMWF seasonal ensemble forecast system as above, but with no representation of simulator uncertainty (ie only initial uncertainty).

Results show that for 7 of the 8 binary forecast events considered (based on climatological temperature and precipitation terciles over all land points), the single-simulator ensemble with stochastic parametrisation outperforms the multi-simulator ensemble. For one of the 8 events, the skill estimates for the stochastic parametrisation ensemble and the multi-simulator ensemble only differ by the third significant digit. It might be imagined that the key reason that the single-simulator stochastic-parametrisation ensemble outperforms the multi-simulator ensemble is that the former has been made with a world-leading weather simulator. However, if we compare the skill of the multi-simulator ensemble with the skill of the same single-simulator ensemble without any representation of simulator uncertainty, then it can be seen from Table 1 that the latter is much the least skilful of the three ensembles for all events considered.

This indicates that the single-simulator stochastic parametrisation ensemble is not more skilful that the multi-simulator ensemble because this particular simulator is somehow inherently better (eg in terms of its deterministic forecast skill) than the other simulators.

In Weisheimer et al (2011), it was also shown that on longer seasonal timescales, stochastic parametrisation still has the edge against the multi-simulator ensemble for precipitation forecasts, but not for forecasts of surface temperature. This suggests (see Section 4 below) that development of the stochastic approach for the land surface and for the oceans is also likely to be required in the future. The skill of a perturbed-parameter ensemble was also tested by Weisheimer et al (2011). The skill scores turned out to be poor, but one cannot rule out the possibility that this was because the simulator in which the parameters were perturbed was not state-of-the-art for monthly and seasonal prediction. Further tests are needed within, eg the ECMWF system, to evaluate the perturbed-parameter method. It is certainly not inconveivable that some combination of perturbed-parameter and stochastic parametrisation techniques may prove optimal.

A key property of stochastic parametrisation is its potential ability to influence the mean state of the simulator and hence reduce the mean bias of the simulator against observations. That is to say, the interaction of the imposed noise with the nonlinearity of the simulator can generate a "rectified" time-mean response. In this way, it is possible that stochastic parametrisation can help alleviate some of the systematic biases of climate simulators. Fig 10 shows an example of such alleviation (from Berner et al, 2011, who also show a positive impact of stochastic backscatter on the mean state of simulations in the tropics).

However, a problem revealed by Fig 10 is that the impact of stochastic parametrisation on simulations of Northern Hemisphere circulation is very similar to the impact of either increasing simulator resolution (ie modifying the dynamical core), or modifying the conventional deterministic parametrisation schemes. Dynamical reasons for this "degeneracy" are discussed in Palmer and Weisheimer (2011). These explain why improving the fidelity of climate simulators has been so difficult over the years, and why it is very easy for a simulator code to contain many sets of "compensating errors". This is a key reason why data assimilation can provide such a powerful tool for enabling simulator development whilst minimizing such compensating-error problems (see Palmer and Weisheimer, 2011 for discussion). This problem of degeneracy is discussed further in Section 4 below.

4. Stochastic parametrisation at the process level

Despite these rather positive results, stochastic parametrisation is still at a rudimentary state of development: the stochastic parametrisation concept described above has only been applied to the atmospheric component of coupled simulators. There is clearly a need to extend the concept to the oceans, the land surface, the cryosphere, the biosphere and so on. The techniques which can be used to develop stochastic parametrisations are manifold, and the logic inductive rather than deductive. A technique of particular relevance is the type of coarse-grain analysis developed in Frederiksen and Kepert (2006) and Shutts and Palmer (2007). Moreover, the sort of experimental programmes advocated by Jakob (2010) are just as important for the development of stochastic parametrisation as for deterministic.

However, even for the atmospheric component of climate simulators, there is a need for uncertainty to be incorporated in the development of parametrisation at the process level, rather than as a “bolt-on extra”. For example, in describing the multiplicative noise parametrisation in Section 3, reference was made to the *ad hoc* parameter μ which clipped the stochastic noise both in the boundary layer and in the stratosphere. The parameter was introduced for plausible reasons, but also because it improved forecast scores. However, one should not introduce parameters purely because of empirical pragmatism: they must additionally have some basis in science. For the stratosphere, the scientific basis is not hard to find. Much of the diabatic heating in the stratosphere is associated with infra-red radiation emitted by carbon dioxide molecules. However, unlike water, carbon dioxide is well mixed in the atmosphere, there is little sub-grid variability. Hence there is no need to represent this process stochastically. It is also conceivable that, at least in sufficiently homogeneous terrain well away from orography, a typical boundary layer “eddy” associated with surface form drag is also sufficiently small in scale that grid scale stochasticity in grid-scale vertical mixing will be relatively small. This argues that, instead of having an *ad hoc* μ parameter, aspects of stochastic parametrisation should be developed at the process level.

The case for stochastic parametrisation at the process level is fairly clear when discussing processes like convection (eg Lin and Neelin, 2003; Plant and Craig, 2008), and imaginative new stochastic schemes for parametrising different convective cloud families are being developed using cellular automata (eg Bengtsson-Sedlar et al, 2011) or stochastic lattice models (Khuider et al, 2003: Frenkel et al, 2011). However, even something as basic (and in principle well known) as radiation needs stochastic treatment; gridbox surface radiative fluxes can depend strongly on poorly resolved near-grid scale circulations. For example, under a region of stratocumulus, surface fluxes will depend strongly on whether in-cloud shallow convection is of the closed cell or open cell type. It is unrealistic to expect these small-scale circulations to be deterministic functions of the large-scale variables; such effects therefore represent a source of uncertainty in forecasts of surface temperature that should be incorporated at the process level into the simulator equations.

However, there is a separate and quite fundamental argument for the need to develop stochastic parametrisation as an inherent part of simulator development, and not as a “bolt-on” extra. In Section 3, it was shown that stochastic parametrisation had an impact on a simulator’s systematic error. Consider the implications of this for setting the parameters α of the deterministic parametrisations $P(X_k, \alpha)$.

For example, the parameter often called “convective entrainment” represents the strength of the process whereby environmental air is entrained laterally into convective plumes. It is well know that climate simulations can be especially sensitive to the value of this parameter

(Stainforth et al, 2005). However, if the notion of a sub-grid ensemble of convective plumes is not well founded due to power-law structure and associated scale invariance, then neither is the existence of a well-defined value for the convective entrainment parameter. As such, and this is universally recognised by the scientists who develop climate simulators, the values of these parameters have to some extent to be "tuned" based on the fit of simulator output to sets of observations of the large-scale structure of the atmosphere (either based on weather forecasts or climate integrations).

However, consider the implications of such tuning exercises if "bolt-on" stochastic parametrisations change the mean state of the simulator. It implies that values of the parameters α_{det} which have been optimally tuned for a deterministic simulator will not be optimal in a stochastic simulator. This implies that α_{det} are not in fact optimal at all.

This situation is familiar in many other areas of physics. Consider the vertical motion of a table-tennis ball with mass m_0 inside water. As found by Green in the 19th Century, the motion of the ball obeys Newton's law $F = ma$ where F is the Archimidean buoyancy force, but where $m = m_0 + M/2$ and M is the mass of water occupying the same volume as the table-tennis ball. In other words, whilst in the absence of randomly fluctuating molecules of water the motion of the ball obeys $F = m_0 a$, in the presence of these random fluctuations the motion of the table-tennis ball behaves as if it were half full of water. This "renormalisation" of mass has measurable effects: the initial acceleration of the ball back towards the surface is about seven times smaller than it would otherwise be. Quantum field theorists recognise in these arguments the difference between the "bare" and "effective" mass of a particle such as an electron in the presence of the fluctuating photonic field.

In the same way, we argue here (as indeed was argued by Frederiksen and Davies, 1997; Frederisksen, 1999) that parameter tuning for weather and climate simulators must be done in the presence of parametrised representations of the inherent stochasticity associated with the scale-invariant properties of the underlying equations. The notion of "bolt-on" stochastic parametrisation (for use when the simulator is run in probabilistic ensemble mode) using deterministically pre-tuned parameter values, is not a scientifically sound procedure.

The focus of attention in this paper has largely been on the parametrisation of physical processes. It has been argued that even when we know the underpinning equations with accuracy, the resulting parametrisations should be considered stochastic. However, for parametrising other processes (eg biological, chemical and perhaps aerosol), where the underpinning equations are not known with accuracy, the need for stochastic representations is no less important and necessary. At the very least, if different deterministic closures $\{A_P, B_P, \cdots C_P\}$ have been proposed for process *P*, and observations cannot rule out any one of these closures over another, a particular closures chould be chosen randomly at a given grid box and time step, using the types of spatially and temporally correlated pattern generators discussed above.

It is important to emphasise that none of the above implies that we must not continue to develop, refine, improve and extend our subgrid parametrisations. This work remains as critical in the future as it has been in the past. However, it is argued here that this development, refinement, improvement and extension, should be performed within an inherently probabilistic, and hence stochastic, parametrisation framework. The author believes that research and development within this more general framework will allow innovative ideas to flourish and parametrisation breakthroughs to occur.

5. Dynamical cores and stochastic processors

It was noted above that stochastic parametrisation inevitably involves the representation of the upscale propagation of subgrid uncertainty onto the resolved grid. This suggests that, just as it may be futile trying to develop precise deterministic parametrisations, so also it is futile to develop precise deterministic dynamical cores, especially for the evolution of scales near the truncation scale.

Whilst this may be the case, if we have accounted statisically for upscale propagation of uncertainty in the parametrisation, then the only disadvantage to retaining a precise deterministic dynamical core is the computational burden. However, how would one go about defining a probabilistic dynamical core which is both consistent with the equations of motion, and would provide a significant reduction in computational cost compared with current deterministic cores? After all, computing a stochastic field which comprises large numbers of pseudo random-number generators is certainly not computationally cheap.

However, there is an emerging technology that may present a way forward here and at the same time provide a new type of synergy between software development (of the high-resolution probabilistic Earth-system simulator) and the very hardware needed to integrate a simulator's equations. This technology (eg Palem, 2005) is motivated by the fact that a significant fraction of a conventional computer's energy consumption is associated with heat dissipation at the chip level. Hence, if the processors of a computer could be designed so that when the voltage across the individual transistors is reduced, the computer would operate with significantly reduced energy consumption but at marginally reduced (eg 99% instead of 100%) accuracy, this capability would certainly be worth exploiting.

Indeed the issue that bit-reproducible computation may become a thing of the past is beginning to be recognised in the supercomputing industry too. In a recent presentation on challenges in application scaling in an exascale environment, IBM's Chief Engineer noted (http://www.ecmwf.int/newsevents/meetings/workshops/2010/high_performance_computing

_14th/index.html) that increasingly there will be “a tension between energy efficiency and error detection”, and asked whether there needs to be a new software construct which identifies critical sections of code where the right answer must be produced – implying that outside these critical sections errors can (in some probabilistic sense) be tolerated.

One can perhaps imagine a future energy-efficient computer with clusters of processors each with different levels of accuracy, integrating a future-generation dynamical core. The more accurate the processor, the larger the scales of motion for which it computes tendencies. The “right answers” will be produced only for the large-scale tendencies. Since, overall, computations are dominated by estimation of tendencies nearer the truncation scale, the synergistically designed probabilistic supercomputer need have relatively few of these slower energy-intensive processors. On top of this, the stochastic parametrisations themselves would be computed using the energy-efficient probabilistic chips.

There is a link here to the work of Lander and Hoskins (1997) who argued that sophisticated and computationally expensive parametrisation schemes should only be applied to the more “believable” scales in a simulator, ie scales far removed from the truncation scale. They propose that simpler parametrisation schemes could be used on the “unbelievable” scales near the truncation scale. This idea has some resonance with the proposal discussed here whereby less-believable computations near the truncation scale could be executed on relatively fast energy-efficient probabilistic processors, leaving computations at the large “believable” scales for traditional energy-intensive bit-reproducible processors.

What would prevent utterly erroneous computations from compromising the validity of the computation (eg due to big errors in the exponents of key real numbers representing the small scales)-?. It could be preventedpartly by performing numerical checks against prior physical bounds, and partly by repeating computations several times, and taking the mode of some small ensemble of obtained values. This has to be a matter for future research, providing hardware developments look sufficiently promising.

