PROCEEDINGS A

rspa.royalsocietypublishing.org

Review

Cite this article: Farmer CL. 2017 Uncertainty quantification and optimal decisions. *Proc. R. Soc. A* **473**: 20170115.
http://dx.doi.org/10.1098/rspa.2017.0115

Received: 16 February 2017
Accepted: 21 March 2017

Subject Areas:
applied mathematics, computer modelling and simulation, statistics

Keywords:
uncertainty quantification, forecasting, stochastic control

Author for correspondence:
C. L. Farmer
e-mail: farmer@maths.ox.ac.uk

An invited contribution to a special feature 'Mathematics in the modern economy'.

THE ROYAL SOCIETY PUBLISHING

Uncertainty quantification and optimal decisions

C. L. Farmer

OCIAM, Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK

CLF, 0000-0003-3783-495X

A mathematical model can be analysed to construct policies for action that are close to optimal for the model. If the model is accurate, such policies will be close to optimal when implemented in the real world. In this paper, the different aspects of an ideal workflow are reviewed: modelling, forecasting, evaluating forecasts, data assimilation and constructing control policies for decision-making. The example of the oil industry is used to motivate the discussion, and other examples, such as weather forecasting and precision agriculture, are used to argue that the same mathematical ideas apply in different contexts. Particular emphasis is placed on (i) uncertainty quantification in forecasting and (ii) how decisions are optimized and made robust to uncertainty in models and judgements. This necessitates full use of the relevant data and by balancing costs and benefits into the long term may suggest policies quite different from those relevant to the short term.

1. Introduction

Making a decision or setting a control entails a choice with the aim of reaching an objective as closely as possible. In turn, a forecast is needed that predicts the outcome of each control setting so that an informed choice can be made. In many cases, there is only sparse knowledge about the starting state of the system of interest and only approximate knowledge of how the system will evolve. In many situations, the task of making decisions and setting controls is left to intuition and experience. However, when a situation goes beyond everyday experience, mathematics can be surprisingly effective.

For example, the production of oil from a geological formation located at a depth of 2–5 km below Earth's surface often involves injecting water into the formation through a system of wells. Oil and gas is thus pushed

© 2017 The Author(s) Published by the Royal Society. All rights reserved.

towards production wells and is recovered from the mixture of oil, gas and water that flows to the surface. The remoteness of the formations means that there is uncertainty in the geometry and properties of the rocks and so the design of the system of wells has to account for many different possible shapes of layers and faults and many different possible patterns and levels of porosity and permeability. The initial design and later decisions involve matters such as the number of wells to be drilled, their location and path and the operating flow rates or pressures. The initial uncertainty, hopefully reduced by making measurements at the large scale using seismic waves, and at the small scale by measuring the properties of rocks along the wells, needs to be quantified. With such a quantified uncertainty, the subsequent decisions can then be made in a way that accounts for the different unknown possible situations. Then, on balance, the recovery plan is optimal on average over the different possible scenarios. These are examples of the optimal decisions to which we refer in the title of this paper.

This example of the oil industry, to be discussed in more detail after an overview in general terms, will indicate how mathematics is used throughout the process of finding oil and gas and the subsequent recovery of oil and gas through the system of wells. We will also make comparisons with other applications such as weather forecasting, precision agriculture and revenue management.

In our discussion of the 'industrial mathematics workflow' the first notion to be explained is that of a mathematical model. We then describe how deterministic and probabilistic forecasts are made, how such forecasts are evaluated, how data are assimilated so as to improve future forecasts and how an optimal decision or control can be calculated.

Throughout the article the term *control policy* will be used to mean either

(i) a *deterministic* control policy, which is a mathematical function taking as its arguments all of the relevant observations and delivering as its value a *decision*, or
(ii) a *stochastic* control policy, which is a stochastic process delivering decisions that are randomly assigned with probabilities that depend on the relevant observations.

Stochastic control theory seeks methods for constructing control policies through the formulation and solution of optimization problems. The classification of types of control policy will be of particular value in helping decision-makers understand better how they might benefit from the assistance of the mathematical sciences.

One topic we will not discuss is that of competition, negotiation or conflict between two or more decision-makers. In some situations, a particular decision-maker can treat the other, competing, decision-makers as part of the system they are aiming to control. In general, though, the existence of competitors requires the introduction of new ideas from game theory, such as those described in [1], which we will not cover. Stochastic control policies tend to be used in the context of competing agents, but could find application in the control situations considered in the following.

An important aim of this article is to emphasize the distinction between *open-loop* control policies that make all the decisions at the outset and *closed-loop* control policies that make the decisions at the time for action, thereby responding to changing circumstances as they are observed.

Throughout the article, the terms *systematic approximation* and *heuristic approximation* will be used. A systematic approximation is an approximation that can be systematically improved by performing further calculations. Heuristics are approximations that have no scientific basis for systematic improvement. Very often a heuristic is adequate and sometimes can even outperform the known systematic methods.

2. Modelling

The first task in mathematical modelling is to describe the state of the system of interest; for example, a physical system such as the atmosphere or an economic system such as a supermarket.

In the following, we will assume that the state of a system can be described by an array of numbers which will be called the *state vector* or just the *state*. We will refer to any particular number in the array as a component of the state vector. The number of components in the state vector is referred to as the *dimension* of the system. We will also assume that the state dimension is finite, and that time proceeds in discrete steps of equal duration, *time*-0, *time*-1, . . . , *time-n*, . . . and so on. In practice, mathematicians often derive such models as approximations to an infinite-dimensional continuous model, but the end result in most applications is a finite-dimensional algorithm to be used on a computer. This decision to frame the discussion in finite-dimensional terms is to avoid extensive use of mathematical notation and the need for technical details. However, on occasion some notation will be used as a summary of the preceding textual explanation.

The developers of simulation software often think only in terms of finite-dimensional models, but it is always useful if we can find exact 'closed-form' solutions for simple infinite-dimensional limits. Such exact solutions, particularly in the form of easily understandable formulae, can provide insight into the behaviour of more complicated models and provide benchmarks for checking computer codes. One example of the interplay between numerical approximations and exact solutions is provided by the advection equation [2]. This equation models the way a contaminant is transported in a fluid, and is a particularly difficult equation to approximate with a finite-dimensional approximation.

In many applications, the state vector is built by dividing the system into a network of connected parts. For physical systems, this involves dividing the space of interest into a grid of 'cells'. Within each cell in a physical system a small set of numbers will be used to quantify variables such as pressure or temperature. In a social system the 'cells' might be people, or households, connected to one another in a social network with variables such as wealth or occupation.

