Forecasting: It’s not about statistics, it’s about dynamics

BY KEVIN JUDD AND THOMAS STEMLER

School of Mathematics and Statistics, The University of Western Australia

In 1963 the mathematician and meteorologist Edward Lorenz published a paper (Lorenz 1963) that changed the way scientists think about the prediction of geophysical systems, by introducing the ideas of chaos, attractors, sensitivity to initial conditions, and the limitations to forecasting nonlinear systems. Three years earlier, the mathematician and engineer Rudolf Kalman had published a paper (Kalman 1960) that changed the way engineers thought about prediction of electronic and mechanical systems. Ironically, in recent years geophysicists have become increasingly interested in Kalman filters, whereas engineers have become increasingly interested in chaos. It is argued that more often than not the tracking and forecasting of nonlinear system has more to do with the nonlinear dynamics that Lorenz considered, than it has to do with statistics that Kalman considered. A position with which both Lorenz and Kalman would appear to agree.

Keywords: forecasting, nonlinear systems, optimal filters

1. Introduction

The landmark papers of Kalman (1960) and Lorenz (1963) are not unrelated. In both cases the authors were looking for a better forecasting alternative to the Wiener filter, which operates in the frequency domain, and both authors adopted a state-based approach. Lorenz considered deterministic nonlinear systems; Kalman considered linear stochastic systems. These two approaches can be viewed as two different extreme special cases of the more general situation of nonlinear stochastic processes. In the paper of Kalman (1960), after stating the objective of obtaining a linear filter to accomplish the tasks of tracking and forecasting of a stochastic process, there is a footnote:

Of course, in general these tasks may be done better by nonlinear filters. At present, however, little or nothing is known about how to obtain (both theoretically and practically) these nonlinear filters.

Kalman clearly understood nonlinear systems posed a challenge and would require something very different to what he proposed. Lorenz, however, revealed the first hints of the difficulties of nonlinearity. The question that might now be asked is, have the goals of Kalman and Lorenz been achieved yet? What theory and practical techniques are now available for tracking and forecasting nonlinear processes? This is the subject of this paper. We only briefly overview the currently available techniques. Our main purpose, as the title and abstract suggest, is to argue that the problem is not yet solved, and that new evidence suggests that future advances lie in the direction that Lorenz initiated, and Kalman considered important.
The problem of tracking and forecasting is old. The essential elements of the theory can be found in the work of Pierre-Simon Laplace on probability and celestial mechanics, in particular, concerning the identification of the orbits of planets and comets. Laplace was first to clearly articulate the notion of state, which is central to forecasting. He conceived of a vast intellect that could measure the position and motion of every particle in the Universe, knew the mathematical laws governing the particles’ interaction, and possessed the capacity to compute them. Laplace’s Universe was deterministic, so to such an intellect all the past and future was known. On the other hand, James Clerk Maxwell, as a result of his study of gases, noted that there was a serious problem for Laplace’s daemon. If the measurement of any particle had the slightest error, then errors would compound with each collision, especially glancing collisions, so that the number and size of errors would grow exponentially, until the forecasts were useless. Meteorologists were therefore aware of this problem for the complex processes of gases and weather. Lorenz showed, using an astonishing simple nonlinear system, that this sensitivity is not a consequence of the complexity of the system, but a consequence of nonlinearity. For weather forecasting systems nonlinearity is unavoidable, but the atmosphere is exceedingly complex too. Traditionally, engineers deal with simpler systems and try to avoid nonlinearity. This is not always possible, for example, in modern power generation and distribution (Kwatny et al. 1995). Nonlinearity may even be exploited to advantage, for example, in optical and electronic devices (Slight et al. 2006).

