Guidelines for the analysis of complex, dynamic systems

by
Erling Moxnes
System Dynamics Group
University of Bergen, Norway

Abstract

Many studies have found widespread misperceptions of dynamic systems. Knowing the complexity of dynamic systems and how little education is offered on this matter, misperceptions and mismanagement should not come as a surprise. Furthermore, due to complexity, analysts face difficulties in setting boundaries for models and in choosing dept of analysis. Consequently, analysts easily misunderstand each other. The purpose of this paper is to point out some basic guidelines for the analysis of complex dynamic systems. The guidelines come from optimisation under uncertainty, Bayesian statistics, and System Dynamics. All three disciplines contribute to a common research program. Taken together, they offer a deep philosophy of dynamic systems and a guide to proper analysis, and they represent a standard which all statistical, optimisation and System Dynamics works could be measured against.

1. INTRODUCTION

Nonlinear, uncertain, dynamic systems are typically hard to analyse and manage. The purpose of this paper is to present guidelines for the modelling and analysis of such systems. The guidelines will be derived from three disciplines of advanced modelling analysis: optimisation under uncertainty, Bayesian statistics, and System Dynamics.

Currently, none of the three types of analysis can be said to be used frequently, when compared to simpler methods and when compared to the frequency of decisions made in complex dynamic environments. According to diffusion analysis, Rogers (1995), the complexity of the methods could be the reason. From a utilitarian point of view, a modelling approach is the more useful the easier it is to understand and comply with. For that reason, this paper focuses more on intuitive explanations than on intricate mathematical analysis, and it points to simplified procedures when warranted and warn against such procedures when inappropriate.

From a philosophy of science point of view, the three disciplines of advanced modelling are close to belonging to the same “hard core” or “research program” in the terminology used by Imre Lakatos. Hopefully this paper will contribute to this view. With a common research program for the analysis of complex, dynamic systems, the problems of the “unavoidable a priori”, Meadows (1980), is reduced to practical issues rather than philosophy. When representatives of the three disciplines end up with different practical approaches; this will to a large extent reflect different assumptions about: purpose, needs of decision makers, costs of analysis, as well as acquired methodological skills. Hence we also keep an eye on these issues.

We limit the discussion to dynamic problems of the infinite horizon type; a class of decision problems which covers a large number of important real life situations. Bertsekas (1987) defines the infinite horizon problem from an optimisation point of view: “First, the number of stages is infinite, and, second, the system is stationary: that is, the system equation, the cost per stage, and the random disturbance statistics do not change from one state to the next. -- The assumption of stationarity is often satisfied in practice, and in other cases it approximates reasonably a situation where the system parameters vary slowly with time.” This is exactly the type of problem that is normally dealt with in System Dynamics studies, where problematic behaviours are produced by the system equations themselves. In other words, the model contains a theory of problem behaviour and therefore it can be used to test policies to solve it.

Before discussing basic requirements we distinguish between two different purposes of modelling: mental model change and policy fine tuning. When it comes to analysis, misperceptions and failures are related either to model structure or to the identification of appropriate policies. First we consider policies relying on all three disciplines of modelling. Note that not only optimal policies matter here; it is also vital to study the consequences of current policies. A proper method must make clear to decision makers why there is a problem in the first place. Without such an understanding, policy makers may not be motivated to listen to policy recommendations; they may be overconfident in current policies. Second, we consider model structure and model testing. Doing so, we again rely on all three disciplines of modelling. The key point here is that prior information is needed; time-series data never suffice. Finally, we conclude.

2. TWO DIFFERENT PURPOSES

We describe two different purposes: mental model change and policy fine tuning. We end with some information economics arguments about the process of modelling where a natural process is from the first to the second purpose.

2.1. Mental model change

Some dynamic problems last for decades and centuries without being properly understood and dealt with, others repeat themselves from one setting to another with little learning from experience, some problems are postulated to occur in the future with limited attention by policy makers. In all such situations there is likely to be a need for better mental models, for conceptual change among policy makers. The main point we want to make is that as a first step in the policy process, simple "qualitative" models have an important role to play. By qualitative we mean models that are consistent with prior structural information, but that are not necessarily very precise and established with much statistical rigour.

We illustrate the usefulness of qualitative models by a well known example from the field of economics. Assume there are clear signs of problematic overproduction in a market. The supply and demand diagram in Figure 1 shows how of a policy of price control produces a new equilibrium with overproduction. This insight is not sensitive to large variations in the parameters that determine the slope of the supply and demand curves. Furthermore, one does not have to be able to predict the equilibrium point (P_e, Q_e) or the exact size of overproduction to understand the effects of the price control. Finally, the analysis points to the current policy as the cause of the problem, and it hints at a removal of the policy as a "new" policy.

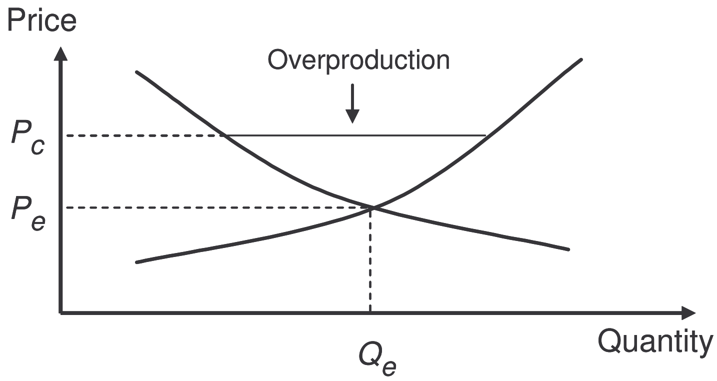

Figure 1. Supply and demand diagram illustrating effects of price control

Modern textbooks in both macro and micro economics make extensive use of similar qualitative models (e.g. Mankiw (2007) and Mansfield (1988)). The models are hardly ever presented with parameters that have been estimated statistically, although the models themselves may have been thoroughly tested. Does this mean that these textbooks are unreliable and useless? Since they are frequently used, the perception is clearly not.

One can make exactly the same observations about models of dynamic problems, although this point may require a longer explanation. Experiments show that most people have difficulties in formulating simple open-loop stock and flow problems and in reasoning about their behaviour (integration), Sweeney and Sterman (2000) and Moxnes and Saysel (revised). In light of these results, it may seem inconsistent that people manage well simple dynamic systems. However, for simple systems, feedback rules are both effective and precise. Think for example of the filling a glass with water. It is when delays and nonlinearities are added, that the simple heuristics start to produce problem behaviours, Brehmer (1989), Sterman (1989), and Moxnes (2004). Consider for example fluctuations caused by feedback adjustments of the water temperature in a shower with a pipeline delay from faucet to showerhead.

Endogenous learning is slow in complex systems, Brehmer (1990) and Paich and Sterman (1993), simple teaching interventions may have little effect, Moxnes (1998), and overconfidence may need to be reduced through information producing “cognitive conflict”, Limon (2001) and Moxnes and Saysel (revised). For these reasons simulations, manually or by the use of computers, are very useful both to be explicit about model structure and to see the behavioural consequences of model assumptions.[1]

To understand behavioural consequences, one does not usually need a model with very accurate parameters, as in the supply and demand example above. A shower model will produce fluctuations in showerhead temperature whenever the operating policy is impatient and does not wait for updated feedback information about the temperature. The model teaches the user to be sufficiently patient, independent of parameter values. Applying this knowledge in a real shower, the user could perform a quick test to find out how long it actually takes for the water to pass from the faucet to the showerhead. With this information, the appropriate patience is established for the decision rule. Without the system knowledge, it is not obvious where the fluctuations come from and

---

1 Phase diagramming methods and eigenvalue analysis are alternative ways to analyse behaviour. These are not pursued here because of an educational barrier of entry.

what information to gather. There is a widespread tendency is to assumed that complicated behaviour is caused by exogenous factors.

