Recent Developments in Bayesian Inference with Applications in Hydrology

James O. Berger* and David Rios Insua**
*Department of Statistics
$PurdueUniversity, USA$
**Decision Analysis Group
Universidad Politecnica de Madrid, SPAIN
CNR-IAMI, Italy

Abstract

This paper describes some fairly new tools for Bayesian inference that are of considerable potential use in hydrology. These tools include Bayesian model selection, new computational techniques, and Bayesian approaches to time series and dynamic linear models. We also illustrate how these tools can be applied to problems in hydrology.

KEYWORDS: Bayesian Inference, Hydrology, Model Selection, Bayes Factors, Bayesian Computation, Markov Chain Monte Carlo, Dynamic Linear Models, Time Series.

1. Introduction

Throughout his career, Jacques Bernier has advocated the use of Bayesian ideas in hydrology, both in its scientific and managerial aspects. Indeed, because of his seminal work, Bayesian methods are starting to permeate hydrological sciences. This paper will describe some fairly recent tools for Bayesian inference which we feel can help to further advance this adoption of Bayesian methods. Our choice reflects in part our personal interests, but we chose them mainly because of their enormous potential in hydrology, their novelty, and their relationship to some of Bernier's interests. We also hope to contribute to the traditional Bayesian/non-Bayesian debate, which affects all sciences in general, and hydrology in particular. Other contributions to this debate can also be found in this volume, such as Munier and Parent (1996) and Duckstein (1996).

Our support for Bayesian ideas is both conceptual and practical : the Bayesian approach provides a coherent framework which facilitates the analysis of decision making problems under uncertainty, see Berger (1985) for a full development. Without entering into much detail, criticisms have centered mainly on three issues:

1. Computations. *Implementing the Bayesian framework leads to difficult computational problems. As a consequence, it is sometimes argued that it is necessary to limit attention to overly simplified models or to undertake a non-Bayesian analysis.* The recent development of Markov chain Monte Carlo methods, and other Bayesian computational machinery, has outdated this criticism, allowing for more realistic (and typically complex) modeling. This will be the topic of Section 3. Rasmussen, Bobée, and Bernier (1994) describes the role of simulation methods in complex Bayesian hydrological models.

2. Imprecision. *It is often argued that the the Bayesian framework demands excessive precision in the Decision Maker's judgements, particularly in regard to specification of the prior distribution.* This corresponds to a too narrow-minded perception of Bayesian ideas, and has led to alternative theories like fuzzy sets or Dempster and Shafer's theory. We remind the reader that the Bayesian framework is normative. When applied, it serves as a guidance for action under uncertainty. However, in early stages of an analysis we may not able to elicit precise information (prior, model, utility). To counter this fact, robust Bayesian methods have been developed. They essentially consist of undertaking a family of Bayesian analyses and basing conclusions on their common ground. If there are too many discrepancies, robust Bayesian tools suggest how to resolve them, guiding elicitation. We shall not pursue this important issue of robustness. The interested reader may follow the review by Berger (1994) and its discussion. Let us mention that robustness issues in hydrological science have been emphasised by Bernier (1991).

In regards to statistical inference, it should also be noted that there exists a well-developed version of Bayesian analysis which utilizes "default" or "noninformative" prior distributions, and hence which requires no more specifications than classical statistical methods. For reviews of this approach, see Berger (1985) and Kass and Wasserman (1995).

3. Descriptive. *It is sometimes argued that Actual Decision Makers do not conform to Bayesian postulates.* Stemming from work by Allais (1953), many experimental studies have pointed out that some decision makers violate the Bayesian postulates in unaided tasks. This suggests weaknesses of the Bayesian approach as a descriptive theory. Some authors interpret this, as well, as threatening its normative status, although such an interpretation carries no logical force. In any case, many new theories which attempt to improve upon Bayesian analysis from a descriptive point of view have appeared, see Rios Insua (1994) for a review.

Whereas experience concerning the descriptive inadequacy of the Bayesian approach has accumulated, recent comparisons among the new theories have painted a somewhat different picture. Note first that by abandoning axioms for descriptive purposes, we have to abandon them for normative purposes : some theories have been proposed too lightly, since they violate principles like transitivity or stochastic dominance. Camerer (1992) observes that all alternative theories so far proposed run into one kind or another of descriptive problems; Bernasconi (1992) observes that performers of experiments seem to conform progressively to Bayesian postulates, after repetitions of experiments; finally, Hey and Orme (1994) observe that, for many subjects, Expected Utility fits equally well as other theories from a statistical point of view, whereas economic implications are not that important for those for which fit is not so good. To sum up, risky decisions do not fully conform to Bayesian postulates, but they approximate them very reasonably.

Some recent descriptive theories are close to this idea. For example, Leland (1990) suggests that paradoxes may not be due to the preference structure but rather to cognitive constraints and experiential limitations, suggesting an approximate EU theory resolution to EU violations. A related issue is that of imprecision in judgements, as in Robust Bayesian Analysis. Experiments described in Rios et al (1994) suggest that paradoxical behavior in experiments may be explained via imprecision in judgments.

This completes our brief defense of Bayesian ideas. Other views may be seen in Edwards (1992). We turn now to the description of some recent useful Bayesian tools. Section 2 considers recent developments in Bayesian model selection. Section 3 considers the powerful new computational tools that have recently become available. Section 4 discusses the attractive and easily implementable tools for Bayesian analysis of time series and dynamic linear models.

2. Bayesian model selection

2.1. Notation

The data, $\underset{\sim}{X}$, is assumed to have arisen from one of several possible models $M_1, \ldots, M_m$. Under M_i, the density of $\underset{\sim}{X}$ is $f_i(\underset{\sim}{x}|\underset{\sim}{\theta}_i)$, where $\underset{\sim}{\theta}_i$ is an unknown vector of parameters of

f_i.