For the purposes of the following discussions it is worth establishing some basic concepts and terminology. In abstract terms tracking and forecasting some aspect of reality requires a model, which could be deterministic, as in difference or differential equations, or stochastic, as in augmenting deterministic equations with random processes. The deterministic component of the model we will refer to as the dynamics. The model of reality provides a simulation of what could happen. Forecasting is more ambitious in that observations of reality are used to try to simulate what will actually happen. The goal in tracking is to use observations of reality to select a model state that is representative of what the system is currently doing. Observations are usually incomplete and inaccurate, so there is uncertainty in the model state and forecast. For some purposes, such as guidance and control, a best guess model state is all that is required. Sometimes the forecast from the best guess state is sufficient, in other situations the uncertainty of the model state and forecast needs to be quantified, for example, in terms of the probability of the occurrence of certain events, or as an ensemble of states or forecasts to reveal the spread of possibilities. Often the theory concerning tracking and forecasting implicitly makes the simplifying assumption that the model and system are identical, the so called perfect model scenario. In practice, the model and system are different; the imperfect model scenario. Dealing with model error is a difficult. Sometimes model error is dealt with through the introduction of stochastic elements into the model, which may be appropriate in some circumstances. Model errors, however, could be in the dynamical component of the model, in which case the errors are not random, they are state dependent and usually correlated over time and space. In the following we use the term filtering to mean obtaining a current state of the model from past observations; in contrast to signal processing filtering which also can have the sense of a non-causal filter or a smoother, which uses past and future observations.

Article submitted to Royal Society
2. Filters from stochastic models

Kalman (1960) considered the tracking and forecasting problems in the context of linear stochastic processes with Gaussian stochastic sources and obtained an optimal filter. This case is the easiest mathematically because all the uncertainties have Gaussian distributions, for which a mean and co-variance matrix is a complete description. Consequently, the filter can be expressed as convenient matrix equations that update the current maximum likelihood estimate of the state and a co-variance of its uncertainty. Kalman & Bucy (1961) introduced an extension of the linear system filter to nonlinear systems using local linearisation of the model. It is clear that Kalman did not consider this extension a complete solution for nonlinear systems.

Although the original filter and its extension are widely used, engineers well understand that if the nonlinearity is significant relative to the noise, then these filtering methods can fail. An obvious and essential problem is that for nonlinear systems the uncertainties are no longer Gaussian, even if the stochastic sources are Gaussian. To overcome this problem, instead of using a single maximum likelihood estimate of the state, one can use an ensemble of state estimates, which better reflect the uncertainty. Unscented Kalman filters (Julier & Uhlmann 1997, 2004) use a minimal ensemble, which avoids having to compute the linearisation of the model, and accounts for first-order bias due to nonlinearity. Ensemble Kalman filters (Evensen 1994, 2003) use an arbitrary large ensemble, and come in many variants. In principle, any non-Gaussian distribution of uncertainty can be represented by a sufficiently large ensemble, but all variants of ensemble Kalman filters effectively only account for second moments. That is, although an ensemble could potentially represent higher-order moments, the filter update rules only involve the co-variances. Of course, high-dimensional models, like those used in weather forecasting, would need impossibly large ensembles just to represent all the co-variances, so the restriction to second moments is not necessarily important or avoidable. Many ensemble Kalman filter variants use transforms and localisations in an attempt to capture the most significant co-variances with an ensemble that is small relative to the model dimension. On the other hand, second moments do not indicate skewness of uncertainty, which is often important, and sometimes identified in small ensembles of high-dimensional systems (Roulston & Smith 2002).

Ensemble Kalman filters are a special case of sequential Bayesian filters; the special case being they are ensemble-based second-moment filters. Assuming that the current knowledge about the system state is represented by a probability distribution, a sequential Bayesian filter provides rules to update that knowledge given a new observation and a stochastic model of the system. These methods are quite general and are implemented as particle filters using Monte Carlo simulation methods (Rubin 1988; Del Moral 1995; Arulampalam et al. 2002).

