Atmospheric Dynamics
Lecture 11: Predictability

1 CONCEPTUAL AND PRACTICAL BACKGROUND

Of all the practical systems studied from the viewpoint of prediction possibly the most studied and most mature is that of the Earth's atmosphere particularly in the mid-latitudes. Practical (and useful) weather predictions have been possible for around fifty years now mainly because of the existence of powerful computers. One of the first to use computers to predict the weather was the mathematician John von Neumann at IAS in Princeton. The accuracy of weather predictions has tended to increase noticeably in the last few decades as a result partly of the availability of increasingly more powerful systems[1] consistent with Moore's Law. In addition to the computer aspect, the MIT meteorologist Edward Lorenz was the first to clearly recognize that atmospheric prediction was very sensitive to initial conditions. This awareness was one of the factors that drove the extensive study of chaotic dynamical systems during the 1960s and 1970s. Our primary focus in this lecture will be on this particular dynamical system.

Predictions made using a dynamical model typically suffer from two deficiencies: Firstly the model used may have certain inadequacies as a representation of reality and secondly initial conditions for a prediction may not be known exactly. Such problems are known as *model errors* and *initial condition errors* respectively.

Progress can be made in improving predictions either by improving physical depictions within models or by improving the observational network and thereby reducing errors in the initial conditions. There is considerable evidence for the atmosphere however that no practical observational network will ever eliminate significant prediction errors due to initial condition errors. Even if one was able to define such conditions to the round-off accuracy of the computing device deployed, at some practical prediction time even these minute errors would grow sufficiently large as to overwhelm the forecast made. Systems exhibiting such a property are popularly known as chaotic. The mid-latitude atmosphere is generally believed to be such a dynamical system but since it is of very high dimensionality it is better considered turbulent rather than chaotic.

In general model errors are almost by definition quite hard to study in a systematic fashion since they are caused by quite diverse factors which are not very well understood. In fact if they *were* better understood they would be removed by improving the dynamical model using this knowledge. Thus the issue of model error tends primarily to be an engineering rather than a theoretical study. We therefore focus our attention here on initial condition errors

[1] The other main factor has been the advent of accurate satellite data measuring in detail the state of the atmosphere.

but it should be borne in mind that model errors can often cause substantial problems for forecasters. It is interesting that the general public often confuses the two issues and attributes errors due to initial condition uncertainty to the inadequacy of meteorologists in reducing model errors. Of course meteorologists are not always averse to using the reverse strategy in response.

Inevitable deficiencies in observing networks imply uncertainty about the initial conditions used in predictions which can therefore be considered to be random variables. The problem of defining the associated probability distribution for these random variables is undertaken by the discipline of data assimilation. In general this problem is very non-trivial since in order to initialize a large dynamical model one requires an effective interpolation of the observed data onto the numerical grid of the prediction system. This is not always straightforward since the models are not perfect representations of reality (model error) and additionally the process of interpolation can introduce spurious dynamical effects. Often the model itself is used as the interpolator in order to avoid this however again the issue of model error intrudes. This study lies beyond the scope of this lecture and we shall suppose that such a definition is possible (albeit difficult!) and focus our attention instead on the temporal evolution of relevant random variables. We shall refer to this study as statistical predictability.

One of the great difficulties involved in studying this problem in realistic systems is the dimensionality of the relevant dynamical system. Typically a global operational numerical weather prediction model has around 10^7 or more prognostic variables primarily due to the need for reasonable horizontal resolution. In order to make the study of statistical predictability within such a system tractable, one typically focuses on relevant reduced state spaces since it is commonly assumed that a much smaller number of "modes" are responsible for much of the overall growth in uncertainty. This assumption is based on considerable practical experience but is supposition only and considerable work needs to be done in confirming it.

A common strategy used in dealing with such large systems is the Monte Carlo technique of ensemble prediction. Thus the initial conditions are sampled in some way and the (deterministic) sample members produced are traced forward in time using the relevant dynamical model. The cloud of predictions that result is then referred to as an ensemble prediction. Obviously in such a program the limiting practical factors are the time integration of the model; the length of prediction needed and the size of the ensemble. Thus, for example, a ten day forecast with 100 ensemble members is going to require around three years of model integration which is can be a considerable computational burden for modern weather prediction models.