The example clearly demonstrates that simulation models can be very useful even though they do not produce very accurate point predictions. Again, this is parallel to the supply and demand example in Figure 1. This is a very important point, because it seems that the focus on policy analysis is easily blurred when analysts get heavily involved in improving models' ability to predict. Econometricians may erroneously demean qualitative insights of a simulation model because their focus is on prediction error and statistical testing. Their own policy analysis may be drowned in technicalities and complaints about low quality of data. System Dynamicists may go on adding model detail to explain every detail of historical behaviour. The model becomes intractable for outsiders, and the analysts end up "modelling the system" rather than a dynamic problem. In both cases, these tendencies could be seen as beginners' problems that go against good practise. A bias towards minimising prediction error could be motivated by human desires to be accurate and clever, as is indicated by investigations of conjunction fallacy and overconfidence, Plous (1993). In these cases people lose sight of the main goal and the final payoffs in their pursuit of short-term goals.

Simple, insightful models can go a long way in motivating policy change. First, simulations or data can be used to produce cognitive conflict, which makes decision makers disbelieve current erroneous mental models. For instance, data showing that the CO_2 concentration in the atmosphere continued to increase during the period 1979 to 1985 when emissions did not increase, challenged high school students' assumptions about a direct, algebraic relationship between emissions and concentration, Moxnes and Saysel (revised). Second, simple models and simulations can provide the analogies and arguments needed to build alternative mental models. Once a fundamental change in mental model has taken place, it is not easily reversed, unless there is sufficient complexity and ambiguity left to allow for denial and wishful thinking. The latter is particularly tempting whenever policy makers are invited to accept that current problems may be due to their own policies of the past, Argyris (1985) and Leibenstein and Maital (1994).

As indicated in the two examples above, considerable improvement could be obtained without much deliberation once mental models have changed. The shower policy was not fine tuned and optimal, however would represent a satisficing policy. The policy of simply removing the price control would most likely cure overproduction while the problem that motivated the price control in the first place may require further analysis. In other cases, there would be an additional benefit for policy fine tuning.

2.2. Policy fine tuning

Because of the costs involved, model improvement and policy fine tuning requires a minimum of motivation. Decision makers must see potential benefits of more complex analyses: to finance such efforts, to spend time trying to understand the results, and to trust whatever they are not prepared to comprehend. Motivation is likely to come from simple models of the type described above.

Normally, fine tuning of policies requires models that are accurate and produce good point predictions of behaviour. Hence, statistical methods are useful to identify the most likely model parameters and to reject inappropriate assumptions. Optimisation is useful to identify accurate policy implications.

However, a model's ability to explain historical behaviour is not always directly linked to its appropriateness as a testing ground for policies. One obvious possibility is that the ideal model changes from its historical version as a reaction to certain policies. For instance, a new policy to ban alcohol is not appropriately analysed in a historically accurate model that does not include the possibility of future smuggling. Another possibility is that of shifting dominance, where historical data are limited to test only the then active part of the model. For instance, in a model of yeast cell growth in wine production, data from the early period will reveal no significant effect of alcohol on yeast growth. Because policies may be more sensitive to structures and parameters that have not been influential in the past than to those that find support in historical time-series data, policies not only improve with models' ability to predict historical behaviours, Moxnes (2005).

2.3. Optimal effort

It follows from Bayesian decision theory that one should consider the expected utility before carrying out further analysis. In practise this would lead to a stepwise procedure, where at each step one decides on further effort. A likely outcome is that one should start with the analysis that produces the greatest expected utility per unit effort. This also seems to be a guiding principle in practice, Lyneis (1999). The socalled 80/20 rule is consistent with this line of reasoning, where 80 percent of the benefits are supposed to follow from 20 percent of the costs of analysis.

Seen in the light of Bayesian decision theory, one sees that the purpose of modelling naturally changes from initial efforts to question mental models, to complex analysis to fine tune policies, and maybe back to simple models to improve and disseminate mental models. These different purposes are important to keep in mind in later sections of this paper.

3. POLICY ANALYSIS

In principle, optimisation problems are infinitely complex. Through simplifications, they become solvable. Underlying models can be simplified or one could search for near-to-optimal or satisficing policies. This means that analysts need to be aware of the consequences of different types of simplification. The main purpose of this section is to give an introduction to the implications of various model attributes for optimal solutions. The insights about optimal policies should be useful for both the purpose of mental model change and for policy fine tuning. They will also be referred to when we discuss model identification in Section 4.

Optimal policies depend on formulations of objective functions (what is to be optimised), restrictions (model formulations), and ability to find optimal (or near-optimal) policies. We use the notation for discrete time systems to be closer to the literature on numerical methods and to simplify the modelling of stochasticity. Insights will be similar for continuous models.

3.1. Deterministic dynamic optimisation

Using formal notation the objective function can be written

$$J = \sum_{k=0}^{\infty} \delta^k \pi(x_k, u_k) \tag{1}$$

where J is to be maximised. J sums up net benefits π over time, where π depends on the state vector (the stocks) x_k and the decision vector u_k. Since we are dealing with infinite horizon problems, time step k ranges from 0 to infinity. The discount factor δ is normally between 1 and 0. In this case the distant future is in practise not weighted. If the discount factor is set equal to 1, there is no discounting and in practise one finds the average or total value of π over a finite time horizon. The objective function could also be designed to capture risk aversion, to value stability etc.

The restrictions are represented by the dynamic model

$$x_{k+1} = f(x_k, u_k) \tag{2}$$

The optimal solution to this dynamic and deterministic optimisation problem can be written

$$u_k = \mu(x_k) \tag{3}$$

First, note that the optimal solution is a *feedback policy*, Bertsekas (1987).[2] Second, in general the optimal policy is a *nonlinear* function of the state vector.[3]

Third, the optimal solution is a nonlinear function of the entire state vector; that is *all the stocks* of the model. For instance, to stabilise a goods inventory, one should not only base the policy on measurements of the inventory, but also include information about the amount of goods already ordered. This is consistent with the statement in System Dynamics that one should not only consider policy parameter changes, but change the structure of the problem by introducing new links.

Fourth, policies could also introduce new stocks in addition to links. For instance, by putting a small rudder on the main rudder of a ship, the ship can be steered with much less use of power. Thus, a creative policy could involve replacing the algebraic function in (3) with a dynamic policy model. With regard to optimisation, one would treat this situation as if the model in (2) was augmented with the extra states and then find an algebraic feedback policy for the augmented state vector in (3). The System Dynamics literature refers to such creative and effective policies as *leverage point policies* (pulling the lever rather than lifting the heavy load directly).

Fifth, the optimal policy depends on the instantaneous values of the state variables; there is no need for *predictions*. The model in (2) is a Markov process for which all information is contained in the current stock values. Hence, a forecast would not contribute with any useful information beyond what is contained in the state variables. Actually, a forecast would complicate the policy in that a two-step procedure would be needed, first produce a forecast, and then find a policy based on that forecast. One

---

2 In this deterministic case the solution could also be written as a function of time, $u_k = u(t_k)$. This is easily seen when the policy (3) is inserted in (2) yielding an equation where x_{k+1} is only a function of x_k. Thus, in this deterministic case the resulting x_k is only a function of time, hence u_k is also a function of time (Bertsekas (1987), p. 23).

3 Only in the case with a quadratic objective function and a linear dynamic model is the optimal policy a linear function of the state vector.

should also keep in mind that it is a challenge in itself to produce accurate point predictions, Forrester (1961) and Sterman (1986).

One may also think of forecasts as a simple way to represent exogenous influences from the rest of the world. However, such forecasts are not consistent with the stationary assumption of infinite horizon problems. Conceptually, the easiest way to deal with forecasts of exogenous variables is to augment the model (2) with the generating process. Doing this the above insights all apply. Adding uncertainty, the below insights also apply.

Case for illustrations

To illustrate some of the points made in this section we find optimal harvesting policies for a fishery model. We want to maximise the objective function, which is the expected net present value over an infinite horizon

$$J = E\left[\sum_{k=0}^{\infty} \rho^k \pi_k\right] \tag{4}$$

where π_k is the yearly net profits and ρ=0.95 is the discount factor. Yearly profits are given by revenues minus costs

$$\pi_k = ph_k - \int_{x_k - h_k}^{x_k} \frac{c}{x_k^{\beta}} dx = ph_k - \frac{c}{(1-\beta)}\left(x_k^{(1-\beta)} - (x_k - h_k)^{(1-\beta)}\right) \tag{5}$$

where p=NOK 6/kg is the price of fish, h_k is the yearly harvest, the per unit operating cost is c=NOK 3/kg, and x_k^{β} denotes the catch per unit effort with β=0.6. Costs are integrated over the yearly change in stock size x_k due to harvesting, because the catch per unit effort decreases when the stock size is reduced. The stock increases with surplus growth and decreases by harvest

$$x_{k+1} = ((x_k - h_k) + a(x_k - h_k) + b(x_k - h_k)^2)w_k \tag{6}$$

where a=0.75 and b=-0.18 are the parameters of the surplus growth, and w_k represents stochastic variation in net recruitment from year to year (lognormal distribution, iid). All parameters reflect data for cod in the Barents Sea, Moxnes (2003).