Despite the power and simplicity of sequential Bayesian filters, they are not uniquely optimal in the sense that Kalman filters are; a fact that is often overlooked, and to which we must return later. In practice, particle filters have a number of deficiencies that are well documented; see for example Arulampalam et al. (2002). Briefly stated, these filters will fail unless the ensemble is sufficiently large and the Monte Carlo simulation sufficiently good. Typically failure manifests itself by the ensemble collapsing onto a few ensemble members, which are not a good representation of the true uncertainty. The potential for this kind of failure is increased.
for small ensembles, poorly implemented simulations, high-dimensional systems, strongly nonlinear systems, and systems close to being deterministic.

There are further deficiencies of sequential filters that are fundamental to a sequential approach, which have only recently been documented (Judd & Stemler 2009). The principle idea here is that a poor state estimate in the past, due to say an unusually large observational error, could propagate to subsequent state estimates, and could even compound over several updates. With sequential filtering there is no possibility of correcting the inevitable past mistakes. In linear systems there is no value in correcting past state estimates, however, for nonlinear systems reassessing all past information can provide better filtering. Reassessment can be realised by non-sequential shadowing filters that are based on a deterministic approach to filtering.

3. Filters from deterministic models

Stochastic models are, in a sense, a worst case model that assumes there are processes of the system that are too complex to be modelled by anything better than a random process. In contrast, deterministic models would seem to be a best case model that assumes that the model is almost perfect. Lorenz appears to have been motivated to consider how good deterministic model forecasts could be. We will see that the situation is more subtle than it first appears, in that, deterministic models can have advantages for tracking and forecasting even when a stochastic model might seem to be more appropriate.

Lorenz (1963) revealed a number of important properties of nonlinear deterministic systems: chaos, attractors, and sensitivity to initial conditions. Lorenz recognised that sensitivity to initial conditions puts fundamental limits on forecasting nonlinear systems. On the other hand, attractors represent the physically realisable states of the model, that is, these are the only states one expects to find the model in; there is the potential here that knowledge of an attractor can be exploited for tracking and forecasting. Sensitivity to initial conditions implies trajectories of nonlinear deterministic models always diverge, but Anosov and Bowen (Anosov 1967; Bowen 1975; Katok & Hasselblatt 1995) showed that models may also have the complementary property of shadowing, which, for the present purposes, we will take to mean, there are trajectories of the model that remain consistent with observations of the system, even when the model is imperfect. Hammel et al. (1988) proposed shadowing as a means of non-causal filtering for signal processing.

Noting that one shadowing trajectory implies the existence of many others, Judd & Smith (2001, 2004) developed shadowing concepts for tracking and forecasting in both perfect and imperfect models. The principle tool is a shadowing filter, which provides an initial shadowing trajectory. By itself, a shadowing trajectory can provide a solution to the tracking problem. From an initial shadowing trajectory many more shadowing trajectories can be easily found, which altogether provide an ensemble of indistinguishable states. The indistinguishable states provide a complete quantification of the uncertainty for tracking and forecasting. On present evidence it appears that imperfect-model shadowing and indistinguishable states come closest to achieving the nonlinear filtering that Kalman and Lorenz envisaged (Judd & Smith 2004; Judd 2008; Judd & Stemler 2009).
4. Shadowing filters

Shadowing filters attempt to find a trajectory of a deterministic model that is consistent with a sequence of observations, that is, the observations could have arisen from observations of the model trajectory under the assumed observational noise. Consequently, shadowing filters are not sequential in the sense of the sequential Bayesian filters previously discussed, because shadowing filters use a long sequence of past observations simultaneously, not just the most recent observation.

A straight-forward method of implementing a shadowing filter is gradient descent of indeterminism (Judd & Smith 2001; Stemler & Judd 2009a), which can be guaranteed to obtain a shadowing trajectory (Ridout & Judd 2002), even with limited gradient information (Judd et al. 2004b). Although shadowing trajectories can in principle be obtained by other variational methods, there are problems in doing so, as discussed later. The gradient descent of indeterminism takes any sequence of states, and attempts to adjust these states iteratively towards a trajectory by minimising the sum of squares of mismatches between each state and the forecast from its preceding state. This follows the method of Hammel et al. (1988), although the principle dates back to the work of Laplace.