The Bayesian approach to model selection begins by assigning prior probabilities, $P(M_i)$, to each model; often, equal prior probabilities are used, i.e. $P(M_i) = 1/m$. It is also necessary to choose prior distributions $\pi(\underset{\sim}{\theta}_i)$ for the unknown parameters of each model; sometimes these can also be chosen in a "default" manner, as will be illustrated later.

The analysis then proceeds by computing the posterior probabilities of each model,

$$P(M_i|\underset{\sim}{x}) = \frac{P(M_i)m_i(\underset{\sim}{x})}{\sum_{j=1}^{m} P(M_j)m_j(\underset{\sim}{x})}, \tag{1}$$

where $m_j(\underset{\sim}{x}) = \int f_j(\underset{\sim}{x}|\underset{\sim}{\theta}_j)\pi_j(\underset{\sim}{\theta}_j)d\underset{\sim}{\theta}_j$. Typically one selects the model (or models) with largest posterior probability.

2.2. Advantages

Curiously, the Bayesian approach to model selection is less widely used than the Bayesian approach to estimation, even though the approach is arguably of even more value in model selection. The most obvious advantage of the Bayesian approach is the simplicity of interpretation of the answers; even those with limited statistical background can easily interpret the conclusion "the (posterior) probabilities that M_1 and M_2 are true are 0.93 and 0.07, respectively." This ease of interpretation is in stark contrast to the situation when classical measures such as P-values or Chi-square are used. Few understand what evidence is actually provided by such classical measures, and misinterpretation is the rule, rather than the exception (cf. Berger and Sellke, 1987; Berger and Delampady, 1987 and Delampady and Berger, 1990).

A second advantage of Bayesian model selection is it is *consistent*, in the sense that, as one obtains more and more data, one is guaranteed to select the true model (or the model closest to the true model if none are true). Classical methods typically fail even this minimal criterion, usually by selecting models that are too complex when there is a large amount of data.

This is related to a third advantage of the Bayesian approach, namely that it acts as an automatic "Ockham's razor," selecting a simpler model over a more complex model if both are compatible with the data. Indeed, Bayesian analysis can be used to quantify Ockham's razor, making precise what has long been viewed as a fundamental, but heuristic, scientific principle. See Jefferys and Berger (1992) for discussion and illustration. Bernier (1991) discusses this as the "parsimony principle."

A fourth advantage of Bayesian model selection is that one can account for model uncertainty. Since each model will have a posterior probability, $P(M_i|\underset{\sim}{x})$, one can maintain consideration of several models, with the input of each into the analysis weighted by the $P(M_i|\underset{\sim}{x})$. Classical analyses, which select one model and base predictions upon this one model, are notorious for providing predicted precisions that are much too small. See Draper (1995) for general discussion. This key point is extensively discussed in the hydrological literature by Bernier (1991, 1994b).

A fifth advantage of Bayesian model selection is that it can be applied to comparison of multiple models, and applies very generally; the models need not be in standard families, and need not be nested.

2.3. Default implementation

The two difficulties in implementing Bayesian model selection are (i) choosing the prior distributions $\pi_i(\theta_i)$, and (ii) computing the $m_i(x)$. A variety of strategies exist for carrying out the integrations necessary to compute the $m_i(x)$; see Kass and Raftery (1995) for discussion. Choosing the $\pi_i(\theta_i)$ is more of a problem.

It may well be the case that subjective knowledge about the θ_i is available, and can be incorporated into subjective proper priors for the θ_i. This is clearly desirable if it can be done. Often, however, the θ_i may be high dimensional and subjective elicitation of all the $\pi_i(\theta_i)$ may be impossible. There are then several possible "default" strategies one can follow.

The simplest default option is to use the approximation typically referred to as BIC (cf. Kass and Raftery, 1995). This is a quite accurate approximation if there is a substantial amount of data. Also, the approximation avoids the computational difficulty mentioned earlier.

For a moderate or small amount of data, BIC can be inaccurate. Sometimes (though not often!) it is possible to use "noninformative" or "objective" priors, $\pi_i^N(\theta_i)$. This can be done for some scenarios in which the dimensions of the vectors θ_i are the same for all models. Here is an example.

Example 1. Suppose we observed the following 30 flood periods : 23, 51, 87, 7, 120, 14, 62, 47, 225, 71, 246, 21, 42, 20, 5, 11, 4, 12, 120, 1, 3, 14, 71, 11, 16, 90, 1, 16, 52, 95. Assume these observations are independent, and consider two models for a datum, x_i.
M_1 : the lognormal model with $\theta_1 = (\mu, \sigma)$, and :

$$f_1(x_i|\mu,\sigma) = \frac{\exp\{(\log x_i - \mu)^2/(2\sigma^2)\}}{\sqrt{2\pi}\sigma x_i}; \tag{2}$$

M_2 : the Weibull model with $\theta_2 = (\gamma, \beta)$, and :

$$f_2(x_i|\gamma,\beta) = \beta x_i^{(\beta-1)}\gamma^{-\beta}\exp\{-(x_i/\gamma)^\beta\}. \tag{3}$$

We choose equal prior probabilities for the models, $P(M_1) = P(M_2) = 1/2$, and thus need only to choose the $\pi_i(\theta_i)$. The best "noninformative" priors for θ_1 and θ_2 are :

$$\pi_1(\mu,\sigma) = \frac{1}{\sigma} \quad \text{and} \quad \pi_2(\gamma,\beta) = \frac{1}{\gamma\beta}. \tag{4}$$

These are the so-called "reference priors" (see Bernardo, 1979, and Berger and Bernardo, 1992, for discussion). Since θ_1 and θ_2 have the same dimension (two), and since f_1 and f_2 can be shown to be transformed "location-scale" models, use of these noninformative priors for model selection is valid.

Using these priors, the $m_i(x)$ can be computed ($m_1(x)$ in closed form; $m_2(x)$ requiring one-dimensional numerical integration). The answers are $P(M_1|x) = 0.31$ and $P(M_2|x) = 0.69$. Thus the data favors the Weibull model by about 2:1.

