Bayesian Analysis of Stochastic System Dynamics

Rudolf Kulhavý

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4, 182 08 Prague, Czech Republic
Tel: +420 26605 2061 Fax: +420 26605 2068
E-mail: kulhavy@utia.cas.cz

ABSTRACT. The paper deals with the system dynamics modeling of a stochastic behavior. The starting point is replacing the traditional system dynamics model with a discrete-time stochastic dynamic model in which state variables are measured indirectly, through noisy and incomplete measurements. The state variables and possible unknown parameters in such a model can be systematically estimated from the available measurements using the Bayesian paradigm. Closed-form solutions exist only for a few special cases, such as a linear normal model with known parameters, otherwise numerical approximations are required. The paper suggests a particle filter algorithm as a particularly appealing approximation that preserves much of the intuitive workings of system dynamics. A practical example illustrates both the stochastic modeling process and the approximate Bayesian analysis.

KEYWORDS. System dynamics; stochastic behavior; Bayesian inference; sequential Monte Carlo methods; particle filtering.

Introduction

The practical appeal of system dynamics modeling stems to a large extent from the conceptual simplicity of its underlying mathematics. It is by no means accidental that most system dynamics publications focus on construction of a particular model rather than technical issues. This is in stark contrast to theoretical publications in statistics, machine learning, control or signal processing that predominantly aim at advancing the technical knowledge while practical models, if considered at all, serve typically as motivational or illustrative ones.

It is a matter of continued discussion (see, e.g., Forrester, 1985; Graham, 2002) whether and how system dynamics theory and practice can benefit from the technical progress made by the other communities, without losing its traditional focus on practical modeling of complex, real-life systems.

One of the issues that has long attracted the attention of modelers of all provenances is the treatment of a stochastic behavior of systems. In fact, there are many practical problems where the stochastic behavior cannot be neglected and needs to be carefully modeled.

One class of such problems appear in connection with *management of financial and operational risks*. The risk management is traditionally based on quantifying a 'value at risk', defined as a proper quantile of the underlying probability distribution of the target variable. Simulation of any single "representative" behavior is not sufficient for calculation of the value at risk; one has to consider the entire collection of possible behaviors to calculate the quantile.

Another class of problems is related to modeling of the *dynamic performance of value networks* in which individual organizations dynamically recombine in order to deliver jointly a tangible value to the end customer. The goods and services produced by one node are consumed by other nodes. Each of the nodes adds value along the network. The uncertainty as to the value added by individual nodes propagates through the network in a way that is difficult to grasp intuitively. Once again, the set of all possible behaviors needs to be considered if one is to understand the range within which the network performance evolves.

The paper demonstrates how the system dynamics paradigm can be generalized so as to cope with the stochastic behavior. First, the standard system dynamics model is extended, step by step, to a stochastic state-space

model. Then, a general Bayesian solution to the problem of estimating the states (and possible unknown parameters) in the model is presented. Its sequential Monte Carlo approximation, known as a weighted bootstrap or particle filter, is introduced and demonstrated on a simplified model of a service company growth. In addition, a Bayesian solution to the problem of testing multiple hypotheses about the model structure is outlined. Finally, the relationship between Bayesian inference and system dynamics is discussed and the benefits of the Bayesian viewpoint are summarized.

Modeling of Stochastic Behavior

The process of generalizing the system dynamics model can be split into several phases.

System Dynamics Model

It is practical to rewrite the traditional system dynamics model (Sterman, 2000) in the following compact form

$$\begin{aligned} \frac{\mathrm{d}x_t}{\mathrm{d}t} &= f(x_t), \\ y_t &= g(x_t), \end{aligned}$$

where x_t stands for a vector of state ("stock") variables and y_t denotes a vector of measurements, both at time t. The introduction of measurements y_t allows for modeling of state variables that cannot be observed directly (as is, e.g., the case of "soft" variables such as team morale or individual experience level). The vector functions $f(\cdot)$ and $g(\cdot)$ are considered known at this stage. Note that the bulk of the system dynamics modeling work is hidden here in specification of the flow vector $f(\cdot)$.

Introduction of Exogenous Input

While the system dynamics methodology typically prefers endogenous explanations of the observed behavior, there are cases where explicit consideration of exogenous inputs is beneficial or even necessary. This is, e.g., the case of a value network modeling where the performance of individual

nodes is combined together via exogenous inputs. Denoting the inputs as u_t, we have

$$\begin{aligned} \frac{\mathrm{d}x_t}{\mathrm{d}t} &= f(x_t, u_t), \quad (1) \\ y_t &= g(x_t, u_t). \quad (2) \end{aligned}$$

Treatment of Unknown Parameters

As a rule, the functions $f(\cdot)$ and $g(\cdot)$ are not known precisely. The uncertainty about their exact forms can be expressed through a vector of unknown parameters θ

$$\begin{aligned} \frac{\mathrm{d}x_t}{\mathrm{d}t} &= f(x_t, u_t, \theta), \\ y_t &= g(x_t, u_t, \theta). \end{aligned}$$

Since in Bayesian analysis the model parameters enter the calculations in the same way as the model states, we prefer treating the unknown parameters as additional state variables augmenting the original state vector. With some abuse of notation, from now on we use the symbol x_t for a vector combining the model states and model parameters. With this convention, the model (1)-(2) applies again. An extra advantage of this approach is that through $f(\cdot)$ we can explicitly model the changes of unknown parameters in time.

Discretization of Continuous-Time Dynamics

The differential equation can be discretized using the Taylor expansion (Heermann, 1990). Under suitable assumptions we can expand the state variable x_t in a Taylor series

$$x_{t+h} = x_t + \sum_{i=1}^{n-1} \frac{h^i}{i!} x_t^{(i)} + O(h^n).$$

For $n = 2$, we have

$$x_{t+h} = x_t + h \frac{\mathrm{d}x_t}{\mathrm{d}t} + O(h^2).$$

After dividing both sides by h and substituting for $\mathrm{d}x_t/\mathrm{d}t$ from (1), we get

$$h^{-1}(x_{t+h} - x_t) = f(x_t, u_t) + O(h).$$

Assuming that the variables x_t, y_t are sampled at regular time instants $t_1, t_2, \ldots$ with a constant sampling period

$$\tau_s = t_{k+1} - t_k,\ k = 1, 2, \ldots$$

and introducing the simplified notation

$$\begin{aligned} x_k &\stackrel{\triangle}{=} x_{t_k},\ k = 1, 2, \ldots \\ y_k &\stackrel{\triangle}{=} y_{t_k},\ k = 1, 2, \ldots \end{aligned}$$

we obtain from (1)–(2) the discretized model

$$\tau_s^{-1}(x_{k+1} - x_k) = f(x_k, u_k), \tag{3}$$

$$y_k = g(x_k, u_k). \tag{4}$$

Note that Eq. (3) can be rewritten as Euler integration

$$x_{k+1} = x_k + \tau_s f(x_k, u_k).$$

The form, although not the most compact one, preserves the meaning of $f(\cdot)$ as (a vector of) flows adding to or subtracting from the stock values at the previous time instant.

