Crowley G (1991) *Dynamics of the Earth's thermosphere*. US National Report to International Union of Geodosy and Geophysics 1987–1990, *Review of Geophysics, Supplement*, pp. 1143–1165.
Hedin AE and Carignan GR (1985) Morphology of thermospheric composition variations in the quiet polar mesosphere from Dynamics Explorer measurements. *Journal of Geophysical Research* 90: 5269–5277.
Killeen TL (1987) Energetics of the Earth's thermosphere. *Reviews of Geophysics* 25: 433–454.
Killeen TL and Roble RG (1988) Thermospheric dynamics: contributions from the first 5 years of the Dynamics Explorer program. *Reviews of Geophysics* 26: 329–367.
Rees MH (1989) *Physics and Chemistry of the Upper Atmosphere*. Cambridge: Cambridge University Press.
Rishbeth H (2001) The centenary of solar-terrestrial physics. *Journal of Atmospheric and Solar–Terrestrial Physics* 63: 1883–1890.
Rishbeth H and Garriott OK (1969) *Introduction to Ionospheric Physics*. London: Academic Press.
Roble RG (1986) Chemistry in the thermosphere and ionosphere. *Chemical Engineering News*. 64(24): 23–38.
Roble RG (1992) The polar lower thermosphere. *Planetary Space Science* 40: 271–297.
Rocket P (1952) Pressure, densities and temperatures in the upper atmosphere. *Physics Review* 88: 1027–1032.
Schunk RW and Nagy AF (2000) *Ionospheres, Physics, Plasma Physics and Chemistry.* Cambridge: Cambridge University Press.

PREDICTABILITY AND CHAOS

L A Smith, London School of Economics, London, UK

Introduction

Chaos is difficult to quantify. The nonlinear dynamic that gives rise to chaos links forecasting on the shortest time scales of interest with behavior over the longest time scales. In addition, the statistics which quantify chaos forbid an appeal to most of the traditional simplifications employed in statistical estimation. Nevertheless, both the concept of chaos in physical systems and its technical relative in mathematics have had a significant impact on meteorological aims and methods. This is particularly true in forecasting. Chaos implies sensitivity to initial conditions: small uncertainties in the current state of the system will grow exponentially-on-average. Yet, as discussed below, neither this exponential-on-average growth nor the Lyapunov exponents that quantify it reflect macroscopic predictability. The limits chaos places on predictability are much less severe than generally imagined. Predictability is more clearly quantified through traditional statistics, like uncertainty doubling times. These statistics will vary from day to day, depending on the current state of the atmosphere. Maintaining the uncertainty in the initial state within the forecast is a central goal of ensemble forecasting (*see* **Weather Prediction:** Ensemble Prediction). Achieving the ultimate goal of meaningful probability forecasts for meteorological variables would be of great societal and economic value. Fundamental limitations in the realism of models of the atmosphere will limit our ability to make probability (PDF) forecasts, just as uncertainty in the initial condition limits the utility of single forecasts even if the model is perfect.

Initially, it appears that chaotic systems will be unpredictable, and this is true in that it is not possible to make extremely accurate forecasts in the very distant future. Yet chaos *per se* does not imply one cannot sometimes make accurate forecasts well into the medium range. And perhaps just as importantly, with a perfect model one can determine which of these forecasts will be informative and which will not at the time they are made. As we shall see below, both the American and the European weather forecast centers have adopted this strategy operationally, with the aim of quantifying day-to-day variations in the likely range of future meteorological variables. Quantifying this range can be of significant value even without a perfect model. Although accurate probability forecasts are likely to require a perfect model, current operational ensemble prediction systems already provide economically valuable information on the uncertainty of numerical weather prediction (NWP) well into 'week two', and research programs on seasonal time scales are underway.

Why is perfect foresight of the future state of the atmosphere impossible? First, it should be no surprise that if our knowledge of the present is uncertain then our knowledge of the future will also be uncertain; the question of prediction then turns to how to best quantify the dynamics of uncertainty. Here the full implications of chaos, or more properly of nonlinearity, mix what are often operationally distinct tasks: observing strategy, data assimilation, state estimation,

ensemble formation, forecast evaluation, and model improvement. This complicates the forecasting problem beyond the limits in which classical prediction theory was developed. First the traditional goal of a single best first guess (BFG) forecast with a minimal least-squares error is not a viable aim in this scenario, nor is a least squares solution desired. Indeed (as discussed below) this approach would reject the perfect model which generated the observations! Second, current models are not perfect. The term 'model inadequacy' is used to summarize imperfections in a given model and the entire model class from which that model is a member. Model improvement and the search for a superior model class, along with the investigation of more relevant measures of model skill, are areas of active research.