To find optimal policies $h_k=h(x_k)$ we use the program package SOPS (Krakenes (2004), Moxnes (2003), and Moxnes (2005)). Figure 2 shows the optimal policy for the deterministic case, w_k=1.0. The policy is the well known target escapement policy, where each year the harvest reduces the stock to the target (2.2 million tons). If the stock is below the target, harvest equals zero.

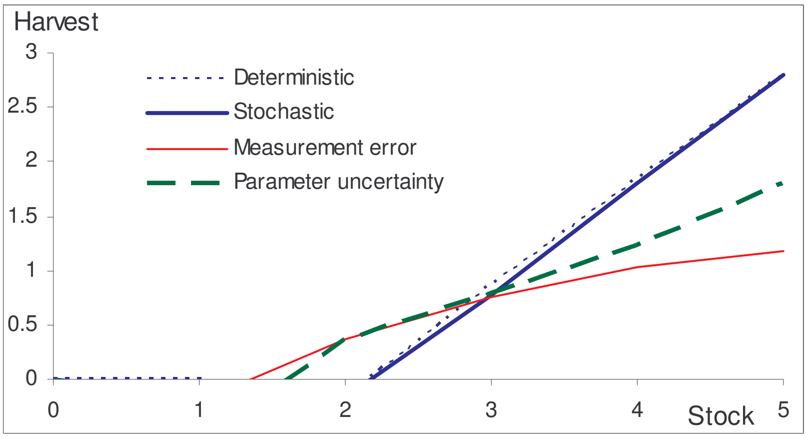

Figure 2: Harvest policies for different cases: deterministic, stochastic, stochastic and measurement error, and stochastic and lasting parameter uncertainty

3.2. Stochastic dynamic optimisation

A next step is to add stochasticity to the model in (2)

$$x_{k+1} = f(x_k, u_k, w_k) \tag{7}$$

by introducing an independently, identically distributed random variable w_k. This variable sums up all the factors that could lead to different outcomes than that predicted by the deterministic model, for instance recruitment variation and predation. With stochasticity, the net benefits vary with the stochastic variable and we have to operate with the expected value in the objective function

$$J = E\left[\sum_{k=0}^{\infty} \delta^k \pi(x_k, u_k, w_k)\right] \tag{8}$$

All the insights from the deterministic model are still valid with the following qualifications. First, the optimal solution is a feedback policy, and in this case the solution can not be expressed as a unique policy path over time. Feedback is needed to correct for the unpredictable outcomes of the stochastic process. Second, the exact relationship between state variables and the policy variable is influenced by the amount of stochasticity; one cannot simply assume the solution from the deterministic case to apply.[4]

Figure 2 shows the optimal policy for the fishery case with stochasticity[5] in recruitment. Again the policy is the well known target escapement policy with only a marginally different target from that obtained in the deterministic case, Reed (1979).

When introducing high dimensionality, stochasticity, and other types of uncertainty, optimisation requires simplification. When there are many state variables, the optimal policy will be of high order. To avoid putting severe restrictions on policies, many grid points are needed in each dimension. This led the inventor of dynamic programming, Richard Bellman, to use the term “the curse of dimensionality”. Bertsekas (1987), pp.143, points out huge needs for both computer storage and CPU time in higher order cases.

---

4 If one tries to fine tune the policy to one (deterministic) scenario, one will normally end up with a biased policy, Rockafellar and Wets (1987).

5 The lognormal distribution has mean=1.0 and standard deviation=0.15, reflecting the residuals when regressing (6) on data for cod in the Barents Sea.

The curse of dimensionality can be dealt with in two ways: either by simplifying models to enable exact optimal solutions, or by seeking approximate solutions to complex models. The first approach is attractive because it produces optimal policies that are internally consistent with model assumptions. It would be surprising if this was not the preferred approach in much theoretical work.

The second approach allows higher external consistency while there will be uncertainty left regarding the policy recommendations. The prime argument for this approach is one of practical policy making. If advanced optimisation methods fail in identifying optimal policies for complex problems, how could one expect policy makers to do so by intuition? Bertsekas (1987), writes: “When everything else fails, one has to settle for a control scheme that can be practically implemented and performs adequately (hopefully close to optimally).”, pp.143. Neuro-dynamic programming, Bertsekas and Tsitsiklis (1996) and SOPS (see references above), are both methods that seek near-to-optimal solutions to complex problems.

3.3. Stochastic dynamic optimisation with uncertain measurements

Measurement error is a further complication. In this case the true stock values x_k are not known with full precision, only uncertain measurements

$$y_k = h(x_k, v_k) \tag{9}$$

are available, here with a stochastic variable v_k. At the outset this complicates tremendously since now the optimal policy becomes a function of not only current values, but of the entire history of stock measurements y_k and decisions u_k. Over time the dimension of the policy function tends towards infinity. In principle there exists a solution where the policy only depends on the last point in time, but where the optimal policy is a function of the probability distribution for all states given the uncertain measurements, Bertsekas (1987), pp.127. A simplification of the latter approach that may work quite well is to let the policy be a function of the expected value of the estimated state $\hat{x}_k$

$$u_k = g(\hat{x}_k) \tag{10}$$

Besides leaving out information about the entire distribution for $\hat{x}_k$, a weakness of this approach is that the estimator requires linearization, for instance if the estimate is produced by a Kalman filter. Another approximation is to let the policy be a direct function of current and past measurements[6]

$$u_k = g(y_k, y_{k-1}, y_{k-2}, \ldots) \tag{11}$$

This way, the estimator with its imperfections is removed from the process. Moxnes (2003) finds that the latter method outperforms the former in a fishery example. Simply assuming that the policy is a function of the last measurement, Figure 2 shows the (very) near-to-optimal policy for the case with measurement error in our stochastic fishery example.[7] The harvest policy is less sensitive to measurements of the stock than to exact information about the stock. In the example, measurement error has a much stronger impact on the policy than stochasticity.

This case exemplifies the above discussion about model versus policy simplification. It is a well known fact that stock estimates are influenced by measurement error. Still nearly all optimisation studies have simplified their models by ignoring measurement error. Our approximate near-to-optimal policy for the complex model clearly outperforms the exact policy for the simplified model.

In case a state cannot be measured, (10) is preferable to (11) as long as the policy is sensitive to information about the unmeasured state. Using (10), the estimate of the unmeasured state will influence decisions.

3.4. Stochastic dynamic optimisation with uncertain parameters

Uncertain information about model parameters α_k is a further complication that is not easily dealt with. The model can be written

$$x_{k+1} = f(x_k, u_k, w_k, \alpha_k) \tag{12}$$

To shed light on this situation, we augment the state vector by the parameter vector α_k, see Bertsekas (1987) pp.162.

$$\alpha_{k+1} = f(\alpha_k, w_k^{\alpha}) \tag{13}$$

---

6 Moxnes (2003) explains how previous decisions u_{k-j} are removed from the equation.

7 Measurements are given as $y_k = x_k v_k$, where v_k is lognormally distributed (mean=1.0 and standard deviation=0.4). To avoid a minor bias we let the model start in year minus 10, see Moxnes (2003) for a more detailed explanation of why.

This system resembles the case with measurement error in that we have less than perfect information about the state variables representing the parameter vector. Assuming that parameters are constants, $\alpha_{k+1} = \alpha_k$, it is only the initial value of α_k that is a random variable.[8] The solution in (10) can be written

$$u_k = g(\hat{x}_k, \hat{\alpha}_k) \tag{14}$$

where $\hat{\alpha}_k$ denotes the estimate of the parameter vector. We distinguish three situations.

