Invited Paper

Fifty years of forecasting chaos and the shadow of imperfect models

Kevin Judd¹a)

¹ University of Western Australia, Perth, 6009, Australia
a) Kevin.Judd@uwa.edu.au

Received January 6, 2014; Published April 1, 2016

Abstract: In this contribution we review progress on the problem of forecasting chaotic dynamical systems. The nature of chaos was initially illusive, having escaped discovery by several great minds who were in a position to find it. Once the essential features of chaos were uncovered, the concept greatly influenced the thinking of physicists and forecasters, in particular, it brought into focus the limitation of forecasting nonlinear dynamical systems. This resulted in a dichotomy between prediction methods for (nearly) linear systems and nonlinear systems; the important features and distinctions of these two approaches are reviewed. An important issue that arises is the role of model error in forecasting; many methods either ignore or over-simplify the nature of model error, largely because this allows adaption of existing techniques. Finally it is argued that the Isis Programme (Imperfect-model Shadowing and Indistinguishable States) provides a theoretical and practical basis for dealing with all sources of uncertainty, including observational noise and model error.

Key Words: modelling, filtering, prediction, shadowing filter, indistinguishable states

1. Chaos discovering chaos

The year 2013, was the fiftieth anniversary of the publication of a paper, “Deterministic Nonperiodic Flows”, by Edward N. Lorenz [1]. This paper clearly illustrated concepts and properties that became known as chaos. Lorenz was trained as a mathematician, but as a result of his war service became a weather forecaster at a time when numerical computation by electronic computers was becoming a possibility. One of the important features of chaos is sensitivity to initial conditions, which has a consequence that a small error in a forecast grows more-or-less exponentially. This property of the weather had been recognised long before by Maxwell and Eddington. Indeed it is a recognised property of many complex systems; as Benjamin Franklin [2] put it:

For want of a nail the shoe was lost.
For want of a shoe the horse was lost.
For want of a horse the rider was lost.
For want of a rider the message was lost.
For want of a message the battle was lost.

This is an invited paper to the “Special Issue on Recent Progress in Nonlinear Theory and Its Applications” in NOLTA, IEICE, Vol. 5, No. 4, 2014.
Lorenz equations demonstrate movement onto an attractor, and sensitivity to initial conditions. These are two key features of chaotic dissipative systems.

For want of a battle the kingdom was lost.
And all for the want of a horseshoe nail.

This is an old idea; a 14th century proverbial tale of this type was collected by Wilhelm Grimm [3]. The great importance of Lorenz’s paper was to demonstrate that sensitivity to initial conditions did not require a complex system, it only required a nonlinear system, and simple one at that. A second important feature of chaos is the attractor. Of all the possible configurations of the system only relatively few are physically realisable. One would only ever expect to see a complex chaotic physical system in one of these attractor states. If you could somehow realise a non-attractor state, the system would quickly move onto the attractor once free to do so. Figure 1 illustrates both of these features. Over the last fifty years, the ideas clearly expounded by Lorenz has opened up completely new ways of looking at physical processes. from electronic circuits, to turbulence, to forecasting weather, and many others.

It is probably in the nature of chaos that several great minds came close to recognising the existence of chaos. The great mathematician Henri Poincaré won from King Oscar II of Sweden a prize specifically for his work on the three body problem. One of the conditions of the winning the prize was the preparation of manuscript setting out his work, but the publication was delayed because of what appeared to be a technical problem: the convergence of series that describe homoclinic orbits. This unexpected impediment is intimately associated with chaos; Poincaré clearly recognised something complex was happening, but never really unravelled the orbits. Much later von Neuman and Ulam were investigating nonlinear maps using the newly invented electronic computers; they investigated the quadratic map as a source of random numbers, but failed to recognise chaos as Lorenz did. Indeed Lorenz’s discovery involved a serendipitous accident when sensitivity to initial conditions was revealed because he recomputed a solution of his differential equations using initial values that were less than the computing machine’s internal precision; the new solution diverged from the previously computed solution.

Kalman, Lorenz and Anosov

Lorenz’s 1963 paper established many of the important concepts and properties of what later became known as chaos, but Lorenz’s primary interest was weather forecasting. Just three years prior the mathematician-turned-engineer Rudolf E. Kalman had published a paper “A New Approach to Linear Filtering and Prediction Problems” [5]. The papers of Lorenz and Kalman are not unrelated; they are two sides of a tossed coin. Both Lorenz and Kalman were interested in forecasting, or prediction,

\[\text{It is sometimes claimed that Poincaré discovered chaos, but this extrapolation by hindsight. Poincaré was a prolific writer and surely would expounded on it in more detail if he had. Neither his students, nor Birkhoff, followed up the ideas.}\]
Fig. 2. The Kalman-Bucy locally-linear filter (extended Kalman filter) compared to a truly non-linear shadowing filter applied to the chaotic Ikeda system [4]. Small black squares indicate the true state, each cross marks a state estimate from different noise realisations. Note how the shadowing filter correctly restricts state estimates to the physically realisable attractor states, and the state estimates have smaller variability about the true state; the fully non-linear shadowing filter is extracting more information from the observations.

and both were looking for a better alternative to the filter of Norbert Wiener that operates in the frequency domain. Both chose a state space approach. Kalman chose a linear model with a stochastic component to account for unknowns, whereas Lorenz chose to investigate the dynamics of deterministic nonlinear models. Kalman and Lorenz were interested in the same problem, but took two very different approaches. They needed to because the problem is very difficult and their two approaches represent two different sets of simplifying assumptions.

Kalman was clearly aware of the limitations of his approach when writing his 1960 paper. There is a footnote to the last sentence of the first column of print that is often overlooked. The sentence reads:

In all these works, the objective is to obtain the specification of the linear dynamical system (Wiener filter) which accomplishes the prediction, separation, or detection of a random signal. 4

Footnote 4 reads:

4 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.

This is an astonishingly prescient aside: Figure 2 illustrates the degree to which Kalman was right to place a caveat on his approach, although Kalman probably did not anticipate the difficulties chaos would present; Lorenz’s discoveries were not yet published. Fifty years later, we are finally able to achieve the goals of Kalman’s footnote; this paper explain how.

