1

Predictability of weather and climate: from theory to practice

T. N. Palmer
European Centre for Medium-Range Weather Forecasts, Reading

1.1 Introduction

A revolution in weather and climate forecasting is in progress, made possible by theoretical advances in our understanding of the predictability of weather and climate on the one hand, and by the extraordinary developments in supercomputer technology on the other. Specifically, through ensemble prediction, whose historical development has been documented by Lewis (2005), weather and climate forecasting is set to enter a new era, addressing quantitatively the prediction of weather and climate risk in a range of commercial and humanitarian applications. This chapter gives some background to this revolution, with specific examples drawn from a range of timescales.

1.2 Perspectives on predictability: theoretical and practical

Predictions of weather and climate are necessarily uncertain; our observations of weather and climate are uncertain and incomplete, the models into which we assimilate this data and predict the future are uncertain, and external effects such as volcanoes and anthropogenic greenhouse emissions are also uncertain. Fundamentally, therefore, we should think of weather and climate prediction in terms of equations whose basic prognostic variables are probability densities $\rho(X, t)$, where X denotes

Predictability of Weather and Climate, ed. Tim Palmer and Renate Hagedorn. Published by Cambridge University Press.

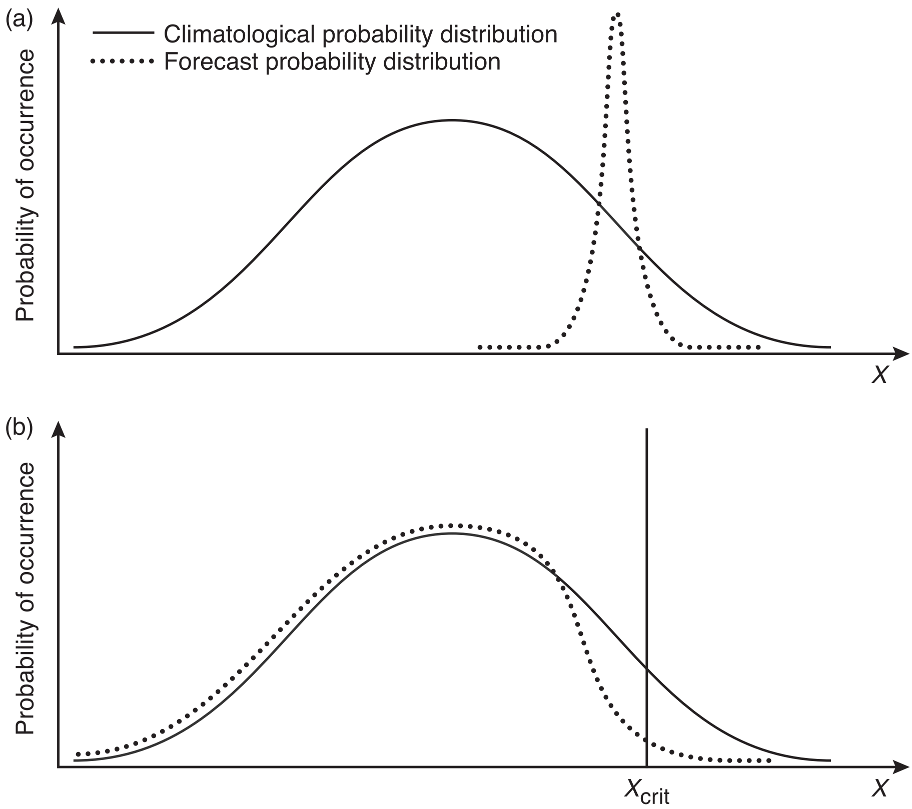

Figure 1.1 Schematic illustration of the climatological probability distribution of some climatic variable X (solid line) and a forecast probability distribution (dotted line) in two different situations. The forecast probability distribution in (a) is obviously predictable. In a theoretical approach to predictability, $\rho(X, t) - \rho_C(X)$ in (b) may not be significantly different from zero overall. However, considered more pragmatically, the forecast probability distribution in (b) can be considered predictable if the prediction that it is unlikely that X will exceed X_{crit} can influence decision-makers.

some climatic variable and t denotes time. In this way, $\rho(X, t)dV$ represents the probability that, at time t, the true value of X lies in some small volume dV of state space. Prognostic equations for ρ, the Liouville and Fokker–Planck equations, are described in Ehrendorfer (this volume). In practice these equations are solved by ensemble techniques, as described in Buizza (this volume).

