Prediction for Decision-Making under Uncertainty

J. P. Norton

Integrated Catchment Assessment & Management, The Australian National University, Canberra, Australia.

School of Engineering (EECE), The University of Birmingham, UK

Abstract: A primary use for mathematical models, in fields such as environmental management, economics and engineering, is prediction. Prediction can aid choice between decisions by assessing their consequences, or between models by comparing their prediction performance. Choice between models raises fewer questions of how to use predictions, and this paper concentrates instead on prediction for decision-making. It starts with an illustration of how systematic but easily overlooked modelling error can spoil prediction. Three radically differing approaches to supporting decision-making with imprecise predictions are then reviewed: Bayesian optimal decision theory, model predictive control and set-membership prediction. Their potential and limitations as aids to environmental decision-making are discussed.

Keywords: *Mathematical models; prediction; decision-making; Bayes estimation; model predictive control; set-membership*

1. INTRODUCTION

Prediction is only one of the possible objectives of mathematical modelling in fields such as environmental management, economics and engineering. A model can be viewed, at the simplest, as just a concise summary of knowledge about a system. Conciseness is valuable as (aside from statistical considerations) it forces selectivity and may expose behaviour buried in the raw data. Construction of a concise model can also reveal deficiencies in data and limits to knowledge. Much may thus be learnt from a model even if it never gets as far as prediction.

Prediction is central, however, in at least two areas. In scientific hypothesis testing, predictions are compared with observation but the prediction model is usually a law or principle, fixed in form and so highly reduced as to leave little room for doubt about how it should be used. The picture is quite different in "what if?" experiments on elaborate simulation models, carried out, for instance, to guide management of environmental systems. Here questions arise about adequacy and economy of model form, sufficient exploration of the state and inputs of the model and the system, reliability and uncertainty of the results, and computing load. In this context, prediction has the ultimate aim of aiding choice, either between management decisions by assessing their consequences or between models by comparing their prediction performance. In choosing between models, the quality of prediction is measured retrospectively. By contrast, decision-making requires that the quality of predictions be assessed in advance or that the outcomes be made insensitive to the prediction uncertainty by monitoring the system and revising decisions as necessary (as in Section 4). Prediction for decision-making is thus more demanding than prediction for model selection.

Section 2 presents a cautionary example of how prediction performance can be ruined by an easily overlooked feature of system behaviour, even in the absence of unpredictable disturbances or system changes. The next three sections offer three very different ways to tackle prediction under uncertainty. Section 3 looks at the much discussed but little used tool of Bayes optimal decision theory. It has widely recognised limitations for practical decision-making but provides a framework in which to examine the basic operations of prediction and optimisation under uncertainty. Another reason for re-examining it is that increased computing capability is making these operations much easier. In sharp contrast to this theory-driven approach, Section 4 summarises model predictive control, a heuristic scheme for sequential decision-making under uncertainty which grew up in process control and may have lessons for other areas of application. By exploiting feedback, it achieves robust performance without asking for uncertainty to be characterised. It can also incorporate constraints on control action and system variables. A third way to attack prediction is outlined in Section 5: a very new set-membership approach to prediction in non-linear dynamical systems, which assumes very little about the form of the model. It is hoped that these widely

differing techniques will suggest ways to make predictive decision-making more systematic.

2. PREDICTION EXAMPLE

2.1. System simulated

The example will show how systematic error in a prediction model is easily overlooked even in favourable conditions, and how common diagnostics differ greatly in ability to detect it.

In the example, alternative models of the discrete-time dynamics of a mildly non-linear system are identified from simulated input-output records. They then generate output predictions, given the future input; input prediction is not considered. The input sequence $\{u\}$ is generated from a white sequence $\{w\}$, uniformly distributed over (-0.25, 0.75), by filtering by

$$u_t = 0.95u_{t-1} + w_t \qquad (1)$$

where the subscripts denote time. For simplicity, the system is single-input, single-output and deviates from linear, time-invariant, first-order dynamics only by one non-linear term. The input is related to the noise-free output sequence $\{y^o\}$ by

$$y^o_t = u_{t-1} + 0.925y^o_{t-1} + 0.1u_{t-1}y^o_{t-1} \qquad (2)$$

Zero-mean, white, u.d. noise is added to $\{y^o\}$. The sample mean-square signal-to-noise ratio is 100 so that noise does not obscure the effects of systematic modelling error; the comparisons below change little for SNR down to 10 or less.

2.2. Model identification

Records (u_t, y_t), t = 1,2,...,400 are generated (Figure 1), having discarded 100 pairs to remove the influence of the (zero) initial conditions.

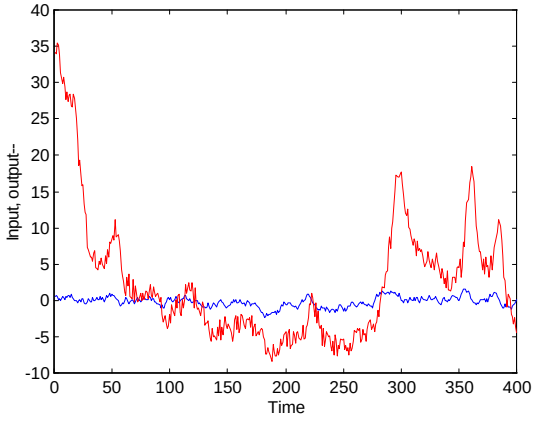

Figure 1. Input-output records for example: full line input, broken line output

The two linear-in-parameters models

$$\begin{aligned} &(1): \; y_t = bu_{t-1} + ay_{t-1} + cv_{t-1} + v_t \\ &(2): \; y_t = bu_{t-1} + ay_{t-1} + pu_{t-1}y_{t-1} + cv_{t-1} + v_t \end{aligned} \qquad (3)$$