As well as being more energy efficient, it is possible that such probabilistic architectures may offer a significant increase in computational speed up for climate simulator codes. If this is so, probabilistic processing may allow a route to cloud-resolved climate simulation much faster than anyone had previously expected, again allowing one to realise the goals of the World Summit on Climate Modelling (Shukla et al, 2010) in the foreseeable future.

The use of probabilistic Earth-system simulators running on machines built with stochastic processors, ie where the inherent quantum-mechanical noise associated with electrons flowing through transistors becomes a resource rather than a nuisance, provides a new synergy between software and hardware design in the field of weather and climate prediction, hitherto unimagined.

6. Probabilistic forecasting and seamless weather prediction – opportunities and possible obstacles.

Following Bjerknes (1904), NWP has historically been considered an example of a deterministic initial-value problem. The notion of probabilistic forecasting using ensemble prediction methods has evolved more recently as a tool to mitigate the effects of chaotic weather variability. Operational ensemble weather prediction systems have been implemented since the mid 1980s (Murphy and Palmer, 1986), long before ensemble methods became commonplace in climate prediction (see the review by Lewis, 2005). Following considerable investment in ensemble prediction at a number of NWP centres since these early days, most NWP centres now develop both a high-resolution deterministic forecast system, and a lower-resolution EPS, and strategic goals are targeted separately on improvements in both deterministic scores for the high-resolution system (eg RMS error or anomaly correlation coefficient) and probabilistic scores for the EPS (eg ranked probability skill score).

These goals are individually challenging and require the determined effort of scientists across a range of disciplines (numerics, parametrisation, data assimilation etc). This presents important questions about allocation of resources. How, for example, should an NWP centre partition its human resources to meet both the deterministic strategic goal on the one hand, and the probabilistic target on the other? A common view is that if most human resources are put into meeting the deterministic goal, the resulting improvements to the deterministic forecast system will necessarily benefit the EPS and help ensure the probabilistic target is also met.

Unfortunately, this concept of "trickle down" does not apply to the development of stochastic parametrisation. Weather simulators with stochastic parametrisations cannot produce forecasts with as low RMS error, or as high anomaly correlation coefficient, as equivalent simulators with deterministic parametrisations (see Fig 5a for an explicit illustration of this in the parametrised Lorenz '96 model). The reason why a probabilistic simulator will not outperform a comparable deterministic simulator in terms of deterministic scores is similar to the reason why the most skilful "deterministic" forecasts are associated with the ensemble-mean forecast (see Fig 11). The reason why an ensemble-mean forecast has especially high deterministic skill is that the relatively unpredictable components of the flow are "damped out" in an ensemble-mean field. However, a penalty is paid for such dynamical smoothing.An ensemble-mean forecast is unlikely to predict the occurrence of a severe weather event, if such an event is relatively unpredictable;the ensemble-mean forecast "hedges" towards climatology and away from such events. Now, as discussed in Sections 2 and 3, a deterministic bulk-formula parametrisation can be considered as providing an estimated mean tendency based on a putative ensemble of inherently unpredictable sub-grid processes, and hence will produce a "damped" simulation of the flow at sub-synoptic scales. In the same sense that an ensemble-mean forecast has low deterministic error, a simulator with deterministic bulk-formula parametrisation will tend to produce forecasts with lower RMS error than an equivalent simulator with stochastic parametrisation, particularly for near grid-scale circulations; recall, cf Fig 3, that each realisation of the stochastic parametrisation is designed to represent a potential realisation of the sub-grid flow, rather than an ensemble average. However, as with the ensemble mean forecast, there is a price to pay for this smoothing – a tendency for the simulator to hedge away from simulating extreme flows. This effect will obviously be strongest for small scales. However, as discussed in Section 2b, small-scale errors can be expected to propagate, intermittently but rapidly, to larger-scale components of the flow.

Put bluntly, stochastic parametrisation is anathema to the strategic goal of maximising deterministic skill! As such, development of stochastic parametrisation at the process level, the type of activity discussed in Section 4, will not naturally emerge from research that is focussed primarily at improving the high-resolution deterministic forecast system.

Should NWP centres therefore start planning for the day where they focus exclusively on developing probabilistic forecast systems, drop their higher resolution deterministic predictions, and measure progress primarily in terms of improvements to probabilistic scores?

Some may argue against this, noting that the enhanced skill of higher-resolution deterministic forecast systems justifies their continued separate development. Unquestionably, this was true in the past: in the early days of operational EPS, the deterministic skill of the unperturbed EPS control forecast was substantially poorer than that of the higher resolution deterministic forecast. However, these days, as shown in Fig 11, the skill of the higher-resolution deterministic forecast is no longer substantially greater than that of the EPS. It should certainly not be concluded from this that there is no need for the development of high-resolution simulators. There is evidence that at T1279 resolution, extreme weather events (such as hurricanes) can be simulated with greater realism than at T639 resolution. Rather, the point of Fig 11 is that the impact of high resolution is more subtle than it was in the past, and much less apparent in headline strategic scores such as 500hPa anomaly correlation coefficient. In Section 2b, the question was raised: How can we ensure that the advantages of integrating simulators at higher and higher resolution will not be somehow be compromised by the intermittent upscale propagation of error? That is to say, how can we produce high-resolution forecast systems that are reliable (the overall theme of this paper)? The author believes that the answer to this question is that future high resolution forecast systems must be explicitly probabilistic.

Others, arguing against this conclusion, may claim that weather forecast offices will continue to require high-resolution deterministic forecasts for the foreseeable future, since weather-forecast customers demand precise deterministic forecasts, and find probabilistic forecasts difficult to understand and difficult to use. This argument becomes yet stronger when one realises that the computational cost of a single high-resolution deterministic forecast is small compared with the cost of a full EPS.

Why is it that the public wants and expects detailed determinstic forecasts? Certaintly, nobody wants an uncertain forecast if a perfect deterministic forecast is available. But the latter is not available and never will be. In the author's opinion, a key reason why the public expects deterministic forecasts is simply because that is what they have been given and hence led to expect, ever since the days of Fitzroy when the first weather forecasts were made available to the general public. However, in cases where uncertainty is routinely expressed to the public, eg in the US National Hurricane Center's "cone of uncertainty" for hurricane track predictions (http://www.nhc.noaa.gov/aboutcone.shtml), the author's own informal research suggests that the public understands, and indeed respects, these uncertain predictions, and consequently no longer demands deterministic predictions.

Independently of whether the public are ready to accept the notion of an explicitly uncertain forecast, perhaps there is an argument that by focussing on probabilistic forecasting methods, the traditional skills of the weather forecaster will somehow be undermined. However, the author believes that the skills of human forecasters will be needed as never before when forecasts are primarily probabilistic in nature. In particular, there will be a need for a greater dialogue between forecasters and customers to help guide individual customers formulate weather-sensitive decision strategies appropriate to their circumstances. A simple (and rather idealised) example is based on the cost/loss model (Murphy, 1969). If a customer incurs a loss L if a particular weather event E (eg based on temperature, precipitation, wind, or some combination thereof) occurs, but can take protective action at cost C, then it makes rational sense to take this protective action on those occasions when the forecast probability of E exceeds C/L. In these circumstances, the job of the forecaster will be to "tease out", at least approximately, the customer's C/L and therefore enable that customer decide on the optimal threshold probability above which preventative action should be taken. Using this cost/loss model, Fig 12 shows the "potential economic value" of the EPS, compared with that the high-resolution deterministic forecast – the latter can be considered as a probabilistic forecast producing only probabilities of one or zero - based on precipitation events at forecast day 4. A "potential economic value" of unity would correspond to a hypothetical perfect deterministic forecast, and a "potential economic value" of zero would correspond to the value obtained by knowing only the climatological probability of E. The value of the EPS is substantially higher than that of the high resolution deterministic forecast – indeed, for a range of users, the high-resolution deterministic forecast, by itself, has no value at all over and above a decision based only on the climatological probability of E. Once again, it should be stressed that this does not at all imply that there is no merit in high resolution. Rather, Fig 12 suggests that the value of high resolution is masked when assessed in deterministic mode. There is no reason to doubt that a T1279 EPS would have higher “potential economic value” than the current T639 EPS, especially for severe events E.

In practice, decision strategies will be much more complex than suggested by a simple cost/loss models, for example requiring knowledge of the customer’s “utility function” which maps, ususally nonlinearly, multiple correlated weather variables to some quantity relevant to the customer (number of ice creams sold, megawatts of electricity produced). It will be the job of tomorrow’s weather forecaster to help the weather-sensitive customer to formulate his or her decision strategy in these realistic circumstances. It is interesting to note that, in this respect, great advances have been made recently in applying ensemble-based probability forecasts to provide flood risk assessments for farmers and community leaders in developing countries in the tropics, and these have been shown to have genuine value in saving lives and property (Webster et al, 2010).

In their interface with the general public, media forecasters need not only to be open about the inherent uncertainty in forecasts, they should routinely relay to the public the fact that techniques exist to quantify this uncertainty. This does not necessarily mean displaying isopleths of probability on TV . However, during media forecasts, forecasters could refer to web sites, even better to interactive displays (“press your red button”), where “fan charts” for temperature and rainfall, similar to those used by the Bank of England in forecasting inflation rate and gross domestic product (http://www.bankofengland.co.uk/publications/inflationreport/irfanch.htm), can be displayed for key cities.

In conclusion then, it is proposed that, in the coming decade, NWP centres should start to focus exclusively on developing probabilistic forecast systems, dropping their separate higher-resolution deterministic forecast systems, and, importantly, measuring progress, and formulating strategic goals, principally in terms of improvements to probabilistic scores. Such a strategy would certainly be consistent with the aims of forecast offices to provide reliable predictions of severe weather: no forecast can be considered reliable without an accurate assessment of forecast uncertainty, and severe weather events are often the most unpredictable and hence uncertain. The benefits of this, in addition to that of improving the reliability of weather forecasting *per se*, would be that improvements made to simulators on the weather timescale would likely also improve the reliability of simulators for longer-term climate prediction.

7. Towards a Seamless Probabilistic Earth-System Simulator for Weather and Climate Prediction

As discussed in the Introduction, output from comprehensive climate simulators informs mitigation policies, climate adaptation strategies and efforts to understand the impacts of climate geoengineering, and generally reduces society's vulnerability to current and future climate. One is hard pressed to think of examples where computer code has such societal relevance! And yet, as discussed above, there are substantial challenges (theoretical, computational, and human) that need to be overcome if we are to progress significantly to the goal of providing society with reliable estimates of future climate, regional and global.

In discussing possible ways to meet these challenges, consider by analogy, the state of the European civil aircraft industry in Europe in the mid 20th Century. At this time, all the major European countries produced their own civil aircraft. However, it was realised that aircraft were becoming too complex and too expensive for individual countries to develop and manufacture independently. Within this milieu, the Airbus consortium (http://www.airbus.com/) was formed. At the time, there must have been much agonising at the national level, as to whether national aerospace industries were doing the right thing getting together in this way. In retrospect, there can be little doubt but that it was. And so, within the Airbus consortium, these same national aerospace industries now focus on specific aspects of the design and production of aircraft in their fight for market share with their great US rival, Boeing.

Hence, by analogy, we can imagine a multi-national Earth-system simulator supported by teams of scientists from national climate and academic institutes. Different teams would focus on different aspects of the simulators: dynamical cores, oceans, clouds, aerosols etc, and on the design of experiments which integrate these aspects together. All should contribute to the analysis and diagnosis of results. To support this, computational resources would be available, not only for operational integrations, but for plentiful research experimentation. Results from the small number of simulators worldwide might continue to be combined in a multi-simulator ensemble, but since each is now based on stochastic-dynamic closure, the resulting ensemble would be much less prone to the type of systemic failure that current generation multi-simulator ensembles are capable. National weather services would still play a crucial role in development work, in conducting scientific experiments, and in communicating the results from the science to their governments and society alike.

Is this a possible framework for the development of future Earth-system simulators? To some extent it already is. For example, within Europe, many climate institutes use the same (NEMO; http://www.nemo-ocean.eu/) ocean simulators. Indeed, development of the EC-Earth simulator (Hazeleger et al, 2010) provides a specific example of how international cooperation can be successful, having been developed from the ECMWF seasonal forecast simulator, ECMWF itself being an outstandingly successful example of international cooperation in the context of NWP.