The question of whether or not X is predictable depends on whether the forecast probability density $\rho(X, t)$ is sufficiently different from some prior estimate $\rho_C(X)$, usually taken as the climatological probability density of X. What do we mean by 'sufficiently different'? One could, for example, apply a statistical significance test to the difference $\rho(X, t) - \rho_C(X)$. On this basis, the hypothetical forecast probability distribution shown as the dotted curve in Figure 1.1(a) is certainly predictable;

but the forecast probability distribution shown in Figure 1.1(b) may well not be predictable.

However, this notion of predictability is a rather idealised one and takes no account of how $\rho(X, t)$ might be used in practice. In a more pragmatic approach to predictability, one would ask whether $\rho(X, t)$ is sufficiently different from $\rho_C(X)$ to influence decision-makers. For example, in Figure 1.1, an aid agency might be interested only in the right-hand tail of the distribution, because disease A only becomes prevalent when $X > X_{\text{crit}}$. On the basis of Figure 1.1(b), the agency may decide to target scarce resources elsewhere in the coming season, since the forecast probability that $X > X_{\text{crit}}$ is rather low.

These two perspectives on the problem of how to define predictability reflect the evolving nature of the study of predictability of weather and climate prediction; from a rather theoretical and idealised pursuit to one which recognises that quantification of predictability is an essential part of operational activities in a wide range of applications. The latter perspective reflects the fact that the full economic value of meteorological predictions will only be realised when quantitatively reliable flow-dependent predictions of weather and climate risk are achievable (Palmer, 2002).

The scientific basis for ensemble prediction is illustrated in Figure 1.2, based on the famous Lorenz (1963) model. Figure 1.2 shows that the evolution of some isopleth of $\rho(X, t)$ depends on starting conditions. This is a consequence of the fact that the underlying equations of motion

$$\dot{X} = F[X] \tag{1.1}$$

are non-linear, so that the Jacobian dF/dX in the linearised equation

$$\frac{d\,\delta X}{dt} = \frac{dF}{dX}\,\delta X \tag{1.2}$$

depends at least linearly on the state X about which Equation (1.1) is linearised. As such, the so-called tangent propagator

$$M(t, t_0) = \exp \int_{t_0}^{t} \frac{dF}{dX}\, dt' \tag{1.3}$$

depends on the non-linear trajectory $X(t)$ about which the linearisation is performed. Hence, the evolved perturbations

$$\delta X(t) = M(t, t_0)\,\delta X(t_0) \tag{1.4}$$

depend not only on $\delta X(t_0)$, but also on the region of phase space through which the underlying non-linear trajectory passes.

It is of interest to note that the formal solution of the Liouville equation, which describes the evolution of $\rho(X, t)$ arising from initial error only (Ehrendorfer, this volume, Eq. (4.49)), can be written using the tangent propagator (for all time in