Kalman filters have been a powerful tool in engineering and other applications. Shortly after Kalman’s original paper, he and Bucy published [6] what is often referred to as an Extended Kalman Filter, that uses local linearisation of a nonlinear model, but this is certainly not what is envisaged in the footnote. The linear Kalman filter is provably optimal for linear systems, but the extended Kalman filter is not. The great usefulness of the Kalman filter and the extended Kalman filter is that if the non-linearity is not too great, then these filters are entirely adequate in many applications. It is only when the non-linearity is significant compared to the noise level that problems occur; uncertainties are no longer Gaussian and the filter is prone to catastrophic failure [7]. Since Kalman’s original paper there have been a plethora of filters that extend Kalman’s approach, for example, the ensemble-based
Kalman filters. There is an engineering adage: if there are many ways of doing something, then none of them is right. There is an optimal nonlinear filter, but it looks nothing like Kalman’s filters.

Kalman as a mathematician clearly understood the limitation of his approach, and recognised something very different was required. A hint at how different an approach is required does not become fully apparent until the consequences of Lorenz’s 1963 paper are understood. Chaos in nonlinear systems makes the forecasting problem a great deal more difficult; probably even more difficult than Kalman originally imagined.

Despite the great deal of work that has pursued extensions of Kalman approach, it is not in the right direction, which is not to say the latter works are not useful under certain assumptions. To fully understand the forecasting problem requires a deep understanding of the dynamics of nonlinear systems. One of the key concepts required did appear until 1967 with the publication of the mathematician Dmitri V. Anosov, “Geodesic flows on closed Riemannian manifolds with negative curvature”. Whereas the papers of Lorenz and Kalman have been visibly influential, with each having over ten thousand citations, Anosov’s paper is relatively unknown and obscure, only a few hundred citations. In fact, it is not even the main results of the paper that has been influential, it is an incidental result called the Shadowing lemma. The concept of shadowing is a powerful tool in the study of chaos, forming the basis of many important results. It is only in the last decade that the importance of shadowing has been recognised in forecasting nonlinear systems.

Reality, models and observations

Before going any further it is worth making a few definitions and distinctions. Economists, engineers, managers, scientists, and others want to have predictions of the future for a multitude of reasons ranging from targeting missiles, to planning infrastructure, to confirming abstract physical theory. What each wants is a prediction of what Reality will do. For our purposes we may assume that reality is arbitrarily complex. This does not mean that Reality is arbitrary; Reality follows rules. The problem for us is Reality is so complex that we can never construct quantitative mathematical equations that can predict all details of Reality; even in the simplest highly controlled laboratory experiment, like swinging pendulum. Our mathematical equations comprise a model of Reality, all models are imperfect.

It is important to make the distinction between the arbitrarily complex Reality, and the necessarily simplified Model. Ideally a model is either perfect, that is identical to Reality, or from a perfect model class, where there exists a perfect model obtained by specifying a few parameters. Often mathematical analysis of prediction and forecasting assume a perfect model scenario, or perhaps a perfect model class scenario with the correct parameters unknown. Such an approach is commonly employed because it has theoretical advantages; essentially reducing everything to a statistical estimation problem. In actuality we must deal with an imperfect model scenario, there is no perfect model. It is important to recognise this because theory and methodology developed in a perfect model class scenario need not apply, neither theoretically nor practically, in an imperfect model scenario. Ideally we want to develop theory that applies in a perfect model scenario, and makes sense, and is useful, in an imperfect model scenario.

If Reality is assumed to be arbitrarily complex, then we cannot know every detail of Reality, even in a classical (non-quantum) universe; there are limitations on what can be measured in practice. Nonetheless, observations are all we know about Reality. Our models are based on past observations, both to formulate their mathematical structure, and to establish values for parameters and variables when making forecasts and predictions. Furthermore, models are falsified and verified, or at least partially confirmed, by comparing predictions with future observations. Observations, however, can be incomplete, unrepresentative, or inaccurate, because of limitations of our measurement devices.

State-based models and forecasting

Perhaps the greatest contribution to science by Isaac Newton is not his laws of motion, gravity, and light, nor his contributions to the invention of calculus, it was his introduction of the notion of state. In a state-based model of Reality there are a set of variables (or a state function in the
perturbations are a sequence of dynamical noise that might appear to play the same role of the model, that is, a stochastic formulation or a deterministic formulation. Can be broadly separated into two principle approaches according to the underlying assumptions of the model, that is, a stochastic formulation or a deterministic formulation.

Taking Lorenz and Kalman as the starting point, the developing theory over the following fifty years can be broadly separated into two principle approaches according to the underlying assumptions of the model, that is, a stochastic formulation or a deterministic formulation.

Both formulations can begin with a time-series of observations $s_t \in \mathbb{R}^k$, $t \in \mathbb{Z}$. We assume observations comprise a finite collection of measurements, $s_t$ is a point in $\mathbb{R}^k$, and that these same measurements are repeated at discrete, equally separated times, which is taken to be the unit of time $t$. Both formulations can also assume a finite dimensional state $x_t \in \mathbb{R}^d$, where the observations are a function of the state, that is, $s_t = g(x_t, \eta_t)$. Here $\eta_t$ is a random variable that represents uncertainties introduced by the limitations of the measurement devices; typically $\eta_t$ is a vector of real numbers, and is referred to as observational noise. Let $s_t$ denote the time-series of observations up to time $t$, that is, the sequence of $s_t$ for $\tau \leq t$, and $x_t$ denote the trajectory of states $x_\tau$ for $\tau \leq t$.

The stochastic formulation of a model can have a form

$$x_{t+1} = f(x_t, \xi_t),$$

where $\xi_t$ is another random variable, typically a vector of real numbers, and is referred to as dynamical noise. Since observations are at discrete times this model is sufficient, although notionally it may be understood as the finite integration of a stochastic differential equation. The underlying model may even have a state space that is notionally a set of vector and scalar fields, but since in practice such partial differential equations are discretised when solved, such a model once again reduces to the above finite dimensional form.

In a perfect model scenario a deterministic model is often viewed as the limiting case of a stochastic model when the dynamical noise goes to zero amplitude, hence reducing to a form $x_{t+1} = f(x_t)$. In an imperfect model scenario, however, one must explicitly account for model error when analysing observations and making forecasts. To do this, one can adopt a model of the form

$$x_{t+1} = f(x_t, u_t),$$

where $u_t$ is typically a vector of real numbers that represent model error. Although the model error $u_t$ might appear to play the same role of dynamical noise $\xi_t$ in (1), it does not. The model error $u_t$ are a sequence of perturbations required to keep the model tracking Reality, actually, tracking the observations $s_t$ of Reality. Suitable values of $x_t$ and $u_t$ can be determined given past observations $s_t$, $t \leq 0$, but the future $u_t$, $t > 0$ are unknowns. Furthermore, there is no assumption that the $u_t$ are random variables that conform to a prescribed distribution; they are simply unknowns, which can be thought of a control variables that steer the imperfect model in the direction Reality takes. Of course, a good model should be expected to require only small control inputs, in some comparative sense.