A Mathematical Framework for Modelling Dynamical Systems

Chaos is a phenomenon found in many nonlinear mathematical models. While one should never forget the distinction between the model and the system being modeled, precise mathematical definitions are more easily made within the perfect model scenario (PMS). Within this useful fiction, one assumes that the model in-hand is itself the physical system of interest. Before moving back to forecasts of the real world, of course, one must recover from this self-deception.

Given a model, an initial condition is simply an assignment of values to all model variables at a particular starting time. Thus the initial condition reflects the current state of the model: it is a vector $\mathbf{x}(t)$ which specifies the value of every variable in the model at time t. For the classical model of a pendulum, the state consists of two variables (the angle and the angular velocity). These two numbers completely define the current state of the model, and so the model is called two-dimensional, or, equivalently, said to have a two-dimensional state space. The famous Lorenz model of 1963 is three-dimensional, as there are three variables: $\mathbf{x} = (x, y, z)$. These low-dimensional systems should be contrasted with the state of an operational NWP model, which may consist of over 10 million variables.

The sequence of states a dynamical system passes through defines the history of the system; this sequence is called a trajectory. For deterministic systems, any single state $\mathbf{x}$ along a trajectory defines all future states of the system. For the classical pendulum, these solutions can be written down analytically; but for all but the simplest nonlinear systems only numerical solutions exist. Thus it is difficult to prove that a given realization of the Lorenz model is chaotic, and even more so modern NWP models.

Chaos

Mathematically, a chaotic system is a deterministic system in which (infinitesimally) small uncertainties in the initial condition will grow, on average, exponentially fast. The average is taken over (infinitely) long periods of time. Of course, finite uncertainties often grow rapidly as well, in which case any uncertainty in the initial condition will limit predictability in terms of a single BFG of the future state, even with a perfect model. But inasmuch as it is defined by the behaviour of (infinitesimally) small uncertainties of (infinitely) long periods of time, chaos *per se* places no limits of the growth of finite uncertainty over a finite period of time. Chaotic systems are often said to display sensitive dependence on initial condition (SDIC), a technical term for systems in which states initially very close together tend to end up very far apart, eventually.

Suppose the true state of the system is $\tilde{\mathbf{x}}$: what is the behaviour of a near-by solution $\hat{\mathbf{x}}$ where $\hat{\mathbf{x}} = \tilde{\mathbf{x}} + \boldsymbol{\varepsilon}$? In the pendulum, a small initial $\boldsymbol{\varepsilon}$ grows slowly, if at all. In a chaotic system, the magnitude of $\boldsymbol{\varepsilon}$ will grow exponentially-on-average; yet this does not imply that the actual magnitude of $\boldsymbol{\varepsilon}$ ever grows exponentially in time. Indeed, since $\boldsymbol{\varepsilon}$ is a distance and $|\boldsymbol{\varepsilon}| > 0$ for any given value of t, one can always define a value

$$\lambda = \frac{1}{t} \log\left[|\boldsymbol{\varepsilon}(t)|/|\boldsymbol{\varepsilon}(0)|\right]$$

for any system, chaotic or not! In this case, observing a value of $\lambda > 0$ for finite t does not even suggest exponential growth. Statistics like λ become interesting only when they approach a constant as $t \to \infty$; by definition, chaos reflects properties only in this limit.

Clearly, chaos includes special cases where magnitude of $\boldsymbol{\varepsilon}$ is growing uniformly, say doubling every second; but it also allows the more common case where the growth of $\boldsymbol{\varepsilon}$ is a function of the state $\mathbf{x}$ and hence changes with time. In general, the growth will not be uniform in time. In fact in some chaotic systems, including the Lorenz 1963 model, there are regions of the state space in which every $\boldsymbol{\varepsilon}$ will decrease! Such regions are said to represent 'return of skill' as forecasts become more accurate in the least-squares sense as time passes.