are identified from the first 300 input-output pairs by extended least squares (which can be generalized for estimation of time-varying parameters: see later). The structured noise model $cv_{t-1} + v_t$, with $\{v\}$ white and zero-mean, covers structure introduced by a and pu_{t-1} as the model is rewritten from output-error to equation-error form. Estimates $\hat{a}, \hat{b}$ or $\hat{a}, \hat{b}, \hat{p}$ give one-step output predictions $\{\hat{y}\}$ from

$$\begin{aligned} &(1): \; \hat{y}_t = \hat{b}u_{t-1} + \hat{a}\hat{y}_{t-1} \\ &(2): \; \hat{y}_t = \hat{b}u_{t-1} + \hat{a}\hat{y}_{t-1} + \hat{p}u_{t-1}\hat{y}_{t-1} \end{aligned} \qquad (4)$$

for t=301,..., 400. Figure 2 shows the large difference in prediction performance between linear model (1) and non-linear model (2).

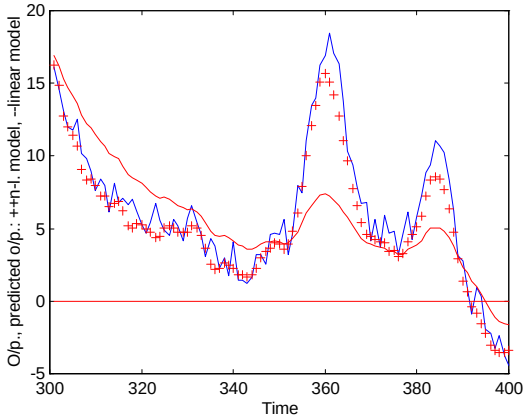

Figure 2. One-step predictions of output: full line output, --- model (1), +++ model (2)

The smooth non-linearity $pu_{t-1}y_{t-1}$ increases the positive peaks. Model (1), although identified from records containing two large peaks and several smaller ones, tracks the peaks near times 360 and 385 poorly and suffers bias elsewhere. Model (2) predicts well.

The poor performance of model (1) is not surprising and model structure (2) would be discovered by a thorough investigation (at this high SNR), but the non-linear term has small enough effects on some common diagnostics available at time 300 to risk being overlooked if only those diagnostics are used. For instance, the sample root-mean-square residual is less than 17% higher for model (1) than for model (2); the

mean-square residuals are only 1.65% and 1.21% of the output variance respectively; and the estimated s.d.'s of $\hat{a}, \hat{b}, \hat{c}$, 0.0083, 0.1011, 0.0595 for model (1) and 0.0089, 0.1057, 0.0598 for model (2), fail to detect any difference. The residual autocorrelation functions, Figure 3, are more informative but would be less so at lower SNR.

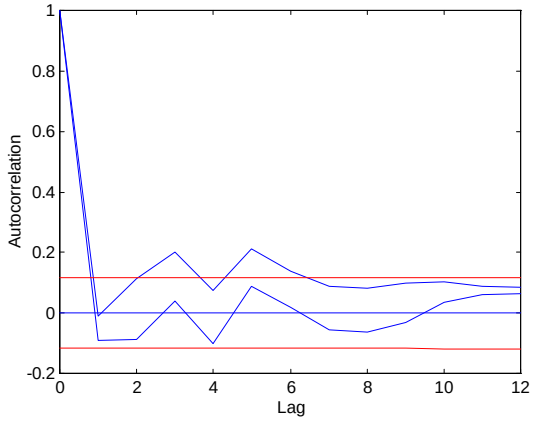

Figure 3. Sample autocorrelations of residual sequence: broken line model (1), full line model (2). Two-standard-deviation lines also shown.

There are better detectors of systematic error. One is simulation-mode one-step prediction error over the whole prior record, Figure 4. It uses predicted output on the right-hand side of (3) rather than the observed value used during parameter estimation.

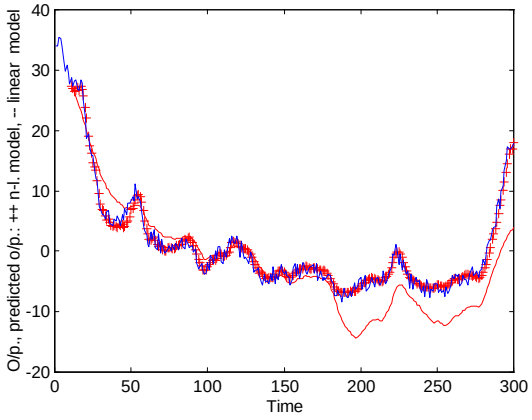

Figure 4. Simulation-mode one-step predictions: full line output, --- model (1), +++ model (2). The predictions from model (2) are virtually indistinguishable from the actual output.

The prediction error in Figure 4 clearly reveals the weakness of model (1) and its nature gives some idea of what is wrong with the model. The trouble can also be revealed by modelling one or more parameters as time-varying, *e.g.* as random walks (Norton and Chanat, 2003). A parameter-estimation version of optimal smoothing (Norton, 1975) yields parameter estimates which, at every point in time, utilise the information from later as well as earlier observations. Figure 5 compares $\hat{b}$, with b modelled as a random walk with r.m.s. increment 0.224, with the actual "coefficient" $b + py_{t-1}$ of u_{t-1}.