Given the merits of poolings resources, why would we not want to go further down this route of rationalisation? The key argument for not adopting the “Airbus” model is that we need extensive simulator diversity in order to estimate prediction uncertainty. However, the stochastic science discussed in the previous sections (and this is why the discussion has been so extensive) suggests that an alternative approach to representing simulator uncertainty is beginning to emerge, and, on timescales where verification data exists, this alternative approach can outperform that provided by conventional multi-simulator ensembles. That is, the argument for maintaining the *status quo* of extensive simulator diversity is being undermined by scientific developments.

It should be stressed that it is not being suggested here that stochastic parametrisation implies that all we need is one “World Weather and Climate Simulator”. Airbus has undoubtedly been successful, not only because it can draw from the pooled resources of European aerospace industries, but also because it has a competitor from another geopolitical grouping. Similarly, one would imagine that if there was some rationalisation of climate simulator development effort, which embraced the notion of stochastic parametrisation as the primary means to estimate simulator uncertainty, then we would still have enough (quasi-) independent Earth-system simulators to foster competition and creativity. What is a desirable number of comprehensive Earth-system simulators? This obviously depends on an assessment of the minimum human and computational resources needed to develop and maintain an Earth-system simulator. However, the author would broadly concur with the findings of the World Summit on Climate Modelling (Shukla et al, 2010), that development of “a small number” based around major geopolitical groupings, might be ideal.

In the course of this paper, evidence has been given how the development of explicitly probabilistic weather and climate simulators will lead to more reliable estimates of uncertainty. At the beginning of the paper, it was also suggested that these methods might be able to actually reduce uncertainty. In considering this possibility, let us focus here on what must surely be the most important, as well as the most uncertain, of all the feedbacks in the climate change problem: that associated with cloud. As is well known (Solomon et al, 2007), even the sign of the cloud feedback is uncertain.

One of the problems in thinking about the notion of “cloud feedback” is that a world without cloud, and hence without cloud feedback, would be utterly alien to us: clouds are absolutely intrinsic to the circulation patterns we observe around us. Not only are clouds determined by the temperature and humidity structure associated with these circulation patterns, clouds in turn are key to determining these circulation patterns, both locally and remotely. For example, anomalous latent heat release in convective cloud systems over the Caribbean may be key to setting up a blocking anticyclone over Europe, whilst the stratus decks that form locally in the vicinity of the blocking anticyclone are key to determining the surface temperature under the block.

This means that we cannot treat the problem of cloud feedback solely as a problem in atmospheric thermodynamics; the problem is as much dynamic as thermodynamic. For the sake of argument, let us consider climate as a dynamical system with distinct nonlinear regime structures (Palmer, 1998; Straus et al, 2007) in both the tropics and extratropics. These regimes will in turn have distinct cloud properties (Williams and Webb, 2009): a blocking anticyclone may be dominated by relatively thin stratus clouds in winter and cirrus clouds in summer, whilst a cyclonic weather regime will contain significant amounts of thick nimbostratus cloud all times of year. From this dynamical perspective, a key element of understanding the cloud feedback problem lies in estimating reliably how anthropogenic forcing will change the frequencies of occurrence of the regimes. (Changes to the structure of the regimes may also be important, but depending on the stability of the regimes, this may be a secondary aspect of the problem.) That is to say, changes in these frequencies of occurrence will be one of the key factors in determining whether upper or lower-level clouds increase or decrease as a result of anthropogenic climate change. Small wonder then that current climate simulators have such difficulty in simulating the sign of cloud feedback with any consistency. As discussed above, these same simulators have difficulty simulating the statistics of observed weather regimes.

Hence, to really make progress in reducing the uncertainty in cloud feedbacks it will be essential that the statistics of weather regimes are simulated correctly: their three dimensional structure, their embedded cloud properties and their frequency of occurrence (see also Stephens, 2005). This is a profoundly challenging dynamical problem, and results suggest that the current generation of climate simulators is not fully up to the challenge.

The same arguments could be applied to another of the important uncertainties in climate prediction: the impact of aerosols. Here the key uncertainties relate to the indirect effect of aerosols, ie through their modification of cloud. Again, this indirect effect will be regime dependent, implying that we will never be able to assess aerosol impact reliably in the atmosphere, without an accurate simulation of structure and frequency of occurrence of weather regimes.

With this in mind, we can suggest why the proposal for inherently probabilistic Earth-system simulators will reduce uncertainty in predictions of climate:

a) As discussed above, representing simulator uncertainty by stochastic parametrisation undermines the inherent need for a large diversity of simulators, meaning that it will be possible to pool human and computational resources. Economies of scale will enable climate scientists to have dedicated access to top-of-the-range supercomputers, enabling key physical processes to be simulated, including *in situ* Rossby wave breaking, key for maintaining weather regimes against dissipation (Woollings et al, 2008), and remote tropical convective systems which help "force" these regimes.

b) Being more consistent with the underlying equations of motion, it could be argued that if there are to be breakthoughs in parametrisation, eg of the effects of unresolved cloud systems, they are more likely to occur within a more general probabilistic framework, than within the traditional deterministic framework.
c) Development of seamless probabilistic weather and climate simulators will enable sophisticated diagnostic tools from data assimilation to be used to reduce climate prediction uncertainty (Rodwell and Palmer, 2007), eg based on studies of biases in analysis increments, composited on specific weather regimes. The use of data assimilation in assessing stochastic parametrisation was illustrated in Fig 8. .
d) There is evidence that stochastic parametrisations can improve directly, estimates of the frequency of occurrence of weather regimes (Jung et al,2005). The reason relates to the rectification of the flow by stochastic noise. As a simple analogy, imagine a ball bearing moving in a potential with multiple minima; an overly damped system will lead to the ball bearing spending too much time in the dominant well and this will be reflected in a bias in the time-averaged position of the ball.

8. Conclusions

Compared with the economists, weather and climate scientists do indeed know their equations, at least as they relate to the physics of weather and climate. However, these equations cannot be solved by pencil and paper. Algorithmic representations of the equations of motion necessarily involve errors and with conventional numerical algorithms based on deterministic closures these errors appear to lead to substantial biases and considerable uncertainty in simulating climate. Some discussion has been given to the possibility that convergence to the "true" underlying equations with increasing resolution, may be exceptionally slow, due to the "-5/3" power law for atmospheric energy. Some technical discussion has been given to an alternative strategy for closing the equations, where the inherent uncertainty in any algorithmic representation of the underlying equations is recognized explicitly. It is suggested that breakthroughs in the parametrisation problem, if they are to occur, will be more likely within a stochastic framework, than in the traditional deterministic framework.

On timescales where verification data exists, these stochastic methods are beginning to outperform conventional multi-simulator ensembles. However, there is much work to be done before all relevant Earth system parametrisations can be said to have been developed in this probabilistic way. Indeed, it has been concluded that focussing excessively on the traditional challenge in NWP, of reducing deterministic measures of forecast error, may increasingly become an obstacle to the seamless development of reliable probabilistic weather and climate simulators. It was argued that it may indeed be time to consider focussing operational weather forecast development entirely on high-resolution ensemble prediction systems.

A key aspect of this paper has been discussion on some of the implications of a move towards probabilistic Earth-system simulation, implications that transcend the technical aspects of stochastic parametrisation. In particular, by underming the argument for a large pool of quasi-independent simulators, the stochastic parametrisation programme provides new support for one of the key conclusions of the World Summit on Climate Modelling (Shukla et al, 2010): for a pooling of human and computational resources amongst climate institutes and for a substantial rationalization of development work towards a very small number of independent Earth-system simulators.

Given the importance and urgency of predicting Earth's climate as accurately as science and technology allows, it is time to give serious thought to such change.

Acknowledgements

My thanks to a number of colleagues who helped produce results shown in this paper, and with whom I had valuable discussions and received helpful comments. These include: Hannah Arnold, Peter Bauer, Judith Berner, Roberto Buizza, Tony Busalacchi, Erland Källén, Chris Farmer, Anna Ghelli, Brian Hoskins, Hugh McNamara, David Richardson, Gregory Seregin, David Straus, Andrew Majda, J. Shukla, Julia Slingo, Rowan Sutton, Jean-Noël Thépaut, Alan Thorpe and Antje Weisheimer. I am also grateful to two anonymous reviewers for their comments. My thanks also to Rob Hine for helping produce the illustrations.

References

Arnold, H., I. Moroz, T.N.Palmer, 2011: Stochastic Parametrisations and Model Uncertainty in the Lorenz '96 System. *Phil. Trans. Roy. Soc.* In preparation.

Arrhenius, S., 1896: On the influence of carbonic acid in the air on the temperature of the groud. The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science Series, 41, 39pp (Reprinted in "The Warming Papers" 2011, edited by David Archer and Raymond Pierrehumbert. Wiley Blackwell, 419pp)

Bengtsson-Sedlar, L., H.Körnich, E.Källén and G. Svensson, 2011: Large-scale dynamical response to sub-grid scale organization provided by cellular automata. *J.Atmos.Sci.* To appear.

Berner, J., G.J.Shutts, M.Leutbecher and T.N.Palmer, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow dependent predictability in the ECMWF ensemble prediction system. *J. Atmos.Sci*., **66**, 603-626.

Berner, J., T. Jung and T.N.Palmer, 2011: Systematic Error Model Error:The Impact of Increased Horizontal Resolution versus Improved Stochastic and Deterministic Parameterizations. *J. Clim.* Submitted.

Bjerknes, V., 1904: Das Problem der Wettervorhersage betrachter von Standpunkte der Mechanik und der Physik. *Meteorol. Z.*, **21**, 1-7.

Bonavita, M. 2011: Ensemble of data assimilations and estimation of uncertainty. ECMWF 2011 Seminar Proceedings.

Buizza, R. and T.N.Palmer, 1995: The singular vector structure of the atmospheric global circulation. *J. Atmos.Sci.*, **52**, 1434-1456.

Buizza, R., M. Miller and T.N.Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc*., **125**, 2887-2908.

Doblas-Reyes, F.J., M.A.Balmaseda, A.Weisheimer and T.N.Palmer, 2011: Decadal climate prediction with the ECMWF coupled forecast system: Impact of Ocean Observations. *J. Geophys. Res*., 116, D19111, doi:10.1029/2010JD015394

Frederiksen, J. S., A. G. Davies, 1997: Eddy Viscosity and Stochastic Backscatter Parameterizations on the Sphere for Atmospheric Circulation Models. *J. Atmos. Sci.*, **54**, 2475–2492.

Frederiksen, J, S., 1999: Subgrid-Scale Parameterizations of Eddy-Topographic Force, Eddy Viscosity, and Stochastic Backscatter for Flow over Topography. *J. Atmos. Sci.*, **56**, 1481–1494.

Frederiksen, J, S., S. M. Kepert, 2006: Dynamical Subgrid-Scale Parameterizations from Direct Numerical Simulations. *J. Atmos. Sci.*, **63**, 3006–3019.

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

Goldstein, M. and J.C. Rougier, 2004: Probabilistic Formulations for Transferring Inferences from Mathematical Models to Physical Systems. SIAM *Journal on Scientific Computing*, **26(2)**, 467-487.

Hasselmann K., 1976: Stochastic climate models, Part I, Theory, *Tellus*, **28**, p.473, 1976.

Hazeleger, W., Camiel Severijns, Tido Semmler, Simona Ştefănescu, Shuting Yang, Xueli Wang, Klaus Wyser, Emanuel Dutra, José M. Baldasano, Richard Bintanja, Philippe Bougeault, Rodrigo Caballero, Annica M. L. Ekman, Jens H. Christensen, Bart van den Hurk, Pedro Jimenez, Colin Jones, Per Kållberg, Torben Koenigk, Ray McGrath, Pedro Miranda, Twan Van Noije, Tim Palmer, José A. Parodi, Torben Schmith, Frank Selten, Trude Storelvmo, Andreas Sterl, Honoré Tapamo, Martin Vancoppenolle, Pedro Viterbo, Ulrika Willén. 2010: EC-Earth: A Seamless Earth-System Prediction Approach in Action. *Bull.Amer. Meteor. Soc.*, **91**, 1357-1363

Isaksen, L., M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher and Laure Raynaud, 2010: Ensemble of data assimilations at ECMWF. ECMWF Technical Memorandum 636.

Jakob, C., 2010: Accelerating progress in global atmospheric model development through improved parametrizations. *Bull. Amer. Meteor. Soc*., **91**, 869-875.