Stochastic Behavior

In a discrete-time model, the stochastic behavior can be modeled easily through introduction of vectors of random variables w_k and v_k into the functions $f(\cdot)$ and $g(\cdot)$, respectively. As a result, we obtain

$$\tau_s^{-1}(x_{k+1} - x_k) = f(x_k, u_k, w_k), \tag{5}$$

$$y_k = g(x_k, u_k, v_k). \tag{6}$$

The variables $\{w_k\}$ and $\{v_k\}$ are assumed mutually independent so that both the state x_{k+1} and the measurement y_k are statistically dependent on the state x_k and the input u_k only.

Probabilistic Formulation

The equations (5)–(6) are well suited for simulation of sample trajectories of the states x_k and the measurements y_k, but they do not provide a compact description of the overall uncertainty. For the latter, we need to introduce probability density functions of x_{k+1} and y_k (assuming both are

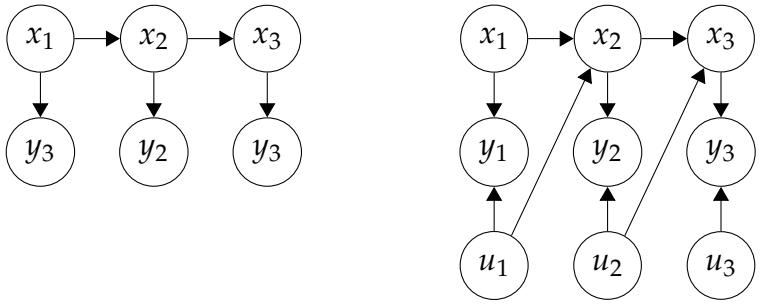

Figure 1: A graphical representation of the structure of dependence among the model variables. The left-hand diagram corresponds to a completely endogenous explanation of the observed behavior, the right-hand diagram introduces exogenous variables.

continuous) conditional on x_k and u_k, i.e.,

$$p(x_{k+1}|x_k, u_k), \tag{7}$$
$$p(y_k|x_k, u_k). \tag{8}$$

The formulation of a discrete-time state-space model in terms of conditional densities stresses the dependence structure of the model. This can be visualized graphically as shown in Fig. 1.

Bayesian Inference

With measurements of y_j and u_j available for $j = 1, \ldots, k$, the uncertainty of the state vector x_k and x_{k+1} can be expressed through their respective densities conditional on the sequences of measurements $y^k = (y_1, \ldots, y_k)$ and $u^k = (u_1, \ldots, u_k)$.

Recursive Formulae

The state estimation process is encapsulated in the sequential updating

$$p(x_k|y^{k-1}, u^{k-1}) \overset{\text{measurement update}}{\longrightarrow} p(x_k|y^k, u^k) \overset{\text{time update}}{\longrightarrow} p(x_{k+1}|y^k, u^k) \tag{9}$$

for $k = 1, 2, \ldots$ The recursion starts at $k = 1$ from the prior density $p(x_1|y^0, u^0) = p(x_1)$.

We present here the measurement and time update steps without proofs, which can be found, e.g., in Peterka (1981, Section 5).

MEASUREMENT UPDATE. Assuming that the generator of the input u_k does not use any other information about the state x_k than information contained in the past measurements y^k, u^k (cf. the natural conditions of control in Peterka, 1981), the posterior density $p(x_k|y^k, u^k)$ is related to the prior density $p(x_k|y^{k-1}, u^{k-1})$ through the Bayes rule

$$p(x_k|y^k, u^k) \propto p(y_k|x_k, u_k)\, p(x_k|y^{k-1}, u^{k-1}). \tag{10}$$

The symbol $\propto$ stands here for proportionality, i.e., equality up to a normalizing factor.

TIME UPDATE. The predictive density $p(x_{k+1}|y^k, u^k)$ follows from the posterior density $p(x_k|y^k, u^k)$ by elementary operations of probability calculus, which result in

$$p(x_{k+1}|y^k, u^k) = \int p(x_{k+1}|x_k, u_k)\, p(x_k|y^k, u^k)\, \mathrm{d}x_k. \tag{11}$$

Kalman Filter

Closed-form solutions to the recursion (10)–(11) exist only for a few special cases. One case of particular importance to the modeling theory and practice is Kalman filter.

MODEL. Assume that the data are modeled by a linear Gaussian state-space model

$$\begin{aligned} x_{k+1} &= A_k\, x_k + B_k\, u_k + w_k, \\ y_k &= C_k\, x_k + D_k\, u_k + e_k, \end{aligned}$$

with matrices A_k, B_k, C_k, D_k of appropriate dimensions and normally distributed random variables

$$\begin{aligned} w_k &\sim N(0, Q_k), \\ e_k &\sim N(0, R_k), \end{aligned}$$

for $k = 1, 2, \ldots$

Recursive estimation of the state vector in the above model is solved by the Kalman filter (Jazwinski, 1970; Anderson and Moore, 1979; Peterka, 1981).

DATA UPDATE. Assuming that the prior density $p(x_k|y^{k-1}, u^{k-1})$ is normal $N(\hat{x}_{k|k-1}, P_{k|k-1})$, the posterior density $p(x_k|y^k, u^k)$ is also normal $N(\hat{x}_{k|k}, P_{k|k})$. Its mean vector and covariance matrix are updated according to the formulae

$$\begin{aligned} \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k\,(y_k - \hat{y}_{k|k-1}), \\ P_{k|k} &= P_{k|k-1} - P_{k|k-1}\,C_k'\,S_k^{-1}\,C_k\,P_{k|k-1}, \end{aligned}$$

with the help of the auxiliary statistics

$$\begin{aligned} S_k &= C_k\,P_{k|k-1}\,C_k' + R_k, \\ K_k &= P_{k|k-1}\,C_k'\,S_k^{-1}, \\ \hat{y}_{k|k-1} &= C_k\,\hat{x}_{k|k-1} + D_k\,u_k. \end{aligned}$$

The recursion starts at $k = 1$ from a normal prior density $N(\hat{x}_{1|0}, P_{1|0})$ with the mean $\hat{x}_{1|0}$ and covariance $P_{1|0}$.

TIME UPDATE. Assuming that the posterior density $p(x_k|y^k, u^k)$ is normal $N(\hat{x}_{k|k}, P_{k|k})$, the predictive density $p(x_{k+1}|y^k, u^k)$ is also normal $N(\hat{x}_{k+1|k}, P_{k+1|k})$. Its mean vector and covariance matrix are updated according to the formulae

$$\begin{aligned} \hat{x}_{k+1|k} &= A_k\,\hat{x}_{k|k} + B_k\,u_k, \\ P_{k+1|k} &= Q_k + A_k\,P_{k|k}\,A_k'. \end{aligned}$$

The structure of a linear Gaussian model with a single state, single input and single measurement is outlined in Fig. 2. Note that in higher dimensions the graphical notation can quickly become overwhelming because of the *"Everything's related to everything else"* generality of the model. While Kalman filtering is critically dependent on the matrix calculus, system dynamics tends to proceed the other way around, by carefully selecting first which of many possible dependencies should be included in the model.