Despite a superficial similarity of mathematical form, the two approaches are quite different conceptually and in subsequent mathematical development. Unfortunately, some researchers assume, or become confused into thinking, that $u_t$ can treated like $\xi_t$. This has lead to instances in the literature where methodologies have been promoted that are incorrect and do not solve the problem they claim to solve.
2. A taxonomy of filters

Figure 3 provides a schematic classification of filters that have been developed over the fifty years since Kalman’s original paper. At the heart of the forecasting problem is the relationship between state and observations: $s_t = g(x_t, \eta_t)$. Given a sequence of observations $s_t$, find the appropriate states $x_t$ assuming the model (1) or (2). This is often referred to as state estimation, data assimilation, or solving the inverse problem. The underlying model (1) or (2) separates the stochastic and deterministic formulations.

### Stochastic formulations

The stochastic formulation is often seen to be subsumed into the wider Bayesian estimation theory. Mathematically, the Bayesian approach to filtering is sequential and can be expressed as follows. Let $M$ be the model state space. Knowledge about the system is represented by a probability density on the model state space. Let $S_t$ represent all observations up to time $t$, and let $\rho(x \mid t', S_t)$ be a probability density for $x \in M$, representing knowledge about the system at time $t'$, given only the observations $S_t$ and the model. The model dynamics allow evolving (present) knowledge $\rho(x \mid t, S_t)$ forward to time $t'$, obtaining $\rho(x \mid t', S_t)$. Usually, the model is such that the evolution of the density $\rho$ can be expressed in terms of transition probabilities $q(x, t, x', t')$, which gives the probability a system represented by model state $x$ at time $t$ arrives in model state $x'$ at time $t'$. In which case,

$$\rho(x' \mid t', S_t) = \int_M q(x, t, x', t')\rho(x \mid t, S_t) dx. \quad (3)$$

Employing Bayes rule and the Chapman-Komolgorov equations, obtains

$$\rho(x \mid t', S_t) = \frac{P(S_t \mid x)\rho(x \mid t, S_t)}{P(S_t \mid S_t)}, \quad (4)$$
where \( P(S_{t'} | x) \) represents the uncertainty about observations given a model state \( x \), and \( P(S_{t'} | S_t) \) is effectively a normalization factor to ensure the left hand side is a probability density. It can be shown that Eq. (4) provides an optimal filter, given some fairly general assumptions about the transition probabilities and observational uncertainties [8].

Kalman’s original filter is a special case of this optimal filter with \( t' = t + 1 \), the model evolution a linear stochastic process, with additive Gaussian perturbations, and observational uncertainties that are Gaussian and additive. The elegantly succinct form of Kalman’s original filter arises because an initially Gaussian density \( \rho \) always evolves into another Gaussian density, and a Gaussian density is completely described by its mean and a covariance matrix. Generalisations of this succinct formulation are moment filters that attempt to capture the effects of non-linearity using higher order moments of the distributions.

If the model is nonlinear, then there is generally no simply described algebra of densities closed under the evolution. The linear-Gaussian situation is quite unique in this respect. In one dimensional state spaces, densities are easily approximated by kernel methods, like Gaussian mixtures. For dimensions of two or more it becomes increasingly difficult to describe and approximate densities; for example, the number of kernels required to achieve a given accuracy grows as the fifth power of the dimension [9]. The practical difficulty with using an optimal sequential Bayesian filter is that one either does not know, or cannot easily represent, the transition probabilities \( q \) or the state density \( \rho \), when the model evolution is nonlinear and the state space more than one dimensional. This gives rise to the great variety of different filters, which attempt to provide useful approximations to an optimal sequential Bayesian filter with minimal resources and computation. These filters typically adopt an ensemble-based technique where a finite sample of states are used to represent, or approximate, the probability distributions. Common approaches are SIR filters and particle filters.

A common property of all such filters is that they are sequential filters: the density \( \rho(x_t | S_t) \) completely encodes all assumed knowledge about the system state \( x_t \) at time \( t \). Equation (4) provides a sequential update of current knowledge when the next observation \( s_{t+1} \) is obtained: explicitly,

\[
\rho(x_{t+1} | S_{t+1}) = \frac{\Pr(s_{t+1} | x_t) \rho(x_t | S_t)}{\Pr(s_{t+1} | S_t)}.
\]

Deterministic formulations

The deterministic approach to tracking and forecasting dates back at least to the mathematician and astronomer Pierre-Simon Laplace and his work on probability theory and celestial mechanics; in particular, determination of orbits of planets and comets. Laplace was first to clearly articulate Newton’s idea of a state space when 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. In Laplace’s conception the universe is 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 like Eddington were therefore aware of this problem for the complex processes of gases and weather. The mathematician George Birkhoff, who was in many respects Poincaré’s intellectual successor, linked the deterministic to the stochastic through ergodic theory and hence laid the foundation for the valid application of stochastic methods to deterministic systems.

Deterministic approaches to tracking and forecasting existed well before Kalman published his paper in 1960. These methods had evolved from Laplace’s early work and were often referred to as least-squares methods. After Kalman’s paper many subsequent papers derived Kalman’s filter from least-squares principles, failing to appreciate that Kalman’s approach provided a much greater wealth of information, such as an analysis of error sensitivity. By the time Jazwinski wrote his accessible book a decade later [8], Kalman’s approach was dominant, especially with the growth of engineering applications in aerospace and process control.
Before and after Kalman’s paper, mathematicians, such as Andronov [10], Arnold [11], Pontryargin and others developed a sophisticated deterministic theory of dynamics, control, and variational methods which generalize least-squares methods. Much of this theory is applicable to tracking and forecasting. Deterministic approaches lived on after Kalman’s paper in optimal control and variational methods. Of particular note are methods developed and implemented by meteorologists for operational weather forecasting models [12, 13]. This is an astonishingly difficult problem: complex highly-nonlinear models with tens of millions of variables. Kalman filters are not practical in this situation without introducing assumptions that make the covariance matrices sparse; for example, by assuming that covariances are zero for variables at locations that are sufficiently spatially separated. Assumptions of this nature are reasonable for some variables, but not all, as the atmosphere has long-range coupling through wave motions. It is ironic that in recent years some meteorologists have become increasingly interested in Kalman filters, whereas engineers have become increasingly interested in chaos in low-dimensional systems like Lorenz considered. Nevertheless, the weather forecasting system used by the European Centre for Medium Range Weather Forecasting (ECMWF), which is indisputably the best in the world, uses a deterministic variational method often referred to as 4D variational assimilation (4DVA) [13].