Jung, T., T.N.Palmer and G.J. Shutts, 2005. Influence of stochastic parametrization on the frequency of occurrence of North Pacific weather regimes in the ECMWF model. *Geophys. Res. Lett*., **32**, L23811

Jung, T., M.J. Miller, T.N. Palmer, P. Towers, N. Wedi, D. Achuthavarier, J.M. Adams, E.L. Altshuler, B.A. Cash, J.L. Kinter III, L. Marx, C. Stan, and K.I. Hodges, 2011: High-resolution global climate simulations with the ECMWF model in Project Athena: Experimental design, model climate and seasonal forecast skill. *J.Clim*. To appear

Kinter III J.L., Cash B, Achuthavarier D, Adams J, Altshuler E, Dirmeyer P, Doty B, Huang B, Marx L, Manganello J, Stan C, Wakefield T, Jin E, Palmer T, Hamrud M, Jung T, Miller M, Towers P, Wedi N, Satoh M, Tomita H, Kodama C, Nasuno T, Oouchi K, Taniguchi H, Andrews P, Baer T, Ezell M, Halloy C, John D, Loftis B, Mohr R, Wong K (2011) Revolutionizing Climate Modeling – Project Athena: A Multi-Institutional International Collaboration. *Bull. Amer. Meteor.Soc*. To appear).

Khouider, B., A.J. Majda and M.A. Katsoulakis, 2003: Coarse-grained stochastic models for tropical convection and climate. *Proc. Nat. Acad. Sci*., **100**, 11941-11946.

Lander, J. and B.J.Hoskins, 1997: Believable and parametrizations in a spectral transform model. *Mon. Wea. Rev.*, **125**, 292-303.

Lewis, J.M., 2005: Roots of ensemble forecasting. *Mon. Wea. Rev.*, **133**, 1865-1885.

Lilly D K, 1973: Dynamic Meteorology ed P *Morel* (Boston, MA: Riedel).

Lin, J. W.-B. and J.D.Neelin, 2003: Toward stochastic deep convective parameterization in general circulation models. *Geophys. Res. Lett.*, **30**, 1162, doi:10:1029/2002GL016203.

Lindborg, Erik, 2007: Horizontal Wavenumber Spectra of Vertical Vorticity and Horizontal Divergence in the Upper Troposphere and Lower Stratosphere. *J. Atmos. Sci.*, **64**, 1017–1025.

Lorenz, E.N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289-307.

Lorenz E.N. 1975. Climatic Predictability. In "The Physical Basis of Climate and Climate Modelling". WMO GARP Publication Series No 16. World Meteorological Organisation. Geneva: 265 pp.

Lorenz, 1996. Predictability – a problem partly solved. Proceedings. Seminar on Predictability. ECMWF. Reprinted in Palmer, T.N. and Hagedorn, R. 2006. Predictability of Weather and Climate. Cambridge University Press.

Majda, A.J and A.L. Bertozzi, 2001: Vorticity and Incompressible Flow. Cambridge Texts in Applied Mathematics. Cambridge University Press, 545pp

Manabe, S. and R.T. Wetherald, 1975: The effects of doubling the CO2 concentration on the climate of a general circulation model. *J.Atmos. Sci.*, **32**, 3-15.

Masson, D. and R. Knutti, 2011: Climate model geneology. *Geophys. Res. Lett*., **38**, L08703

Matsueda, M. and T.N.Palmer, 2011: Accuracy of climate change predictions using high resolution simulations as surrogates of truth. *Geophys. Res. Lett.*, **38**, L05803, doi:10.1029/2010GL046618

Meehl. G.A., C. Covey, T.Delworth, M.Latif, B;McAvaney, J.F.B. Mitchell, R.J. Stouffer and K. E. Taylor: 2007. The WCRP CMIP3 multimodel dataset. *Bull. Amer. Meteor. Soc*., **42**, 950-960

Murphy, A. H., 1969: Measures of the Utility of Probabilistic Predictions in Cost-Loss Ratio Decision Situations in which Knowledge of the Cost-Loss Ratios is Incomplete. *J. Appl. Meteor.*, **8**, 863–873.

Murphy, J.M. and T.N.Palmer, 1986: Experimental monthly forecasts for the United Kingdom. Part II: A real-time long-range forecast by an ensemble of numerical integrations. *Met. Mag*., **115**, 337-344.

Murphy, J.M. et al, 2004: Quantification of modelling uncertainties in a large ensemble of 265 climate change simulations. *Nature,* **430**, 768-772.

Nastrom, G.D. and K.S. Gage, 1985: A climatology of atmosphere wavenumber spectra of wind and temperature observed by commercial aircraft. *J. Atmos.Sci*., **43**, 857-870.

Palem, K. V. 2005. Energy aware computing through probabilistic switching: A study of limits. *IEEE Trans. Comput.* ***54***, 1123–1137.

Palmer, T.N. and P.J. Webster, 1993: Towards a unified approach to climate and weather prediction. Proceedings of 1st Demetra Conference on Climate Change. European Community Press.

Palmer, T.N., 1997: On parametrising scales that are only somewhat smaller than the smallest resolved scale, with application to convection and orography. Proceedings of the ECMWF Workshop on new insights and approaches to convective parametrisation. ECMWF. Shinfield Park. Pp 328-337.

Palmer, T.N., 1998: Climate change from a nonlinear dynamical perspective. *J.Clim*.,**12**, 575-591.

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

Palmer, T.N., F.J. Doblas-Reyes, A.Weisheimer and M.J.Rodwell, 2008:Towards seamless prediction: Calibration of climate change projections using seasonal forecasts. *Bull. Amer. Meteor. Soc*., **89**, 459-470.

Palmer, T.N., R. Buizza, F.Doblas-Reyes, T.Jung, M. Leutbecher, G.J.Shutts, M.Steinheimer and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. Technical Report 598. European Centre for Medium-Range Weather Forecasts.

Palmer, T.N., 2011: A CERN for climate change. *Physics World*, **24**, 14-15

Palmer T.N. and A. Weisheimer, 2011: Diagnosing the causes of bias in climate models – why is it so hard? Geophys*. and Astrophys. Fluid Dynamics*, **105**, 351-365.

Pennell, C. and T. Reichler, 2011: On the effective number of climate models. *J. Clim*., **24**, 2358-2367.

Phillips, N.A., 1956: The general circulation of the atmosphere: a numerical experiment. *Quart. J. Roy. Meteor. Soc.*, **82**, 123-64

Plant, R. S., G. C. Craig, 2008: A Stochastic Parameterization for Deep Convection Based on Equilibrium Statistics. *J. Atmos. Sci.*, **65**, 87–105.

Pollard, D., 2002: A User's Guide to Measure Theoretic Probability. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. 351pp

Rodwell, M. J. and T.N.Palmer, 2007: Using numerical weather prediction to assess climate models. *Quart. J. Roy. Meteor. Soc*., **133**, 129-146.

Rossow, W.L. and B. Cairns, 1995: Monitoring Changes of Clouds. Climate Change, **31**,305-347

Rotunno, R. and C. Snyder, 2008: A generalization of Lorenz's model for the predictability of flows with many scales of motion. *J. Atmos. Sci.*, **65**, 1063-1076.

Scaife A.A., T. Spangehl, D. Fereday, U. Cubasch, U. Langematz, H. Akiyoshi, S. Bekki, P. Braesicke, N. Butchart, M. Chipperfield, A. Gettelman, S. Hardiman, M. Michou, E. Rozanov and T.G. Shepherd, 2011:Climate Change and Stratosphere-Troposphere Interaction. *Clim. Dyn.*, DOI 10.1007/s00382-011-1080-7.

Schertzer, D., and S. Lovejoy, 1993: *Lecture Notes: Nonlinear Variability in Geophysics 3: Scaling and Mulitfractal Processes in Geophysics*, 292 pp., Institut d'Etudes Scientifique de Cargèse, Cargèse, France.

Schertzer. D. and S. Lovejoy, 2004: Uncertainty and unpredictability in geophysics: chaos and multifractal insights. State of the planet: frontiers and challenges in geophysics. *Geophysical Monograph Series,* **150**, 317- 334. American Geophysical Union.

Shukla J., T.N.Palmer, R. Hagedorn, B. Hoskins, J. Kinter, J. Marotzke, M. Miller, J. Slingo 2010: Toward a new generation of world climate research and computing facilities. *Bull. Amer. Meteor.Soc., **91,*** 1407–1412.

Shutts, G., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079-3102.

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

Simmons, A.J. and A. Hollingsworth, 2002: Some aspects in the improvement in skill in numerical weather prediction. *Quart. J. Roy. Meteor. Soc*., **128**, 647-677.

Slingo, J. and T.N.Palmer, 2011: Uncertainty in weather and climate prediction. *Phil Trans Roy Soc.* To appear.

Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.), 2007: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, 2007. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Stainforth, D. et al., 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases, *Nature,* **433**, 403-406.

Stephens, G.L., 2005: Cloud feedbacks in the climate system:a critical review. *J. Clim*., **18**, 237-273.

Straus, D. M., S. Corti and F. Molteni, 2007: Circulation Regimes: Chaotic Variability versus SST-Forced Predictability. *J. Clim.*, **20**, 2251–2272.

Straus, D M., 2011: Macroturbulence in very high resolution atmospheric models. J. Atmos.Sci, in preparation.

Thuburn, J. , J. Kent and N. Wood, 2011: Energy and enstrophy cascades in numerical models. Proceedings of ECMWF Workshop on Representing Model Error in Weather and Climate prediction.

WCRP, 2005: The World Climate Research Programme Strategic Framework, 2005-2015 Coordinated Observation and Prediction of the Earth System (COPES). World Meterological Organisation Technical Document No 1291 (WCRP- 123) 59pp.

Webster P.J., Jun Jian, T.M.Hopson, C.D. Hoyos, P.A. Agudelo, Hai-Ru Chang, J.A. Curry, R.L.Grossman, T.N.Palmer, and A.R.Subbiah, 2010: Extended-Range Probabilistic forecasts of Ganges and Brahmaputra floods. *Bull. Amer. Meteor. Soc.*, **91**, 1493- 1514

Weisheimer, A., Doblas-Reyes, F.J., Palmer, T.N., Alessandri, A., Arribas, A., Deque, M., Keenlyside, Noel, MacVean, M., Navarra, A. and Rogel, P., 2009: ENSEMBLES: a new multi-model ensemble for seasonal-to-annual predictions: Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs *Geophys. Res. Lett.*, **36 (21)**, L21711.

Weisheimer, A., T. N. Palmer, and F. J. Doblas-Reyes, 2011: Assessment of representations of model uncertainty in monthly and seasonal forecast ensembles. *Geophys. Res. Lett*., **38**, L16703, doi:10.1029/2011GL048123.

Williams, K. and M. Webb, 2009: A quantitative performance assessment of cloud regimes in climate models. *Clim. Dyn.*, **33**, 141.

Wilks, D.S., 2005: Effects of stochastic parametrizations in the Lorenz '96 system. *Quart. J. Roy. Met. Soc.*, **131**, 389-407.

Wilks, D.S., 2006: Statistical methods in the atmospheric sciences, vol 1 of International Geophysics Series, Elsevier.

Woollings, T. , B.J. Hoskins, M. Blackburn and P. Berrisford, 2008: A New Rossby Wave–Breaking Interpretation of the North Atlantic Oscillation. *J. Atmos.Sci*., **65**, 609-626.

Figure Captions

1. Variance power spectra of wind and potential temperature based on aircraft observations. The spectra of meridional wind and temperature are shifted by one and two decades to the right, respectively. Lines with slopes -3 and -5/3 are entered at the same relative coordinates for each variable, for comparison. From Nastrom and Gage (1985).

2. Seasonal forecast reliability diagrams for the ENSEMBLES multi-simulator ensemble. Based on 1980-2001 hindcasts initialised on May 1st and for forecast period June-August. a) seasonal mean NINO3 sea surface temperature above upper climatological tercile. b) seasonal mean precipitation anomalies in Amazon Basin in lower climatological tercile. c) as b) but for Northern Europe. d) as b) but for Sahel. The dotted lines show the climatological frequency of the event and the size of the grey dots is indicative of the relative sample size within that probability bin.