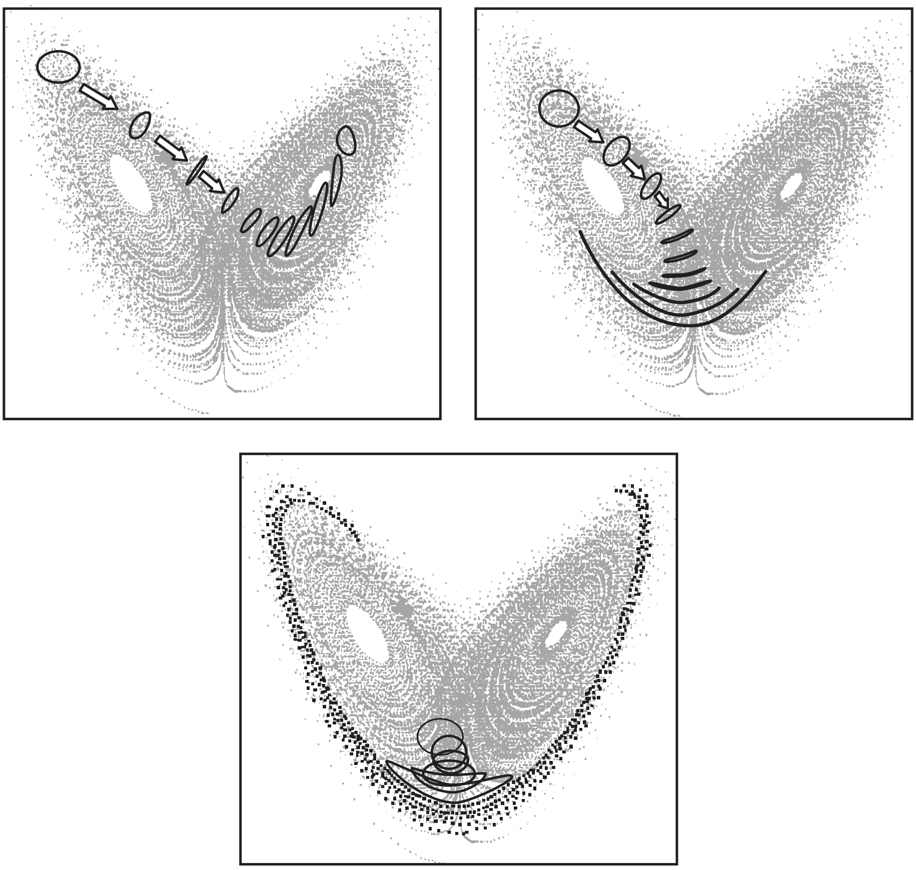

Figure 1.2 Finite time ensembles of the Lorenz (1963) system illustrating the fact that in a non-linear system, the evolution of the forecast probability density $\rho(X, t)$ is dependent on initial state.

the future, not just the time for which the tangent-linear approximation is valid). Specifically

$$\rho(X, t) = \rho(X', t_0)/|\det M(t, t_0)| \tag{1.5}$$

where X' corresponds to the initial state which, under the action of Eq. (1.1), evolves into the state X at time t. Figure 1.2 shows solutions to Eq. (1.5) using an ensemble-based approach.

To illustrate the more practical implications of the fact that $\rho(X, t)$ depends on initial state, I want to reinterpret Figure 1.2 by introducing you to Charlie, a builder by profession, and a golfing colleague of mine! Charlie, like many members of my golf club, takes great pleasure in telling me when (he thinks) the weather forecast has gone wrong. This is mostly done in good humour, but on one particular occasion Charlie was in a black mood. ‘I have only four words to say to you,’ he announced, 'How do I sue?' I looked puzzled. He continued: 'The forecast was for a night-time minimum temperature of five degrees. I laid three thousand square yards of concrete. There was a frost. It's all ruined. I repeat – how do I sue?'

If only Charlie was conversant with Lorenz (1963) I could have used Figure 1.2 to illustrate how in future he will be able to make much more informed decisions about when, and when not, to lay concrete! Suppose the Lorenz equations represent part of an imaginary world inhabited by builders, builders' customers, weather forecasters and lawyers. In this Lorenz world, the weather forecasters are sued if the forecasts are wrong! The weather in the Lorenz world is determined by the Lorenz (1963) equations where all states on the right-hand lobe of the attractor are 'frosty' states, and all states on the left-hand lobe of the attractor are 'frost-free' states. In this imaginary world, Charlie is planning to lay a large amount of concrete in a couple of days' time. Should he order the ready-mix concrete lorries to the site? He contacts the Lorenzian Meteorological Office for advice. On the basis of the ensemble forecasts in the top left of Figure 1.2 he clearly should not – all members of the ensemble predict frosty weather. On the basis of the ensemble forecasts in the bottom left of Figure 1.2 he also should not – in this case it is almost impossible to predict whether it will be frosty or not. Since the cost of buying and laying concrete is significant, it is not worth going ahead when the risk of frost is so large.