There are many ways to formulate a variational equation for forecasting purposes. The following is a typical shadowing filter: given a finite sequence of observations \((s_t)_{t=0}^{n}\), with additive Gaussian noise \(\eta_t\), and a deterministic model (2), then minimise,

\[
\begin{align*}
\text{minimize} & \quad \sum_{t=0}^{n} \| s_t - g(x_t) \|^2 \\
\text{subject to} & \quad x_{t+1} = f(x_t, u_t), \quad 0 \leq t \leq n, \\
\text{with} & \quad \| u_t \| \leq K, \quad 0 \leq t \leq n.
\end{align*}
\]

The parameter \(K\) is an upper bound on the expected model errors \(u_t\). The solution of this optimisation is a sequence of states \(x_t\); in a perfect model scenario \(K = 0\) and the sequence of states is a trajectory of a deterministic model. A variational approach like this is a best-guess method; the last state of the trajectory provides a best-guess state estimate and forecast.

**Stochastic vs deterministic**

An important distinction between the stochastic approach to filtering and the deterministic approach is the object of focus: a stochastic approach to filtering focuses on a probability density of the current state \(\rho(x_t \mid S_t)\); a deterministic approach to filtering focuses on the trajectory up to the present time. As a consequence, variational methods, like 4DVA and shadowing filters, require no intrinsic uncertainty information and as such are best-guess methods, in the sense that they obtain a single state that represents the best-guess of the true state in a perfect model scenario. On the other hand, most stochastic approaches are ensemble-based, like particle filters and ensemble-Kalman filters. Although Kalman’s original filter requires an auxiliary covariance estimate, it provides an explicit best-guess.

Uncertainty information is vital to many forecasting needs, so stochastic approaches would appear to have an advantage by intrinsically providing uncertainty information, but any best-guess method can be augmented with an ensemble formation scheme, which typically employ some perturbation of the best-guess or trajectory to obtain an ensemble of states. For example, ECMWF augments the 4DVA with carefully controlled perturbations of the best-guess in directions of singular vectors of the linearised flow: the singular vectors with maximum singular values represent the directions of maximum linear growth, which are assumed to capture the directions of most sensitivity to initial conditions.

The difference in focus between the stochastic and deterministic approaches has non-trivial consequences. In the stochastic approach to filtering the primary object of concern is a density of one state \(\rho(x \mid t, S_t)\), whereas in the deterministic approach the equivalent object of concern, in the perfect model scenario, is a density of trajectory segments \(\rho(x(\tau) \mid 0 \leq \tau \leq t, S_t)\), where \(x(\tau)\) represents a segment of a trajectory of the model. It is true that one can obtain from \(\rho(x(\tau) \mid 0 \leq \tau \leq t, S_t)\) a marginal
density of the final state \( x(t) \), which is equivalent to a density \( \rho(x \mid t, S_t) \), however, optimization of the marginal density \( \rho(x \mid t, S_t) \) is not the same as optimizing the joint density \( \rho(x(t) \mid 0 \leq t \leq T, S_t) \). For example, the methods obtain different maximum likelihood final states. Consequently, stochastic and deterministic methods are both valid, but give different results because they optimize different criteria.

**Convenient untruths**

Before leaving this overview of the taxonomy of filters represented schematically in Fig. 3, it is worth mentioning what not to do. Not displayed in Fig. 3 are a couple of methods often referred to as *Weakly Constrained 4DVA* or *Kalman smoothers*. These methods purport to apply variational methods to the stochastic formulation (1). For example, if observational noise \( \eta_t \) is assumed to be Gaussian and additive with variance \( \sigma^2 \), and dynamical noise \( \xi_t \) assumed to be Gaussian and additive with variance \( \zeta^2 \), then given a finite sequence of observations \( (s_t)_{t=0}^n \) some folks propose finding \( (x_t)_{t=0}^n \) to minimise,

\[
\frac{1}{\sigma^2} \sum_{t=0}^n \|s_t - g(x_t)\|^2 + \frac{1}{\zeta^2} \sum_{t=0}^n \|x_{t+1} - f(x_t)\|^2.
\]  

(7)

At first sight this is an appealing and convenient method; it appears to obtain a best guess taking into account both observational noise \( \eta_t \) and dynamical noise \( \xi_t \). Unfortunately, it does not, because the observations have insufficient degrees of freedom. In order for this method to obtain the desired goal one needs sufficient observations to establish the unknowns \( x_t \) and the implicitly eliminated unknowns \( \xi_t \); otherwise the implied values of \( \xi_t \) are not independent, which violates an implied assumption in the formulation of the objective (7).

Compared to (6), minimising (7) is equivalent to assuming \( x_{t+1} = f(x_t) + u_t \) and replacing the uniform bounds on \( u_t \) with a sum of squares constraint \( \sum_{t=0}^n \|u_t\|^2 \leq n\zeta^2 \). In a deterministic imperfect model context this is a reasonable and legitimate assumption, and there is no contradiction in the optimal corrections \( u_t \) being not independent. However, it is important not to be confused by the form of (7): it is not equivalent to a stochastic formulation, nor is it equivalent to assuming model errors are Gaussian, which would be unreasonable.

When correctly interpreted, variational equations such as (6), of which minimising (7) is a special case, provide a deep insight into imperfect model scenario when combined with the powerful theory that has been developed from Anosov’s shadowing lemma ideas. A duality feature explains why, for example, gradient descent methods are unexpectedly successful.