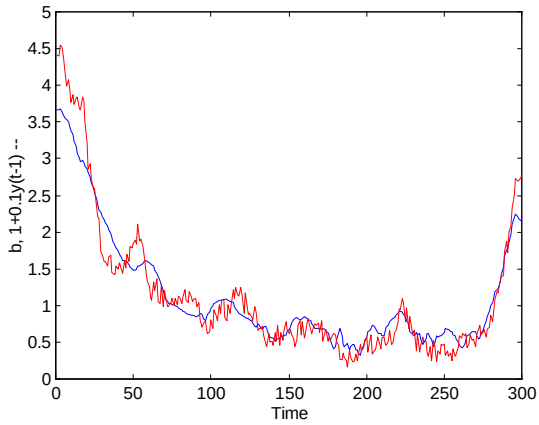

Figure 5. Effective coefficient of u_{t-1} (broken line) and time-varying estimate $\hat{b}$ (full line)

Viewed as an output-dependent gain applied to the input, $\hat{b}$ gives a clear picture of the nature and effect of the non-linearity. Alternatively, a time-varying estimate $\hat{a}$ of the coefficient of y_{t-1} correlates fairly well with $a - pu_{t-1}$, but the more rapid variations of the input are followed less well by $\hat{a}$.

To summarise, omission of a well behaved and quite small non-linearity may render a predictor useless yet affect common diagnostics of model performance so little as to escape detection. Better-chosen diagnostics, such as simulation-mode prediction error over the whole previous record or estimates of selected parameters represented as time-varying, are much more effective.

3. OPTIMAL PREDICTION-BASED DECISION-MAKING?

3.1. Motivation

Decision-making based on uncertain prediction, and according to a single criterion, can be posed as a mathematical optimisation problem (Raiffa, 1968). A solution amounts to automated decision-making, long recognised as impracticable for complicated systems (Holling, 1978, p.119 and refs.). An attempt to pose and solve a decision-optimisation problem may be frustrated by irreducibility of the decision objective to a simple criterion, inability to provide quantified uncertainties for a predictive model and the

measurements feeding it, or failure to gain stakeholders' trust. Even so, the attempt may yield insight into what model predictions can do for decision-making, and what would make decision-making more objective and its risks clearer. In particular, a mathematical formulation of the problem clarifies the roles of prior information (including uncertainty specification), measurements and model. A more immediate reason for another look at decision theory is the increasing computational feasibility of implementing its fundamental operations of Bayesian uncertainty-updating and associated optimisation under uncertainty, by Monte Carlo methods.

It is assumed that prediction for decision-making supplies uncertain values of either the outcomes of alternative decisions or, less directly, variables affecting the outcomes. A scalar loss function measures the outcomes. The optimal decision minimises the risk, defined as expected loss, averaged over the prediction uncertainty. Prediction and loss evaluation require running of a model. In the simplest cases this might be an analysable state-space model but in environmental applications it will usually be a digital simulation.

3.2. Description of uncertainty; pdf tails and bounds

A framework for considering uncertainty in the model and its predictions must be chosen. A Bayes probabilistic framework is reasonably broad yet can be specialised. In it, the uncertain predictor parameters are characterised by their joint probability density function (pdf). All other uncertain items influencing the prediction (*e.g.* unmeasured disturbances) are also described by their pdfs. It may seem unrealistic to assume that a pdf is available for every uncertain item affecting the prediction, as prior experience or scientific knowledge of some items may be very limited. However, a Bayes framework can accommodate some types of very limited knowledge; in particular, if only the range of an item is known, it defines the support but not the shape of the pdf. If all contributing items have known, finite support, Bayes estimation yields bounds on the prediction. Bounds on predicted outcomes may be enough to rank costs and permit a decision. This is so, for instance, if the cost is monotonic over disjoint ranges of each outcome. Bounds also match worst-case cost criteria, as discussed in Section 5.

When little is known of the pdf of an uncertain quantity, reluctance to assume a finite range, on grounds of ignorance of extreme behaviour, would be understandable. The choice is between uncertain estimates of pdf tails (or cumulative probabilities of being in a tail) and an imprecise range. Unconvincing assumptions about pdf tail shapes can be avoided by adopting a tentative range, encompassing all previous experience, and accepting the unknown probability that a sample will fall outside it. The assumed range may be revised if its consequences turn out to be implausible. It is easier to assess the credibility of this sort of uncertain quantile, by reference to the observed extremes in the limited historical record, than that of an assumption about the shape of unbounded pdf tails. This is particularly so when extremes are predominantly due to few enough factors for an appeal to the Central Limit Theorem to be unwarranted.

Everday, informal human decision-making tends to binary categorisation ("likely" and "unlikely"), treating an extreme event as ignorable until its perceived risk (combining likelihood of occurrence and seriousness of consequences) becomes high enough. It is also notable that assumed ranges of uncertain quantities, with the risks of exceeding them often unknown but deemed acceptable, are used very effectively in much of engineering design, in preference to assuming pdf tail shapes. Consideration of the probabilities of extreme behaviour is avoided by executing toleranced, worst-case design so that the system will not be close to the specified performance boundaries for any combination of component parameter values within specified ranges. With proper monitoring of manufacturing tolerances and checking of subsystems ("marginal testing" in electronic design, testing to failure in mechanical engineering), this process can produce highly reliable systems. Environmental decision-makers have much less control and much poorer information, but can benefit from knowing, at least roughly, how the extremes of predicted behaviour, over given ranges of the contributing influences and with unknown but small risks of exceeding them, compare with the boundaries of acceptable outcomes. That is, judgements and actions may be facilitated by bounded-uncertainty predictions.