Nevertheless the large and often explosive growth of uncertainty in weather prediction means that this issue needs to be addressed in the practical context. A common technique used is to identify a set of modes of "maximal instability" and then use these to construct a prediction ensemble. Two particular modal calculations are popular:

1. Singular vectors. Here the model is linearized about the initial condition

means and modes calculated that show the greatest growth with respect to a norm of interest (often energy). This technique will identify maximally growing modes providing that linearity is a reasonable assumption i.e. the perturbations do not grow "too big" in some sense. Such an assumption is questionable for medium range forecasts (i.e. 5-10 days) but reasonable for short range predictions (1-2 days). More detail can be found in [3]. This method is used at the main European weather prediction agency (ECMWF in Reading UK).

2. Bred vectors. Errors from a previous prediction are added to the initial condition means and then are repeatedly (i.e. iteratively) integrated forward over a short time interval (typically of order a day) using the dynamical model until modes of "maximal" growth are identified. Again it is somewhat questionable whether modes of maximal growth for medium range forecasts will be identified since the iteration time period used is usually considerably shorter than a medium range prediction. The precise methodology can be found in [4]. This method is used at the main US weather prediction agency (NCEP in Washington DC).

We briefly discuss the concept of singular vectors since they are also often used to theoretically analyze systems which are weakly non-linear and they can also be extended to study the stochastic forcing of such systems.

1.1 Singular Vectors

Suppose we have a non-linear dynamical system governed by the equation

$$\boldsymbol{u}_t = \boldsymbol{C}(\boldsymbol{u}) \tag{1}$$

If one is interested in studying the growth of small perturbations which might in the present context be thought of as representing small initial condition uncertainties then it makes sense to linearize this system about the (mean) state of the system applicable at the initial condition time. The linearized equations can then be written as

$$\boldsymbol{v}_t = \boldsymbol{A}\boldsymbol{v}$$

where $\boldsymbol{A}$ is the matrix which results from the linearization of equation (1) and $\boldsymbol{v}$ is the vector of small perturbations. Consider now the evolution of such a linear system forward in time. We may write

$$\boldsymbol{v}(T) = \exp\left(t\boldsymbol{A}\right)\boldsymbol{v}(0) \equiv \boldsymbol{B}(T,0)\boldsymbol{v}(0)$$

where $\boldsymbol{B}(T,0)$ is referred to as the propagator (of the state vector) from time zero to time T. Now suppose that we have a particular positive definite quantity which is defined on this system and is of practical relevance in evaluating uncertainty growth. Further suppose that such a "metric" can be formulated as a norm from an inner product defined on the vectors of our system. A common choice here is the energy of the system but there are many other possible choices

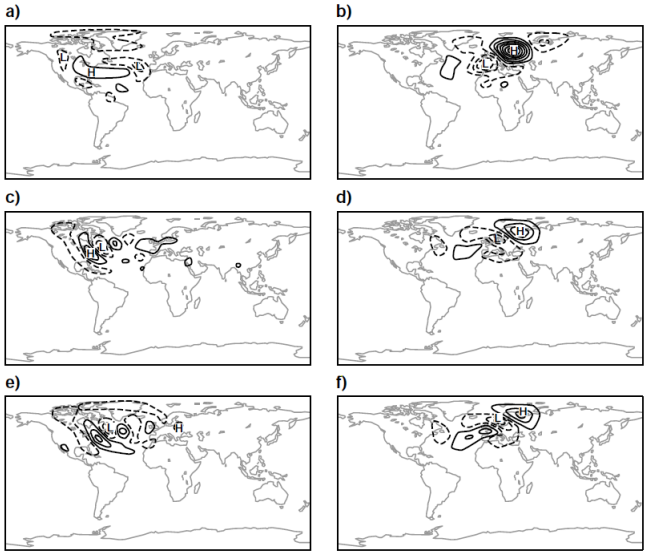

Figure 5. Streamfunction of the dominant atmospheric singular vector calculated using a primitive equation numerical weather prediction model for a three-day trajectory portion made from initial conditions of 9 January 1993 at: (*a*), (*d*), 200 hPa; (*b*), (*e*), 700 hPa; (*c*), (*f*), 850 hPa. The quantities in (*a*)–(*c*) are at initial time, in (*d*)–(*f*) at final time. The contour interval at optimization time is 20 times larger than at initial time. The figure is taken from Buizza and Palmer (1995).