**3. The Isis Programme**

The Isis Programme (Imperfect-model Shadowing and Indistinguishable States) is an ambitious research programme that began around 10 years ago by the author and Leonard A. Smith. The overall goal of this programme is to provide theory and tools to deal with the three challenges of modelling, forecasting, and management of complex non-linear systems. A central principle of the Isis Programme is acceptance of the fact that models of complex systems are necessarily imperfect and observations are always inadequate to achieve anything usefully approaching perfection. We argue that if one truly accepts that a model is imperfect, then one must accept that many methods of classical theory are inadequate or inapplicable. One of the insights of the Isis Programme is that there are considerable rewards in viewing complex systems in terms of imperfect nonlinear deterministic models [14], introducing stochastic elements only as a last resort, or for specific purposes [15, 19]; this contrasts with viewing complex systems as stochastic systems from the outset.

The Isis Programme has created a wealth of new theory that provides easily implemented practical techniques [16] that are superior to presently popular statistical techniques often seen as the “new way forward” [7, 14, 19]. Our methods have been shown to be applicable to highly nonlinear chaotic systems [15, 19], to larger complex systems for which computational resources impose severe limitations on practical algorithms [17], to unprecedented application to operational weather forecasting models [18, 19]. Explanations for this success of our methods have been derived from theory [14,
19, 20, 22], along with new insights into convergence rates of our algorithms and the effects of model error [20, 22]. One of the great surprises of the Isis Programme is just how effective our methods are; bettering classical methods that are supposed to be optimal.

The two central components of the Isis Programme are Shadowing filters, that obtain best-guess trajectories, and Indistinguishable states, that provide an efficient methodology for understanding and quantifying uncertainty.

**Shadowing and shadowing filters**

Shadowing is a concept introduced into dynamical systems theory by Dmitri Anosov in 1967 [21], which has been greatly extended since. In general, shadowing theorems prove that for dynamical systems with certain special properties, if a sequence of states \((y_t)_{t=0}^n\) is close to being a trajectory, say \(\|y_{t+1} - f(t)\| \leq \alpha\), then there is an actual trajectory \((x_t)_{t=0}^n\) close by \(\|x_t - x_t\| \leq \beta\); we can say that \(x_t\) shadows \(y_t\). Although shadowing is stated here for a finite sequence of states, for some very special properties, like hyperbolicity, shadowing occurs for infinite sequences; on the other hand, shadowing is a generic property of continuous maps. Shadowing theorems often state explicitly the relationship between \(\alpha, \beta, \) and \(n\).

Shadowing is a powerful tool. It can be used to prove the existence of trajectories with prescribed properties, by linking together snippets of shorter segments of the trajectory. It can be used to show that although numerical computation always have round-off errors, there is a actual trajectory close to the numerically computed sequence. It can be used to show that if two dynamical systems have similar dynamics \(f_1(x) \approx f_2(x)\), then they have the same trajectories. It can be used to show that an imperfect model is consistent with finite-accuracy measurements of a system.

A shadowing filter takes a sequence of states \(y = (y_t)_{t=0}^n\) and a perfect model, and obtains a new sequence of states \(x = (x_t)_{t=0}^n\) that is close to \(y\) and close to being a trajectory of the model — strictly speaking we would prefer a trajectory, but since algorithms typically employ iterative methods, one never attains complete convergence, nor typically needs to. Shadowing filters naturally generalise to an imperfect model scenario: instead of aiming for \(x\) to be a shadowing trajectory, one only aims to remove the effects of observational noise from \(y\), and have \(\|x_{t+1} - f(x_t)\| \leq \gamma\) as well as the shadowing conditions involving \(\alpha, \beta\) and \(\gamma\) previously stated. Shadowing theorems exist that relate the quantities \(\alpha, \beta\) and \(\gamma\) in this imperfect model scenario [22]. The quantity \(\gamma\) is a measure of the model error, and \(u_t = x_t - f(x_{t-1})\) are the implied optimal corrections to keep the imperfect model \(f\) tracking reality.

There are various ways to implement a shadowing filter, one of the most straight forward approaches is gradient descent of indeterminism, \(I(x) = \sum_{t=0}^{n-1} \|x_{t+1} - f(x_t)\|^2\), that is, initially let \(x = y\), then move \(x\) down the gradient of \(I(x)\). This is a remarkably effective scheme: it works well for simple strongly non-linear chaotic systems (see Fig. 2) up to operational weather forecasting models [18, ? , ]; details for implementing such a shadowing filter are found elsewhere [16].

**Stochastic vs deterministic again**

From the point of view of forecasting, one clearly cannot expect a weather forecasting model to make reliable four-day forecasts if it cannot shadow previously observed weather for four-day periods. The ability to shadow past observations is a prerequisite of forecasting. Furthermore, if one has a model and a solution \(x\) that shadows recent observations \(s\) up to the present time, then the final state of the shadowing \(x\) ought to be a good candidate state for running out a forecast. Clearly, in the perfect model scenario, shadowing filters can provide a best guess of the trajectory leading up to the present, and hence a best-guess state from which to launch a forecast. As Fig. 2 illustrates, for non-linear systems these deterministic filters perform better than stochastic filters when the non-linearity is significant relative to the observational noise.

One of the greatest surprises is using deterministic shadowing filters on non-linear stochastic systems obtains best-guess state estimates that are better than stochastic filters using optimal sequential update: that is, one obtains better state estimates by assuming the system is deterministic, even though the system is known to be stochastic.

This apparent contradiction occurs in situations when the observational noise is larger than the
dynamical noise. With a little thought this begins to make sense. Recall the dynamical noise perturbs the state at every time step, for example, $x_{t+1} = f(x_t) + \xi_t$ and the observational masks the true state, for example, an observation $s_t = x_t + \eta_t$. Clearly, if the state perturbations $\xi_t$ are larger, on average, than observational errors $\eta_t$, then most of the uncertainty about the state in due to dynamical noise perturbations $\xi_t$; under these circumstances a sequential updating stochastic filter is appropriate. On the other hand, if the observational errors $\eta_t$ are larger than the dynamical noise perturbations $\xi_t$, then most of the uncertainty about the state is due to the observational noise; under these circumstances a shadowing filter can out perform a sequential stochastic filter. The success of the shadowing filter in this case derives from use of information from the past, that is, the shadowing filter finds the joint density of a sequence of states is more useful than the marginal density of the final state used by a sequential stochastic filter.