For instance, consider the case where, on average, half the time $\boldsymbol{\varepsilon}$ is constant and the other half of the time it grows by a factor of four. This will yield in the same exponential-on-average growth as doubling every time step, yet there will be times when prediction is easy. Or for a more extreme case, consider where $\boldsymbol{\varepsilon}$

shrinks by a factor of two 9 times out of 10, but once in 10 grows by a factor of 2^{19} (about half a million); again this is exponential-on-average growth equivalent to doubling every time step. The question, then, is whether or not these variations of predictability can be identified in advance. As discussed below, ensemble forecasts aim to do just that.

The Statistics of Chaos: Lyapunov Exponents and Doubling Times

Given a deterministic system which remains in a bounded region of state space, chaos is defined by a statistic called the Lyapunov exponent. In a one-dimensional system, the Lyapunov exponent reflects the logarithm of the (geometric) average growth of infinitesimal uncertainties. In the limits $\boldsymbol{\varepsilon} \to 0$ and $t \to \infty$, the geometric average of $|\boldsymbol{\varepsilon}(t+1)|/|\boldsymbol{\varepsilon}(t)|$ defines the Lyapunov exponent, usually called λ. Note that nowhere is there the suggestion that $|\boldsymbol{\varepsilon}(t)| \approx |\boldsymbol{\varepsilon}(0)|e^{\lambda t}$. A system will have as many Lyapunov exponents as the dimension of the state space. The largest Lyapunov exponent is often called the 'leading' Lyapunov exponent, and if the leading Lyapunov exponent of a bounded deterministic system is positive, then the system is 'chaotic'. Hence the three systems noted above – one in which $\boldsymbol{\varepsilon}$ doubles every time step, and the other two in which $\boldsymbol{\varepsilon}$ sometimes grows and sometimes does not – each have the same Lyapunov exponent. Typically, the logarithm is taken with base two, so that if, for example, the uncertainty doubles every second, the Lyapunov exponent is one bit per second, thus the Lyapunov exponent is an average rate.

Also note that for every state **x** on the attractor, there corresponds a unique direction in state space associated with the leading Lyapunov exponent. If the state space has a dimension greater than one, then estimating Lyapunov exponents involves matrix multiplication along a trajectory. Matrix multiplication does not commute, thus when dealing with statistics like Lyapunov exponents one has to apply multiplicative ergodic theorems; this makes many of their properties appear counterintuitive. Many intuitive methods of statistical estimation fail when applied to chaotic systems. None of this is surprising, since most statistical intuition is developed in the context of more familiar ergodic theorems.

If the sum of all the Lyapunov exponents is negative, then the trajectories will evolve towards a set whose dimension is less than the dimension of the state space; this set is called an 'attractor'. An attractor may be something as common as a fixed point, a periodic orbit, or a torus; in such cases the attractor has simple geometry. Alternatively, an attractor may have a strange geometry: it may consist of a fractal set of points in state space, in which case it is called a 'strange attractor'. Note that being chaotic reflects a property of the dynamics of the system, while strangeness reflects the geometry of the set on which the system evolves, not the dynamics of the evolution itself. Given a chaotic system with a strange attractor, the choice of initial conditions should reflect the local structure of the attractor, yet this structure is determined by the long-time behavior for the system. In this way, chaos links the longest time scales of the system to the shortest time scales of interest.

The attractor of the Lorenz 1963 model with typical parameter values is shown in **Figure 1**. It is believed that there are parameter values for which the Lorenz 1963 model has chaotic dynamics on a strange attractor, but, as noted above, such properties are difficult to prove even in this fairly simple system of equations.

In meteorology, the doubling time, τ_2, provides a more traditional measure of predictability than the

