Limits to Predictability in 2000 and 2100

Leonard A. Smith
London School of Economics
London, WC2A 2AE, UK
lenny@maths.ox.ac.uk

Abstract

Deterministic chaos is widely thought to place the ultimate limit on our ability to forecast. While chaos certainly limits our ability to predict precise outcomes in the perfect model experiments of which theorists are most fond, this paper explores the role of uncertainty in real physical systems: a simple nonlinear circuit, the weather next week, and the Earth's climate system. Here model error, not uncertain observations, may pose the more fundamental limit to prediction; this is true whether one prefers stochastic or deterministic models. A crucial difference between the circuit and the atmosphere is one of time scales: the duration over which we can observe the circuit seems long in terms of all of its natural periods, never-the-less it is error in our model(s) that prevents reliable forecasts. How should we model physical systems on time scales over which we know our models are flawed? How might we predict the weather, climate and nonlinear circuits in 2100? Will we have deployed 'improved techniques'? Or will we have altered our aims in understanding and predicting nonlinear systems?

Laws, where they do apply, hold only *ceteris paribus.*
Nancy Cartwright

1.0 Predicting Chaos

It is commonly held that if we understand a phenomena, then we should be able to predict what will happen in a given set of circumstances. Interestingly, prediction is often a priority in just those cases where we are unable to know, much less set, the circumstances; what confidence can we have in our models then? Or under new circumstances, noting that, given our limited experience, the current circumstances may not appear new.

This paper aims to discuss the roles uncertainty plays in models of reality. We will identify several distinct effects and claim that similar roles arise when modeling *any* physical system, be it as simple as an electric circuit or as complicated as the Earth's atmosphere-ocean system. The fundamental question of interest here is to ask what limits our understanding of a given system? And how should we distribute limited resources to best improve our understanding. Are our Laws consistent with our observations in the particular circumstances of interest? Or in terms of predictability: are our limits set by not knowing where to start (that is, an uncertain initial condition), problems in our Laws of physics (model error), a lack of computational power (approximation error), or do they lie elsewhere? After a brief introduction to modeling, these questions will be addressed and the ideas of tracking our uncertainty via ensemble forecasting illustrated. We speculate that even given vast increases in computational power, observational accuracy, and much longer periods of observation, both uncertainty and error will play major roles in forecasting physical systems in 2100. This suggests we alter our aims when modeling physical systems [1].

2.0 State Space Models

Faced with a real physical system, physicists tend to write down 'first principles' models by noting the relevant physical quantities that describe the state of the system, and then employing the laws of physics to quantify how a given initial state will change with time. The evolution of the entire system is then represented by a trajectory in this model-state space. The true state space, if such a thing exists, is almost certainly inaccessible since it contains, among other things, us. The fact that the model-state space and the true state space differ is a crucial point if we are interested in studying physical systems in addition to models: at best, states in our model-state space will correspond to projections from the true state space. Physical systems are often fairly well modeled in the regions of state space over which they have been observed, in part because we can keep adding variables to the model-state space until this is the case.

Extrapolating into the unknown is risky, unless we are in the rosy scenario where no new regions of state space are encountered. Just as Newton's Laws predicted Neptune, they predicted Vulcan; both of these expected planets were observed for decades, although only Neptune is seen today.

Figure 1A is a voltage trace recorded from a nonlinear circuit [2]. Applying Kirchhoff's Laws yields a model for this system: a third order ordinary differential equation which is deterministic and chaotic [3]. By chaotic, we mean that small uncertainties in the initial state of the model grow exponentially fast, on average. If our model were perfect, any uncertainty in the current state *might* quickly grow to ruin a forecast. Of course it *might* not: exponential on average doesn't imply fast everywhere. Chaos doesn't even imply that uncertainty always increases with time, only that it will grow when averaged over long enough times; often *every* initial uncertainty will shrink over finite times even in paradigm systems like the 1963 Lorenz system [4]. Determining how fast the uncertainty in each individual forecast is likely to grow (or shrink) is the aim of ensemble forecasting [5, 6].