First, no new information about the parameters is obtained over time; the estimate is constant and equal to the estimate at the time when the policy analysis is carried out, $\hat{\alpha}_k = \hat{\alpha}_0$. This simplifies (14) considerably since $\hat{\alpha}_k$ disappears as a variable; the situation resembles that in (10). Hence, once again we can use the simplified procedure of (11) when using SOPS.

Figure 2 shows the effect of uncertainty in the parameters of the growth function in our fishery example.[9] The effect of parameter uncertainty is quite similar to the effect of measurement error. Again, the policy is clearly different from the case with stochasticity only.

Second, learning about the parameters is exogenous. That is, the distributions of the parameter estimates are narrowed over time as a result of efforts to improve models and estimation procedures. Over time, $\hat{\alpha}_k$ moves in the direction of α_k with a certain probability. Ideally, this process must be modelled, and the updated estimates can no longer be neglected in (14). Intuitively and roughly, one should expect the optimal policy to change over time. Initially, it should be quite similar to the policy with no learning. In the long run it should move towards the case with known parameters.[10]

Third, the most complex case of all is where parameter estimates are updated (reestimated) each time new measurements y_k arrive. The main problem in this case is not to update the parameter estimates; a recursive Bayesian estimation technique could provide a solution. The problem lies in the fact that the policy at one point in time

8 Parameters that change over time could be seen as model variables.

9 The two parameters in (6) are uniformly and independently distributed, $a \in (0.5, 1.0)$ and $b \in (-0.24, -0.12)$.

10 Bertsekas (1987) pp.162 comments this case by saying: "...in the great majority of cases it is practically impossible to obtain an optimal controller by means of a DP [dynamic programming] algorithm. Suboptimal controllers are thus called for...."

influences the data that is used for estimation in ensuing time periods. Hence the policy influences not only the current net benefits π_k, but also the quality of the model in later periods. This means that there is a tradeoff between short term maximisation of the criterion and the learning objective. This is commonly referred to as dual control. The final value of the objective function is optimised only when this tradeoff is optimal.

The latter problem can only be solved to optimality in extremely simple cases. Still, the dual problem is of considerable interest to modellers. Consider a model of a renewable fish resource, for which the initial parameter estimates for the surplus growth function are biased. The thin curve in Figure 3 illustrates the biased growth curve where both the maximum sustainable yield and the carrying capacity are underestimated compared to the correct curve (thick line). A policy, which maximises the yield according to the biased curve, will keep the size of the resource close to 1.5 units. The policy generates data points that are of little help to distinguish between the biased and the correct growth curves. To explore the growth curve at higher resource values, the harvest rate must be reduced. If the initial estimate had been correct, this would lead to a short-term loss. If, as assumed here, the correct curve has the higher maximum sustainable yield, harvests could be raised to a higher level than before and the net benefits would be higher in the long-term.

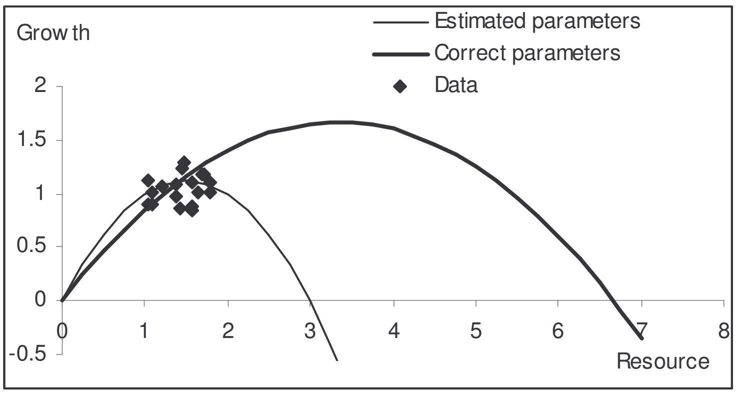

Figure 3: Surplus growth functions for estimated and correct parameters together with data points.

3.5. Conclusions regarding policy design

Clearly, dynamics, nonlinearity, and various types of uncertainty can lead to very complicated decision problems. Whether, in such systems, one makes decisions based on an intuitive trial-and-error approach or on exact optimisation with highly simplified models, there is obviously a potential for biased decisions.

However, the optimisation literature gives some guidelines such as: Explore nonlinear feedback policies; policies need not be complicated by forecasts. Consider all model stocks as potential information sources for the policy. Then search for the leverage points. Possibly add new structure to the policy, for instance a dynamic filter to deal with measurement error. Be aware that deterministic optimisation or simulation will normally lead to biased policy conclusions. Be aware of the tradeoff between data generation and short-term optimisation.

Finally, improvement is valuable even though the ideal, global optimum is not reached. If the underlying problem is erroneous mental models among decision makers, considerable improvement could result from rather simple analysis. Simple models are also likely to be needed to create cognitive conflicts to reduce overconfidence and to motivate further analysis and policy change. The more fine-tuned one wants the policy to be, the more effort must be put into the policy analysis. Those who have the most efficient tools should normally be the ones to identify the most successful policies. Hence, analysis could improve through competition.

4. MODELLING AND TESTING

Modelling is hypothesis formulation and testing against data. According to the discipline of Bayesian statistics[11], data can be split into prior information and data. For the case of dynamic models, we will usually think of data as time-series data. Both sources of data are of great value. We start by discussing hypothesis formulation, where prior information is the key source of data. Next, we discuss model testing and parameter estimation, where Bayes' theorem is used to combine prior information and time-series data.

We distinguish between three types of prior information: minimally informative priors, empirical, and subjective priors. Regarding testing methods we discuss both advanced Bayesian techniques and simple, manual heuristics. We keep the two purposes in mind: mental model change and policy fine tuning, here through model improvement.

4.1. Model formulation based on prior information

---

11 Although Bayesian statistics is not widely used, it gains popularity in some niches. Ashby (2006) writes in a 25 year review: "From sparse beginnings, where Bayesian statistics was barely mentioned, Bayesian statistics has now permeated all the major areas of medical statistics --."

For model building, prior information is a prerequisite. Consider a situation without prior information. To explain a change in A, one would have to consider all possible and impossible influences B_i. Out of millions of explanatory factors B_i, many would give statistically significant explanations of the change in A. Therefore, prior information is needed for model formulation.

This is also a point Smith (2002), who has been central in the development of experimental economics, makes: “The purpose of theory is precisely one of imposing much more structure on the problem than can be inferred from the data. This is because the assumptions used to deduce the theory contain more information, - -, than the data. The next time you report experimental data supporting a hypothesis, someone may note that the result might be due to “something else”. Of course, this is necessarily, and trivially, true; there are an infinite number of them.”

According to Berger (2006): “Model-building is not typically part of the objective/subjective debate, however - in part because of the historical success of using models, in part because all the major philosophical approaches to statistics use models and, in part, because models are viewed as “testable", and hence subject to objective scrutiny. It is quite debatable whether these arguments are sufficient to remove model choice from the objective/subjective debate --.” In other words, model formulations could range from pure subjective speculation to well founded theories based on much prior information. The latter seems preferable.

Dynamic models formulated in differential or difference equations are used in nearly all fields of study. Much experimentation and observation has led to the accumulation of large amount of prior information. There are laws of physics, chemistry, biology, medicine etc. and many of these laws are supplied with universal constants. When modelling for instance commodity cycles in the hog market, Meadows (1970), the modeller can build on precise prior information related to the modelling of growth of hogs with details about gestation time, offspring per livestock, time to mature, weight growth etc. These are typically well established biological relationships that need little testing against time-series data. Furthermore, people in the business have information about business rules and regulations, market structure, availability and accuracy of information etc.

Formulating dynamic models is not a trivial task; both laboratory experiments and experiences from teaching suggest that considerable knowledge and training is needed. Very useful in this regard are the principles and guidelines for modelling developed in the discipline of System Dynamics, Forrester (1961), Forrester (1968), Forrester and

Senge (1980), and Sterman (2000). For instance: diagramming techniques and equations that clearly distinguish stocks and flows, and that distinguish in- and outflows; practical definitions of stocks; the feedback concept where causality is circular and not only one way; formulations of nonlinearities; distinction between goal and realisation; boundary adequacy; real life interpretations of variables and parameters; and numerous tests to ensure that theories are internally consistent, robust, and consistent with prior information. These guidelines enable students to rediscover scientific discoveries of the past, given that the proper questions are raised. Hence they should also be helpful in guiding future discoveries.