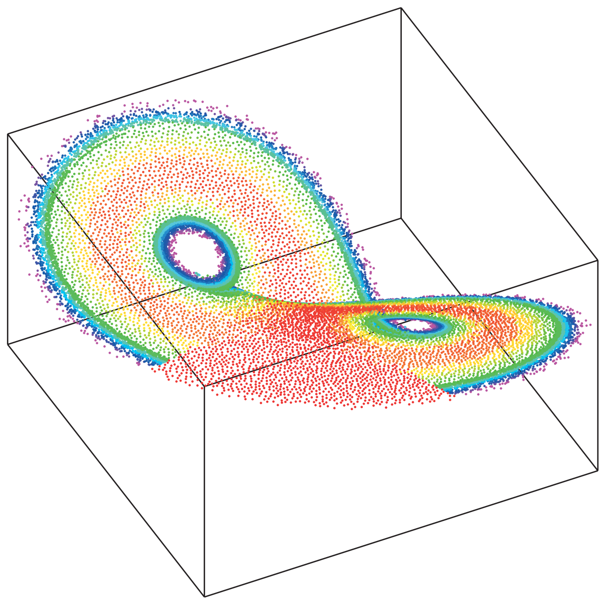

Figure 1 The distribution of uncertainty doubling times on the Lorenz attractor. Points colored red double in less than one Lorenz second. Points colored red have a $\tau_2 < 1$ Lorenz time step, orange points $\tau_2 < 2$, and so on through yellow, light green, dark green, blue, and purple. The mauve points on the inner and outer edges of the attractor for which $\tau_2 < 7$. The density of points with $\tau_2 < 5$ has been reduced for clarity. The visible line in the foreground which separates red points from each of the other colors shows the location of points which double just as they enter the region in which all uncertainties decrease, referred to in the text. (Adapted with permission from Figure 1 of Smith LA (1994) Local optimal prediction. *Philosophical Transactions of the Royal Society of London Series A* **348**: 371–381.)

Lyapunov exponents. It is also a more relevant measure, although if the growth is not uniform then the time required for an initial uncertainty at **x** to increase by a factor of four will not be twice the doubling time at **x** (or, more generally, $\tau_{q^2} \neq 2\tau_q$). In practice, one just has to look at the statistics. An average time is a more relevant measure of predictability than an average rate. Computing an average rate requires stating the duration over which to average *a priori*, while the relevant time scale is itself the quantity of interest. In fact, one can generate a family of chaotic systems each of which has a leading Lyapunov exponent greater than one, yet containing members with an average τ_2 as large as desired! Indeed, estimating a time-like statistic with the inverse of an average rate is a dubious endeavor. To convince yourself of this, consider estimating the average value of a by one over the average of $1/a$, when a is uniformly distributed between zero and one.

In **Figure 1**, points on the Lorenz attractor are colored by the doubling time of an infinitesimal uncertainty aligned initially in the local orientation corresponding to the leading Lyapunov exponent. The coloration is neither uniform nor random. Note the line separating the red points on one side from the band of each color on the other. The origin of this demarcation will be explained in the next section. Red points have a τ_2 of less than one second, for orange points it is less than two, and so on through yellow, light green, dark green, blue, and purple. The mauve points on the inner and outer edges of the attractor for which $\tau_2 > 7$. This is a clear illustration that predictability will vary with initial condition in an organized way! Which, in turn, suggests that predictability will vary in a predictable way: quantifying this in practice is the goal of ensemble forecasting. Yet even within the PMS, one is interested in finite initial uncertainties and forecasts over a finite duration. The accuracy of such forecasts need not reflect the Lyapunov exponents of the system in any way. Thus chaos *per se* places few restrictions on predictability.

True Limits of Predictability

So what are the limits to predictability of a chaotic system? The answer depends on the use to which the forecast is to be put.

Linear prediction theory aims to identify the optimal least-squares predictor: the model which, on average, yields a BFG future state with the smallest (squared) prediction error. This is a coherent approach to Gaussian uncertainties evolved under linear models, but not when applied to nonlinear systems with uncertain initial conditions.

If the initial condition is uncertain, then this uncertainty will evolve nonlinearly. It can be proven that given a series of uncertain observations of a chaotic system, there will always be uncertainty in the current state. This is the case even if a perfect model is in hand and the observations extend into the infinite past. Even then, there will be a set of indistinguishable states, each consistent with the series of observations and with the long-time dynamics of the system. The ideal forecast is then an ensemble forecast, where the members of the ensemble are drawn from the set of indistinguishable states, and each member weighted with its likelihood given the available observations. In the limit of infinitely large ensembles, this forecast can accurately quantify the relative probability of different events and the decay of predictability, correctly reflect the variations in each from day to day.

In practice, ensemble forecasting is a Monte Carlo approach to estimating the probability density function (PDF) of future model states given uncertain initial conditions. An ensemble forecast for the Lorenz 1963 system is shown in **Figure 2**. The vertical axis is time, the horizontal axis is the variable x from the Lorenz system, and each line at constant time represents the probability density function of x at that time.