In the following, the equations used to relate the state vector at one time with the state vector at the preceding time will be called a *forward model*. Note that the term 'forward model' is not universally used in this general sense, and is only used to mean the equation that gives the state vector at *time-n* as a function of the state vector at *time*-$(n-1)$. In practice, the equations are not usually an explicit set of instructions but a large set of simultaneous equations connecting the state components at a particular time in a particular cell with the state components of the neighbouring cells, and the corresponding state components at the previous time step. These equations must then be solved in some iterative fashion on a computer, and the software that we use to perform these calculations is called a *simulator*.

As time evolves, some of the components of the state will change, but many might be left unchanged. The ones that change, following the terminology of weather forecasters, we will call *variables*. Components of the state vector that are constant in time will be called *deterministic parameters* and stochastic components that only appear in the equations at one particular time will be called *stochastic parameters*. The *control variables*, which represent physical devices, such as valves, or design parameters in a system to be manufactured, are adjustable parameters that may be used to optimize some part of the output of the forward model.

When forward models, at each time step, depend on a large subset of components of the state vector that are stochastic and independent from each other and independent from the components at other times we say that the model is a *stochastic model*. When such stochastic components are absent from the model, we say that the model is *deterministic*.

This notion of a state vector is not standard in all domains, and such a state vector would be referred to in some subjects as an *augmented state vector*.

Let us use (i) the symbol v_n to denote the variables, the time changing components of the state vector at *time-n*, (ii) the symbol u_n to denote the time-independent components, the deterministic parameters where the forward model for the deterministic parameters is simply $u_n = u_{n-1}$, and (iii) the symbol w_n to denote the stochastic parameters. If the system is stochastic the complete state vector at *time-n* is $\psi_n = (u_n, v_n, w_n)$. If the system is deterministic the state vector is just $\psi_n = (u_n, v_n)$.

We can summarize the notion of a forward model by noting that in the most general case it provides the probability density of a state given the immediately preceding state. This can be written as the formula $\pi(\psi_n \mid \psi_{n-1})$, where π is a general symbol for a probability density function and where the subscript n is a time index. This formula is to be read as the 'probability density of the state at *time*-n conditioned on the state at *time*-$(n-1)$' and the vertical bar is to be read as 'conditioned on' or as 'given'. In the following where we use notation, we follow the convention of the statistics literature where the same symbol, in this case π, is used to denote different functions depending on their arguments.

Many models have hierarchical structures, wherein at the largest scale the number of parts is relatively small but each part is composed of smaller parts that respond on shorter time scales. In turn, the smaller parts could be composed of yet smaller parts and so on. If a state vector can be assigned to the largest parts alone, then a simulation through time is possible even on a modest computer. However, the initial state vector might have been specified on a smaller scale, so that in the larger grid cells a detailed description using smaller cells is available. However, if the detailed state, using information on all of the scales, is used in the simulator, the amount of computer memory required to solve the simulator equations may be larger than that which is available. In such cases, it is therefore necessary to develop averaging techniques[1] and perhaps even modify the simulator equations so that the model only uses the state specification at the largest space and time scales.

Simulation software, either open source or commercial, is now widely available. For many applications such as weather or oil reservoir forecasting, the effort needed to write software can be of the order of a hundred person-years. However, the value of such software to the economy arises more from the value to users, through forecasting, system optimization and decision-making, than from the immediate value from licensing or selling the simulation software. Experience has shown that advances in numerical simulation are as much down to improvements in algorithms as to improvements in computer technology.

3. Forecasting

In the special case of *deterministic forecasting*, when we know the initial state vector for a forward model in its entirety and the model is deterministic, a computer can calculate a unique update of the state at each time step to provide a forecast.

However, in most applications, we do not know the initial state vector and, in many applications, the model is also sometimes stochastic. Since the presence of stochastic parameters in the forward model and the absence of knowledge of the initial state are both instances of not knowing something we are concerned in general with *probabilistic forecasting*, which we now discuss.

If the initial state of a system is not known, we collect all the relevant information in order to build a probability density function—called our *prior* density—that summarizes our initial knowledge. We could, for example, assume that knowledge of one particular component is independent from all the other components. Then our uncertainty as to the value of that component could be summarized by assigning a mean and a variance to quantify our degree of certainty regarding the mean. The Gaussian density function with its familiar bell-shaped graph is a useful and common starting assumption, and the joint density for the entire state vector would then be given as a product of the Gaussian densities for the individual components. In a more general case, we might believe there to be correlations between the components. To model this, we could assign cross-covariances to build a multivariate Gaussian density.[2]

An important property of the Gaussian density is that its maximum value occurs when the state is equal to the mean state. Another property is that the covariance matrix of the Gaussian

[1]This is the subject known variously as *homogenization*, *upscaling* or *parametrization* [3–6].

[2]When we believe that densities are not Gaussian, one approach is to transform our state components and characterize them as functions of some other state vector that does have a Gaussian probability density.

density is the inverse of the matrix of the second derivatives of its negative logarithm. These two properties play a key role in the next stage of the workflow, as described later on.

If we wish to forecast from an initial state prescribed by a prior density, rather than a particular state vector, we might try to forecast using the mean state of the prior as a starting state and if the model is stochastic setting the stochastic parameters to their mean values. This is possible, but, except in the special case that the forward model is linear, it is not a good idea, as when we make a forecast we want to know the mean state of the future and not the future of the system starting at the mean state. A better approach is to sample from the prior density, and use a large ensemble of different state vectors for the initial conditions. Most sampling algorithms will ensure that each state in the ensemble is selected with equal probability and so the future ensemble constructed by processing each ensemble member using our simulator will again have states with equal probability. We can then compute the mean and variance of the future ensemble and so obtain approximate forecast statistics. The forecast will have a sampling error that could be reduced by increasing the size of the ensemble.

It is possible to estimate the probability density of a state given an ensemble of states with prescribed probabilities in a process called *density estimation* [7]. Such estimated densities will often be in the form of a weighted sum of Gaussian densities, in general multivariate. Each ensemble member becomes the mean of a contribution to the Gaussian sum and some appropriate covariance matrix, such as a diagonal matrix, is needed for each contribution. As the number of ensemble members is increased and the covariance matrices have decreasing variance the estimated density is a systematic approximation to the original density from which the ensemble has been drawn [7].