How about the situation shown in the top right of Figure 1.2? If we took the patronising but not uncommon view that Charlie, as a member of the general public, would only be confused by a probability forecast, then we might decide to collapse the ensemble into a consensus (i.e. ensemble-mean) prediction. The ensemble-mean forecast indicates that frost will not occur. Perhaps this is equivalent to the real-world situation that got Charlie so upset. Lorenzian forecasters, however, will be cautious about issuing a deterministic forecast based on the ensemble mean, because, in the Lorenz world, Charlie can sue!

Alternatively, the forecasters could tell Charlie not to lay concrete if there is even the slightest risk of frost. But Charlie will not thank them for that either. He cannot wait forever to lay concrete since he has fixed costs, and if he doesn't complete this job, he may miss out on other jobs. Maybe Charlie will never be able to sue, but neither will he bother obtaining the forecasts from the Lorenzian Meterorological Office.

Suppose Charlie's fixed costs are C, and that he loses L by laying concrete when a frost occurs. Then a logical decision strategy would be to lay concrete when the ensemble-based estimate of the probability of frost is less than C/L. The meteorologists don't know Charlie's C/L, so the best they can do is provide him with the full probability forecast, and allow him to decide whether or not to lay concrete.

Clearly the probability forecast will only be of value to Charlie if he saves money using these ensemble forecasts. This notion of 'potential economic value' (Murphy, 1977; Richardson, this volume) is conceptually quite different from the notion of skill (in the meteorological sense of the word), since value cannot be assessed by analysing meteorological variables alone; value depends also on the user's economic parameters.

The fact that potential economic value does not depend solely on meteorology means that we cannot use meteorological skill scores alone if we want to assess whether one forecast system is more valuable than another (e.g. to Charlie). This is relevant to the question of whether it would be better to utilise computer resources to increase ensemble size or increase model resolution. As discussed in Palmer (2002), the answer to this question depends on C/L. For users with small C/L, more value may accrue from an increase in ensemble size (since decisions depend on whether or not relatively small probability thresholds have been reached), whilst for larger C/L more value may accrue from the better representation of weather provided by a higher-resolution model.

In the Lorenz world, Charlie never sues the forecasters for 'wrong' forecasts. When the forecast is uncertain, the forecasters say so, and with precise and reliable estimates of uncertainty. Charlie makes his decisions based on these forecasts and if he makes the wrong decisions, only he, and lady luck, are to blame!

1.3 Why are forecasts uncertain?

Essentially, there are three reasons why forecasts are uncertain: uncertainty in the observations used to define the initial state, uncertainty in the model used to assimilate the observations and to make the forecasts, and uncertainty in 'external' parameters.

Let's consider the last of these uncertainties first. For example, the aerosol content of the atmosphere can be significantly influenced by volcanic eruptions, which are believed to be unpredictable more than a few days ahead. Also, uncertainty in the change in atmospheric CO_2 over the coming decades depends on which nations sign agreements such as the Kyoto protocol.

In principle, perhaps, 'stochastic volcanoes' could be added to an ensemble prediction system – though this seems a rather fanciful idea. Also, uncertainties in humankind's activities can, perhaps, be modelled by coupling our physical climate model to an econometric model. However, we will not deal further with such uncertainties of the 'third kind' but rather concentrate on the first two.