3. a) Schematic of hypothetical situation where there is some scale separation between resolved and unresolved flow, justifying the notion of deterministic parametrisation . b) Schematic of the more realistic situation where there is no scale separation between resolved and unresolved flow, justifying the notion of stochastic parametrisation.

4. Probability distributions of the tendency term in the (“large-scale”) X equations, due to the (“small scale”) Y variables in the Lorenz (1996) dynamical system with c=4. Thick solid line: prior climatological distribution. Dashed line: distribution conditioned on $-6<X<-5$. Dashed line: distribution conditioned on $13<X<14$. The fact that the distribution is broader when X is constrained to large values, than when constrained to small values, provides some explanation for why the multiplicative noise parametrisation is more skilful in Fig 5. The thin solid lines define hat functions associated with the deterministic parametrisation scheme for $-6<X<-5$ and $13<X<14$. That these are much narrower than the conditional probability distributions shows the “overconfidence” of the deterministic parametrisation.

5. a) Solid and dashed lines show the Ranked Probability Skill Scores for 75 initial condition ensemble forecasts at t=0.6, based on differences between the Lorenz (1996) dynamical system and various parametrised versions of the system – see text for details - with c=10 (solid) and c=4 (dashed). The dotted line shows the ensemble mean RMS error for c=4 (with values given on the right hand side of the diagram) b) Hellinger Distance between the climatological probability distribution of the Lorenz (1996) dynamical system and the various parametrised versions, with c=10 (solid) and c=4 (dashed). Based on integrations over 400 time units.

6. Realisations of the stochastic pattern generator used in the ECMWF Stochastically Perturbed Parametrisation Tendency scheme (Palmer et al, 2009). Solid (dotted) lines correspond to positive (negative) values.

7. Continuous Ranked Probability Skill Score for 850hPa temperature in the tropics based on the ECWMF Ensemble Prediction System with no representation of model uncertainty (dotted line); the original “stochastic physics” scheme of Buizza et al (1999) (dashed line); the 2-time Stochastically Perturbed Parametrisation Tendency scheme described in Palmer et al (2009) (solid line). See Palmer (2009) for details.

8. Relationship between T=6hr ensemble spread and T+6hr ensemble-mean error of ensembles of data assimilations of the ECMWF forecast system, for tropospheric vorticity in the northen hemisphere, binned on error and based on: (dashed line) an ensemble with representation of observation uncertainty but not model uncertainty; (solid line) an ensemble with representation of observation and model uncertainty (based on stochastically perturbed tendencies and stochastic kinetic energy backscatter). The dotted line shows the ideal relationship. From Bonavita (2011).

9. Reliability diagram for the ECMWF Ensemble Prediction System at t=4 days for prediction of rainfall exceeding 1mm/day over the European domain, based on verification from March-May 2011. The values against the dots give the number of occasions where probability forecasts within a given 10% range (and at 0% exactly) were made.

10. Mean systematic error of 500 hPa geopotential height fields for extended boreal winters (December–March) of the period 1990-2005. Errors are defined with regard to the observed mean field (contours), consisting of a combination of ERA-40 (1990- 2001) and operational ECMWF analyses (2002-2005). Shown are the systematic error of experiments (a) low-resolution T95 simulator, (b) T95 simulator with stochastic kinetic-energy backscatter, (c) high-resolution T511simulator and (d) T95 simulator with improved deterministic parameterizations. Contour interval 2 decametres. (From Berner et al, 2011)

11. 500hPa geopotential anomaly correlation coefficient over Northern Hemisphere extratropics for March-May 2011. Light solid line: high resolution (T1279) ECMWF deterministic forecast. Dashed line: unperturbed control forecast from the (T639) ECMWF ensemble prediction system. Dotted line: deterministic forecast based on the ensemble average over the members of the ECMWF ensemble prediction system.

12. Potential economic value (Murphy, 1969) as a function of user cost/loss ratio, based on prediction of rainfall exceeding 1mm/day over the European domain for March-May 2011 (1=value of a perfect deterministic forecast, 0=value associated with a climatological probability forecast). Solid line for ECMWF Ensemble Prediction System. Dotted line associated with ECMWF high resolution deterministic forecast.

Table Caption

1. Brier Skill Scores for probabilistic predictions for all global land area 2m temperature and precipitation grid points, based on exceeding upper climatological tercile (warm/wet) and not exceeding lower tercile (cold/dry) events for: the ENSEMBLES multi-simulator ensemble (MSE), an ensemble using the ECMWF simulator with stochastic parametrisation (SPE) and an ensemble using the ECMWF simulator without any representation of simulator uncertainty. Bold figures indicate the system with the highest score. (From Weisheimer et al, 2011)

Figures

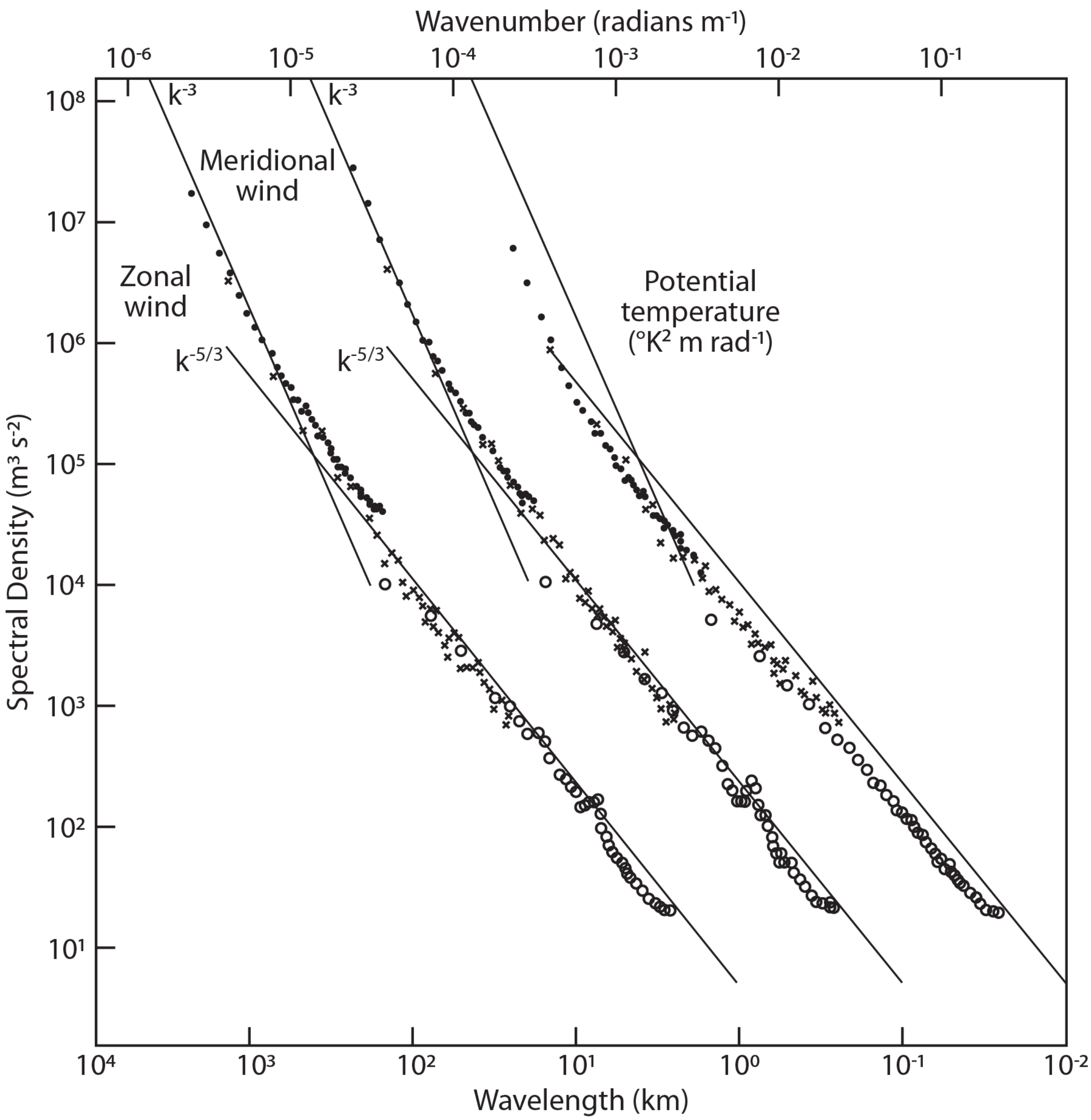


Figure 1.

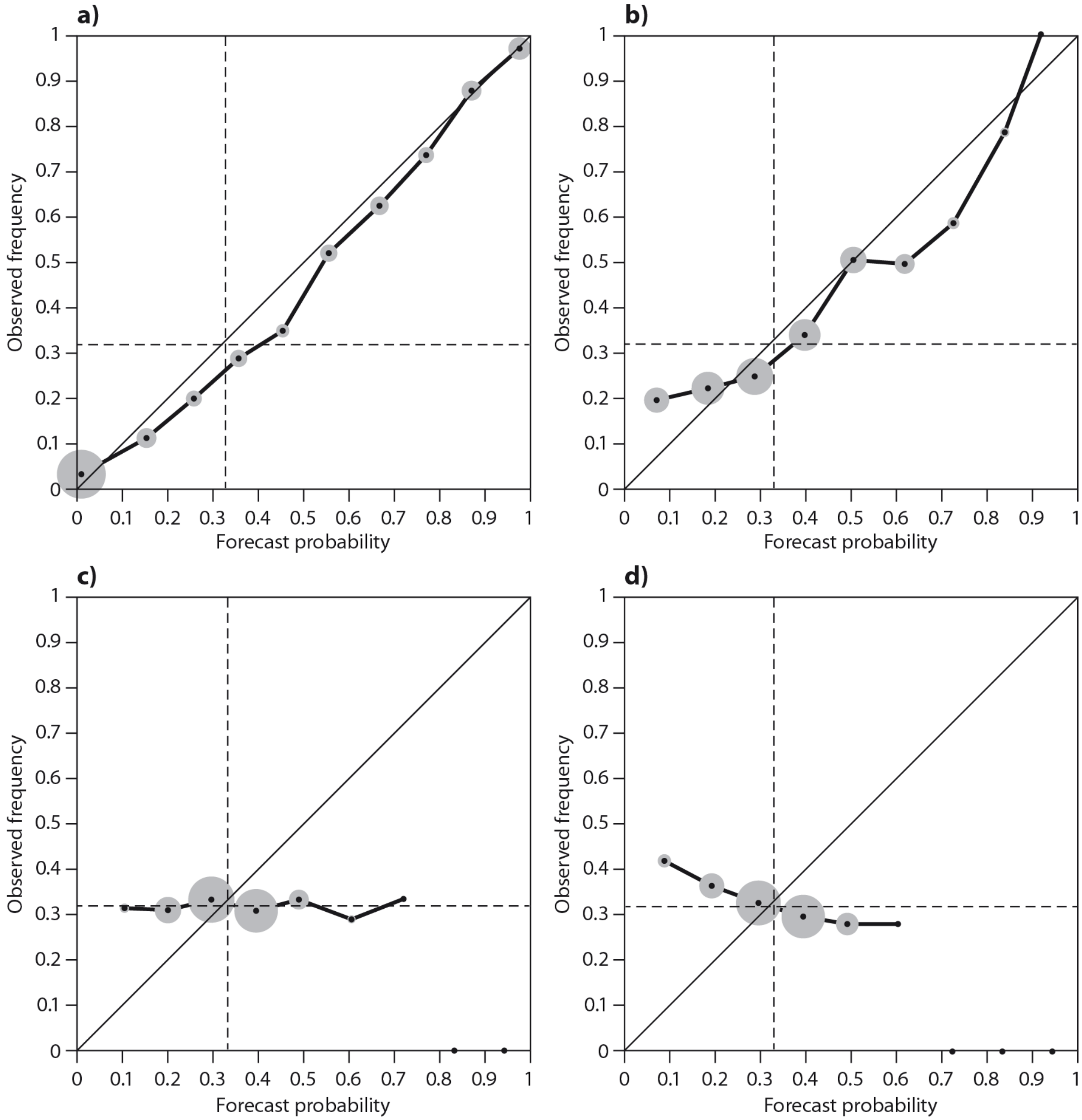

a)
b)
c)
d)
Observed frequency
Forecast probability
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Figure 2