Unfortunately, it is somewhat rare for use of noninformative priors in model selection to be valid. Recently, however, two very general default Bayesian model selection tools have been developed, the "Intrinsic Bayes factor" approach of Berger and Pericchi (1996), and the "Fractional Bayes factor" approach of O'Hagan (1994, 1995). These approaches are too involved to describe here, but it is noteworthy that they apply to almost any model selection problems, and operate without the need for subjective proper priors. Although we cannot present the algorithms here, it is of value to look at an application, to see the nature of the conclusions that arise.

Figure 1 here

Fig. 1 Time series data

Example 2. Figure 1 presents some time series data typical of a variety of hydrological applications. It is decided to model this as a stationary autoregressive process with drift. For instance, the AR(1) model with a linear drift would be described as :

$$Y_t = \beta_1 t + \phi_1(Y_{t-1} - \beta_1(t-1)) + \epsilon_t, \tag{5}$$

where Y_t is the observation at time t, β_1 is the unknown linear coefficient, ϕ_1 is the unknown autocorrelation, and the ϵ_t are i.i.d. $\mathcal{N}(0, \sigma^2)$ errors, σ^2 also unknown.

It is decided to consider autoregressive models of order 1, 2, 3 and 4, and also to consider constant (C), linear (L), and quadratic (Q) drift. Thus the AR(j) model with drift of polynomial order k ($k = 0, 1, 2$) can be written :

$$Y_t = \sum_{\ell=0}^{k} \beta_\ell t^\ell + \sum_{r=1}^{j} \phi_r (Y_{t-r} - \sum_{\ell=0}^{k} \beta_\ell (t-r)^\ell) + \epsilon_t. \tag{6}$$

We are thus considering twelve models (any of the four AR models together with any of the three polynomial drifts).

The "intrinsic Bayes factor" approach applies directly to this problem. It utilizes only standard noninformative priors for the parameters (constant priors for the β_i and ϕ_i, and $1/\sigma^2$ for the variance, σ^2). Note that, because the models being considered are of differing dimensions, one cannot use these noninformative priors directly, but must use them through the "intrinsic Bayes factor" algorithm. There is also a computational complication: because of the stationarity assumption, $\phi = (\phi_1, \phi_2, \ldots, \phi_j)$ is restricted to the "stationarity region," and so the integration in computation of the $m_i(x)$ must be carried out over this region. Methods of doing this, as well as the relevant intrinsic Bayes factor algorithm, can be found in Varshavsky (1996). The results are summarized in Table 1.

TABLE 1. Posterior probabilities of models assuming equal prior probabilities.

Model	$P(M_i \mid x)$	Model	$P(M_i \mid x)$
AR(1), C	~ 0	AR(3), C	0.740
AR(1), L	~ 0	AR(3), L	0.001
AR(1), Q	~ 0	AR(3), Q	0.001
AR(2), C	0.161	AR(4), C	0.076
AR(2), L	0.011	AR(4), L	0.006
AR(2), Q	0.001	AR(4), Q	0.001

There is clearly no support for a nonconstant drift. Among the models with constant drift, the AR(3) model is the clear winner, although the AR(2) model receives some support. It is of interest that classical model selection procedures choose substantially more complex models, such as the {AR(4), C} or even the {AR(4), Q} models. This is the "overfitting" of classical methods that was referred to earlier.

3. Advances in Bayesian computation

3.1. Introduction

Recent computational tools have allowed application of Bayesian methods to highly complex and nonstandard models. Indeed, for complicated models, Bayesian analysis has arguably now become the simplest (and often only possible) method of analysis.

Although other goals are possible, most Bayesian computation is focused on calculation of posterior expectations $E^*[g(\theta)]$, where E^* represents expectation with respect to the posterior distribution and $g(\theta)$ is some function of interest. For instance, if $g(\theta) = \theta$, then $E^*[g(\theta)] = E^*[\theta] \equiv \mu$, the posterior mean; if $g(\theta) = (\theta - \mu)^2$, then $E^*[g(\theta)]$ is the posterior variance of θ; and, if $g(\theta)$ is 1 if $\theta > C$ and 0 otherwise, then $E^*[g(\theta)]$ is the posterior probability that θ is greater than C.

3.2. Traditional numerical methods

The 'traditional' numerical methods for computing $E^*[g(\theta)]$ are numerical integration, Laplace approximation, and Monte Carlo Importance Sampling. Brief introductions to these methods can be found in Berger (1985). Here we say only a few words, to place the methods in context and provide references.

A successful general approach to **numerical integration** in Bayesian problems, using adaptive quadrature methods, was developed in Naylor and Smith (1982). This was very effective in moderate (e.g., 10) dimensional problems.

Extension of the **Laplace approximation** method of analytically approximating $E^*[g(\theta)]$, leading to a reasonably accurate general technique, was carried out in Tierney et al. (1989). The chief limitations of the method are the need for analytic derivatives, the need to redo parts of the analysis for each different $g(\theta)$, and the lack of an estimate of the error of the approximation. For many problems, however, the technique is remarkably successful.

Monte Carlo importance sampling [see Geweke (1989) and Wolpert (1991) for discussion] has been the most commonly used traditional method of computing $E^*[g(\theta)]$. The method can work in very large dimensions, and carries with it a fairly reliable accuracy measure. Although one of the oldest computational devices, it is still one of the best, being nearly 'optimal' in many problems. It does require determination of a good 'importance function', however, and this can be a difficult task. Current research continues to address the problem of choosing a good importance function; for instance, Oh and Berger (1993) developed a method of selecting an importance function for a multimodal posterior.