There are other reasons for not worrying too much, in Bayesian decision-making, about where to truncate a pdf when the probability thus excluded is unknown or only upper-bounded,. First, the optimal decision nominally minimises average loss, weighting extreme outcomes with their small probability densities and large losses. In practice, a decision-maker simply wants to avoid any behaviour beyond some fairly well defined boundaries. The transition from acceptable to unacceptable is often quite sharp, so an optimisation over the truncated, "likely" range,

together with separate exploration of what range of circumstances will keep the outcome acceptable under each decision, seems preferable to a once-and-for-all optimisation which integrates over regions where the probability density is low but uncertain and the cost high but fairly arbitrary. Second, the ensemble of outcomes may itself be very uncertain for various reasons: the system has no close comparators, so the ensemble of its behaviour is guessed from its limited observed past; the predictions extrapolate into regions of state space where the model is untested; the model is wrong in some respect which has not *yet* clashed with observation; or the forcing of the system is inhomogeneous in time or space, with isolated large shocks. Extremes not observed so far still cannot be excluded. It is a scientific truism that model "validation" is only failure to falsify it. Similarly, a prediction-based decision cannot be justified, but only not demonstrably unjustified.

3.3. Prediction in Bayes framework

Returning to Bayesian prediction-guided decision-making, assume that by one means or another, prior pdfs, perhaps truncated, have been arrived at for all items (state variables, model parameters and unknown forcing) influencing the future value of a scalar system output. At time t prior knowledge and measurements X_t have yielded a pdf $p(q_t \mid X_t)$ for the state and parameters q_t of a predictor model

$$y_{t+k} = h_{tk}(q_t) \qquad (5)$$

for the scalar output y_{t+k}. Analytical derivation of the density estimate (Tay and Wallis, 2000) $p(y_{t+k} \mid X_t) \equiv p(h_{tk}(q_t) \mid X_t)$ is generally not feasible even if the underlying state-space model is linear in state and parameters and $p(q_t \mid X_t)$ is Gaussian, for two reasons: products of state variables and parameters appear, and recursion of a linear state equation over k time steps to give a k-step prediction produces terms of total degree k in the parameters. Linearity in the parameters can be retained by parameterising and identifying the linear predictor model directly rather than through a state-space model, but easy interpretation of the parameters (*e.g.* as rate constants) is thereby lost. Also, complicated correlation, not necessarily well described by a covariance, then exists among the k-step predictor parameters.

As an alternative to analytical derivation, Monte Carlo numerical approximation of $p(y_{t+k} \mid X_t)$ has the attraction that $h_{tk}(q_t)$ need not be written explicitly. The model is run once for each of a large set of samples from $p(q_t \mid X_t)$, producing a sample set from $p(h_{tk}(q_t) \mid X_t)$ large enough for the local density of samples to approximate the probability density adequately. Unknown forcing with known distribution can be included by applying a sample of the forcing to each sample. Time-structured forcing such as rainfall would require an auxiliary model, but the forcing need not be additive and the technique does not rely on linearity or on any special properties of the pdfs.

Pdf prediction by propagating samples is well matched to Monte Carlo Bayes estimation of $p(q_t \mid X_t)$. For i=1,2,...,t, a large set of samples of q_i is alternately time-updated, using the model to propagate them from time i-1 to time i, and observation-updated, using measurements at time i to evaluate the observation likelihood at time i for each sample. A sample q_{i-1} from $p(q_{i-1} \mid X_{i-1})$ is time-updated by substitution into the state equation or, if the model is complicated, executing a simulation run over one time step. The time intervals need not be uniform; the final time update is the prediction from t to $t+k$. Such Monte Carlo Bayesian state-estimation schemes have received a great deal of attention in the past decade (Doucet, de Freitas and Gordon, 2001), because of their much greater flexibility than classical state estimators in applications such as target-tracking (Gordon, Salmond and Smith, 1992), where non-linearity, poor observability and ambiguity may make Kalman-filter trackers cumbersome and unreliable. Their ability to handle any given pdfs for initial state, forcing and observation error is also valuable. Propagation of the state pdf rather than the mean and covariance allows asymmetry and ambiguity (multiple maxima) to be registered. Derivation of marginal densities and quantiles is straightforward.

These state estimators, regrettably named "particle filters", incorporate the observation likelihoods through Bayes' rule, multiplying the equal probability masses of the time-updated "particles" by the likelihoods, then resampling according to the likelihoods, *i.e.* importance resampling (Smith and Gelfand, 1992), to obtain a new sample set for the next time update. The danger is that when new observations are processed, the great majority of samples have low associated likelihoods. They then have low probability of appearing in the resampled set, which may soon collapse to a single sample. This

is particularly a danger when new observations indicate a change from what was predicted, precisely when it is crucial that the pdf should track the new conditions. Much effort has gone into ways of avoiding collapse, for instance by introducing jitter, using stratified sampling or generating new samples when inadequate coverage is detected (Carpenter, Clifford and Fearnhead, 1999; Veres and Norton, 2001).