In turn, the ensemble of states can be used to produce an ensemble of forecasts of the observations to be made. This can be done by using a forward model of the measuring apparatus to calculate a vector of predicted observations for each state in the ensemble. Again, this ensemble can be used to construct a joint probability density for the vector of observations. This construction of a probability density is essential in the next step where the forecasts are evaluated for accuracy. Note that we can make probabilistic forecasts of any combination of state vector components and observations, but it is only the forecasts of the observations themselves that can be tested against reality.

4. Forecast evaluation

An interesting question now arises. If an ensemble of forecasts has been made, and one or two of the forecasts turn out to have been accurate (when compared with the actual observations as they become available) and the rest of the ensemble made mediocre forecasts, was the ensemble good or bad as a forecast? A literature has grown around this question and there are many methods of scoring a forecast. An example of such a scoring method, for a probabilistic forecast, is to first of all perform density estimation to recover a forecast probability density for the observations from the ensemble. The *logarithmic score* is then given by the negative logarithm of the density evaluated using the observation values. The smaller the logarithmic score, the more accurate is the forecast. By accumulating a sum of such scores over a sequence of time steps, we can compare one forecasting system with another, using the same observation values. The difference in scores can then be used to indicate which models and priors provide the best forecasts.

A detailed discussion of such scoring has been given in the edited volume of Jolliffe & Stephenson [8]. An extensive investigation into forecast evaluation, in the context of insurance pricing, has been undertaken by Maynard [9], who suggests that several different types of score should be used together when assessing the fidelity of probabilistic forecasts.

5. Observations and data assimilation

Once a forecast has been made and scored, we are in a position to assimilate the data from the latest observations by updating our probabilities for each ensemble member before moving

to the next forecast. The terminology in this subject is not standardized, and, in this paper, we use the term *uncertainty quantification* for the whole process of (i) making a probabilistic forecast, (ii) scoring or evaluating a forecast, and (iii) data assimilation. When we talk of 'data assimilation' we mean the process of building and analysing the probability density of the states given particular values of the observations.

(a) Making an observation

Measurement is rarely a simple matter of reading a number from a digital display. Indeed, many measurements are indirect and can only be inferred where a mathematical model is available. For example, if we are measuring the permeability or elasticity of rock, we need a model of the measuring instrument and its interaction with the system under observation. Such measurements are always affected by stochastic noise and so measurement models are inherently probabilistic.

In the following, the words 'observation' and 'measurement' are often synonymous. However, there is sometimes a difference. In the usage followed here, an observation is always direct in that it is the actual recording made by a measuring instrument. By contrast, a measurement, as already indicated, can be indirect in that the numbers that are reported are the result of an inference using an observation, and a model that relates the observation to the quantity to be measured. In many cases such measurements are localized in space and time and can be completely decoupled from the more general problem of measuring a state vector. An example of this is provided by the measurement of the permeability of a porous material. The measurement involves flowing a fluid through a sample of the material and observing the pressure drop for a controlled flow rate. The permeability is then inferred using a simple formula derived from combining the principle of mass conservation with Darcy's law that relates the flow rate, the pressure drop and the fluid viscosity. In turn, observation of the pressure drop or the flow rate might involve a further layer of modelling and inference. We thus see a complicated process that we usually ignore and take for granted. So, in the following we will use the words 'observation' and 'measurement' to mean much the same but bearing in mind the hidden depths of the scientific process.

The measurement model can often be thought of as predicting a measurement value that is a known function of the system state plus some stochastic noise with zero mean and known variance. Thus, the measurement model provides the probability density of the observation conditioned on a particular system state. However, only the probability density of the system state is given and hence we need the joint probability density of the observation and state which can be forecast by taking the product of the two densities.

Let us denote the vector of observations at *time-n* by s_n. The observation is then a function, h, of the state vector plus some random disturbance, ξ_n, of the form $s_n = h(\psi_n) + \xi_n$, where ξ_n is sampled from a density depending on some of the control variables. This can be summarized in the formula $\pi(s_n \mid \psi_n)$, which can be read as the 'probability density of the observation given the state'. Once the observation has been made, the formula $\pi(s_n \mid \psi_n)$ reduces to just a function of the state.

(b) A framework for data assimilation

Once we have chosen a forward model, an observation model and an initial probability density for the state we can construct the most general statement of the outcome in the form of the joint probability density of all of the states and all of the observations. This can then be reduced by integrating over various choices of the states at different times to provide, for example, the 'filtering density', which is the probability density of the current state given all of the observations up to and including those at the current time. Another density of interest is the 'smoothing density', which is the probability density of some combination of past states (or even of particular components of past states) depending upon all of the observations up to some time later than the current time. The forecasting density, which we discussed in the previous section, is the

probability density of a future state given observations up to some fixed past time. (In the earlier discussion this past time was simply the initial time.)

We can summarize this using some notation, which makes the structures we are discussing much easier to comprehend. The general expression for the joint density can be written as $\pi(\psi_{0:N}, s_{0:M})$, where $\psi_{0:N}$ is the sequence of all the state vectors from *time*-0 to *time*-N and $s_{0:M}$ is the sequence of the observations from *time*-0 to *time*-M.

By integration over all state vectors other than ψ_N one can find $\pi(\psi_N, s_{0:M})$. When $M = N$ we have the *filtering density*, when $M < N$ we have the *forecasting density* and when $M > N$ we have the *smoothing density*. For further discussion, see [10–12].

(c) Bayes' theorem

The joint density of the observations and the states can be decomposed in two different ways. These are:

(i) the *probability density of the observations given the states, multiplied by the probability density of the states*, or
(ii) the *probability density of the states given the observations, multiplied by the probability density of the observations*.

Bayes' theorem [13] is the statement that these two decompositions are equal.

Thus, our general density function can be written in the two equivalent forms,

$$\begin{aligned}\pi(\psi_{0:N}, s_{0:M}) &= \pi(s_{0:M} \mid \psi_{0:N})\pi(\psi_{0:N}) \\ &= \pi(\psi_{0:N} \mid s_{0:M})\pi(s_{0:M}).\end{aligned}$$

The first line of this equation is known from the forward model, the observation model and the prior, but the second equation, in general, is not known directly because the term $\pi(s_{0:M})$ is very expensive to calculate.