3.3. Markov chain simulation techniques

The newest techniques to be extensively utilized for numerical Bayesian computations are Markov chain simulation techniques, including the popular Gibbs Sampling. [Certain of these techniques are actually quite old — see, e.g., Hastings (1970); it is their application and adaption to Bayesian problems that is new.] A brief generic description of these methods is as follows:

Step 1. Select a 'suitable' Markov chain on Θ, with $p(\cdot,\cdot)$ being the transition probability density (i.e., $p(\theta,\theta^*)$ gives the transition density for movement of the chain from θ to θ^*). Here 'suitable' means primarily that the posterior distribution of θ given the data x, $\pi(\theta|x)$, is a stationary distribution of the Markov chain, which can be assured in a number of ways.

Step 2. Starting at a point $\theta^{(0)} \in \Theta$, generate a sequence of points $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(m)}$ from the chain.

Step 3. Then, for large m, $\theta^{(m)}$ is (approximately) distributed as $\pi(\theta|x)$ and :

$$\frac{1}{m}\sum_{i=1}^{m} g(\theta^{(i)}) \cong E^*[g(\theta)]. \tag{7}$$

The main strengths of Markov chain methods for computing $E^*[g(\theta)]$ are :

(1) Many different g can simultaneously be handled via Step 3, once the sequence $\theta^{(1)}, \ldots, \theta^{(m)}$ has been generated.

(2) Programming tends to be comparatively simple.

(3) Methods of assessing convergence and accuracy exist and/or are being developed.

The main weaknesses of the Markov chain methods are :

(1) They can be quite slow. It is not uncommon in complicated problems to need m to be in the hundreds of thousands, requiring millions of random variable generations if the dimension of θ is appreciable.

(2) One can be misled into prematurely judging that convergence has obtained.

The more common Markov chain methods, corresponding to different choices of $p(\cdot,\cdot)$, will briefly be discussed. A recent general guide to these methods, and their use in practice, is Gelman et. al. (1995). See also Smith (1991) and Besag et. al. (1995).

Metropolis-Hastings algorithm : One generates a new θ^* based on a 'probing' distribution, and then moves to the new θ^* or stays at the old θ according to a certain 'accept-reject' probabilities, see Hastings (1970).

Gibbs sampling : The Markov chain moves from $\theta^{(i)}$ to $\theta^{(i+1)}$ one coordinate at a time (or one group of coordinates at a time), the transition density being the conditional posterior density of the coordinate(s) being moved given the other coordinates. This is a particularly attractive procedure in many Bayesian scenarios, such as analysis of hierarchical models, because the conditional posterior density of one parameter given the others is often relatively simple (or can be made so with the introduction of auxiliary variables). Extensive discussion and illustration of Gibbs sampling can be found in Gelfand and Smith (1990), Gelman and Rubin (1992), Raftery (1992) and Smith and Gelfand (1992).

Example 3. The following posterior density is a somewhat simplified version of posterior densities which occur commonly in Bayesian analysis, and which are particularly amenable to Gibbs sampling. Suppose the posterior density is :

$$\pi(\theta_1, \theta_2|\text{data}) = \frac{1}{\pi}\exp\{-\theta_1(1+\theta_2^2)\} \tag{8}$$

on the domain $\theta_1 > 0$, $-\infty < \theta_2 < \infty$. Many posterior expectations cannot be done in closed form. Gibbs sampling, however, can easily be applied to this distribution to compute all integrals of interest.

Note, first, that the conditional distribution of θ_2, given θ_1, is Normal with mean zero and variance $1/2\theta_1$; and, given θ_2, θ_1 has an exponential distribution with mean $1/(1+\theta_2^2)$. Hence the Gibbs sampling algorithm can be given as follows :

Step 0. Choose an initial value for θ_2; for instance, the maximizer of the posterior, $\theta_2^{(0)} = 0$.

Step i(a). Generate $\theta_1^{(i)} = \mathcal{E}/(1+[\theta_2^{(i-1)}]^2$, where $\mathcal{E}$ is a standard exponential random variable.

Step i(b). Generate $\theta_2^{(i)} = Z/\sqrt{2\theta_1^{(i)}}$, where Z is a standard normal random variable.

Repeat Steps i(a) and i(b) for $i = 1, 2, \ldots, m$.

Final Step. Approximate the posterior expectation of $g(\theta_1, \theta_2)$ by :

$$E[g(\theta_1,\theta_2)] \quad = \quad \int_{-\infty}^{\infty}\int_{1}^{\infty} g(\theta_1,\theta_2)\pi(\theta_1,\theta_2|\text{data})d\theta_1 d\theta_2$$

$$\cong \frac{1}{m}\sum_{i=1}^{m} g(\theta_1^{(i)}, \theta_2^{(i)}). \tag{9}$$

For instance, the typical estimate of θ_1 would be its posterior mean, approximated by $\hat{\theta}_1 = (1/m)\sum_{i=1}^{m}\theta_1^{(i)}$. Table 2 presents the results of this computation for various values of m. Note that the true posterior mean here is 0.5.

TABLE 2. Approximate values of posterior mean of θ_1 from Gibbs Sampling

m	100	500	1,000	10,000	50,000
$\hat{\theta}_1$	0.43761	0.53243	0.48690	0.49857	0.50002

Hit and run sampling : The idea here is roughly that one moves from $\theta^{(i)}$ to $\theta^{(i+1)}$ by choosing a random direction and then moving in that direction according to the appropriate conditional posterior distribution. This method is particularly useful when Θ is a sharply constrained parameter space. Extensive discussion and illustration can be found in Belisle et al. (1993) and Chen and Schmeiser (1993).

Hybrid methods : Complex problems will typically require a mixture of the above (and other) methods. Here is an example, from Müller (1991), the purpose of which is to do Gibbs sampling when the posterior conditionals [e.g., $\pi(\theta_i|x$, other $\theta_k)$] are not 'nice'.

Step 1. Each step of the Markov chain will either :

- generate $\theta_j^{(i)}$ from $\pi(\theta_j|x$, other $\theta_k^{(i)})$ if the conditional posterior is 'nice' or
- generate $\theta_j^{(i)}$ by employing one or several steps of the Metropolis-Hastings algorithm if the conditional is not nice.