3.4. Minimum-risk decision

Once found, $p(y_{t+k} | X_t)$ forms the basis for an action $a_t(p(y_{t+k} | X_t))$ which together with output y_{t+k} determines the loss *via* a loss function $L(a_t, y_{t+k})$. Selection of a loss function matching the decision-maker's concerns may be difficult or impossible (Smith, 1988), but let us assume that at least some preferences about the outcome can be encapsulated in one. The loss function generally depends directly on a_t, since actions incur immediate costs, as well as indirectly on a_t through its influence on y_{t+k}. The decision-maker aims to choose, from a set A of possible actions, the one $\hat{a}_t^*$ which minimises the expected loss (cost minus benefit) over all possible outcomes, *i.e.* the risk

$$\hat{R}(a_t | X_t) \circ E[L(a_t, y_{t+k}) | X_t] = \int_{-\infty}^{\infty} L(a_t, y_{t+k}) p(y_{t+k} | X_t) dy_{t+k}. \quad (6)$$

Use of samples to represent the pdf turns the integration into straightforward summation. To find the best action

$$\hat{a}_t^*(X_t) = \arg\min_{a \in \mathsf{A}} \hat{R}(a_t | X_t) \quad (7)$$

the risk must be evaluated at a number of possible values of a_t. If A consists of a few (perhaps only two) discrete alternatives, the minimisation is trivial. If on the other hand a continuous range of actions is available, numerical search for the optimum is required. For broad classes of loss functions and pdf shapes, the optimum is at a readily found point such as the conditional mean $E[y_{t+k} | X_t]$, as noted long ago in optimal estimation theory (Deutsch, 1965).

3.5. What computation is required?

The relative simplicity of the expression for $\hat{R}(a_t | X_t)$ conceals a variety of possible situations and corresponding computational loads in environmental applications (which do not necessarily arise in other areas, such as economics and finance, where minimum-risk decision theory has been considered). Let us look for especially easy cases. One is where the uncertain y_{t+k} does not depend on a_t but does affect the cost and/or benefit of the outcome. For example, y_{t+k} might be rainfall in the period *t* to *t+k*, affecting water availability beyond the otherwise known consequences of an irrigation water allocation a_t. Here there are two separate models, probabilistic for y_{t+k} and deterministic for the outcome of a_t, so evaluation of $L(a_t, y_{t+k})$ requires a single Bayesian pdf prediction and as many deterministic predictions as possible actions. Optimisation of risk $\hat{R}(a_t | X_t)$ does not involve repetition of the pdf prediction. Another example is where y_{t+k} is a list of future prices, affected by market uncertainties, applied to the crop (treated as certain) resulting from a given land-use choice a_t, for stated climatic conditions. However, this relative simplicity is easily lost. In the first example, irrigation water allocation probably depends in fact on flow and hence rainfall, by policy. In the second, the crop may affect the price distribution. If so, $p(y_{t+k} | X_t)$ becomes $p(y_{t+k} | X_t, a_t)$ and there are as many pdf predictions to carry out as candidate actions.

Another situation where only a single pdf prediction seems to be needed is where a_t is an advance commitment from a resource (say investment or irrigation water) to satisfy a future demand with a known payback, and y_{t+k} is the aggregate of uncertain competing demands on the same resource. The loss function $L(a_t, y_{t+k})$ is simply the sum of the (sign-inverted) certain benefit $z_{t+k}(a_t)$ of satisfying the first demand and the uncertain benefit of satisfying the others, subject to a_t. The catch is that in practice the benefit from the forward commitment may also have to be predicted under uncertainty. It would be predicted as a pdf $p(z_{t+k} | a_t, X_t, W_{t+k})$ where W_{t+k} is the set of uncertain conditions determining the benefit of commitment a_t of resource for use from time *t+k*. There are then

two types of pdf prediction to make, a single one for $p(y_{t+k} \mid X_t)$ and as many as possible actions for $p(z_{t+k} \mid a_t, X_t, W_{t+k})$. We have to integrate over both to find the risk $\hat{R}(a_t \mid X_t)$.

The next complication arises from treating $p(y_{t+k} \mid X_t)$ as if it were correct in finding the optimal action a_t^*. The risk $\hat{R}(a_t \mid X_t)$ found by (6) has a circumflex as a reminder that it only approximates the true risk

$$R(a_t) \circ E[L(a_t, y_{t+k})] = \int_{-\infty}^{\infty} L(a_t, y_{t+k}) f(y_{t+k}) dy_{t+k}. \quad (8)$$

which would result from the actual pdf $f(y_{t+k})$. The actual pdf would lead to an optimal action

$$a_t^o = \arg\min_{a \in A} R(a_t) \neq a_t^*(X_t) \quad (9)$$

Proper design of a robust, Bayesian, prediction-based, decision-optimising procedure would examine the sensitivity of the predicted risk function $\hat{R}(a_t \mid X_t)$ to error in the pdf $p(y_{t+k} \mid X_t)$. The procedure is robust so long as the information X_t determines $p(y_{t+k} \mid X_t)$ well enough for (7) to yield the same "optimal" action (within acceptable bounds, if the choice is not discrete) as would be obtained with perfect knowledge at time *t* of $f(y_{t+k})$.

It seems that optimal decision theory may involve heavy computing even if the decision problem can be suitably formalised and advantage is taken of modern Monte Carlo techniques for propagating uncertainty. The next section describes a more heuristic approach to predictive decision-making. It was developed for automatic control in the process industries, which also has to contend with poor models and uncertainty about future disturbances and ambient conditions.

4. SEQUENTIAL DECISION-MAKING WITH UNANALYSED UNCERTAINTY

4.1. Predictive decision-making with feedback

The Bayesian scheme outlined above takes uncertainty explicitly into account but depends on a detailed uncertainty prescription and on Monte Carlo implementation of state and parameter estimation to obtain $p(q_t \mid X_t)$, prediction of $p(y_{t+k} \mid X_t)$ and integration of probability-weighted loss over y_{t+k}. For once-and-for-all decisions and long prediction intervals, the effects of uncertainty should clearly be analysed as fully as possible. However, if an initial decision can be followed by periodic corrections as the consequences unfold, the problem becomes one of feedback control, with much reduced need for uncertainty analysis. A crucial property of feedback is that, with proper design, it can reduce the sensitivity of overall system behaviour to variations or errors in the system response and to disturbances. For linear systems, it is easy to show that if the controller gain can be increased enough to make the loop gain high, there are nice consequences for robustness: overall (closed-loop) gain from input disturbances to output becomes insensitive to the input-output open-loop (forward-path, compensator plus plant) gain, ultimately tending to 1/*feedback gain*; and the effects of output disturbances are reduced by a factor equal to the loop gain.