Since we do not know the present exactly, it would be foolhardy to expect to know the the future exactly. The idea behind ensemble forecasting is to propagate our uncertainty in the initial condition forward in time, so that we have a reasonable estimate of our uncertainty in a forecast. This is illustrated in Figure 1B, where 32 different initial conditions are followed, all starting at the same time ($t = 0$). Each starting point is consistent with our observations. In this case two different models were employed with 16 different initial conditions for each model. The trajectories of one model are marked by a '+', the other by a '$\times$'. Note how they bracket reality (the solid line) fairly well until time ≈ 220 when the entire ensemble of $\times$'s make a transition to negative voltages, while the other ensemble, and reality, remain positive. We will return to the question of how to interpret multi-model ensemble forecasts below.

While Kirchhoff's Laws yield a model which is useful for mathematical study, that model does not agree particularly well with the observations. For this reason different models built from the data itself are used below[1]. For the aims of this paper, the origin of the model is irrelevant: *every* model whether derived from so-called 'first principles' or data-based will suffer from the shortcomings noted below.

The evaluation of ensemble forecasts is a bit tricky: we always forecast a distribution but verify with a single realization. This is discussed in [1] and references therein.

[1] These are radial basis function models in delay space as in [7]; more details can be found at [8] and references therein.

How can we know if we were right? In short, we cannot, but there is an approach which allows us to falsify our model. We can look at all the initial conditions consistent with our initial uncertainty: if not even *one* of them gives us a model trajectory which is consistent with the observations (a shadow) then we know model error it to blame. If our model cannot shadow the data in this way, then there is no initial condition in the model-state space which does the right thing. Over the time scales on which the model can shadow, we can evaluate the ability of the entire ensemble prediction system to produce probability forecasts by contrasting predicted probabilities with observed relative frequencies [5, 1]. Alternatively, if we observe the system long enough that it comes back near its original location in model-state space, then we can say more.

3.0 Recurrence in (Model) State Space

When contrasting the circuit with systems like the atmosphere, we have one real advantage which is likely to persist until the year 2100, and beyond. The circuit is recurrent in model-state space. The model defines a state space and our observational accuracy defines a distance in this space, effectively placing a sphere of initial conditions all consistent with our 'best guess' state at each point in time[2]. If the likely duration of observations is long enough so that a trajectory returns close to itself (to within the sphere about an observation made in the distant past), then the system can be considered recurrent. While there are other ways to define recurrence, this one will suffice here.

The recurrence time of the Earth's atmosphere has been estimated to be 10^{30} years - as this is longer than the lifetime of the planet (and just about everything else) we are unlikely to observe a near return to a previous state of the entire atmosphere. Of course, as our understanding of the physics of the circuit grows, its model state space may also grow until it too ceases to be recurrent in 2100. At present, however, we observe very near returns to previous model states. How does this help?

4.0 State Dependent, Systematic Error

A major complication of ensemble forecasting is that one estimates a distribution of future states, each of which is consistent with our current uncertainty, and then attempts to verify this forecast with the single set of observations of what really happened. Since the starting point

[2] Of course, a different radius may be defined for each different confidence level, but that need not concern us here.

is different every day, each forecast distribution is also different. We can never verify a single probability forecast, although we can test for statistical consistency over many ensemble-verification pairs [1]. But if we only look for consistency between our model and reality at many different locations in state space, how might we identify local (state dependent) model error? systematic errors which vary with location in state space? In general we cannot. In a recurrent system, things are a bit easier since near returns can be used to detect model error.