Step 2. For the probing function in the Metropolis-Hastings algorithm, use the relevant conditional distribution from a global multivariate normal (or t) importance function, as typically developed in Monte Carlo importance sampling.

Step 3. Adaptively update the importance function periodically, using estimated posterior means and covariance matrices.

Other discussions or instances of use of hybrid methods include Geyer (1992), Gilks and Wild (1992), Tanner (1991), Smith and Roberts (1993), Berger and Chen (1993) and Tierney (1994).

3.4. Software existence and development

Availability of general user-friendly Bayesian software would rapidly advance use of Bayesian methods. A number of software packages do exist, and are very useful for particular scenarios. An example is BATS (cf. Pole, West and Harrison, 1994 and West and Harrison, 1989), which is designed for Bayesian time series analysis. A listing and description of pre-1990 Bayesian software can be found in Goel (1988) and Press (1989).

Four recent software developments are BAIES, a Bayesian expert system (see Cowell, 1992); [B/D], an 'expectation based' subjective Bayesian system (see Wooff, 1992);

BUGS, designed to analyze general hierarchical models via Gibbs sampling (see Thomas et al., 1992); and XLISP-STAT, a general system with excellent interactive and graphics facilities, but limited computational power (see Tierney 1990).

Two of the major strengths of the Bayesian approach create certain difficulties in developing generic software. One is the extreme flexibility of Bayesian analysis, with virtually any constructed model being amenable to analysis. Classical packages need contend with only a few well-defined models or scenarios for which a classical procedure has been determined. Another strength of Bayesian analysis is the possibility of extensive utilization of subjective prior information, and Bayesians tend to feel that software should include an elaborate expert system for prior elicitation. This is hard, in part because much remains to be done empirically to determine optimal ways to elicit priors. Note that such an expert system is not, by any means, a strict need for Bayesian software; it is possible to base a system on use of noninformative priors.

4. Bayesian forecasting through dynamic linear models

In this section, we describe a class of forecasting models that may be very useful for practitioners in hydrological forecasting, say, of inflows to reservoirs. Dynamic Linear Models (DLMs) actually stem from work by Harrison and Stevens (1976). However, during the late eighties and early nineties, numerous modeling and computational enhancements and the development of the user-friendly software BATS (Pole et al., 1994) have made them readily available for applications. Here, we shall describe the main ideas of DLMs, and illustrate them with the forecast of inflows to Kariba Lake, used to manage the reservoir. For a thorough review of Bayesian forecasting, see West and Harrison (1989). West (1995) provides recent developments and applications, whereas specific applications to hydrology may be seen in Rios Insua and Salewicz (1995), Salewicz et al. (1996) and Muster and Bardossy (1995).

Apart from the usual benefits of Bayesian modeling, we see many advantages that make DLMs potentially useful for hydrologists. One that is especially important is that they allow for moving away from stationarity assumptions, since process parameters are time varying. This is important in hydrology, as Bernier (1994a,b) has been recently emphasizing. Also, they are flexible enough to model usual behavior of hydrological time series like seasonal patterns and trends, and permit the incorporation of covariates, such as rainfall for inflows. They are also computationally fast, facilitating their use in real time decision making and the large-scale simulations habitual in hydrology. Finally, they allow for the incorporation of all prior information, including that due to interventions, hence incorporating a principle of *management by exception*, fundamental in the Bayesian forecasting philosophy (West and Harrison, 1989) : a set of models is routinely used for processing information, making inferences and forecasting, unless exceptional circumstances arise. Examples, would include a sudden rainfall or a big release from a reservoir upstream. In this case, the system is open to external intervention, typically by inclusion of additional subjective information. Forecasting is performed sequentially based on all available information.

Our problem is to forecast the next r values of a variable y_t, from instant T to instant $T+r$, given the available information D_T. For that we use DLMs, which in their simplest case have the following structure, for every instant of time t, $t = 1, 2, 3, ...$:

- *Observation equation :*

$$y_t = F_t z_t + v_t, \quad v_t \sim N(0, V_t) \tag{10}$$

where y_t denotes the observed value, which depends linearly on the values of the state variables z_t, perturbed by a normal noise.

- *System evolution equation :*

$$z_t = G_t z_{t-1} + w_t, \quad w_t \sim N(0, W_t) \tag{11}$$

describing the evolution of the state variables, linearly dependent on the variables in the previous state plus a random perturbation.

- *Initial information :*

$$z_0 | D_0 \sim N(m_0, C_0) \tag{12}$$

describing the prior beliefs of the forecaster.
The error sequences v_t and w_t are independent, and mutually independent. Moreover, they are independent of $(\theta_0 | D_0)$.

Updating procedures and the use of this model for forecasting are described in detail by West and Harrison (1989). Essentially, inferences about both parameters and forecasts, one or more steps ahead, are based on a normal model, with corresponding parameters computed recursively.

The model specification requires that $F_t, G_t, V_t, W_t, m_0, C_0$ are known. The modeling of these is fully described in West and Harrison (1989). Concerning F_t and G_t, the key idea is the *Superposition principle*, which states that linear combinations of independent DLM's lead to a DLM. This suggests a model building strategy based on blocks, representing polynomial trends, seasonal patterns and dynamic regression, if covariates are available. One of us (Rios Insua) has been using this modeling strategy in our hydrological consulting work.

Concerning W_t, the use of the *discount principle* (West and Harrison, 1989) allows for a semiautomatic modeling approach to that variance, based on the idea of information discounting. Finally, a typical strategy concerning V_t is to consider it constant, but unknown, and introduce a procedure for learning adaptively about it.

Example 4. The example we consider illustrates in part these ideas. We are interested in forecasting the inflows to Lake Kariba, a hydropower reservoir in the Zambezi river, as part of a management system. Figure 2 represents part of the time series of monthly inflows to the Lake available, after logarithmic transformation.