The snag is that most linear systems become unstable as the feeedback gain is increased. Well before the onset of instability, the response to disturbances is poorly damped. The beneficial effects of feedback can be retained if the loop gain is high over the frequency range containing most of the disturbance power, yet low enough over the frequency range giving greatest proximity to instability; this is a primary aim of control system design. For non-linear systems, stability is generally no longer determined entirely by the system parameters but depends on the input and state. However, successful application of feedback around a non-linear system may reduce the effect of non-linearity by reducing the variation of the state variables.

The prospect of obtaining the benefits of feedback in non-engineering applications of predictive models justifies a look at an automatic control technique which might contribute ideas for sequential decision-making.

4. 2. Model predictive control

Model predictive control (MPC) (Garcia *et al.*, 1989; Morari and Zafiriou, 1989; Soeterboek, 1992; Maciejowski, 2002) is a flexible approach to robust control in the face of modelling errors and unpredictable disturbances. It has been developed largely in industry and is appealing for its heuristic nature and proven effectiveness in difficult cases. It originated in the late 1970's in the process industries, where the processes are typically hard to model, having distributed, non-linear and stiff dynamics, and the control system must cope with rapidly varying input conditions. Moreover, control actions and state variables are heavily constrained by physical limits (*e.g.* valve

opening and closing rates, pump speed, heater power), safety considerations (*e.g.* boiler level, reactor temperature and pressure) and product quality. Analytical control design methods are unable to deal with this combination of difficulties, so the basis of MPC is *constrained numerical optimisation using a predictive model* of the plant. At regular time intervals, a sequence of future control actions is optimised deterministically, to match as closely as possible a desired output trajectory. For example, the integral of the squared error between the controlled output and desired trajectory might be minimized, over an interval from the minimum cost horizon up to the prediction horizon. Most MPC schemes use linear models and quadratic optimisation criteria, so that the optimisation can be manipulated into a standard form for which efficient solvers exist: LP, QP, quadratically constrained QP or linear matrix inequalities (Boyd *et al*., 1994).

The repetitive on-line use of a model allows adaptation to disturbances, measured input changes, evolving control priorities and changes in the system itself. The model may be updated, so that the overall scheme combines model identification and control synthesis. The combination of model and controller which jointly optimizes a stochastic control performance criterion, the dual controller, can be synthesised only in very simple special cases, so instead the scheme uses the generally suboptimal combination of the "best" model according to a fitting criterion and "optimal" control obtained by treating the model as true.

The second important feature of MPC is *receding-horizon control*, in which the optimal control sequence is recomputed periodically, at intervals shorter than that covered by the optimisation. In other words, only the initial part of each optimal sequence is actually applied. The reason is to maintain flexibility in the face of unforeseen changes, rather than relying on a complete precomputed control sequence which may soon be out of date. At every recomputation, the most recent measurement of the output provides some feedback.

The prediction horizon is a compromise between speed of computation, needing a short horizon, and robustness (low sensitivity of stability and performance to variations in system and environment) (Bemporad and Morari, 1999), which increases as the horizon is lengthened. Indeed, stability can be ensured, in principle, by meeting a terminal constraint on state at the end of an infinite prediction horizon (Maciejowski, 2002). It turns out that speed can be improved at low cost in performance by constraining the control sequence to be constant beyond some point before the prediction horizon, optimising the sequence only up to this control horizon. It is intuitively obvious that the shorter the control horizon, the greater the control effort up to it. Taking strong control action over the early part of the sequence to approach the desired trajectory leaves later control (in updated control sequences) more scope for dealing with later disturbances or changes in the system or the control objective.

What, if anything, does the MPC approach tell us that has implications for decision-making in other fields? Broadly,

(1) feedback can mitigate the need for uncertainty to be considered, by making the controlled system less sensitive to disturbances and model error;

(2) receding-horizon control, revising the control at intervals much shorter than the prediction horizon, takes a long-term view yet incorporates feedback;

(3) designing the control policy as if it were to cover an interval well beyond the next revision makes the control more circumspect;

(4) pretending that the control action will be constant after some time well short of the prediction horizon increases the short-term action, which if successful brings the system closer to the desired condition and leaves more scope for later action to meet unforeseen changes.

4.3. Implications for sequential decision-making?

There are obvious limits to how far techniques from control engineering can provide helpful analogies for other fields where regulation or "servo control" to achieve a desired change is required, such as environmental management. The existence of stakeholder groups, regulatory frameworks, political agendas, budgetary constraints, conflicting interests and much larger spatial and time scales will prevent most such problems from being formulated or tackled as tidily as most control-engineering problems. The decision-maker may well have severely restricted ability to act on the system. In particular, regular monitoring and revision of the control actions may be too expensive or unreliable, or impracticable for other reasons. Resistance to trying a scheme developed in other circumstances would be natural, although simulation studies may be able to provide some confidence in the applicability of MPC-style sequential control in new areas. Other potential difficulties, mostly faced by any decision-making strategy, include formulation of a performance criterion and constraints suitable for numerical optimisation, especially when the outcome has a spatial

dimension; unintended, unmodelled and unacceptable consequences of fierce control action due to high feedback gain; non-stationarity (*e.g.* seasonal or other climate-driven behaviour) which must be taken into account in the control optimisation; invalidation of the design rules which have grown up for MPC by non-linearities which show up as severe deviations from the model; incomplete state observability and/or reachability resulting in unacceptable behaviour of internal variables; delay between output and control action hindering or preventing stabilisation of the system, let alone good control; practical difficulties in maintaining control consistently at suitable update intervals and over a long enough period; and moving targets as social, environmental and economic priorities change.