1.3.1 Initial uncertainty

At ECMWF, for example, the analysed state X_{a} of the atmosphere is found by minimising the cost function

$$J(X) = \tfrac{1}{2}(X - X_{\mathrm{b}})^T B^{-1} (X - X_{\mathrm{b}}) + \tfrac{1}{2}(HX - Y)^T O^{-1} (HX - Y) \quad (1.6)$$

where X_{b} is the background state, B and O are covariance matrices for the probability density functions (pdf) of background error and observation error, respectively, H is

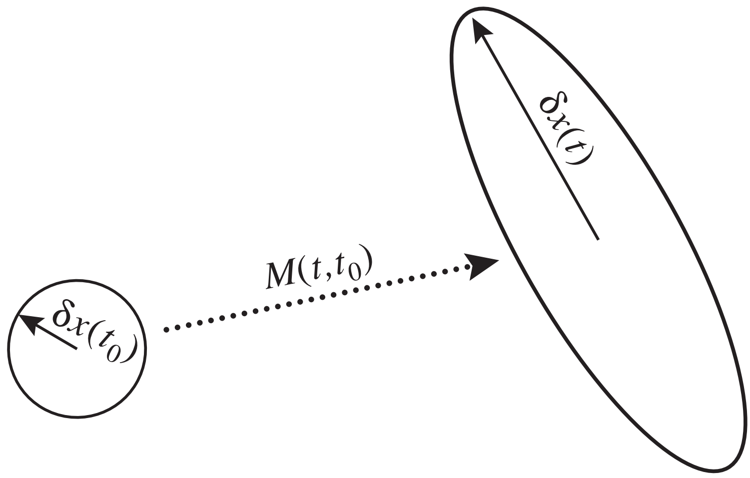

Figure 1.3 Isopleths of probability that the region enclosed by the isopleths contains truth at initial and forecast time. The associated dominant singular vector at initial and final time is also shown.

the so-called observation operator, and Y denotes the vector of available observations (e.g. Courtier *et al.*, 1998). The Hessian

$$\nabla\nabla J = B^{-1} + H^T O^{-1} H \equiv A^{-1} \tag{1.7}$$

of J defines the inverse analysis error covariance matrix.

Figure 1.3 shows, schematically, an isopleth of the analysis error covariance matrix, and its evolution under the action of the tangent propagator M (see Eqs. 1.3 and 1.4). The vector pointing along the major axis at forecast time corresponds to the leading eigenvector of the forecast error covariance matrix. Its pre-image at initial time corresponds to the leading singular vector of M, determined with respect to unit norm in the metric given by A. The singular vectors of M correspond to the eigenvectors of $M^T M$ in the generalised eigenvector equation

$$M^T M\, \delta x(t_0) = -\lambda A^{-1} \delta x(t_0). \tag{1.8}$$

Given pdfs of uncertainty based on Eq. (1.6), we can in principle perform a Monte Carlo sampling of the Hessian-based initial pdf and produce an ensemble forecast system based on this initial sampling.

There are three reasons for not adopting this strategy.

Firstly, there is the so-called 'curse of dimensionality'. The state space of a weather prediction model has about 10^7 dimensions. Many of these dimensions are not dynamically unstable (i.e. are not associated with positive singular values). In this sense, a random sampling of the initial probability density would not be a computationally efficient way of estimating the forecast probability density. This point was made explicitly in Lorenz's analysis of his 28-variable model (Lorenz, 1965):

> If more realistic models . . . also have the property that a few of the eigenvalues of MM^T are much larger than the remaining, a study based upon a small ensemble of initial errors should . . . give a reasonable estimate of the growth rate of random error. . . . It would appear, then, that the best use could be made of computational time by choosing only a small number of error fields for superposition upon a particular initial state.

Studies of realistic atmospheric models show that the singular values of the first 20–30 singular vectors are indeed much larger than the remainder (Molteni and Palmer, 1993; Buizza and Palmer, 1995; Reynolds and Palmer, 1998).