Figure 2 here

Fig. 2 Monthly inflows to Kariba Lake in mln.m^3. Oct '30-Sep '65

Hence, if i_t designates the inflow to the lake, we shall forecast $y_t = \log i_t$, decomposing the series into a level and a seasonal part with annual cycle. For the level, we use a first order polynomial term. For the seasonal part, we use a Fourier decomposition of the pattern, see West and Harrison (1989). To improve short term forecasting, we add also a low coefficient first order autoregressive term. The precise model is as follows :

- *Observation equation*

$$y_t = (1,1,0,1)'(z_{t1}, z_{t2}, z_{t3}, z_{t4}) + v_t, \tag{13}$$

  where z_{t1} designates the level, (z_{t2}, z_{t3}) refer to the seasonal component, and z_{t4} to the autoregressive term. v_t is a normal observation error with constant but unknown variance V.

- *System equation*

$$\mathbf{z}_t = G\mathbf{z}_{t-1} + \mathbf{w_t}, \tag{14}$$

  with matrix G given by :

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\pi/6) & \sin(\pi/6) & 0 \\ 0 & -\sin(\pi/6) & \cos(\pi/6) & 0 \\ 0 & 0 & 0 & 0.4 \end{pmatrix} \tag{15}$$

  and observation error distributed as :

$$\mathbf{w_t} \sim N(\mathbf{0}, \Sigma_t) \tag{16}$$

  with Σ_t defined allowing for discount, with a discount factor δ_1 for the level and a discount factor δ_2 for the seasonal part. Note that the evolution of the seasonal part is defined in terms of periodic functions of period 12. They actually correspond to the first harmonic of the Fourier decomposition. If necessary, other harmonics may be introduced. The coefficient of the autoregressive part is 0.4.

- *Prior information*

$$\mathbf{z}_0|\phi \sim N(\mathbf{m}_0, VC^*) \tag{17}$$

$$\phi \sim Gamma(n_0/2, d_0/2) \tag{18}$$

  with $\phi = V^{-1}$. This provides a model for learning about the variance.

The assessment of the prior was done judgmentally, based on the beliefs of an expert, and sensitivity thoroughly checked. Figure 3 provides an indication of the forecasting performance of the model.

Figure 3 here

Fig. 3 Forecasts for log inflows to Kariba Lake. Dotted lines are uuper and lower limits of .95 predictive intervals. Dots represent actual log inflows

A key issue in Bayesian forecasting is that the output is the entire predictive distribution, not just summaries. Thus we can use this distribution for any purposes, taking expectations or simulating the future. As an aside, the standard method of : i) estimating model parameters; ii) plugging the estimates into the model; and iii) using the estimated model for prediction or simulation; typically greatly underestimates uncertainty in predictions, since the uncertainty in the model parameters is not taken into account.

To turn back to our illustration, we shall describe the actual use of this forecasting model for determining the optimal operating policy for Lake Kariba hydropower system, in terms of regulating the flow through the dam. Hence, the problem is to determine how much water to release through turbines and spillgates.

Assume that, at the beginning of a month, the reservoir operator makes the decision to release u_{1t} volume units of water for energy production and, additionally, u_{2t} units of volume to control the level of the reservoir. Priority is given to energy release: if there is enough water, commitments are fulfilled; if there is too much water, part of it is released or spilled. Then, given u_{1t}, u_{2t}, the control strategy proposed is as follows :

- If there is not enough water to release u_{1t}, all available water is released for energy production to satisfy the first objective of the reservoir operation. Otherwise, u_{1t} is released for energy production.

- If, after the release of u_{1t}, there is still water available, some water is additionally released to control reservoir storage level. If there is not enough water to release the volume u_{2t} defined above, all available water is released. Otherwise, u_{2t} is released. In the event that, after the two releases, the remaining water would exceed the maximum storage M, all excess water is spilled.

We need to determine the optimal controls.

¿From the operational and managerial viewpoint, the factors that characterize the consequences of a given operating policy at the end of every month are :

- the existence of energy deficit,

- the amount of water spilled,

- the value of the reservoir storage level at the end of the month.

These are easily computed, given the dynamics of the reservoir, the storage at the beginning of the month and the inflow.

Next, we assume that we can can specify the value of a storage level which secures 'satisfactory' operation of the reservoir over a long term time horizon. Such an assumption is reasonable, since 'traditional' methods of reservoir operation are based on the concept of *rule curves.* Consequently, each month one could maximize the expected value of a utility function which depends on the existence (or not) of deficit, the amount of water spilled and the deviation from a given 'ideal' (or reference) state x^*_{T+1}, i.e. :

$$\int f(u_{2,T+1}, k_{T+1}, \delta(s_{T+1}, x^*_{T+1}))dH(i_{T+1}|D_T) \tag{19}$$

where $\delta(s_{T+1}, x^*_{T+1})$ represents the deviation of the final state from the 'ideal' final state. Intuitively, if the ideal state is defined taking into account the dynamics of the system, we would not lose too much by following this approach.

For simplicity of calculations, and because assessments indicated that it was a good approximation, we used the utility function :

$$f(u_2, k, s) = \lambda f_1(k) + (1-\lambda)f_2(u_2) + \rho(s - x^*)^2. \tag{20}$$

Since k may attain only two values, and value 0 is better than 1, we may write

$$f_1(k) = 1 - k. \tag{21}$$

In order to assess f_2, expert information was used to estimate the risk aversion of the system's management. Assuming constant risk aversion, see Clemen (1991), to the amount of water spilled, one can take as utility function :

$$f_2(u_2) = a + b\exp(-cu_2), \tag{22}$$

with $b, c \geq 0$, f_2 being nonincreasing. An expert provided the information necessary to assess the values of parameters of the utility function, with standard assessment techniques. The following values of parameters were obtained : $\lambda = .75, \rho = -10^{-10}, b = 1.08365, a = -.07171, c = .0001415$.