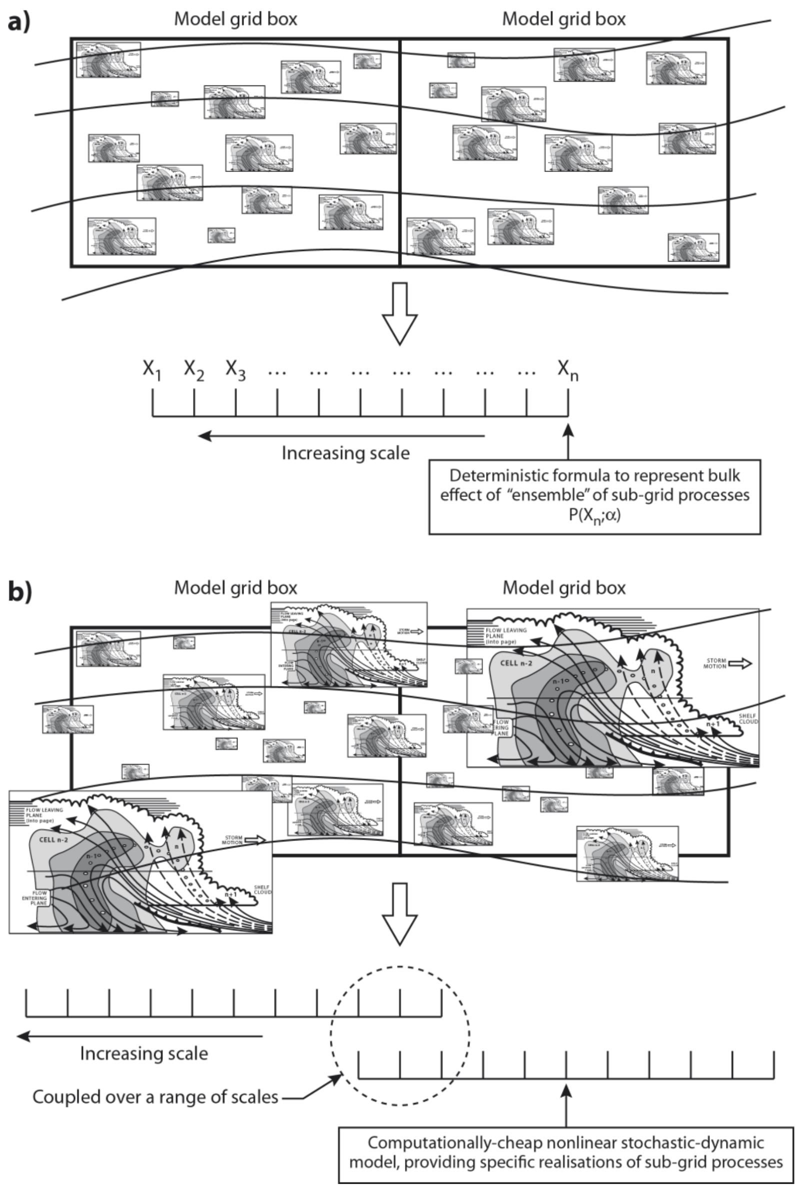
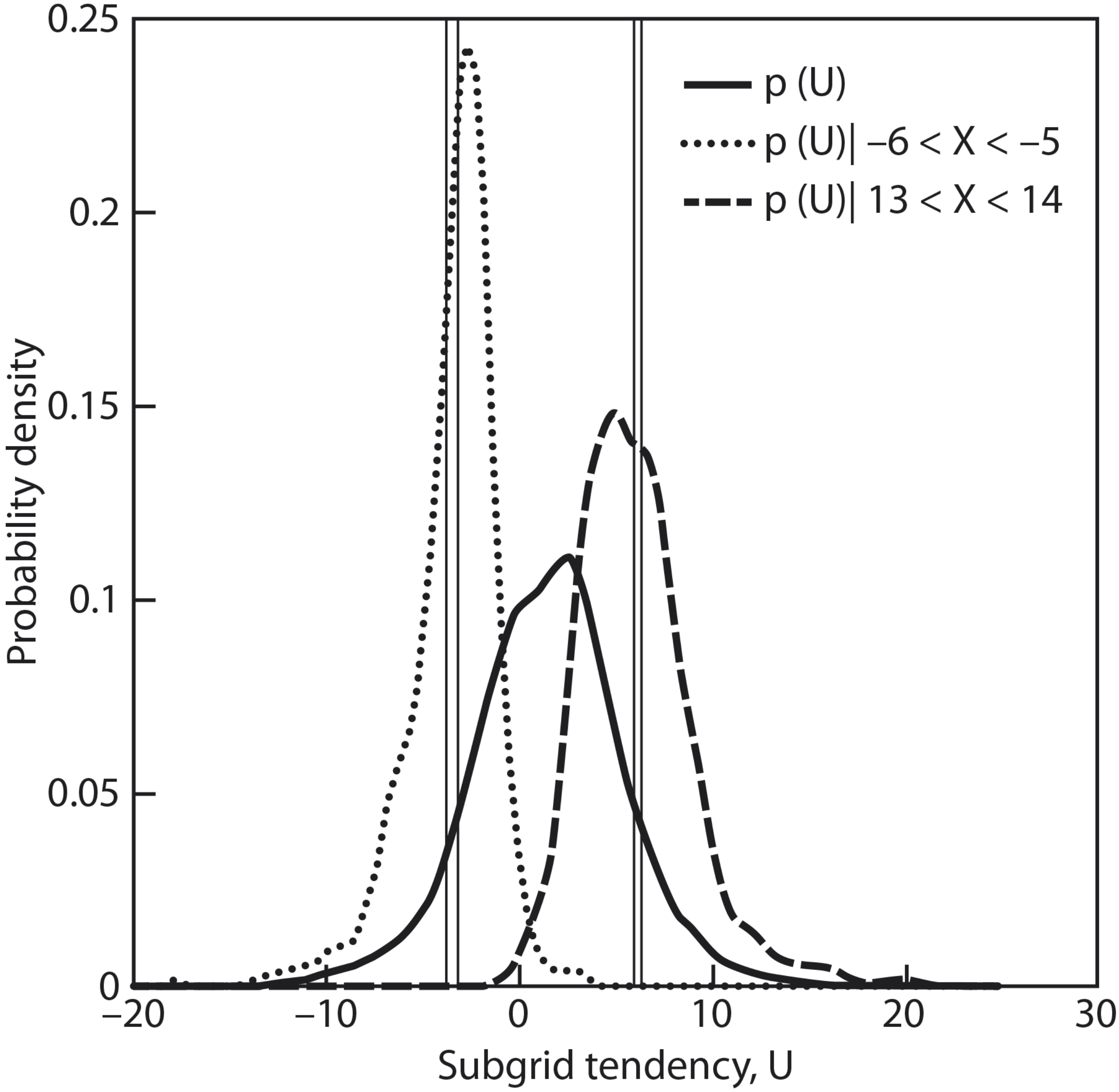

a)
Model grid box
Model grid box
X1 X2 X3 … … … … … … … Xn
Increasing scale
Deterministic formula to represent bulk effect of "ensemble" of sub-grid processes
P(Xn;α)
b)
Model grid box
Model grid box
FLOW LEAVING PLANE (into page)
CELL n-2
STORM MOTION
SHELF CLOUD
FLOW ENTERING PLANE
n+1
Increasing scale
Coupled over a range of scales
Computationally-cheap nonlinear stochastic-dynamic model, providing specific realisations of sub-grid processes

Figure 3



0.25
0.2
0.15
0.1
0.05
0
Probability density
−20
−10
0
10
20
30
Subgrid tendency, U
p (U)
p (U)| −6 < X < −5
p (U)| 13 < X < 14

Figure 4

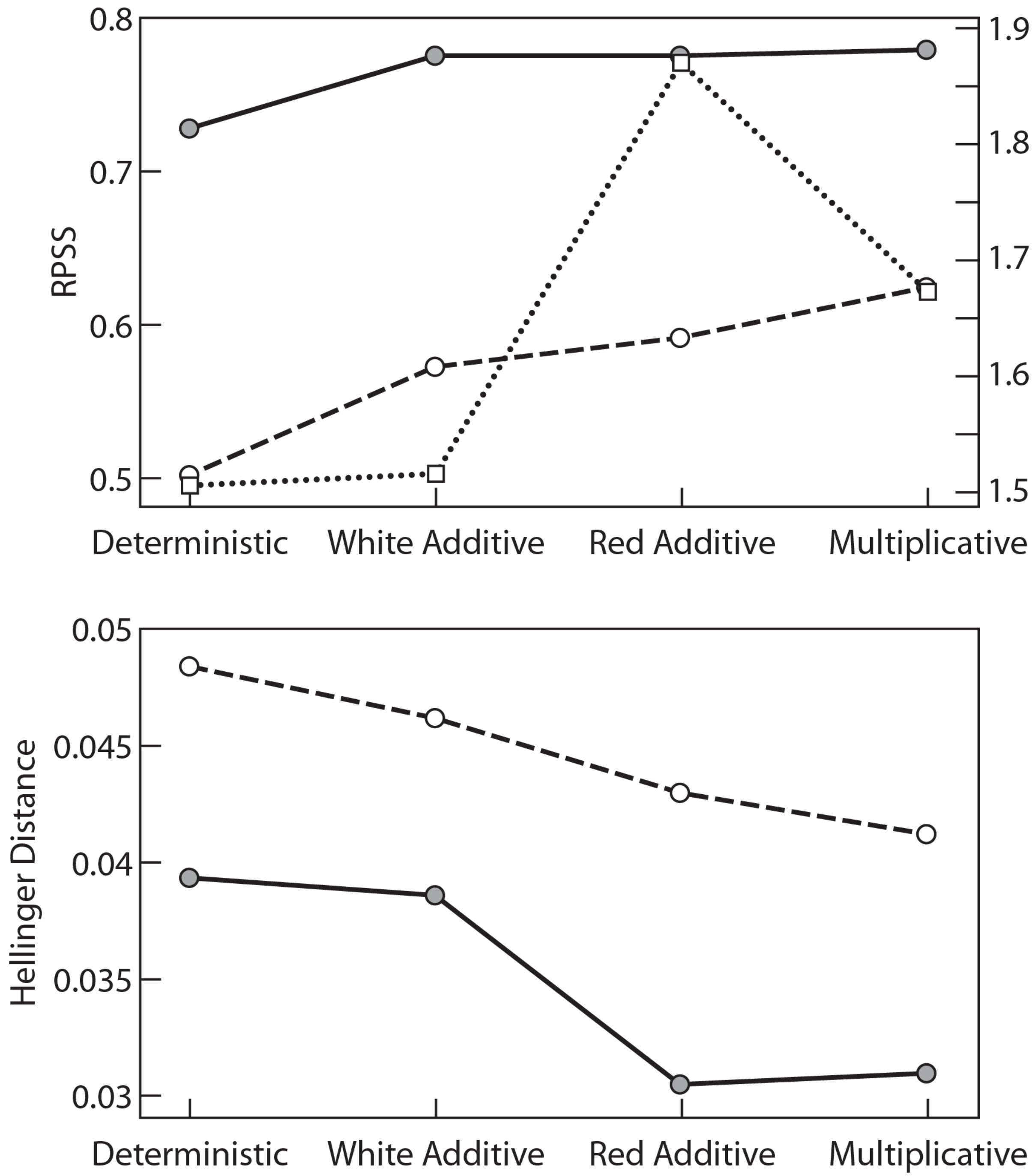

0.8
0.7
0.6
0.5
RPSS
1.9
1.8
1.7
1.6
1.5
RMSE
Deterministic
White Additive
Red Additive
Multiplicative
0.05
0.045
0.04
0.035
0.03
Hellinger Distance
Deterministic
White Additive
Red Additive
Multiplicative