Figures 1C and 1D indicate how choosing to study a recurrent system eases the identification of state dependent model error. The dashed line connects a series of 15 step ahead forecasts, the solid line the corresponding observations. The points scattered about these two lines are *not* ensembles but the images (+) of observed near returns and the forecast ($\times$) based on that same near return. In other words, the observations up to $t = 7810$ are used to estimate the state of the system at $t = 7810$, the model is then used to make a forecast for 15 steps later (that is, $t = 7825$), the process is then repeated for $t = 7811$. And so on. The forecasts give rise to the dashed line, while the solid line shows the true images 15 steps later. Next we scan the entire data set, looking for other times at which the state of the system is near our estimated state for $t = 7810$; the 15 step ahead image of each of these neighboring states is plotted as a (+) at $t = 7825$, while the model forecast for each of these neighboring states is plotted as a ($\times$), also at $t = 7825$. Note how the spread of points changes with the initial condition: the images look very predictable at $t \approx 7810$, but the same level of initial uncertainty yields a much broader distribution near 7830. This is the hallmark of chaos. But look more closely at the zoom of Figure 1D: while the targets (+) are spread more widely indicating sensitivity to the initial condition, the forecasts ($\times$) are as well. And the two tend to agree fairly closely. Contrast this situation with that near $t = 7827$ where the forecasts are systematically too low. We can see (and correct) this local systematic error because the system returns to this region of model-state space. In short, how you go wrong depends on where you are at the moment, and if you never come back then it is hard to learn from your mistakes.

In systems in which there is no principle of superposition of solutions, assigning a single cause to a forecast error is a dubious undertaking. Expecting to build a single near-perfect model appears naive in the same way that expecting to extract a single deterministic forecast from an uncertain initial condition is naive. If the details of the model matter, then the best we can hope for is to sample the possible behaviors. In Figure 1B, the ensembles under the two different models are quite similar up until $t \approx 160$, and both reflect the general behavior of the system. By $t = 240$ the two ensembles are quite different, and in this case neither bounds the true value. By forecasting with the best erroneous models available, using ensembles over both initial conditions and models offers the chance to capture sensitivity in state space where it is important, while exploiting differences between the ensembles of each model to yield an indicator of model error. We can then hope to determine whether initial uncertainty or model error dominate a given forecast.

5.0 Operational Weather Forecasts

For many years now, operational weather centers have performed ensemble forecasts employing a (few) dozen starting points every day and following them out about 2 weeks into the future [6]. At present, it is unclear how long the operational models can shadow [9, 10], nor is the relative importance of model error and initial uncertainty well understood [1, 11]. Even in a perfect model scenario, with another century of accurate observations and the computer resources of the year 2100, there will still be ambiguity in the correct initial condition [12] and even ambiguity in the ensemble. And if the initial probability distribution is in error then the forecast probabilities will be in error even if the model is perfect.

Study of the circuit clearly shows the importance of addressing the issues of error and of uncertainty separately. Employing initial perturbations larger than the observational uncertainty in the hope of capturing effects of model error degrades forecasts where the model was accurate while regions of systematic local error are not much improved. A unified approach to using multi-model ensembles provides an promising way forward [1], especially if the development of the individual models has been as isolated as possible (so that they share as few common flaws as possible). Most other techniques, including attempts to recalibrate probability forecasts, are based on the assumption of some degree of recurrence in model-state space; it is not at all clear this can be justified for weather forecasts. The situation is even more clearcut when we consider climate.

Weather is often distinguished from climate in three ways. First by the time-scales involved (days as opposed to centuries), second in the desired detail in forecast (monthly mean values) and third by the nature of the boundary conditions. Weather is usually concerned with a ‘fixed’ system consisting of the atmosphere and ocean, while climate studies often include the response of this system to a change, like ‘suddenly’ doubling the amount of CO_2 in the atmosphere. While researchers in weather forecasting have been arguing how to best bal-

ance resources between model complexity (*e. g.* higher resolution) and quantifying uncertainty (*e. g.* larger ensembles), the climate community has striven primarily for more complicated models over a detailed analysis of the existing uncertainties in the current 'best' model.