The notion of shadowing derives from deterministic models, but the idea can be generalised to stochastic models. Indeed, gradient descent can be modified for this purpose (Judd 2008); the algorithms obtained being closely related to those developed for the imperfect model scenario (Judd & Smith 2004). Case studies have so far shown that the simple gradient-descent shadowing filters so obtained out-perform Kalman-Bucy filters (Judd 2003a) and particle filters (Judd & Stemler 2009). These two studies focused on the tracking problem, but evidence currently being gathered suggests that the superiority extends to ensemble forecasting (Judd & Stemler 2009; Stemler & Judd 2009b).

Shadowing filters have many advantages over sequential Bayesian filters that results from the exploitation of nonlinearity and the simultaneous use of many past observations. Shadowing filters avoid propagation of unavoidable past errors, and can obtain representations of non-Gaussian uncertainty using an ensemble of indistinguishable states. Ensembles of indistinguishable states are generated from a shadowing trajectory. The performance of the shadowing filter is not affected by the ensemble size, which is arbitrary and can be varied at any time as required. There is no possibility of ensemble collapse that plagues particle filters. Since indistinguishable states correspond to shadowing trajectories, the entire ensemble does not need to be regenerated at each forecast time, because many useful ensemble members are simply extensions of existing shadowing trajectories. Shadowing filters are valid at large and small noise levels, including the deterministic limit, without ad hoc modification that Bayesian filters require at the small noise limit (Judd 2003b).

The case study of Judd & Stemler (2009) revealed a rather surprising result: under most circumstances the shadowing filter derived from the assumption of a deterministic system, performed better than a particle filter, and a shadowing filter derived from the assumption of a stochastic system, even when the system was stochastic. The implications of this are quite profound. It implies that for tracking, the stochastic element of a model is not important, it is the nonlinear dynamics that provides the important information, that is, the stochastic element of the model only comes into play when the uncertainty is considered. A further implication is...
that when deciding on a model to employ, it is more important to provide a good
dynamical description first, and only introduce stochastic elements as a last resort
and only if demonstrated to be necessary. The study shows that the only case where
a shadowing filter derived from a deterministic model is not the best choice, is when
the stochastic elements of the system have larger variance than the observational
errors. If this situation occurs, it can be easily identified.

5. Filters from variational methods

Shadowing filters we have employed have similarities to filters derived from vari-
arional principles, and are often confused with them. Our shadowing filters are,
however, quite distinct, and avoid the deficiencies of variational methods. For more
details on the points raised below, see Judd (2008).

The method of 4D variational assimilation is in fact a method for finding shad-
owing trajectories, and hence is a type of shadowing filter. In this filter shadowing
trajectories are found using a shooting method, whereby an initial condition is
varied to obtain a trajectory close to the observations. The difficulty with this
technique is that sensitivity to initial conditions means that the time window over
which shadowing can be achieved is limited. Gradient-descent shadowing filters
have no such limitations.

Another variational method is the so called weakly constrained 4D variational
assimilation (WC4DVA), which attempts to apply variational methods to a stochas-
tic model; although the literature concerning WC4DVA is confused about the for-
mulation of the problem. Often discussion begins with a cost function, leaving the
underlying stochastic model unstated, which leads to conflicting, or inappropriate,
interpretations of what the terms in the cost function represent. At best, WC4DVA
is sometimes presented as an attempt to achieve shadowing (or words to that effect)
in the stochastic model context, but this would be incorrect. A correct stochastic
formulation of WC4DVA results in available observations having insufficient degrees
of freedom to solve the problem as stated, that is, the variational method does not
achieve a solution of the problem it is supposed to be solving (Judd 2008). An
alternative formulation, holds that what would be the innovations in a stochastic
model are model errors, or at least, attempts are made to treat them as such. Why
this should be a valid formulation of the problem is not explained, and seems to be
entirely incorrect to the authors. In any case, this prescriptive view of model error
is both unhelpful and unnecessary.