Figure 1: Singular vectors from the ECMWF weather forecast model.

which depend on the subjective requirements of those evaluating uncertainty. The magnification of this uncertainty measure for the system between time zero and time T is given by

$$\lambda = \frac{(\boldsymbol{v}(T), \boldsymbol{v}(T))}{(\boldsymbol{v}(0), \boldsymbol{v}(0))} = \frac{(\boldsymbol{B}(T,0)\boldsymbol{v}(0), \boldsymbol{B}(T,0)\boldsymbol{v}(0))}{(\boldsymbol{v}(0), \boldsymbol{v}(0))} = \frac{(v(0), \boldsymbol{B}^*(T,0)\boldsymbol{B}(T,0)\boldsymbol{v}(0))}{(\boldsymbol{v}(0), \boldsymbol{v}(0))}$$

where the star operation indicates the adjoint with respect to the particular inner product. The eigenvectors of $\boldsymbol{B}^*(T,0)\boldsymbol{B}(T,0)$ are called the forward singular vectors of the propagator matrix $\boldsymbol{B}(T,0)$.

Since $\boldsymbol{B}^*(T,0)\boldsymbol{B}(T,0)$ is a non-negative and Hermitean matrix all its eigenvectors (the singular vectors) are orthogonal and the corresponding eigenvalues η_i are non-negative and called the singular values. This implies that an arbitrary perturbation of the initial conditions may be decomposed uniquely into a linear combination of singular vectors and each vector contributes its singular value multiplied by the projection coefficient to the total value of λ. If the spectrum of the singular values is reasonably peaked (as it usually tends to be) then restricting perturbations to the singular vectors with largest singular values will give the maximally linearly unstable part of state space.

Of course these arguments only work when the linearization is valid which

tends to be on order of perhaps 2 to 3 days at maximum for a normal weather forecast. An example of the dominant singular vector from a standard weather prediction model (see [3]) can be seen above.

2 THEORETICAL APPROACHES TO PREDICTABILITY

The current practical approach to statistical predictability is limited by the linearization assumption as well as the truncation of state space to a rather small subspace (typically of dimension 20 at most). We consider now more fundamental approaches which are not limited with respect at least to the linearization. Conceptually it is clear that predictability should be viewed as the time evolution of probability distributions for the random variables which serve to define the dynamical system of interest.

There are two probability distributions of importance to predictability. The first is the *prediction* distribution which is the probability distribution for the random variable which we wish to predict at a particular time. In general, as we have seen above, we assume that at the initial time that this is specified due to the nature of the observation network. A dynamical model of some kind is used then to evolve this specified distribution forward in time. For very large times we shall assume that the prediction distribution converges asymptotically to the second important kind of distribution namely the *equilibrium* distribution which we further assume is unique[2]. If the dynamical system is ergodic (a common hypothesis for realistic systems) then this latter distribution will also be the *historic* or *climatological* distribution. Note that the equilibrium distribution may not necessarily be time invariant however we shall assume that the dynamical system under consideration is subject only to external *periodic* forcing and so the equilibrium distribution is also periodic. The earth system closely approximates such a system with the dominant external forcings being the annual and diurnal cycles caused by the Earth's rotation about the sun and it's axis.

In general if one knows nothing concerning the initial conditions of a dynamical system then the best assumption concerning relevant random variables is that they have the equilibrium or historical distribution. Statistical prediction is then the process of using initial condition information to modify this distribution. In a perfect prediction one would modify it to be a delta function about a particular value.

This process is analogous to so-called Bayesian paradigms for learning: Suppose that before a learning experience occurs, our best estimate about a particular random variable X, based on all previous learning, is that it has a probability function of q. Following another learning experience we revise our estimate to

[2]These assumptions are based on empirical observation of the behaviour of numerical models rather than on rigorous results. The complexity of realistic dynamical systems generally forbids the latter.

p. The change in the probability function as a result of this experience is clearly a measure of the amount of learning that has occurred. The prior ("before enlightenment") in the case of prediction is reasonably identified as the equilibrium distribution while the posterior ("after enlightenment") is the prediction distribution. Clearly the concept of "distance" between probability distributions is crucial to this view of predictability. In Bayesian learning theory the functional most commonly used for this purpose is the so called relative entropy which plays a central role in the field of information theory. In the context of predictability this functional can be identified intuitively with the importance or utility of the statistical prediction process (see [1]).