Once we have made a measurement and know the values of the observations, Bayes' theorem enables us to calculate the *posterior density*, which is the *probability density of the states given the observations*. The technical term *likelihood* is used for the function obtained by evaluating the probability density of the observations given the state evaluated at the particular value of the observations after they are known. Bayes' theorem then reveals that the posterior probability density of the states given the observations is proportional to the product of the likelihood and the prior density.[3] Thus, the consequence of Bayes' theorem is that we can find, in principle,

$$\pi(\psi_{0:N} \mid s_{0:M}) = \frac{\pi(s_{0:M} \mid \psi_{0:N})\pi(\psi_{0:N})}{\pi(s_{0:M})}.$$

There are methods, such as Markov chain Monte Carlo (MCMC) [16], for sampling from the posterior, $\pi(\psi_{0:N} \mid s_{0:M})$, that avoid the need to calculate the term $\pi(s_{0:M})$.

We emphasize that the application of Bayes' theorem is not a way of extracting theories or models from data: it is just a procedure for updating probability densities. As we have indicated in the previous section, we need to perform forecast evaluation to quantify the fidelity of models and probability densities. There is no systematic procedure for improving models. Indeed, constructing models at the start of a project and then improving models as the need arises requires imagination, experience, skill and collaboration.

[3]Bayes' theorem leads to some surprising and counter-intuitive results. These often relate to asking questions about the probability that something has happened given a related observation in the context of prior information. For example, the *base-rate* fallacy [14] and the *Monty Hall* brain teaser [15].

(d) Smoothing methods for data assimilation

The posterior density, as is apparent from the preceding discussion, is a very complicated mathematical object. It depends on very many—perhaps millions—variables and parameters, and in general has a very complicated shape with many local maxima. Thus, a summary of the information encapsulated in the posterior density is required for conceptual and also for computational reasons. These summaries are in the form of a *deterministic summary* given by, for example, a state that is the position of a local maximum of the density, or a *stochastic summary* given by an ensemble of samples.

(i) Deterministic summaries

The term *mode* is used to mean a local maximum of a probability density and the term *objective function* is used to mean the negative logarithm of the density. In the case of the Gaussian density, the mode is unique, a global maximum of the density and equal to the mean. If there is only one mode, then we say that the density is mono-modal and the second derivatives of the objective function give an estimate of the inverse of the covariance matrix. It turns out that it is practical for us to compute the mode of a mono-modal density and also a local mode when there is more than one mode. In general, there are many modes and this non-uniqueness needs to be quantified and understood if data assimilation is to be useful in practice. The calculations involved in finding modes of a density rely on efficient optimization methods such as the quasi-Newton method [17] that only require the first-order gradient of the objective function with respect to the states at each time. In practical problems where the state dimension is very large it is fortunately feasible to compute gradients using a method known as the *adjoint method*, which uses a forward run of the simulator, storing the state vectors at each time step, and a backwards in time run of a related linear model that can be coded in an efficient way. The adjoint method enables the computation of gradients with a small percentage overhead on a forward run and is one of the most important numerical algorithms in applied mathematics [18]. Without the adjoint method many methods for data assimilation (and, as we shall see, for making decisions) would be infeasible.

For completeness let us note that the deterministic summary approach to data assimilation is, in many cases, equivalent to the older method of *regularized inverse problems* as developed in the applied mathematics literature [10]. The advantage of the Bayesian formulation is, however, that there is a clear method—through summarizing prior knowledge in the prior density—for deriving the regularization term that is, in fact, the negative logarithm of the prior density.

Once the mode has been computed we require at least an estimate of the diagonal terms in the covariance matrix. These diagonal terms are the individual variances in the posterior density of the initial state vector. Progress has been made [19] in finding approximate covariance matrices but it would seem that the difficulty of this problem leads the majority of researchers to think that ensemble methods are the way forward.

(ii) Ensemble summaries

The usual method for summarizing a density in statistics and theoretical physics is to draw an ensemble of samples from the density using a controlled random walk in a method known as MCMC [16]. In statistics, the models can be of low enough state dimension, and, in theoretical physics, the models can be sufficiently simple (even though of very large dimension), for the MCMC method to be practical. In most uncertainty quantification problems such as oil reservoir data assimilation and weather data assimilation, MCMC is not practical unless a very good starting state is available. This might be a state obtained using a conventional optimization method, as used for deterministic summaries, but suitably roughened by adding random disturbances in a controlled way to the observation values and the prior mean. This is the heuristic method of *maximum randomized likelihood* [18], which in some applications might be adequate even without the refinement of the rigorous MCMC method [20].

(e) Sequential methods for data assimilation

If, at any time step, we have a representation of the latest probability density of the state given all previous observations, our forward model enables us to find the forecast density for the next observations, and the posterior density once the observations are made. In principle this sequential update, which is called *filtering*, solves a central problem in uncertainty quantification. When the forward model is linear, and the prior and all stochastic parameters have Gaussian probability densities, filtering can be shown to generate a sequence of Gaussian probability densities for the prior and posterior filtering densities at each stage of the sequence. This sequence, which for the linear-Gaussian case is known as the *Kalman filter*, is of fundamental importance to many disciplines. However, when the state dimension is very large, the covariance matrices that are generated are dense matrices and cannot be stored without compression and loss of information.

When the forward model is nonlinear the Kalman filter is not applicable, but heuristic approximations based on linearization of the forward model can be used to generate various versions of the *extended Kalman filters* [21]. Although these filters are sometimes useful, in general they are unreliable and can be unstable in ways that are hard to understand and difficult to cure. Thus, in recent years attention has turned to a class of methods that use an ensemble to represent the probability densities. These methods go by the name of *ensemble Kalman filters* and have many variants [21]. Most of the variants are heuristic but some are systematic [22]. The key ideas in these methods are that (i) an ensemble of state vectors is available that can be used to find the ensemble mean and the covariance of the probability density, (ii) the forward model can be used to generate a forecast by updating each ensemble member over one time step, and (iii) the ensemble covariance is used in a linear operation to adjust each ensemble member to better agree with the latest observations.

Unfortunately, limitations in computer memory force ensemble methods to use ensembles that are too small. As a result the ensemble estimates of the correlations between different components of the state vector are in error. This then leads to the need for heuristic techniques to obtain reasonable results, particularly for problems where large parts of the state vector (such as all of the components of a particular physical type) are not observed. When the problem is largely one of interpolation the method is quite useful, but in general ensemble Kalman filters are not reliable. In particular, the changes made in adjusting the ensemble of states to agree with the observations can generate artificial transients. Sometimes the transients lead to instability and destroy the predictive skill of the forecast step and so destroy the value of the filter. Thus, much current research aims to improve the situation in data assimilation. For example, into the use of sequential ensemble methods for solving the smoothing formulation of the data assimilation problem [23].