Finally, the expected utility function had to be maximized with respect to control variables and subject to constraints on the controls (releases from the reservoir) : the amount of water released has to be nonnegative and the amount of water released for energy production is limited by the capacity of the turbines. The optimization problem is thus given as :

$$\begin{array}{rl} & \max \quad \Psi(u) \\ s.t. & 0 \leq u_1 \leq m \\ & 0 \leq u_2 \end{array} \tag{23}$$

where $\Psi(u)$ is the expected utility.

The analysis of the results suggest the possibility of operating the reservoir at much more efficient and secure levels than is currently done. A full description of the study may be seen in Rios Insua and Salewicz (1995).

Acknowledgments This work was financed by grant DMS–9303556 from the National Science Foundation and by a grant from the Iberdrola Foundation. We are grateful to discussions with Andrzej Salewicz and Concha Bielza.

Bibliography

Allais, M. (1953). 'Le comportement de l'homme rational devant le risque : critique des postulats et axiomes de l'école Américaine'. Econometrica, 21, p.503–546.

Belisle, C., Romeijn, H. E. and Smith, R. (1993). 'Hit-and-run algorithms for generating multivariate distributions'. Mathematics of Operation Research, 18, p.255–266.

Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis* (2nd edition). Springer-Verlag, NY.

Berger, J. (1994). 'An overview of robust Bayesian analysis'. Test, 3, p. 5–124.

Berger, J. and Bernardo, J. (1992). 'On the development of the reference prior method'. In J. Bernardo, J. Berger, A. Dawid and A. F. M. Smith (editors), *Bayesian Statistics 4*, Oxford University Press, London.

Berger, J. and Chen, M. H. (1993). 'Determining retirement patterns: prediction for a multinomial distribution with constrained parameter space'. The Statistician, 42, p. 427–443.

Berger, J. and Delampady, M. (1987). 'Testing precise hypotheses (with discussion)'. Statist. Science, 2, p. 317–352.

Berger, J. and Pericchi, L. (1996). 'The intrinsic Bayes factor for model selection and prediction'. J. Amer. Statist. Assoc., 91, p. 109–122.

Berger, J. and Sellke, T. (1987). 'Testing a point null hypothesis: the irreconcilability of P values and evidence'. J. Amer. Statist. Assoc., 82, p. 112–122.

Bernardo, J. (1979). 'Reference prior distributions for Bayesian inferenc'e. J. Roy. Statist. Soc. B, 41, p. 113–147.

Bernasconi, M. (1992). 'Different frames for the independence axiom: an experimental investigation in individual decision making under risk'. J. Risk and Uncert., 5, p. 159–174.

Bernier, J. (1991). 'Bayesian analysis of robustness of models in water and environmental sciences '. NATO ASI on Risk and Reliability in Water Resources and Environmental Engineering, Porto Karras, Greece, J. Ganoulis (Ed.), Springer-Verlag Berlin Heidelberg, vol. G29, p. 203–229.

Bernier, J. (1994a). 'Statistical detection of changes in geophysical series'. In *Engineering risk and reliability in a changing physical environment.* L. Duckstein and É. Parent (Eds.), Kluwer Academic Publishers, the Netherlands. NATO ASI Series E: Applied

Sciences, vol. 275, p. 159–176.
Bernier, J. (1994b). 'Quantitative analysis of uncertainties in water resources'. In *Engineering risk in natural resources management with special references to hydrosystems under changes of physical or climatic environment.* L. Duckstein and É. Parent (Eds.), Kluwer Academic Publishers, the Netherlands. NATO ASI Series E: Applied Sciences, vol. 275, p. 343–357.
Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995). 'Bayesian computation and stochastic systems'. Statistical Science, 10, 1-58
Camerer, C. (1992). 'Recent tests of generalizations of Expected Utility Theory', in Edwards (ed) *Utility Theories: Measurement and Applications*, Kluwer
Chen, M. H. and Schmeiser, B. (1993). 'Performance of the Gibbs, hit-and-run, and Metropolis samplers'. Journal of Computational and Graphical Statistics, 2, p. 1–22.
Delampady, M. and Berger, J. (1990). 'Lower bounds on posterior probabilities for multinomial and chi-squared tests'. Annals of Statistics, 18, p. 1295–1316.
Draper, D. (1995). 'Assessment and propogation of model uncertainty'. J. Roy. Statist. Soc. B, 57, p. 45–98.
Clemen, R. (1991). *Making Hard Decisions.* Wadsworth: New York.
Cowell, R. G. (1992). 'BAIES: A probabilistic expert system shell with qualitative and quantitative learning'. In: *Bayesian Statistics 4* (J. Bernardo, J. Berger, A. Dawid and A. F. M. Smith, Eds.). Oxford University Press, Oxford.
Duckstein, L. (1996). 'Bayes and fuzzy logic modeling of engineering risk under dynamic change'. In this volume.
Edwards, W. (1992). *Utility Theories: Measurement and Applications*, Kluwer.
Gelfand, A. E. and A. F. M. Smith (1990). 'Sampling based approaches to calculating marginal densities'. J. Amer. Statist. Assoc., 85, p. 398–409.
Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian Data Analysis.* Chapman and Hall, London.
Gelman, A. and Rubin, D. B. (1992). 'On the routine use of Markov Chains for simulation'. In J. Bernardo, J. Berger, A. Dawid, and A. F. M. Smith (editors), *Bayesian Statistics 4*, Oxford University Press, London.
Geweke, J. (1989). 'Bayesian inference in econometrics models using Monte Carlo integration'. Econometrica, 57, p. 1317–1340.
Geyer, C. (1992). 'Practical Markov chain Monte Carlo'. Statistical Science, 7, p. 473–483.
Gilks, W. R. and P. Wild (1992). 'Adaptive rejection sampling for Gibbs sampling'. In J. Bernardo, J. Berger, A. Dawid, and A. F. M. Smith (editors), *Bayesian Statistics 4*, Oxford University Press, London.
Goel, P. (1988). 'Software for Bayesian analysis: current status and additional needs'. In: *Bayesian Statistics 3*, J. M. Bernardo, M. DeGroot, D. Lindley and A. Smith, (Eds.). Oxford University Press, Oxford.
Harrison, P. J. and Stevens, C. F. (1976). 'Bayesian forecasting', J. Roy. Statist. Soc. B, 38, p. 205–247.