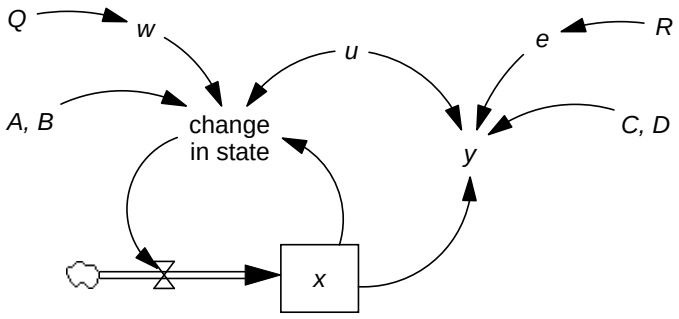

Figure 2: A system dynamics representation of linear Gaussian dynamic model with a single state, single input and single measurement.

Historical Remarks

While the use of probability for modeling of a stochastic behavior appears as a rather natural option, its application to quantification of the modeler's uncertainty (as to the most plausible model for a given purpose and evidence at hand) has met with mixed reactions. In fact, the conceptual and philosophical disputes between the proponents of frequentist and Bayesian schools of statistical thought occupied much of the last century.

The Bayesian approach uses probabilistic description *symmetrically* for both quantities we observe and quantities about which we wish to learn. Whether one proceeds from a model to a sample or from a sample to a model, the same (product and sum) rules of probability calculus apply. This leads to the fundamental principle of Bayesian inference: To form a judgment about the likely truth or falsity of any proposition, the correct procedure is to calculate the *probability* that the proposition is true *conditional on* all the evidence at hand (Jaynes, 2003, Chapter 4).

The name 'Bayesian' comes from the frequent use of Bayes' theorem in the inference process. What we know today as Bayes' theorem appeared first, in a very special context, in Bayes (1763). It was, however, Laplace (1774) who saw the result in generality and showed how to use it in real problems of hypothesis testing.

Much later, Keynes (1921), Jeffreys (1939) and in particular Cox (1946, 1961) demonstrated that probability theory is in fact the only theory of plausible reasoning under uncertainty that abides by logical consistency. Cox did so through a functional derivation of probability theory as the unique extension of Boolean algebra where probability represented a sub-

jective degree of plausibility.

Numerical Approximation

Since most system dynamics models include nonlinear feedback loops, closed-form solutions to the Bayesian estimation problem are rare. A suitable numerical approximation is required otherwise.

Monte Carlo Simulation

A particular appealing option for our purpose is Monte Carlo approximation of Bayesian inference, which is based upon essential duality between a sample $\{x^{(i)}, i = 1, \ldots, M\}$ and the density $p(x)$ from which the sample is generated (Smith and Gelfand, 1992): the density generates a sample, and from a sample we can approximately recreate the density.

In terms of samples, the Bayesian recursion (12) translates into sequential updating of a sample set

$$\{x_{k|k-1}^{(i)}\} \overset{\text{measurement update}}{\longrightarrow} \{x_{k|k}^{(i)}\} \overset{\text{time update}}{\longrightarrow} \{x_{k+1|k}^{(i)}\} \tag{12}$$

for $k = 1, 2, \ldots$. The recursion starts at $k = 1$ from the prior sample $\{x_{1|0}^{(i)}\}$.

The use of Monte Carlo simulation in Bayesian inference has been studied in various contexts since the 1940s. But, it was the application of sampling-importance resampling (Rubin, 1988) and weighted bootstrap (Smith and Gelfand, 1992) to state estimation of dynamic systems (Gordon *et al.*, 1993) that gave birth to a whole new family of recursive nonlinear filtering algorithms, known as *sequential Monte Carlo* (Liu and Chen, 1998) or *particle filtering* (Doucet *et al.*, 2001).

The approximate Bayesian recursion proceeds as follows.

Particle Filter Algorithm

1. *Initialization:* Draw M samples from the prior distribution

$$x_{1|0}^{(i)} \sim p(x_1),\ i = 1, \ldots, M$$

and set $k := 1$.

2. *Data Update:* Collect the data u_k, y_k. If the output y_k is measured at time k, then evaluate the importance weights

$$\pi_i = \frac{p(y_k|x_{k|k-1}^{(i)}, u_k)}{\sum_{j=1}^{M} p(y_k|x_{k|k-1}^{(j)}, u_k)},\ i = 1, \ldots, M \tag{13}$$

and draw M samples from a kernel approximation to the posterior distribution

$$x_{k|k}^{(i)} \sim \sum_{j=1}^{M} \pi_j K(x_k - x_{k|k-1}^{(j)}),\ i = 1, \ldots, M.$$

If the output y_k is not measured, set

$$x_{k|k}^{(i)} = x_{k|k-1}^{(i)},\ i = 1, \ldots, M.$$

3. *Time Update:* Draw M samples from a mixture approximation to the predictive distribution

$$x_{k+1|k}^{(i)} \sim \frac{1}{M} \sum_{j=1}^{M} p(x_{k+1}|x_{k|k}^{(j)}, u_k),\ i = 1, \ldots, M.$$

4. *Iteration:* Increment $k := k + 1$ and iterate from data update.

The function $K(x)$ stands for a suitable kernel function (Silverman, 1986; Hastie *et al.*, 2001), here normalized so as to integrate to one. The use of a kernel density estimate in the particle filter – known as *kernel smoothing* (Kulhavý and Ivanova, 1999), *regularized particle filter* (Musso *et al.*, 2001) or *smooth bootstrap* (Stavropoulos and Titterington, 2001) – prevents the degeneracy of a sample set.

Illustrative Example

It was shown that coupling of as few as two nonlinear feedback loops can yield a rather complex dynamic behavior. In particular, Forrester (1968, Section 2.5) considered sales growth and delivery delay modeled through self-reinforcing and balancing feedback loops, respectively, and interlinked through the actual orders made. We modify the original example as follows (cf. Figure 3).

NONLINEAR DYNAMICS. Consider a service company whose economic results depend on the performance of both its sales and service staff.