On the other hand, some of the features listed above are already present in informal, common-sense management decision-making. They have made MPC commercially successful in a variety of applications, now beginning to extend into aerospace. The ever-increasing availability of numerical models which can be run in many trials of alternative decisions invites more systematic, sequential optimisation and control in non-engineering problems with features (such as non-linearity, distributedness, poor modelling, time-varying ambient conditions) addressed by MPC.

5. SET-MEMBERSHIP PREDICTION

5.1. Bounded-norm setting for prediction

Up to this point prediction has been model-based and thus dependent on identification of a suitable model structure and parameter values. If prediction is the sole motive for modelling, an alternative exists: set-membership prediction (Novara and Milanese, 2001b; Milanese and Novara, 2002). It is of interest for two reasons: it does not rely on choice of a specific model structure, sidestepping the identification problem, and it formulates the prediction problem in a bounded-error, worst-case-optimal context quite different from the probabilistic formulation in Section 3 or the deterministic, feedback-oriented approach of Section 4.

The problem is prediction at time t of scalar output y_{t+k} from noise-affected measurements

$$y_i^o = y_i + v_i,\ f_i^o = f_i + \mathbf{w}_i,\ i = 1,2,...,t \qquad (10)$$

of the output and explanatory variable vector $f \in \Re^n$ (*e.g.* earlier input and output samples). The noise is known only to be bounded:

$$|v_i| \le d,\ \|\mathbf{w}_i\|_2 \le e,\ \ i = 1,2,...,t \qquad (11)$$

and the relation $y_{t+k} = f(f_t)$ is assumed only to be differentiable throughout some region $\mathsf{D} \subseteq \Re^n$ about f_t^o and to have a bounded gradient:

$$\|\nabla f(f)\|_2 \le g\ \ \forall f \in \mathsf{D} \qquad (12)$$

The objective is a predictor $\hat{y}_{t+k} = \hat{f}(f_t^o)$ which is close to minimising the worst-case error

$$\tilde{\tilde{y}}_{t+k} \in \sup_{f \in \mathsf{F}_t} \sup_{f_t \in \mathsf{D}_t} \left| y_{t+k} - \hat{f}(f_t^o) \right| \qquad (13)$$

where

$$\begin{aligned} \mathsf{D}_t &\in \{f \mid \|f - f_t^o\|_2 \le e\}, \\ \mathsf{F}_t &\in \{f \mid y_{i+1}^o = f(f_i^o + \mathbf{w}_i) + v_{i+1}, \\ &\quad |v_{i+1}| \le d,\ \|\mathbf{w}_i\|_2 \le e,\ i = 1,...,t\} \end{aligned} \qquad (14)$$

That is, the predictor aims to minimise the largest error over all feasible f_t and over the feasible set F_t of all *f*'s compatible with the specified bounds and the observations to date.

5.2. Noise-bound specification and almost-optimal prediction algorithm

Although the functional form of $\hat{f}(f)$ need not be identified, values must be specified for the bounds g, d, e. Their values can themselves be bounded, without prior information, through the conditions for the feasible set F_t to be non-empty, *i.e.* for there to be no clash between the values of g, d, e and the observations of y and f. Necessary and sufficient conditions, respectively, are (Milanese and Novara, 2003)

$$\hat{f}(f_i^o) \ge \breve{h}_i \text{ and } \breve{f}(f_i^o) > \breve{h}_i,\ \ i = 1, 2,..., t-1 \qquad (15)$$

where

$$\left.\begin{aligned} \breve{f}(f) &= \max_{i=1, 2,..., t-1} (y_{i+1}^o - e - gd - \|f - f_i^o\|) \\ \hat{f}(f) &= \min_{i=1, 2,..., t-1} (y_{i+1}^o + e + gd + \|f - f_i^o\|) \end{aligned}\right\} \qquad (16)$$

By default, the smallest g leaving F_t non-empty can be found for trial values of d, e, thus defining a surface in (g, d, e) space above which any point is valid. In practice, the output-noise

bound e may be obtainable from knowledge of the measurement process, or the gradient bound g estimated from an approximate model. The two-dimensional feasibility boundary in the other two bound values can then be explored and a point a little above it selected; application examples seem to indicate that the choice is not usually critical.

Once g, d, e are chosen, a prediction algorithm based on F_t has to be devised. A general worst-case optimal algorithm is not available, but it can be shown (Novara and Milanese, 2001a) that the simple algorithm $\hat{f}(f_t^O) = \breve{f}(f_t^O) + \bar{f}(f_t^O)/2$ is almost optimal, with prediction error bounded by $gd + (\bar{f}(f_t^O) + \breve{f}(f_t^O))/2$. The algorithm has been applied to the Wolf sunspot series (Novara and Milanese, 2001b) and a river daily flow series (Milanese and Novara, 2002), giving maximum errors better than, and r.m.s. errors closely comparable to, a wide selection of alternative predictors including some using neural networks and local linear approximation.

This approach has potential where measurement uncertainties and model-output error are best characterised by bounds, non-linearities are smooth overall and can be bounded, a parametric model is hard to identify and a conservative, worst-case criterion for predictor performance is appropriate.