At $t = 0$ the system is near $x \approx 0$ and the initial ensemble consists of 512 initial states, each of which is both indistinguishable from the true state given the observations and also consistent with the long-term dynamics of the system (that is, 'on the attractor'). This constitutes a perfect ensemble. While only the value of the x component is shown, each member of the ensemble is a complete state of the system, and corresponding figures could be drawn for y and z.

Initially the distribution spreads out as might be expected, while the average value of x increases. At $t \approx 0.4$, however, the volume of the convex hull of the ensemble shrinks, showing a true 'return of skill' as the ensemble enters a region where all uncertainties decrease! Here a BFG forecast at $t \approx 0.4$ is expected to be more accurate than the corresponding forecast at $t \approx 0.2$. This is the origin of the discontinuity in doubling times noted above in **Figure 1**: red points to one side of the line double just before entering the region, while points in the rainbow bands just across the line enter the shrinking region before they double, and must wait a finite time to be advected out of the shrinking region before they might double. Hence the discontinuity.

Returning to **Figure 2**, notice that as distribution returns near $x = 0$ at $t \sim 0.7$, a small fraction of the ensemble members switch to the wing of the attractor with negative values of x, while the majority make another circuit with $x > 0$. Owing to the symmetry of the attractor, there is a somewhat artificial return of

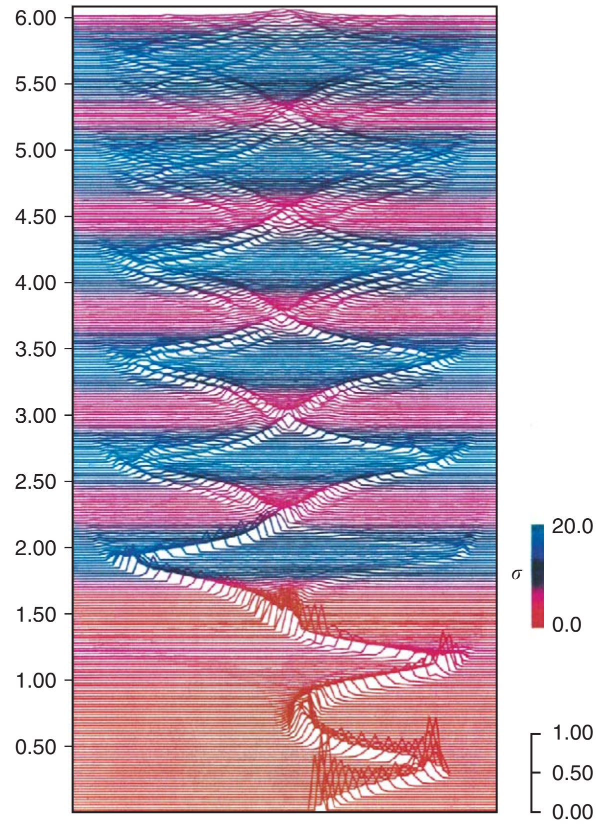

Figure 2 Evolution of a perfect ensemble under a perfect model, showing the probability distribution of one component of the model state, the variable x, in the Lorenz system. Time is denoted along the ordinate, while the abscissa is centered about zero. Each line reflects a particular time: the height reflects the probability density at that value of x, while the color of each line reflects the standard deviation of the ensemble at that time. (Adapted with permission from Figure 23 of Smith LA (1997) The maintenance of uncertainty. *Proceedings of the International School of Physics "Enrico Fermi"*, course CXXXIII, pp. 177–246, Società Italiana di Fisica, Bologna, Italy.)

skill at $t \sim 1.5$. After this, however, the ensemble members divide more evenly between the two wings of the attractor, and the distribution turns blue. The color here reflects the standard deviation of the forecast ensemble.

At this point, the standard deviation of the forecast is greater than that of a set of points taken at random from the attractor (that is, the *climatology*). Many classical measures of predictability would not see the significant information content that the forecast obviously continues to possess; good skill scores should reflect the information content of the ensemble. Even more worrying, the 'optimal' least-squares forecast at $t = 2.7$ would be near $x = 0$ where there is zero probability of observing the system. One important point illustrated here is that knowing the mean value exactly is often of much less value than knowing the likely distribution of values even approximately. A second point is that tuning nonlinear models with the aim of a 'better' average least-squares error will make the models worse, as it systematically forces model parameters away from more realistic, but heavily penalized, behavior. Such models are expected to be underactive rather than realistic.