**Forecasting and uncertainty**

The great 20th century philosopher of science Karl Popper argued that science cannot reveal *naked truth*, there is always objective uncertainty, and consequently a scientific theory, or model, cannot be validated as true by consistent observations, only falsified by inconsistent observations. Popper went on to argued that scientists should make *accountable* models by quantifying the model’s objective uncertainty; useful models are accountable in that their predictions are consistent with their objective uncertainty. Weather forecasters certainly perceived the need for accountable forecasting models and introduced ensemble forecasting schemes in an effort to quantify the uncertainty of their forecasts. A seeming advantage of stochastic filters and ensemble-based filters, like particle filters, is that they explicitly work with the probability distribution of the underlying state estimate, hence, the uncertainty of the current state and forecasts is built-in to the theory and methodology. Shadowing filters, which work with a best-guess trajectory, do have this feature. As it turns out, shadowing provides a natural means to incorporate objective uncertainty, which in practice is more informative than ensemble-based filters, like particle filters.

Consider a perfect model scenario first. Suppose we have a forecast model $f(x)$, a trajectory $x = (x_t)_{t=0}^n$ that shadows a sequence of observations $s = (s_t)_{t=0}^n$, and we know statistics of our observational errors $\Pr(s_t|x_t)$, which is a testable property of our measurement instruments. The shadowing trajectory should be consistent with the observations in the sense that given the known statistics of the observational errors, then this trajectory could have given these observations. In particular, we can state the probability of the observations occurring $\Pr(s|x) = \prod_{t=0}^n \Pr(s_t|x_t)$.

It is important to keep in mind that a shadowing trajectory is not unique; there are many other shadowing trajectories $y = (y_t)_{t=0}^n$ that are also consistent with the observations $s$. Furthermore, recall the two key features of chaotic systems of sensitivity to initial conditions and the physical realisable states lying on the attractor. If $n$ is suitably large, then an alternate shadowing trajectory consistent with the observations can be found by choosing $y_0$ close to $x_0$. Furthermore, if $n$ is sufficiently large, all the shadowing trajectories consistent with the observations should lie in, and spread out across, the attractor.

In classical statistics one would aim to find a maximum-likelihood trajectory $\hat{x}$, that is, the trajectory that maximises $\Pr(s|x)$. One might have devised a shadowing filter to do this, but often there is not a great advantage in striving for this goal: clearly, once one suitably long shadowing trajectory $x$ has been found others can be easily obtained by suitable perturbations of the initial state $x_0$. Without too much trial and error and ensemble of shadowing trajectories can be formed that are weighted according to $\Pr(s|x)$ and provide objective uncertainty information.

In a practical forecasting scheme, one would evolve forecasts from each ensemble trajectory, reweight the ensemble trajectories when new observations arrive, discard trajectories that become too unlikely, and initialise new trajectories as needed. However, unlike a Bayesian sequential ensemble filter that requires a large ensemble to accurately maintain a representation of the underlying probability distribution, the shadowing trajectory approach does not: one instead maintains just one central shadowing trajectory that is updated using the shadowing filter at each step, and then perturbs this when necessary to maintain an ensemble. That is, the shadowing trajectory contains all the important
In the space of observations the observations of the model states $x$, $y$ and $z$ have probability densities $\rho(s, x)$, $\rho(s, y)$ and $\rho(s, z)$ for an observation $s$ are represented by shaded regions; observations outside the shaded regions have zero probability of occurring for those states. Any observation $r$ distinguishes $z$ from $x$ and $y$, because there is zero probability this observation for states $x$ and $y$. By the observation $p$ the state $x$ can be distinguished from state $y$, because $p$ has zero probability for state $y$, but given an observation $q$ the two states $x$ and $y$ are indistinguishable.

Indistinguishable states

The concept of shadowing leads to the implementation of shadowing filters that when given a stream of observations obtain a core shadowing trajectory and allow construction of an ensemble of shadowing trajectories to assess forecast uncertainty. This is an entirely practical forecasting scheme applicable to simple chaotic systems, up to enormously complex operational weather forecasting models. To properly implement such a forecasting scheme requires theoretical tools to guide the construction of ensembles in practical and efficient ways. Furthermore, such a scheme should adapt easily to an imperfect model scenario. An important concept for doing this is indistinguishable states.

We can ask first what an observation can tell us about possible model states, for example, when can an observation distinguish which of two states is a possible state. Assume a (perfect) model and knowledge of the nature of observational error, in particular, if the model state is $x$, then an observation $s$ of $x$ should have a probability density $\rho(s, x)$ on observation space. Figure 4 illustrates that an observation $s$ can distinguish which of two model states $x$ and $y$ could have occurred only if $\rho(s, x)\rho(s, y) = 0$; otherwise, the states are indistinguishable by the observation. This idea can be generalised to model trajectories $x$ and $y$ given any expected sequence of observations $s$: we say that $x$ and $y$ are indistinguishable, if

$$\lim_{n \to \infty} \prod_{\tau=0}^{n} \int \rho(s, y_{t-\tau})\rho(s, x_{t-\tau}) \, ds > 0. \quad (8)$$

Note that if any one factor in the product is zero, then the entire product is zero, that is, a single observation can be sufficient to distinguish two trajectories. If no factors are zero, then the limit is almost always zero, that is, one expects that either there is eventually a single observation that distinguishes the two trajectories, or the evidence accumulates to distinguish trajectories in the limit. On the other, the limit can be non-zero, that is, there is a non-empty set of trajectories $y$ that cannot be distinguished, regardless how many observations are made. In the simplest case of additive Gaussian observational noise it can be shown [15] that $x$ and $y$ are indistinguishable if
Fig. 5. Forecasting a black swan of the Ikeda system. The true trajectory states marked by large cross-hairs. Observations are circles, where the large circle shows the standard deviation of Gaussian error. A two-sigma black swan event occurs at $t = 4$. (a) A particle filter using 1000 particles (small dots) and their forecasts for $t = 5$. (b) Crosses mark computed shadowing trajectory for $t = 1$ to 4 and forecast for $t = 5$. (c) A small indistinguishable state ensembles that were calculated at each observation arrival, and the forecast for $t = 5$. (d) The shadowing filter recovers easily given $t = 5$ observation and corrects the states at $t = 1, 2, 3, 4$.

$$\sum_{\tau = -\infty}^{t} \|y_{\tau} - x_{\tau}\|^2 < \infty,$$

which requires that indistinguishable trajectories must converge running backward in time. This implies the set of all $y$ indistinguishable from $x$ lie in the unstable manifold of $x$.