(f) Hybrid sequential–smoothing methods

In practice, a hybrid method is used, sometimes called *long-window 4D-Var*. 'Long-window' refers to a window of time that is chosen to be short enough that there are not too many observations during the window, but long enough for the posterior density to be relatively insensitive to approximations in the prior. '4D-Var' refers to the use of an optimization method for finding a mode of the density—sometimes called a variational (Var) method—in three-dimensional space and one-dimensional time. The prior for each window is an approximation to the posterior at the end of the previous window. By making the window long enough we can mitigate the effect that the approximate 'prior' is not consistent with the observations used for the earlier windows. This long-window method represents the state of the art in uncertainty quantification [24]. By increasing the length of the window we obtain a systematic method that combines the virtues of both sequential and smoothing approaches.

The modern literature on these hybrid methods demonstrates many incremental refinements in our capacity to solve uncertainty quantification problems. However, there is still much room for improvement, for example if the system is chaotic or has wave-like behaviour. Nevertheless, very

high-dimensional uncertainty quantification problems are solved routinely, albeit approximately, in a way consistent with Bayesian principles.

6. Decision and control in the presence of uncertainty

(a) Guidelines for stochastic control

In practical situations, our theoretical aim is to define policy as functions of the potentially available observational data. Optimal decisions at a particular time are made by setting control variables to the values of the control policies evaluated using the data. In the previous section concerned with data assimilation, it was assumed that the observations had been designed and made and that any control variables in the dynamical system had specified values. In this section, we consider the problems of (a) how we choose the values of the controls for the target system and also (b) how to control the observations. The control of the observations involves (i) when and where to make measurements, (ii) what measurements to make, and (iii) the accuracy with which to make the measurements.

As most practical applications involve elements of uncertainty, the problem of making optimal decisions is mathematically equivalent to an application of stochastic control theory.[4]

To determine an optimal policy we first need to define a function that measures the cost, in some sense, of implementing any given policy. This is another situation that calls for mathematical modelling: in this case driven by economic and statistical theory. The simplest measure of cost is to sum the costs over the time horizon of interest and then take the expectation value with respect to the stochastic quantities in the dynamics, the observations and the policies. This particular expected cost is a basic *risk-neutral cost* and from now on we will assume that minimization of the total expected cost is the stochastic control problem to be solved.[5] We note that the length of the time horizon enables the decision-maker to balance short-term with long-term costs in a principled manner and that the choice of time horizon is an important factor in decision-making.

Let us suppose that we have designed a sequence of control policies where each policy is parametrized by a vector of *policy parameters*. Thus, the policy function will be a function of the past observations and the policy parameters. We can then sample the initial system state, the deterministic and stochastic parameters and the observation noise. If the policy is to be a stochastic policy we also need to sample the probabilities inherent in the policy. Then, using the forward models for the system and the observations we can, for any selection of policy parameters, by taking an average over the ensemble of simulations compute the expected cost for those policy parameters. Note that, in the cost at each time step, the cost of operating the system and making observations can be included. The problem is then to choose the policy parameters so that the total expected cost is a minimum. This last problem is a problem in the numerical optimization of functions and we say a little more about this in the later section on optimizing measurement costs.

When making a decision it is best to follow, *as closely as is practical*, the guidelines:

— build a mathematical model of your system;
— choose the cost function to be minimized;
— choose the control variables in the system;
— choose the control variables for the observations you are going to make;
— use uncertainty quantification to estimate a probability density for the current state of the system;

[4]However, once again, terminology is not standard, and the terms *sequential decision theory* (statistics), *dynamic programming* (operational research), *reinforcement learning* (artificial intelligence), *real options theory* (mathematical finance) and *recursive economics* (economics) are used to refer to stochastic control. Powell [25] discusses the problem of terminology in some detail.

[5]In practice, economic considerations relating to risk-aversion or risk-seeking predispositions require the use of *disutility functions* [26] rather than basic costs. Disutility functions are closely related to utility functions [27], of which there is much discussion in the literature on how to decide which utility functions to use and how to include discounting in the total utility [28,29].

— use forecast evaluation to quantify the quality of your forward models and priors; in particular, look at the probabilistic forecasts of the costs that are to be controlled and evaluate the accuracy of the forecast of the cost;
— evaluate forecasting scores and control policies using a variety of forward models and priors in an exercise of model comparison, criticism and sensitivity analysis [13];
— assume that the decisions to be made at each time step are values of functions—the *control policies*—of the relevant observations at that time and perhaps some randomness (if probabilistic policies are to be used);
— choose the control policies that minimize the expected value of the total cost; in many cases, this can be done efficiently using the adjoint method [18]; and
— work on improving models and priors in the light of poor forecasting scores or undue sensitivity to subjective judgements of probability.

Any control policy that gives us the minimum possible expected cost is said to be *optimal* and any other policy that leads to a higher expected cost than the minimum possible is said to be *suboptimal*. Each possible policy will have an associated expected cost, so the above exercise provides a ranking of the policies from which we can select the best. It thus follows that the forecasts only need to be accurate enough to provide a useful ranking of the possible policies.

Of outstanding importance in the decision-making workflow is the need to use imagination and research to explore the range of control options available and to assess the costs and consequences of each option [30]. In many applications, our understanding of the system is sparse, and so the critical aspects of the decision-making workflow are crucial.

In practice, however, the procedure presented above is exceedingly difficult to follow. There are several obstacles arising from (i) the generally high dimension of the state vectors, (ii) the difficulty of parametrizing the control policies so that the number of parameters is relatively low, (iii) the enormous quantity of observational data, and (iv) the enormous computational task of estimating the expected costs for a given set of policy parameters. It is thus necessary for us to employ heuristics in computing the control policies just as in modelling and uncertainty quantification. Nevertheless, when the guidelines are translated into mathematical notation the statement of the problem is succinct and clear [31,32]. It is evident, even from the informal description above, that there will be an enormous variety of different heuristics for making sequential decisions. Stochastic control theory is a rigorous mathematical theory, but in applications we need to explore the performance of various families of heuristics in choosing an algorithm for a particular application. The literature on the topic of stochastic control is vast. However, there are some key ideas that dominate, and in the remainder of this section we will describe one such idea—that of stochastic model predictive control (SMPC)—that is beginning to be used in large-scale engineering applications such as oil reservoir engineering. We believe that methods such as SMPC have wide applicability and relevance.