Of course the aim of climate modeling may differ from that of weather forecasting; a good climate forecast need give only the big picture, not the details. Then again, a good ensemble weather forecast will do exactly this by giving the best picture (distribution) available while not specifying details that cannot be determined at present. Never-the-less, climate forecasts may aim at an even broader picture, attempting only to get, say, an accurate monthly average temperature. While it is not clear that is can be done in a coupled nonlinear simulation without getting the details accurately, it is clear that one can define a *credibility ratio*,

$$\tau_{\text{cred}} = \frac{\Delta t}{\tau_{\text{ave}}} \tag{1}$$

where Δt is the smallest time step in the model and τ_{ave} is the smallest duration over which a variable has to be averaged before it compares favorably with observations. Given only a single climate model run under a transient forcing scenario, there are good reasons to argue that τ_{ave} must be fairly large on statistical grounds. Most of these reasons vanish, however, if an ensemble run is made: consider 520 different climate model experiments, each starting with an initial condition corresponding to a different Monday from the 1950's. Each one run until Jan 2000. Where would the average temperature of western Canada during January 1976 fall relative to the corresponding distribution from the model runs? And if the model distribution cannot capture (bound) reality in this data, the observations used to construct the model, what faith should we have that it can do so for parameters at which the system has never been observed?

When extrapolating into the unknown, we wish both to use the most reliable model available and to have an idea of how reliable that model is for the extrapolation being attempted, say a climate forecast over the next 100 years. The argument against the ensemble experiment above is that only a model 500 times simpler (faster) than our best current model could be used. But what faith do we have in our more complicated model? Other than the fact that in 10 years time, this model will be condemned as too simple to be worthy of serious study?

6.0 Uncertainty of the Second Kind

Lorenz distinguishes predictions of the first kind where the time order of individual forecasts is important, from those of the second kind where the goal is the distribution of final time states of the system. To a large extent, the two merge together when making ensemble forecasts under a perfect model. In the year 2100, computer power should allow huge ensembles of high resolution models, but as noted above quantifying uncertainty of the first kind as an accurate probability distribution may not have been achieved. Uncertainty of the second kind considers not the uncertainty in the likely state of the system, but the uncertainty in the likely distribution of states, effectively our uncertainty in the climate of the distant future. Model error provides a major source of this uncertainty.

A 'best shot' climate model experiment must assume that the trajectory is (a) realistic, and (b) representative, and also that (c) the future is rosy. Ensemble climate experiments relax point (a) and test (b), but do not mitigate (c). All climate experiments assume that the rosy scenario holds: nothing horrible or unexpected happens. Examples where the rosy scenario fails for the circuit include a short-circuit due to an unseen 'flaw', or a theorist tripping over the equipment; in the Earth's climate system examples include a large perturbation (nuclear winter), but also any unknown flaw in the model: we assume that no important process, important interaction or important external forcing has been left out.

Given a good physical model with only small uncertainty in the parameters, and knowing the initial condition of atmosphere and ocean for every Monday in the 1950's, what is the uncertainty in the distribution of climate variables in the year 2000? What most contributes to this uncertainty? Uncertainty in the the initial conditions? in the parameters? in the boundary conditions?

Define a package as the collection of a particular model structure, a particular set of parameterizations and a fixed set of parameter values. For each package, run an ensemble of initial conditions from 1950 to 2000. Each package then has its own distribution: uncertainty of the second kind is reflected in how much these distributions differ. The two extreme options are (1) that they are each rather peaked with little overlap or (2) that they are rather similar. It seems we *must* aim for (2), our competing model structures must be so good that the details are irrelevant (within the rosy scenario). If the details matter, we are sunk.

While anyone can download the circuit data from [8], working with climate models is not feasible. Or is it? Allen [13] has proposed distributing a climate model (one package) and an initial condition to interested individuals over the world wide web, allowing a single PC to compute a single trajectory over the period 1950 to 2050. Details (and a sign-up sheet) can be found at [14]. There

are questions of experimental design still to be resolved: how should resources be distributed between sampling different initial conditions and different packages? Some trajectories are likely to become blatantly 'unphysical' before 2000, how much resource should go to quantify this probability for a given package? How might one detect a more subtle 'unphysical' trajectory in 2000-2050?

Regardless of what we learn about the future, such an experiment will teach us a great deal by contrasting the distributions of the ensemble of models with the observed climate over the past decades. It will make it possible to see which variables (and credibility ratios) can be captured in the data used to construct the models, which is surely a necessary (if not sufficient) condition for considering the same variable from a forecast.