- The stocks of *Sales Capacity* and *Service Capacity* are measured in multiples of full-time equivalents (FTE) of an *average* sales or service person. This is to relate the labor capacity to the total performance of a team rather than the number of physical persons. Hence, hiring an additional person can increase the stock by more or less than one, depending on the actual person's productivity.
- The *Cost of Sales* and *Cost of Service* are defined by the cost of an average sales and service person, respectively. Combined with the previous point, the definition captures the fact that the team of highly knowledgeable and experienced persons cost more than an equally large but less productive team.
- The sales performance, *Sales Generation Rate*, is expressed in man-days of service work that one unit of *Sales Capacity* is capable to generate per unit of time.
- The actual *Signing Rate* is bounded by the limited *Service Demand*. As the sales capacity increases, the actual sales growth slows down so that it never exceeds the demand (see the upper left plot in Fig. 4).
- The *Revenue* generated through sales is partially used to cover the *Labor Cost*. The available funding is distributed among *Sales Capacity* and *Service Capacity*. The actual split between the two, expressed through *Sales Cost Ratio*, is governed by a company policy.
- The *Sales Capacity* and *Service Capacity* are adjusted dynamically towards the optimum capacities through negative feedback loops.
- The service *Delivery Rate* is determined by the available *Service Capacity* and the average *Workload* per one unit of *Service Capacity*.
- The *Service Backlog* is built up with *Signing Rate* and depleted with *Delivery Rate*.
- The ratio of *Service Backlog* and *Delivery Rate* yields *Delivery Delay*.
- The *Perceived Delay* as recognized by the client follows dynamically, through a first-order filter, the *Delivery Delay*.
- As the *Perceived Delay* increases, the market demand drops. The *Relative Drop in Demand* is a nonlinear function of the *Perceived Delay* (see the upper right plot in Fig. 4).

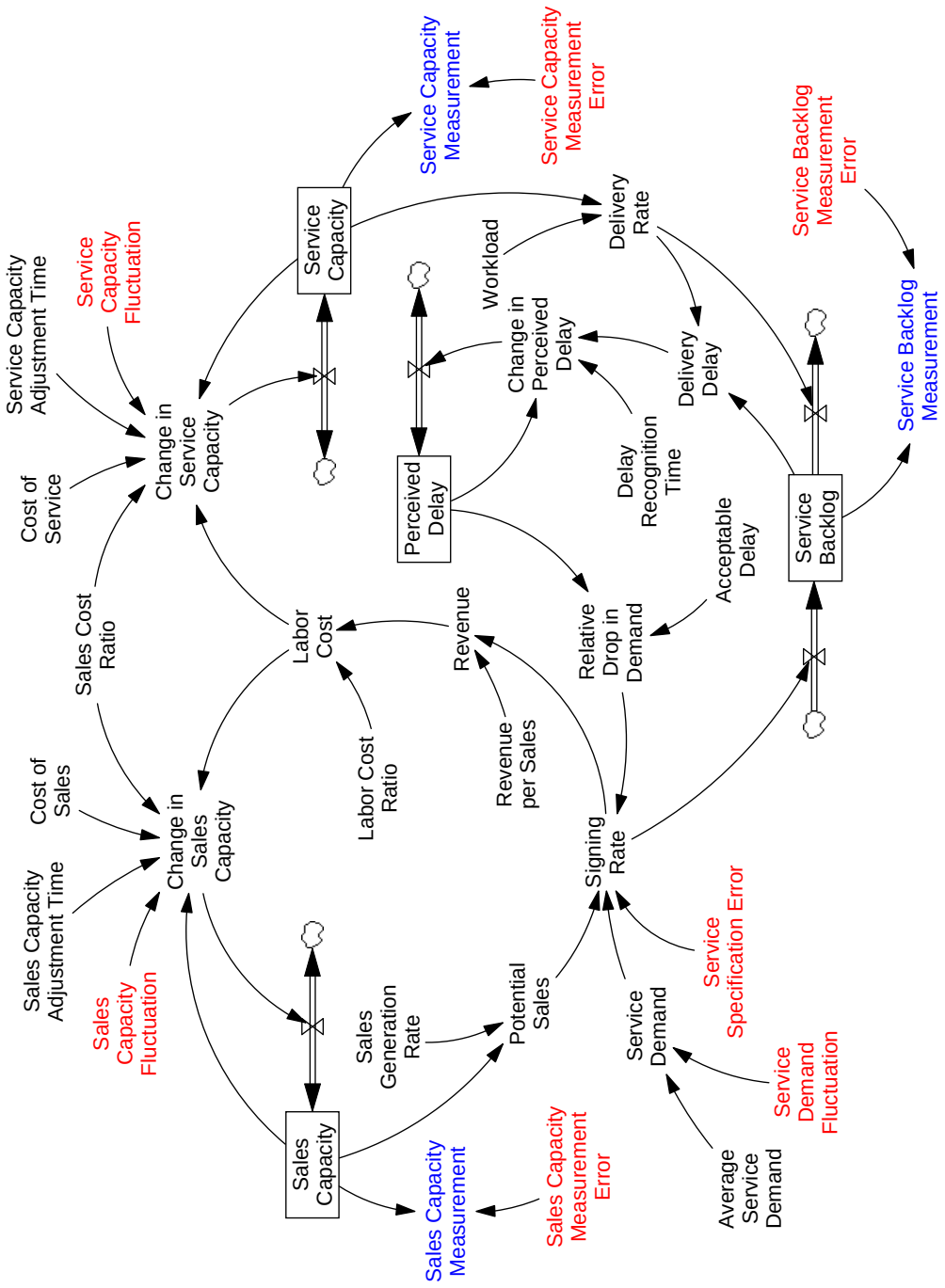

Figure 3: The structure of a service company dynamic performance model.

- In addition to the limited *Service Demand* and the nonlinear *Relative Drop in Demand,* the *Signing Rate* is affected by the *Service Specification Error*. This accounts for the uncertainty as to the actual service work that the service company commits to.

Figure 5 shows the prediction of the state variables for one particular combination of the model parameters. The sampling period is $\tau_s = 1$ week. The length of simulation is 100 weeks. The oscillatory behavior results from the interplay between the nonlinear feedbacks and dynamic delays.

STOCHASTIC BEHAVIOR. Since most variables in the model are positive by definition, we model the sources of random fluctuations via log-normal distribution $\mathrm{LN}(\mu, \sigma^2)$ with the density

$$p(z; \mu, \sigma) = \frac{1}{z\sigma\sqrt{2\pi}} \exp\Big(-\frac{1}{2\sigma^2}(\ln z - \mu)^2\Big).$$

In particular:

- *Service Demand* is modeled as log-normal with $\mu = \ln$ *Average Demand* and $\sigma =$ *Service Demand Fluctuation*.
- *Service Specification Error* is modeled as log-normal with $\mu = 0$ and appropriate standard deviation σ.
- Both *Sales Capacity* and *Service Capacity* are modeled in terms of logarithm of the state variable $\xi_{j,k} = \ln x_{j,k}$ (with index j pointing at the appropriate entries of the state vector) as

  $$\frac{\xi_{j,k+1} - \xi_{j,k}}{\tau_s} = \frac{\xi^*_j - \xi_{j,k}}{\tau_j} + \varepsilon_{j,k},\ \varepsilon_{j,k} \sim N(0, \sigma_j^2)$$

  where $\xi^*_{j,k} = \ln x^*_{j,k}$ is logarithm of a target value of the state variable and τ_j stands for the adjustment time constant. After taking the exponential function of both sides, we obtain the state equation in a multiplicative form, including multiplicative log-normal error

  $$x_{j,k+1} = x_{j,k} \left[\frac{x^*_{j,k}}{x_{j,k}}\right]^{\frac{\tau_s}{\tau_j}} \delta_{j,k},\ \delta_{j,k} \sim LN(0, \tau_s^2 \sigma_j^2). \tag{14}$$