(b) Stochastic closed-loop model predictive control

Suppose that we are part of the way into a project, having already made a sequence of decisions concerning the control of system operation and the control of observations. Suppose, too, that a complete record of all decisions and all observations has been kept. Then at the time we need to make our next decision we:

— perform data assimilation to construct an ensemble of current states conditioned on the past observations;
— use one of the many numerical optimization methods [17] to find a sequence of future control values that finds a local minimum, or, even better, a global minimum of the total expected cost;
— implement the first of the future control values in the real system, discarding all of the subsequent control values;

— observe the system as the next time step unfolds;
— evaluate the forecast of the observations; and
— repeat the process.

Note that, in the above, the control policies appear to be specific numerical values and not functions as in the ideal workflow. However, since the process takes into account the history of the past observations and decisions it is a *feedback control policy* in that the decisions are influenced by observations as they occur. Such a policy is known as a *closed-loop policy*. In other words, contrary to first appearances, we have computed the value of a policy function without having an explicit formula for the policy. Such a procedure takes into account our uncertainty about the present state and any uncertainty in the future evolution arising from uncertainty in the system. The method is, however, suboptimal, as it does not provide a method for controlling the cost of the future measurements.

For completeness, we remark that sometimes we might not discard all of the subsequent control values and thus not need to repeat the process. We might simply implement the controls that were calculated at the initial time. Such a policy, known as an *open-loop policy*, is far from optimal and ignores the future observations. However, in order to save on computing costs we might consider using a mix of open-loop and closed-loop policies. This is how a multitude of different approaches can be developed such as the simplest heuristic of deterministic model predictive control, where uncertainty is ignored and the above procedure is used with an ensemble of just one state vector. The importance of computing such closed-loop policies is that they respond to the actual observations as they arrive and provide a degree of robustness to mis-specification of the forward model. Thus, the emphasis in the industrial mathematics workflow should be more on the accuracy of the relative rankings of the costs associated with the different policies than on the detailed accuracy of forecasts of the observations.

(c) Optimizing measurement costs

If we have implemented a closed-loop model predictive control system we still need to make decisions about where and when to make observations. Suppose you have decided on a particular control policy for such decisions. For example, you might decide to make more observations or fewer observations depending upon the posterior variance of your ensemble. Such control policies might even be parametrized and the problem is then generalized to include the observation control policies in the list of optimal policies to be determined. For any particular value of the policy parameters the total expected cost of such a policy can be estimated. This could be done by (i) sampling from the prior and (ii) using that sample as the 'truth' in a numerical experiment where the observations are made using the 'truth' state, updated using our forward model. Such an experiment can be repeated several times with different samples of the 'truth'. One can then estimate the expected cost or even the probability density of the total cost of operating that particular policy.

In principle, this approach can be used as the basis of an optimization procedure. In this case, the optimization might be far too difficult even for the adjoint method. However, if the number of parameters can be chosen to be quite small we can use the method, sometimes known as the method of *optimization of expensive functions*[6] [33,34]. Here one uses one more outer loop of Bayesian decision-making in order to choose the optimal parameters for the actual observation controls. Indeed, the approach can in principle be used for choosing any feedback control policy for the observations and the system, provided that the number of parameters in the feedback policy is not too large. This has not, as far as the author is aware, been exploited but is a possible addition to heuristic methods such as SMPC, enabling one to optimize the control of the observations. An important application of such ideas could be to the problem of designing inspection schedules for risk-based asset management [35]. Indeed, this application is

[6]Also known as *optimization using surrogates* or *optimization using emulators*.

one of several occurring in the area of uncertainty quantification and management in high-value manufacturing, where all elements of the mathematical workflow are present.[7]

7. The oil industry

As a further illustration of the mathematical workflow in action, we again cite the example of the oil industry [4]. Oil and gas exploration begins with seismic surveys where, for example, at sea a seismic boat sends pulses of sound into the sea and recordings are made of the small fraction of sound energy reflected from the rocks below the seabed. The first task is to unravel these reflected sound vibrations into a three-dimensional map of the rock properties. This is the first of the uncertainty quantification problems that must be solved. The state vector for the propagation of elastic waves in the subsurface is the first part of the state vector that must be inferred. The controls in seismic exploration include the choice of the initiating sound source and the source location. The measuring instruments, in the seismic case the geophones attached to long streamers at the stern of the seismic boat, number in the thousands and their location is determined from satellite global positioning.

In this particular inverse problem, a mathematical model is needed that describes how sound energy propagates. A candidate model would be the classical theory of elasticity as developed by mathematicians in the eighteenth and nineteenth centuries. To a first approximation, the classical wave equation can be used. Generalizations have been developed in which the effect of fluid in the pores is also modelled. However, whichever model is used, the procedure is that one postulates a description of the spatially varying rock properties and then one makes a forecast of what would be observed by the geophones. This forecast requires the use of massive computing power in which the seismic components of the state vector are adjusted in an iterative process until reasonable agreement with observations is achieved. The catch though, as with most inverse problems, is that there are many possible state vectors that predict the observations with equal accuracy, as was discussed in the section on deterministic summaries of posterior densities. Traditionally, and even in the current state of the art, only one state vector is found. However, there is a growing realization that more needs to be done in quantifying the uncertainty in the seismic state by building an ensemble of possible state vectors [20,36] for input to other stages of the workflow.

Given one or more solutions to the seismic inversion problem, the interpretation of the state vectors requires the knowledge of petroleum geologists and reservoir engineers to identify volumes that might contain commercial quantities of hydrocarbons.

The most basic recovery technology—known as *primary recovery*—involves wells from which fluid can flow under the natural pressure in the reservoir. However, this is not an efficient recovery method if the aim is to recover as much hydrocarbon as economically viable. In *secondary recovery*, water or gas is injected in some wells, and the hydrocarbons are pushed towards other wells, the production wells. This process maintains the reservoir pressure serving to (i) control the physical chemistry of the reservoir fluids and (ii) prevent subsidence of the seabed. *Tertiary recovery* techniques involve more exotic injection fluids such as surfactants, but these are not economic unless the oil price on world markets is at a high level.