The information in the initial ensemble will slowly diffuse away, and whether or not the information in the forecast at any given time is useful depends on the aims of the user. Eventually, any finite ensemble will itself become indistinguishable from a random draw from the climatology. At this point the forecast is useless, but this time is unrelated to the Lyapunov exponent, or the doubling time, or any other measure of infinitesimal uncertainties.

Accountable Ensemble Forecasts

Corresponding to each probability forecast there is only a single verification; thus no single forecast can be evaluated. Rather, the quality of a (long) series of probability forecasts must be considered. And inasmuch as nonlinearity will mix aspects of data assimilation (*see* **Weather Prediction:** Data Assimilation), ensemble formation and model inadequacy, the ensemble prediction system (EPS) can be evaluated only as a whole. This should not come as a surprise, since in a nonlinear system one expects to lose the benefits of linear superposition.

While the absolute accuracy of the EPS will vary with the level of initial uncertainty, ensemble forecasts under a perfect model using perfect initial ensembles are 'accountable': the uncertainty in any forecast variable computed from this ensemble will reflect the true value with an accuracy limited only by the finite number of members in the ensemble. Karl Popper introduced the notion of accountability for BFG forecasts in order to illustrate that a good model should indicate how accurately the initial condition must be measured in order to guarantee the accuracy of a forecast at any fixed lead time. The notion is easily extended to ensemble forecasts, in that an accountable ensemble forecast system should indicate how large an initial ensemble should be in order to reflect events accurately with a given level of probability.

Of course, the detailed shape of each forecast distribution will differ from day to day, **Figure 2** shows the probability distribution for a particular initial condition and set of observations. Yet if the EPS is accountable, then as the number of members in the ensemble is increased, the probability forecast will grow more accurate in a predictable way. For example, every time the ensemble size is doubled, the frequency

with which any particular variable in the verification will fall outside the ensemble will be cut in half. [In fact, the ensemble size must be increased from N members to $(2N + 1)$ members, since the probability that the next random draw falls outside the current range is $2/(N + 1)$.]

In practice, ensembles are not drawn from a set of indistinguishable states; there are a number of competing methods now used operationally, and other methods are soon to join them. Current formation schemes include sampling directions of forecast errors of the recent past, or the directions of fastest growth in the near future. Neither approach attempts to sample the initial uncertainty accurately, and thus accurate probability forecasts could not be expected from the raw forecasts, even were the models to be perfect.

Operational ensembles typically consist of between 10 and 100 members, evolved over a duration of two weeks, although seasonal ensembles are a current topic of research. Recalling that operational model-state spaces typically have ten million dimensions gives an indication of just how difficult sampling the initial uncertainty may prove to be. Despite the technical difficulties, the value of operational ensembles is reflected in **Figure 3**, a 42-hour forecast for 26 December 1999. The three panels in the top left corner show the low-resolution (control) and high-resolution (T319) BFG forecasts, and the more colourful analysis (1 0 5) which serves as the verification. The color in the analysis reflects the intense winter storm that swept across Europe. The other 50 panels in the figure each show a member of the ensemble forecast at $t = +42$ hours. This collection of 'postage stamp' maps is analogous to a single PDF at constant time in **Figure 2**. Note that about 20% of the ensemble members contain storms, and that even though there is no known way to extract an accountable probability estimate from this operational forecast, there is significantly more information than is provided by the control forecasts. In its present state, this information is already of significant societal and economic value.

Physical Systems and Mathematical Models

Arguably that, no physical system is ever isolated, and perturbations from outside the system imply that no dynamical system can be perfectly modeled as deterministic. What then does one mean by saying that a physical system is chaotic? Lorenz (1993) suggests a physical system should be called chaotic if its behavior would be chaotic were it to be isolated. This, of course, assumes there is perfect mathematical description of the hypothetical isolated system, but it is similar to the manner in which other mathematical terms are interpreted in physics; for example, the definition of periodicity in a physical system. Periodicity is a useful concept in physics, although arguably no physical system is truly periodic. Similarly, chaos may be a useful concept within physics, even if no physical system is truly deterministic.

One property that distinguishes periodic and chaotic systems is that periodic systems eventually return to exactly the same state $\mathbf{x}$ in state space. While this never happens in chaotic systems, near returns do occur for all points on the attractor; the longer is the duration of the observations, the closer are the nearest returns. Such systems, like the Lorenz 1963 model, are said to be 'recurrent'. What does it mean to say a physical system is recurrent.