The second reason for not adopting a Monte Carlo strategy is that in practice Eq. (1.6) only provides an estimate of part of the actual initial uncertainty; there are other sources of initial uncertainty that are not well quantified – what might be called the 'unknown unknowns'. Consider the basic notion of data assimilation: to assimilate observations that are either made at a point or over a pixel size of kilometres into a model whose smallest resolvable scale is many hundreds of kilometres (bearing in mind the smallest resolvable scale will be many times the model grid). Now sometimes these point or pixel observations may be representative of circulation scales that are well resolved by the model (e.g. if the flow is fairly laminar at the time the observation is made); on other occasions the observations may be more representative of scales which the model cannot resolve (e.g. if the flow is highly turbulent at the time the observation is made, or if the observation is sensitive to small-scale components of the circulation, as would be the case for humidity or precipitation).

In the latter case, the practice of using simple polynomial interpolation in the observation operator H in Eq. (1.6) to take the model variable X to the site of the observation, is likely to be poor. However, this is not an easily quantified uncertainty– since, ultimately, the uncertainty relates to numerical truncation error in the forecast model (see the discussion below). Similarly, consider the problem of quality control. An observation might be rejected as untrustworthy by a quality-control procedure if the observation does not agree with its neighbours and is different from the background (first-guess) field. Alternatively, the observation might be providing the first signs of a small-scale circulation feature, poorly resolved by either the model or the observing network. For these types of reason, a Monte Carlo sampling of a pdf generated by Eq. (1.6) is likely to be an underestimate of the true uncertainty.

The third reason for not adopting a Monte Carlo strategy is not really independent of the first two, but highlights an issue of pragmatic concern. Let us return to Charlie, as discussed above. Charlie is clearly disgruntled by the occasional poor forecast of frost, especially if it costs him money. But just imagine how much more disgruntled he would be, having invested time and money to adapt his decision strategies to a new weather risk service based on the latest, say, Multi-Centre Ensemble Forecast System, if no member of the new ensemble predicts severe weather, and severe