Figure 6 shows the prediction of the state variables for the same combination of the model parameters as in Figure 5, except for a multiplicative

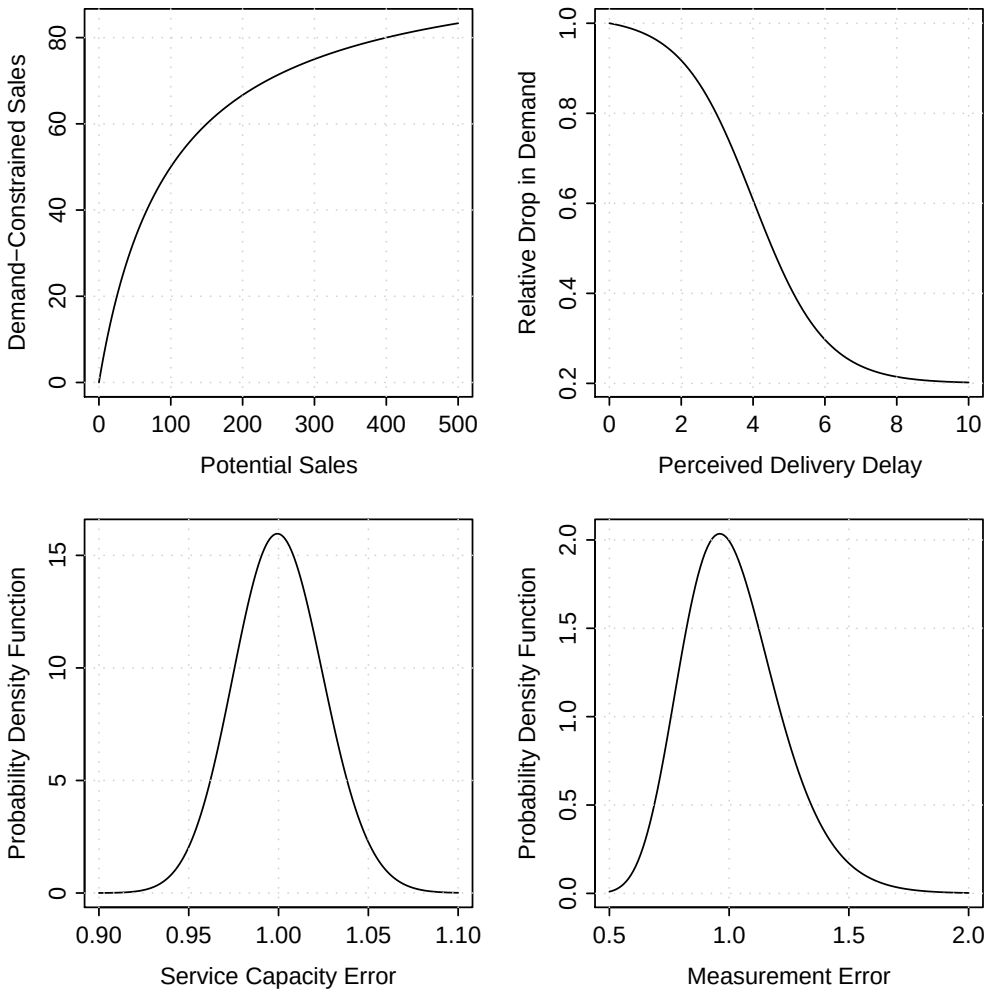

Figure 4: The dominant elements of the model. The upper left plot shows the effect of limited demand upon the generated sales. The upper right plot shows the relative drop in demand due to the perceived delivery delay. The lower left plot shows the density of multiplicative error affecting the service capacity. The lower right plot shows the density of multiplicative error for the state measurements.

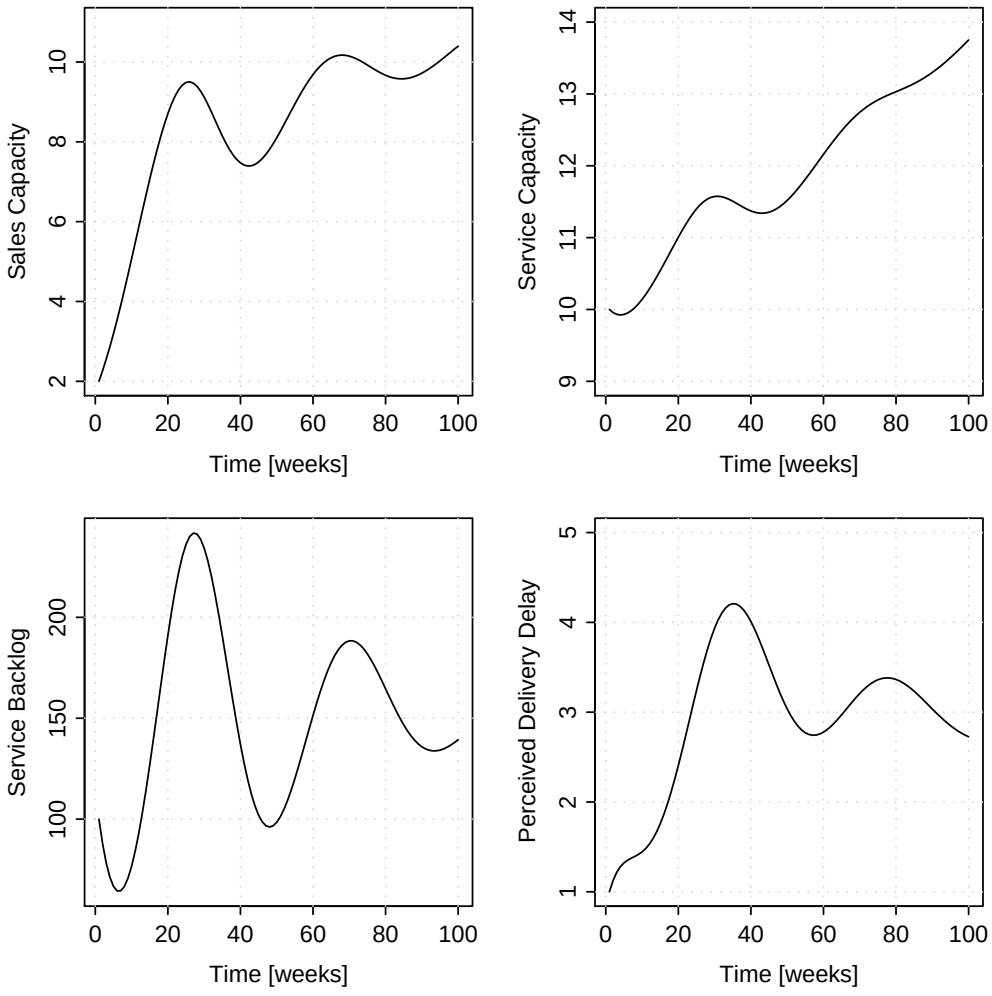

Figure 5: The prediction of state variables for a deterministic model, with no sources of random fluctuations present.

log-normal error ($\sigma = 0.025$) affecting *Service Capacity* through Eq. (14) (see the lower left plot in Fig. 4). The shades of gray indicate the sample quantiles 0%, 5%, 10%, ..., 95%, 100%; the black line marks the sample median. The particle filter uses 1,000 samples.