Once the decision is made to develop a field, the first well might be positioned to intersect the largest connected volume of hydrocarbon so that some initial high production is possible with potential to offset the huge costs already incurred in the oilfield development. Even while drilling a well, many measurements of the rock properties are possible. After drilling, even more measurements are possible, and if you can imagine a physical property, then the chances are that it can be measured. This is the fascinating subject of well logging [37], which is a major industry in its own right. Electrical, nuclear and direct fluid mechanical measurements are routine. Also sampling of rock and fluids is possible even at depths of kilometres. Each type of measurement

[7]See the Uncertainty Quantification and Management Special Interest Group website (https://www.ktn-uk.co.uk/interests/uncertainty-quantification-and-management).

requires a mathematical model and some level of data assimilation. Once again mathematical theory is an essential component of the workflow and the key to integrating multiple types of measurement into a coherent whole for application in later decision-making. After some flow in the wells is established, even more observations become available to be used in updating the relevant components in the state vectors [18].

An oilfield-operating consortium has to decide on the position and well trajectories for perhaps hundreds of wells over the life of an oilfield. This is a demanding example of a decision and control problem where until recently the deterministic model predictive control approach held sway. There is now more interest in SMPC [38]. The application of stochastic control theory enables oil companies to balance short-term income against long-term income while allowing for uncertainty. Indeed, by injecting water into a reservoir at higher rates, short-term oil production can be increased, but this leads to lower than optimal recovery factors as a result of earlier breakthrough of water in the production wells. This in turn means that more drilling or more water injection over a longer period of time is needed to generate the same total recovery that would have pertained had the earlier rate of recovery been lower. This is of importance to the oil companies and perhaps of even more importance to governments that wish to encourage oil companies to maximize the total recovery of oil from an oilfield.

8. Conclusion

Now we know how mathematics is used, in general, and in the particular example of oilfield management, we can see how the same ideas, albeit with differences of emphasis, occur in many domains. One might speculate that mathematics has a role in any policy-making where numerical quantification has been or should be used.

For example, in the practice of weather forecasting there are many sources of data, ranging from satellite-based measurements, balloons, rockets, ground-based observations, ship-based observations and floats in the oceans. Weather forecasting has driven the very idea of simulation [39,40], a drive that continues to this day. Inverse problems have always been of central concern [41] and the topic of forecast evaluation had its origins in weather forecasting. The topic of decision and control is less prominent although the problem of choosing the density and/or frequency of measurements has been studied [42,43].

The methods of precision agriculture where science and mathematics play a central role in agricultural management are of increasing interest [44,45]. The problems where remote measurements are made from instruments carried by drones, samples are taken from the fields, simulators of crop growth are constructed and stochastic control problems about crop rotation and the application of fertilizers and pesticides are solved are very similar to problems encountered in oil recovery.

As a final example, we mention the topic of revenue management where the language of the modelling is so different from that of oil recovery or agriculture. However, the mathematics induces a one-to-one correspondence between any two of the different areas. In revenue management, where the problem is to devise a control policy for setting prices and other control variables in the sales process, we again see most of the different topics of modelling, forecasting and control from the industrial mathematics workflow [46]. A specific example is that of how to reduce prices when selling surplus inventory—sometimes called the 'markdown problem' [47].

One might suggest that, when a particular topic from the ideal mathematical workflow is lacking development or exploitation in a particular domain, it is a signal for further investment to strengthen that topic within the domain.

Mathematics is one of many factors in the provision of goods and services. Mathematics has an irreplaceable role in the economy, where its value is partly from the software and services it provides but is mainly from the added value to those industries that make use of such software and services. Perhaps, too, mathematics as a key ingredient in decision-making has potential that is yet to be fully exploited.

In the previous sections, we have reviewed some of the background for concluding that

— *Effective simulation needs mathematics*. Making decisions and choosing control policies is central to economic activity. Simulation plays an essential role in evaluating the effect of choosing particular options, and mathematics is at the heart of useful and practical simulation.
— *Knowing the ideas of statistical decision theory is empowering*. Everyone involved in decision-making and in supporting decision-making should be aware of the basic principles of uncertainty quantification and optimal decision-making. Knowledge of engineering control theory is valuable too. Models and forecasts only need to be good enough for us to rank our options. That is, we should emphasize policy evaluation rather than forecast evaluation.
— *Collaboration is essential*. The formulation and solution of problems requires experience and knowledge of several domains. By working collectively, through mechanisms such as study groups[8] and workshops [48], where particular practical problems are the focus, we can benefit from the skills of others, and gain encouragement to develop our own skills and knowledge.

Competing interests. I declare I have no competing interests.
Funding. I received no funding for this study.
Acknowledgements. I wish to thank my colleagues in industry, the Oxford Mathematical Institute and the Oxford Martin School for their encouragement and their criticism. I also wish to thank three reviewers for their comments and suggestions. The University of Oxford provided library and office facilities.

References

1. Raiffa H. 2007 *Negotiation analysis: the science and art of collaborative decision making*. Cambridge, MA: Harvard University Press.
2. Leveque RJ. 2002 *Finite volume methods for hyperbolic problems*. Cambridge, UK: Cambridge University Press.
3. Farmer CL. 2002 Upscaling: a review. *Int. J. Numer. Methods Fluids* **40**, 63–78. (doi:10.1002/fld.267)
4. Farmer CL. 2005 Geological modelling and reservoir simulation. In *Mathematical methods and modeling in hydrocarbon exploration and production* (eds A Iske, T Randen), pp. 119–212. Heidelberg, Germany: Springer.
5. E W. 2011 *Principles of multiscale modeling*. Cambridge, UK: Cambridge University Press.
6. Stensrud DJ. 2007 *Parameterization schemes: keys to understanding numerical weather prediction models*. Cambridge, UK: Cambridge University Press.
7. Silverman BW. 1986 *Density estimation for statistics and data analysis*. London, UK: Chapman and Hall.
8. Jolliffe IT, Stephenson DB (eds). 2012 *Forecast verification: a practitioner's guide in atmospheric science*, 2nd edn. New York, NY: Wiley.
9. Maynard T. 2016 Extreme insurance and the dynamics of risk. PhD thesis, LSE, London, UK. See http://etheses.lse.ac.uk/3297/.
10. Kaipio JP, Somersalo E. 2004 *Statistical and computational inverse problems*. Berlin, Germany: Springer.
11. Law K, Stuart A, Zygalakis K. 2015 *Data assimilation: a mathematical introduction*. Berlin, Germany: Springer.
12. Reich S, Cotter C. 2015 *Probabilistic forecasting and Bayesian data assimilation*. Cambridge, UK: Cambridge University Press.
13. O'Hagan A, Forster J. 2004 *Kendall's advanced theory of statistics*. Vol. 2B. *Bayesian inference*, 2nd edn.Paris, France: Arnold Publishers.
14. Kahneman D. 2011 *Thinking, fast and slow*. Harmondsworth, UK: Penguin Books.
15. Vos Savant M. 1996 *The power of logical thinking*. New York, NY: St Martin's Press.