It will undoubtedly be argued that the model(s) chosen for such an experiment are too simple, that showing a simple model fails for a particular variable does not imply a newer bigger better model will fail. This argument is as irrelevant as it is true. If model results are being used as forecasts, and not only for pure research, then they are incomplete without a reliable estimate of the forecast uncertainty of every variable discussed; and in pure research mode itself the question of predictability is much more interesting than that of any one particular forecast. Arguably, we are in need of a better baseline for which variables and time scales are reliable in *any* climate modeling scenario, establishing this for any good model strengthens our confidence in its 'newer bigger better' offspring. The Casino-21 experiment can provide confidence in more complicated models, until such time as the reliability of those models can be tested directly.

But these are short term arguments; by 2100 the outcome will be known, at least in one realization. What about climate forecasts made in 2100 for 2200? There will still be uncertainty in our modeling, but assuming that the available computer resources increase faster than the complexity of the models, quantifying the uncertainty in climate forecasts in 2100 will prove as common for state-of-the-art climate models then as it is for operational weather models today.

Models, when they do apply, will only hold in certain circumstances. Belief in extrapolation outside observed circumstances is largely a question of faith: we cannot know *a priori* if we are discovering Neptunes or Vulcans. We may, however, be able to identify shortcomings of our model even within the known circumstances and thereby increase our understanding.

7.0 Acknowledgements

I am grateful for discussions and assistance from D. Orrell, J. Hansen, P. McSharry M. Roulston and D. Broomhead, and particularly to M. Allen, T. Palmer, A. Thorpe and Z. Toth for decreasing my uncertainty regarding large models. This work was supported by ONR DRI grant N00014-99-1-0056.

References

[1] Smith, L. A. *Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems*, in *Nonlinear Dynamics and Statistics*, edited by A. I. Mees, Birkhauser, Boston, (in press).

[2] Rulkov, N., L. Tsimring, and H. Abarbanel, *Tracking unstable orbits in chaos using dissipative feedback control*, Phys. Rev. E, **50**, 1, 314-324, 1994.

[3] Abarbanel, H. *Challenges in modelling nonlinear time series*, in *Nonlinear Dynamics and Statistics*, edited by A. I. Mees, Birkhauser, Boston, (in press).

[4] Smith, L. A. , C. Ziehmann, and K. Fraedrich *Uncertainty Dynamics and Predictability in Chaotic Systems* Quart. J. Royal Meteorological Soc., **125**, 2855–2886, 1999.

[5] Smith, L. A. *The Maintenance of Uncertainty* in *Proc International School of Physics "Enrico Fermi"*, Course CXXXIII, 177–246, Società Italiana di Fisica, Bologna, Italy, 1997.

[6] Palmer, T.N., *Predicting uncertainty in forecasts of weather and climate*, Rep. Prog. Phys., **63**, 71-116, 2000.

[7] Smith, L. A. *Identification and Prediction of Low-Dimensional Dynamics*, Physica **58 D**, 50-76, 1992.

[8] http://y2k.maths.ox.ac.uk

[9] Gilmour, I. *Nonlinear model evaluation: ι-shadowing, probabilistic prediction and weather forecasting* D.Phil thesis, *University of Oxford* 1998.

[10] Gilmour, I. and L. A. Smith *Enlightenment in shadows* in *Applied nonlinear dynamics and stochastic systems near the millennium*, edited by J. B. Kadtke and A. Bulsara. AIP, New York, 1997.

[11] Orrell, D. *Forecast error growth, imperfect models, and climatology of the two-scale Lorenz system*, Transfer of status thesis. *University of Oxford*, 1999.

[12] Judd, K. and L. A. Smith *Indistinguishable States I* Physica D, (in review Dec 1999).

[13] M. Allen, *Do-it-yourself climate prediction*, Nature, **401**, 642, 1999.

[14] www.climate-dynamics.rl.ac.uk/ hansen/casino21.html