At this point one has to leave the perfect-model scenario behind. Observations of a physical system are at best uncertain measurements of variables in the system's state space (if such a thing exists); in order to use them in the model the observations must be cast into a model-state space. Mathematically, a data assimilation scheme (*see* **Weather Prediction:** Data Assimilation) is simply a projection operator which accomplishes this task. Whatever the projection operator may be, the fact that forecasts are made in the model-state space holds deep consequences for attempts to make accountable probability forecasts. Estimates of predictability reflect the limitations of our models, while the underlying physical system is not so constrained.

Once some method of data assimilation is adopted so that the observations of the system can be projected into the model-state space, one can ask if a physical system is likely to be recurrent within a particular model-state space. Will two similar states be observed during the likely duration of the observations? Over the lifetime of the system? Often the answer is yes.

Many physical systems are also recurrent within the model-state space over the time of a typical experiment. Near recurrence in the model-state space opens up many modeling possibilities, the simplest being to use (local) linear regression (*see* **Data Analysis:** Time Series Analysis). It also introduces the possibility that we can learn from past mistakes, improving the model by identifying state-dependent systematic errors. Of course, doing so may increase the dimension of the model-state space to the extent that given the available observations it is no longer recurrent!

Given that the estimated recurrence time of the Earth's atmosphere is longer than the lifetime of the Solar System (longer, in fact, than the expected lifetime of the Universe!), this remedy is not available to meteorologists modeling the global circulation.

Figure 3 Ensemble forecast for the French storm of 1999. Each 'postage stamp' is a weather map of southern England and France. The three panels in the top left corner show two best-guess forecasts made at different model resolutions and the analysis, which indicates that the verification was rather different from either of these forecasts. Each of the 50 members of the ensemble at the time of the verification is also shown. (Reproduced with permission of the European Centre for Medium-Range Weather Forecasts.)

Of course, one may be able to exploit recurrence in building parameterizations (*see* **Boundary Layers:** Modeling and Parameterization) or in applications on smaller scales over shorter forecast horizons. But chaos is defined for the system (or model) as a whole. There is a clear and useful intuition of what is meant by the concept of 'approximately periodic', which is lacking for the phrase 'approximately chaotic'. Further, it is not obvious how to interpret physical systems as chaotic, if they are not expected to exist over the time scales required for chaos to manifest itself in their mathematical analogs.

Loss of Predictability: Model Inadequacy and Shadowing

In addition to uncertainty in the initial condition and uncertainty in parameter values, meteorologists must contend with 'model inadequacy': there are aspects of the real physical system that our model is simply unable to mimic. When no model in the available class of models is structurally adequate to duplicate the observed phenomena, it is unclear what is meant by the 'correct' initial condition or the 'true' parameter values. While the Bayesian agenda provides a principled scheme for handling uncertainty in initial condition and parameter value, no systematic approach is available for handling model inadequacy. Progress here requires having a good idea.

Recall that in the discussion of 'uncertainty' in the initial condition above, it was assumed that in addition to a best-guess initial condition, $\hat{\mathbf{x}}$, there was also a true initial condition, $\tilde{\mathbf{x}}$. The error in the initial condition was defined as the difference between these two points. When the model is imperfect, there is no 'true' initial condition (even if the model variables have the same names as the system variables) and the very concept of 'uncertainty in the initial condition' has to be reconsidered.

As an example, note that since the current resolution of NWP models is at best tens of kilometers, many different states of the atmosphere (with different futures) will be mapped onto the same state of the model. This is but one example of the projection effects noted above: the model initial conditions are, at best, projections of the true system state into the model-state space. The model cannot, then, be expected to reproduce the evolution of every atmospheric state, simply because there are more atmospheric states than model states! Of course, the model may have trajectories which shadow the observed atmospheric states, remaining indistinguishable from the trajectory of the atmosphere given the observational uncertainty. How might meteorologists distinguish forecast failures due to limitations in the ensemble formation scheme from those due to model inadequacy? One approach is to look for 'shadowing trajectories' within the historical observations.

Given an imperfect model, there may or may not be a model trajectory that stays close to the series of observed states, no matter which data assimilation method is used to translate the observations into model states. In this context, 'close' must be interpreted in relation to the uncertainty in the observations. A model trajectory that remains near a set of target states is said to 'shadow' the target states. Each analysis will have an associated shadowing time, just as it has an associated value of τ_2. The distribution of shadowing times reflects the relevance of model inadequacy.