Notice how the relatively small error in the *Service Capacity* produces significant uncertainty in all state variables (and, of course, all dependent variables, incl. *Revenue*).

BAYESIAN INFERENCE. Finally, let us assume that measurements of *Sales Capacity*, *Service Capacity* and *Service Backlog* are available and modeled via multiplicative log-normal errors

$$y_{j,k} = x_{j,k}\ \delta_{j,k},\ \delta_{j,k} \sim LN(0, \sigma_j^2) \tag{15}$$

with $\sigma_j = 0.1$ (cf. Fig. 3 and the lower right plot in 4), for all j pointing at the corresponding entries of x and y vectors. The simulation of measurements proceeds in two steps – first, the state variables are simulated using the same model as considered for estimation, then the simulated state values are perturbed through Eq. (15).

The importance weights (13) in the particle filter follow from the product of log-normal densities

$$\pi_i \propto \prod_j \frac{1}{y_{j,k}\,\sigma_j}\ \exp\Big(-\frac{1}{2\sigma_j^2}(\ln y_{j,k} - \ln x^{(i)}_{j,k|k-1})^2\Big).$$

The combined estimation (until $t = 80$ weeks) and prediction (starting from $t = 81$ weeks) of the state variables is shown in Fig. 7.

The Bayesian analysis of the illustrative model has been implemented in the R programming language (a full code is available from the author upon request).

Model Comparison

In the preceding sections, we dealt with the problem of estimating state (and parameter) values within a single model with preassigned structure. We show now how Bayesian inference can be extended to the case of multiple model structures. Consistently with the Bayesian paradigm, we measure the plausibility of a model structure, labeled by a discrete index l, by

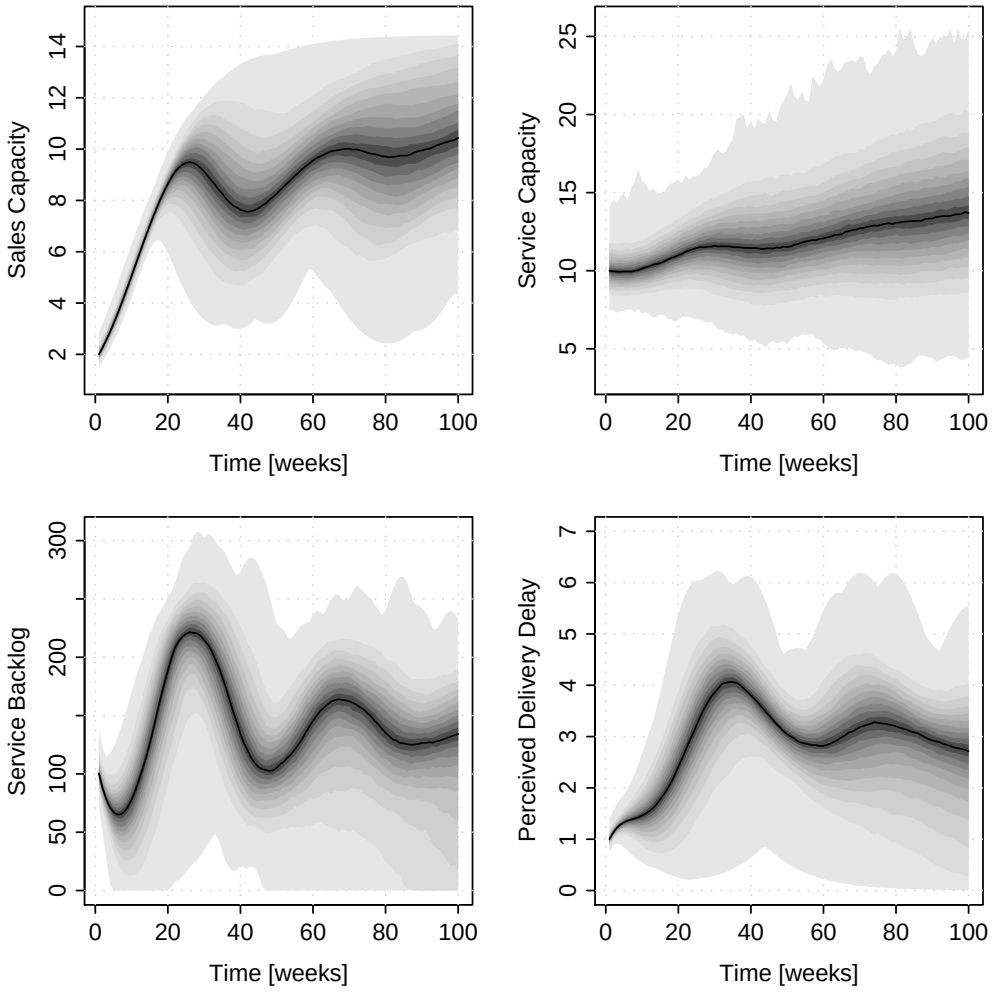

Figure 6: The prediction of state variables for a stochastic model, with a single source of random fluctuation represented by a log-normal error in the service capacity equation (14).