One of the greatest difficulties is to capture human decision making. How humans actually combine information sources to make decisions can be quite complicated. In the tradition of economics, a way out of this difficulty has been to assume that people make "rational decisions". However, in the light of the earlier discussion of optimisation under uncertainty, it is not at all clear what rational decisions are in complex, uncertain, dynamic systems. The choice of a policy formulation somewhere between state of the art in the optimisation literature and the simplest of heuristics must be subjective, unless it builds on prior information.

Optimisation under uncertainty points to policies that rely on feedback. Laboratory experiments suggest that people rely on simple heuristics that work well in simple systems but cause problems in more complex systems for which people are not able to construct appropriate models. Oscillating showerhead temperatures exemplify; the heuristic that works well when filling a glass with water fails in a system with a pipeline delay. With precise and frequent feedback, learning and policy adjustment is likely to take place in simpler systems. Without, successful policies seem to rely on abilities to analyse dynamic systems. If so, it should be possible to obtain some prior information about this.

4.2. Bayes' theorem

Bayes' theorem or Bayes' rule prescribes how prior information and data should be combined to give more precise posterior estimates and hypothesis tests. The main purpose of the presentation is to convey the ideas and intuition behind.

According to Zellner (1984): "Ideally, it would be desirable to have a unified set of principles for making inferences and decisions which can be readily applied in a broad range of circumstances and fields to yield good results. One of the main points of this paper, and hardly a novel one, is that the Bayesian approach approximates this ideal much more closely than do non-Bayesian approaches currently in use in econometrics.", p.187. This praise of Bayesian statistics does not necessarily mean that it is, or will be widely used for all purposes. One important reason for this is that Bayesian statistics appears more difficult to learn and apply than other statistical methods, Moore (1997). Even when Bayesian statistics are not used formally, the underlying philosophy should provide guidelines for model testing.

The continuous version of Bayes' theorem reads:

$$f_{\Theta|X}(\theta \mid x) = \frac{f_{X|\Theta}(x \mid \theta) f_{\Theta}(\theta)}{\int f_{X|\Theta}(x \mid \theta) f_{\Theta}(\theta) d\theta} = c \cdot f_{X|\Theta}(x \mid \theta) f_{\Theta}(\theta) \qquad (15)$$

Here $f_{\Theta|X}(\theta \mid x)$ denotes the *posterior* distribution for an unknown parameter θ given new data x. The posterior distribution is formed by multiplying the likelihood function $f_{X|\Theta}(x \mid \theta)$ by the *prior* distribution $f_{\Theta}(\theta)$ for θ. The *likelihood* function denotes the probability of obtaining the new data x given the parameter θ. The denominator is the integral of the numerator over θ. This is just a constant which ensures that the entire expression becomes a proper probability distribution where the cumulative distribution tends towards 1.0 as θ tends towards infinity. To simplify the expression, the denominator is replaced by $1/c$ in the last part of (15).

In the following we illustrate how Bayes' theorem works and what the challenges are. At the outset there are three challenges. First, a prior distribution must be specified, second a likelihood function must be determined, and third, the posterior distribution must be found. To simplify, we assume that we will test a model with only one parameter θ.

We start by assuming that the prior is uniformly distributed. Then we obtain the likelihood function from a regression where we estimate θ from time-series data. The distributions are illustrated in Figure 4. Using Bayes' theorem, we find a posterior distribution that overlaps perfectly with the likelihood function. Thus, in this case the prior distribution does not contribute to a more precise determination of θ. The example is constructed such that the expected value of θ based on the likelihood is not significantly greater than zero. This is a quite frequent observation when testing dynamic models on time-series data of limited length and quality.

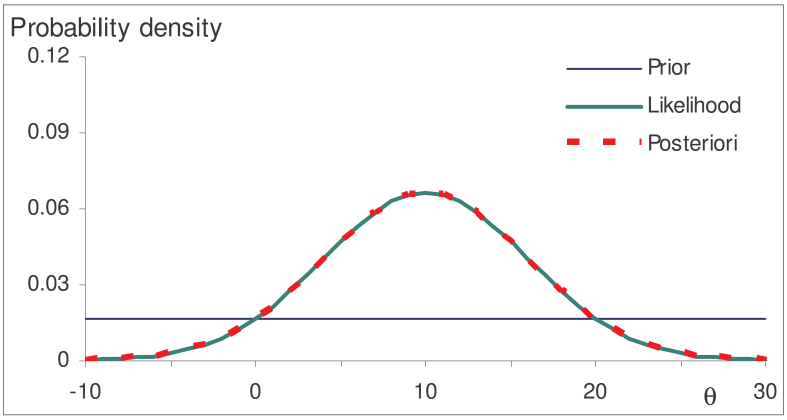

Figure 4: Example - uniform prior, binomial likelihood and posterior.

Next, consider a situation where we do have prior information, illustrated by the prior distribution in Figure 5. In this illustration, the prior estimate has an expected value of 8, two units below the expected value of the likelihood function. Thus one effect on the posterior distribution is to shift the expected value of θ to a lower value. The distribution of the prior is nearly as wide as that for the likelihood, still the prior leads to a considerably more narrow posterior distribution. In spite of the lower expected posterior value for θ, the parameter estimate is now significantly different from zero. The multiplication of the likelihood with the prior in (15) explains why we get these effects.

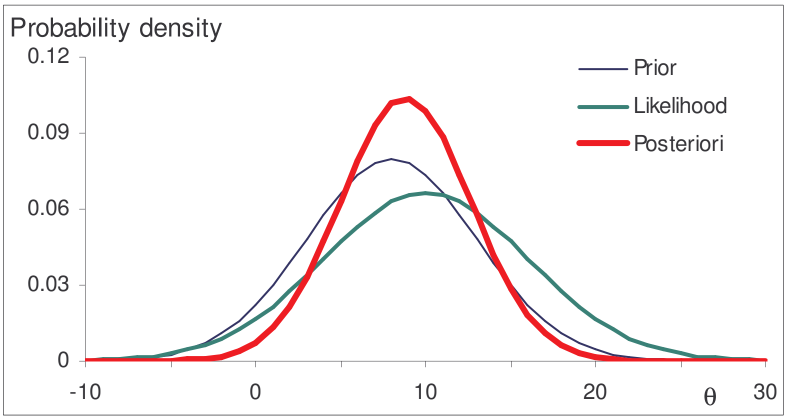

Figure 5: Example - binomial prior, binomial likelihood and resulting posterior.

Bayes' theorem can also be used to blend the results of previous studies with new data, irrespective of what type of experiment or study the prior comes from. This requires that all studies focus on the distribution of the same θ. This requirement is relaxed in meta-analysis, Glass (2000).

The illustration shows that the accuracy of the posterior estimate can be improved by both better prior information and by the use of more time-series data. Where to search for more information should be guided by the expected improvement obtained per unit of money spent. There is no general answer to where, sometimes prior information is nearly perfect and sometimes will repeated experiments provide cheap and long time-series.

4.3. Three different priors

Depending on the type of information source that is used, the literature operates with different types of priors. Here we focus on minimally informative, empirical, and subjective priors.

Minimally informative priors do not require much specific information about the unknown parameter θ. For that reason minimally informative priors tend to be noncontroversial. A uniform distribution exemplifies.[12] A major reason for using minimally informative priors is that they allow the analyst to use the Bayesian framework without appearing to be subjective. In the literature on Bayesian statistics, minimally informative priors are usually referred to as objective priors. That definition is misleading since empirical priors should also fall into the objective prior category.

There is a philosophical divide between Bayesian statistics and the more standard frequentist approach.[13] However, in practise and as long as one uses minimally informative priors, the two approaches often lead to quite similar results. Berger (2006) shows examples where on average Bayesian confidence intervals are somewhat smaller than the intervals established by frequentist methods. Probably these differences are also small compared to the biases introduced by misspecifications, measurement errors, correlated data etc. Berger continues: “Bayarri and Berger (2004) review the vast literature showing that objective Bayesian methods are the most promising route to the unification of Bayesian and frequentist statistics.”