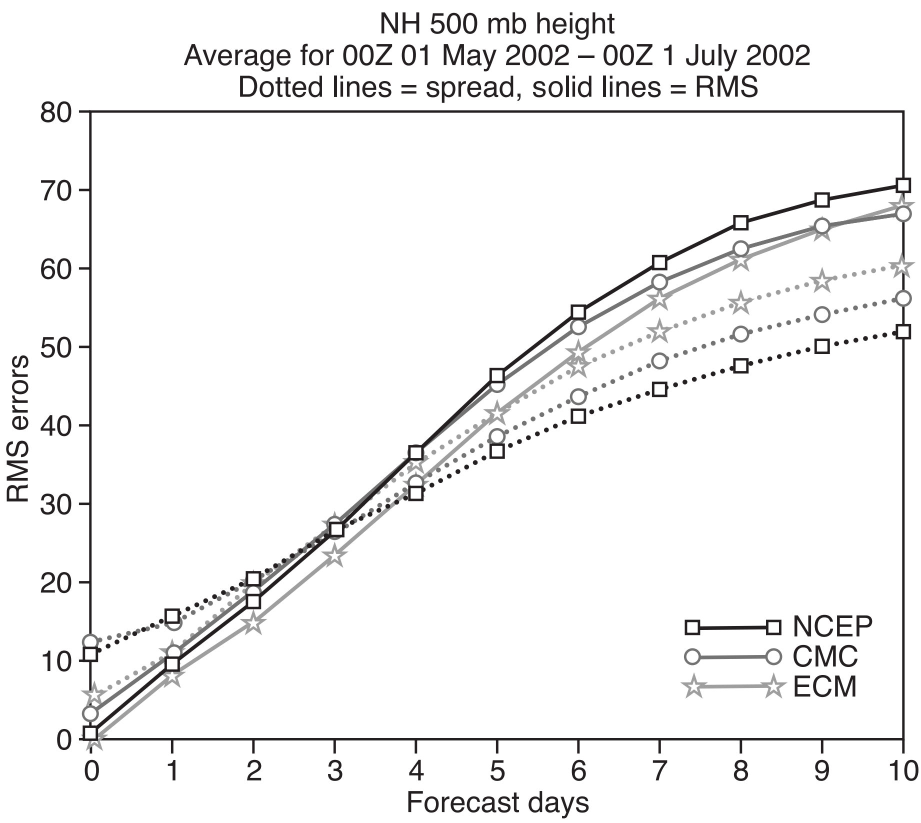

Figure 1.4 May–July 2002 average root-mean-square (rms) error of the ensemble-mean (solid lines) and ensemble standard deviation (dotted lines) of the ECMWF, NCEP and MSC ensemble forecast systems. Values refer to the 500 hPa geopotential height over the Northern Hemisphere latitudinal band 20–80 N. From Buizza *et al.* (2003, 2005).

weather occurs! Just one failure of this sort will compromise the credibility of the new system.

To take this into account, a more conservative approach to sampling initial perturbations is needed, conservative in the sense of tending towards sampling perturbations that are likely to have significant impact on the forecast.

For these three reasons (together with the fact that instabilities in the atmosphere are virtually never of the normal-mode type: Palmer, 1988; Molteni and Palmer, 1993; Farrell and Ioannou, this volume and Ioannou and Farrell, this volume), the initial perturbations of the ECMWF ensemble prediction system are based on the leading singular vectors of *M* (Buizza, this volume).

The relative performance of the singular-vector perturbations can be judged from Figure 1.4 (Buizza *et al.*, 2003), based on a comparison of ensemble prediction systems at ECMWF (Palmer *et al.*, 1993; Molteni *et al.*, 1996), NCEP (US National

Centers for Environmental Prediction; Toth and Kalnay, 1993) and MSC (Meteorological Service of Canada; Houtekamer *et al.*, 1996); the latter systems based on bred vectors and ensemble data assimilation respectively. The solid lines show the ensemble-mean root-mean-square error of each of the three forecast systems, the dashed lines show the spread of the ensembles about the ensemble mean. At initial time, both NCEP and MSC perturbations are inflated in order that the spread and skill are well calibrated in the medium range. The growth of perturbations in the ECMWF system, by contrast, appears to be more realistic, and overall the system appears better calibrated to the mean error.

1.3.2 Model uncertainty

From the discussion in the last section, part of the reason initial conditions are uncertain is that (e.g. in variational data assimilation) there is no rigorous operational procedure for comparing a model state X with an observation Y. The reason that there is no rigorous procedure is directly related to the fact that the model cannot be guaranteed to resolve well the circulation or weather features that influence the observation. In this respect model error is itself a component of initial error. Of course, model error plays an additional role as one integrates, forward in time, the model equations from the given initial state.

Unfortunately, there is no underlying theory which allows us to estimate the statistical uncertainty in the numerical approximations we make when attempting to integrate the equations of climate on a computer. Moreover, an assessment of uncertainty has not, so far, been a requirement in the development of subgrid parametrisations.

Parametrisation is a procedure to approximate the effects of unresolved processes on the resolved scales. The basis of parametrisation, at least in its conventional form, requires us to imagine that within a grid box there exists an ensemble of incoherent subgrid processes in secular equilibrium with the resolved flow, and whose effect on the resolved flow is given by a deterministic formula representing the mean (or bulk) impact of this ensemble. Hence a parametrisation of convection is based on the notion of the bulk effect of an incoherent ensemble of convective plumes within the grid box, adjusting the resolved scales back towards convective neutrality; a parametrisation of orographic gravity-wave drag is based on the notion of the bulk effect of an incoherent ensemble of breaking orographic gravity waves applying a retarding force to the resolved scale flow.

A schematic representation of parametrisation in a conventional weather or climate prediction model is shown in the top half of Figure 1.5. Within this framework, uncertainties in model formulation can be represented in the following hierarchical form:

- the multimodel ensemble whose elements comprise different weather or climate prediction models;