6. CONCLUSIONS

The example in Section 2 illustrated the importance of selecting suitable diagnostics of model quality. The following three sections outlined contrasting approaches to prediction as an aid to decision-making, making very different demands on the prediction model.

Classical, probabilistic Bayesian optimal decision theory offers a standard framework into which prediction and optimisation under uncertainty fit neatly, without unduly limiting the form of the model. On the other hand, it demands a large amount of prior probabilistic information, imposes heavy (though increasingly tolerable) computational load and assumes that a satisfactory cost function can be formulated, a vain hope for many applications.

Model predictive control exploits the ability of feedback, in the shape of regular revision of the prediction and optimisation, to confer robustness in the face of modelling error and to avoid the need for detailed characterisation of uncertainty. It caters for heavily constrained problems and may be able to achieve computational efficiency by keeping the optimisation in a standard form, such as quadratic programming. Although it is beginning to find application outside the process industries where it was developed, it has not yet been tried in environmental applications. Here the need for the model structure to fit the optimisation algorithm may be a significant limitation, as strong non-linearity, spatial distributedness and conflicting objectives may raise new difficulties.

The recently developed set-membership prediction technique offers worst-case, near-optimal prediction with very modest demands on uncertainty specification and model properties. Initial trials, which include river flow prediction, have shown it to perform well compared with a wide range of alternatives. More experience of its capabilities in larger problems is needed.

Finally, let us not lose sight of the other considerations in selecting and testing a prediction model. In the three approaches above, prediction and optimisation are the aims. For these purposes, the model should be economical and able to satisfy the assumptions of the prediction and optimisation techniques. However, effective decision-making also depends critically on good insight into the system, and a good prediction model may be poor at yielding insight. Familiarity, ease of interpretation, ease of testing and degree of detail may all be more important for the latter than economy or compatibility with standard assumptions.

7. REFERENCES

Bemporad, A. and M. Morari, Robust model predictive control: a survey, in *Robustness in Identification and Control*, Eds. A. Garulli, A. Tesi and A. Vicino, *LNCIS 245*, Springer-Verlag, London, 207-226, 1999.

Boyd, S., L. El Ghaoui, E. Feron and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, SIAM, Philadelphia, PA, 1994.

Calafiore, G. and M. C. Campi, Interval predictors for unknown dynamical systems: an assessment of reliability, *Proc. 41st Conf. on Decision & Control, Las Vegas*, 2002.

Carpenter, J., P. Clifford and P. Fearnhead, An improved particle filter for non-linear problems, *Institution of Electrical Engineers Proceedings on Radar, Sonar & Navigation*, 146, 2-7, 1999.

Deutsch, R., *Estimation Theory*, Prentice Hall, Englewood Cliffs, NJ, 1965.

Doucet, A., N. de Freitas and N. J. Gordon (eds.), *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, New York, 2001.

Garcia, C., D. M. Prett and M. Morari, Model Predictive Control-theory and practice-a survey, *Automatica*, 25, 335-348, 1989.

Gordon, N. J., D. J. Salmond and A. F. M. Smith, Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *Institution of Electrical Engineers Proceedings Part F*, 140, 107-113, 1993.

Holling, C. S. (ed.), *Adaptive Environmental Assessment and Management*, IIASA Int. Series on Applied Systems Analysis 3, John Wiley & Sons, Chichester, 1978.

Maciejowski, J. M., *Multivariable Feedback Design*, Addison-Wesley, Wokingham, 1989.

Maciejowski, J. M., *Predictive Control with Constraints*, Prentice Hall, New York, 2002.

Milanese, M. and C. Novara, Nonlinear set membership prediction of river flow, *Proc. 41st Conf. on Decision & Control, Las Vegas*, 2002.

Milanese, M. and C. Novara, Learning complex systems from data: the set membership approach, in L. Giarré and B. Barmish (eds.) *Multidisciplinary Research in Control: the Mohammed Dahleh Legacy*, Springer, 2003.

Morari, M. and E. Zafiriou, *Robust Process Control*, Prentice Hall, Englewood Cliffs, NJ, 1989.

Norton, J. P., Optimal smoothing in the identification of linear time-varying systems, *Proceedings of Institution of Electrical Engineers*, 122(6), 663-668, 1975.

Norton, J. P. and J. G. Chanat, Assessing rainfall-runoff records for time invariance using transfer function models with time-varying parameters, *Int. Congr. on Modelling & Simulation MODSIM 2003*, Townsville, 2003.

Novara, C. and M. Milanese, Set membership prediction of nonlinear time series, *Dipartimento di Automatica e Informatica, Politecnico di Torino internal report*, 2001a.

Novara, C. and M. Milanese, Set membership prediction of nonlinear time series, *Proc. 40th Conf. on Decision & Control*, Orlando, Fl, 2131-2136, 2001b.

Raiffa, H., *Decision Analysis*, Addison-Wesley, Reading, MA, 1968.

Smith, A. F. M. and A. E. Gelfand, Bayesian statistic withoutn tears: a sampling-resampling perspective, *The Ameican. Statistician*, 46 (2), 84-88, 1992.

Smith, J. Q., *Decision Analysis: a Bayesian Approach*, Chapman & Hall, London, 1988.

Soeterboek, R., *Predictive Control: A Unified Approach*, Prentice Hall, Hemel Hempstead, 1992.

Tay, A. S. and K. F. Wallis, Density forecasting: a survey, *Journal of Forecasting*, 19, 235-254, 2000.

Veres, G. V. and J. P. Norton, Weighted particle filter for state estimation, *Proc. IASTED Int. Conf. on Control & Applications*, Banff, Canada, 115-120, 2001.