Hastings, W. K. (1970). 'Monte-Carlo sampling methods using Markov chains and their applications', Biometrika, 57, p. 97–109.

Hey, J. and Orme, C. (1994). Circles, triangles and straight lines. To appear in *Econometrica.*

Jefferys, W. and Berger, J. (1992). 'Ockham's razor and Bayesian analysis', American Scientist, 80, p. 64–72.

Kass, R. and Raftery, A. (1995). 'Bayes factors and model uncertainty'. J. Amer. Statist. Assoc., 90, p. 773–795.

Kass, R. and Wasserman, L. (1995). 'The selection of prior distributions by formal rules'. To appear in *J. Amer. Statist. Assoc.*

Leland, J. (1990). Indiscrimination in evaluation: An approximate expected utility theory resolution to expected utility violations, *presented at FUR-V, Duke University.*

Müller, P.(1991). 'A generic approach to posterior integration and Gibbs sampling'. Technical Report 91-09, Department of Statistics, Purdue University.

Munier, B. and Parent, E. (1996). 'Le développement récent des sciences de la décision: un regard critique sur la statistique décisionnelle Bayesienne'. In this volume.

Muster, H. and Bardossy, A. (1996). 'Precipitation forecasts for flood management in river basins'. In this volume.

Naylor, J. and Smith, A. F. M. (1982). 'Application of a method for the efficient computation of posterior distributions'. Appl. Statist., 31, p. 214–225.

O'Hagan, A. (1994). *Bayesian Inference*, Edward Arnold, London.

O'Hagan, A. (1995). 'Fractional Bayes factors for model comparisons'. J. Roy. Statist. Soc. B, 57, p. 99–138.

Oh, M. S. and Berger, J. (1993). 'Integration of multimodal functions by Monte Carlo importance sampling'. J. Am. Statist. Assoc., 88, p. 450–456.

Pole, A., West, M., and Harrison, J. (1994). *Applied Bayesian Forecasting.* Chapman and Hall: London.

Press, J. (1989). *Bayesian Statistics.* Wiley, New York.

Raftery, A. (1992). 'How many iterations in the Gibbs sampler?' In J. Bernardo, J. Berger, A. P. Dawid, and A. F. M. Smith (editors), *Bayesian Statistics 4*, Oxford University Press.

Rasmussen, P.F., Bobée, B. and Bernier, J. (1994). 'Une méthodologie générale de comparaison de modèles d'estimation règionale de crue'. Revue des Sciences de l'Eau, 7, p. 23–41.

Rios Insua, D. (1994). 'Ambiguity, imprecision and sensitivity in Decision Theory', in Puri and Vilaplana (eds) *New Progress in Probability and Statistics*, SVP.

Rios Insua, D. and Salewicz, A. (1995). 'The operation of Lake Kariba', *J. Multicriteria Decision Analysis.*

Rios, S., Rios-Insua, S., Rios Insua, D. and Pachon, J. (1994). 'Experiments in robust decision making', in Rios (ed) *Decision Theory and Decision Analysis: Trends and Challenges*, Kluwer.

Salewicz, K., Rios Insua, D., Mueller, P., and Bielza, C. (1996). 'Bayesian methods in reservoir operations'. To appear in French and Smith (eds.). *Case Studies in*

Bayesian Analysis, Arnold.

Smith, A. (1991). 'Bayesian computational methods'. Phil. Trans. Roy. Soc., 337, p. 369-386.

Smith, A. F. M. and Gelfand, A. E. (1992). 'Bayesian statistics without tears: a sampling-resampling perspective'. American Statistician, 46, p. 84–88.

Smith, A. F. M. and Roberts, G. O. (1993). 'Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods'. J. Roy. Statist. Soc. B, 55, 3–23.

Tanner, M. A. (1991). *Tools for Statistical Inference: Observed Data and Data Augmentation Methods*, Lecture Notes in Statistics **67**, Springer Verlag, New York.

Thomas, A., Spiegelhalter, D. J. and Gilks, W. (1992). 'BUGS: A program to perform Bayesian inference using Gibbs sampling'. In: *Bayesian Statistics* **4** (J. Bernardo, J. Berger, A. Dawid and A. F. M. Smith, Eds.). Oxford University Press, Oxford.

Tierney, L. (1994). 'Markov chains for exploring posterior distributions'. Ann. Statist., 22, p. 1701-1762.

Tierney, L. (1990). *Lisp-Stat, an Object-Oriented Environment for Statistical Computing and Dynamic Graphics.* Wiley, New York.

Tierney, L., Kass, R. and Kadane, J. (1989). 'Fully exponential Laplace approximations to expectations and variances of non-positive functions'. J. Am. Statist. Assoc., 84, p. 710–716.

Varshavsky, J. (1996). Intrinsic Bayes factors for model selection with autoregressive data. in J. Bernardo et. al. (editors), *Bayesian Statistics 5*, Oxford University Press, London.

West, M., Harrison, J. (1989). *Bayesian Forecasting and Dynamic Models.* Berlin: Springer.

West, M. (1995). 'Bayesian forecasting'. Discussion paper, ISDS, Duke University.

Wolpert, R. L. (1991). 'Monte Carlo importance sampling in Bayesian statistics'. In: *Statistical Multiple Integration* (N. Flournoy and R. Tsutakawa, Eds.). *Contemporary Mathematics*, Vol. 115.

Wooff, D. A. (1992). '[B/D] works'. In: *Bayesian Statistics* **4** (J. Bernardo, J. Berger, A. Dawid and A. F. M. Smith, Eds.). Oxford University Press, Oxford.