---

12 Berger (2006) provides a list of minimally informative priors with desirable properties.

13 In the Bayesian framework the parameter θ is seen as a random variable. Thus both the prior and the posterior denote distributions from which one can establish means, standard deviations and confidence intervals; all useful measures for decision making. The framework of frequentist statistics holds that θ is an unknown constant, which either lies or does not lie in an interval, there is no probability involved. Probability is only used about the relative frequency of occurrence of an event. This probability should be defined as the limit of its relative frequency in a large number of trials. According to Zellner (1984, p.193): “However, this is not to say that non-Bayesians do not engage in considerations concerning whether a particular hypothesis is probably true. They do so, but only in an informal and subjective manner.”

Empirical priors[14], as we define them here, build on data other than time-series data. Priors that are not minimally informative are usually referred to as subjective. For priors that build on generally available numerical data, this seems inappropriate. We have already mentioned biological data about for instance gestation and maturation times. These are estimates based on much numerical information and with no particular element of subjectivity.

Forrester (1980a) makes the point that there is vastly more information in people's mental data bases than in written data bases, which again exceed that of numerical data bases. Many sciences have developed and used techniques to elicit data from the mental and written data bases.[15] Some of these elicited estimates may be subject to well known biases in judgement, Wolfson et al. (1996).[16] To the extent that biases are systematic and well know, biases could be reduced by researchers by making standard corrections for underestimations of delay times, standard deviations etc. When elicited priors are based on collected information, the elicitation methods are not necessarily subjective in the sense that they are biased by the researcher's own views. Furthermore, the fact that the estimates may be biased does not necessarily rule them out as prior information. Likelihood functions are also likely to be biased by measurement errors, model misspecifications, and overly simplified estimation techniques.

In most cases elicited information is used as it is. In our case the elicited data is treated as prior information and will be updated by likelihood functions from time-series data. This automatically limits the potential damage of highly biased priors. Furthermore, great deviations between prior estimates and expected values of likelihood functions should stimulate to critical appraisals of both priors and likelihoods. It should also be kept in mind that when there is no valuable prior information about a parameter, a minimally informative prior should be used.

By subjective priors we think of prior parameter estimates that are guesstimates made by the researcher, possibly influenced by inputs from others, but not arrived at through a documented and replicable procedure. The first associations that come to mind may range from prejudice to fraud. For instance subjective parameters seem totally unacceptable in a study designed to find the effect of a new drug. However, when

---

14 Note that empirical priors should not be confused with and has nothing to do with "Empirical Bayes methods".

15 For the use of elicitation techniques in connection with simulation models see: Luna-Reyes and Andersen (2004), Ford and Sterman (1998), and Vennix et al. (1992).

16 There is also a danger that prior estimates are influenced by the same data that are used in the likelihood function. In this case the same data will erroneously be used twice to reduce the variance of the posterior.

dealing with dynamic systems, there are some interesting exceptions. First, for the purpose of mental model change, rough parameter guesses may be sufficient to learn about the dynamics of a system. Our previous example of showerhead temperature illustrates; the oscillations that occur due to the pipeline delay could be observed for a wide range of pipeline parameters and decision rules.

Second, to determine optimal research efforts, a model with subjective parameters may be used to perform first rough tests of the hypothesised model structure to see if it reproduces the behaviour patterns seen in time-series data. Furthermore, sensitivity tests on such a model will indicate which parameters are most important and are most in need of precise prior estimates. Different from open-loop models, models with negative feedback loops (also referred to as counteracting or balancing loops) tend to be quite insensitive to large variations in many parameters.

Third, clients may want to test models based on their own subjective prior information. This is the information they will base their decisions on anyway, such that subjectivity will be the same. If the client tries to deceive someone, it will be him- or herself. Model simulations or optimisations provide extra checks on the client's mental models and the corresponding behavioural implications.

Because empirical and subjective priors can be very useful, the divide between the objective and the subjective Bayesian schools may not be very deep. Berger (2006) writes: "Note that, in practice, I view both objective Bayesian analysis and subjective Bayesian analysis to be indispensable, and to be complementary parts of the Bayesian vision." Then he adds: "--I feel that there are a host of practical and sociological reasons to use the label "objective" for priors of model parameters that appropriately reflect a lack of subjective information."

4.4. Methods for model testing

As for optimisation, complex dynamic models present great challenges for model testing. We start by discussing the augmented Kalman filter, which represents a Bayesian approach. Then we comment on a series of tests that go beyond the standard statistical tests. Finally, we present a light version of Bayesian statistics for model calibration and testing.

As mentioned before, most statistical methods are frequentist methods and do not take account of prior information. Many of these methods have been discussed and used by

System Dynamicists to analyse dynamic models, e.g. Peterson (1980), Graham (1980), Hamilton (1980), and Eberlein and Wang (1985). However, because dynamic models consist of coupled, nonlinear equations; time-series are short, autocorrelated, and with measurement errors; there is a considerable need for prior information. Hence there are good reasons to think that Bayesian methods are superior to frequentist methods for this purpose.

One such Bayesian method is the augmented Kalman filter where parameters and system states are estimated simultaneously. Parameters and states are updated from prior values. This updating slows down if there is measurement error present. A weakness of the Kalman filter is that it is ideal only for linear systems. An extended version deals with nonlinear systems, relying on linearisation. The augmented filter is also an extended filter since parameters and state variables normally form nonlinear relationships in model equations. Another weakness of the augmented and extended Kalman filter is that it does not always converge.

Using Bayesian methods it is important to be consistent when it comes to the definition of priors. Estimation procedures typically operate with discrete time difference models. Then, for instance, an interest rate will be defined as the interest accrued over one time step. Therefore, using the instantaneous interest rate as a prior would introduce a bias. In general, estimation models with coupled difference models represent the one time step solution to the corresponding coupled differential equations model. In these one step solutions all future states will in principle depend on the current state of all state variables, even when the differential equation model has only a few links from states to flows. Hence, using instantaneous priors in a difference equation model could lead to both parameter biases and model errors (many priors are set equal to zero, overly simplifying the discrete time model[17]). It is not sufficient to adjust each and every prior parameter from instantaneous to time step values.

The above is only a major problem when the time step of the estimation model is long compared to the shortest time constant in the model. Since the time step is typically defined by the time step of available data, discrete time Bayesian techniques typically depend on sufficiently rapid updating of data. If data are not reported sufficiently frequently, this is an argument for using differential equations, where the simulation step length can be set short enough to allow for the use of instantaneous parameter definitions.

---

[17] This is an argument in favour of methods proposed by Henry and Richard (1982) and Henry and Richard (1983) of letting the data speak by allowing for multiple explanatory factors.

To simplify Bayesian estimation one could estimate and test models equation by equation. This is also what is typically done when using frequentist methods. This is often referred to as a problem because one neglects feedback from dependent to independent variables when regressing. However, just as stocks decouple otherwise simultaneous equations, they also remove the direct effect of the left-hand side variable on the right-hand side variables in difference equations. Hence, one may wonder if this particular problem is often exaggerated. However, all other deviations from the idealised conditions required by statistical methods could lead to parameter biases. Thus, when all equations are put together, the model may fail to endogenously explain historical behaviour. Hence, model testing is not necessarily finished when the individual regressions have been made.

A last complication to be discussed here is the one where no measurements exist of a variable. Soft variables like expectations and attitudes exemplify. Many or most decisions are influenced by these types of variables. Even if they cannot be measured, they can be quantified and modelled. Given that the unmeasured variable is observable (meaning that it can be estimated based on information about other state variables), the model can be reformulated such that the unmeasured variable disappears. The parameters of the original equation for the unmeasured variable will however show up in all equations that the unmeasured variable influences. As a consequence, these equations will be complicated by more parameters to be estimated, additional nonlinearities, and more complex dynamics. With limited length time-series data and measurement error, the likelihood of obtaining statistically significant estimates deteriorates.