6. Model error

It is the nature of model error that its exact properties are unknown. Sometimes the
system involves complex processes for which it is reasonable to substitute a random
process in a model. For other systems, parameter values are unknown, or the exact
nature of the dynamics is unknown, that is, the deterministic mapping or vector
field, is unknown. It is common practice to assume that deficiencies of models can
be replaced by random processes. This is clearly incorrect for the latter type model
errors because these errors are non-random, state dependent, and highly correlated.
Furthermore, in §4 it was reported for tracking that the stochastic component of
the model is usually not important, because a deterministic model usually performs

Article submitted to Royal Society
as good or better. Hence, there is no need to introduce random processes into the model for tracking purposes; the majority of effort should be invested in getting the dynamics right. Random processes are, however, useful in representing uncertainty in forecasts due to model error (Judd et al. 2007).

Two significant deficiencies of sequential Bayesian filters, and weakly constrained 4D variational assimilation, are that they attempt to represent model error by random processes alone, and that the nature of model errors must be prescribed. The nature of model error is generally unknown. An advantage of shadowing filters is that model errors are discovered, not prescribed (Judd et al. 2008). Geometrical analysis of the performance of shadowing filters can be used to reveal the nature of model errors. The results of this analysis can then be used to inform better tuning of the filter and the generation of indistinguishable states to quantify uncertainty.

7. Conclusion

Sequential Bayesian filters, of which the family of Kalman filters is a special case, are highly successful filters. However, they have deficiencies both in implementation and at the fundamental level of not being able correct inevitable past errors. They also require that the properties of model error have to be prescribed. Case studies have shown that non-sequential shadowing filters usually perform better than particle filters. Shadowing filters do not have the theoretical and practical deficiencies of sequential Bayesian filters. Furthermore, because shadowing filters do not prescribe the nature of model error, they can provide insight into the nature of model error. The successes and additional benefits of shadowing filters stems from their exploitation of the dynamics, and ignoring processes that are effectively random. Although, sequential filters employ dynamics, and shadowing filters employ statistics, it is the emphasis on dynamics in the shadowing filters that gives them their advantages.

The many advantages of shadowing filters suggests that they should be useful, if not superior to sequential Bayesian filters, in situations in engineering where tracking and forecasting is required. The ease of implementation and reasonable computational requirements of gradient descent shadowing filters have enabled their implementation for a operational weather forecasting model (Judd et al. 2004a, 2008). Although significant effort has been made to implement Ensemble Kalman filters for global weather forecasting, it will always be an impossible goal, because in such high-dimensional systems it is not possible to adequately represent uncertainty with any reasonable sized ensemble. Of course, Ensemble Kalman filters have been implemented, for example, in meso-scale systems, but only with additional assumptions (Houtekamer & Mitchell 2005). As we have indicated, the key limitation of these filters is that the performance depends on having a sufficiently large ensemble. In contrast, the performance of shadowing filters is unaffected by the size of the ensemble of indistinguishable states used to quantify uncertainty; the ensemble can be as small as necessary, or as large as practical.

The authors predict that despite their advantages it will be some time before shadowing filters become widely adopted. Ironically, this is a result of dynamics; in the form of cultural inertia. There is currently a significant investment in Kalman filter variants and sequential Bayesian techniques, and belief in them that at times approaches religious fervour. Although the ensemble of Bayesian filterers is large,
the ensemble has collapsed onto perturbations of a few ideas, an example of the classic failure of applying Bayesian methods.

**Acknowledgements**

This work has had essential support from ARC grants DP0662841 and DP0984659, and from ONR/FO through NICOP award N000140510668. The authors also acknowledge the support of Leonard Smith and the London School of Economics.

**References**


*Article submitted to Royal Society*