Figure 5

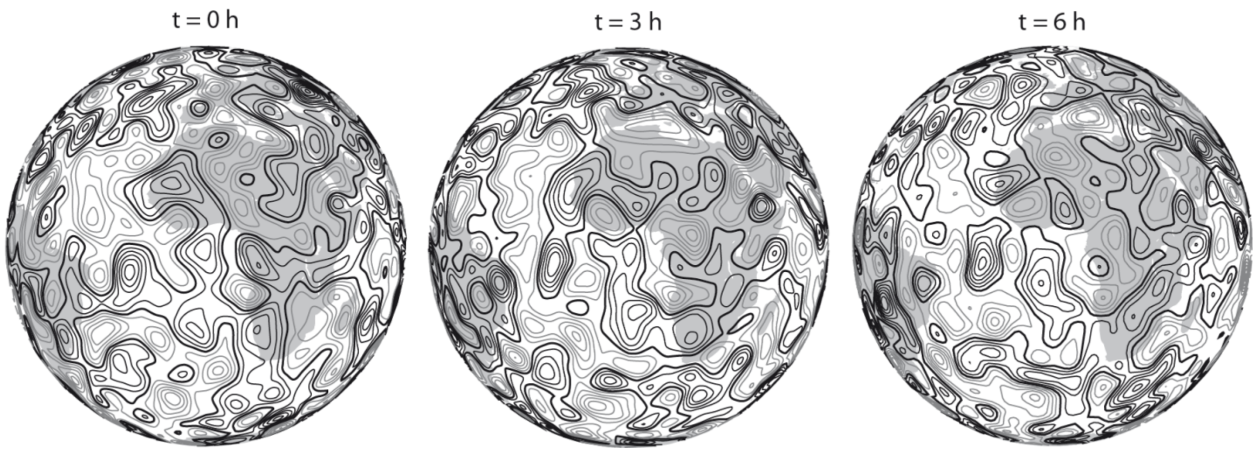
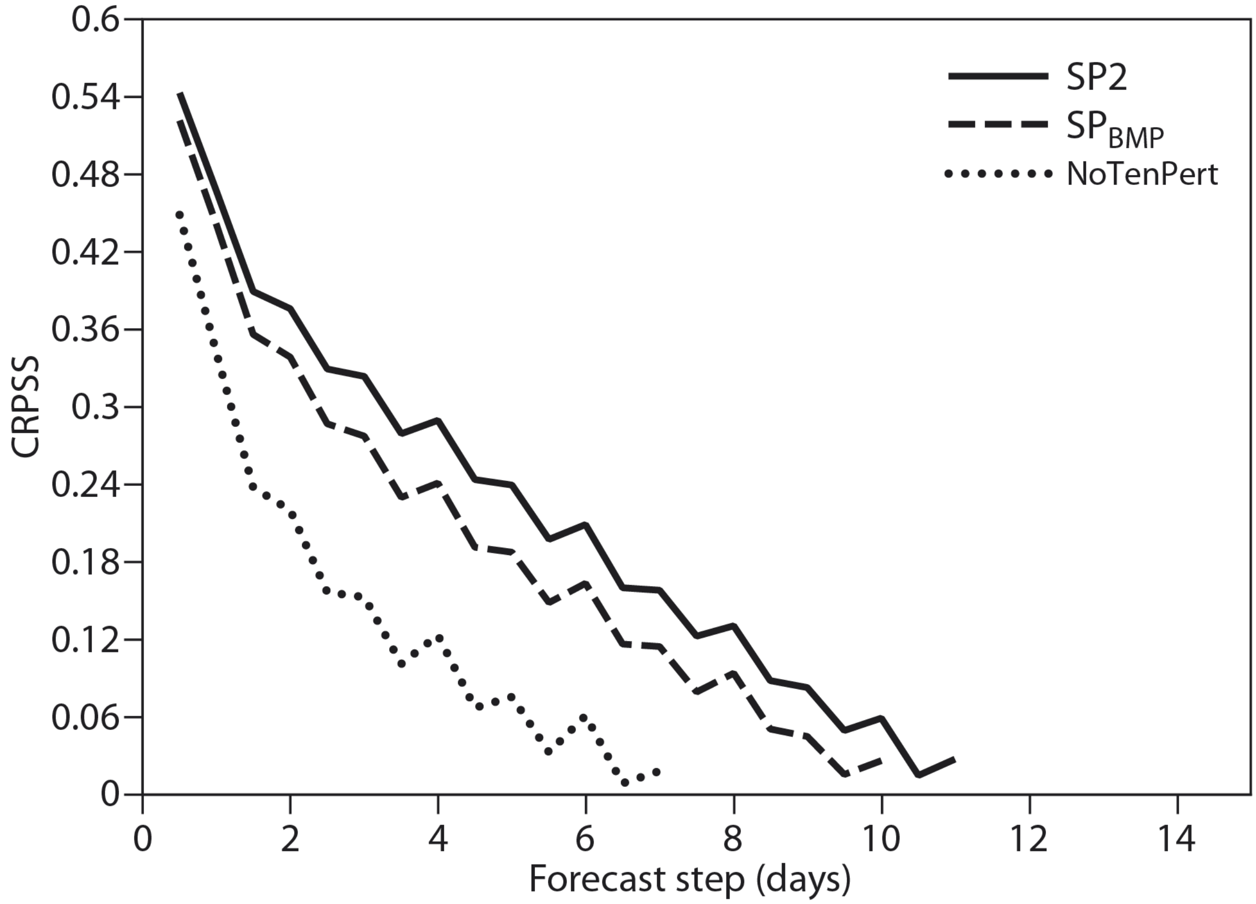


Figure 6



0.6
0.54
0.48
0.42
0.36
0.3
0.24
0.18
0.12
0.06
0
CRPSS
0
2
4
6
8
10
12
14
Forecast step (days)
SP2
SP$_{BMP}$
NoTenPert

Figure 7.

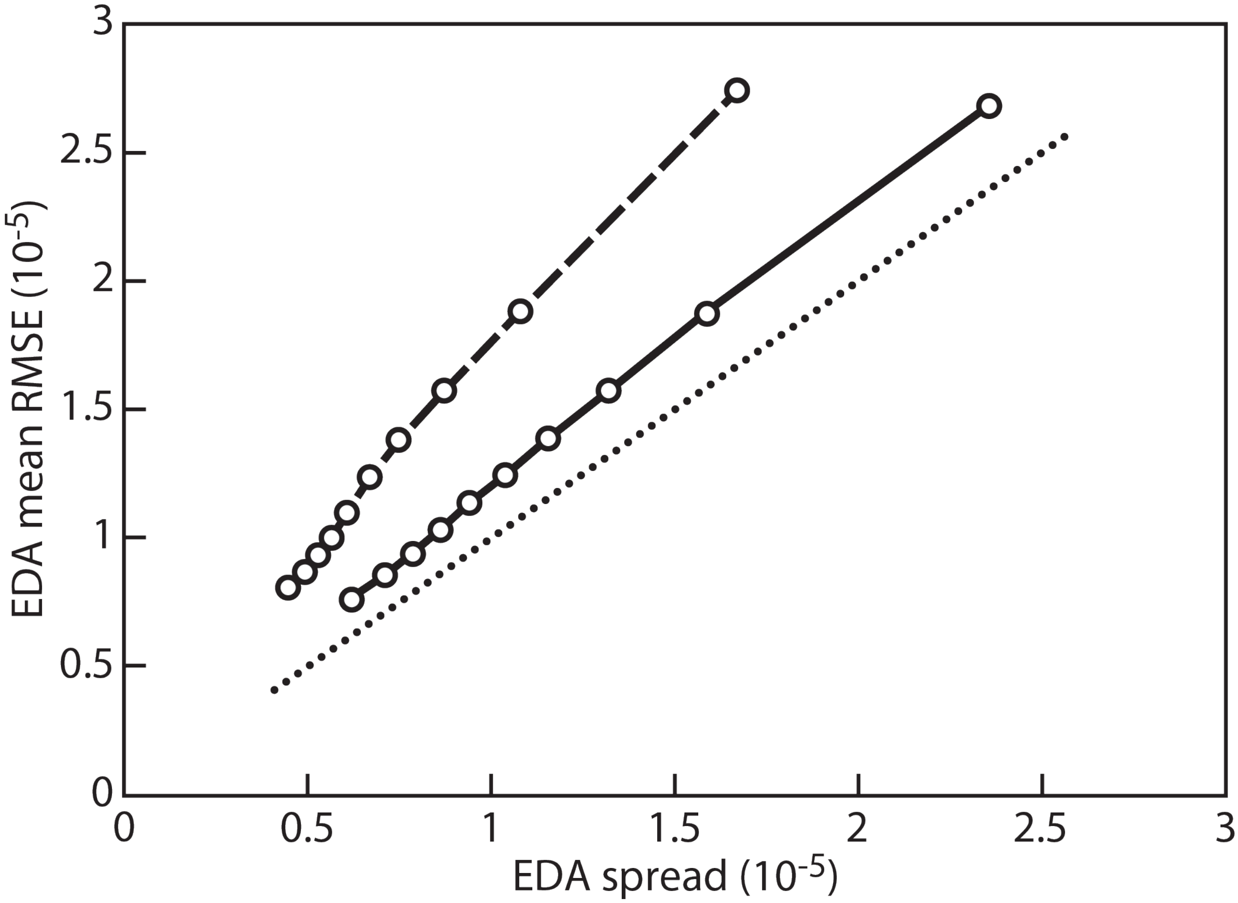


Figure 8

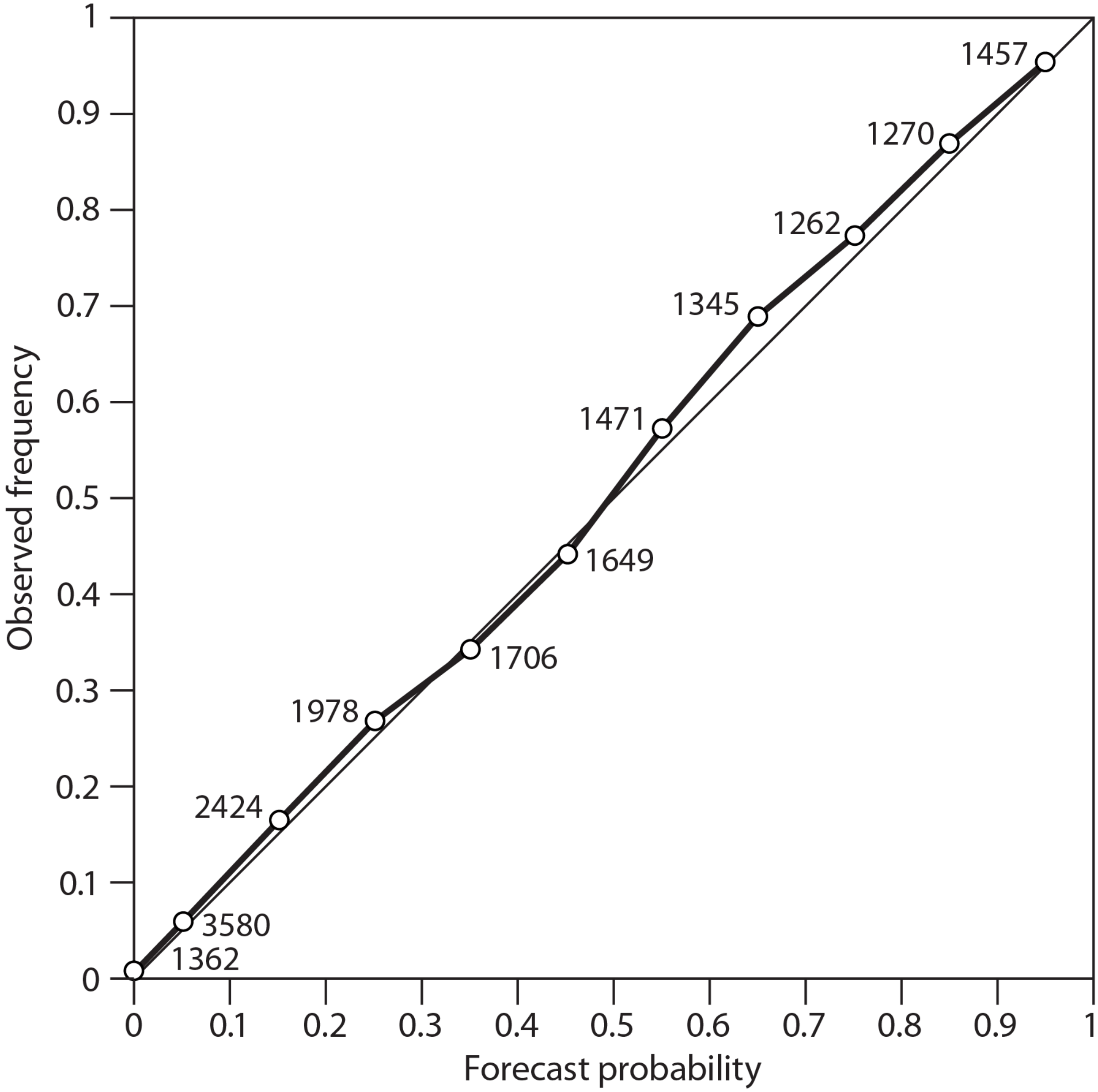

Observed frequency
Forecast probability
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1362
3580
2424
1978
1706
1649
1471
1345
1262
1270
1457

Figure 9.