If shadowing trajectories exist, then initial condition(s) which shadow may be cast in the role of 'truth' (that is, the role of $\tilde{\mathbf{x}}$) when computing uncertainty in the initial condition, at least for forecasts that are short relative to the duration over which the model can shadow. This suggests that our very definition of 'observational noise' will itself depend on the quality of the model in hand. Indeed, many data assimilation schemes are based on the assumption that long shadowing trajectories exist almost everywhere in state space.

If no shadowing trajectory exists on the time scale of interest, then the model mixes 'uncertainty in the initial condition' and 'model inadequacy' to the extent that the former cannot be unambiguously defined. On these time scales, all model trajectories differ significantly from the observations: the set of indistinguishable states is empty, and there is no optimal method of ensemble formation. Indeed, outside PMS the issue of model improvement is linked to that of forecast usage; there need be no unique best way forward. Nevertheless, current ensemble forecasts are of great value in identifying when the forecasts are sensitive to uncertainties in the initial condition, since any single BFG forecast can be identified, at the time it is issued, as unlikely to be an accurate anticipation of reality. Hence they can be expected to provide useful identification of when the forecast will be unreliable; empirical studies suggest they are also useful in identifying forecasts which are likely to have high skill. In addition, when two members of the same ensemble lead to radically different forecasts in the medium range, determining what distinguishes them at short lead times can suggest valuable observations for improving the forecast.

In addition to ensembles over initial conditions, research is underway aimed at determining how to better include stochastic effects into nonlinear models. Such stochastic effects are commonly referred to as 'dynamical noise' to distinguish them from observational noise; the latter alters the observations

but not the trajectory. A major difficulty here is formulating a relevant state-dependent dynamical noise, as the traditional approaches tend to spread the forecast into unphysical directions. Using multiple models provides one approach, stochastic parameterizations is another.

Current research is also exploring the use of ensembles over distinct models, or even ensembles over trajectories each of which uses a variety of distinct models. Ideally, these models should be independent, so that they share as few common inadequacies as possible. Methods for allocating resources among models, and for the evaluation of the distributions so obtained as forecasts, provide yet other interesting areas of current research. The traditional goal of identifying the 'optimal' least-squares predictor need no longer be a desirable end for any real forecast user. Modern forecast users, in particular industrial users, are quite capable of exploiting probability forecasts.

Since the introduction of the electronic computer, indeed since L. F. Richardson's computations early in the last century, weather prediction has been at the forefront of research into the predictability of nonlinear dynamical systems. One safe forecast is that it will remain there for the foreseeable future.

See also

Boundary Layers: Modeling and Parameterization. **Data Analysis:** Time Series Analysis. **Weather Prediction:** Data Assimilation; Ensemble Prediction.

Further Reading

Lorenz (1993) provides a general introduction to both the history and physics of chaos, while Smith (1998) is an overview of the implications chaos holds for philosophy as well as a general introduction. Introductions to both operational and theoretical ensemble forecasts can be found in Palmer (2000) and Smith *et al*. (1999), respectively. A general discussion of the role of model inadequacy in predictablity from philosophical, physical, and Bayesian perspectives can be found in Cartwright (1983), Smith (2001), and Kennedy and O'Hagan (2001), respectively and reference thereof.

Cartwright N (1983) *How the Laws of Physics Lie*. Oxford: Oxford University Press.

Kennedy M and O'Hagan A (2001) Bayesian calibration of computer models. *Journal of the Royal Statistical Society Series B* 63: 425–464.

Lorenz E (1993) *The Essence of Chaos*. London: University College London Press.

Palmer TN (2000) Predicting uncertainty in forecasts of weather and climate. *Reports on Progress in Physics* 63: 71–116.

Smith LA (2001) Disentangling uncertainty and error: on the predictability of nonlinear systems. In: Mees AI (ed.) *Nonlinear Dynamics and Statistics*, pp. 31–64. Boston, MA: Birkhauser.

Smith LA, Ziehmann C and Fraedrich K (1999) Uncertainty dynamics and predictability in chaotic systems. *Quarterly Journal of the Royal Meteorological Society* 125: 2855–2886.

Smith P (1998) *Explaining Chaos*. Cambridge: Cambridge University Press.