Because of the many possible complications, all formal statistical methods produce results with biases and remaining uncertainty. Hence a theory that has been falsified by one method may get support from some other method (Duhem-Quine problem). This calls for alternative tests that also go beyond Bayesian statistics.[18] Numerous such tests are proposed and discussed in Mass and Senge (1978), Forrester and Senge (1980), Zellner (1981), Barlas and Carpenter (1990), and Chapter 21 in Sterman(2000).

---

[18] Some of these additional tests only imply minor changes in the standard statistical methods. To test for cyclical behaviour, one could for example use a criterion with the deviation between the frequency distribution of the model and that of the data (Kullbach-Leibler's method) rather than the prediction error. Doing this one implicitly also questions the existence of cycles in the data by explicitly considering the frequency distribution. In practise, however, it is not easy to establish frequency distributions in non-stationary time-series.

Based on the many potential problems and weaknesses of Bayesian and frequentist methods[19], as well as considerable costs in terms of education and analysis, it is tempting to explore simplified techniques. This has been tried in System Dynamics with critical remarks from statisticians as a result. For instance, Zellner (1980) kindly suggested that System Dynamics might benefit from a greater emphasis on statistical techniques. Forrester (1980b) replied: “He might be correct; however, so many people are already working with statistical methods that perhaps some of us should emphasize alternatives.” Below, we take both suggestions seriously and indicate how principles and insights from Bayesian statistics could be used to justify and to improve simplified techniques for model testing and parameter calibration.

First consider the case where there exists perfect prior information about all parameters and the model structure. Simulations will be needed only to make explicit the behaviour of the model and eventually to test policies. If time-series data are influenced by measurement errors, this would be the only cause of deviation between simulated and observed behaviour. The certainty of the prior estimates implies that parameters should not be adjusted to improve the fit.

Second, some prior estimates are lacking or are not significantly different from zero. These parameters could be calibrated to improve the fit between simulated and observed behaviour while the more certain prior parameters estimates are kept as they are.[20] This could be seen as a special case of Bayesian statistics where either prior estimates or data are used. Figure 4 illustrates the case with no prior information. Figure 5 illustrates a case were both likelihood and prior counts, a case that is missed with a rough either-or approach. If the number of parameters without prior information is limited, one could of course also make changes in parameters for which uncertain information exists to improve the fit to time-series data, however with an eye to the prior. Compared to single equation statistical tests, the simulation test captures all model interactions. Compared to discrete time Bayesian models, the simulation model can always be run with a sufficiently short time-step to ensure that prior information about both structure and parameters can be used directly and correctly.

---

[19] According to Leontief (1971) in his presidential address to the American Economic Association: “In no other field of empirical inquiry has so massive and sophisticated a statistical machinery been used with such indifferent results.”

[20] Moxnes (1990) provides an example where a dynamic model of fuel shares reproduces historical developments when using prior estimates for equipment costs and lifetimes, while uncertain “constant convenience premiums” are found by calibration. Just as obtained statistical estimates of parameters are often judged ex post by their signs, the obtained premiums were judged by their magnitudes.

In the second case simulated behaviour could be sufficiently close to observed behaviour that a trained statistician would regard the probability of model rejection as highly unlikely. Figure 6 exemplifies. The thick solid line shows the historical development of anthropogenic CO_2 in the atmosphere. The thin solid line shows simulated behaviour of a dynamic model where CO_2 is represented as a stock, the inflow is the historical emission rate, and the outflow equals the stock divided by an average lifetime of 40 years. The lifetime is the only unknown parameter and it has been adjusted manually. The fit is very good and there is no reason to reject the model based on this fit. In this simple case, a statistical test would also find a highly significant parameter for the lifetime.

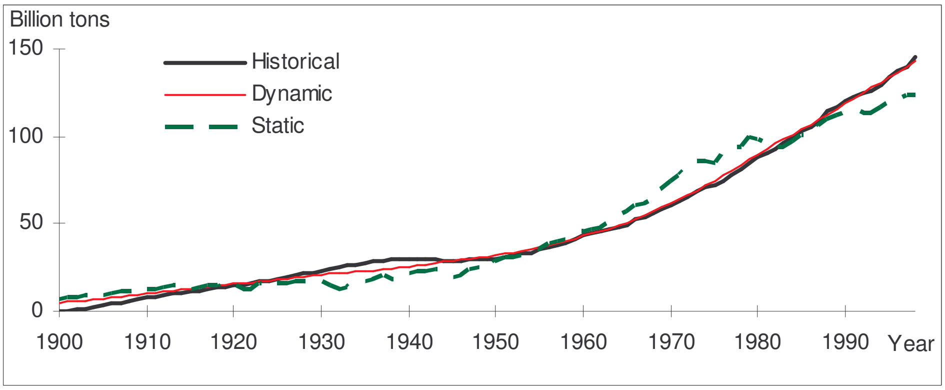

Figure 6: Historical development of anthropogenic CO_2 in atmosphere compared to simulated developments with a dynamic and a static model.

To illustrate the difficulty of rejecting a false hypothesis by standard statistical methods, Figure 6 also shows the behaviour of a static model. Here the CO_2 concentration is a linear model of the historical emission rate. A regression gives the impressive t-ratio of 45.4 for the slope coefficient, clearly implying no rejection.[21] The example shows that quite large deviations from historical data may not lead to rejection. Probably, a person using simulations would be more sceptical of the linear model than a person seeing only a high t-ratio. Comparing simulations and historical behaviours also allow for other comparisons than just prediction error. Most important, does the model reproduce the problem behaviour of interest?

In case many prior estimates are highly uncertain or do not exist at all, manual calibration to fit historical data becomes increasingly unjustifiable. This is because many sets of adjustable parameters could possibly make the model fit the historical data. Statistical methods would normally warn that obtained estimates are not

---

[21] A second order static model obtains an even better fit to the historical data with t-ratios higher than 4.2 for both the first and the second order term.

statistically significant. Manual calibration in this case could also be highly subjective and influenced by the modeller's biases. Hence, testing of parameter rich dynamic models through simulation is likely to depend on a minimum of prior information. This is a main reason for System Dynamics textbooks to insist on real life interpretations of model variables and parameters.

Prior information about parameters could also come from previous studies. That could be valuable information to improve problem solving. Testing a decision rule with its parameters could also reduce the uncertainty about it. For instance, Meadows (1970) used the same decision rules in three different commodity market models. Other than the decision rules, the models built on prior information about the biology of chicken, hogs, and cattle. The three models produced cycles with distinct frequencies, similar to the frequencies observed in time-series data for the three respective markets.

4.5. Conclusions regarding model testing

Like for optimisation, nonlinear, uncertain, dynamic models present numerous challenges for model testing. In the light of model complexity and lack of long and reliable time-series data, prior information is of great potential value. Bayesian statistics provide a formal apparatus to blend priors and likelihoods obtained from time-series data. Hence, Bayesian statistics allow for the use of more information than more standard frequentist methods. However, practical use of Bayesian statistical methods requires simplifications that could introduce biases. For the moment, the methods also seem to be quite costly in use. This creates interest in a simplified Bayesian approach where as much prior information as possible is introduced before the model is calibrated to replicate the problem behaviour modes of interest in historical time-series. The more parameters without prior information, the more subjective the simplified Bayesian approach becomes and the larger the need for Bayesian or frequentist statistical methods.

5. CONCLUSIONS

Bayesian statistics and advanced optimisation under uncertainty provide important guidelines for model construction, testing, and policy identification. Blended with the more practical philosophy and guidelines from the discipline of System Dynamics, a powerful and workable methodology emerges. It may seem like a paradox when we claim that in complex dynamic systems, manual model testing and manual search for policies could outperform rigorous methods from statistics and optimisation. There are good arguments for this claim, and there are limitations and traps. We have also pointed out the natural progress in studies from simple systems to challenge current mental models to comprehensive, well tested models for policy fine tuning. In this broader view advanced optimisation, Bayesian statistics and System Dynamics could be seen to belong to the same research program. The practitioners of the latter two disciplines probably have a bias towards prediction accuracy to the detriment of a clear problem focus. In optimisation there is likely to be a bias toward exact solutions, to the detriment of more realistic models.