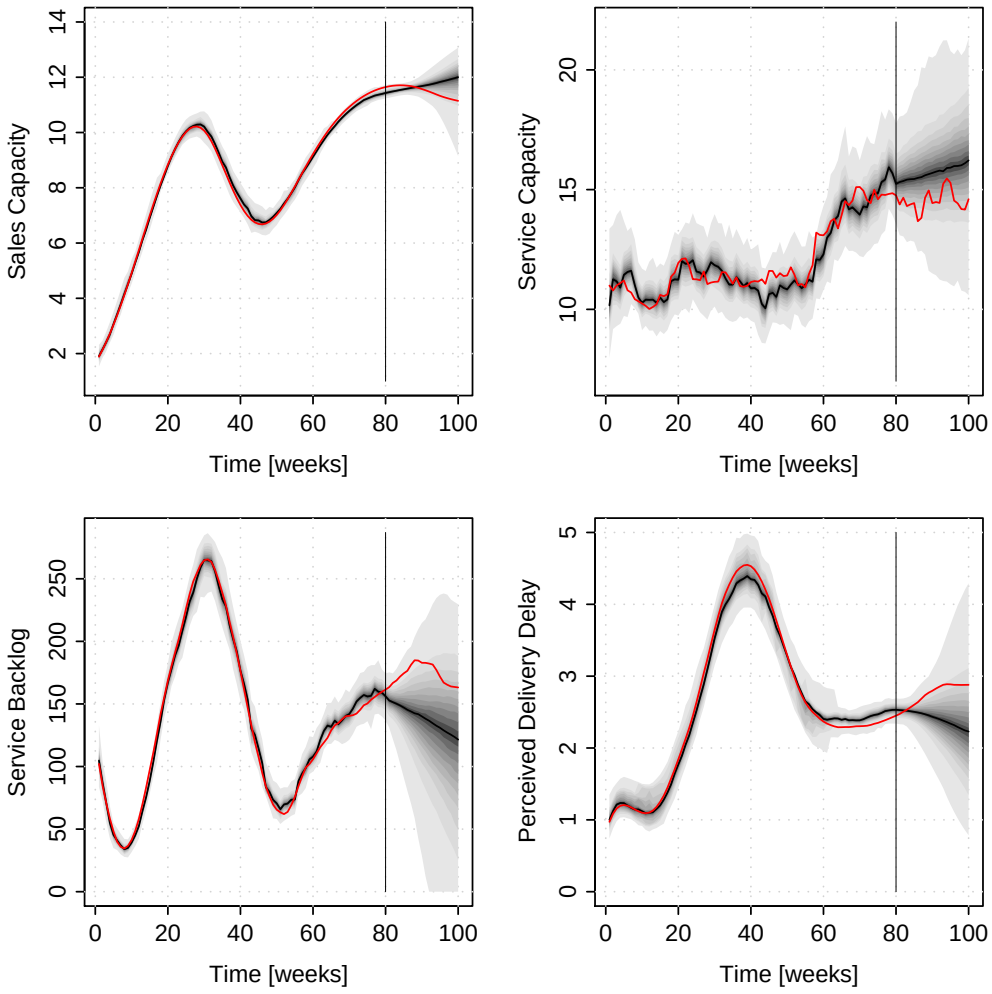

Figure 7: The results of Bayesian estimation of the state variables. The measurements of the underlying (unobserved) stocks are shown in red color. The measurements are considered available until time $t = 80$ weeks (indicated by a thin vertical line); after that, the distributions of the state variables are propagated through time updates only. Note the rapid increase of the state variable uncertainty as a result of the lack of measurements.

the probability of l conditional on the data sequences y^N and u^N. We assume that the observed data y_k and u_k are common to all model structures, whereas the state vectors $x_{k,l}$ can vary with the model structure l.

Assuming further that the generator of the input u_k does not use any other information about the model structure l and the state $x_{k,l}$ than information contained in the past measurements y^k, u^k (cf. again the natural conditions of control in Peterka, 1981), the posterior density $p(l|y^N, u^N)$ is related to the prior density $p(l)$ through the Bayes rule

$$p(l|y^N, u^N) \propto p(l) \prod_{k=1}^{N} p(y_k|y^{k-1}, u^k, l) \tag{16}$$

where the predictive density is defined as

$$p(y_k|y^{k-1}, u^k, l) = \int p(y_k|x_{k,l}, u_k, l)\, p(x_{k,l}|y^{k-1}, u^{k-1}, l)\, \mathrm{d}x_{k,l}. \tag{17}$$

The calculation of the predictive probabilities $p(y_k|y^{k-1}, u^k, l)$ for given data y^k, u^k can be approximated using Monte Carlo integration

$$p(y_k|y^{k-1}, u^k, l) \approx \frac{1}{M} \sum_{i=1}^{M} p(y_k|x_{k,l}^{(i)}, u_k, l) \tag{18}$$

with M samples drawn from the posterior density

$$x_{k,l}^{(1)}, \ldots, x_{k,l}^{(M)} \sim p(x_{k,l}|y^{k-1}, u^{k-1}, l)$$

using the particle filter algorithm described earlier.

Careful analysis of the formula (17) reveals that Bayesian inference has an inherent tendency to favor – among models of similar predictive power – those that are more parsimonious or "simpler" in terms of the dimension of a state (and parameter) space. This behavior, often referred to as Occam's razor, is brilliantly elucidated in Jaynes (2003, Chapter 20), see also MacKay (2003, Chapter 28). The "default" behavior of Bayesian model comparison can be easily overridden by assigning higher prior probability to a more complex model if prior information indicates that such model is more plausible for a given purpose.

System Dynamics vs. Bayesian Inference

System dynamics and Bayesian inference have much to offer to each other. On the one hand, system dynamics can equip Bayesian modeling with a practical methodology for converting *prior* information into a dynamic model structure. The reach of system dynamics is determined by its distinct features:

- All stocks, flows and parameters in a model are required to be *interpretable* in terms of a real system.
- A particular *problem* behavior is modeled rather than the underlying system.
- Data is *aggregated*, for a purpose; the microscopic behavior is of no or little interest.
- Any information that is considered *relevant* to the modeling problem at hand can in principle be incorporated (including process, business, equipment, and human factor information).
- A *closed-loop* system behavior, subject to implemented or contemplated policies, is modeled rather than an open-loop, 'to be controlled' system.

The attention that system dynamics pays to careful modeling of a structure generating the problem behavior distinguishes it from the empirical (or black-box) approaches based on statistical extrapolation from data. The focus that system dynamics puts on modeling of a particular behavior rather than the system itself makes it different from the theoretical (or mechanistic or first-principle or white-box) approaches. Being positioned between the two extremes, system dynamics has much in common with phenomenological modeling in natural and social sciences (Turchin, 2003) and grey-box modeling in engineering sciences (Bohlin, 2006).

In problems that match the above characteristics, system dynamics offers effective means for constructing highly informative priors by interrelating individual pieces of evidence. This is no small thing, compared to the general advice to "make use of all available information" that the practitioners typically get from the Bayesian statistics textbooks.

On the other hand, there are several ways how system dynamics can benefit from Bayesian inference. First, the Bayesian framework helps clarify the roles and relationships of individual ingredients of dynamic modeling.

Let us mention only two points here.

- By allowing for a random behavior in both state and measurement equations, one can study models of different complexity and level of detail within a *single* framework. One can consider even *highly simplified* models, capturing only the dominant dynamics of the target behavior. Obviously, such models make a poor "pointwise" fit to the actual behavior. This has stimulated in the past much discussion as to the methodological differences between system dynamics and statistics (or econometrics). Embedding system dynamics models within a broader class of *stochastic* models reconciles the apparent conflict. Simpler models naturally tend to exhibit more uncertainty in their states and/or measurements, but as long as this uncertainty is consistently accounted for, there is no reason to seek after a new inference mechanism.
- By introducing probability as a modeling language, one can give a precise meaning to *causality*. The term, often used rather vaguely, is substituted in a probabilistic setting with the concept of *conditional independence* (Dawid, 1979). A specific instance of it is the *Markov property* of a system state. As Jaynes (2003) stressed, physical causation is not an essential ingredient of Bayesian inference; what matters is the assumption of a *logical connection* between hypotheses and data. It is the model purpose that decides on what logical connections are important to consider. For instance, if the purpose of a model is to optimize a particular policy, it is crucial to capture the logical connection between the policy and the observable behavior of a system, but it is not necessary to include in the model the whole chain of microscopic causes and effects.