The relative entropy is defined by

$$D(p||q) \equiv \int_S p(x) \log(p(x)/q(x)) dx \tag{2}$$

and satisfies three theoretically appealing properties

- It is always non-negative and only vanishes when p and q are identical on all sets of non-zero measure.
- It is invariant under all non-degenerate transformations of state space variables.
- In a closed dynamical system the relative entropy of two different realisations of the defining random state variables is non-increasing in time.

In addition to it's interpretation as the utility of the process of statistical prediction, relative entropy also measures the asymptotic equilibration process of the prediction distribution as it converges toward the equilibrium distribution. When the relative entropy falls to a value close to zero then one can conclude that initial condition data is basically useless for a prediction.

It is important to stress that the interpretation of relative entropy given above is valid under the assumption that the model is perfect i.e. not subject to model error. Naturally this assumption is approximately true only in a realistic situation. The point is that one is able to analyze utility drops due only to initial condition error growth.

3 APPLICATION TO REALISTIC MODELS

There are a number of interesting technical issues involved in applying the theoretical approach above to realistic models of the atmosphere. These include defining the relevant probability distributions; dealing with the high dimensionality of the realistic system and finally calculating the entropic functionals. It is fortunate that distributions within such systems are very often approximately Gaussian. This allows their approximate definition from ensembles using the means and covariances as well as the analytical calculation of entropic functionals from equation (2). In addition it turns out that it is possible to define

reduced state spaces of dimension 100 or so which serve to describe a very large fraction of all variability within the system. Recently the lecturer applied these ideas to a reasonable but still simplified model of the mid-latitude atmosphere (see [2]) with the typical result shown in Figure 2.

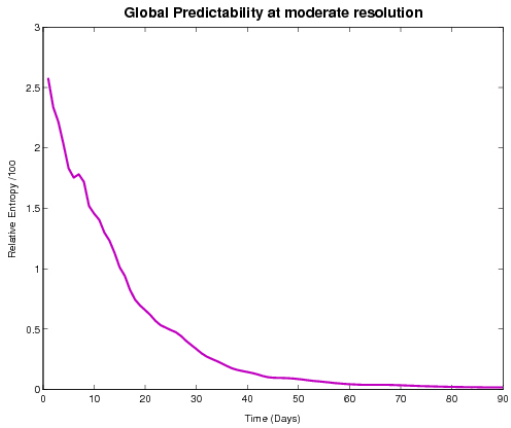

Figure 2: Global prediction distribution relaxation with time.

Notice that relaxation/equilibration is complete by about 45 days so after that time initial condition data is irrelevant to the statistical prediction.

The configuration of the model used here is for the Northern winter when the jet stream is strongest in the Northern hemisphere (see Lecture 7 Figure 2). The relaxation rate is dependent on the degree of baroclinic instability present. This is illustrated in Figure 3 which shows slower relaxation and hence greater potential predictability in more stable regions (the summer hemisphere).

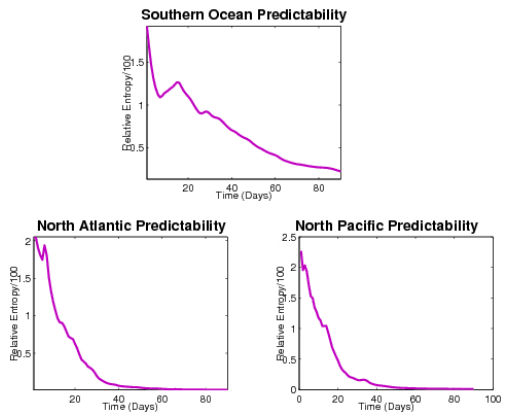

Figure 3: Prediction distribution relaxation for various regions of the globe. The results are for the Northern winter.

References

[1] R. Kleeman. Measuring dynamical prediction utility using relative entropy. *J. Atmos Sci*, 59:2057–2072, 2002.

[2] R. Kleeman. Limits, variability and general behaviour of statistical predictability of the mid-latitude atmosphere. *J. Atmos Sci*, 65:263–275, 2008.

[3] T. N. Palmer. Predicting uncertainty in forecasts of weather and climate. *Rep. Progr. Phys.*, 63:71–116, 2000.

[4] Z. Toth, Y. Zhu, and T. Marchok. The use of ensembles to identify forecasts with small and large uncertainty. *Weather and Forecasting*, 16(4):463–477, 2001.