The indistinguishable states provide a link between uncertainty and dynamics. Define $Q(y|x)$ to be the left-hand side of (8). With suitable normalisation, $Q(y|x)$ is the expected probability that trajectory $y$ is indistinguishable from the trajectory $x$. Hence, if $x$ is a putative shadowing trajectory, then $Q(y|x)$ provides the expected uncertainty associated with this shadowing trajectory given the observational uncertainty $\rho$. This tool has many useful functions: it provides an appropriate prior for ensemble generation, it allows effective statement of forecast uncertainty, it allows easy and efficient exploration of extreme events, and it generalises naturally to imperfect models.

Black swans and beyond

The proverbial tale “For the want of a nail” or the more modern form of a butterfly flapping its wings in Amazon causing a Typhoon to landfall in Japan, captures the concept of sensitivity to initial conditions. The medieval origin of the “For the want of a nail” story reveals that the concept existed long before Lorenz mathematically captured the true essence of the phenomenon in chaos. There is another concept that pervades our thinking that appears in common phrases and notions such as, “out of the blue” and “left field”; this is the concept of a Black Swan, an event that is unexpected, surprising and out of the ordinary [23]. The notion of a black swan dates back to ancient time; before the discovery of Australia it was a statement of obvious truth that “all swans are white”. Mathematically, this idea has been expressed as a three, or five, sigma event, which expresses the possibility of an event far out in the tail of a Gaussian bell curve. Such events have small, but nonethe-less non-zero probability of occurring. As Oscar Wilde put it: “expect the unexpected” [24]. We will look at the implications of black swans for forecasting in a moment, but there is a modern variant of this ideal that goes one step further. Referred to as unknown unknowns, or unk-unk, the idea arose in aerospace engineering and the military; unk-unk was famously referred to by Donald Rumsfeld, there are known knowns, there are known unknowns, but there are also unknown unknowns [25]. This statement sets out the three important aspects of forecasting: known knowns represent our model, known unknowns represent uncertainties that can be represented by probability distributions, and
unknown unknowns (unk-unk) represent uncertainties of model error. Black swans can cause havoc for the unprepared, and can even utterly undermine a forecasting scheme, but model error can be much worse.

In the perfect model scenario a sequential Bayesian filter appears to work well, provided the system is low-dimensional and the ensemble that represents the probability distribution is large. In high-dimensional systems it can be shown that surprising small ensembles can capture truth most of the time [26]. Problems arise when the ensemble does not capture truth, because then there is a poor forecast, but worse still the entire filter can fail and needs to be restarted, because truth is so far from the ensemble. Figure 5(a) illustrates how even with a large ensemble of 1000 points a particle filter fails to capture the true state; at this point the filter collapses and needs to be restarted. Making the ensemble larger helps to avoid failures, but this example is not a particularly rare black-swan and only a two dimensional model; to capture low probability events requires vast ensembles, whose size grows exponentially with the model’s dimension. Panel (b) shows that as expected a shadowing filter too is misled by the black-swan observation, making a poor forecast, although, panel (c) shows how only a small ensemble of indistinguishable states is required to reveal the range of uncertainty. Panel (d) shows that, unlike the particle filter, a shadowing filter easily recovers when the new observation arrives. Note how the shadowing filter corrects the entire trajectory, recognising a black-swan as the source of the error. Since an indistinguishable state ensemble can be assessed by the quantity $Q(y|x)$, one can efficiently explore forecast uncertainty and the nature of possible black-swan events before they occur, without a vast ensemble.

Low-probability events are a nuisance, but they can be dealt with in a mathematically rigorous way; the unknown unknowns of model error cannot. Model errors cannot be assumed to be random; often they are systematic and state dependent. Often model error is explored in the perfect-model class scenario, where parameterised model class contains a perfect model, but the correct parameters are unknown. Although useful, this approach does not capture the possibility that reality often has levels of physical processes not considered in the model; that is, the system is coupled to an entire subspace of dynamics that is ignored. Taking a deterministic modelling perspective these two cases can be generalised to a simpler case of model whose attractor is in the wrong shape and position compared to reality, and a more complex case where the model only attempts to approximate a projection of Reality. In both cases shadowing theory tools can detect and counteract model errors. Anosov and others following him have developed theory about what an imperfect model can and cannot do.

It is important to appreciate that if one accepts a model has errors, then there is no sensible concept of a true state, nor any sense of estimating the true state; the problem of state estimation is replaced by a less clear-cut problem of identifying useful states. In the perfect model scenario the true state is optimal in every sense for every purpose; with an imperfect model different states may be optimal for different purposes, for example, for forecasting as opposed to guidance control. To be useful in any sense, however, a deterministic model must be able to shadow reality at least for short periods of time. Alternatively, one can allow small adjustments to states to enable the model states to track reality. Ideally each adjustment is precisely the accumulated model error, but, of course, such adjustments are unknown unknowns. In practice, one can obtain optimal corrections, that is, is some sense the smallest adjustments that keep the model tracking reality to within some accuracy. The existence of such optimal corrections is closely related to shadowing theory. At the present time this theory is being actively developed.

4. Conclusion
Kalman created a state-based filter that is optimal for linear stochastic systems. The Kalman-Bucy extension remains useful for local linearisation provided the non-linearity is small relative to the noise. Kalman clearly saw that non-linear systems would require something quite different from his filter design. It was Lorenz who revealed the extent of the difficulties in dealing with non-linear systems by investigating deterministic forecasting. Generalisation of the Kalman filter leads to the sequential Bayesian filters, which assume all the important information about a the system is captured in a probability distribution $Pr(x_t|s_t)$ for the current state $x_t$ given past observations $s_t = (s_{\tau})_{\tau \leq t}$. 247
Although optimal in theory, such filters have limitations in practice because of the impossibility of representing continuous probability distributions with finite ensembles. We now know that shadowing filters based on a deterministic model perform better in practice if the observational noise level exceeds the dynamical noise level. An important difference between shadowing and Bayesian approaches is that the Bayesian approach works with the marginal distribution $Pr(x_t|s_t)$ of the current state given the past observations $s_t$, while the shadowing approach works with the joint distribution $Pr(x_t|s_t)$ of a trajectory of states $x_t = (x_\tau)_{\tau \leq t}$, which effectively allows the filter to correct past mistakes caused by misleading observations. Furthermore, although a shadowing filter implements a best-guess, the existence of an attractor that manifests the physically realisable states, and sensitivity to initial conditions, implies that there exists a set of indistinguishable states, which can be easily constructed and analysed to determine the uncertainty of the current state and forecasts.