Second, Bayesian inference yields a coherent framework for consistently updating the *prior state of knowledge* about unknown or uncertain states and parameters of the model with *numerical evidence* at hand. As an information updating mechanism, the Bayesian scheme offers no definite answers, no absolute truths. It remains the modeler's responsibility to elicit prior information from available sources and to convert it to probabilities, represented in our problem setting by the densities $p(x_1)$, $p(x_{k+1}|x_k, u_k)$ and $p(y_k|x_k, u_k)$. Two modelers, even with the same evidence at hand, may end up with different models because of different prior information they have been able to extract from available resources or because of different interpretations of the same piece of information. Only the models that have been assigned positive prior probability can be eventually selected;

failure to include a plausible model among candidate models cannot be corrected with any amount of data.

Third, the Bayesian paradigm enables the modeler to capture and combine *uncertainty* in all its forms, including system state fluctuations, measurement errors, unknown model parameters and unknown model structure. The consistent treatment of uncertainty is essential if one is to prevent issues like overfitting or overly optimistic prediction. It is crucially important in making decisions as to the most appropriate model. The Bayesian literature distinguishes three major tasks in this respect:

- *Model comparison* selects the most plausible model out of candidate models.
- *Model assessment* checks that the model provides adequate fit to the data and behaves sensibly even under extreme conditions.
- *Model robustness analysis* tests the sensitivity of inference results to the modeling assumptions.

Much has been published on these topics. We have only briefly touched on the first task in the previous section and refer the interested reader to Carlin and Louis (2000, Chapter 6) and Gelman *et al.* (2004, Chapter 6) for more information and further insight.

Conclusion

Embedding the system dynamics modeling process within the Bayesian inference framework enables the modeler to formulate and solve – in a unified and consistent manner – the tasks of

- *modeling* – testing of hypotheses about the model structure,
- *estimation* of stock variables and model parameters from the available measurements,
- *prediction* of stock and flow variables beyond the period of available measurements.

All the tasks are solved by calculating the probability distribution of uncertain variables of interest conditional on available measurements.

The price paid for the extra power and comfort is twofold:

- a more complex model that has to account for all significant sources of uncertainty, be it the stochastic behavior of the underlying system or the uncertainty of the model structure and parameters;
- a more computationally intensive algorithm that replaces simulation of a single trajectory in the state space with simulation of a much larger set of trajectories approximating the probability distributions of interest.

In this paper, we have attempted to show that

- careful modeling of the stochastic behavior can be beneficial in its own right,
- the progress made in sequential Monte Carlo, coupled with the ever increasing performance of computing have made Bayesian inference an attractive option, especially for problems where quantification of the state uncertainty is a key requirement.

Acknowledgments

The author gratefully acknowledges research support from the Grant Agency of the Academy of Sciences of the Czech Republic through Grant No. IAA700750701. The author wishes to thank anonymous reviewers for their valuable comments and helpful suggestions.

References

Anderson BDO, Moore JB. 1979. *Optimal Filtering*. Prentice-Hall: Englewood Cliffs, NJ.

Bayes T. 1763. An essay toward solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society* : 370–418.

Bohlin T. 2006. *Practical Grey-box Process Identification: Theory and Applications*. Springer-Verlag: London.

Carlin BP, Louis TA. 2000. *Bayes and Empirical Bayes Methods for Data Analysis* (2nd edn.). Chapman & Hall/CRC: London.

Cox RT. 1946. Probability, frequency, and reasonable expectation. *American Journal of Physics* **14**(1–13).

Cox RT. 1961. *Algebra of Probable Inference*. Johns Hopkins University Press: Baltimore, MD.

Dawid AP. 1979. Conditional independence in statistical theory. *Journal of the Royal Statistical Society, Series B* **41**(1): 1–31.

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

Forrester JW. 1968. *Principles of Systems*. Wright-Allen Press: Cambridge, MA.

Forrester JW. 1985. Future development of the system dynamics paradigm. *Department memo D-3715*, System Dynamics Group, MIT, Cambridge, MA.

Gelman A, Carlin JB, Stern HS, Rubin DB. 2004. *Bayesian Data Analysis* (2nd edn.). Chapman & Hall/CRC: London.

Gordon NJ, Salmond DJ, Smith AFM. 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F* **140**(2): 107–113.

Graham AK. 2002. On positioning system dynamics as an applied science of strategy. In *Proceedings of the 20th International Conference of the System Dynamics Society*. Palermo, Italy.

Hastie T, Tibshirani R, Friedman J. 2001. *The Elements of Statistical Learning*. Springer-Verlag: New York.

Heermann DW. 1990. *Computer Simulation Methods in Theoretical Physics* (2nd edn.). Springer-Verlag: New York.

Jaynes ET. 2003. *Probability Theory: The Logic of Science*. Cambridge University Press.

Jazwinski AH. 1970. *Stochastic Processes and Filtering Theory*. Academic Press: New York.

Jeffreys H. 1939. *Theory of Probability*. Clarendon Press: Oxford.

Keynes JM. 1921. *A Treatise on Probability*. MacMillan: London.

Kulhavý R, Ivanova P. 1999. Quo vadis, Bayesian identification? *International Journal of Adaptive Control and Signal Processing* **13**(6): 469–485.

Laplace PS. 1774. Mémoire sur la probabilité des causes par les

événements. *Mémoires de l'Académie royale des Sciences de Paris* **6**: 621–656.

Liu JS, Chen R. 1998. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association* **93**(443): 1032–1044.

MacKay DJ. 2003. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press: Cambridge.

Musso C, Oudjane N, Gland FL. 2001. Improving regularised particle filters. In Doucet *et al.* (2001), chap. 12, pp. 247–272.

Peterka V. 1981. Bayesian approach to system identification. In *Trends and Progress in System Identification*, Eykhoff P (ed.). Pergamon Press: Elmsford, NY. chap. 8, pp. 239–304.

Rubin DB. 1988. Using the SIR algorithm to simulate posterior distributions. In *Bayesian Statistics 3*, Bernardo JM, DeGroot MH, Lindley DV, Smith AFM (eds.). Oxford University Press: Oxford. pp. 395–402.

Silverman BW. 1986. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall: London.

Smith AFM, Gelfand AE. 1992. Bayesian statistics without tears: a sampling-resampling perspective. *American Statistician* **46**(4): 84–88.

Stavropoulos P, Titterington DM. 2001. Improved particle filters and smoothing. In Doucet *et al.* (2001), chap. 14, pp. 295–318.

Sterman JD. 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. McGraw-Hill: New York, NY.

Turchin P. 2003. *Complex Population Dynamics: A Theoretical/Empirical Synthesis*. Princeton University Press: Princeton, NJ.