[8]See the Mathematics in Industry website (http://www.maths-in-industry.org/future/).

16. Brooks S, Gelman A, Jones GL, Meng X-L. 2011 *Handbook of Markov chain Monte Carlo*. London, UK: Chapman & Hall/CRC Press.
17. Nocedal J, Wright SJ. 2006 *Numerical optimization*, 2nd edn. Berlin, Germany: Springer.
18. Oliver DS, Reynolds AC, Liu N. 2008 *Inverse theory for petroleum reservoir characterization and history matching*. Cambridge, UK: Cambridge University Press.
19. Bui-Thanh T, Ghattas O, Martin J, Stadler G. 2013 A computational framework for infinite-dimensional Bayesian inverse problems. Part 1. The linearized case, with application to global seismic inversion. *SIAM J. Sci. Comput.* **35**, A2494–A2523. (doi:10.1137/12089586X)
20. Lim SWX. 2017 Bayesian inverse problems and seismic inversion. DPhil Thesis, University of Oxford, Oxford, UK. See https://ora.ox.ac.uk/.
21. Evensen G. 2009 *Data assimilation: the ensemble Kalman filter*, 2nd edn. Berlin, Germany: Springer.
22. Stordal AS, Karlsen HA, Nævdal G, Skaug HJ, Vallès B. 2011 Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter. *Comput. Geosci.* **15**, 293–305. (doi:10.1007/s10596-010-9207-1)
23. Chan Y, Oliver DS. 2013 Levenberg-Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Comput. Geosci.* **17**, 689–703. (doi:10.1007/s10596-013-9351-5)
24. Trémolet Y. 2006 Accounting for an imperfect model in 4D–Var. *Q. J. R. Meteorol. Soc.* **132**, 2483–2504. (doi:10.1256/qj.05.224)
25. Powell WB. 2015 The jungle of stochastic optimization. *SIAM News* **48**. See https://sinews.siam.org/Details-Page/the-jungle-of-stochastic-optimization.
26. Whittle P. 1996 *Optimal control: basics and beyond*. New York, NY: John Wiley and Sons Ltd.
27. Pratt JW, Raiffa H, Schlaifer R. 1995 *Introduction to statistical decision theory*. Cambridge, MA: MIT Press.
28. Keeney RL, Raiffa H. 1993 *Decisions with multiple objectives: preferences and value tradeoffs*. Cambridge, UK: Cambridge University Press.
29. Traeger CP. 2009 Recent developments in the intertemporal modeling of uncertainty. *Annu. Rev. Resour. Econ.* **1**, 261–85. (doi:10.1146/annurev.resource.050708.144242)
30. Lindley DV. 1985 *Making decisions*, 2nd edn. New York, NY: Wiley.
31. Bertsekas DP. 2017 *Dynamic programming and optimal control*, vol. I, 4th edn. Belmont, MA: Athena Scientific.
32. Powell WB. 2011 *Approximate dynamic programming: solving the curses of dimensionality*, 2nd edn. New York, NY: Wiley.
33. Jones DR. 2001 A taxonomy of global optimization methods based on response surfaces. *J. Glob. Optim.* **21**, 345–383. (doi:10.1023/A:1012771025575)
34. Fowkes J. 2011 Bayesian numerical analysis: global optimization and other applications. DPhil Thesis, University of Oxford, Oxford, UK. See https://ora.ox.ac.uk/objects/uuid:ab268fe7-f757-459e-b1fe-a4a9083c1cba.
35. Bharadwaj UR, Silberschmidt VV, Wintle JB. 2012 A risk based approach to asset integrity management. *J. Qual. Maint. Eng.* **18**, 417–431. (doi:10.1108/13552511211281570)
36. Martin J, Wilcox LC, Burstedde C, Ghattas O. 2012 A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM J. Sci. Comput.* **34**, A1460–A1487. (doi:10.1137/110845598)
37. Luthi SM. 2001 *Geological well logs: their use in reservoir modeling*. Berlin, Germany: Springer.
38. Barros EGD, Van den Hof PMJ, Jansen JD. 2016 Value of information in closed-loop reservoir management. *Comput. Geosci.* **20**, 737–749. (doi:10.1007/s10596-015-9509-4)
39. Richardson LF. 1922 *Weather prediction by numerical process*, 1st edn (2007, 2nd edn.). Cambridge, UK: Cambridge University Press.
40. Lynch P. 2008 The origins of computer weather prediction and climate modeling. *J. Comput. Phys.* **227**, 3431–3444. (doi:10.1016/j.jcp.2007.02.034)
41. Rodgers CD. 2000 *Inverse methods for atmospheric sounding: theory and practice*. Singapore: World Scientific.
42. Dando ML, Thorpe AJ, Eyre JR. 2007 The optimal density of atmospheric sounder observations in the Met Office NWP system. *Q. J. R. Meteorol. Soc.* **133**, 1933–1943. (doi:10.1002/qj.175)
43. Eyre JR, Weston PP. 2014 The impact of the temporal spacing of observations on analysis errors in an idealised data assimilation system. *Q. J. R. Meteorol. Soc.* **140**, 1441–1452. (doi:10.1002/qj.2227)

44. Stafford JV (ed.). 2013 *Precision agriculture '13: Papers presented at the 9th Eur. Conf. on Precision Agriculture, Lleida, Catalonia, Spain*. Wageningen, The Netherlands: Academic Publishers.
45. Institute for Mathematical Innovation. 2017 *Agri-Food Mathematical Science Study Group with Industry, Bath, UK, 16–18 January 2017*. Bath, UK: University of Bath. See http://www.bath.ac.uk/imi/events/Agri-Food-Study-Group.html.
46. Talluri KT, van Ryzin GJ. 2004 *The theory and practice of revenue management*. Boston, MA: Kluwer Academic Publishers.
47. Ramakrishnan R. 2012 Markdown management. In *The Oxford handbook of pricing management* (eds Ö Özer, R Phillips), ch. 25. Oxford, UK: Oxford University Press.
48. International Centre for Mathematical Sciences. 2015 *The role of inverse problems and optimisation in uncertainty quantification, Edinburgh, UK, 17–18 June 2015*. See http://www.icms.org.uk/workshops/UQ2015.