Of course, in practice, one should always use the appropriate tool for the job. Taking into account information acquired, the observational errors, the dynamical noise, the non-linearity of the model, and the nature of model error, then one can choose the appropriate filter and the computational resources required. One should not choose to use a simple Kalman filter, just because it is easy to compute, because this risks catastrophic failure. Nor should one implement an extensive ensemble-based scheme, when only a reliable best-guess is required, such as in guidance controllers.

Choosing an appropriate filter can be reduced to four parameters: $\sigma_{\text{obs}}$ the standard deviation of the observational noise, $\sigma_{\text{dyn}}$ the standard deviation of the dynamical noise, $\rho$ the radius of curvature of the non-linearity, and $\epsilon$ a bound on the anticipated model error. If $\epsilon \ll \sigma_{\text{obs}}, \sigma_{\text{dyn}}, \rho$, that is, in a very nearly a perfect model scenario, then the appropriate filter is indicated by the following table:

| Conditions          | Filter         |
|---------------------|----------------|
| linear ($\rho = \infty$) | Kalman         |
| $\sigma_{\text{obs}}, \sigma_{\text{dyn}} \ll \rho$ | Kalman–Bucy   |
| $\sigma_{\text{dyn}} < \rho < \sigma_{\text{obs}}$ | Shadowing      |
| $\rho, \sigma_{\text{obs}} < \sigma_{\text{dyn}}$ | Bayesian        |

If $\epsilon > \sigma_{\text{dyn}}$ and is not small, then one is in the imperfect model scenario and must take the model error into account. Simply assuming that model error can be treated like dynamical noise is not appropriate. At present, only a shadowing approach seems to offer a consistent and justifiable methodology to operate in the imperfect model scenario.

The Isis Programme with its emphasis on non-linear dynamics provides a simple set of tools to implement accountable forecasting of nonlinear systems that is effective, robust and can be generalised to the imperfect model scenario. At present it appears to be the only approach that offers a consistent and justifiable methodology to account for model error, although the theory is still in development.

References
[1] E.N. Lorenz, “Deterministic nonperiodic flow,” J. Atmos. Sci., vol. 20, pp. 130–141, 1963.
[2] B. Franklin, Poor Richards Almanack, Peter Pauper Press, 1758/1950.
[3] W. Grimm, Vridankes Bescheidenheit, Dieterich-schen Buchanblung, Göttingen, 1834.
[4] K. Ikeda, “Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system,” Optical Communications, vol. 30, p. 257, 1979.
[5] R.E. Kalman, “A new approach to linear filtering and prediction problems,” Journal of Basic Engineering (Series D), vol. 82, pp. 35–45, 1960.
[6] R.E. Kalman and R.S. Bucy, “New results in linear filtering and prediction problems,” Journal of Basic Engineering (ASME), 83D, 1961.
[7] K. Judd, “Nonlinear state estimation, indistinguishable states and the extended Kalman filter,” Physica D, vol. 183, pp. 273–281, 2003.
[8] A.H. Jazwinski, Stochastic Processes and Filtering Theory, vol. 64, Academic Press, New York, 1970.
[9] B.W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman and Hall, London, 1988.
[10] A.A. Andronov, A.A. Vitt, and S.E. Khaikin, *Theory of Oscillators*, Pergamon, London, 1966.
[11] V.I. Arnold, *Geometric Methods in the Theory of Ordinary Differential Equations*, Springer-Verlag, New York, 1982.
[12] P. Courtier, E. Andersson, W. Heckley, J. Pailleux, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, and M. Fisher, “The ECMWF implementation of three-dimensional variational assimilation (3d-var) i: Formulation,” *Q. J. R. Meteorol. Soc.*, vol. 124, pp. 1783–1808, 1998.
[13] F. Rabier, H. Jarvinen, E. Klinker, J.F. Mahfouf, and A. Simmons, “The ECMWF operational implementation of four-dimensional variational assimilation — part I: Experimental results with simplified physics,” *Q. J. R. Meteorol. Soc.*, vol. 126, pp. 647–677, 2000.
[14] K. Judd and T. Stemler, “Failure of sequential Bayesian filters and the advantages of shadowing filters,” *Physical Review E*, vol. 79, 066206, 2009.
[15] K. Judd and L.A. Smith, “Indistinguishable states I : perfect model scenario,” *Physica D*, vol. 151, pp. 125–141, 2001.
[16] T. Stemler and K. Judd, “A guide to shadowing filters for forecasting and state estimation,” *Physica D*, vol. 238, pp. 1260–1273, 2009.
[17] K. Judd, L.A. Smith, and A. Weisheimer, “Gradient free descent: shadowing and state estimation with limited derivative information,” *Physica D*, vol. 190, pp. 153–166, 2004.
[18] K. Judd, C.A. Reynolds, and T. Rosmond, “Toward shadowing in operational weather prediction,” Technical Report NRL/MR/7530-04-18, Naval Research Laboratory, Monterey, CA, 93943-5502, USA., 2004.
[19] K. Judd, “Chaotic-time-series reconstruction by the Bayesian paradigm: Right results for the wrong reasons,” *Physics Review E*, vol. 67, 026212, 2003.
[20] K. Judd, “Forecasting with imperfect models, dynamically constrained inverse problems, and geometrically modified gradient descent,” *Physica D*, vol. 237, pp. 216–232, 2008.
[21] D.V. Anosov, “Geodesic flows and closed Riemannian manifolds with negative curvature,” *Proc. Steklov Inst. Math.*, 90, 1967.
[22] K. Judd, “Shadowing pseudo-orbits and gradient descent noise reduction,” *Journal of Nonlinear Science*, vol. 18, pp. 57–74, 2008.
[23] N.N. Taleb, *The Black Swan: The Impact of the Highly Improbable*, Random House, 2007.
[24] Oscar Wilde, *An Ideal Husband*, Dover, 2001.
[25] Donald Rumsfeld, *Known and Unknown: A Memoir*, Penguin, 2011.
[26] K. Judd, L.A. Smith, and A. Weisheimer, “How good is an ensemble at capturing truth?,” *Q. J. Roy. Met. Soc.*, vol. 133, pp. 1309–1325, 2007.