For further research, it would be very interesting to arrange practical "competitions" between formal and more practical methods. Synthetic data experiments could be used where the data generating process is known. Different cases should be chosen that are thought to favour either formal or practical approaches. Furthermore, new simple to use Bayesian statistical tests for dynamic models are likely to be very beneficial, and would probably bring the disciplines of System Dynamics and Bayesian statistics closer together.

REFERENCES

Argyris, C. (1985). *Strategy, Change, and Defensive Routines*. Boston: Pitman.

Ashby, D. (2006). "Bayesian statistics in medicine: A 25 year review." *Statistics in medicine* **25**(21):3589-3631.

Barlas, Y., and Carpenter, S. (1990). "Philosophical Roots of Model Validation: Two Paradigms." *System Dynamics Review* **6**(2):148-166.

Bertsekas, D.P. (1987). *Dynamic Programming. Deterministic and Stochastic Models*. New Jersey: Prentice-Hall.

Bertsekas, D.P., and Tsitsiklis, J.N. (1996). *Neuro-Dynamic Programming*. Belmont Massachusetts: Athena Scientific.

Brehmer, B. (1989). "Feedback Delays and Control in Complex Dynamic Systems." Pp. 189-196 in *Computer Based Management of Complex Systems*, edited by Milling, P. and Zahn, E. Berlin: Springer Verlag.

Brehmer, B. (1990). "Strategies in Real Time, Dynamic Decision Making." Pp. 262-279 in *Insights in Decision Making*, edited by Hogarth, R.M. Chicago: University of Chicago Press.

Eberlein, R.L., and Wang, Q. (1985). "Statistical estimation and system dynamics models." Pp. 206-222 in *1985 International Conference of the System Dynamics Society*, edited by Warkentin, M.E. Keystone, Colorado: http://www.systemdynamics.org/.

Ford, N.D., and Sterman, J.D. (1998). "Expert knowledge elicitation to improve formal and mental models." *System Dynamics Review* **14**(4):309-340.

Forrester, J., and Senge, P. (1980). "Tests for building confidence in system dynamics models." *TIMS Studies in the Management Sciences* **14**:209-228.

Forrester, J.W. (1961). *Industrial Dynamics*. Cambridge: MIT Press.

Forrester, J.W. (1968). *Principles of systems*. Cambridge: Wright-Allen Press.
Forrester, J.W. (1980a). “Information Sources for Modeling the National Economy.” *Journal of the American Statistical Association* **75**(371):555-571.
Forrester, J.W. (1980b). “Information Sources for Modeling the National Economy: Rejoinder.” *Journal of the American Statistical Association* **75**(371):572-574.
Glass, G.V. (2000). “Meta-Analysis at 25.”.
Graham, A.K. (1980). “Parameter estimation in system dynamics modeling.” in *Elements of the System Dynamics Method*, edited by Randers, J. Cambridge: MIT Press.
Hamilton, M.S. (1980). “Estimating lengths and orders of delays in system dynamics models.” in *Elements of the System Dynamics Method*, edited by Randers, J. Cambridge: MIT Press.
Henry, D.F., and Richard, J.F. (1982). “On the Formulation of Empirical Models in Dynamic Econometrics.” *Journal of Econometrics* **20**:3-33.
Henry, D.F., and Richard, J.F. (1983). “The Econometric Analysis of Economic Time Series.” *International Statistical Review* **51**:111-163.
Krakenes, A. (2004). “SOPS - a tool to find optimal policies in stochastic dynamic systems.”. Bergen: System Dynamics Group, Department of Information Science, University of Bergen.
Leibenstein, H., and Maital, S. (1994). “The Organizational Foundations of X-Inefficiency. A Game-Theoretic Interpretation of Argyris´ Model of Organizational Learning.” *Journal of Economic Behavior and Organization* **23**(3):251-268.
Limon, M. (2001). “On the cognitive conflict as an instructional strategy for conceptual change: a critical appraisal.” *Learning and Instruction* **11**(4-5):357-380.
Luna-Reyes, L.F., and Andersen, D.L. (2004). “Collecting and analyzing qualitative data for system dynamics: methods and models.” *System Dynamics Review* **19**(4):271-296.
Lyneis, J.M. (1999). “System dynamics for business strategy: a phased approach.” *System Dynamics Review* **15**(1):37-70.
Mankiw, N.G. (2007). *Brief Principles of Macroeconomics*: Thomson South Western.
Mansfield, E. (1988). *Microeconomics*: W W Norton & Co Ltd.
Mass, N.J., and Senge, P.M. (1978). “Alternative Tests for the Selection of Model Variables.” *IEEE Transactions on Systems, Man and Cybernetics* **SMC-8**(No 6):450-480.
Meadows, D.H. (1980). “The Unavoidable A Priori.” in *Elements of the System Dynamics Method*, edited by Randers, J. Cambridge: MIT Press.
Meadows, D.L. (1970). *Dynamics of commodity production cycles*. Cambridge, Mass.: Wright- Allen Press.
Moore, D.S. (1997). “Bayes for Beginners? Some Resons to Hesitate.” *American Statistical Association* **51**(3):254-264.
Moxnes, E. (1990). “Interfuel substitution in OECD-European electricity production.” *System Dynamics Review* **6**(1):44-65.
Moxnes, E. (1998). “Overexploitation of renewable resources: The role of misperceptions.” *Journal of Economic Behavior and Organization* **37**(1):107-127.
Moxnes, E. (2003). “Uncertain measurements of renewable resources: Approximations, harvest policies, and value of accuracy.” *Journal of Environmental Economics and Management* **45**(1):85-108.
Moxnes, E. (2004). “Misperceptions of basic dynamics, the case of renewable resource management.” *System Dynamics Review* **20**(2):139-162.

Moxnes, E. (2005). “Policy Sensitivity Analysis: simple versus complex fishery models.” *System Dynamics Review* **21**(2):123-145.
Moxnes, E., and Saysel, A.K. (revised). “Misperceptions of basic climate change dynamics.” *Climatic Change*.
Paich, M., and Sterman, J. (1993). “Boom, Bust, and Failures to Learn in Experimental Markets.” *Management Science* **39**(12):1439-1458.
Peterson, D.W. (1980). “Statistical tools for system dynamics.” in *Elements of the System Dynamics Method*, edited by Randers, J. Cambridge: MIT Press.
Plous, S. (1993). *The Psychology of Judgment and Decision Making*. New York: McGraw Hill.
Reed, W.J. (1979). “Optimal escapement levels in stochastic and deterministic models.” *Journal of Environmental Economics and Management* **6**:350-363.
Rockafellar, R.T., and Wets, R.J.-B. (1987). “Scenarios and Policy Aggregation in Optimization Under Uncertainty.” WP-87-119: IIASA.
Rogers, E.M. (1995). *Diffusion of innovations*. New York: The Free Press.
Smith, V. (2002). “Method in Experiment: Rhetoric and Reality.” *Experimental Economics* **5**:91-110.
Sterman, J.D. (1986). “Modeling the Formation of Expectations: The History of Energy Demand Forecasts.” *International Journal of Forecasting*.
Sterman, J.D. (1989). “Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment.” *Management Science* **35**(3):321-339.
Sterman, J.D. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Boston: Irwin/McGraw-Hill.
Sweeney, L.B., and Sterman, J.D. (2000). “Bathtub dynamics: initial results of a systems thinking inventory.” *System Dynamics Review* **16**(4):249-286.
Vennix, J.A.M., Andersen, D.F., Richardson, G.P., and Rohrbaugh, J. (1992). “Model-building for group decision support - issues and alternatives in knowledge elicitation.” *European Journal of Operational Research* **59**(1):28-41.
Wolfson, L.J., Kadane, J.B., and Small, M.J. (1996). “Bayesian environmental policy decisions: two case studies.” *Ecological Applications* **6**(4):1056-1066.
Zellner, A. (1980). “Information Sources for Modeling the National Economy: Comment.” *Journal of the American Statistical Association* **75**(271):567-569.
Zellner, A. (1981). “Philosopy and Objectives of Econometrics.” in *Macroeconomic Analysis: Essays in Macroeconomics and Econometrics*, edited by Currie, D., Nobay, R., and Peel, D. England: Croom Helm.
Zellner, A. (1984). *Basic Issues in Econometrics*. Chicago: The University of Chicago Press.