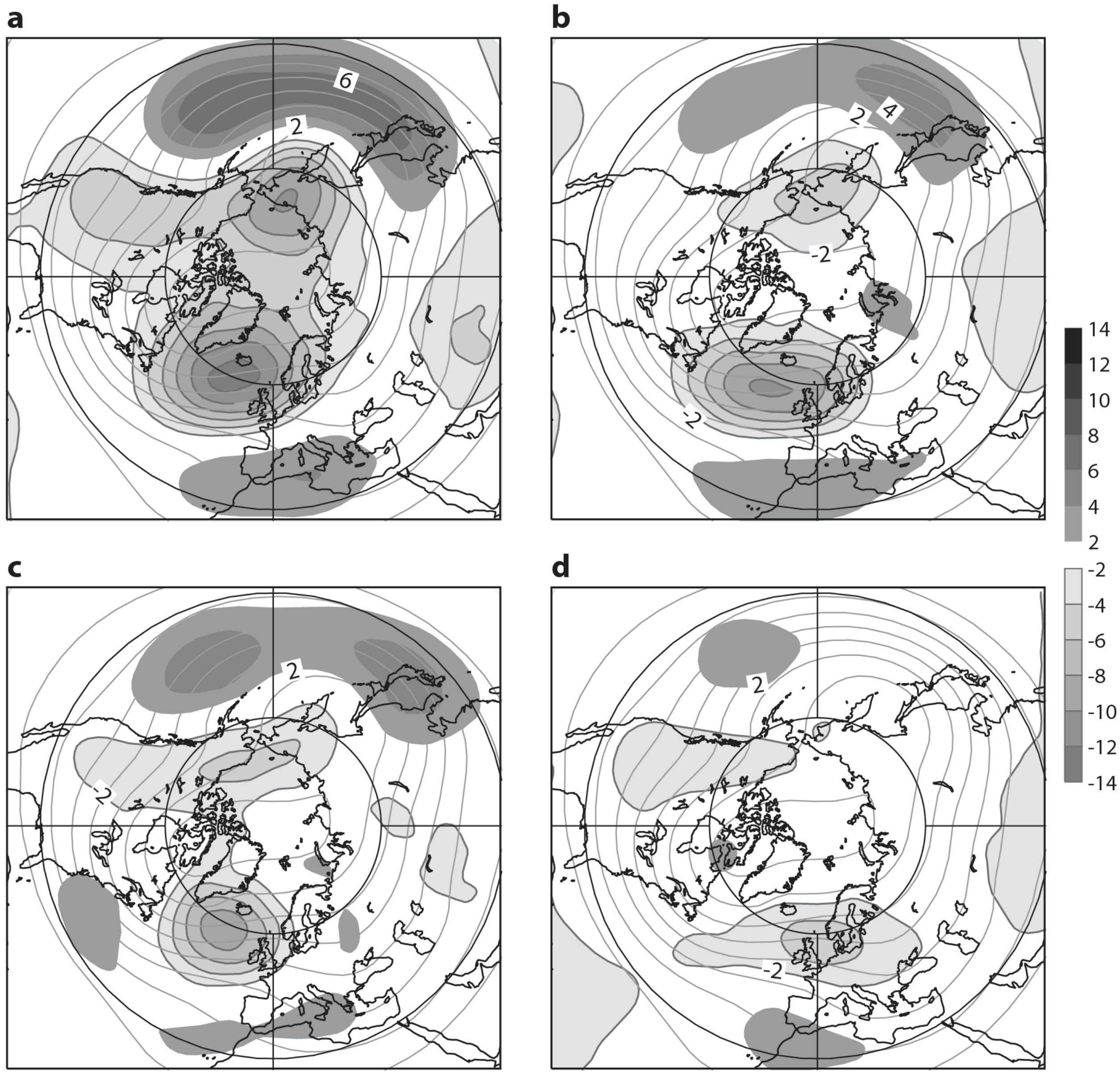


Figure 10.

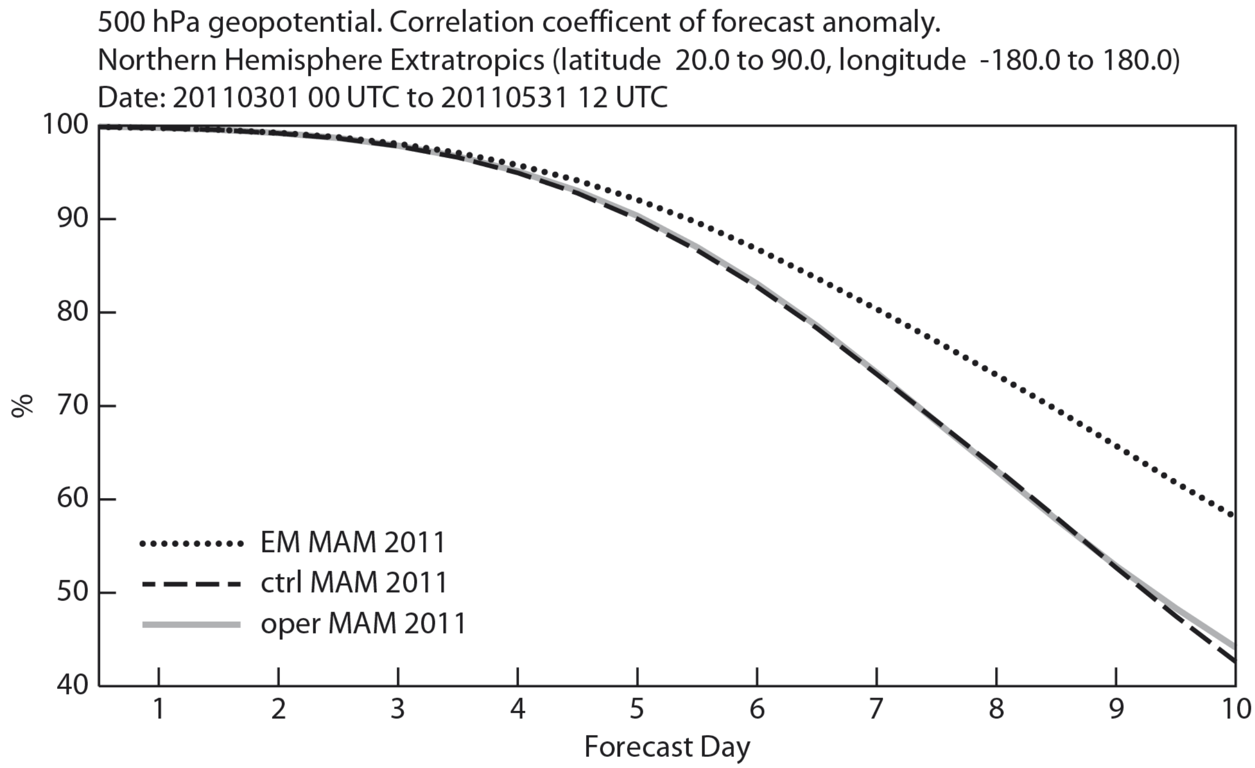

500 hPa geopotential. Correlation coefficent of forecast anomaly.
Northern Hemisphere Extratropics (latitude 20.0 to 90.0, longitude -180.0 to 180.0)
Date: 20110301 00 UTC to 20110531 12 UTC
100
90
80
70
60
50
40
%
1
2
3
4
5
6
7
8
9
10
Forecast Day
EM MAM 2011
ctrl MAM 2011
oper MAM 2011

Figure 11

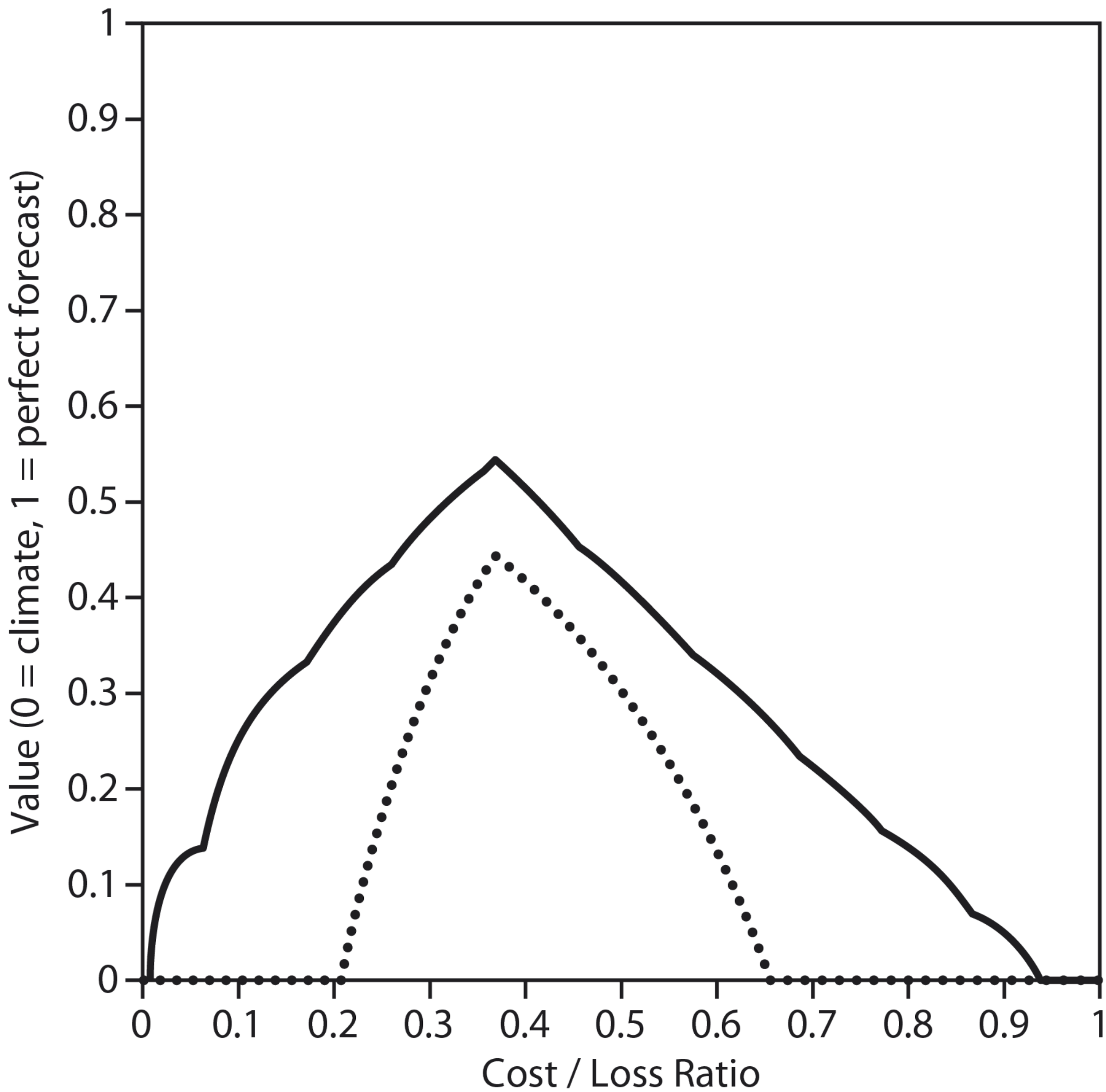

Value (0 = climate, 1 = perfect forecast)
Cost / Loss Ratio
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Figure 12

Table

	T2m				Precipitation			
	May		Nov		May		Nov	
	Cold	Warm	Cold	Warm	Dry	Wet	Dry	Wet
MSE	0.178	**0.195**	0.141	0.159	0.085	0.079	0.080	0.099
SPE	**0.194**	0.192	**0.149**	**0.172**	**0.104**	**0.118**	**0.095**	**0.114**
CTRL	0.147	0.148	0.126	0.148	0.044	0.061	0.058	0.075